Augmented Reality Based Actuated Monitor Manipulation from Dual Point of View

44161627-8 REN YING

Supervisor: Prof. Jiro TANAKA

Graduate School of Information, Production and Systems
Waseda University

A thesis submitted for the degree of
Master of Engineering

September 2018

Abstract

Recent years, as movable monitors are increasingly used in domestic space for room surveillance, they require simple, intuitive manipulation system to control their motion. Some previous researchers have realized mobile monitor manipulation. However, the previous way of joystick based controller dose not realize direct and intuitive manipulation. In addition, the way of using single viewpoint, third-person view or first-person view, for monitor control may have respective disadvantages in different situations.

In this study, we aim to propose a monitor manipulation system which enables users to control each part of actuated monitor intuitively with interaction of augmented reality models from dual point of view. Third-person view, provided by a fixed setting camera in working space, allows user to understand the surrounding environment of monitor for specifying the distance of movement or direction of rotation and avoiding collision with obstacles. First-person view, seen by the camera mounted on monitor, can work in some specific situations, such as the dead zone of third-person view. Moreover, augmented reality models are applied into monitor control which allow users to manipulate each part of monitor by directly dragging the corresponding models to target position or direction on dual point of view. With these models, we can realize the spatially consistent relation between virtual world and real world.

Through this system, we have realized the concept of user-centered manipulation, which means the action of target object is described in user's coordination system. Thus users do not need to consider the position difference between controller and monitor which can avoid mapping errors between users' input and monitor's resulting movement. In addition, the controller using a portable computer has advantages over the conventional switch/button interface since it provides unlimited interaction designed by software. Lastly, 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, dual point of view, augmented reality, direct manipulation

Acknowledgements

Firstly, I would like to express my deepest appreciation to my supervisor,Prof. Tanaka. For me, he is such a knowledgeable mentor, who provides me with many professional instructions and guide me to build research thinking independently. With his instruction, I know how to explain a research problem or method in a academic way, how to use figures and exact words to support my research, how to present my topic with passion and regulate my paper in a more standard way. In the process of working my graduate project, Prof. Tanaka office was always open whenever I ran into a trouble spot or had a question about my research or writing. It is my great honor to be supervised by Prof. Tanaka. With his profound viewpoints and thoughtful suggestions, I am able to do my research and write my thesis. In addition, I would like to thank the volunteers who joined the experiment of this research. Without their passionate participation and feedback, the evaluation of the system could not have been successfully conducted.

Finally, I would like to thank my classmate for providing me with strong support and continuous encouragement throughout the process of system working and writing thesis. This accomplishment would not have been possible without them. For me, The time spent together with them is precious and unforgettable.

Table of contents

List of figures

Chapter 1

Introduction

1.1 Domestic Robot Control

Recent years, the mobile robots are entering into domestic setting for service utilization[1]. The increasing presence of these robots in daily living activities requires the development of more natural Human-Robot Interaction capabilities [2]. For instance, robot with pan-tilt and caterpillar band generally has a multi-DOF (degree of freedom) so these products require not only simple on/off control but also complicate motion control in space[3]. Joystick based controllers are commonly used to control robot. Recent years, however, robot control by mobile phone and PC is widely used since it enables human to connect with the robot via mobile communication network regardless of time and space (Fig 1.1). In addition, compared to conventional switch/button controller, video-based interface like smartphone and PC can provide unlimited interaction designed by software [1].

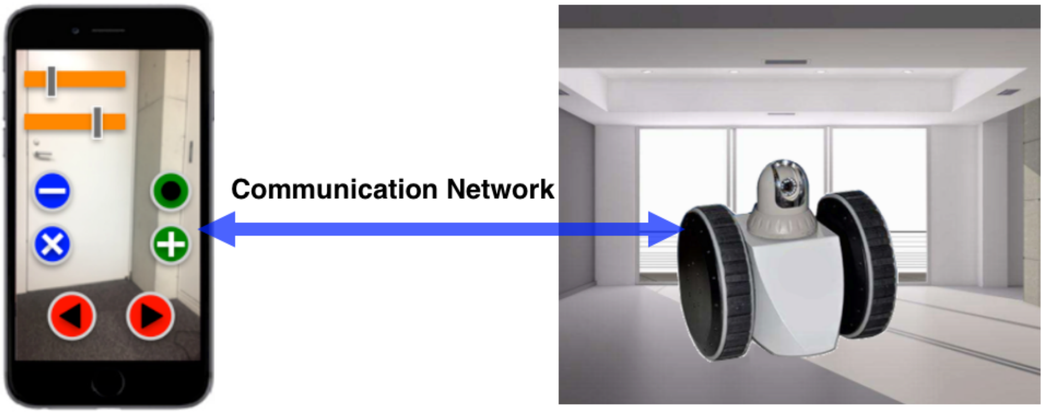


Fig. 1.1 Domestic robot manipulation system

There are various point of view used for controlling robot in video-based interfaces. Top-down view has been widely used since it is easy for understanding the locomotion of the working space. Sakamoto et al. proposed a video-based Tablet PC interface to command home robots by using stroke gestures on a computer screen [4]. In this system, ceiling mounted cameras provide the user a top-down view. On the one hand, it can give users a more realistic feeling of controlling the robot. On the other hand, top-down view is useful for detecting the location of robots and objects because it is not necessary to apply coordinate transformation. Kato et al. developed a multi-touch interface with a top-down view from a ceiling camera for controlling multiple mobile robots,with which users can see the global situation at once[5]. Guo et al. presented two innovative interfaces that allow a single operator to interact with a group of remote robots using toys and touch on a large tabletop display showing a top-down view of the working space in real time[6].

Besides, a number of interfaces for robot control are based on the first-person view (robot's-eye view). Fritsche et al proposed using augmented reality to provide the operator with first-person vision and a natural interface to directly control the camera, and at the same time the robot [7]. Pittman et al explored the capabilities of head tracking combined with head mounted displays as an input for robot navigation [8]. In this system, the first-person view is used to make their controls more meaningful to users given the difference in relative viewpoints. Lui et al proposed a novel semi-autonomous control framework for two mobile robots teleoperated by a single remote user via a first-person view camera attached on the over-seeing wheeled mobile robot with some visuo-haptic feedback[9]. Correa et al. proposed a multimodal framework for interacting with an autonomous robotic forklift through a combination of spoken utterances and sketched gestures on the robot's-eye view displayed on the interface [10].

Third-person view is also used in video-based remote robot control systems. Ruffaldi et al presented a stereo augmented reality eye-wear integrated in a working helmet for HRC with a humanoid robot for collaborative applications,which creates a third point of view augmented reality feedback based on robot state[11]. Hosoi et al. proposed a system called Shepherd that allows a user to manipulate robots such as radio controlled cars from his viewpoint[12]. In this system, the user holds a mobile device in his/her hand which can capture images of robots through a camera mounted on a mobile device.

1.2 Augmented Reality

1.2.1 Overview

Augmented reality is the technology which integrates digital information with the user's environment in real time. Unlike virtual reality, which creates a totally artificial environment, augmented reality uses the existing environment and overlays new information on top of it. With the help of AR technology, the information about surrounding real world becomes interactive and digitally manipulable [13].

Augmented reality system has three prominent features[14]:

- combines real and virtual objects in a real environment
- registers real and virtual objects with each other
- runs interactively, in three dimensions, and in real time

Currently, there are two general definitions of augmented reality. Ronald Azuma, from University of North Carolina, believes that augmented reality includes three aspects: combining virtual object with reality, real time interaction and three dimension[15]. Paul

Milgram Fumio Kishino put forward another definition in 1994, which is named Milgram's Reality-Virtuality continuum[16].

1.2.2 Technology

Commonly, augmented reality is the combination of the hardware and software. Currently, many software development kits have appeared to promote the development of augmented reality applications. In the aspect of hardware, an AR platform combines processors, displays, sensors, and input devices.

Various technologies are used in augmented reality rendering. For example, Head-mounted displays (HMDs) are commonly used to improve the user's sense of immersion.A head-mounted display renders images of both the physical world and virtual objects over the user's field of view and adjust accordingly with the user's head movements.

Chapter 2

Research Goal and Approach

2.1 Target Problem

As we known, mobile robot control is a fundamentally important problem in the robotics society. Some previous researchers have designed various video-based interfaces for robot control from single view, like third-person view or the first-person view. However, we find that these works have some problems.

In the third-person view provided by the fixed setting camera, the first problem is related to the difference between human's view point and the robot's view point. When users want to move a robot to the left from the user's viewpoint, they need to consider the relative position and direction between the robot and themselves. For example, When users stand with the robot face-to-face, in the view of user, the right of themselves is the left of robot. Especially for inexperienced users, the mapping between users' input and resulting robot motion raises the level of their cognitive load, and even cause mapping errors in operation commands. The second problem is the manipulation in dead zone of the view from the fixed camera. In the invisible area of the camera, it is hard to control the robot just from only one view.

In the first-person view, catching the sight of working space is not so easy since the narrow view provided by the robot-mounted camera. Therefore, on the one hand, collisions with obstacles on the side or behind robot when moving seems unavoidable. On the other hand, it may cost more time to find target objects for observing and specifying the distance of movement and angle of rotation.

The other problem is about manipulation interface. The most popular method for control multi-DOF robots is the joystick based controller. However, for inexperienced users they may demand training time to understand the functions of different buttons. Moreover, for joystick, the number of controllable DOF is limited by the number of buttons. The controllable DOF can be increased by using some combinations of 2 or 3 keys, but it will make operation more difficult and demand a longer training time for users. In addition, for general joystick based controller, the moving velocity of the robot is related in proportion to the timing or degree of key pressing. This also not so intuitive especially for inexperienced user.

2.2 Research Goal

In this study, we aim to 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. In addition to the third person view captured by the fixed setting camera in working space, the monitor-mounted camera provides the first-person view which can be used for manipulation in dead zone. Moreover, the way of augmented reality based manipulation realizes intuitive operation since users can control each part of robot by directly control augmented reality model instead of understanding functions of different buttons for manipulation. Furthermore, augmented reality based manipulation can eliminate the difference of view point between users and robot because user do not need to consider relative position and direction between the robot and themselves and just need to drag the augmented reality model to target place or direction.

2.3 Research Approach

As the Fig 2.1 shows,we propose a robot control system which allows users to manipulate a multi-DOF robot intuitively from dual point of view. 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.

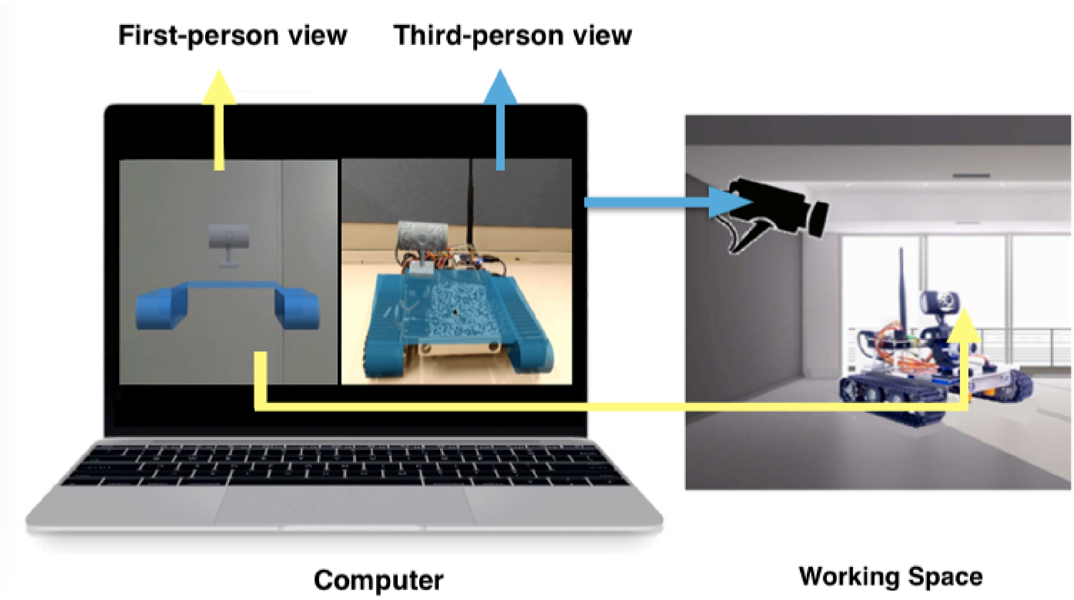


Fig. 2.1 Domestic robot manipulation system

In the third-person view, users can catch the sight of entire working space including controlled monitor and obstacles, which is not only easy to avoid collisions but also easy to find target objects for observing. For controlling monitor, the first step is the recognition of controlled part of monitor in view point. After that, the controlled objects showing on the screen are superimposed with augmented reality models which build in virtual coordinate system. As users drugging the AR model directly to the desired position or direction, the changes of position and direction can be detected and transferred to corresponding manipulation command. Through wireless communication equipment, the chip on monitor receives the command and translate it to the signal that can drive the motor or engine on monitor to follow the AR model to move and rotate.

However, in specific situation such as dead zone of the third-person view, user need to switch to first-person view for manipulation. Taking out the step of object recognition, the AR models representing different controlled parts of monitor will be built and shown in the fixed position of the first-person view. The user select and drag the different AR models to rotate or move. In the virtual coordinate system, the difference value between initial position or direction and current position or direction can be calculated for determining corresponding manipulation command. Through wireless communication equipment, the lower computer on monitor receives the command and control the monitor to follow the AR models to move and rotate.

Chapter 3

System Design

3.1 System Overview

Our system is an augmented reality manipulation interface for robot control. The system has two manipulation methods, one is the third-person view manipulation. The fixed camera captures the image of the working space in real-time, which shows on the screen with AR models overlaid on the real monitor. Following users' control, the system drives the monitor to match the motion of AR models. Similarly, in the first-person view manipulation, AR models shown in fixed position of view is used to determine the motion of monitor.

3.2 The Software Structure of System

The software structure of system in Fig 3.1 is composed of two parts, one is AR model control, the other is actuated monitor control. In the first part, AR models representing pan-tilt and caterpillar band will be superimposed on the dual point of view which can be drag to move and rotate. The changes of position and direction of AR model in virtual coordinate will be stored and written into the shared file, which can be read by the upper computer. In the section of actuated monitor control, upper computer read the position and direction data from the shared file. Then the data will be calculated and transformed into the manipulation command. Through wireless communication equipment, the command is received by lower computer on monitor for controlling pan-tilt and caterpillar band respectively.

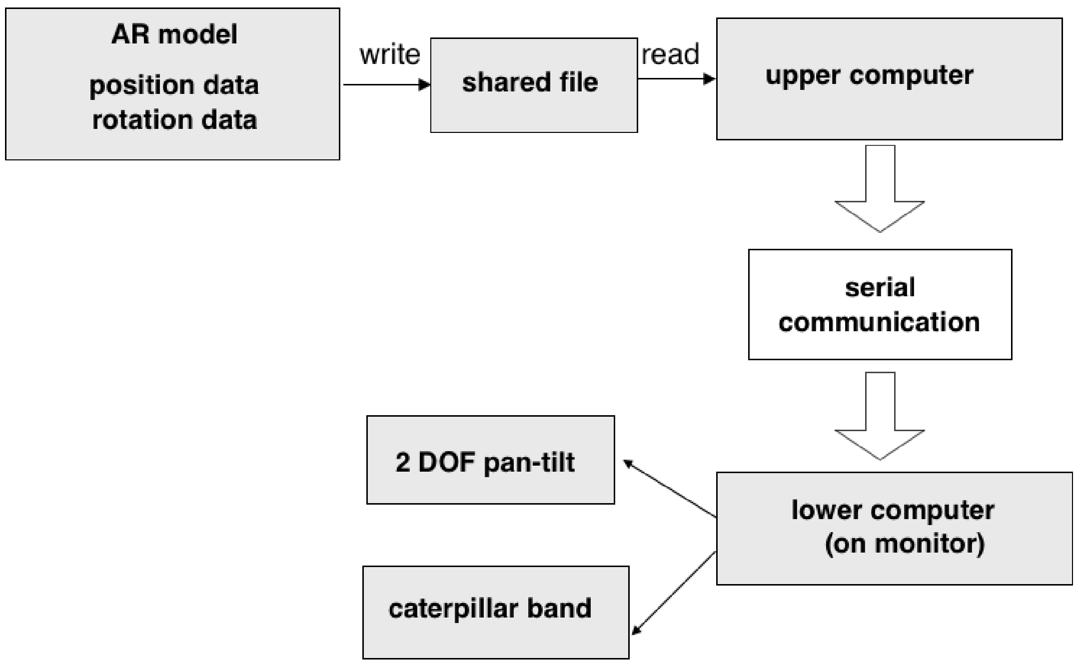


Fig. 3.1 The software structure of the system

3.3 The Hardware Structure of System

The hardware structure of the system (Fig 3.2) contains two parts, one refers to the actuate monitor, which consists of single chip microcomputer, camera, servo motor, caterpillar band and electric motor, the other is bluetooth module for wireless communication. In order to realize vertical rotation and horizontal rotation of camera , two servo motors are used to assemble 2 DOF pan-tilt. In addition, the caterpillar band is equipped with electric motor which can make translation and left/right handed rotation. In terms of communication part, bluetooth ia used to realize serial communication for transmitting manipulation command from host computer to single chip microcomputer

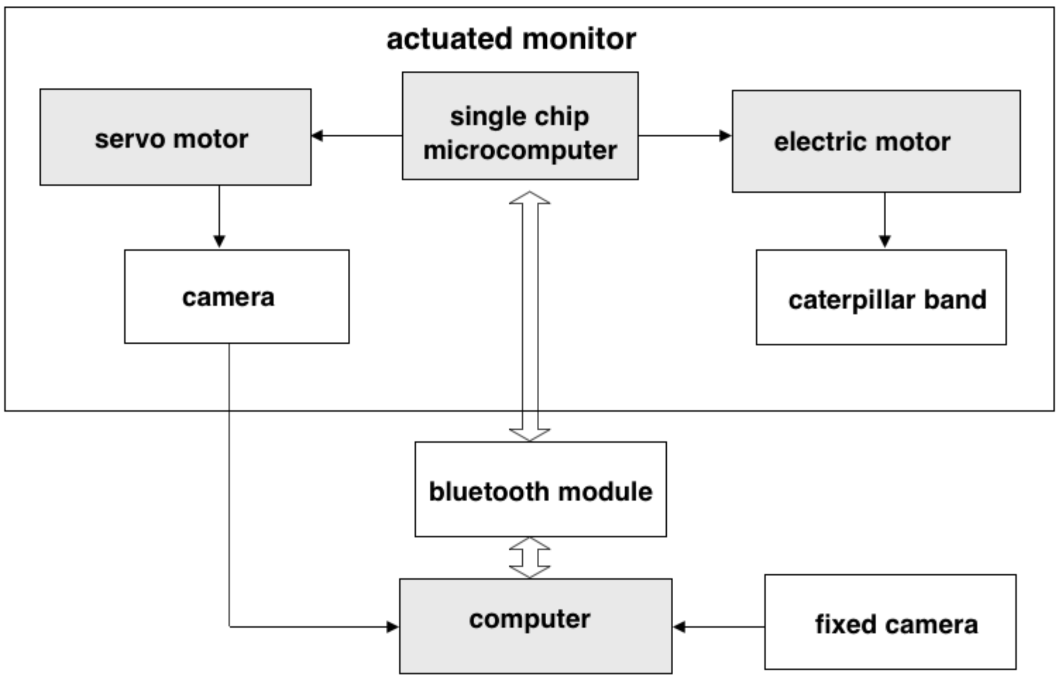


Fig. 3.2 The hardware structure of system

3.4 Dual Point of View

Our system allows monitor manipulation from two point of view : third person view and rear first-person view.

3.4.1 First-person View

We define the view from the camera on monitor as the first-person view (shown in Fig 3.3). When monitor is employed in a working space, the mounted camera can follow the movement of the monitor to surveilling different places. In addition, the 2DOF pan-tile means even in the fixed position, the direction of vision can also be adjusted to various angle, which provide user with all around vision.

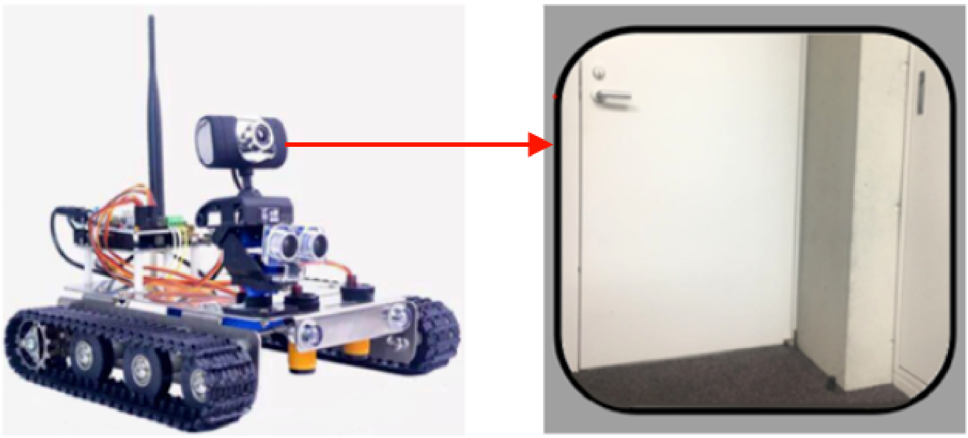


Fig. 3.3 First-person view

3.4.2 Third-person View

we install a fixed camera in the working space. The view from this camera refers to the third-person view (shown in Fig 3.4).The advantage of fixed camera is that it gives the user a good stable view for understanding the entire surrounding environment to find target objects for observing. However, the movement of the fixed camera is limited, making it difficult to solve the problem of dead zone.

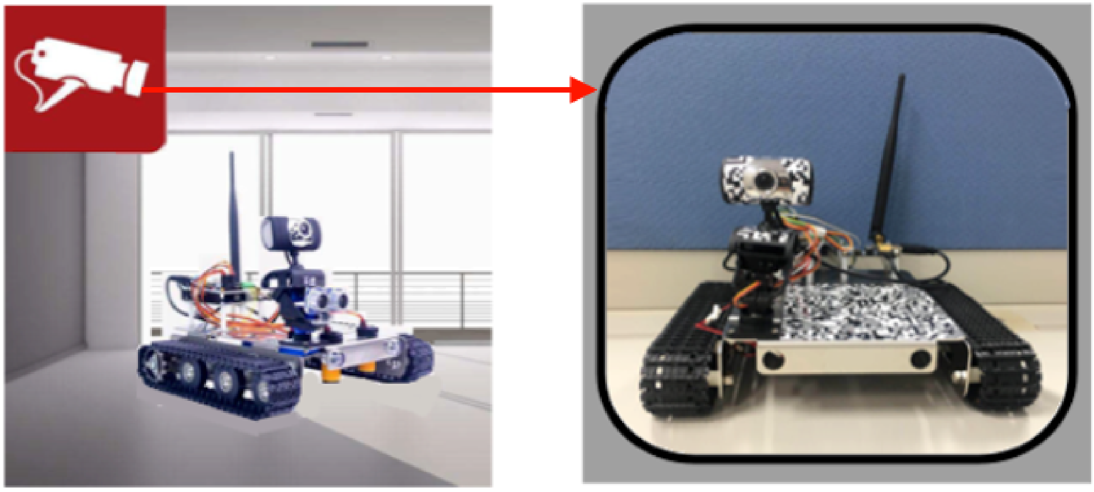


Fig. 3.4 Third-person view

3.5 The Controlled Part of Actuated Monitor

In our design, there are two controlled parts of monitor : 2 DOF (degree of freedom) pan-tilt and caterpillar band.

In order to adjust the vision of camera, the 2 DOF pan-tilt is connecting for change its direction. Two servo motors in pan-tilt are installed perpendicularly to each other, in which way the pan-tilt can make horizontal rotation or vertical rotation around different axises(Fig 3.5). Moreover, two motors can not work simultaneously, which means the pan/tilt can only rotate around one axis at a time.

Fig. 3.5 The two-direction rotation of pan-tilt

The caterpillar band, equipped with electric motor, can be controlled to go forward/back and rotate in left/right handed (Fig 3.6). For translation, two caterpillar bands rotate in the same direction at the same speed. However, for rotation, in order to decrease the turning radius, two caterpillar bands rotate in opposite direction at the same speed.

Fig. 3.6 The translation and rotation of caterpillar band

3.6 Augmented Reality Model Manipulation

3.6.1 Augmented Reality Model

For intuitive manipulation, the augmented reality models are built in dual point of view for determining the direction and position of the controlled part.

In the third-person view, when the camera and caterpillar band come into the view of the fixed camera, they can be recognized automatically. Aiming to indicate the relation between virtual models and controlled parts, the real objects showing on the live video will be overlapped with AR model. The AR model is always described in the fixed camera's coordinate system independent from the object posture. The Figure 3.7 shows the AR models of the camera and caterpillar band respectively, with the the same shape and size of the real one. In addition, the AR models are set to semi-transparent material to avoid shading vision.

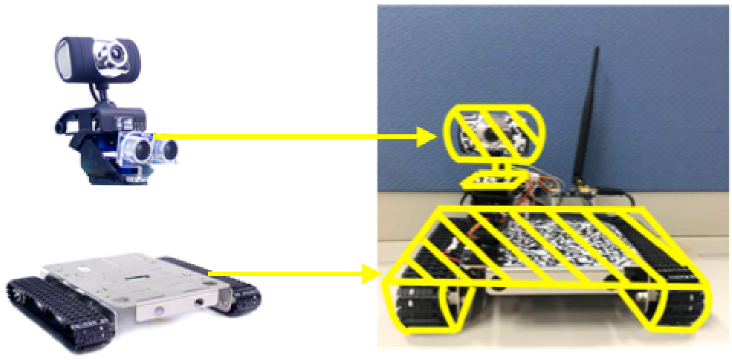


Fig. 3.7 Augmented Reality model in third-person view

In the first-person view, since the angle of the mounted camera, the pan-tilt and caterpillar band can not show on the view. However, we need to build semi-transparent models with the same shape of the real objects for control ((shown in Fig 3.8)). Therefore, we renamed this view as the rear first-person view.

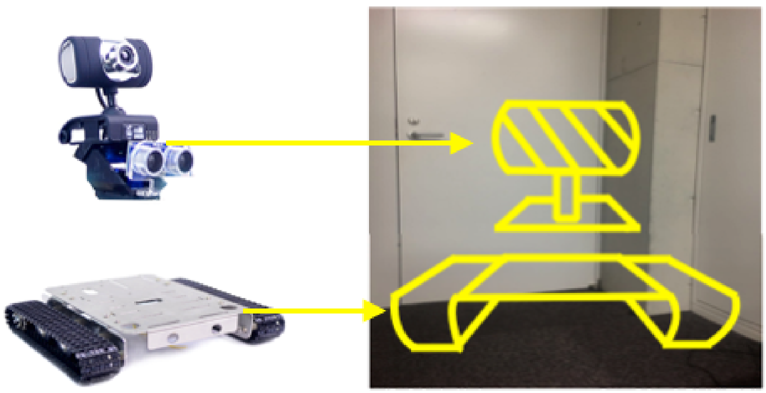


Fig. 3.8 Augmented Reality model in rear first-person view

3.6.2 Model Manipulation

On both of view, AR models can be dragged to make same motion and rotation as the actuated monitor.

In the third-person view, because the coordinate system of AR models is independent from the users' coordinate system, we do not need to consider the relative direction of the monitor and user. As shown in Fig 3.9, the model representing pan-tilt can be dragged to make horizontal or vertical rotation around different axises. In the same way, the model representing caterpillar band can go forward/back and rotate in left/right/handed.

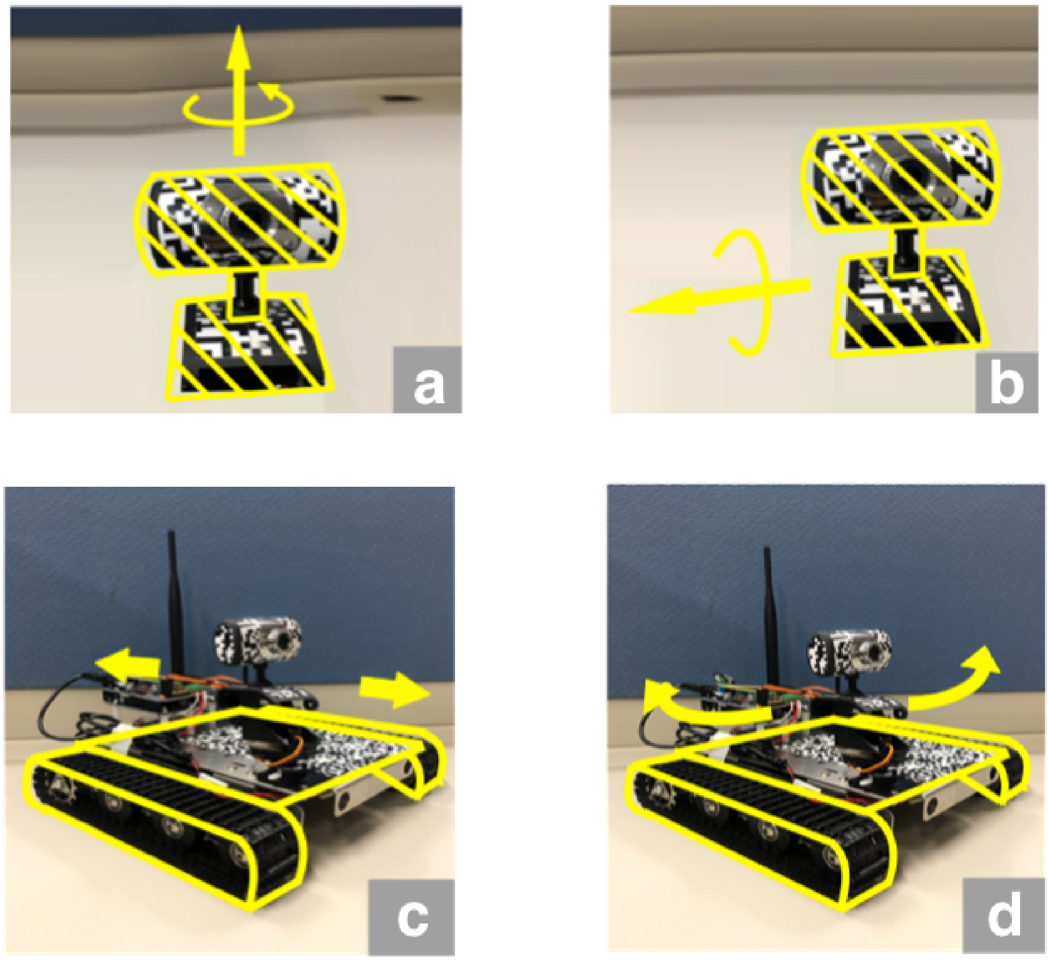


Fig. 3.9 The augmented reality model manipulation in third-person view (a)horizontal rotation of camera (b)vertical rotation of camera (c)translation of caterpillar band (d) left/right handed rotation of caterpillar band

In the first-person view, the coordinate system of the AR model is same as the users' coordinate system. As shown in Fig 3.10, the model representing pan-tilt can be moved up or down for vertical rotation, and right or left for horizontal rotation. In the same way, the user control the AR model representing caterpillar to translate or rotate in left/right handed .

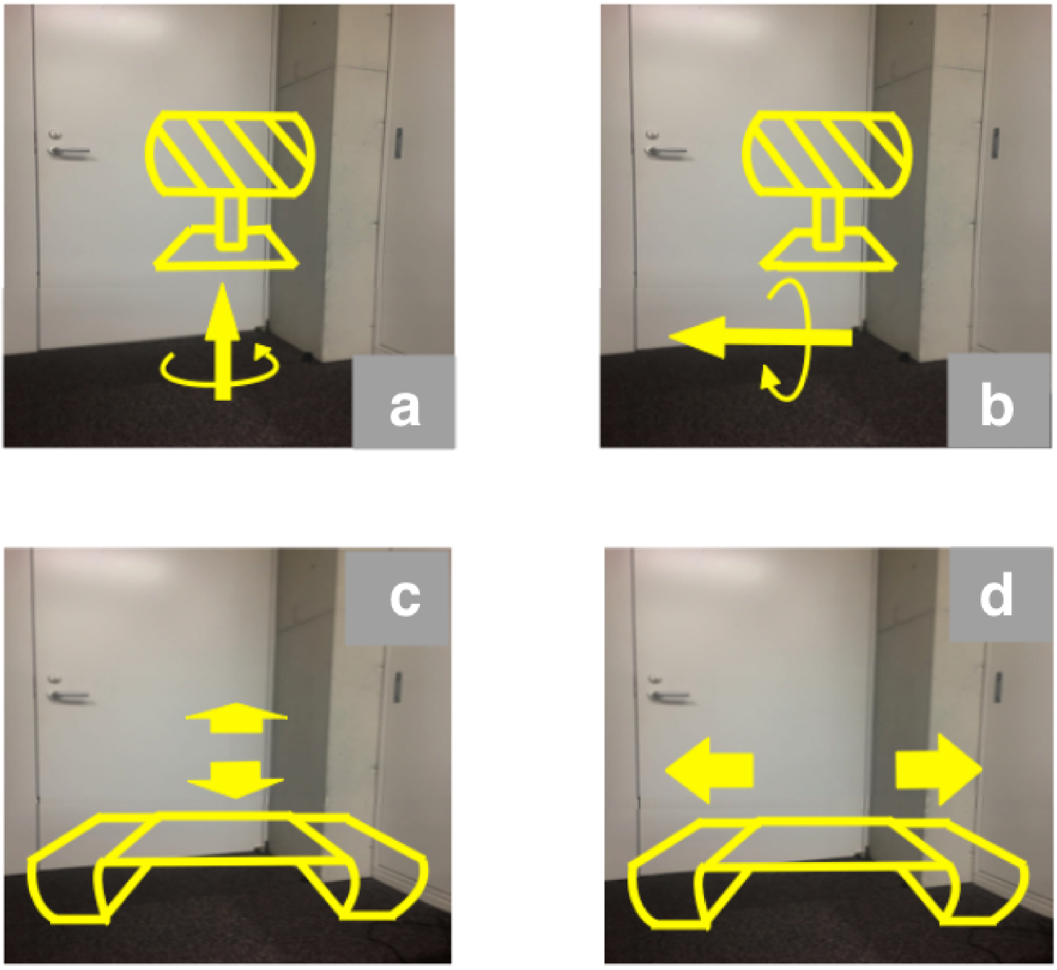


Fig. 3.10 The augmented reality model manipulation in first-person view (a)horizontal rotation of camera (b)vertical rotation of camera (c)translation of caterpillar band (d) left/right handed rotation of caterpillar band

3.7 Wireless Communication

In terms of the communication part, we aim to transmit manipulation command and device status between monitor and upper computer by bluetooth module. After simple pairing between two bluetooth, the data we want to transmit will be divided into multiple packets. The number of port and transition speed are set in the upper computer, and the transmission method of data package is serial transmission. The wireless transmission flow of system shows in the Fig 3.11.

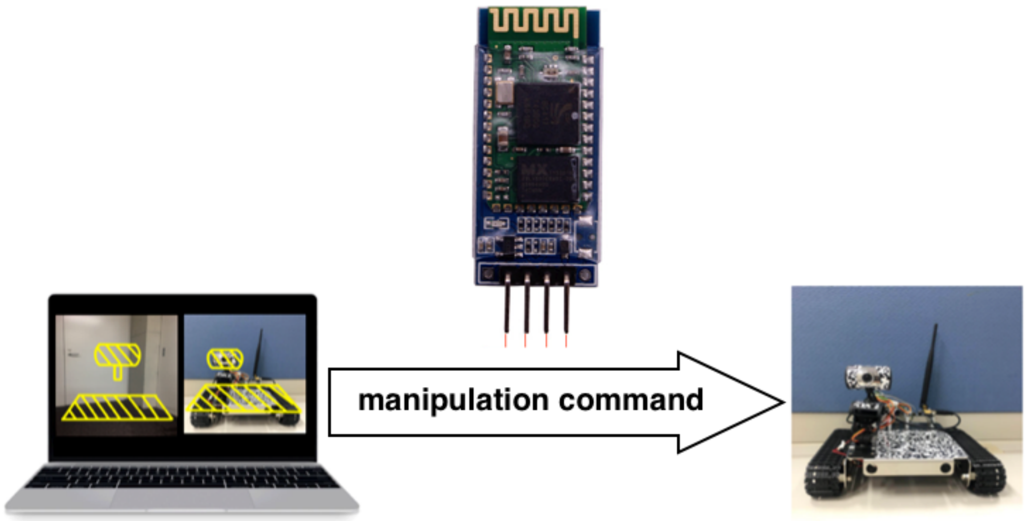


Fig. 3.11 Wireless communication

Chapter 4

Implementation

4.1 Hardware Overview

The hardware of our system consists of the three parts: the actuated monitor, the wireless communication equipment and the host computer. Figure 4.1 shows the system hardware and control flow in system.

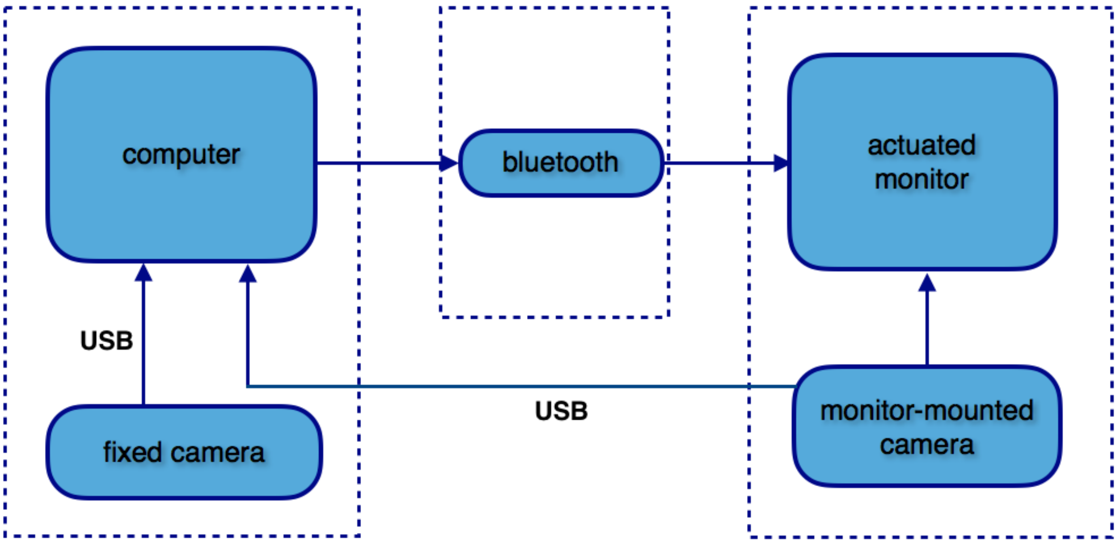


Fig. 4.1 The hardware overview of system

The robotic vehicle, whose microprocessor is Arduino, is equipped with a camera connecting 2DOF pan-tilt to work as the actuated monitor. Fig 4.2 shows our monitor and its

DOF. The vehicle has a mechanism for locomotion using caterpillar band which allows the monitor to rotate and move forward or backward (2DOF). In addition, the pan-tilt with two servo motors connecting with camera has 2DOF, vertical rotation and horizontal rotation.

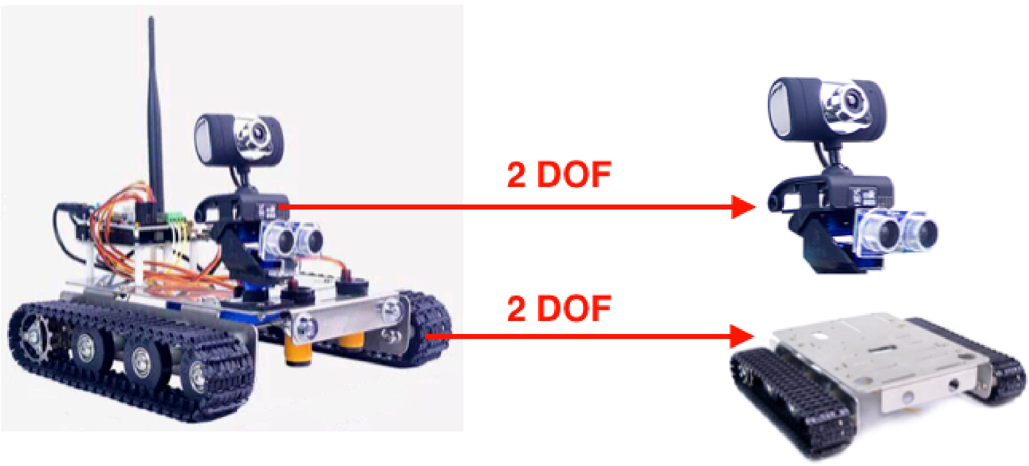


Fig. 4.2 The monitor and its DOF

4.2 Development Environment

In the section of augmented reality, We develop this part using Unity 4.0 and programming language is C#. In terms of monitor control, the upper computer is developed in Visual Studio in C#.

4.3 Third-person View Manipulation

The structure of third-person view manipulation is shown in Fig 4.3. The live video captured from the fixed camera sends to Unity in real-time. By using Vuforia Object Scanner, the controlled objects will be scanned into a package imported into Unity, which means the object can be recognized automatically when it comes into the view of camera. Then the AR models are built and superimposed on live video. As users dragging the AR models, the translation data and rotation data will be sent to upper computer, in where the data will be transformed to corresponding manipulation command. In the monitor, the lower computer on chip receives the command from upper computer by bluetooth module and translate to the signals for controlling pan-tilt and caterpillar band.

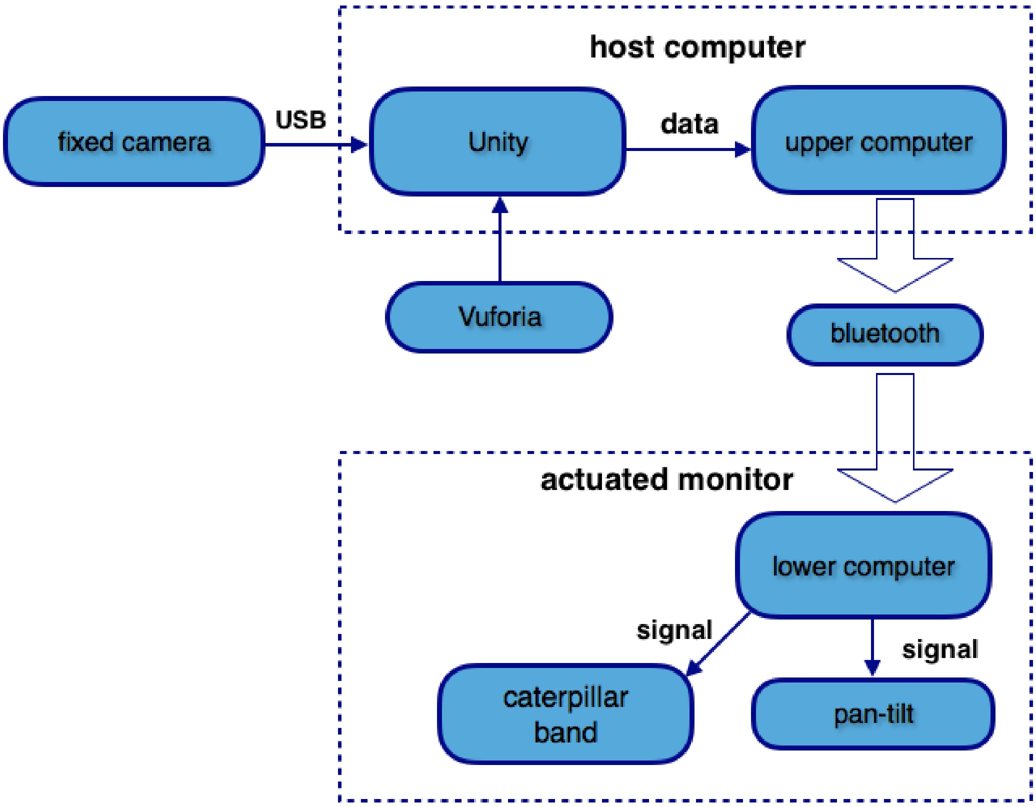
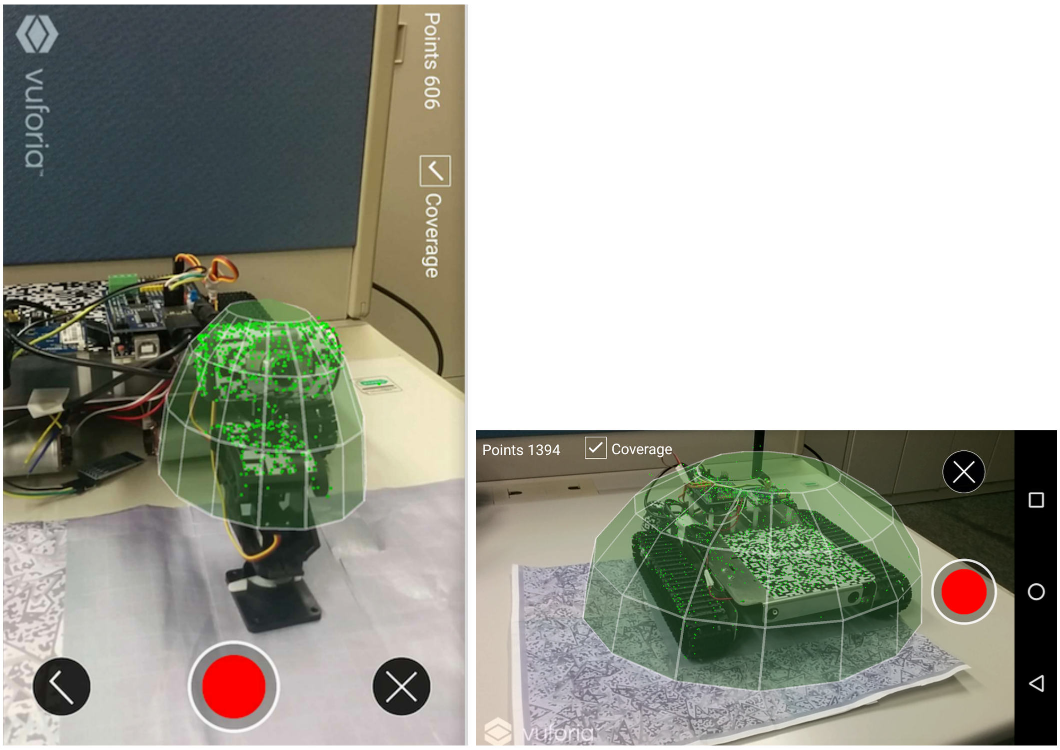


Fig. 4.3 The structure of third-person view manipulation

4.3.1 Vuforia Object Scanner

The Vuforia Object Scanner is an Android application that is used to scan a physical 3D object[17]. The Object Scanner produces an Object Data file including the source data required to define an Object Target. In addition. The Object Data file can also be imported into Unity for recognizing objects automatically.

As shown in Fig 4.4, once users Launch the the Vuforia Object Scanner, the target object need to be placed in the grid region of the 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 capture the vantage points on the surface of the object. Initially, all surfaces 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. From the picture we know that after all areas of polyhedron full turn to green, the vantage points of camera is 606, and vantage points of caterpillar band is 1394.

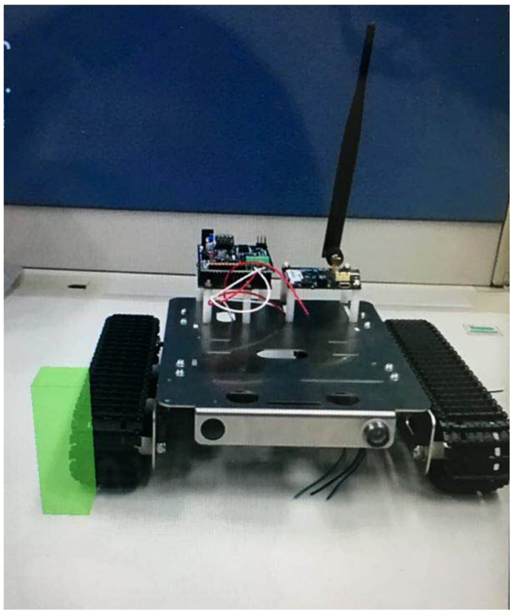


(a) The scan of camera (b) The scan of caterpillar band

Fig. 4.4 The scan process of controlled objects

After scanning, the results can be tested by pressing the test button. As showing in Fig 4.5, when we use the camera on smartphone to capture the camera and caterpillar band again, it will present an augmentation (green cube) when the object is recognized successfully. In addition, the green cube shows in different size and direction since the corresponding distance and direction between target object and camera on smartphone. The Object Data file can be download from database as a package and import into Unity. So when the Unity project is running, the objects will be recognized in the view of camera.

(a) The recognition test of camera

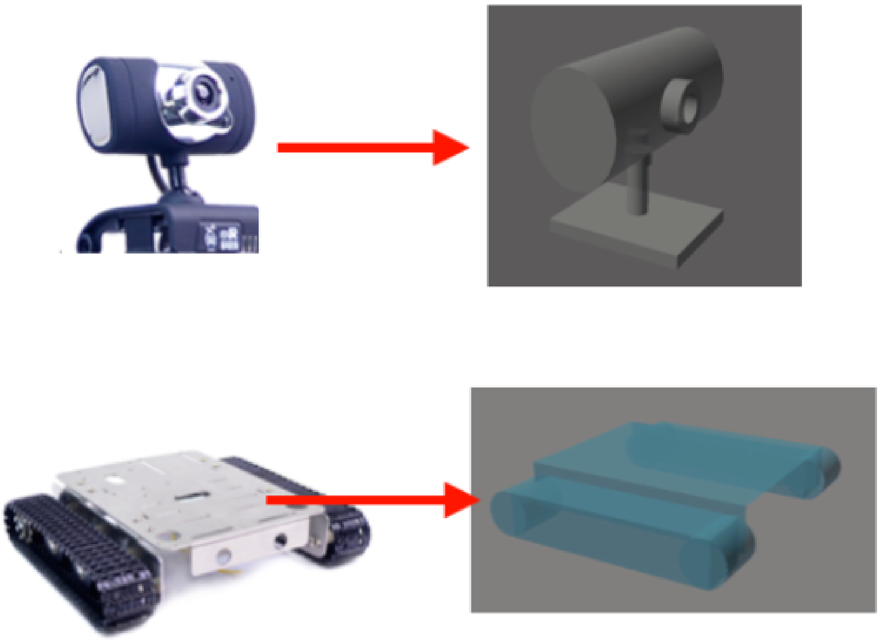


(b) The recognition test of caterpillar band

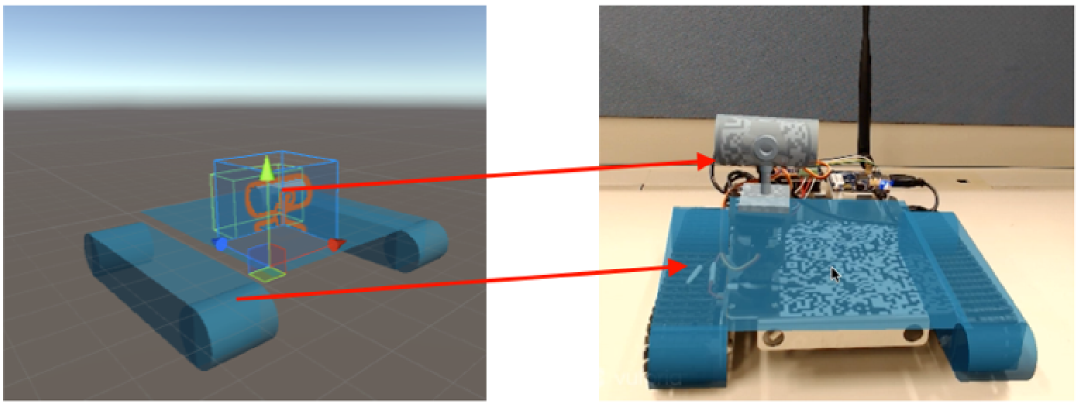
Fig. 4.5 The recognition test of controlled objects

4.3.2 Augmented Reality Model manipulation

After object recognition, the AR models superimposed on camera and caterpillar band can be built in Unity. MAYA, a 3D computer graphics application is used to draw AR model which can be imported into Unity in the format of FBX. In order to indicate the controlling relation between AR model and real object, the AR model will be completely overlapped on the camera and caterpillar band with the same size and shape. In addition, the AR model shows in semi-transparent for avoiding shading the vision of user. The Fig 4.6 shows the AR model superimposed on camera and caterpillar band.



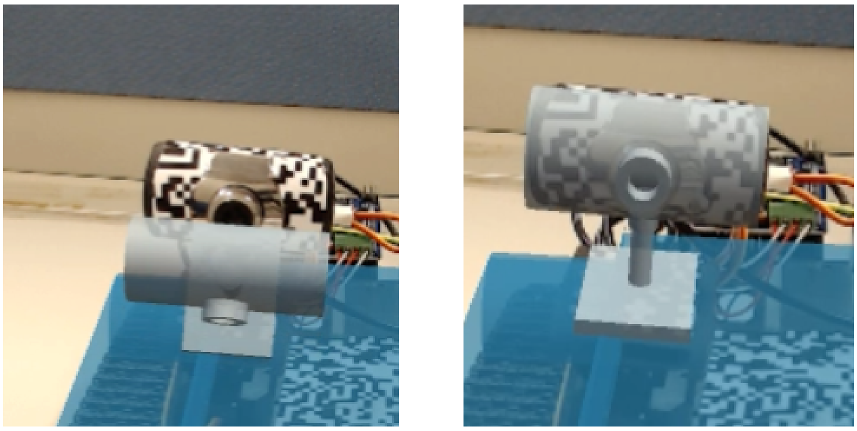
(a) The augmented reality model of camera and caterpillar band



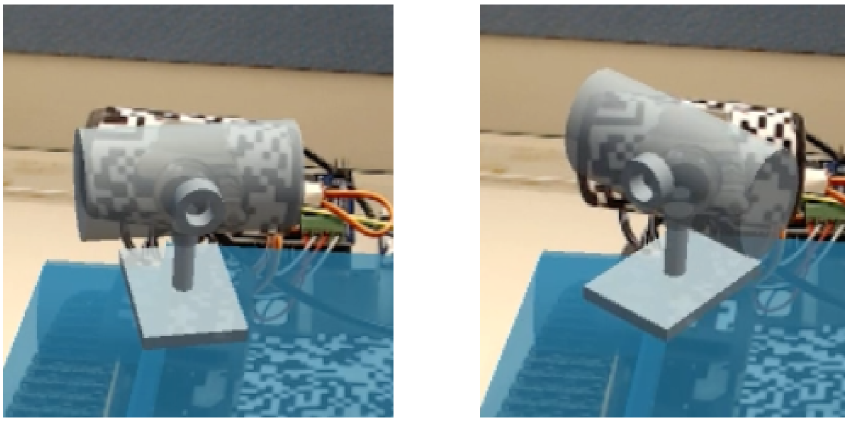
(b) The real objects superimposed with augmented reality model in the Unity scene

Fig. 4.6 The augmented reality model in third-person view

Next, two models can be dragged to move and rotate respectively. The model of camera, indicating the rotation angle of 2DOF pan-tilt, enables horizontal rotation (left-to-right), or vertical rotation (up-to-down) (Fig 4.7).



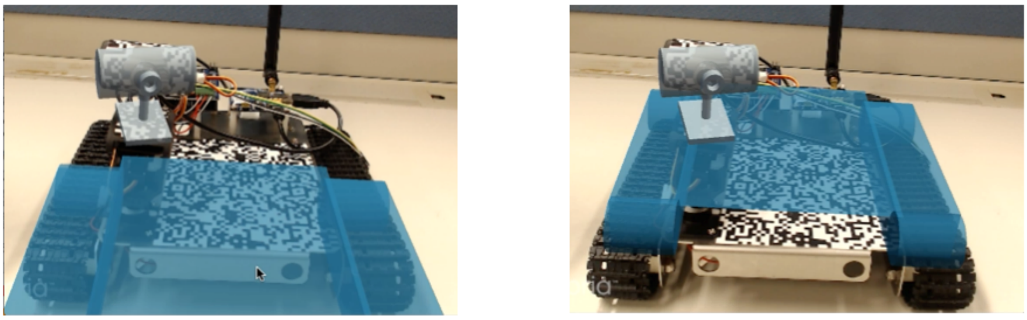
(a) Vertical rotation



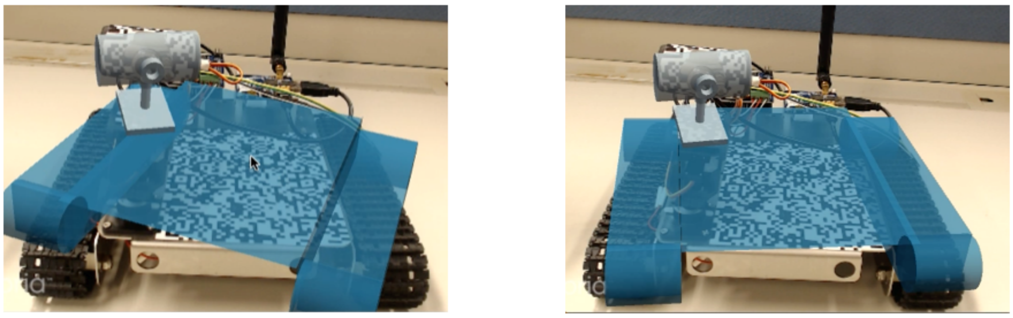
(b) Horizontal rotation

Fig. 4.7 The rotation of AR model representing camera

Similarly, the model of caterpillar band, indicating the position and direction of the monitor, can be dragged to go forward/ back, or make right/left handed rotation(Fig 4.8).



(a) Go forward/back



(b) left/right handed rotation

Fig. 4.8 The rotation of AR model representing caterpillar band

Moreover, for distinguishing different models, recent approaches for direct selection of the target device are based on hand gestures, and laser pointers [18]. Another approach for device selection is using a touch screen with camera captured image[19]. In my system, Raycast is used in Unity to distinguish different target objects. The essence of Raycast is to "draw" a line between two points in the game scene to detect any physical object that collides on this line. In our system, when user clicks the object on screen, the ray from AR camera to click position is built to detect collision with different objects. Since the different angles of the camera and caterpillar band, the two objects can be distinguished. Besides, the clicked position refers to the screen coordinates instead of the world coordinates.

4.3.3 Actuated Monitor Manipulation

After controlling the AR model in Unity, the change of direction and position will be wrote into the shared file. The upper computer, reading the data from the shared file, transfer the motion of AR model to manipulation command. In the upper computer (Fig 4.9), the number of cluster communication port and Baud rate are set for sending message. Users can click the button of "OpenPort" to establish connection between upper computer and lower computer.

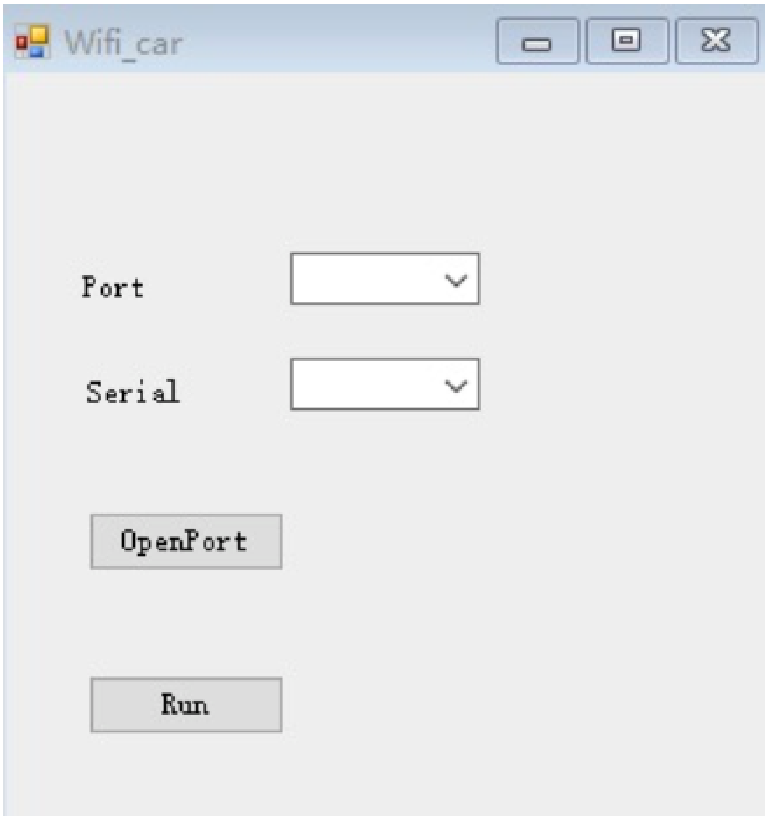
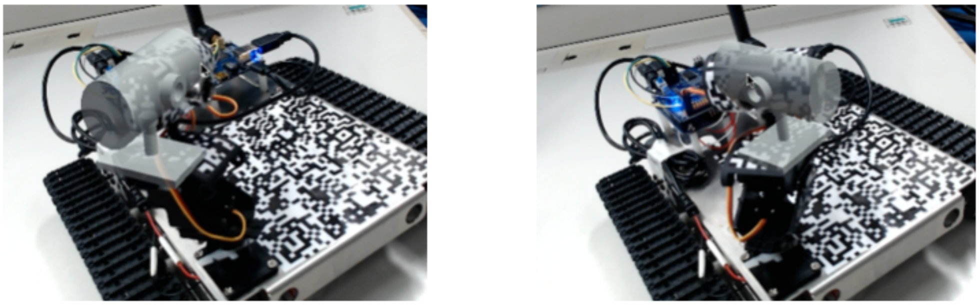
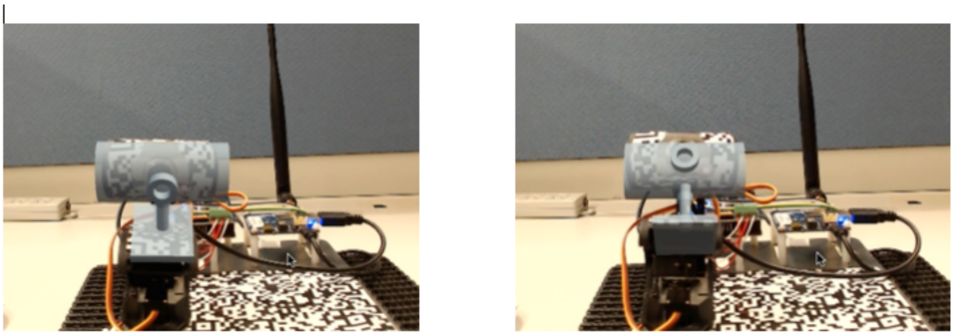


Fig. 4.9 The upper computer

The 2 DOF pan-tilt equipped with servo motors, can realize horizontal rotation about 120 degrees with 60 degrees on right side and 60 degrees left side respectively, and vertical rotation about 90 degree with 10 degree on front side and 80 degree 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 the upper computer received it can determine the corresponding command for control. In the Fig 4.10 we can see that the pan-tilt follow the AR model to rotate in different directions.



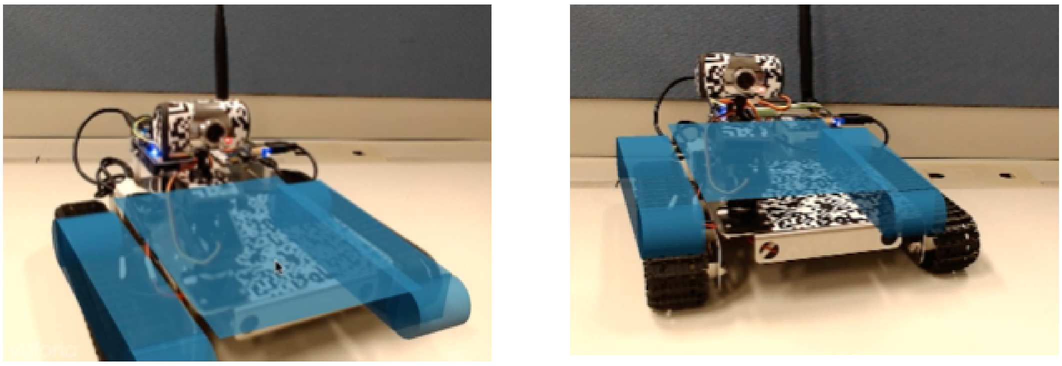
(a) Horizontal rotation



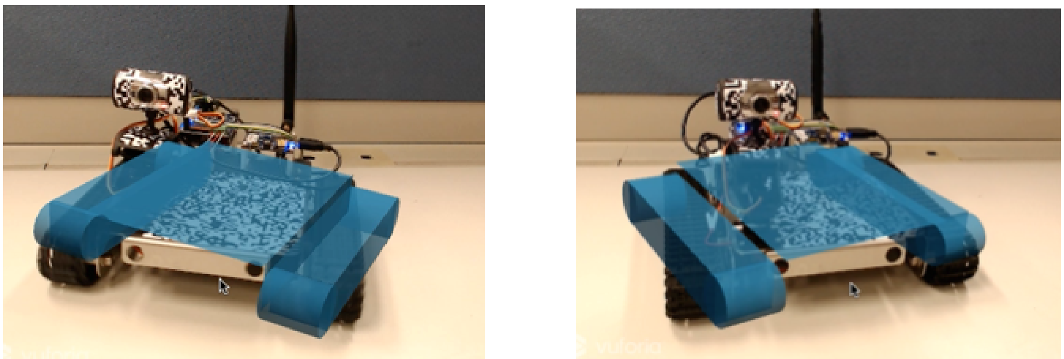
(b) Vertical rotation

Fig. 4.10 The rotation of 2 DOF pan-tilt

The caterpillar band equipped with electric motor, can go forward/ back in any distance or horizontal rotation degree with 360 degree on right and left side respectively. For translation, the upper computer reads the initial position and current position of AR model in the coordinate system in Unity and calculates the movement 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 number to determine corresponding command and send to the lower computer. For rotation control, the difference value of current direction and previous direction will be calculated to positive or negative corresponding to the right handed or left handed and transfer to command. After each dragging, the new position and direction data will overlap the old one as the origin point of the monitor's coordinates system. The Fig 4.11 shows the translation and rotation of caterpillar band.



(a) Go froward/back



(b) left/rigth handed rotation

Fig. 4.11 The rotation of model representing caterpillar band

4.3.4 Wireless Communication of Bluetooth Module

The bluetooth module is employed to send manipulation command from upper computer to lower computer. After paring, the command will be divided into several packages consist of head ,tail and useful information(Fig 4.12). The data is represented as hexadecimal number, with 2 digit of head, 2 digit of tail and 6 digit for useful information. The content of head and tail is hexadecimal number "FF", which indicates the start and end of each package.

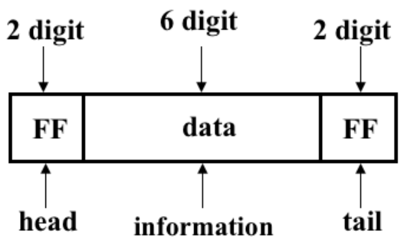


Fig. 4.12 The format of bluetooth package

4.4 Rear-person View Manipulation

The structure of rear first-person view manipulation shows in Fig 4.13. The monitor-mounter camera captures the live video and sends to the host computer via USB wired connection. In Unity, the AR models representing camera and caterpillar band will be placed in fixed position for control. Similar to the third-person view manipulation, the changes of models will transmit to the upper computer which generates manipulation command. Then the lower computer in monitor receives the command and translates to the signals for controlling pan-tilt and caterpillar band.

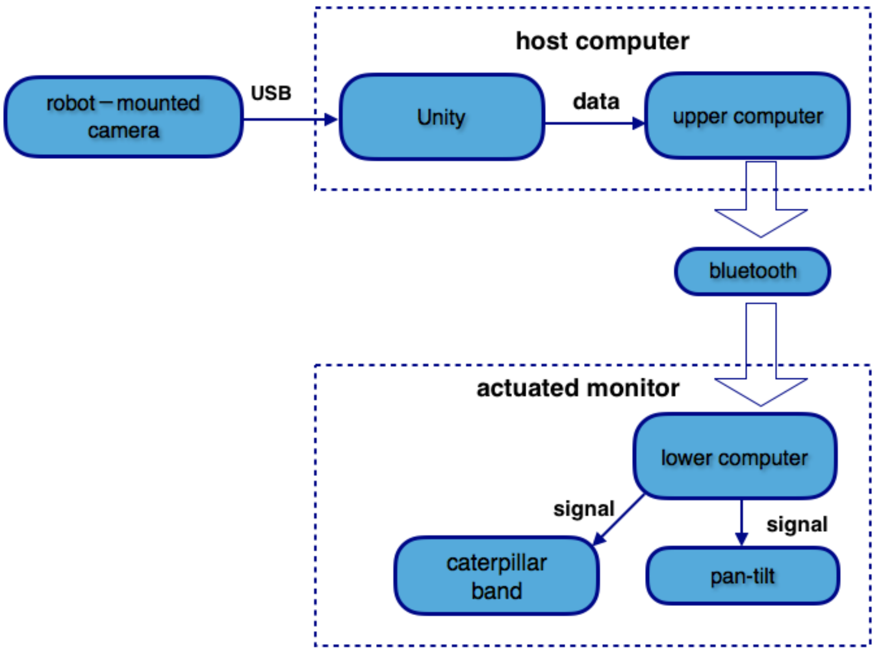
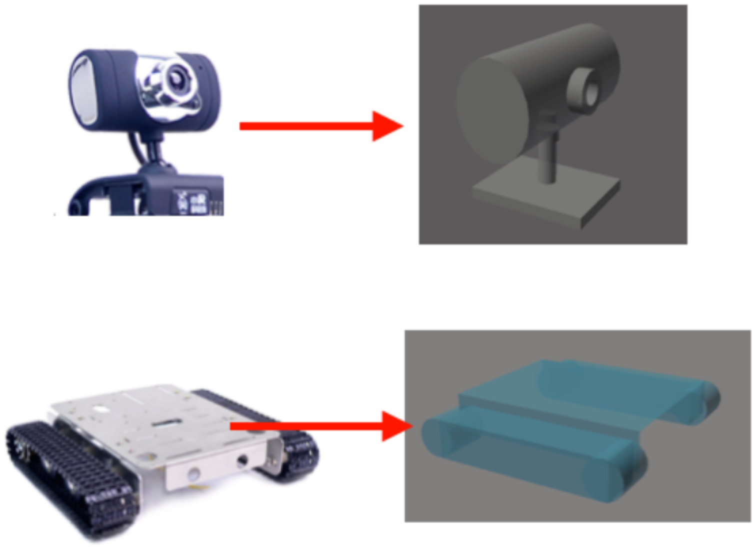


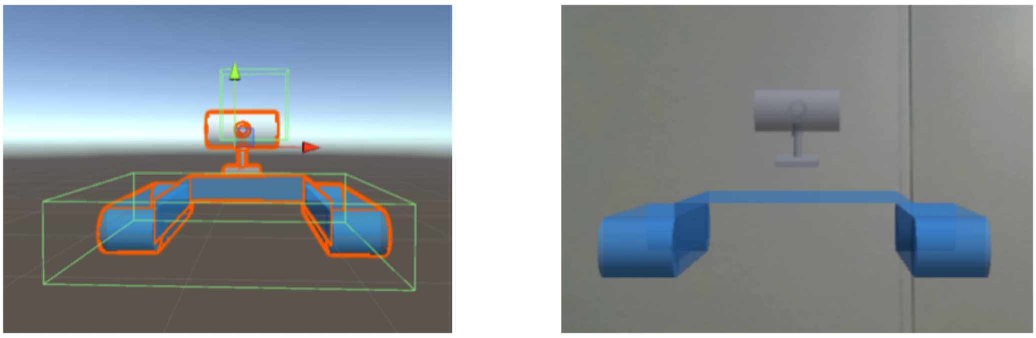
Fig. 4.13 The structure of rear first-person view manipulation

4.4.1 The Augmented Reality Model

In order to realize consistent manipulation interface in dual point of view, the same AR model used in third-person view can be imported into Unity with the format of FBX in the rear first-person view. The AR models representing camera and caterpillar band will be place in the fixed location initially for control(Fig 4.14). To avoid shielding the view of user, the two AR models reveal in semi-transparent.



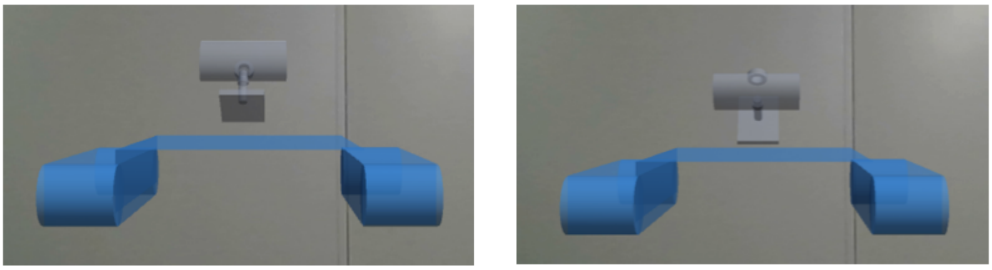
(a) The augmented reality model of camera and caterpillar band



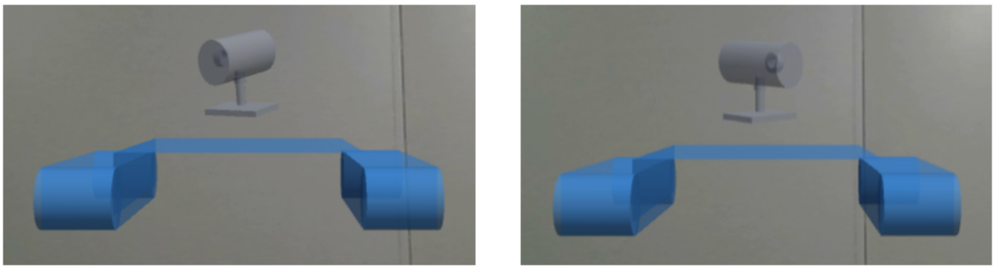
(b) The augmented reality model in unity scene

Fig. 4.14 The augmented reality model in rear first-person view

Following the movement of mouse, the model of camera, indicating the direction of 2DOF pan-tilt, can be dragged to up and down for vertical rotation(Fig 4.15(a)), or right and left for horizontal rotation (Fig 4.15(b)).



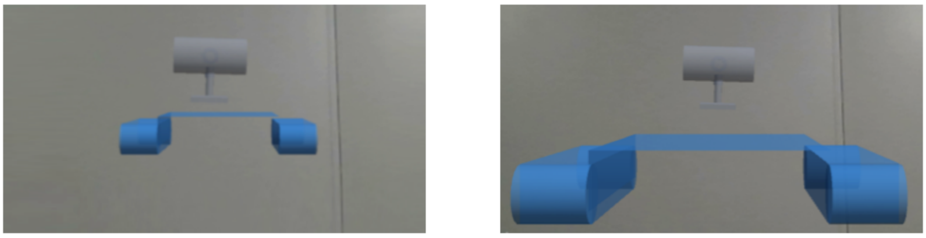
(a) vertical rotation



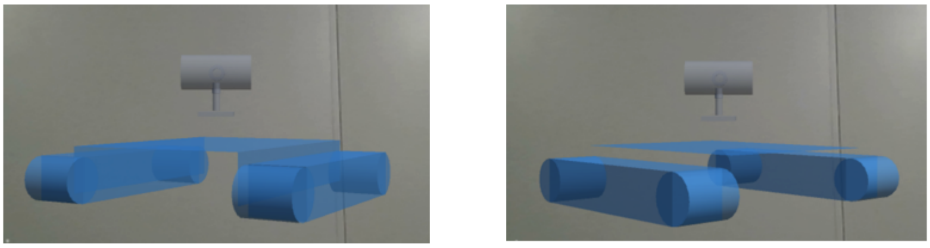
(b) horizontal rotation

Fig. 4.15 The rotation of model representing camera

Similarly, the model of caterpillar band, indicating the position and direction of the monitor, can be controlled to go forward and back(Fig 4.16(a)), or make right left handed rotation(Fig 4.16(b)).



(a) Go forward/back



(b) left/right handed rotation

Fig. 4.16 The rotation of model representing caterpillar band

4.4.2 Actuated Monitor Manipulation

The manipulation part of rear first-person view is as same as the third-person view. As shown in the Fig 4.17, in order to indicate the rotation of pan-tilt clearly, we put a picture of "smile face " in the center area of the first-person view, which shows the initial direction of camera.

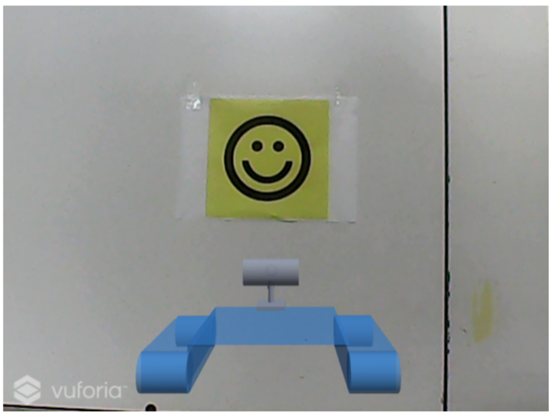
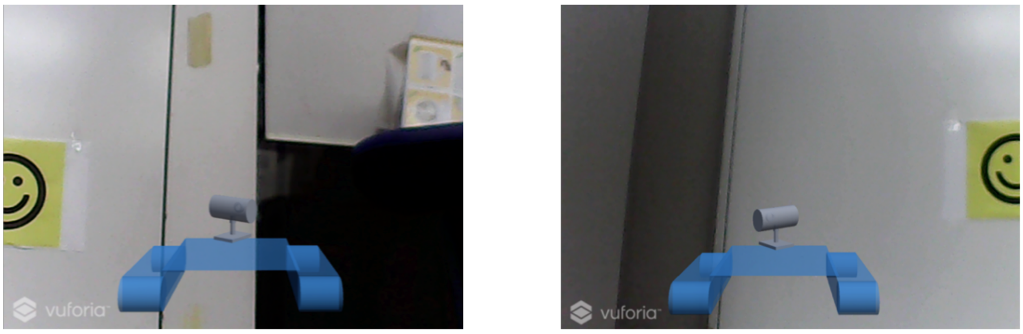


Fig. 4.17 The initial direction of camera

Later, we drag the AR model to various directions with different angles. In the Fig 4.18, we can find that the position of smile face has changed in the view of camera after we control the pan-tilt, which means the change of camera is in step with the change of AR model.



(a) Horizontal rotation

(b) Vertical rotation

Fig. 4.18 The rotation of 2 DOF pan-tilt

The control method of caterpillar band in rear first-person view is also same with that in third-person view. The monitor can be controlled to go forward/ back in any distance or horizontal rotation degree with 360 degree on right and left side respectively. In order to show the control result of caterpillar band, we set the initial position and direction of monitor, in which we place a cup in the center area of the first-person view as Fig 4.19 shows.

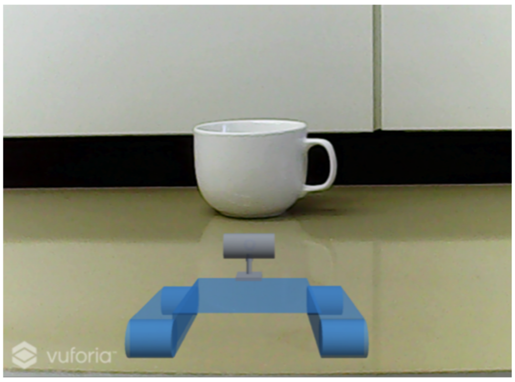
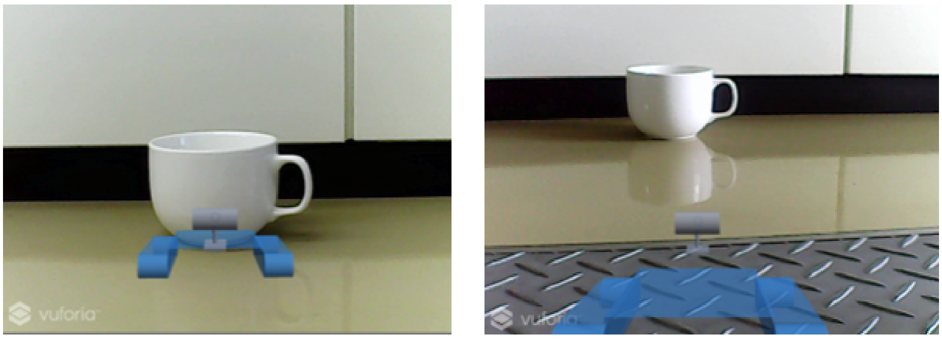
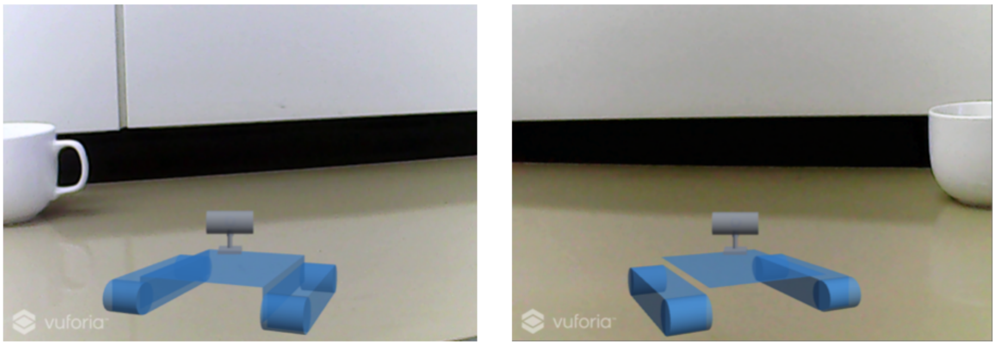


Fig. 4.19 The initial position of caterpillar band

Next, we drag the AR model of caterpillar band to different position and direction, the change of the position of the cap in the view of camera indicates the change of caterpillar band in real world like the Fig 4.20 shows.



(a) Go forward/back



(b) left/right handed rotation

Fig. 4.20 The rotation of caterpillar band

Chapter 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 method that overlays a color-coded coordinate system on the end-effector of the robot using augmented reality to help the user to understand the key mapping of a joystick [20].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[21]. 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 enables the operator to understand the robot internal statuses intuitively, which is helpful for debugging and actual operation.

Chen et al. also developed a mixed reality environment for performing robot simulations based on the concept of Mixed Reality[22]. 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 an augmented reality through graphic overlaying on a stereo video [23]. In their application, the user wearing a data glove controls a robotic arm by manipulating a virtual cursor overlaid on the video image.

Xiong et al also developed a tele-robotic system based on augmented reality to control a six DOF robotic arm [24]. In this system, a virtual robot works as an interface between the operator and the real robot, mitigating the problem of time-delay between user operation and real robot action. This idea is also used in their research, but they use a touch screen for the interface and they empirically compare three touch interaction methods, while they use a head-mounted display, a data glove, and voice commands for their interface. In addition, they present the advantages of predictive display. Simulation of virtual robot's tasks in the augmented environment improves the safety of the telerobot when it executes the planned tasks.

Our work is closely related to the previous research called "TouchMe", a tele-operating system which allows the user to manipulate a multi-DOF robot intuitively with touch interaction from a third-person view[2]. TouchMe defines 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 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 can not solve the problem of dead zone manipulation. In addition, four fiducial markers (ARToolKit [25]) 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 specified position or rotates to specified 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 can not 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 control. The two point-of-view manipulation systems are running simultaneously and user can switch according to the different situation. 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 recognize 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 from different angles automatically.

Chapter 6

Preliminary Evaluation

We conducted a 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.

6.1 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 minutes.

6.2 Method

We conducted our user experiment in a prepared working space with the monitor like the Fig 6.1 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.

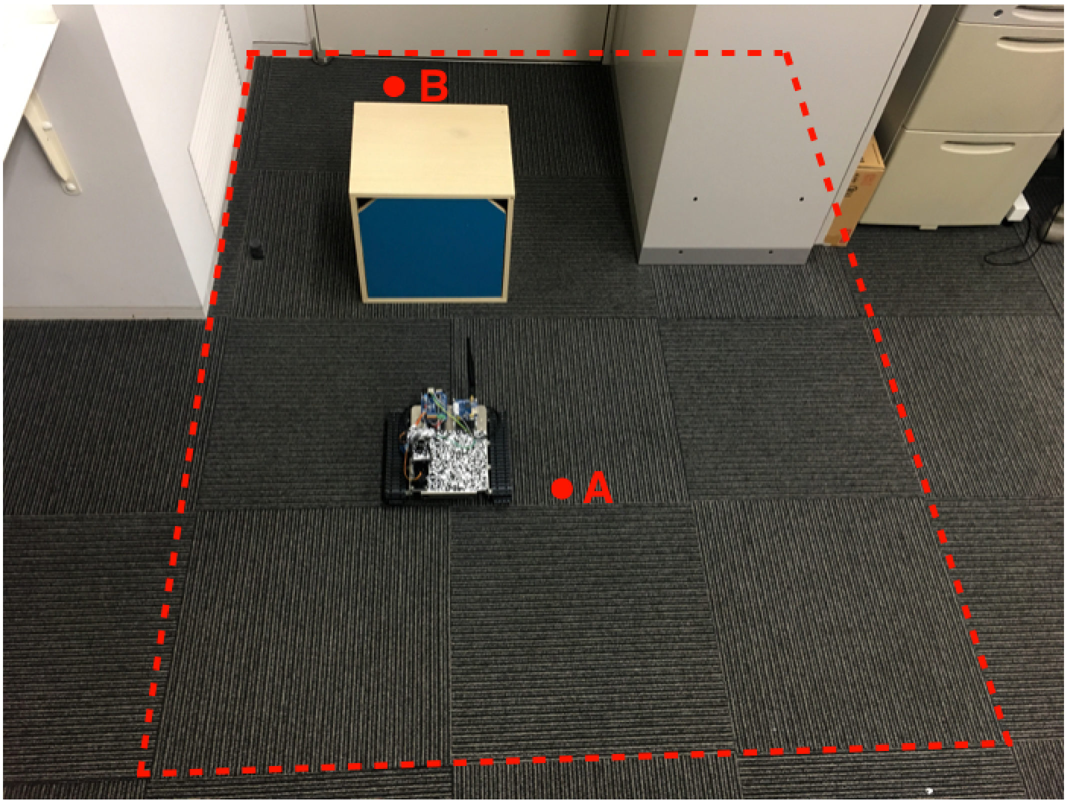


Fig. 6.1 The experimental working space

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 minutes. During the experiment, 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 minutes to use our system for controlling each part of monitor from dual point of view to complete the two tasks as follows:

Task 1:

Control the caterpillar band to translate from Place A to Place B.

Task 2:

Control the caterpillar band to translate from Place A to Place B.

Control the caterpillar band for horizontal rotation

Control the 2 DOF pan-tilt for horizontal rotation

Control the 2 DOF pan-tilt for vertical rotation

After finishing the task, all participants were asked to fill a questionnaire (Fig6.2). They needed to answer the 6 questions by grading from 1 to 5 (1=very negative, 5=very positive) for dual point of view manipulation and finally chose the prefer one in two point of view.

1. Do you think this interface is intuitive to manipulate monitor in domestic working space?

2. Do you think using augmented reality model is better than button to control monitor?

3. Do you have difficulty to get accustomed to our system to manipulate monitor?

4. Which point of view do your prefer to manipulate monitor?

Questionnaire

Welcome to this Questionnaire which we researcher want to evaluate our design.
Thank you for filling it all out !

Researcher Remarks

No. ___ Group .
No. ___ Participant .

Questions

Third-person view manipulation (Each question is graded from 1 to 5)

1. Do you think this interface is intuitive to manipulate monitor in domestic workspace ?
   far from intuitive ☐—☐—☐—☐—☐ very intuitive
2. Do you think using Augmented Reality model is better than button to control monitor ?
   much worse ☐—☐—☐—☐—☐ much better
3. Do you have difficulty to get accustomed to our system to manipulate monitor ?
   many difficulties ☐—☐—☐—☐—☐ no difficulty

Rear first-person view manipulation (Each question is graded from 1 to 5)

4. Do you think this interface is intuitive to manipulate monitor in domestic workspace ?
   far from intuitive ☐—☐—☐—☐—☐ very intuitive
5. Do you think using Augmented Reality model is better than button to control monitor ?
   much worse ☐—☐—☐—☐—☐ much better
6. Do you have difficulty to get accustomed to our system to manipulate monitor ?
   many difficulties ☐—☐—☐—☐—☐ no difficulty

7. Which point of view do your prefer to manipulate monitor?
   - ☐ a. Third-person view
   - ☐ b. Rear first-person view
   - ☐ c. Using both view

Fig. 6.2 Questionnaire

6.3 Result

All 8 participants succeeded in completing the two tasks 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 Figure 6.3.

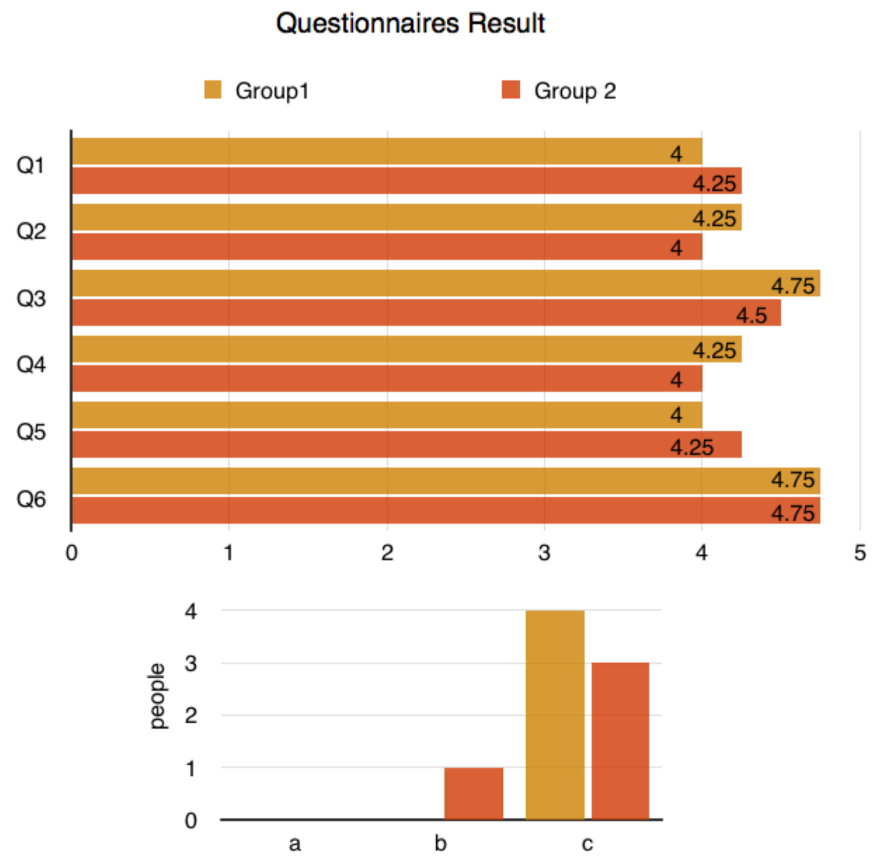


Fig. 6.3 The results of the questionnaire

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

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 control 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 surrounding environment and target place for translation. However, for participant in group 2 who did not know the working space before, since the relatively narrow view of the robot-mounter camera, 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. For each question it shows the positive result which 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. When asked about the experience of controlling model, 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.

Chapter 7

Conclusion and Future Work

7.1 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 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 AR models described in the user's coordination system.

Our system have received a positive feedback from the preliminary experiment. The result indicates that the 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 .

7.2 Future Work

In the future, we plan to improve our work in some aspects. Firstly, we can extent the manipulation interface from PC to touch screen. With the function of touch, it is more easy and more precise for user to drag models for advanced manipulation. In addition, currently the robot-mounted camera and the fixed setting camera in working space transmit the live video to host computer via USB wired connection. In the future we can try to realize the wireless connection between them.

References

[1] Shunichi Kasahara, Ryuma Niiyama, Valentin Heun, and Hiroshi Ishii. extouch: Spatially-aware embodied manipulation of actuated objects mediated by augmented reality. In *Proceedings of the International Conference on Tangible, Embedded and Embodied Interaction*, TEI '13, pages 223–228. ACM, 2013.

[2] Alessandra Rossi, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L Walters. Human perceptions of the severity of domestic robot errors. In *International Conference on Social Robotics*, pages 647–656. Springer, 2017.

[3] Sunao Hashimoto, Akihiko Ishida, Masahiko Inami, and Takeo Igarashi. Touchme: An augmented reality based remote robot manipulation. In *Proceedings of the International Conference on Artificial Reality and Telexistence*, ICAT'11, pages 61–66, 2011.

[4] Daisuke Sakamoto, Koichiro Honda, Masahiko Inami, and Takeo Igarashi. Sketch and run: A stroke-based interface for home robots. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 197–200. ACM, 2009.

[5] Jun Kato, Daisuke Sakamoto, Masahiko Inami, and Takeo Igarashi. Multi-touch interface for controlling multiple mobile robots. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, CHI EA'09, pages 3443–3448. ACM, 2009.

[6] Cheng Guo, James Everett Young, and Ehud Sharlin. Touch and toys: New techniques for interaction with a remote group of robots. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 491–500. ACM, 2009.

[7] Lars Fritsche, Felix Unverzag, Jan Peters, and Roberto Calandra. First-person teleoperation of a humanoid robot. In *Humanoid Robots (Humanoids), 2015 IEEE-RAS 15th International Conference on*, pages 997–1002. IEEE, 2015.

[8] Corey Pittman and Joseph J LaViola Jr. Exploring head tracked head mounted displays for first person robot teleoperation. In *Proceedings of the 19th international conference on Intelligent User Interfaces*, pages 323–328. ACM, 2014.

[9] Kent Yee Lui, Hyunjun Cho, ChangSu Ha, and Dongjun Lee. First-person view semi-autonomous teleoperation of cooperative wheeled mobile robots with visuo-haptic feedback. *The International Journal of Robotics Research*, 36(5-7):840–860, 2017.

[10] Andrew Correa, Matthew R. Walter, Luke Fletcher, Jim Glass, Seth Teller, and Randall Davis. Multimodal interaction with an autonomous forklift. In *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction*, HRI '10, pages 243–250. IEEE Press, 2010.

[11] Emanuele Ruffaldi, Filippo Brizzi, Franco Tecchia, and Sandro Bacinelli. Third point of view augmented reality for robot intentions visualization. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pages 471–478. Springer, 2016.

[12] Kazuhiro Hosoi and Masanori Sugimoto. Shepherd: A mobile interface for robot control from a user's viewpoint. In *Proceedings of the The IEEE International Conference on Robotics and Biomimetics*, ROBIO'06, pages 908–913, 2006.

[13] Syed Mohsin Abbas, Syed Hassan, and Jongwon Yun. Augmented reality based teaching pendant for industrial robot. In *Proceedings of the The International Conference on Control, Automation and Systems*, ICCAS '12, pages 2210–2213. IEEE, 2012.

[14] Rick van Krevelen and Ronald Poelman. A survey of augmented reality technologies, applications and limitations. *International journal of virtual reality*, 9(2):1–21, 2010.

[15] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent advances in augmented reality. *IEEE computer graphics and applications*, 21(6):34–47, 2001.

[16] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.

[17] https://library.vuforia.com/articles/training/vuforia-object-scanner-users-guide.

[18] Kentaro Ishii, Shengdong Zhao, Masahiko Inami, Takeo Igarashi, and Michita Imai. Designing laser gesture interface for robot control. In *IFIP Conference on Human-Computer Interaction*, pages 479–492. Springer, 2009.

[19] Masayuki Tani, Kimiya Yamaashi, Koichiro Tanikoshi, Masayasu Futakawa, and Shinya Tanifuji. Object-oriented video: Interaction with real-world objects through live video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, pages 593–598. ACM, 1992.

[20] Aditya Nawab, Keshav Chintamani, Darin Ellis, Gregory Auner, and Abhilash Pandya. Joystick mapped augmented reality cues for end-effector controlled tele-operated robots. In *Proceedings of the IEEE Conference on Virtual Reality*, VR'07, pages 263–266. IEEE, 2007.

[21] Kazuhiko Kobayashi, Koichi Nishiwaki, Shinji Uchiyama, Hiroyuki Yamamoto, Satoshi Kagami, and Takeo Kanade. 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 '07, pages 1–2. IEEE Computer Society, 2007.

[22] Ian Yen-Hung Chen, Bruce MacDonald, and Burkhard Wunsche. Mixed reality simulation for mobile robots. In *Proceedings of the International Conference on Robotics and Automation*, ICRA'09, pages 232–237. IEEE, 2009.

[23] David Drascic, Julius J. Grodski, Paul Milgram, Ken Ruffo, Peter Wong, and Shumin Zhai. Argos: A display system for augmenting reality. In *Proceedings of the INTERACT and CHI Conference on Human Factors in Computing Systems*, CHI '93, page 521. ACM, 1993.

[24] Youjun Xiong, Shiqi Li, and Ming Xie. Predictive display and interaction of telerobots based on augmented reality. *Robotica*, 24(4):447–453, 2006.

[25] Hirokazu Kato and Mark Billinghurst. 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'99, pages 85–94. IEEE, 1999.