Hitachi Review

January 2025 Technical Information

AI Robotics Extending Autonomous Capabilities for On-site Work

#Healthcare, QoL #Generative AI #Robotics #Research & Development

Author

Naoaki Noguchi, Ph.D.

Connective Automation Innovation Center, Research & Development Group, Hitachi, Ltd.
Current work and research: Management of research into fields such as robot systems and mechatronics.
Society memberships: The Robotics Society of Japan (RSJ) and the Japan Society of Mechanical Engineers (JSME).

Kenjiro Yamamoto

Robotics Research Department, Connective Automation Innovation Center, Research & Development Group, Hitachi, Ltd.
Current work and research: Research and development of robot systems and application development.
Society memberships: RSJ and JSME.

Hiroshi Ito, Ph.D.

Robotics Research Department, Connective Automation Innovation Center, Research & Development Group, Hitachi, Ltd.
Current work and research: Research and development of intelligent robots.
Society memberships: RSJ, JSME, the Society of Instrument and Control Engineers (SICE), and the Japanese Society for Artificial Intelligence (JSAI).

Hideyuki Ichiwara, Ph.D.

Nuclear Energy Systems Research Department，Decarbonized Energy Innovation Center, Research & Development Group, Hitachi, Ltd.
Current work and research: Research and development of robot systems and fuel debris removal techniques for Fukushima Daiichi decommissioning.
Society memberships: RSJ, JSME, and SICE.

Highlight

In addition to facilitating DX in infrastructure, transportation, and manufacturing, Hitachi is also seeking to improve the wellbeing of frontline workers. Meanwhile, the fusion of generative AI and robotics over recent years has prompted interest in the automation of workplace tasks that have been problematic in the past. This led Hitachi to enhance a deep predictive learning model and develop an AI technique for combined learning of both information on robot operations and multimodal information including vision and force-sensing. For the use of imitation learning in which robots learn actions performed by a worker, Hitachi has also demonstrated the utility of a new type of robot it developed that combines an omnidirectional movement mechanism and a dual-arm manipulator.

This article describes how, by extending these robot work capabilities, Hitachi is taking steps to reduce the demands of both dangerous tasks and routine work, while also providing workers with new ways of working that transcend time and place.

1. Introduction

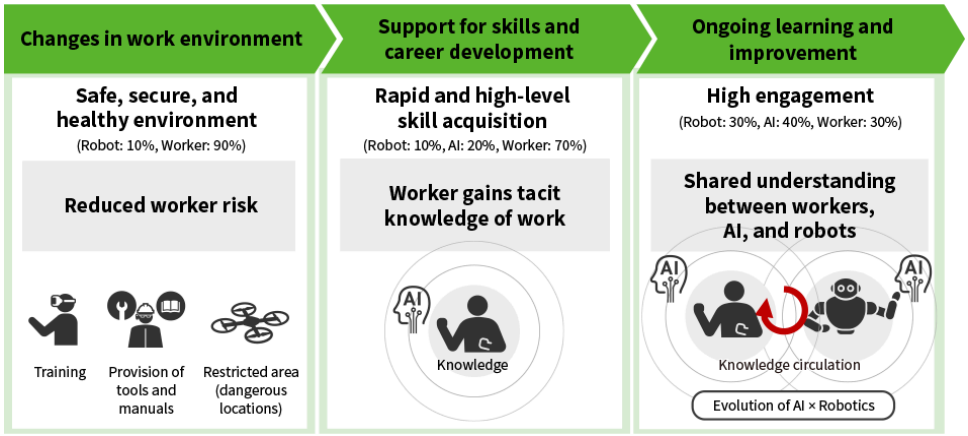
In the past, the research and development of industrial and service robots and associated topics such as autonomy, cooperation, and communication with workers has been undertaken in partnership between industry, government, and academia. In addition to the automation of factories and warehouses, Hitachi has also been working on field robotics, which has applications in construction and maintenance in sectors such as urban development and infrastructure provision[1]. Underlying this, the wellbeing of frontline workers and how to deal with their uneven distribution are issues for society to address. In order to complete their work, the workers vital to these workplaces draw on diverse work-related capabilities and experience to make judgements and decisions based on the surrounding conditions. At the same time, as there is a need to reduce workloads and shrink the labor requirements for this work, it is also anticipated that the future will bring a society that seeks to have human beings doing those things for which we are best suited. Robotics is expected to be one of the keys to creating such a society. Development is currently underway on new functions such as having robots utilize generative artificial intelligence (AI) to determine their own operations based on relevant scenarios and natural-language instructions issued by human workers.

This article describes core technologies that will accelerate practical deployment for cutting-edge robotics.

2. Challenges and Activities for Practical Deployment of Robots

Figure 1 illustrates example steps for putting robots into service. This involves successive stages, such as applications where robots take over dangerous tasks to reduce worker hazards or other risks, where generative AI is used to acquire tacit knowledge from workers so that workers and robots can work together, and where ongoing learning occurs by having both parties share intentions and experience with one another. Meanwhile, workers draw on capabilities such as their senses, communication and critical thinking, and work capabilities to address societal expectations such as those relating to safety, quality, and the environment as they go about diverse workplace tasks. It is anticipated that the above process of putting robots into service can be accelerated by using advanced technologies to extend the capabilities that human workers possess in areas like thinking, sensing, and working, underpinned by communication between workers. This section describes techniques for achieving robot autonomy and the integration of workers and robots, areas that have a close relationship with the extension of these human workers' capabilities.

Figure 1—Example Steps for Putting Robots into Service



AI: artificial intelligence

The figures use the example of the relative percentages of work performed by workers and robots to show the changes for both work and workers that occur when robots are introduced, and how they relate to one another.

2.1 Research Targets and Challenges for Practical Deployment

Extensive study will be needed if robots are to be put to practical use in open environments, including study on topics such as maintaining safety when robots come into contact with humans and their surroundings, flexibility of robot operation in response to changes in the surrounding environment, and the ability to recover from breakdowns in communication or when operations fail. There are also many cases where robots need to be taught to perform workers' actions, and where the effort and coordination required to do so takes considerable time. Recent years have also seen a lot of research going into general-purpose robots. This involves using humanoid robots to perform a wide range of different work.

This section looks at equipping robots with the autonomous ability to perform the sort of actions that workers do without thinking and teaching them to perform the deliberate actions that workers take in full awareness of what is happening around them. Among the robots being equipped with these capabilities are humanoid robots and robots with multi-jointed arms as well as mobile machines with less freedom of operation. All these technologies involve advanced robot control with capabilities such as real-time feedback.

2.2 Use of Multimodal Practices for Performing Complex Actions

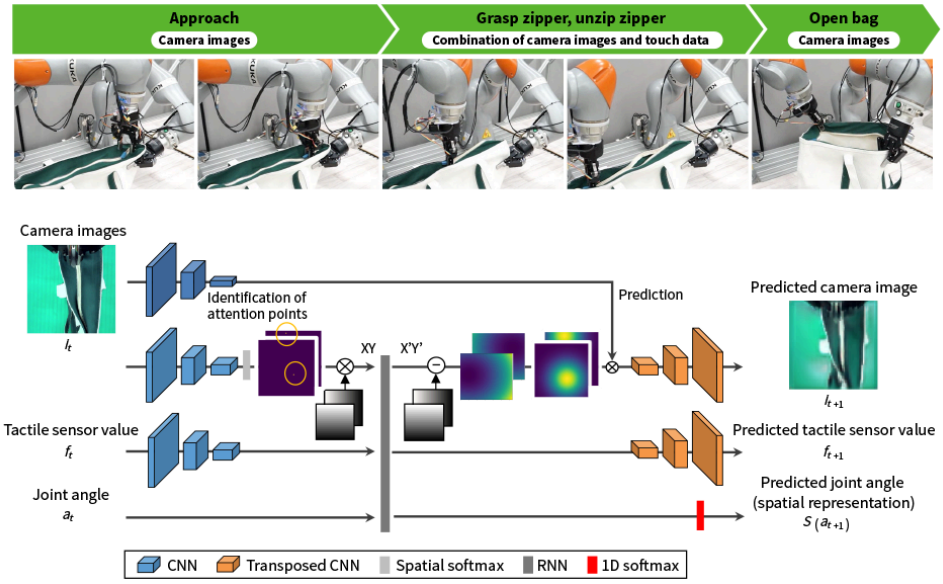
One way to perform complex actions is to predict what will happen in the near future based on multi-modal information and actions that include vision and force-sensing as well as instructions for position and joint information, and then generate the next action to perform in order to minimize the deviation from reality. In particular, weighting the importance of multimodal sensor information in real time based on the environment or task being performed is critical for making learning more efficient, as explained below. There is also a need for high-speed real-time processing using a small amount of training data.

Hitachi has made further enhancements to its existing deep predictive learning method[2], having developed a technique that progressively generates action instructions based on the next desired state. This is done by inputting multimodal sensor information into multiple time-series network models (deep predictive learning models), learning which sensory information to pay attention to, and selecting it accordingly. The learning performance of this technique can be further enhanced by using a simulator to perform many different actions in a virtual environment.

2.3 Deep Predictive Learning Integrating Multimodal Information

Figure 2 shows an example operation that involves using vision or tactile senses to unzip a floppy cloth bag. The lower part of the figure shows the deep predictive learning model used for this task[3]. A feature of this system is that it incorporates a visual attention model for reliable and robust action generation in response to dynamic changes in images obtained by camera. The pixel coordinates indicating where in the image to focus attention are obtained by using spatial softmax[*1] to highlight the feature map and soft argmax[*2] to identify the coordinates of the point with the maximum value. The technique also looks at where the tactile sensor data resembles the image, using a convolutional neural network (CNN) for tactile processing. As the CNN is robust with respect to positional shifting of the image, a similar level of robustness can be anticipated with respect to deviations in fingertip (tactile sensor) position. Learning efficiency has also been improved by adopting a spatial representation that probabilistically expresses continuous joint angle values with multiple neurons.

Figure 2—Example Task of Unzipping Cloth Bag and Overview of Deep Learning Model



CNN: convolutional neural network, RNN: recurrent neural network

Learning efficiency is improved by measures that include feature map highlighting and using separate viewpoint attention models for dynamic changes in the camera images and deviations in tactile sensor position when grasping the zipper.

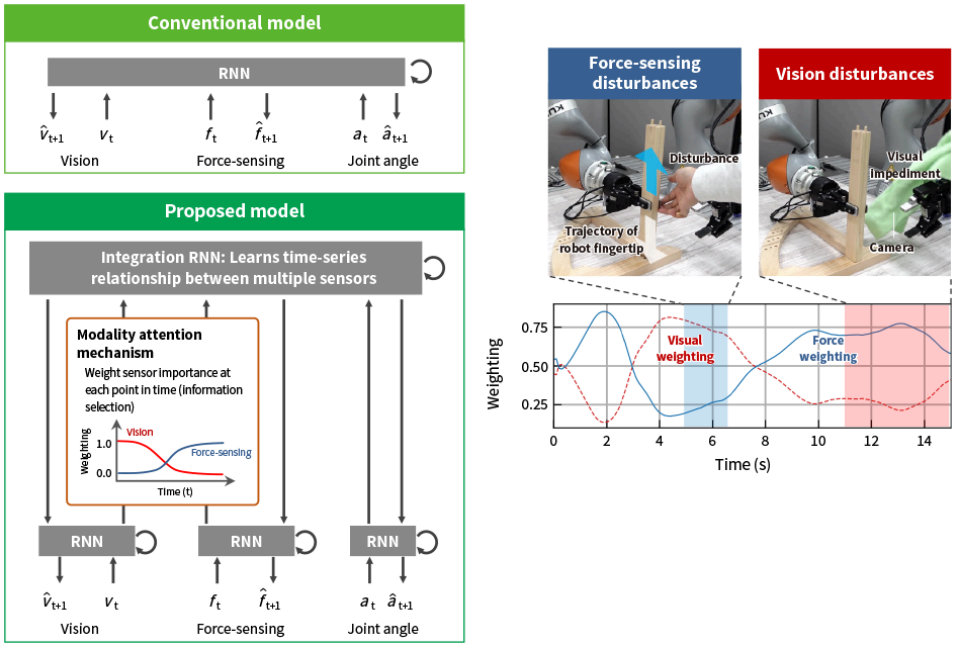
*1 A function that extends the softmax function to enable its use on two-dimensional data such as images. Softmax is a function that converts multiple inputs to probability (0 to 1) values such that the multiple outputs sum to one.

*2 A function used internally by neural networks that approximates the argmax function in a way that makes it differentiable. Argmax (arguments of the maxima) is a function that returns the index of the maximum value.

2.3 Enhancements to Multimodal Integration Learning Model

Robots should be able to operate reliably even in unstructured situations such as when there is something obstructing their field of view and they experience a force not anticipated in their planning. Accordingly, Hitachi made improvements to a deep predictive learning model to implement a function that directs their attention to force-sensing information when their vision is interfered with. The conventional method used for combining different modalities involved integrating multiple input modalities in a single module. The problem in this case is that even when sensor information is not relevant to the task it is still treated as necessary for the robot to select and perform actions, thereby turning this information into noise that diminishes the robustness of robot operation. In response, Hitachi looked at how sensory processing is distributed and integrated in the brain, adopting a design that adjusts the importance weighting of the different modalities by incorporating an integration recurrent neural network (RNN) to integrate and combine the single-modality RNNs that handle sensing and movement. This is called a modality attention mechanism. Figure 3 shows the chair assembly task used to verify that this will work; it calls for controlling the position and level of force exerted by the fingertips that perform the action[4), 5)]. The test demonstrated that a heavy weighting was placed on vision information when aligning parts with one another and on force-sensing information when fitting them together. By providing an indication of these weightings as the robot performed the task, the test was also able to keep track of which sensor information was being used to determine its action.

Figure 3—Concept of Modality Attention Mechanism and Example Chair Assembly Task



The robot learned how to weight the different sensor information simply by having a worker demonstrate the desired actions multiple times. This enabled it to pay attention to different sensor information as it completed the task, depending on the action it was performing and its surroundings.

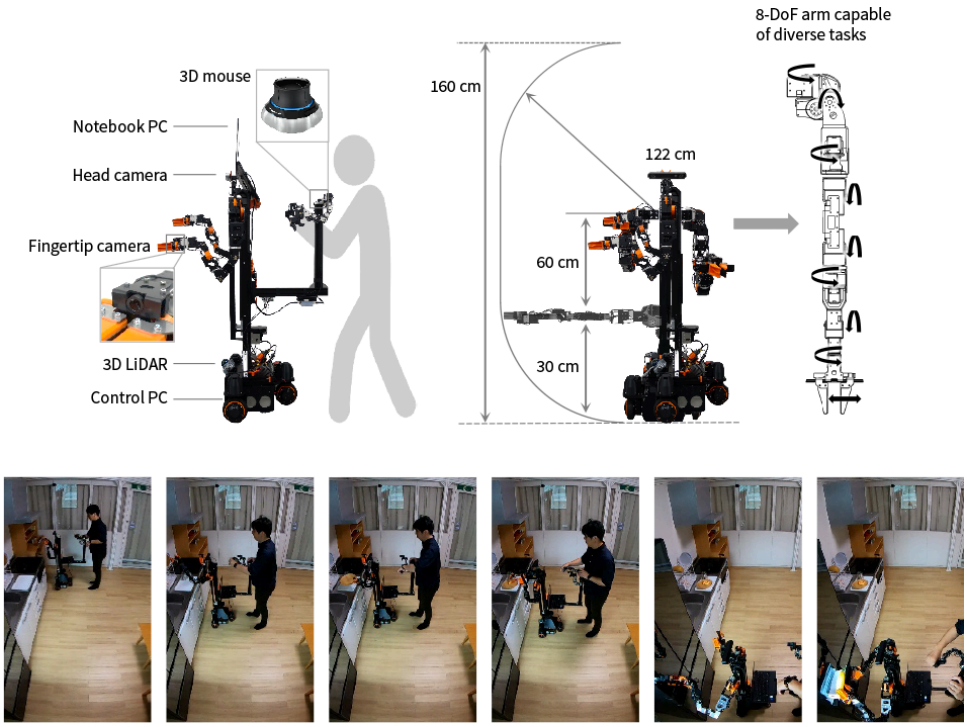
2.4 Imitation Learning for Closing Gap with Reality

This section describes the challenges that arise with real-world robot operation when simulation is used to train robot actions. A “reality gap” invariably occurs during actions such as touching. This takes the form of numerous physical quantities diverging between simulation and reality and there are cases when the intended action cannot be completed without taking these divergences into account.

One way to overcome this is for a worker to show the robot how to perform the task and to perform action learning using the sensor data acquired during this time. Generalized capabilities acquired from previous teaching and learning depend heavily on the quality of the data, with data quality becoming even more important as tasks get more complicated. Accordingly, Hitachi has developed its own robot featuring action teaching that is low-cost and intuitive while still offering considerable flexibility[6]. It is equipped with a leader-follower mechanism that enables it to perform elaborate and delicate tasks with only a small amount of learning data.

Figure 4 shows a dual-arm mobile manipulator, with the dual-arm manipulator being fitted on top of an omnidirectional movement mechanism. The robot is made up of a head, body, tractor unit, and remote control unit. The body is comprised of the dual-arm manipulator, which is equipped with 8-degrees-of-freedom (DoF) arms and 1-DoF grippers, and a mechanism that moves up and down. The tractor unit has independent three-wheel steering. The remote control unit used for teaching is made up of a leader arm and a three-dimensional (3D) mouse and notebook PC. Visual information acquired by the robot camera is displayed on the PC. The quality of the teaching data can be degraded by workers’ unconscious tendency to use information from outside the robot’s field of view. However, as is done on this robot, it is possible to acquire data of sufficiently high quality for imitation learning by having the teacher operate the robot like a glove puppet, viewing what the robot is seeing on the PC screen as they train it. Hitachi has demonstrated that use of this teaching method enables a wide range of actions to be taught in a short span of time, such as opening and closing doors or taking objects out of cupboards in a kitchen, making it possible for the robot to perform these actions autonomously.

Figure 4—Dual-arm Mobile Manipulator and Associated Action Instructions



3D: three dimensional, DoF: degrees of freedom

By viewing what the robot is seeing during a short period of training, a worker can teach it to perform a wide range of actions, such as opening and closing doors or taking objects out of cupboards in a kitchen. This enables the robot to perform these actions autonomously.

3. Conclusions

A lot of work is currently going on in the field of robotics research with the goal of enabling robots to perform diverse and complex tasks. This includes humanoid robots and general-purpose foundation models intended for use with them, remote control using virtual reality (VR), human-robot interaction featuring natural language and AI, and the manipulation of objects that move unpredictably[7]. This article has focused on multimodal information and operations such as vision and force-sensing, describing an AI learning platform that achieves high learning efficiency with a small amount of data and a robot that can collect data of sufficiently high quality for imitation learning in which a worker shows the robot what to do. In the future, Hitachi will accelerate the practical deployment of these robots by using them to collect data from a wide variety of environments. By collating and utilizing the data collected by the robots, Hitachi also intends to establish an ecosystem-style business and accelerate social implementation.

Acknowledgements

Considerable assistance was received from Professor Tetsuya Ogata of Waseda University in the development of the technique described in this article for using deep learning for robot action training and action generation. The authors would like to take this opportunity to express their deep gratitude.

REFERENCES

1) K. Yamamoto et al., “Field Robotics Bringing Innovation to Digital Solutions for Field Work,” Hitachi Review, 70, pp. 470-475 (Jun. 2021).
2) H. Ito et al., “Efficient multitask learning with an embodied predictive model for door opening and entry with whole-body control,” Science Robotics, Vol. 7, Issue 65 (Apr. 2022).
3) H. Ichiwara et al., “Contact-Rich Manipulation of a Flexible Object based on Deep Predictive Learning using Vision and Tactility,” 2022 IEEE International Conference on Robotics and Automation (ICRA), pp. 5375-5381 (2022).

4) H. Ichiwara et al., ”Multimodal Time Series Learning of Robots Based on Distributed and Integrated Modalities: Verification with a Simulator and Actual Robots,”2023 IEEE International Conference on Robotics and Automation (ICRA), pp. 9551-9557 (Jun. 2023).

5) H. Ichiwara et al., “Modality Attention for Prediction-Based Robot Motion Generation: Improving Interpretability and Robustness of Using Multi-Modality,” IEEE Robotics and Automation Letters, Vol. 8, Issue 12, pp. 8271-8278 (Dec. 2023)

6) H. Ito et al., “Motion Generation Method for Object Manipulation with Uncertainty Development of a multi-tasking robot for human life support (1),” Japan Society of Mechanical Engineers (JSME), Robotics and Mechatronics Conference (May 2024) in Japanese.

7) K. Yamamoto et al., “Real-time Motion Generation and Data Augmentation for Grasping Moving Objects with Dynamic Speed and Position Changes,” Proceeding of 2024 IEEE/SICE International Symposium on System Integration (SII) (Jan. 2024).

Hitachi Review

Hitachi Review is a technical medium that reports on Hitachi's use of innovation to address the challenges facing society.

The *Hitachi Review* website contains technical papers written by Hitachi engineers and researchers, special articles such as discussions or interviews, and back numbers.

Hitachi Hyoron (Japanese) website	*Hitachi Review* (English) website
https://www.hitachihyoron.com/jp/	https://www.hitachihyoron.com/rev/

Hitachi Review Newsletter

Hitachi Review newsletter delivers the latest information about Hitachi Review when new articles are released.