Aircraft Operation Support by Learning Expert Know-how Based on Instrument Information and Visual Information

YUSUKE HAZUI[*1] HIROFUMI BEPPU[*1]
HITOI ONO[*2] TAKESHI TSUCHIYA[*3]

The rearing of aircraft pilots requires a lot of cost and time, so it is highly important to improve the training efficiency. To efficiently improve pilot abilities, appropriate instruction by experts is necessary, but the shortage of such experts is becoming an issue due to the aging of pilots. Therefore, for the purpose of training, we have developed an inverse reinforcement learning method to learn piloting know-how based on information from instruments and visual information that experts pay attention to during a piloting operation. We applied this technology to aircraft landing problems and confirmed that it enables learning of experts' landing skills. This report also describes a pilot training system that utilizes the obtained piloting models.

1. Introduction

In recent years, the declining birthrate and aging population in Japan have affected various industries. In the aviation industry, the shortage of aircraft pilots, the rearing of which requires time and costs, has become serious, and rapid rearing of pilots through efficient training is one of the challenges there. Normally, after completing classroom lessons, trainees experience and learn basic control operations such as takeoff/landing, altitude maintenance, turning, etc., using a simulator before actual flight training with actual aircraft. At this stage, appropriate guidance by an expert instructor is the key to rapid rearing, but it is becoming increasingly difficult to secure sufficient human resources and time for such guidance. As such, based on the concept that the development and application of a training support system that can automatically generate appropriate advice on behalf of instructors and enable trainees to learn on their own will lead to the solution of the issue, we generated an expert AI that imitates the piloting operations of an expert and constructed an evaluation index that clearly shows the difference between the expert and the trainee and makes the trainee aware of it. This report describes, for a landing operation as an example, the simultaneous learning of an expert AI and a reward function[(1)], which is an evaluation index, based on instrument and visual information obtained from an expert's operation using an adversarial inverse reinforcement learning method, one of the reinforcement learning methods. **Figure 1** shows the framework of this technology.

Hereafter, Chapter 2 presents an overview of the developed method, Chapter 3 describes the results of applying the developed method to an aircraft landing problem, Chapter 4 discusses the applicability of the developed method to aircraft training support, and Chapter 5 provides a conclusion.

*1 CIS Department, Digital Innovation Headquarters, Mitsubishi Heavy Industries, Ltd.
*2 Chief Staff Engineer, CIS Department, Digital Innovation Headquarters, Mitsubishi Heavy Industries, Ltd.
*3 Professor, Department of Aeronautics and Astronautics, University of Tokyo

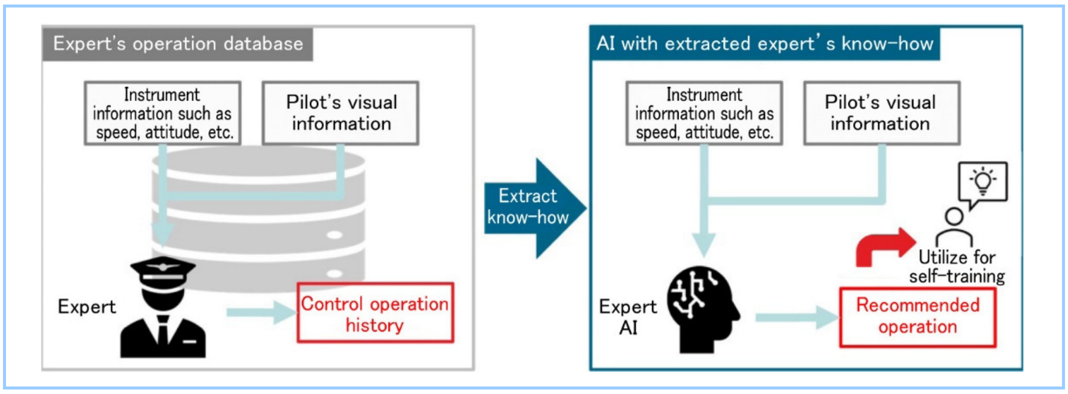

Figure 1 Framework of this technology

2. Inverse reinforcement learning technology based on multimodal information

2.1 Reinforcement learning and inverse reinforcement learning

Reinforcement learning is known as a method to autonomously learn how to perform operations depending on the situation in a dynamically changing environment. Reinforcement learning is one of the machine learning methods, for which an index to evaluate operations, called a reward function, is designed in advance to learn operations that maximize the rewards to be obtained.

On the other hand, in applying reinforcement learning to learning aircraft operation, there are issues such as the fact that it is not easy to design rewards that formulate evaluation criteria of experts, and that operations that experts cannot assume are difficult to accept in the field.

Inverse reinforcement learning, which is a technology that enables the estimation of the reward function and the learning of the operation method simultaneously by using an expert's operation data, can address these issues. Supervised learning is a well-known method for mastering an expert's operations under certain conditions, but it requires preparation of the expert's operation data that cover all the conditions of the desired operation. On the other hand, by learning anew using the estimated reward function even under conditions different from the one in which the expert's operation data were obtained, inverse reinforcement learning enables learning under conditions where no expert's operation data exists without designing a reward function. As a result, learning a control operation method applicable to a wide range of conditions can be made with a small amount of data and a small number of trials. **Figure 2** shows the difference between reinforcement learning and inverse reinforcement learning.

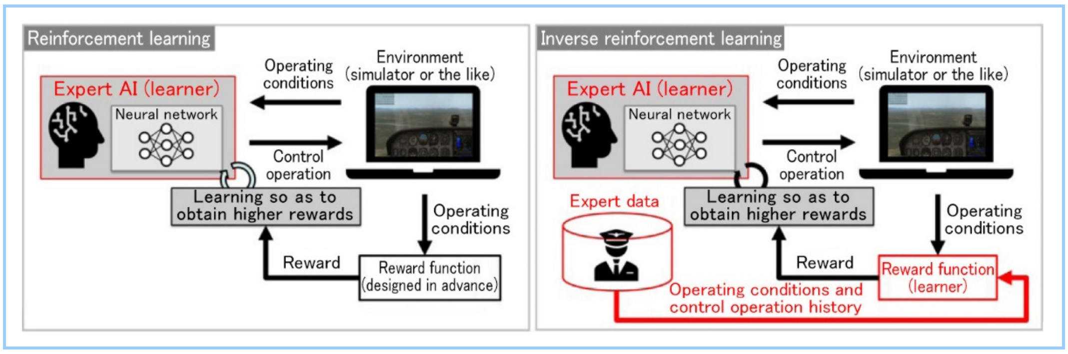

Figure 2 Summary of reinforcement learning and inverse reinforcement learning

2.2 Inverse reinforcement learning based on multimodal information

Normally, a person perceives a situation for decision-making based on multiple types of information, such as measurement values indicated by instruments and visual information. For example, in piloting an aircraft, experts make decisions based on instrument indications, visual information outside the cockpit window, engine noise, and perceived gravity. Therefore, for learning the expert AI, an inverse reinforcement learning method that can handle the multimodal information is needed. Section 2.2 describes the developed multimodal information-based inverse reinforcement learning technology.

Figure 3 shows a structural diagram of the developed inverse reinforcement learning method. In this report, AIRL (Adversarial Inverse Reinforcement Learning[(2)]), which incorporates an adversarial generative network, one of the deep learning technologies, to learn more complex actions, is employed as the base inverse reinforcement learning method.

AIRL consists of two neural networks: a generator and a discriminator. The generator learns the expert's operations using reinforcement learning and generates operation data using the expert AI. On the other hand, the discriminator identifies whether the data is generated by the expert AI (false) or is an expert data (true) prepared as training data. The generator learns operations that are closer to the training data in an attempt to fool the discriminator, whereas the discriminator learns to better distinguish between false and true data so that it is not fooled by the generator. When the generator learns the expert's operations completely, the discriminator will not be able to discriminate between operation data generated by the generator and the expert's operation data, and consequently it can be said that the generator has learned to imitate the actions of the expert.

The multimodal information consists of sensor values (measurement values indicated by instruments) and image information obtained from the simulator environment. We attempted to synthesize sensor values and image information by performing preprocessing such as CNN (Convolutional Neural Network) or geometric transformations on the image information. Expert pilots recognize the altitude of the aircraft and the distance from the current position to the runway by using the visually obtained runway geometry information. In contrast, the technology developed in this report transforms geometrically the runway width and the angle between the runway sides into visual information (the way it looks on the display)[(3)] shown in **Figure 4** to obtain visual information necessary for learning.

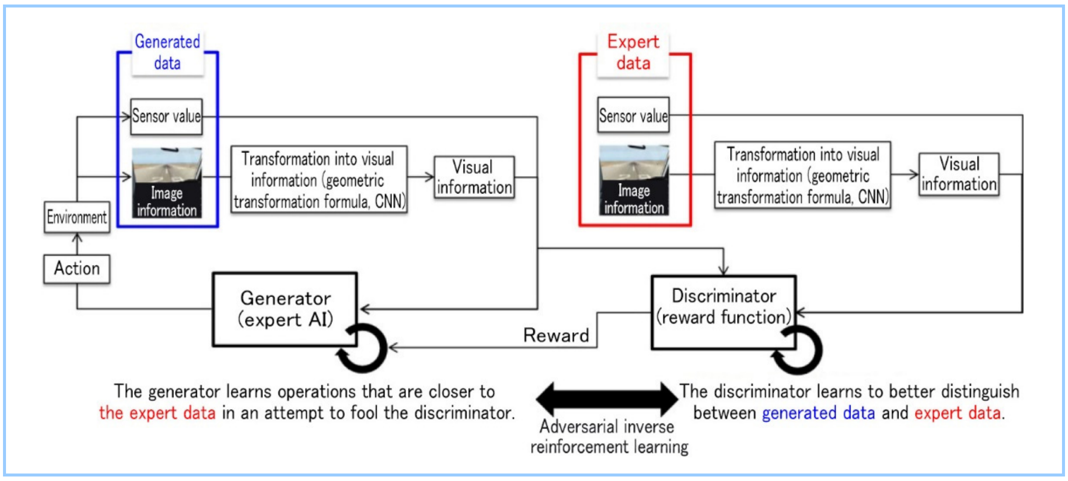

Figure 3　Structure of developed method

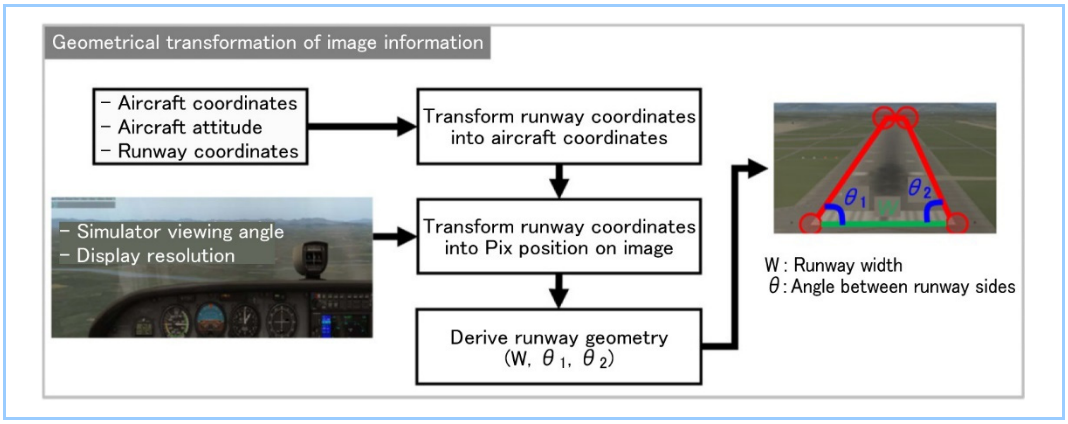

Figure 4　Example of transformation into visual information

3. Application to aircraft landing problem

3.1 Problem setting

Using the commercial flight simulator X-PLANE11[4], we assumed landing of an aircraft on a runway from initial conditions of the aircraft position and attitude where the runway is in view while meeting the altitude and airspeed requirements. **Figure 5** shows the view from the cockpit in the initial conditions. During flight, an aircraft is subject to disturbance due to headwinds and must be controlled accordingly to land in the desired state. In this simulation, altitude and speed control by the elevator and throttle are the learning target.

An expert's landing operation data required for learning in the developed method described in Section 2.2 were collected from an actual landing operation performed by an expert on the X-PLANE11 using a commercially available yoke (control stick) and throttle as shown in **Figure 6**.

Figure 5 Initial cockpit view on X-PLANE11

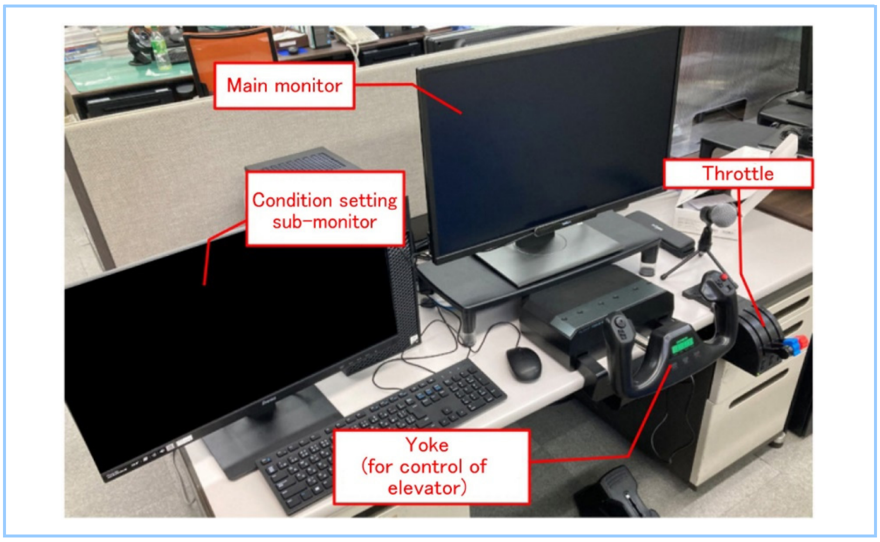

Figure 6 Configuration of equipment setup used for expert data collection

3.2 Learning results

To monitor the learning process in the developed method, an evaluation function was designed so that higher scores are given to flights that satisfy the requirements that need to be met for a successful landing, such as the allowable altitude and airspeed. **Figure 7** shows the relationship between this evaluation function value and the learning episodes. It is indicated by Figure 7 that as the number of trials increased, the learning in the developed method progressed well, and finally the landing operation was learned after 207 learning episodes. The 207 learning episodes took just over 17 hours.

Figure 8 shows the results of landing operation performed by the learned expert AI (policy function). The expert data used for the learning are shown as the red line, and the results obtained with the developed method are shown as the blue line. It was confirmed that the requirements for the allowable altitude and airspeed were met.

In addition, the elevator and throttle operation history of the expert and that of the developed method were similar. For example, the actions of reducing the elevator and throttle to maintain the aircraft's pitch angle and airspeed immediately after wind disturbance are common to data of both actions. The actions of increasing the throttle compared to the initial setting to compensate for the lack of propulsive power due to the wind disturbance are also common. In particular, throttle operation is known to be difficult for beginners to master because it not only increases or decreases the airspeed, but also affects the aircraft attitude, causing changes in the altitude and further changes in the airspeed. In a trial using a conventional reinforcement learning method conducted prior to this trial, it was not possible to obtain throttle operations similar to those of an expert. Thus, it was confirmed that the developed method can perform learning of an expert AI that can imitate the landing operations of an expert.

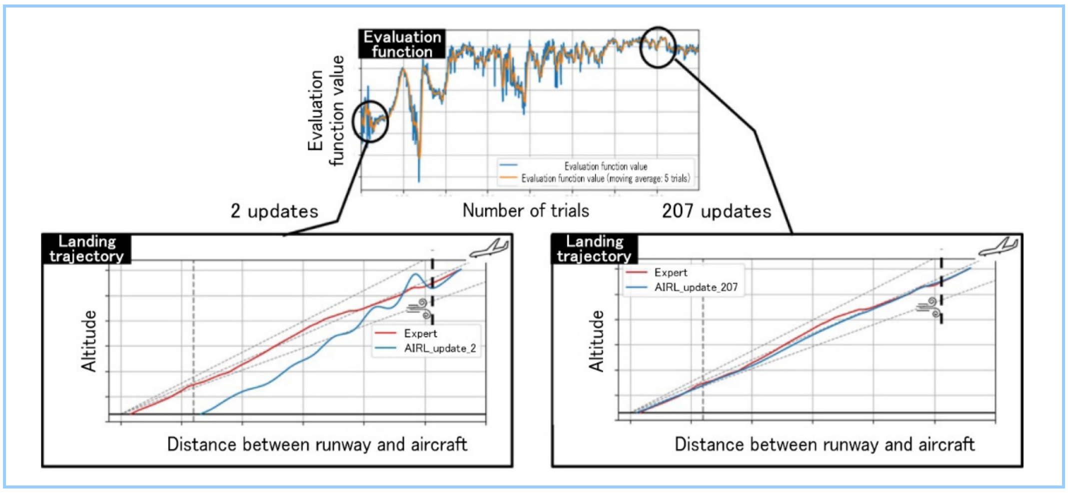

Figure 7 Relationship between evaluation function and number of trials

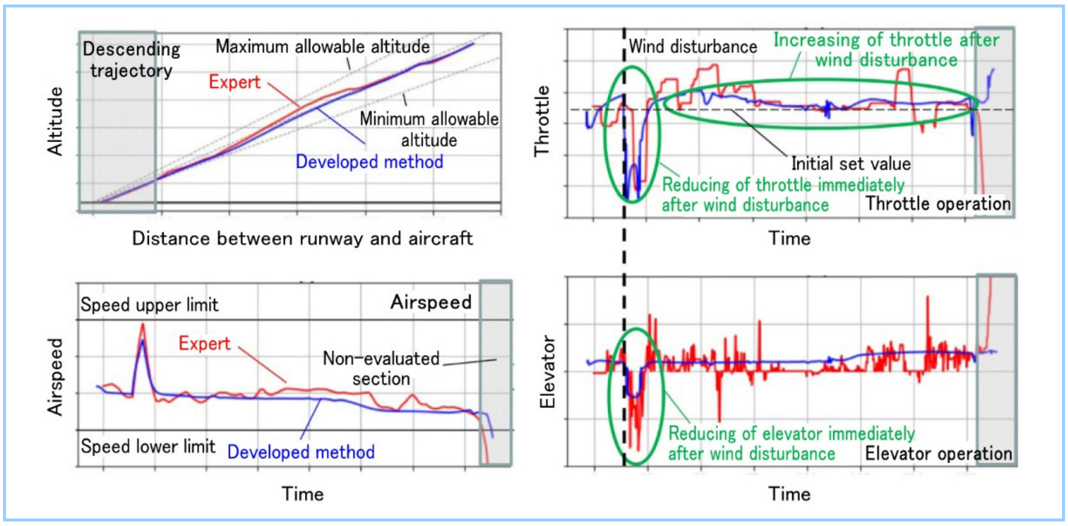

Figure 8 Expert AI's landing operation data

3.3 Robustness verification of expert AI

The expert AI evaluated in Section 3.2 was obtained by performing learning under conditions where the altitude and speed of the wind disturbance were kept constant, and the expert data used for the learning was data for a single landing operation under the same conditions. On the other hand, the expert AI is intended to be used for pilot training, and is required to be able to provide appropriate operations under various aircraft conditions and wind disturbance conditions. In other words, the expert AI needs to be robust to environmental changes. To verify the robustness of the expert AI, evaluation of its behavior under different wind disturbance conditions was conducted.

Specifically, the wind speed and the wind disturbance timing were changed. **Figure 9** shows the behavior when the wind speed was changed, and **Figure 10** shows the behavior when the wind disturbance timing was changed. It was confirmed that in both cases the landing was successful while satisfying the necessary requirements. These results indicated that an expert AI that is robust to environmental changes can be generated without prior collection of expert data under all assumed conditions.

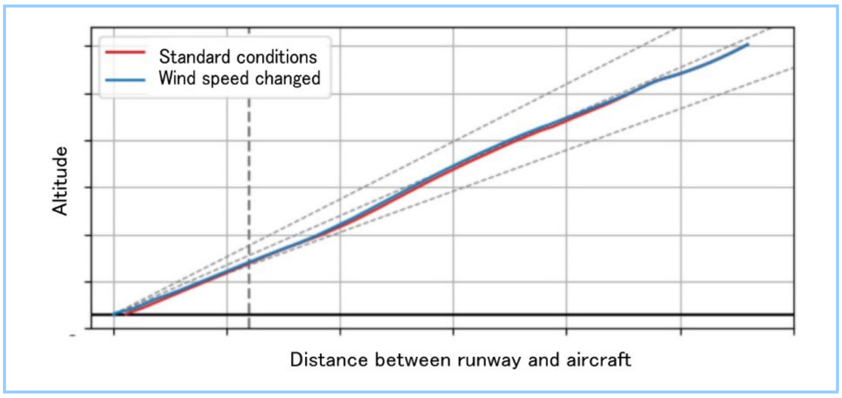

Figure 9 Landing trajectory by expert AI in case of changing wind speeds

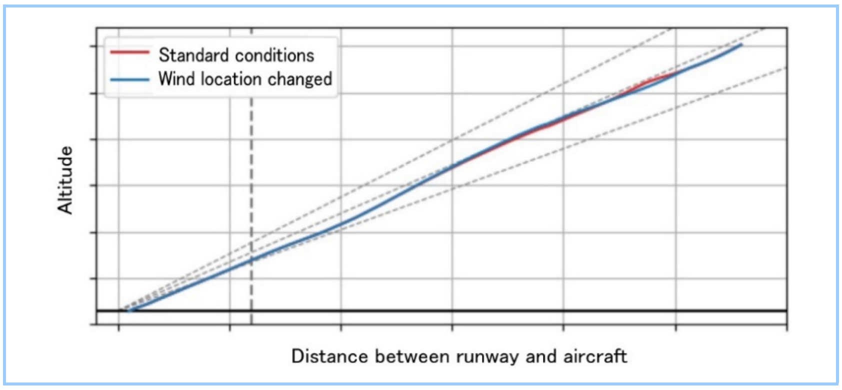

Figure 10 Landing trajectory by expert AI in case of changing wind generation timings

4. Application to future flight training systems

We aim to realize a self-training system in the future, in which a trainee can learn with feedback from an expert's recommended operations and intentions, as if an instructor pilot is supporting the trainee's practice. This section introduces a prototype system that utilizes recommended operations and evaluation of flight state values by an expert AI.

Figures 11 and **12** show screen-capture images of a feedback video to a trainee's flight operations on the aircraft simulator. The cockpit view image on the left of Figure 11 shows the simulator screen, and the chart on the lower right of the figure compares the trajectories of the trainee and the expert, in which the red dot represents the trainee's current flight position. The green bar in the chart on the upper right of the figure shows the trainee's current elevator (control stick) and throttle (power) control amounts, and the red line represents the operation recommended by the expert AI. For example, in the state shown in Figure 11, the expert AI is recommending to push the control stick, but the trainee positions the control stick in the neutral position, which is an undesirable operation. The expert AI is also recommending to control the throttle to reduce the output, but the trainee is not doing so. The charts on the upper left and lower left of Figure 12 are similar to the above described charts of Figure 11, and the state value heat map on the lower center of the figure shows the value of the flight state calculated by the expert AI. In this state value heat map, the high value region means the state that the expert AI is aiming for. The red circle and cross also indicate the flight state of the expert and trainee, respectively, so the trainee can learn what kind of flight state to aim for in response to ever-changing conditions. For example, in the state shown in Figure 12, the expert AI calculates and indicates that a state with reduced airspeed and pitch angle compared to the current flight state is preferable, and outputs the operations necessary to transition to that state of recommended operations. We believe that feedback of the above information to trainees may lead to correction of the trainee's control, and we will continue to study this possibility in the future.

Figure 11 Screen-capture images of feedback video to trainee's flight operations (i)

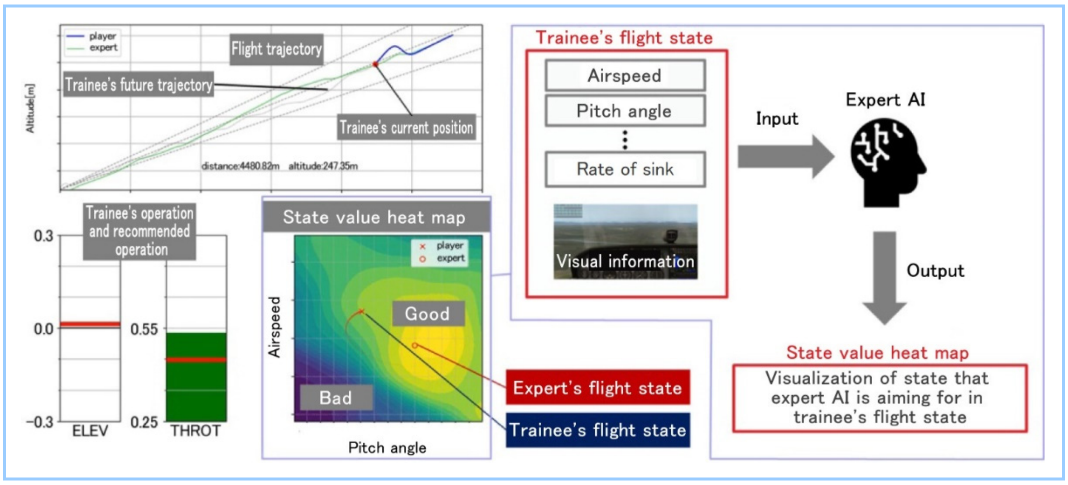

Figure 12 Screen-capture images of feedback video to trainee's flight operations (ii)

5. Conclusion

This report presented the developed inverse reinforcement learning method for learning the know-how of experts by incorporating multimodal information consisting of instrumental and visual information. In addition, the developed method was applied to an aircraft landing problem, and it was confirmed that a robust expert AI can be generated that can control the aircraft with the same operations as an expert. Furthermore, a prototype system that supports flight training was introduced, and a support use case in which operations recommended by the expert AI and a state value heat map were used was shown.

In the future, we will develop a hierarchical inverse reinforcement learning technology and a method with higher explainability to deal with more complex problems required in actual flight operations, and we will also consider applying this technology to flight training systems.

References

(1) H. Beppu, et al., Learning of Piloting Skills for Aircraft by Adversarial Inverse Reinforcement Learning with Multimodal Information, Proceedings of 2023 62nd Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), 2023

(2) J. Fu, et al., Learning Robust Rewards with Adversarial Inverse Reinforcement Learning, International Conference on Learning Representations, 2018

(3) Ryota Mori, et. al, Study on Analysis Method of Pilot Maneuver During Landing Phase, Doctoral Dissertation, University of Tokyo, 2009

(4) Laminar Research, X-PLANE 11 Professional, 2017, https://www.x-plane.com/product/x-plane-11-professional-use-digital-download/