The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Physics Projection: Intelligence with Physical World

Naoto Iwahashi*1 Hideaki Negoro*1 Soichi Kawano*2

*1Okayama Prefectural Univ *2Luke System

This paper presents a new approach named *physics projection*, through which robots can learn the physical world and predict the effects of their actions actively and online. Physics projection consists of three components: a robot, physical world model, and physics engine. The process of physics projection has a double loop structure comprising (1) a learning loop of the physical world model and (2) a simulation search loop. Experiments were performed using the TurtleBot3 mobile robot and Unity graphic engine. The results clearly showed that the robot predicted the effects of its various actions under the given physical conditions and successfully executed the tasks of carrying a wine glass without dropping it and a cup filled with water without spilling. The robot could predict a catastrophic effect that could not be predicted by a human operator.

1. Introduction

Intelligence arises from the coupling of the dynamics of internal (mental) states and external (world) states. The nature, mechanisms, and functions of such coupling have long been discussed in the fields of philosophy, science, and technology [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. Not only low-level sensory-motor activities but high-level cognitive activities, such as language and communication, are also established in the coupling.

We have been studying the coupling of internal states of human and robot in linguistic and physical communication, which reflect the external states, where mutual beliefs play an important role [15, 16, 17, 18]. Mutual beliefs can be considered as the communication state that is used to generate and understand linguistic (utterances) and physical actions. By focusing on such characteristics, we constructed a computational model of human–robot linguistic and physical communications (Figure 1). The model has a double loop structure containing (1) a mutual-belief learning loop (outer red-line loop) and (2) a simulation search loop (inner blue-line loop). The mutual-belief learning loop is executed when the robot is acting and making observations both linguistically and physically during communication, and a mutual belief model is learned actively and online. The simulation search loop is executed as a process in the mutual-belief learning loop to search for an appropriate action by evaluating numerous candidates by using a probabilistic-inference-based simulation.

Linguistic and physical communication is based on various constraints such as physical, sensory-motor, psychological, conceptual, and experiential constraints. The human–robot linguistic-and-physical-communication model makes it possible to incorporate these constraints into the mutual-belief model and use them in the simulation search. Among all the constraint types, physical constraints might be an important basis. However, the aforementioned human–robot model cannot apply physical constraints, such as gravity, collision, and stability, directly. Nevertheless, it can statistically learn the concepts of the physical relationship between objects and use it in the simulation search. Note that no previous high-order cognitive systems can incorporate physical constraints directly [19]. In addition,

Contact: Naoto Iwahashi, iwahashi@c.oka-pu.ac.jp, www-ail.c.oka-pu.ac.jp

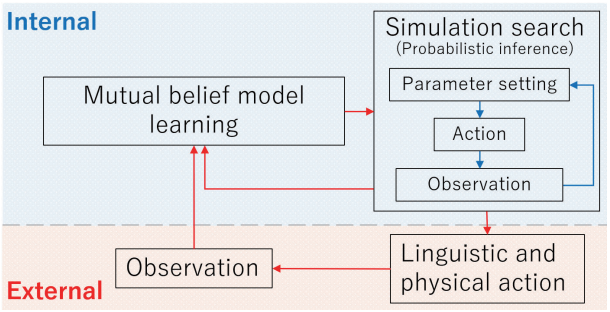

Figure 1: Linguistic and physical communication model with mutual beliefs

even without the connection of physical constraints with high-order cognitive activities, it is extremely difficult for state-of-the-art robots to predict the effects of their actions on the physical world. Therefore, robots cannot safely carry wine glasses on top of a tray or place unknown objects on a table without knocking it down.

This paper presents a new approach named *physics projection* that enables robots to learn the physical world model and predict the effects of their actions actively and online. Many previous works are closely related to this approach. The prediction of the effects of actions by physics engines was intensively studied in the fields of virtual reality, augmented reality, robotics, machine learning, and cognitive science [20, 21, 22, 23]. The comparison and integration of learned and analytic physical models were discussed in [24, 25]. The recent study of automatic 3D modeling [26] is also in this line with the current research, and could be applied to these research fields. Compared with these studies, physics projection is a novel idea, and makes the following contributions:

1. It comprises active and online loops of the learning of the physical world model and simulation search.
2. The proposed method can be integrated with high-level cognitive systems.

2. Physics Projection

2.1 Incorporating the physical world

Physics projection is a new approach for incorporating the physical world into artificial intelligence; It operates

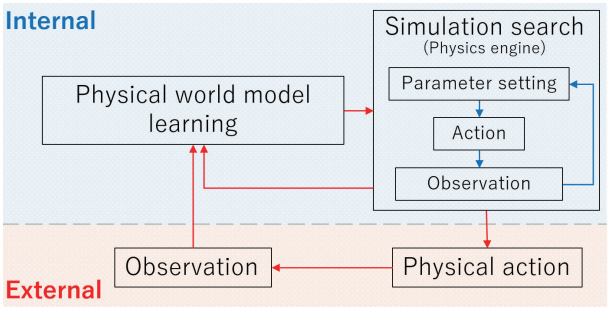

Figure 2: Physics Projection

the internal (mental) model of the external (physical) world to understand the physical world. The approach consists of three components: a robot, physical world model, and physics engine (Figure 2). The process of physics projection has a double loop structure comprising (1) a learning loop of the physical world model (outer red-line loop) and (2) a simulation search loop (inner blue-line loop). The learning loop of the physical world model is executed when the robot is performing actions and making observations in the physical world; thus, the physical world model is learned actively and onlie. The simulation search loop is executed as a process in the learning loop of the physical world model to evaluate numerous candidates and search for an appropriate action. Eventually, the loop structures of physics projection and above-mentioned human–robot liguistic-and -physical-communication model are the same, although they operate at different spatiotemporal scales and handle different conceptual granularity. Therefore, the physical constraints of physics projection can be easily incorporated into the human–robot linguistic-and-physical-communication systems.

2.2 Learning of the physical world model

The physical world model is represented by a set of physical entities with attributes, such as shape, color, mass, center of mass, inertia tensor, friction, softness, viscosity, velocity, acceleration, and gravity. These attributes are learned by robots through passive or active sensory-motor observation. Now, let us consider the robotic task carrying a wine glass on top of a tray. In this case, the attributes of the glass, floor and interaction between the robot and floor can be useful. The glass attributes include shape, mass, center of mass, and inertia tensor, whereas the floor attributes include friction, surface shape, and slope declination. The attributes of floor–robot interaction include the relationship between action-control settings and robot acceleration. However, only a few of these attributes can be observed through image processing. Although the mass, center of mass, and inertia tensor of the glass cannot be directly observed, they can be inferred through cross-modal prediction methods, such as the multimodal learning method [27]. The acceleration realized by a particular action setting cannot be obtained in a passive manner but can be observed by performing actual actions.

2.3 Simulation search

Simulations are run several times using the computational physics engine with different parameter settings of action control and the physical world model. The simulation search has two functions: physical-world adjustment, and action-control optimization. These functions are explained as follows with respect to the task of carrying a wine glass.

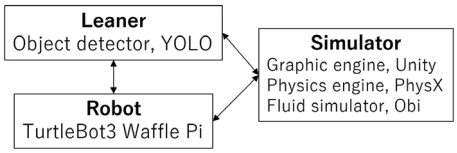

Figure 3: System implementation

2.3.1 Physical-world adjustment

Simulations are run by setting the position values of the center of mass of the glass. The simulation search determines an appropriate value of the floor friction to achieve consistency between reality and simulation of whether the glass remains on the tray or falls down. According to this value, the parameters of the physical world model are adjusted, and the adjusted physical world model is used at subsequent simulation searches.

2.3.2 Action control optimization

Here, the simulations are run by setting several different values of velocity for action control. The simulation search selects an appropriate velocity value so that the robot carries the wine glass without dropping it in a reasonable time. According to this value, the actual robot can execute the action, and this is expected to produce a desirable result.

2.4 System implementation

We implemented the physics projection system by using the TurtleBot3 Waffle Pi mobile robot, Unity graphic engine, PhysX physics engine, Obi fluid simulator, and Yolo object detector (Figure.3) .

3. Experiments

3.1 Tasks and conditions

The following three tasks were set.

STEP The robot descends one step while carrying a wine glass. The height of the step was set at 7 mm. The conditions were that the robot must not drop the wine glass; however, the wine glass may fall with the impact of descending the step. Here, the robot judges the speed at which it should descend the step.

SLOPE The robot rotates 180° on a slope with a declination of 15 % while carrying the wine glass. Here too, the robot must not drop the wine glasse; however, the glass may fall because of the centrifugal force, which changes depending on the direction of rotation because the floor is tilted. Here, the robot judges in which direction it should turn.

WATER The robot moves over an obstacle with a height and width of 7 and 120 mm, respectively, while holding a cup of water. Here, the robot must not spill water, and thus it judges the speed at which it should move over the obstacle.

The heights of the step and obstacle, and the slope inclination were measured manually, and were input into the physical-world model. The wine glass and cup were recognized by the YOLO object detection software, with a camera attached on TurtleBot3. The shape of the glass and

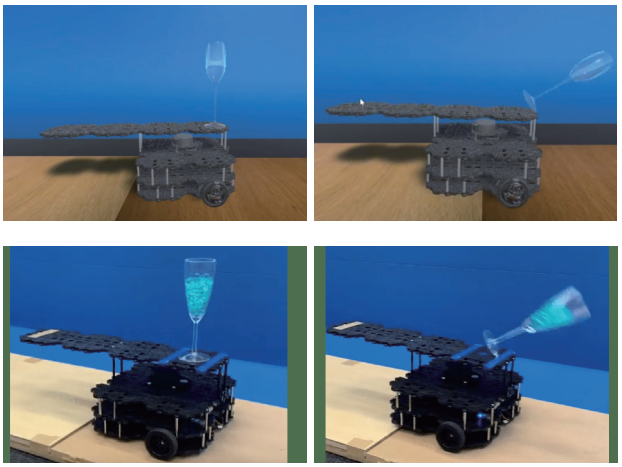

Figure 4: Resultant action effects in theSTEP task (velocity and height of center of mass). **Upper left:** Success in the prediction (15 cm/s and 13 cm). **Upper right:** Failure in the prediction (20 cm/s and 14 cm). **Bottom left:** Success in the actual task (15 cm/s and 13 cm). **Bottom right:** Failure in the actual task (20 cm/s and 14 cm).

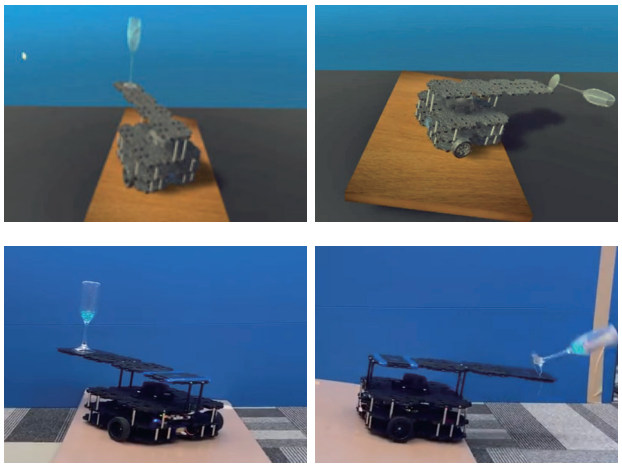

Figure 5: Resultant action effects in the SLOPE task. **Upper left:** Success in the prediction (clockwise). **Upper right:** Failure in the prediction (counterclockwise). **Bottom left:** Success in the actual task (clockwise). **Bottom right:** Failure in the actual task (counterclockwise).

cup, as well as the amounts of wine and water were calculated through subsequent image processing. The mass and position of the center of mass of the wine glass were inferred through a cross-modal prediction mechanism.

3.2 Results

3.2.1 STEP

Learing

In the initial setting, when the simulated and actual robots descended the step, their velocities changed in different ways. The impact of the robot's tire contacting the floor caused the actual robot to decelerate owing to motor characteristics and floor friction. Therefore, the deceleration in the simulation was set by considering consistency between the simulated and actual action effects at velocities of 10, 15, and 20 cm/s with the height of the center of mass of the glass 13 cm.

Prediction

After the learning, physics projection accurately predicted the action effects at velocities of 10–20 cm/s with heights of 12, 13, and 14 cm of the center of mass of the glass. These conditions were different from the condition under which the learning was done (*i.e.* cross-situational setting); this shows the generalization capability of physics projection. Figure 4 illustrates the predicted and actual action effects.

3.2.2 SLOPE

Prediction

The physics projection predicted that if the robot rotated in the clockwise direction, the wine glass did not fall; in contrast, with rotation in the counterclockwise direction, the glass kept falling down. The similar phenomena were observed for the actual robot, and the resultant effects were same as the predicted effects, as shown in Figure 5.

3.2.3 WATER

Learning

The learning was executed for the situation that the robot descended a step with the height of 7 mm, which was different from that in WATER task. At initial setting of the liquid attributes in the simulator, the physics projection prediction was not consistent with the actual effects; the simulation robot spilled water when moving at velocities of 20 and 22 cm/s, while the actual robot spilled and retained water when moving at velocities of 20 and 22 cm/s, respectively. The liquid attributes were learned in the simulation search loop to minimize the inconsistency between reality and simulation.

Prediction

After the learning, physics projection was executed for WATER task conditions. Several situations were used with different combinations of velocities of 20, 22, and 24 cm/s and the position of cup positions of 6, 0, -6, and -12 cm. The cup position is the horizontal distance in the forward direction from the center position. The physcs projection predicted that when the position of the cup moved forward, water tended to spill. Figure 6 shows the predicted and actual action effects.

Unexpected effect

When the cup was positioned at the rear end of the robot (cup position of -12 cm), the catastrophic effect that could not be expected even by a human operator was predicted through physics projection (Figure 7). When the robot moved on the obstacle, its right front wheel, which is one of two powered wheels, floated from the floor. Therefore, the robot lost control, turned unexpectedly to the right, dropped off the path, and spilled water. We confirmed that the actual robot acted similarly.

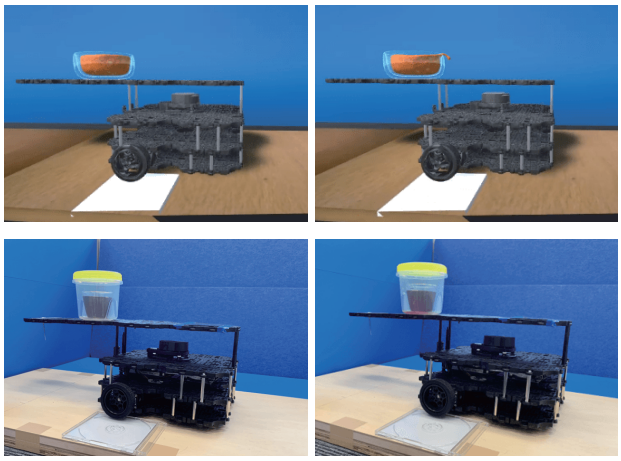

Figure 6: Resultant action effects in the WATER task (velocity and position of cup). **Upper left:** Success in the prediction (24 cm/s and 6 cm). **Upper right:** Failure in the prediction (26 cm/s and 6 cm). **Bottom left:** Success in the actual task (24 cm/s and 6 cm). **Bottom right:** Failure in the actual task (26 cm/s and 6 cm).

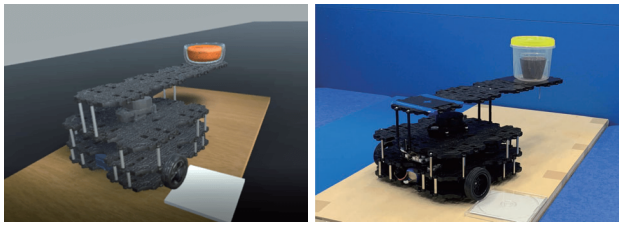

Figure 7: Unexpected action effects in the WATER task (velocity, 24 cm/s; position of cup, -12 cm). **Left:** Prediction. **Right:** Actual task.

4. Discussion

Physics projection is a very simple idea, and will attract not only technological but also scientific interests. Several research topics could be discussed, illustrated as follows:

1. When should the physical world model be learned?
2. How should the attributes be selected for learning?
3. How can incomplete physical world models work for prediction?
4. How widely should the simulation search be evaluated?
5. How would differentiable physics engines improve physics projection?
6. How can physics projection be integrated with high-order cognitive systems?

For further improvement and refinement of this approach, it is necessary to incorporate rapidly developing machine learning technology and apply recent scientific findings.

Acknowledgements

This work was partially supported by JST CREST (Grant number JPMJCR15E3, ‘‘Symbol Emergence in Robotics for Future Human-Machine Collaboration”) and JSPS KAKENHI Grant Number 18K11359.

References

[1] Martin Heidegger. *Being and Time*. 1927.

[2] Kenneth Craik. *The nature of explanation*. Cambridge University Press, 1963.

[3] Michael Polanyi. The tacit dimension. 1966.

[4] Terry Winograd. Understanding natural language. *Cognitive psychology*, 3(1):1–191, 1972.

[5] James Jerome Gibson. The ecological approach to visual perception. 1979.

[6] Humberto Maturana and Francisco Varela. Autopoiesis and cognition: the realization of the living. *Dordrecht: Reidel*, pages 2–62, 1980.

[7] David Marr. Vision: A computational investigation into the human representation and processing of visual information. mit press. *Cambridge, Massachusetts*, 1982.

[8] Dan Sperber and Deirdre Wilson. *Relevance: Communication and Cognition*. Blackwell, Oxford, 1986,1995.

[9] Stevan Harnad. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1-3):335–346, 1990.

[10] Rodney A Brooks. Intelligence without representation. *Artificial intelligence*, 47(1-3):139–159, 1991.

[11] John R. Searle. *Mind: A Brief Introduction*. Oxford University Press, 2004.

[12] Elizabeth S Spelke and Katherine D Kinzler. Core knowledge. *Developmental Science*, 10(1):89–96, 2007.

[13] Karl Friston. The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11(2):127, 2010.

[14] Georg Northoff. *The spontaneous brain: from the mind-body to the world-brain problem*. The MIT Press, 2018.

[15] Naoto Iwahashi. Language acquisition by robots. *Journal of Artificial Intelligence Society of Japan*, 18(1):49–58, 2003.

[16] Naoto Iwahashi. Robots that learn language: A developmental approach to situated human-robot conversations. In *Human–robot interaction*, chapter 5. I-Tech, 2007.

[17] Naoto Iwahashi et al. Robots that learn to communicate. In *AAAI Fall Symposium: Dialog with Robots*, 2010.

[18] Komei Sugiura, Naoto Iwahashi, et al. Situated spoken dialogue with robots using active learning. *Advanced Robotics*, 25(17):2207–2232, 2011.

[19] Brenden M Lake, Tomer D Ullman, Joshua B Tenenbaum, and Samuel J Gershman. Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40, 2017.

[20] Peter W Battaglia et al. Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, pages 18327–18332, 2013.

[21] Lars Kunze and Michael Beetz. Envisioning the qualitative effects of robot manipulation actions using simulation-based projections. *Artificial Intelligence*, 247:352–380, 2017.

[22] Nicholas Watters et al. Visual interaction networks: Learning a physics simulator from video. In *Advances in Neural Information Processing Systems 30*, pages 4539–4547. 2017.

[23] Marc Toussaint et al. Differentiable physics and stable modes for tool-use and manipulation planning. *Proc. of Robotics Science&Sysem*, 2018.

[24] Alina Kloss et al. Combining learned and analytical models for predicting action effects from sensory data. *arXiv*, 2018.

[25] Niko Sünderhauf et al. The limits and potentials of deep learning for robotics. *The International Journal of Robotics Research*, 37(4-5):405–420, 2018.

[26] Ting-Chun Wang et al. Video-to-video synthesis. In *Advances in Neural Information Processing Systems*, pages 1152–1164, 2018.

[27] Takaya Araki et al. Online object categorization using multimodal information autonomously acquired by a mobile robot. *Advanced Robotics*, 26(17):1995–2020, 2012.