別紙４－１（課程博士（英文））

Date of Submission (month day, year) : July 6th, 2021

Department of Computer Science and Engineering	Student ID Number	D189301	Supervisors	Jun Miura, Shigeru Kuriyama
Applicant's name	Chandra Kusuma Dewa			

Abstract (Doctor)

Title of Thesis	Policy Integration for Person-Following Robot Training Using Deep Reinforcement Learning

Approx. 800 words

Recently, robots are widely used in numerous fields to support humans in many tasks. Thanks to the rapid growth of technology, the automated machines are becoming more powerful and more beneficial in their roles to help performing some tasks that humans even cannot do. Likewise, a particular type of partner robots which can interact closely to humans is also required to support their daily needs. In such cases, there are also demands for specialized robots which have the main ability to follow and attend specified humans in a safe close distance continuously. However, making them able to perform that kind of skills is not trivial since person-following cannot be considered as a simple task to be performed by robots.

In this thesis, we tried to train a mobile robot for performing the person-following task. Here, we see that the task is a complex task which can be broken down into several simpler sub tasks. When the robot is away from the target person, it should be able to perform the navigation task safely in order to make its position is close enough to him. Afterwards, the robot should also be able to perform the attending task properly once its position is close to the target person. In our study, we consider several previous studies which tried to make the robot able to accompany the target person at his left side or at his right side instead of following him from behind. In order to make the robot able to master the complex person-following task appropriately, we utilize deep reinforcement learning (DRL) approach for both obtaining the optimal policy for each sub task and integrating all those optimal policies into one strong optimal person-following meta-policy.

To obtain the optimal navigation policy, we employ the soft actor-critic (SAC) learning algorithm for making the robot able to approach the target person well. Moreover, we propose a specific framework which is intended to train a mobile robot to navigate quickly but safely. The framework utilizes a novel state transition checking method to ensure that the training environment provided for the robot always follows the Markov decision process properly. Furthermore, it also employs a novel velocity increment scheduling technique during the training process. The technique follows a curriculum learning strategy by setting a small value of velocity for the robot at the beginning of the training episode. As the number of episodes increases, the robot's velocity is increased gradually so that the robot can gradually learn the complex task of fast but safe navigation in the

training environment form the easiest level, such as the one with the slow movement, to the most difficult level, such as the one with the fast movement.

To obtain the optimal attending policies, we also use the SAC algorithm to make the robot able to attend the target person when its position is close to the target person. During the training process, we propose the U-shaped reward function which can guide the robot to attend the target person at his left side or at his right side. Moreover, we propose a novel weight-scheduled action smoothing technique so that the robot can generate smooth and safe trajectories for the attending task. To make the robot can better portray the surroundings we also propose a novel policy network architecture which employs one dimensional convolutional neural network to extract features from laser scans automatically.

Finally, to integrate all the optimal navigation and attending policies, we also propose a framework which employs the double deep Q-network which can make the robot learn to choose the most appropriate policy given the current state of the person-following environment. Inside our framework, we introduce the action generator module which can adjust the state of the person-following environment for each sub policy appropriately. Furthermore, the module is also able to smooth the actions generated by the robot using the action smoothing strategy to prevent the robot hitting the target person when it is close to him and when the robot's actions are generated from changing policies.

From all experiments that we conduct in our study, we can conclude that the proposed navigation training framework is able to make the robot navigate approaching the target person faster with lower collision rate compared to other DRL-based navigation baseline frameworks. We confirm that the U-shaped reward function and the weight-scheduled action smoothing that we propose can make the robot attend the target person both at his left and right side with smooth and safe trajectories. We also confirm that our proposed framework can integrate all the navigation and attending policies for obtaining the meta-policy for the person-following task. For future work, we consider using dynamic environments for the training of the navigation and the attending tasks so that more robust policies can be obtained. We also plan to propose another method so that the robot can switch its attending position easily when it is close to the target person. Furthermore, we also will use computer vision techniques for detecting and tracking the target person so that the policies can be deployed for real robots.