



# Augmented Reality Based Actuated Monitor Manipulation from Dual Point of View

Ying Ren<sup>(✉)</sup> and Jiro Tanaka

Graduate School of Information, Production and System, Waseda University,  
Kitakyushu, Japan  
m18302962396@163.com, jiro@aoni.waseda.jp

**Abstract.** The mobile robots are increasingly used in domestic space for room surveillance. However, joystick based controller dose not realize direct and intuitive motion control for monitor. To solve this problem, we propose an augmented reality based interface to control monitor from dual point of view, referring to the third-person view and rear first-person view. In this system, augmented reality models used to represent monitor are superimposed on the dual point of view, which enables users to control each part of actuated monitor intuitively with interaction of augmented reality models on screen. Through augmented reality models, our system realized the concept of user-centered manipulation since the control of target object is described in user's coordination system. In addition, we carried out a preliminary user study to evaluate our design and the performance of the system, and a positive feedback has been received.

**Keywords:** Robot control · Augmented reality · Dual point of view · Direct control

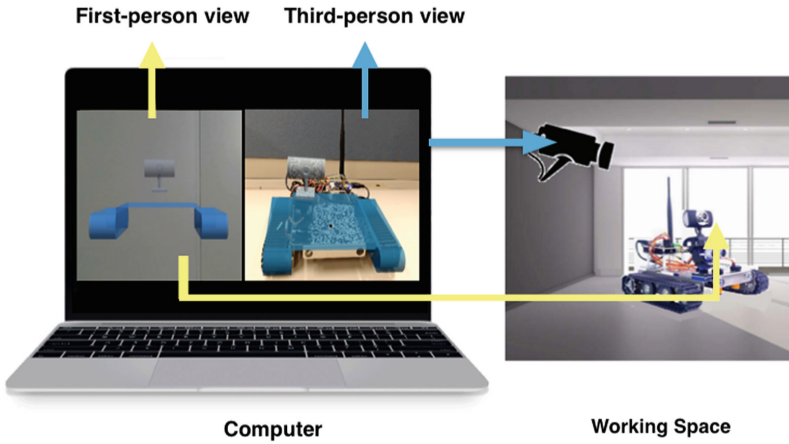
## 1 Introduction

Recent years, mobile robots are increasingly utilized in domestic space for room surveillances. Since the complex motion control of monitor in diverse working environments, it is necessary to develop a natural Human-Robots Interaction system to realize intuitive manipulation [8]. Especially for inexperienced user, understand the function of button may burden their thinking.

Traditionally, joystick based controller is most commonly used for robot control. However, the limitation of button operation cannot enable user to realize direct and intuitive operation. On the one hand, complicated motion manipulation implemented on the multiple buttons is not a direct way to indicate the controlling relationship between controller and robot. Especially for inexperienced user, it may demand more effort for them to get familiar with the function of controller. Moreover, by using joystick based control, the gap between user's coordinate system and robot's coordinate system may led to mapping

errors between users' input and robot's motion. Therefore, extra calculation is required for user to map the relative direction between themselves and robot.

Recent years, however, compared to conventional button controller, video-based interface like smartphone and PC are widely used which can provide unlimited interaction designed by software [4]. Commonly, their control is based on single view, such as third-person view or first-person view. For operation on these two views respectively, however, some disadvantages cannot be solved. Specifically, in the third-person view provided by fixed camera, robot cannot be manipulated in dead zone of view. In the first-person view, since the narrow vision of robot-mounted camera, user cannot understand the situation of entire working space.



**Fig. 1.** Augmented reality based robot manipulation system.

In this study, we propose a robot manipulation system which enables users to control the actuated monitor intuitively with interaction of augmented reality model from dual point of view, which refers to the third-person view and rear first-person view. As shown in Fig. 1, the screen of PC displays two live videos acquired from fixed setting camera observing the robot from third-person view, and robot-mounted camera observing the working space from first-person view. On dual point of view, augmented reality models are applied to monitor control which allow users to manipulate each part of monitor by directly dragging the corresponding models to target position or direction. With these models, we can realize the spatially consistent relation between virtual world and real world for more intuitive operation. Furthermore, augmented reality based manipulation eliminates the difference of view point between users and monitor since manipulation is described in user's coordinate and it is unnecessary to consider relative position and direction between monitor and themselves.

In addition, our system is based on dual point of view. In the third-person view, users can catch the sight of entire working space including controlled

monitor, target objects for observing and surrounding obstacles. With this, it is more efficient to determine the motion of monitor and much easier to avoid collisions. However, in specific situation such as dead zone of the view from fixed setting camera, user need to operation in first-person view. In practical work, two point of view are complementary to each other in diverse settings to provide users with all-around view.

## 2 User Interaction

Our system is an augmented reality interaction interface for users' operation on monitor. The two cameras capture the live videos of the working space in different view point which show on the screen simultaneously. Physical monitor in live video is overlaid with augmented reality models on dual point of view. In the way of mouse control, the user select the part of virtual model they want to control, and then drags it to the target position and direction. As user control the virtual model in real time, the system drives the physical monitor so that it matches with the motion of augmented reality model.

### 2.1 Dual Point of View

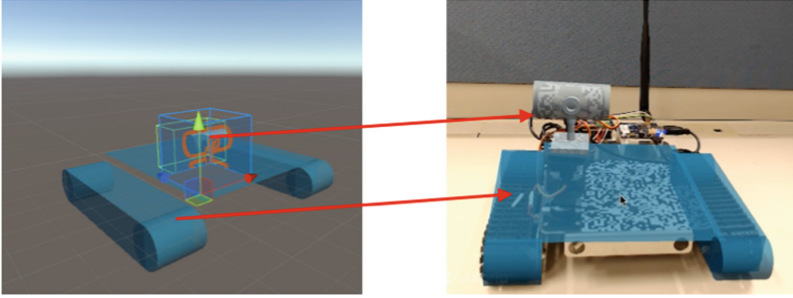
In order to provide user with all-around view, our system allows the operation on monitor from two point of view: third-person view and first-person view.

**Third-Person View.** Surveillance by fixed setting camera in different places is the most commonly way for safety supervision. Usually, the camera is mounted in appropriate position without much obstructions on view. In addition, the height of camera is exceeded the person's eye level in order to obtain a broad vision. In our system, a fixed camera is installed in the working space, the view from this camera is considered as the third-person view. With this, a stable view can be obtained for knowing the layout of space and finding target objects for observing efficiently. However, the camera is limited to its fixed position without flexible movement. Although the wide-angle camera can be applied and the camera can be rotated in a given angle range, the dead zone of view and the problem of occlusion are inevitable and difficult to resolve.

**First-Person View.** The actuated monitor is equipped with a camera, and we define the view from this camera as the first-person view. The camera is mounted on a pan-tilt, which can be drove to rotate in two perpendicular orientations. When monitor is employed in a working space, the camera can follow the movement of the monitor to surveilling different places. In addition, the 2DOF (degree of freedom) pan-tilt means even in the fixed position, the direction of vision can also be adjusted to various angle, which provide user with all-around vision.

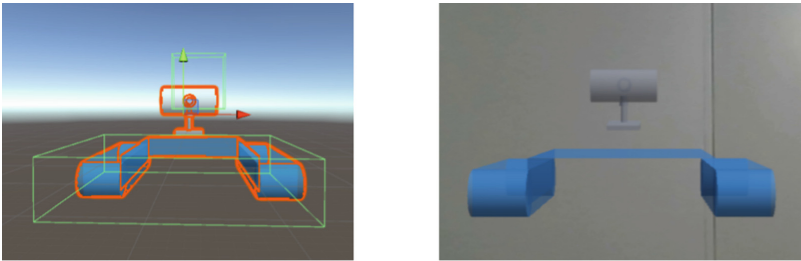
## 2.2 Augmented Reality Model

For intuitive manipulation, in dual point of view the augmented reality models are built on screen to operate the direction and position of two controlled parts on monitor: 2 DOF pan-tilt and caterpillar band.



**Fig. 2.** Augmented reality model in third-person view.

In the third-person view, aiming to indicate the relation between virtual models and controlled parts, the real objects showing on the screen will be overlapped with augmented reality model described in the fixed camera's coordinate system. As shown in Fig. 2, the virtual models are completely superimposed on camera and caterpillar band with the comparable size and same shape. In addition, the virtual model shows in semi-transparent for avoiding shading the vision of user.



**Fig. 3.** Augmented reality model in rear first-person view.

In the first-person view, in order to realize consistent manipulation interface, the same augmented reality model is made use of on screen for operation (Fig. 3). In general, however, in the vision of the monitor-mounted camera, the pan-tilt and caterpillar band cannot show on its view. In our system, the virtual model reveals on screen just like the view from the back side of the first-person view.

Therefore, we rename this view as the rear first-person view. In each operation, the virtual models representing camera and caterpillar band will be place in the fixed location initially for control. Furthermore, when user manipulate the model of caterpillar band, the virtual model of camera will make synchronous motion since these two parts are connected with each other.

In dual point of view, following the control of mouse, the virtual model of camera on screen, which indicates the direction of 2 DOF pan-tilt, can be dragged to up and down for vertical rotation, or right and left for horizontal rotation. In the same way, the model of caterpillar band, revealing the position and direction of the physical monitor, can be controlled to go forward and back, or make right/left handed rotation.

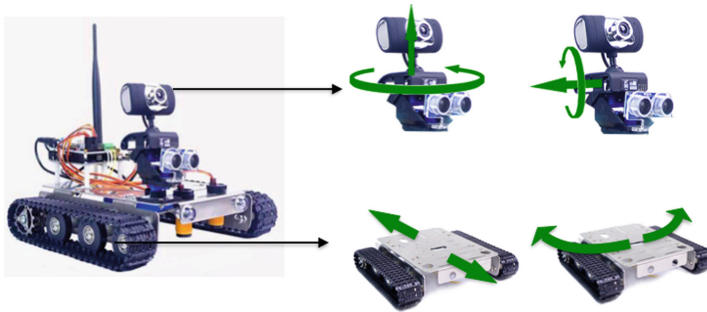
In real world, the controlled parts of monitor follow the on-screen models to make corresponding motion. User can see the results through the live video on screen.

### 3 System Implementation

Our system is an augmented reality based interface for robot manipulation. By using augmented reality technology, the operation of models' position and orientation on screen can be projected to three-dimensional space for physical monitor control. Moreover, the system has two manipulation methods, one is the third-person view manipulation, the other is the first-person view manipulation.

#### 3.1 System Hardware Overview

The hardware of our system consists of the three parts: the actuated monitor, the wireless communication equipment and the host computer. As shown in Fig. 4, the robotic vehicle, whose microprocessor is Arduino, is equipped with a camera connecting with 2 DOF pan-tilt to work as the actuated monitor. The vehicle has a mechanism for locomotion by using caterpillar band, which allows



**Fig. 4.** The actuated monitor.

the monitor to rotate and move forward or backward. In addition, the pan-tilt with two servo motors connecting with camera has 2 degree of freedom, vertical rotation and horizontal rotation. In terms of communication part, bluetooth is used to realize serial communication for transmitting manipulation command from host computer to single chip microcomputer.

### 3.2 Vuforia Object Scanner

In the third-person view manipulation, in order to register the position between the real monitor and virtual model, an Android application, Vuforia Object Scanner, is used to scan a physical 3D object [9].

Once user Launch the Vuforia Object Scanner, the target object need to be placed in the grid region of a paper named Object Scanning Target, and the coordinate system will be shown on the screen. According to the size of the target object, it will be a polyhedron covering on the object. Move the camera around the object to scan the vantage points on the surface of the object. Initially, all surface regions are in gray. When a surface region has been successfully captured, it will turn green. Once all of the surface areas are captured, we can press the stop button to terminate the scanning process. Since then, once the target object comes into the view of camera, it can be recognized automatically and its position and orientation in three-dimensional space can be mapped on screen. In addition, the accuracy rate of recognition is related to the number of vantage point on surface, which means the more vantage points you get, the higher accuracy it will be.

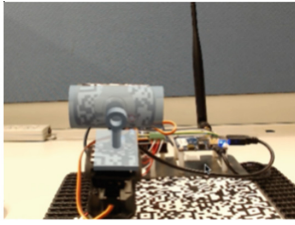
### 3.3 Actuated Monitor Manipulation

The system provide users with two manipulation method: third-person view manipulation and rear first-person view manipulation.

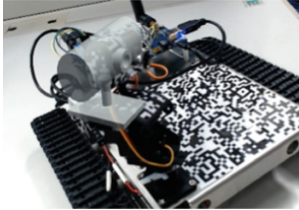
**Third-Person View Manipulation.** After controlling the AR model, the change of direction and position will be transferred to manipulation command for monitor control.

The 2 DOF pan-tilt equipped with servo motors, can realize horizontal rotation about  $120^\circ$  with  $60^\circ$  on right side and  $60^\circ$  left side respectively, and vertical rotation about  $90^\circ$  with  $10^\circ$  on front side and  $80^\circ$  on back side. Once the upper computer receives the rotation data, the angle will be judged into vertical or horizontal. After that, according to the specific angle we drag the corresponding command can be determined for control. In the Fig. 5 we can see that the pan-tilt follow the augmented reality model to rotate in different directions.

The caterpillar band equipped with electric motor, can go forward/back in any distance or horizontal rotation with  $360^\circ$  on right and left side respectively. For translation, the upper computer reads the initial position and current position of AR model in coordinate system in Unity and calculates the movement

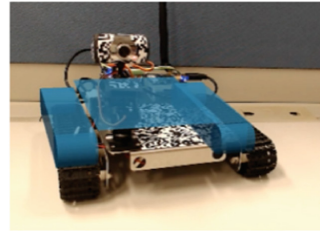
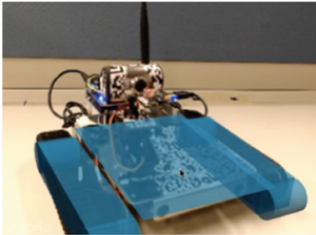


(a) Vertical rotation

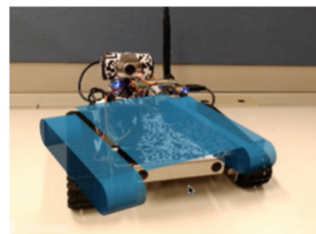
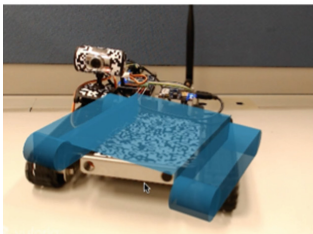


(b) Horizontal rotation

**Fig. 5.** The rotation of 2 DOF pan-tilt



(a) Go forward/back



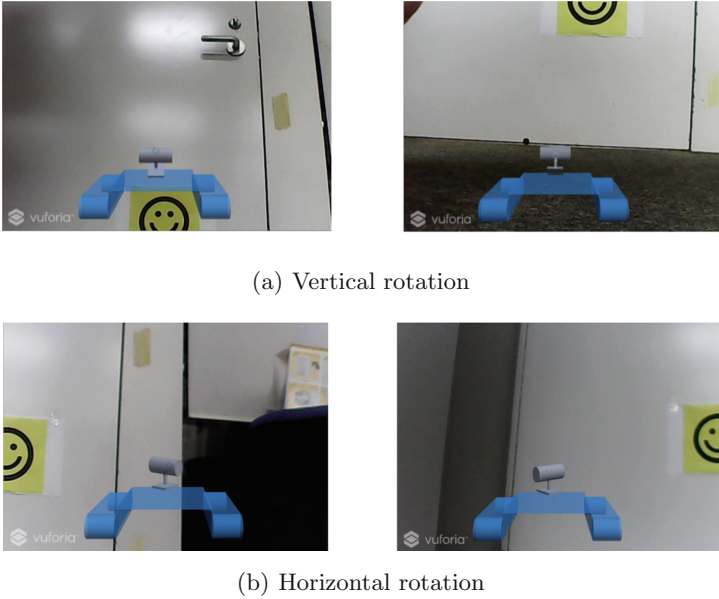
(b) Left/right handed rotation

**Fig. 6.** The rotation and translation of caterpillar band

distance. If the value is positive, it means the model moves to front side. If negative, the model moves to the back side comparing the previous position. Then the upper computer according to the specific value to determine the corresponding command and send to lower computer. For rotation control, the difference value of current orientation and previous orientation will be calculated to positive or negative corresponding to the right handed or left handed. Then based on the specific angle the upper computer will determine the operation command. After each dragging, the new position and direction data will overlap the old one as the origin point in the monitor's coordinates system. The Fig. 6 shows the translation and rotation of caterpillar band in third-person view.

**Rear First-Person View Manipulation.** In the same way, the system realize the rear first-person view manipulation of two controlled part.

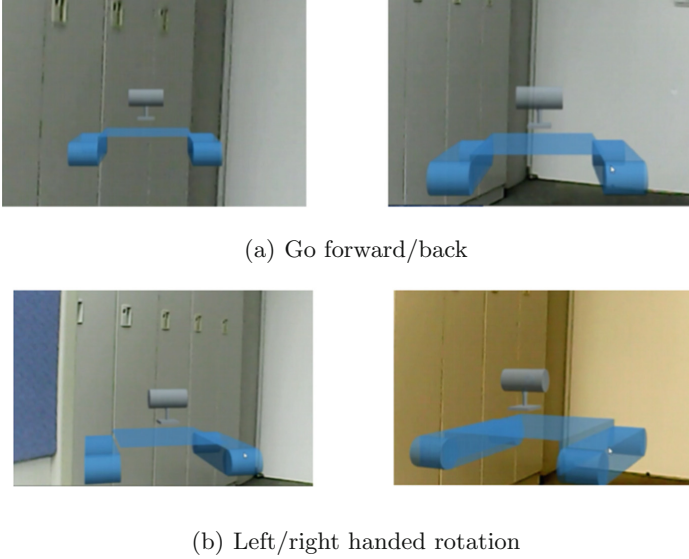
For 2 DOF pan-tilt, we drag the AR model in different direction around two axes, and the change of angle indicates the target direction of camera. As shown in Fig. 7, the smile face shows in different positions in these four pictures, which illustrate the rotation of camera.



**Fig. 7.** The rotation of 2 DOF pan-tilt

Follow the virtual model on screen, in Fig. 8 the caterpillar band can be drove to go forward/back in any distance or horizontal rotation degree with  $360^\circ$  on right and left side respectively.





**Fig. 8.** The rotation and translation of caterpillar band

## 4 Preliminary Evaluation

We conducted a preliminary user experiment to evaluate the performance of our system. Our target is to test whether our system could achieve an intuitive manipulation of monitor by using augmented reality model in dual point of view.

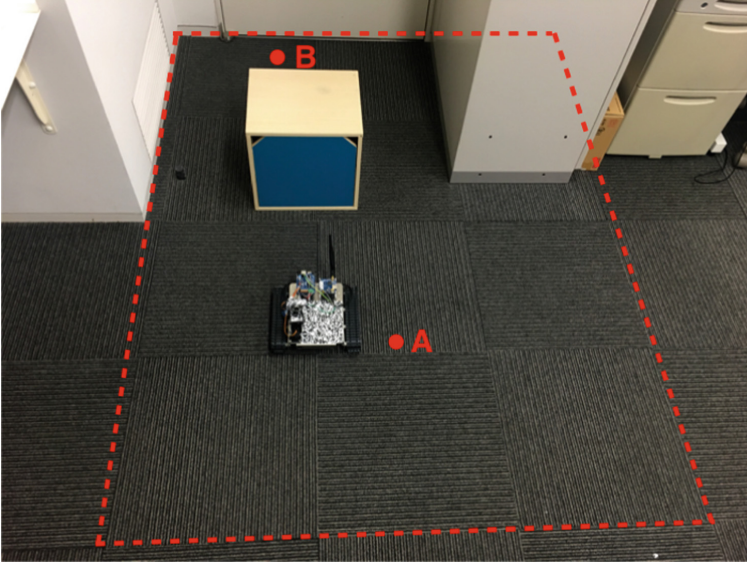
### 4.1 Participants

We recruited 8 participants including 4 females and 4 males. Before experiment, we asked each participant to confirm all of them have the experience of using joystick to control robot, and they were not familiar with our robot. The experiment took approximately 25 min.

### 4.2 Method

We conducted our user experiment in a prepared working space with the monitor like the Fig. 9 shows. The fixed setting camera was placed at 80 cm high from the floor. No participant was allowed to enter or see the working space before experiment, therefore the working space was a completely unknown environment for them. All objects and the monitor were placed in their initial positions for each trial.

When the test began, we explained to all participants how to control the robot, and the DOF of the robot. Before taking the experiment, the participants were asked to practice using our system for about 15 min. During the experiment,



**Fig. 9.** The experimental working space.

participants were divided into 2 groups. In group 1, they were asked to use third-person view manipulation at first and next rear first-person view. In group 2, the participants were allowed to use two point of view in reverse order. All participants in each group were given 10 min to complete the following task:

- (a) Control the caterpillar band to translate from Place A to Place B.
- (b) Control the caterpillar band for horizontal rotation.
- (c) Control the 2 DOF pan-tilt for horizontal rotation.
- (d) Control the 2 DOF pan-tilt for vertical rotation.

After finishing the task, all participants were asked to fill a questionnaire in Table 1. They needed to answer the Q1–Q3 by grading from 1 to 5 (1 = very negative, 5 = very positive) for third-person view manipulation, and Q4–Q6 for rear first-person view manipulation. Finally, in Q7 they need to chose the prefer one in two point of view with three options: (a) third-person view; (b) rear first-person view; (c) using both view.

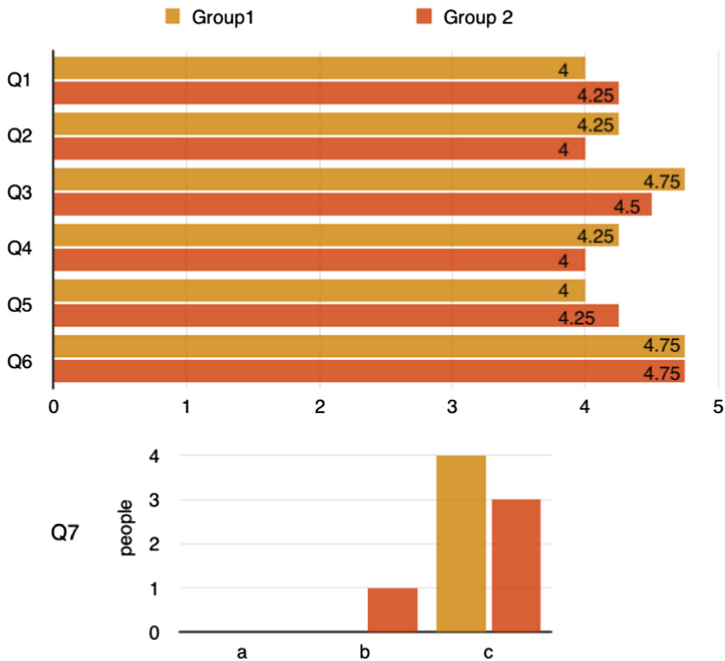
### 4.3 Results

All participants succeeded in completing task within the stipulated time. After collecting the questionnaires result from all participants, we calculated the average scores of each question from participants in each group respectively. The results from the questionnaire are shown in Fig. 10.

Question 1 to question 6 are related to the practicability of augmented reality based manipulation in two point of view. In each question, the average score of

**Table 1.** Questionnaire

Third-person view	
Q1	Do you think this interface is intuitive to manipulate monitor in domestic working space
Q2	Do you think using augmented reality model is better than button to control monitor?
Q3	Do you have difficulty to get accustomed to our system to manipulate monitor?
Rear first-person view	
Q4	Do you think this interface is intuitive to manipulate monitor in domestic working space
Q5	Do you think using augmented reality model is better than button to control monitor?
Q6	Do you have difficulty to get accustomed to our system to manipulate monitor?
Dual point of view	
Q7	which point of view do you prefer to manipulate monitor?

**Questionnaires Result****Fig. 10.** The results of the questionnaire.

two groups are higher than 4 points, which prove that our design can realize intuitive manipulation to some extent.

For question 1 and question 4, the results are opposite. In the post-task interviews we find that it may related to the inverse manipulation order of the two groups. In group 1 participants operated the monitor in the third-person view at first, by which they can understand the entire working space. After that, when they use the rear first-person view later, they can better know the target place for translation. However, for participant in group 2 who did not know the working space before, they may spend longer time to determine target place for translate or rotate. The question 2 and question 5 is related to the comparison between our method with the way of button manipulation. The positive result means our system enable users better manipulation experience than the previous way of joystick control.

Question 3 and question 6 are used to judge the ease of use of our system. The result suggests that the user can easily learn how to use and get accustomed to our system without much difficult. In the post-task interviews, the majority of participants thought other than joystick based control by which user need to know the function of various buttons, using augmented reality model is more direct and simple. Besides, the user indicated that they have successfully learned how to use it in a very short time and our system is very suitable for inexperienced people.

Question 7 is regarding to the preference of two point of view for different people. For 8 participant, only one people selects the rear first-person view manipulation control only. The other people think the combination of dual point of view is better since they can choose more efficient view in different situations.

## 5 Related Work

There have been several studies for intuitive manipulation system to control multi-DOF robot based on augmented reality and mixed reality techniques.

Nawab et al. proposed a system which allows user to understand the mapping of the joystick based controller by overlaying a color-coded coordinate system on the end-effector of the robot by using augmented reality technology [7]. Moreover, this paper reports the positive effects of augmented reality visual cues on operator performance during end-effector controlled teleoperation using only camera views.

Kobayashi et al. developed a novel environment for robot development, in which intermediate results of the system are overlaid on physical space using Mixed Reality technology [6]. Real-time observation enables the developers to see intuitively, in what situation the specific intermediate results are generated, and to understand how results of a component affected the total system. Their method allows the operator to obtain the internal statuses of robot intuitively, which is useful for operation in practical work.

Chen et al. proposed a mixed reality environment for performing robot simulations based on the concept of Mixed Reality [1]. Robot developers can create

scenarios for evaluating robot tasks by mixing virtual objects into a real physical environment to create an MR simulation with varying level of realism. The simulation environment can be displayed to users in both an AR and an AV view.

Drascic et al. developed a display system named ARGOS (Augmented Reality through Graphic Overlays on Stereo-video) [2]. In order to enhancing human-robot interaction, an augmented reality through graphic overlaying on a stereo video. In their design, user controls a robotic arm by operate the virtual cursor superimposed on the video image. A tele-robotic system based on augmented reality to control a robotic arm is introduced by Xiong et al. [10]. They designed a virtual robot works as an interface to control the physical robot by using the operator, which can reduce the time-delay between user operation and the action of robot. In addition, they present the advantages of predictive display. Simulation of virtual robot's tasks in the augmented environment improves the safety of the robot when it executes the planned tasks.

Hashimoto et al. propose "TouchMe", a tele-operating system which allows the user to manipulate a multi-DOF robot intuitively with touch interaction from a third-person view [3]. TouchMe de has two elements to realize more intuitive control: (1) Using a third-person view camera because it allows the user to understand the situation of the entire work space. With this, it is easy to avoid collisions with obstacles on the side or behind the robot when the robot is rotating or moving backwards and clear to find target objects for observing and specify the distance of movement and angle of rotation. (2) Building computer graphics (CG) model superimposed on multi-DOF robot to help the user predict how the robot will move and understand the controllable direction of the mounted part. In this work, the camera captures the image of the workspace in real-time, and the image is shown on the touch screen with a CG model overlaid on the real robot. The user controls the robot by touching the part of the CG model where he/she wants to move, and he/she then drags it to the desired position and direction. Although the system eliminates the difference of view point between user and robot, the TouchMe cannot solve the problem of dead zone manipulation. In addition, four markers (ARToolKit [5]) on the top of robot are used to locate the initial state of the robot and also for visual feedback when the robot moves to the specific position or rotates to specific direction. However, marker recognition can be influenced by the shooting angle of the camera. For example, in the angle that four makers are shield by obstacles, robot cannot be recognized successfully.

Comparing with TouchMe, our system has advantages in following two aspects. Firstly, based on third-person view our system add first-person view provided by a robot-mounted camera, which means when robot moves into the dead zone of the third-person view, user can switch to the first-person view for operation. The two point-of-view manipulation systems are running simultaneously and user can switch according to the different situations. What's more, instead of using marker for registration between the real robot and the CG model, an application named Vuforia Object Scanner are used to scan and rec-

ognize the robot automatically. It means after scanning the target object and getting enough feature points from its surface, when the target object comes into the view of the camera it can be recognized automatically.

## 6 Conclusion

In this paper, we have presented the design, implementation and an preliminary evaluation of an augmented reality based system for controlling a multi-DOF monitor. Our system allows the user to manipulate each part of the monitor by directly dragging corresponding augmented reality models from dual point of view, which refers to the third-person view provided by a fixed setting camera in working space, and the rear first-person view seen by the camera mounted on monitor. In addition, by using augmented reality models we can realize the spatially consistent relation between virtual world and real world, which means the action of target object is consistent in augmented reality models described in the user's coordination system. Our system have received a positive feedback from the preliminary experiment. The result indicates that user could achieve an intuitive manipulation of monitor by using augmented reality model to some extent. Although in this paper we test our system in a working space with simple environment, it is also suitable for other domestic space with more complex environment.

## References

1. Chen, I.Y.-H., MacDonald, B., Wunsche, B.: Mixed reality simulation for mobile robots. In: Proceedings of the International Conference on Robotics and Automation (ICRA 2009), pp. 232–237. IEEE (2009)
2. Drascic, D., Grodski, J.J., Milgram, P., Ruffo, K., Wong, P., Zhai, S.: ARGOS: a display system for augmenting reality. In: Proceedings of the INTERACT and CHI Conference on Human Factors in Computing Systems (CHI 1993), p. 521. ACM (1993)
3. Hashimoto, S., Ishida, A., Inami, M., Igarashi, T.: TouchMe: an augmented reality based remote robot manipulation. In: Proceedings of the International Conference on Artificial Reality and Telexistence (ICAT 2011), pp. 61–66 (2011)
4. Kasahara, S., Niiyama, R., Heun, V., Ishii, H.: exTouch: spatially-aware embodied manipulation of actuated objects mediated by augmented reality, pp. 223–228 (2013)
5. Kato, H., Billinghurst, M.: Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In: Proceedings of the IEEE and ACM International Workshop on Augmented Reality (IWAR 1999), pp. 85–94. IEEE (1999)
6. Kobayashi, K., Nishiwaki, K., Uchiyama, S., Yamamoto, H., Kagami, S., Kanade, T.: Overlay what humanoid robot perceives and thinks to the real-world by mixed reality system. In: Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2007), pp. 1–2. IEEE Computer Society (2007)

7. Nawab, A., Chintamani, K., Ellis, D., Auner, G., Pandya, A.: Joystick mapped augmented reality cues for end-effector controlled tele-operated robots. In Proceedings of the IEEE Conference on Virtual Reality (VR 2007), pp. 263–266. IEEE (2007)
8. Rossi, A., Dautenhahn, K., Koay, K.L., Walters, M.L.: Human perceptions of the severity of domestic robot errors. In: Kheddar, A., et al. (eds.) ICSR 2017. LNCS, vol. 10652, pp. 647–656. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-70022-9\\_64](https://doi.org/10.1007/978-3-319-70022-9_64)
9. Vuforia Object Scanner (2018). <https://library.vuforia.com/articles/Training/Vuforia-Object-Scanner-Users-Guide>
10. Xiong, Y., Li, S., Xie, M.: Predictive display and interaction of telerobots based on augmented reality. *Robotica* **24**(4), 447–453 (2006)