VISUAL SUPPORT SYSTEM FOR REMOTE-CONTROL CONSTRUCTION MACHINE BASED ON AUTONOMOUS CAMERAS

無人化建設機械の操縦を支援する自律的映像提示システムの構築

February 2016

Junjie YANG

楊　俊傑

Visual Support System for Remote-Control Construction Machine Based on Autonomous Cameras

無人化建設機械の操縦を支援する自律的映像提示システムの構築

Junjie YANG

楊　俊傑

Development of Modern Mechanical Engineering, Research on Intelligent Machine

Graduate School of Creative Science and Engineering

Waseda University

This dissertation is submitted for the degree of Doctor of Engineering

February 2016

ABSTRACT

Considering the high risk of rescue and recovery work in disaster scenes, unmanned construction has been widely used based on remote control. Compared with manned construction, the efficiency of unmanned one is about 40% lower. According to the number of conventional research, insufficient visual information, communication delay, and lack of tactile sense are represented as the three main reasons. Among them, insufficient visual information is ranked first. Blind area, narrow angle of view, no sense of distance and eyestrain are four major reasons causing visual problems. In conventional researches, the most common method is to update the hardware, such like high-resolution monitor, 3D videos, wide-angle lens and etc. By using such method, current visual problems cannot be thoroughly solved, even if the efficiency can be raised indeed. Thus, I decided to take over this research and try to find a method which can support advanced vision information and solve the visual problems fundamentally and thoroughly.

This research is conducted in both software aspect and hardware aspect. It is mainly realized in three steps according to the order of information process. The first is acquiring most of the necessary visual information in the work site. And then, it will be processed by cameras according to different demand of photography. At last, the processed visible visual information will be displayed together with invisible information by using augmented reality prompts. It can make views understandable and enable operator easy to imagine and reconstruct the work site in his/her brain.

Three systems are designed to achieve the three steps. Camera placement system is used to cover most of the work site, which makes required visual information accessible to environmental cameras. By using Autonomous camera control system, environmental cameras can track respective targets by suitable photography method. And videos will be displayed in different viewports properly to make operator easy to understand the meaning and the role of each. Finally, augmented reality (AR) vision prompt system will be embedded in autonomous camera control system. It will calculate the invisible information and display it in the form of prompts with processed videos. In order to verify the effect of later two systems, a virtual reality (VR) simulator was developed.

In camera placement system, environmental cameras should not only be applicable to complicated work site but also meet different demands of photography. By using conventional method, it is possible to conduct camera placement on a complicated work

site. However, the local height of different cameras is fixed, which makes it difficult to fulfill the photography demand. Thus, I proposed a novel model, which enable the local height of camera variable. By using the camera model, it is possible to conduct camera placement in 3D world. And it can meet different demand of photography better. It start by loading the terrain condition and required areas in advance. Then, it slides the terrain into meshes and gives the possible vehicle position with the consideration of tilt and smoothness of each mesh. Combining the perspective properties of environmental camera, all meaningful camera states are calculated including the position and posture. At last, the system arranges special camera-carried vehicles by trade-off between cost and coverage. By using the system, most of the required area can be covered by environmental cameras. It can supply operator with reasonable views to understand work space better. In addition, camera can be changed quantitatively and modified manually to fulfill photography requirements better according to the result of the system.

In order to supply operator with the views according to the real-time work situation, environmental cameras are designed to be adjustable autonomously. The system is named autonomous camera control system. In the system, the motion of machine is classified into mobility and manipulation. 4 roles of cameras are defined to describe them. 2 of the roles are used to photo the machine the front/back and side. The others are used to photo the manipulator from the front and the side. Environmental cameras positioned in work site in advance are arranged to play these roles. Thus making observation targets be tracked and framed according to the motion of machine. Because the position of each environmental camera is fixed, one camera cannot play a special role all the time. Therefore, cameras need to switch with others if some of them cannot play the occupied roles. The system is realized through reading the real-time position and posture of construction machine as well as the preset environmental information first. These states are used to adjust the postures and magnification of cameras to play the four roles. Sometimes, the exchange or rearrangement of camera is conducted to ensure the 4 roles can be always played. By using the system, operators can observe reasonable views of construction machine and those of its manipulator to imagine and reconstruct the work site in their brain.

AR vision support system is embedded in autonomous camera control system. It is used to inform operator the invisible information to reinforce the sense of distance. AR prompt items such like arrow and wireframe are displayed in relative views respectively by calculating the situation around construction machine. Distance between obstacle

and machine, drop point of machine, reachable area of manipulator and relationship between joystick and rotation direction are displayed by different augmented reality prompts. In addition, a danger view is introduced to autonomous camera control system, in order to inform operator of the potential danger with AR prompts. By using the system, operators can understand the work condition better by stronger sense of distance. Moreover, they can concentrate on the views corresponding to their current action without ignoring the danger situation which can maximize their concentration and raise the work efficiency.

As a result, camera placement system succeeds in using the least cameras to show most of the necessary areas of work site. Autonomous camera control system succeeds in processing different visual information according to the demand of photography. It is also able to arrange cameras to ensure that visual information can be always given properly. In AR vision support system, danger view is introduced and AR prompts is able to be displayed with related videos. Eye sight is also be attracted to suitable views as expected according to the situation around controlled machine, which maximize operator's concentration with being aware of dangers. Thus making the performance like efficiency, accuracy and safety of operation better and increasing the user experience while decreasing the misoperation. In addition, the principle of human-friendly AR prompts is made clear, which enable us to design and utilize better prompts in the future work.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Junjie YANG - February 2016

1 INTRODUCTION

In our daily life, human beings always use their eyes to obtain visual information. Visual information from different perspectives is combined to reconstruct the real world. Nevertheless, we are still unclear about how human beings achieve it. In many cases, we try to increase the amount of vision to do that. However, the effect does not rise as the number of vision increases.

This paper is intended to reveal some facts in the field of human cognition. And these facts will be applied to solving practical engineering problem.

1.1 Background

No matter the natural disasters or the manmade disasters, they do destroy living environment [64][66]. And after that, no one knows whether a secondary disaster occurs or not. Usually, we send rescue team to save injured people and dispatch construction team to recover the environment immediately [72]. However, it is tantamount to gambling these team members' lives. For example, a dangerous goods warehouse was on fire in China's Tianjin in the night of August 12th 2015. Over 600 firemen were sent to extinguish the flames immediately. Half an hour later, a series of explosions that killed over 17 and hurt more than 400. Until September 14th 2015, 165 persons were dead including 110 firemen. And 8 persons are still missing.

Even if the rescue team and construction team can come back alive, their mind will probably be tormented by post-traumatic stress disorder (PTSD) in the future. According to [1], people from these teams are easy to suffer PTSD for the following four reasons:

- They are obliged by their occupation.

- They are filled with sense of mission.

- They have to obey orders from their commanders.

- They can devote their lives, which is thought by the public.

In order to realize the same rescue and recovery without endangering more people, we have to utilize some methods which can protect rescue members and construction members from dangerous scene.

There are two oblivious methods which can achieve that. One is to a shelter which can resist all damage, the other is to keep them far from the dangers or potential dangers. The previous is hard to achieve. Any one of the existing shelters can only resist several types of damage rather than realize full protection. Thus, tele-operation becomes a more realistic and safer option.

Nowadays, small-scale machines are developed and widely used all around world to search the wounded after earthquake, remove explosive or radioactive materials and etc. For instances, [2, 3, 4]. Compared with small-scale machines, large-scale machines for recovery and rescue are less developed.

In this field, the corresponding technologies mainly include civil engineering technology, mechanical technology, imaging technology, wireless communication technology and information communication technology.

In Japan, unmanned tele-operation construction technology was started to be used in 1969. At that time, the pier of a bridge over Jōganji River collapsed. A tele-operated amphibious bulldozer was utilized to conduct recovery work (Figure 1.1). The operator stood on the beach and controlled the machine by wired remote control. The distance is not quite big. As a result, it succeeded to conducting excavation and dozing. [5], [6]
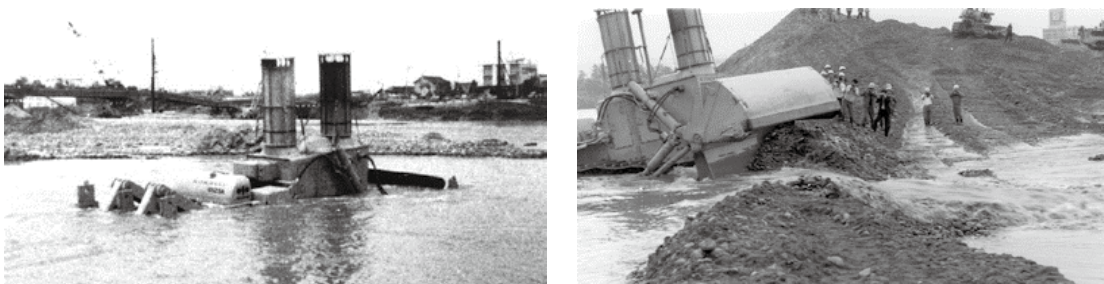
*Referred from [5]

Figure 1.1 Tele-operated amphibious bulldozer

In 1983, tele-operated backhoe was developed and introduced to conduct excavation tasks where there maybe fell some solid or stone. At that time, wireless tele-operation technology was introduced (Figure 1.2). [5], [6]

Receiver

Transmitter

*Referred from [5]

Figure 1.2 First wireless tele-operated construction machine

From 1970, tele-operation based on video started to be developed. Operator was able to watch the video about first-person perspective of an excavator (Figure 1.3). The video came from the camera which was carried in the cab of excavator by wired data transmission. In 1980s, the technology was practically realized. The excavator could successfully excavate and transport landslides. [5], [6]

*Referred from [5]

Figure 1.3 Tele-operated construction based on videos

Until 1980s, the development of fundamental technologies in tele-operation field had already been basically completed.

In the beginning of 1990s, tens of people died for the pyroclastic flow. In order to avoid secondary disaster, tele-operated recovery has been conducted since 1993. From that time on, tele-operation construction was widely used to conduct disaster response in Japan. These tele-operated machines have to adapt to extreme environments, such like potentially extreme temperature (more than 100°C) or long-distance tele-operation (more

than 100m). At that time, the problems of tele-operated construction gradually revealed, and researchers try to use different methods to solve them. [6]~[13]

In 2000, Mount Usu erupted. To conduct recovery far from poisonous gas, new wireless communication technologies were introduced, which keeps operators work more than 2km far from the work site. High-power transmitter (2W) was used instead of regular one (0.01W), so that the control signal can be directly transmitted to work machine. In addition, videos were transmitted through relay mobile (Figure 1.4).

*Referred from [5]

Figure 1.4 Recovery after eruption of Mount Usu

Since that, magnificent improvement of tele-operated construction has not appeared. Most of the improvement can be classified into 2 types. One is to use better devices such as high-resolution monitors instead of old ones [75]. The other is to make work machines do dedicated work more precisely such as piling up bricks [13][76]. Almost no technologies are developed to deal with the problems in regular work. Therefore, operators have to handle their regular work with low efficiency. According to [7], [14] and [15], efficiency of unmanned construction work is about 60 percent compared with manned one. In order to make it possible to keep rescue and recovery not only safe but also efficient, the current tele-operation system for large-scale machine has to be improved.

In [15], current research themes can be classified as Figure 1.5. According to information transmission classification method, they can be roughly classified into information acquisition (including "vision" and "tactile sense"), information communication (including "wireless and radio" and "operation delay"), information manipulation (done by operators) and information output (including "machine and attached facilities" and "operation method"). Among them, information acquisition takes up 48 percent, which

is far more than other parts. And 60 percent of information acquisition is about vision. As the basic information of planning and judgment, vision becomes the most important factor compared with others. According to the questionnaire of [16], I found that the main reason causing low efficiency is vision as well. Thus, I believe that vision is important and worthy of study.

[15][15]

Research themes

*Referred from [15]
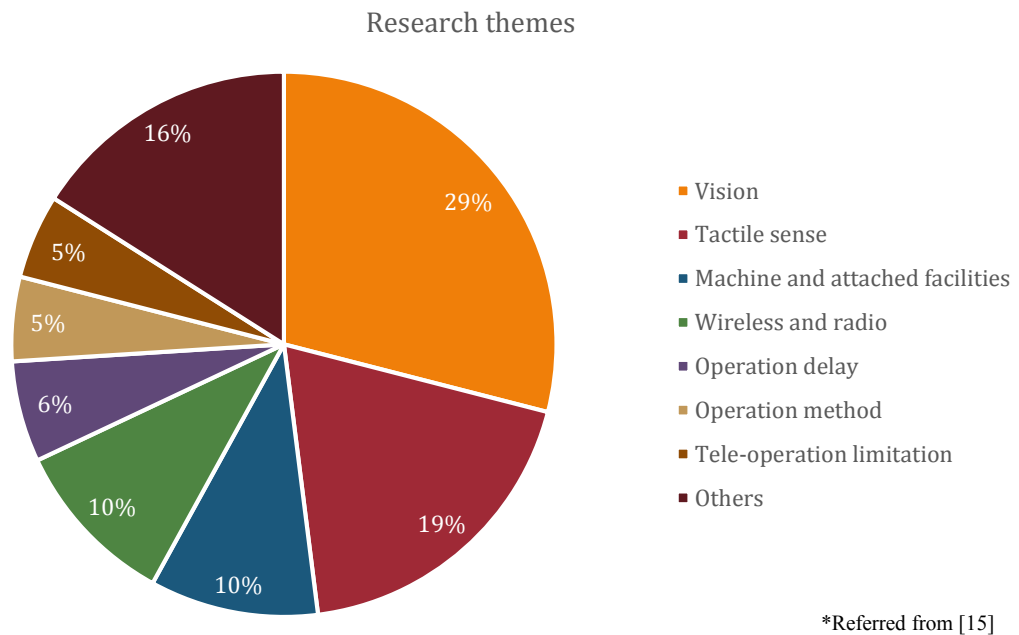
Figure 1.5 Classification of themes

1.2 Visual support system

In tele-operation field, visual support system can be divided into the two sections:

  i.   Information acquisition section

In acquisition section, GPS, radar, laser and different types of cameras are always used derive the position of controlled object. Here, GPS, radar and laser can capture the detailed position of object or the rough environment of object. However, it is hard for them to describe the other detailed properties of object, such like posture and color. By using these devices, operators can handle the rough position of object in the whole environment. In addition, a terrain map of the work site or a rough obstacle position can be also available. In Figure 1.6, it shows the visualized data acquired from corresponding devices. Here, devices of Figure 1.6 (a) and (c) are traditional GPS (Global Positioning System) and radar; device of Figure 1.6 (b) is a 3D LiDAR (Laser Imaging Detection

and Ranging) product developed by Velodyne Company, which is widely used in auto-
motive, UAV (Unmanned Aerial Vehicle) [8]-[10].
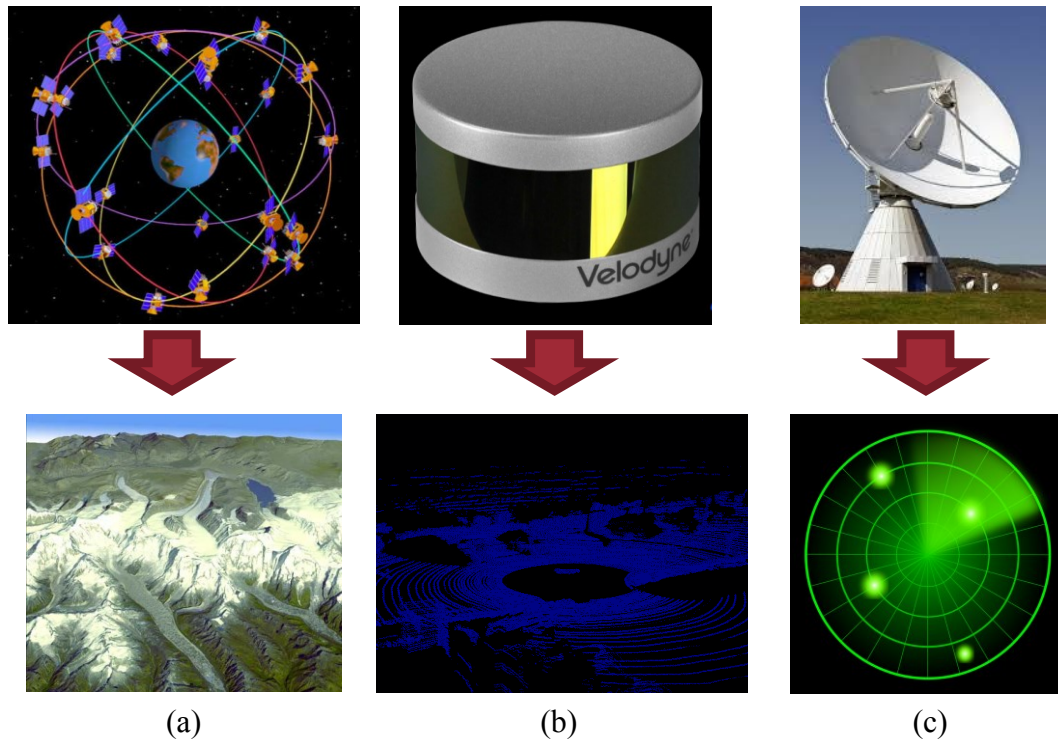
| (a) | (b) | (c) |

Figure 1.6 Terrain map, environmental map and positioning map

Compared with the above three types of devices, cameras are more commonly used, be-
cause the data from it is more intuitive and more detailed. Different kinds of cameras
can be used in different scene. The differences among cameras are mainly lens and
zooming method. Except the types of lens for regular photography like wide-angle lens
or long focus lens, the combination of regular lens and a special mirror is always used to
acquire an omni-vision (Figure 1.7). In [11], the device is used to record the environ-
ment around the robot.
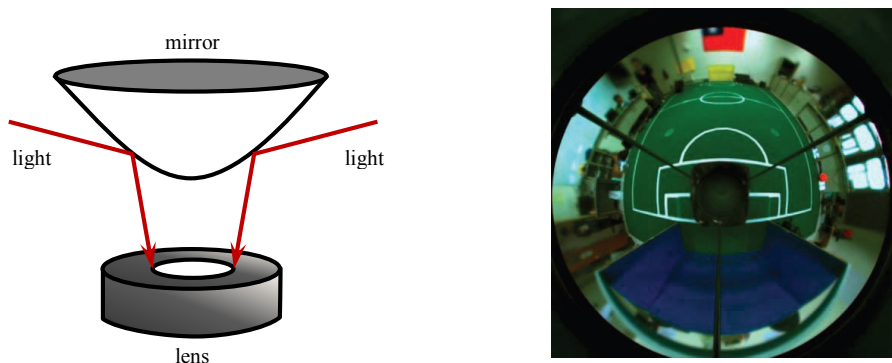
*Referred from [11]

Figure 1.7 Omni-vision device and acquired image

Junjie YANG - February 2016

There are two types of zooming method. One is optical zooming and the other is digital zooming. Optical zooming is realized by changing the view angle of lens, which makes the image after zooming has the same resolution as that before zooming. It is mainly limited by the corresponding property of lens. Digital zooming is implemented by cropping an image down to a centered area with the same aspect ratio as the original, which sacrifices the resolution of image. By using digital zooming, the limitation depends on the noise and amount of resolution after zooming. Low-resolution image cannot describe the object precisely. Usually camera has an optical zoom lens and digital zoom will be automatically applied only if its longest optical focal length has been reached [12].

ii.   Information output section

Generally, visual information is always be displayed by monitors or projectors. And sometimes, it can be displayed in different ways. For example, monitor array or curved projection. By using these methods, images or videos can be displayed for dedicated use. Commonly, monitor arrays are usually used show single extremely big scene or monitor multiple scenes (Figure 1.8 (a) and (c)). Curved projection is mainly used to show panoramic view around the observer (Figure 1.8 (b)). In addition, stereo monitor is another choice even if there are still some problems (Figure 1.8 (d)). For example, the overlapped images.

(a)

(b)

(c)

(d)
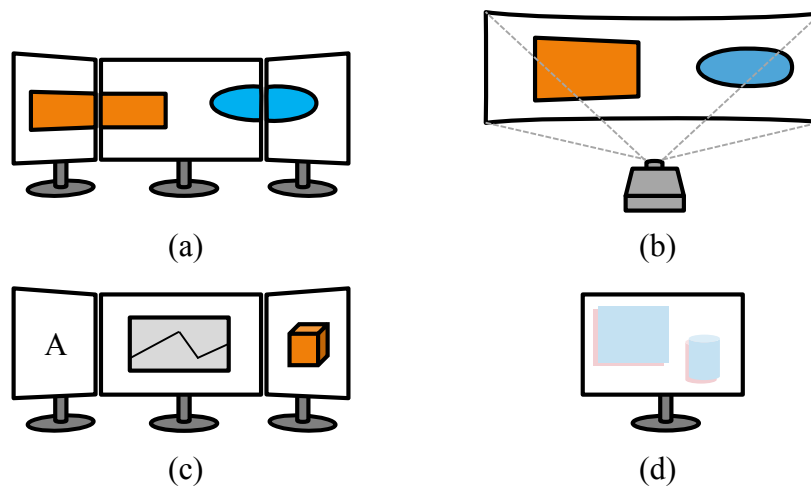
Figure 1.8 Different kinds of display method

What's more, how to represent the acquired data is another problem. Visual information can be always displayed as word, graph, image or the combination of them (Figure 1.8 (c)).

In unmanned construction case, the combination of GPS, cameras and monitor array is a regular way to show visual information.

1.3 Requirements

The Art of War said "if you know the enemy and know yourself, you need not fear the result of a hundred battles." In my case, it means if operators know the site and machine itself, they can rescue and recover safely and efficiently. In order to achieve that, we have to know the requirement of vision and related factors.

According to [13] and [35], "blind area", "narrow angle of view" and "no sense of distance" are the three main problems of insufficient vision. Thus, it is easily to be considered that "high coverage of work area", "wide angle of view" and "more visual information about distance" are the solution.

In conventional methods, wide-angle lens, big monitors and 3D videos are thought as the solution [14], [35], [69], [75]. By using each method, time efficiency can be raised to some extent. However, both of these methods cannot solve these three problems at the same time. When wide-angle lens is used, the vision area becomes bigger and the wanted visual information seems easier to be available. As the same time, operator's focus is maybe attracted by other unrelated things. In addition, wide-angle lens cannot make blind area seen and enhance the sense of distance. The later one can enhance the sense of distance, but the other two problems cannot be solved still. What's more, 3D videos make operators' eyes easy to feel tired.

Thus, I decided to use software method to solve these problems fundamentally and completely.

1.4 Objective and solution

As the ultimate objective of my method, the above three problems will be solved, and the time efficiency, safety and accuracy can increase obviously. In addition, the visual support system will be improved holistically instead of improving a partial section.

According to my solution, most of work site should be covered as the first step, which means most of the required visual information can be accessible from environmental cameras. And then, each camera will be designed to track object and adjust its magnification rate according to the size of the object. As the last step, the videos from active cameras will be reasonably processed and arranged on the monitors. At that time, aug-

mented reality elements will be displayed together with related videos, which can enhance the sense of depth and sense of distance.

In order to realize the objective. Several things should be done.

At first, cameras should be arranged properly according to the balance of cost and coverage rate with the consideration of environmental constraints, such like the smoothness of the ground and the possible position of the camera. Then, visual information have to be processed and described properly. Thus clear observation targets and human-friendly expression technique are needed. While acquiring visual information, the camera should make clear what should observe and how to observe it according to photography demand. When monitor displays the visual information, the system should tell it what should render and how to render it in a human-friendly way.

Here, three systems shown in Figure 1.9 are designed to achieve each step of my solution, which will be introduced in following chapters. Before that, the detail of each step will be discussed.
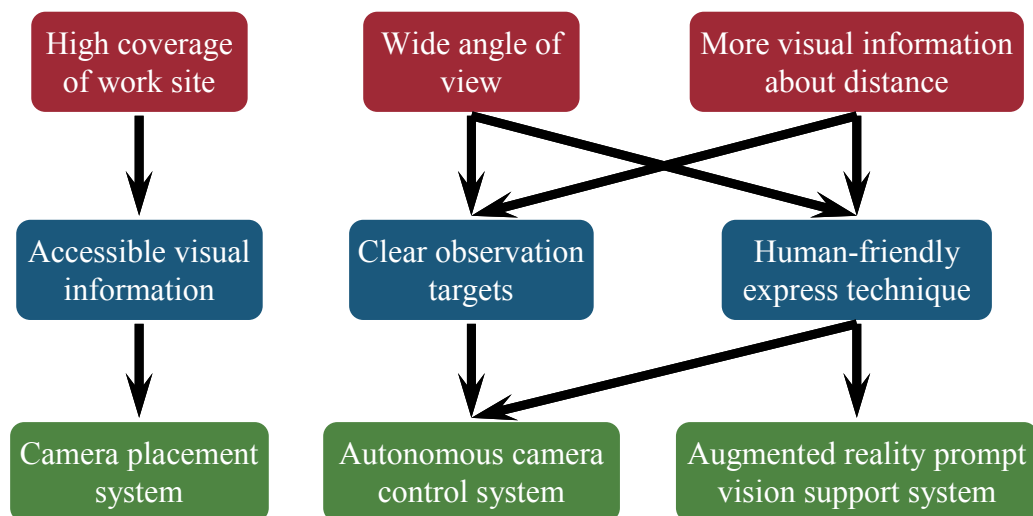
Figure 1.9 Entire system diagram

1.4.1 Accessible visual information

In order to make visual information accessible from environmental cameras which are installed in advance, the following essential factors should be considered.

Cameras are a kind of line-of-sight based sensors with limited resolution. Thus, the photography range is limited. And the shape of the photography area should be an over-

lapped section of a pyramid and a ball. They are used to meet different kinds of photography demands in work site. In most of cases, the work site is not always plain. Sometimes, the terrain is quite complicated. Thus, these cameras have to be applicable to complicated terrain. These factors imply that both of the posture and the position of cameras should be thought about before camera placement.

1.4.2 Clear observation targets

When operators control the construction machine, the entire task is always divided into several sub-tasks. For example, when an operator is asked to transport an object and the machine is far from it, the task is always divided into two parts: driving the machine to approach the object and doing the transportation. It is a strategy way of thinking. In such cases, the machine is thought as a particle and the observation target should be the entire work site or a part of it (Figure 1.10 (a)), such like a mini map.

In order to complete each sub-task, it will be divided into several actions. For instance, when the object is reachable, the transportation sub-task will be divided into three actions: picking the object, transporting it and releasing it. This is a tactical way of thinking. In this case, the machine is treated as an object with volume instead of a particle. Here the observation target should be a bigger scope including some parts of machine or the entire machine (Figure 1.10 (b)). Except that, a cab view is also thought as a tactical view which is designed to observing the front area of machine. By using this view, operator will feel like setting in the machine. And the view is a common selection as well, such as [13], [14][35].
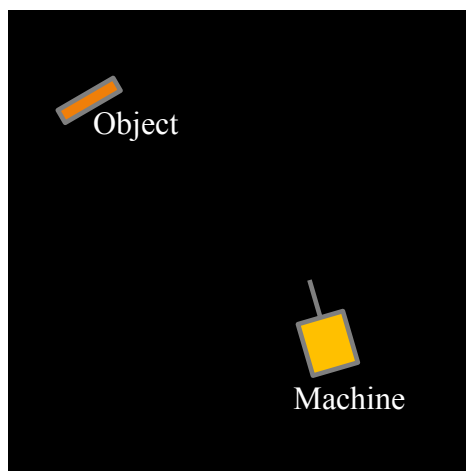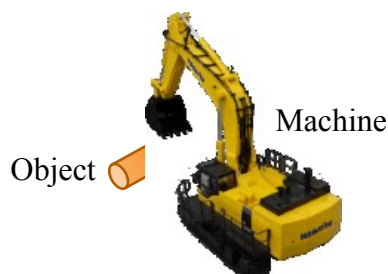
(a) Strategy target   (b) Tactical target

Figure 1.10 Different observation targets

In order to make target easily understandable, the scope of target has to be discussed here. In any case, scene near the machine is the most important. It means the scene in the maximum scope of action with the consideration of operators' reaction time and response time. The scope of action is also related to the possible range of motion of the machine during its response time. In Figure 1.11, I give the scope of motion of a car in different conditions. In construction machine case, the scope is more complex and more difficult to predict.

(a) Car runs slowly   (b) Carr runs fast
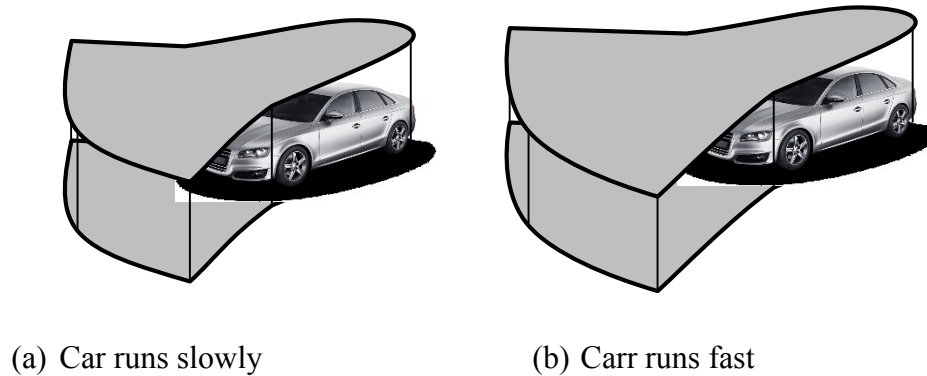
*Referred from www.drivingfutures.com

Figure 1.11 Maximum scope of motion of a car in different conditions

A simple method is to give an extremely large scene, which contains the scope of maximum action of machine. However, this may cause the influx of extra visual information. Because of the human being's limitation of processing information, it may increase the difficulty of identifying the related information. It implies that the time cost of processing visual information increases, which will lead to the low efficiency in both time and accuracy. Contrarily, if the scene is smaller than the scope of action, operator may slow down the manipulation or movement of the machine voluntarily for fear of the unknown. Thus I think a suitable scope of the scene should be one of the most important requirements while developing the vision system.

1.4.3 Human-friendly expression technique

After making visual information accessible and making clear observation targets, the required information should be transformed in a human-friendly way. In this way, operator should observe appropriate views and reconstruct the work site condition in their brain.

By using conventional method, some of views used are not understandable because of the unsuitable scope. And sometimes, some distance is hard to be calculated which may

cause misjudgement. Thus, a good human-friendly expression method should include not only proper real-time videos but also some prompts which can describe invisible information (Figure 1.12).

(a)   (b)
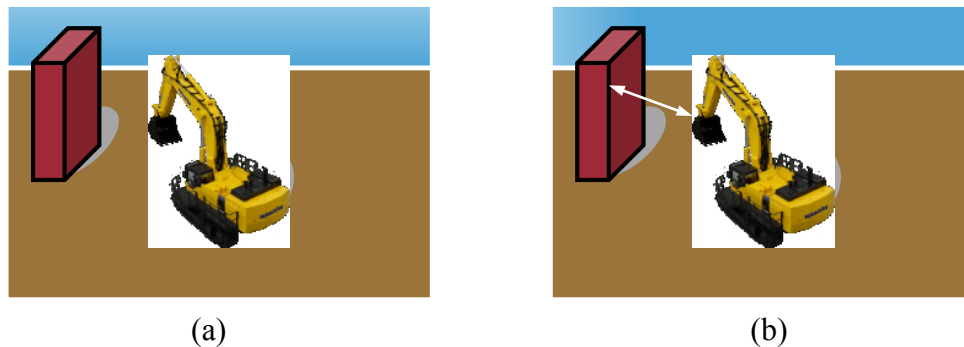
Figure 1.12 Visible information and invisible information

i.   Visible information

In order to making the targets understandable when they are recorded, the number of views and the types of views are quite important. Besides them, how to arrange these views in a monitor is another question.

For cab view and mini map, dedicated devices can be used. In addition, the cab view is mainly designed to make operator immersive and the different directions of a same mini map is meaningless, so only one view is enough to describe either of them.

On the other hand, scope of entire machine or that of some parts of the machine is always recorded from third-person perspective. Several views from different aspects for same scope can be used to describe work condition, which makes operator easily imagine and rebuild the scene in his/her brain.

Referring to the views in static cases, we always use at least 2 views like orthographic projections to describe all necessary features of an object, when we make engineering drawing [17] (Figure 1.13). The number of views increases with the increasing complexity of object. Sometimes, auxiliary projection, axonometric projection, and section views are also used to illustrate the complicated object.
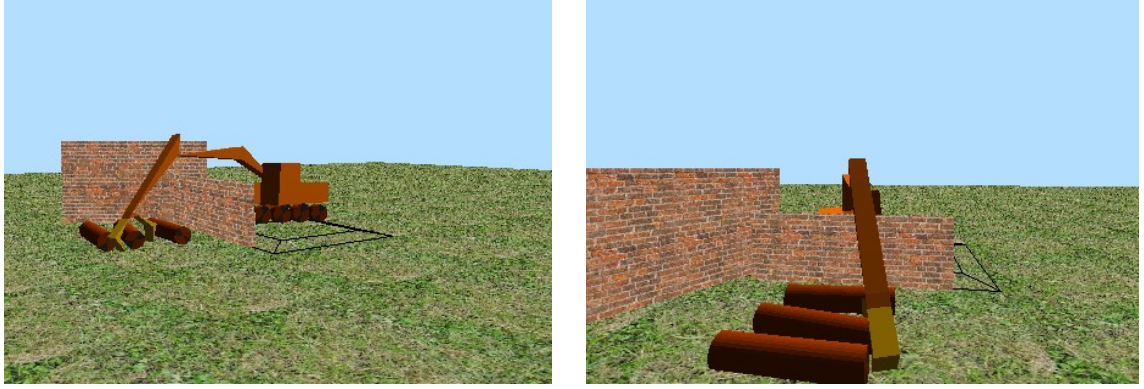
Figure 1.13 Side view and front view of work condition

Compared with cases in static conditions, views in dynamic conditions should be more complex. The unrelated information from camera cannot be ignored. And the video changes real-time, which means that operator has to process large amount of visual information all the time during his task. In order to reduce the operator's pressure of processing visual information, videos have to be intuitive and understandable. Moreover, operator pays more attention on holistic scene instead of small details. Thus, at least one perspective should be used to describe each scene respectively. In some cases, motion of target is very small. In order to understand the work condition in such cases, 2 views from different directions are recommended.

Types of views should be designed according to the photography requirements for targets. According to [16], [67], cameras carried on construction machine itself are fixed and used to record first-person perspectives about entire front scene including end-effector (Figure 1.14 (b)). And the mini map from GPS system gives the overlook view of the entire work site (Figure 1.14(a)). Thus, the types of views for these two targets have been decided.

(a)   (b)
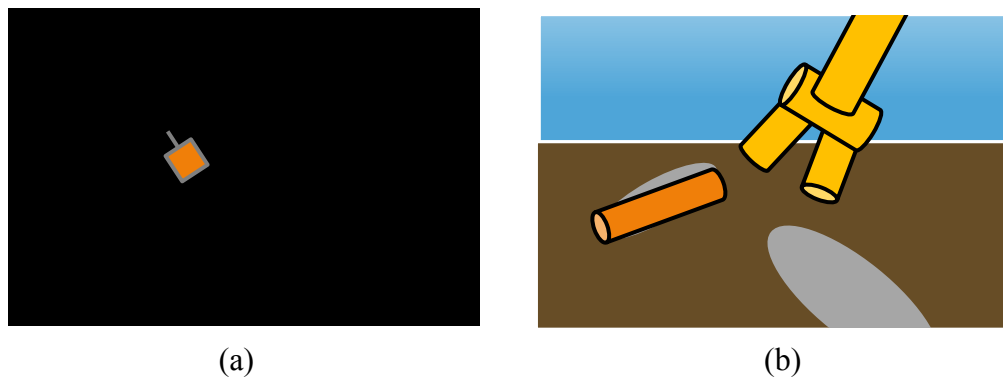
Figure 1.14 GPS view and cab view

For the other targets, same target can be recorded from several different directions. In the last section, 2 views are recommended. Therefore, two views from different directions will be used.

Besides the two factors mentioned above, a good arrangement pattern is quite important. It makes operator easy to understand the function of each view and handle the work site condition. In principle, views for same scope should be close, and more important views should be closer to cab view. Here cab view is considered as the main view.

   ii.   Invisible information

Sometimes, it is very difficult for operators to imagine some of invisible data like distance or orientation by using videos only. For example, it is easy to be aware that the machine is near a wall in Figure 1.12 (a). However, it is hard to figure out the rough distance between the wall and the machine without prompts like the arrow in Figure 1.12 (b). Like the arrow here, prompts make operator easy to understand the important invisible information without calculation.

However, prompts should not appear everywhere. Too many visual information will make views hard to be understand which may influence operators' judgment. So, not only the number and the types of prompts but also the corresponding views of them should be thought about.

1.5 Outline of the following chapters

- Chapter 2: I will propose a camera model which enable camera to be put in 3D world. According to the model, required observation objects will be designed in 3D world as well. With the consideration of balance between coverage rate and cost, camera placement algorithm will be conducted. And the result shows it is possible to conduct camera placement in a complicated terrain and most of the required area is covered.

- Chapter 3: In order to confirm my proposed systems, a simulation platform is built. The software part and the hardware part will be introduced.

- Chapter 4: To make camera track the wanted objects autonomously by adjusting its posture and magnification rate, work condition is analyzed at the beginning. According to that, different observation objects and photography methods are decided. Thus, the roles of cameras are designed. In order to make all roles can be always played, a camera arrangement method is proposed. Comparison

experiment confirmed the good performance in time efficiency, safety, accuracy and user experience.

- Chapter 5: Different kinds of augmented reality elements are used to enhance the sense of depth and sense of distance. An danger view is introduced to make the sence more understandable. The result of comparison experiment showed that the good performance in time efficiency, safety, accuracy and user mental pressure. Sight is also be successfully guided to views related to operator's task.

- Chapter 6: Conclude my work and point out the future works.

-

Junjie YANG - February 2016

2 CAMERA PLACEMENT SYSTEM

In this chapter, I proposes a novel sensor model in three dimensional (3D) space. By using this model, it is possible to do line-of-sight-based sensor placement where the height of each camera above ground is variable and the required observation area cannot be treated as a flat surface.

2.1 Background

Unmanned construction machine is not only be used on plains, but also be used on complicated terrains such as in mountain area [18]. In such complicated cases, the attitude difference can be very big and maybe there are some steep slope like cliff. Attitude everywhere is different. Compared with such cases, there is almost no attitude difference is relatively small. However, in some cases, objects are put on different height above the same attitude. Thus, it is almost impossible for us to treat required observation areas like that in 2D cases [19], [20]. Meanwhile, cameras should be also put on different height above local attitude to meet different televising demand (Figure 2.1).

Until now, camera placement in complicated terrain is always conducted by experience. Therefore, sometimes blind angle appears which makes efficiency in unmanned construction case much lower than that in manned construction case [14], [15]. So more observable area are required to solve the problem, and it can be better done by computer instead of human being's experience.
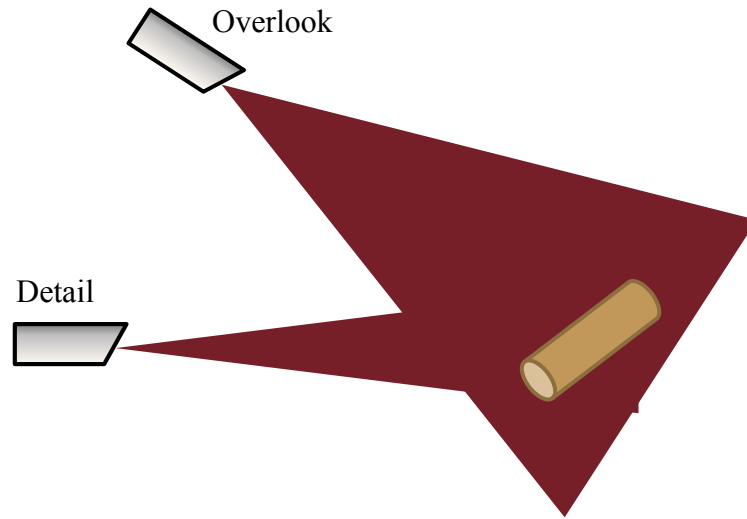
Figure 2.1 Different camera televising demand

2.2 Related works

In traditional method, observation targets and line-of-sight-based sensors such as cameras are always treated as points putting on a planar surface, which makes sensing range treated as a planer sector [19], [20]. It can be only used in 2D environment (Figure 2.2 (a)). Compared with this, there exists another improved method, through which observation targets and cameras can be put on non-planar surface [21]. Therefore, the sensing range of a camera becomes a cone instead of a sector. By using this method, it is possible to conduct camera arrangement on complicated terrain if height above local attitude is fixed. It is a pseudo-3D method (Figure 2.2 (b)). However, both of them cannot meet the requirement of unmanned construction, so I decided to make a novel line-of-sight-based model to enable observation targets and cameras put on variable height on complicated terrain to realize 3D camera placement (Figure 2.2 (c)).

(a)   (b)   (c)
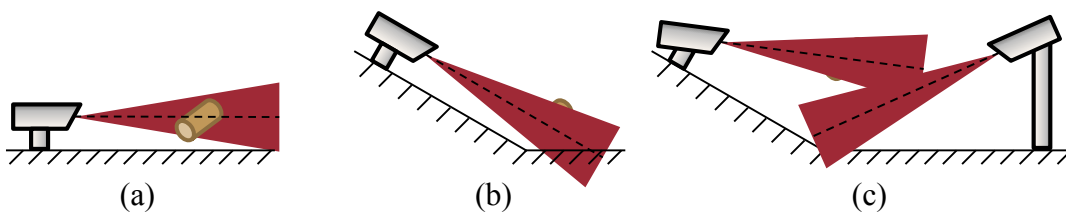
Figure 2.2 line-of-sight-based sensor model

2.3 Proposed model

2.3.1 Camera model

In order to simplify the camera model in 3D environment, the volume of camera is ignored as the models in 2D or pseudo-3D environment [19], [20], [21]. Thus we can treat our cameras in 3D model as a particle. Besides, it should own the following properties.

- Basic properties of a 3D object

    o Position

    o Orientation

- Perspective

    o Angle of view

    o Aspect ratio

    o Sensing distance

    o Up direction

2.3.1.1 Basic properties

In traditional 2D models, the position is always defined as (1).

$$p = (x, y) \tag{1}$$

In pseudo-3D models, it is described as (2), where z is defined in (3) [21].

$$p = (x, y, z) \tag{2}$$

$$z = g(x, y) \tag{3}$$

Here, z is the height above local position corresponding to x and y, which means that height of camera is not variable when x and y are fixed. In order to make it variable, height of camera above local position should be introduced. Thus I modify function g and rewrite it as (4).

$$z = g(x, y) + h \tag{4}$$

h means the height of camera above the local ground, which enables it to be positioned in different height at the same position.

The orientation of camera includes the yaw angle θ around the vertical axis and pitch angle ξ around the horizontal axis. Therefore, the property of camera i can be described as (5).

$$c_i = (p_i, \theta_i, \xi_i) \tag{5}$$

Thus, the basic property c_i of camera i in my novel 3D model is like Figure 2.3.
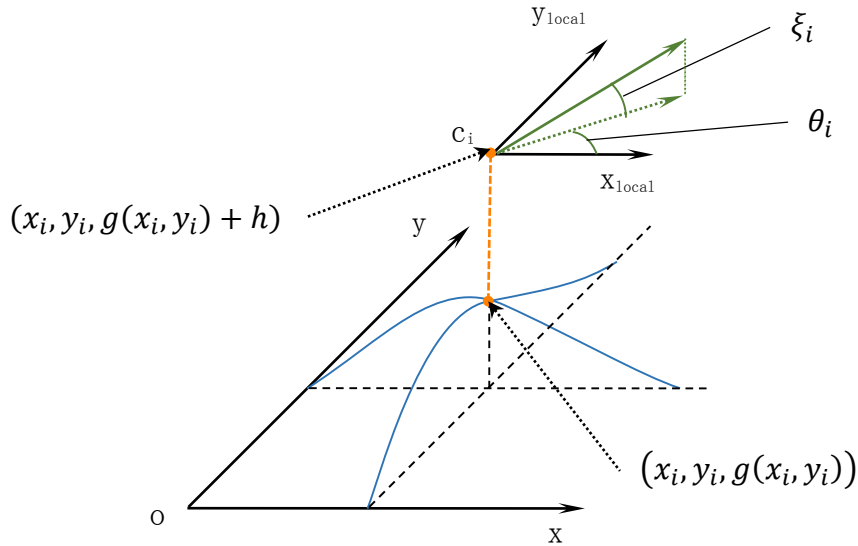
Figure 2.3 Basic property c_i of camera i

2.3.1.2 Perspective

The common process for taking a photograph of a desired view with a camera is shown in Figure 2.4.
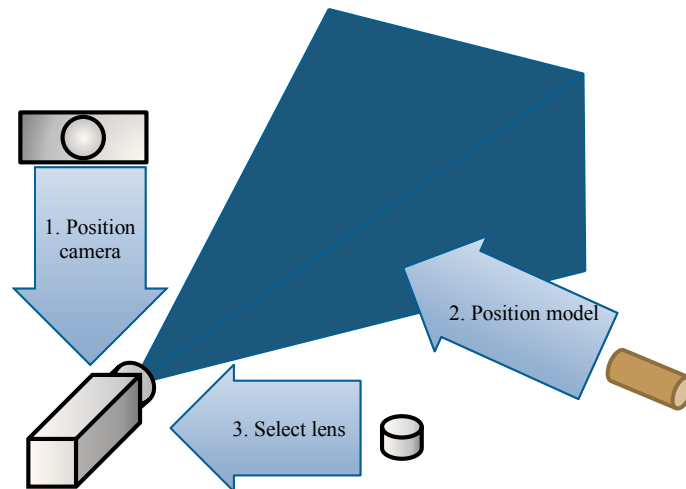
Figure 2.4 Steps in configuring and positioning the viewing frustum

The position and orientation of camera is not only decided according to those of object, but also decided by lens selected. As shown in Figure 2.5, when cameras with different perspective properties produce the desired view, position of cameras are different. In Figure 2.6, four factors of perspective is defined. Volume, shape of it, direction of view are decided by these four factors.
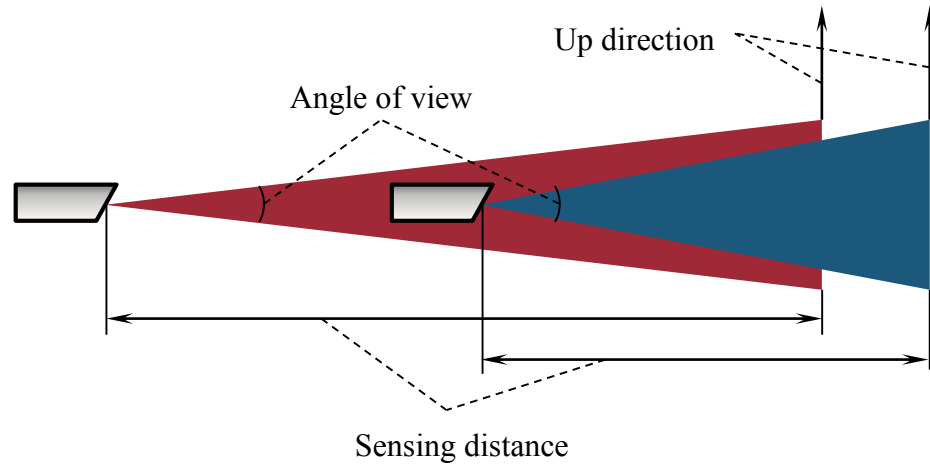
Figure 2.5 Different perspective properties in 2D view

$$\text{Aspect ratio} = \frac{Width}{Height}$$
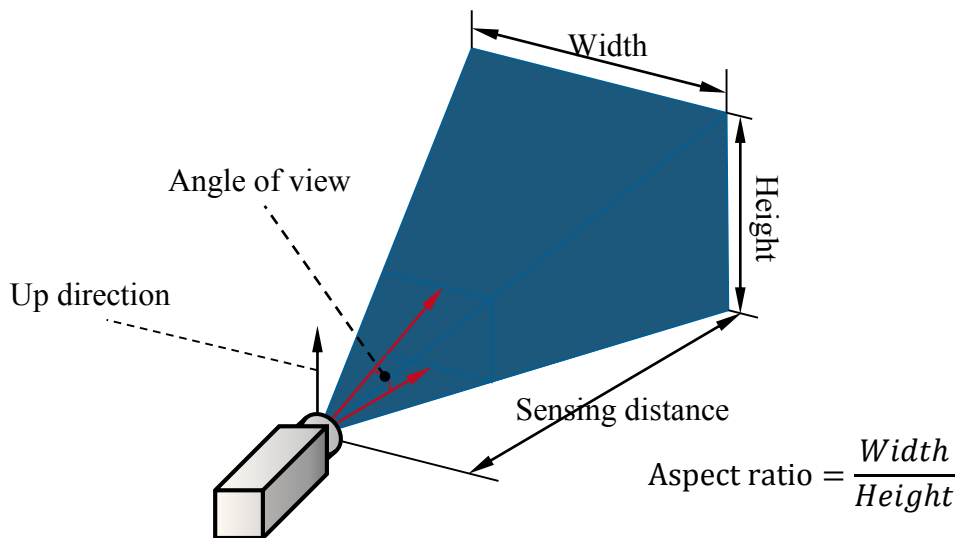
Figure 2.6 Perspective of a camera

Commonly, sensing distance is related to the resolution, the higher the bigger. It totally depends on the development of hardware. Aspect ratio depends on the properties of hardware as well. Here, we set the distance as 30 meters and aspect ratio as traditional 4:3. Up direction is always set as the reverse direction of gravity, which meets human habits. Moreover, angle of view depends on the requirement of tasks. In unmanned construction case, wide-angle lenses, standard lenses and long focus lenses are commonly

used [18]. Since telephoto lens sub-type are usually used instead of long focus lenses, they are referred to parlance as "medium telephoto lenses" and "super telephoto lenses" [24]. The previous one is widely used. Each angle of view is as follows [25].

- Wide-angle lenses: 64°~114°

- Standard lenses: 40°~62°

- Medium telephoto lenses: 10°~30°

These angles of view are always diagonal angles. They should be transformed to vertical ones to be used in computer graphics. After transformation, they become as follows.

- Wide-angle lenses: 41°~85°

- Standard lenses: 25°~40°

- Medium telephoto lenses: 6°~18°

In order to avoid the distortion of wide-angle lenses and simplify the computation, multifunctional lenses are used. The vertical angle of view is defined from 10°~70°.

2.3.1.3 Coverage

After deciding the basic properties and perspective, the coverage area of camera c_i can be determined. The coverage judgment of camera c_i observing point q can be described as a Boolean function (6).

$$Cov(c_i, q) \qquad (6)$$

The Boolean value is corresponding to the distance between camera c_i and point q, pan angle and tilt angle between c_i and q and visibility from p_i to q. They are written in (7~10)

$$d(c_i, q) = \|p_i - q\| \qquad (7)$$

$$p(c_i, q) = \text{atan2}\big((y_q - y_{p_i})/(x_q - x_{p_i})\big) - \theta_i \qquad (8)$$

$$t(c_i, q) = \text{atan}\big((z_q - z_{p_i})/\|p_i - q\|\big) - \xi_i \qquad (9)$$

$$v(c_i, q) \in \{0,1\} \qquad (10)$$

In (10), 0 stands for invisible and 1 stands for visible. Visibility judgment is shown in Figure 2.7.
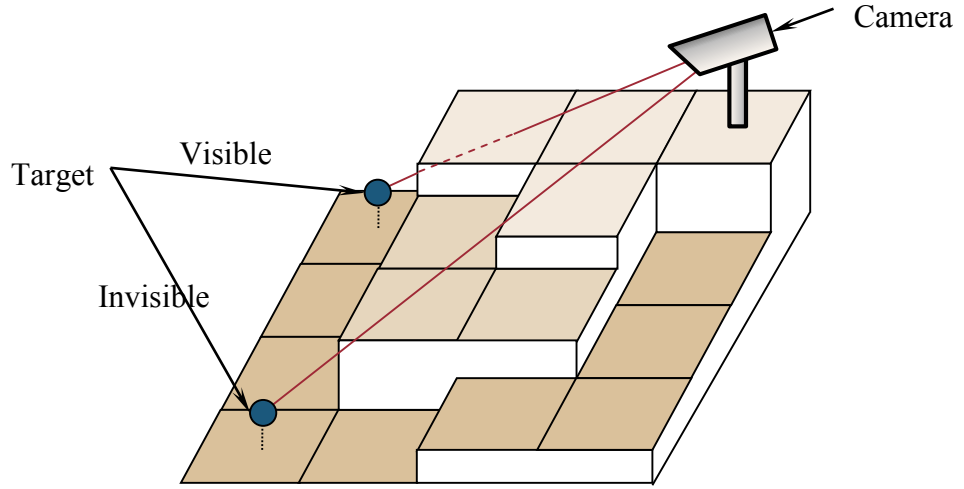
Figure 2.7 Visibility of target from camera

Thus, function (6) can be written as equation (11)

$$Cov(c_i, q) = f[\mu_d(c_i, q), \mu_p(c_i, q), \mu_t(c_i, q), v(c_i, q)] \tag{11}$$

According to the sensing distance, angle of view in pan plane and that in tilt plan, μ_d, μ_p, μ_t are defined as Boolean function. 1 stands for meeting the demand and 0 stands for not meeting the demand. Then, (11) can be represented as the product of μ_d, μ_p, μ_t and v in equation (12).

$$Cov(c_i, q) = \mu_d(c_i, q) \cdot \mu_p(c_i, q) \cdot \mu_t(c_i, q) \cdot v(c_i, q) \tag{12}$$

2.3.2 Target model

Observation targets are always treated as points with no volume [22]-[24]. Their positions are set above the local attitude, which is similar to that of cameras. We set the position of target q_i as (13), where z is defined in (14). Here, h depends on the height above ground.

$$q_i = (x, y, z) \tag{13}$$

$$z = g(x, y) + h \tag{14}$$

2.3.3 Environment model

In order to make the model adapt to most cases with variant complexity, the terrain is treated as a surface and described as (15). Here, z is described (16) and means the attitude of local position.

$$t = (x, y, z) \tag{15}$$

$$z = g(x, y) \tag{16}$$

Because the carrier of a camera is a vehicle, the area covered cannot be ignored. There-fore, I divide the terrain into $(a\text{-}1)(b\text{-}1)$ meshes averagely, where a stands for the number of rows and b stands for the number of columns. Each mesh is a little bigger than the covered area of carrier. These camera carriers are set at the center of meshes respectively. Each mesh can only set one carrier. In some cases, camera carriers are perhaps not stable on the local meshes, so tilt and smoothness of each mesh have to be taken into consideration.

Four vertexes of each mesh are rarely on the same plane only if the mesh is completely flat. In order to calculate the tilt and smoothness, all meshes are all treated as tetrahedrons.

In my proposed model, smoothness depends on the distance $\|\vec{h}\|$ between two diagonals $\vec{v_1}$ and $\vec{v_2}$, which are bolded in Figure 2.8 (a). Tilt depends on the angle τ between the green plane and black plane in Figure 2.8 (b). The green plane here should parallel to the two bolded diagonals.

(a)   (b)
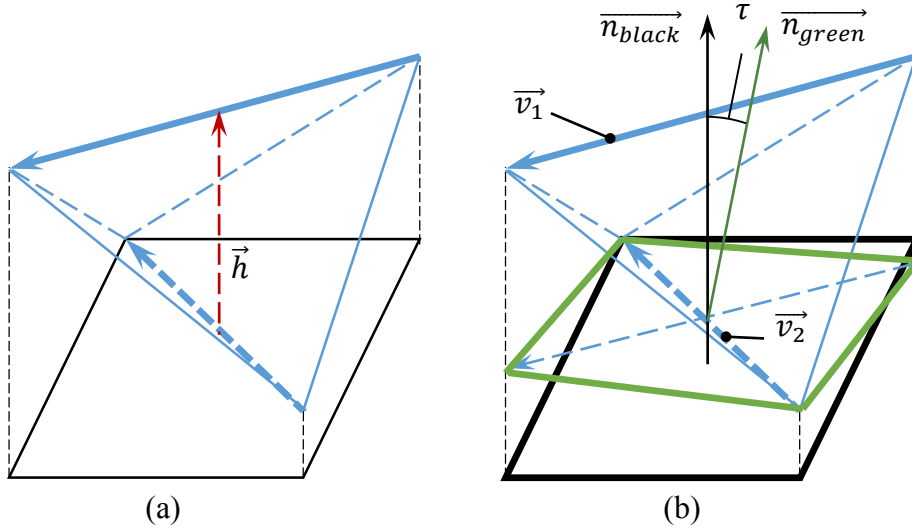
Figure 2.8 Smoothness and tilt of a mesh

The direction of unit normal vector $\overrightarrow{n_{black}}$ of the black plane is the reverse direction of gravity, thus it can be written as (17). Thus, the tilt and smoothness of the mesh can be measured by (17~20), where $\overrightarrow{n_{green}}$ stands for the unit normal vector of the green plane.

$$\overrightarrow{n_{black}} = (0, 0, 1) \tag{17}$$

$$\vec{n}_{green} = (\vec{v_1} \times \vec{v_2})/(\|\vec{v_1}\| \cdot \|\vec{v_2}\|) \tag{18}$$

$$\|\vec{h}\| = |\vec{n}_{green} \cdot \vec{h}| \tag{19}$$

$$\tau = \text{acos}(\vec{n}_{green} \cdot \vec{n}_{black}) \tag{20}$$

When four vertexes are almost on the same plane $\|\vec{h}\|$ can be small, which means the mesh is smooth. In this case, the tilt angle between the plane and the ground plane will be calculated. Large τ stands for large tilt angle; large d stands for low smoothness. It is possible to position a camera carrier on the center of the mesh only if both $\|\vec{h}\|$ and τ are small.

2.4 Evaluation method

The coverage possibility of a single required observation target q by camera c_i can be described as (12). Therefore, the coverage possibility of a single required observation target by all cameras can be written as follows:

$$Cov(C, q) = 1 - \prod_{i=1, \cdots, n}(1 - Cov(c_i, q)) \tag{21}$$

Here, C stands for the set of all possible cameras and n stands for the number of all candidate cameras. In traditional method, the position of cameras is not limited, so all targets can be observed if there are enough cameras. However, in the cases with limitation, some targets are maybe not observed by all cameras. Therefore, we propose the following equations to evaluate the coverage rate more precisely.

$$N_{observable} = \sum_{i=1, \cdots, m} Cov(C, q_i). \tag{22}$$

$$N_{observed} = \sum_{i=1, \cdots, m} Cov(N, q_i). \tag{23}$$

$$C_O = N_{observed}/N_{observable} \tag{24}$$

$$D = N_{observable}/m \tag{25}$$

N is the set of all used cameras; m stands for the number of observation targets required; $N_{observable}$ means the number of total observable targets; $N_{observed}$ means the number of observed targets by used cameras. The rate C_O of the observed targets in observable ones is used to evaluate the coverage rate. Meanwhile, the rate D of the observable targets in required ones can describe the complexity of environment observation

2.5 Optimization Strategy of Camera Placement

According to our proposed models, there are more limitation for camera placement. Also, camera placement computation and evaluation standard are extended to 3D space.

We cannot thus simplify the environment as a narrow flat plane or a terrain surface as before. Traditional methods like iteration depend on initial values too much, making simulation difficult to find an optimal or suboptimal solution. To solve the problem and verify the feasibility of our proposed models, we make some assumption to simplify the computation for our case.

2.5.1 Preparation

In order to reduce the amount of computation in camera placement, we calculate the possible configuration of each camera in advance. Each possible camera configuration should cover at least 1 required target point, or it is unacceptable. For each camera ci, we have the equation as follows:

$$t_i = \sum_{j=1,\cdots m} Cov(c_i, q_j). \tag{26}$$

Here, t_i means the number of observable targets by camera c_i. If ti is larger than 0, each q, which makes $Cov(s_i, q) = 1$, will be stored in set T_i, which is described like:

$$T_i = \{q_1, \cdots q_{t_i}\} \tag{27}$$

Here, q_1 to q_{ti} are all observable targets from camera c_i. Considering each camera should shoot required target in its center area, if all measured targets are compressed, we can treat the possible orientation of the camera as the vector from a camera to a required observation target, to simplify the computation by avoiding probable unacceptable property. Therefore, the total observed targets can be described as the union T of T_i

$$T = \bigcup_{i \in N} T_i \tag{28}$$

2.5.2 Proposed placement strategy

From the analysis of disadvantages of traditional strategies, we propose a practical camera placement strategy. The algorithm of the proposed strategy is shown in Fig. 4. In the pseudo-code of camera placement, *num* stands for the number of cameras used; *flag* stands for the number of patterns when the number of cameras is decided.

```
for t = 1 to n
    C_candidate(1, n) = {c_i}
end for
flag = 1, num = 1, m = n, P_req = 30%
do (if P_req > 0)
    flag = 0
    for  t = 1 to m
        for r = 1, …, n
            if sensor n meet the all environment requirement and the relationship with each sensor in C_candidate (num, t)
                then
                    calculate P_r
                    if P_r > P_req
                    then
                        flag = flag + 1, C_candidate(num + 1, flag) = C_candidate(num, t) + {C_r}
                    end if
                end if
        end for
    end for
    if flag > 0
    then
        num = num + 1
    else
        if C_O < 90%
        then
            P_req = P_req – 5%
        else
            P_req = P_req – 2.5%
        end if
    end if
end do
```

Figure 2.9 Pseudo-code of camera placement in our strategy

2.5.2.1 Traditional strategy

In traditional method, iteration is always used with no practical limitation, which makes the result undependable. There are two main disadvantages [25][30]. First, the initial property of some cameras is not dependable. In some cases, some cameras are still used even if it can hardly observe any required targets. This phenomenon always occurs in large maps with a lot of large valueless observation areas. Random initial values probably set some cameras in such areas, which causes that these cameras have no chance to exchange with better options. Second, the initial number of cameras are always determined by the size of maps. Such iteration methods are always used in the situation where all targets are required to be observed. Random initial cameras are thus nearly

homogeneously distributed in the scene. However, in many practical cases, not all of the targets should be observed and cameras cannot be placed in some areas. Therefore, the initial cameras should depend on the scene and the result should reflect the requirement. Besides, some other practical requirements should be concerned, such as the safe distance between camera and nearest target, the required distance between each two cameras, the landform and etc. We thus try to propose a strategy to obtain the optimal placement of cameras. And, it should adapt to most cases mentioned above.

2.5.2.2 Proposed strategy

In our proposed strategy, the objective should take two things into consideration, the coverage rate and the profit when increasing a new camera. The definition of coverage rate is already mentioned in the previous section. We here set the target coverage rate C_O 90% to ensure a high observability. The profit means the rate of value of increased targets and the total value of observable targets when a new camera c_i is placed in the scene. The profit P_i is given by (29)

$$\begin{cases} P_i = \sum_{j=1,\cdots,n_1}(q_j w_j)/\sum_{k=1,\cdots,n_2}(q_k w_k). \\ q_j \in T_i \\ q_j \notin T \end{cases} \qquad (29)$$

The w_i means the value of required target q_j, which always concerns the importance of q_j. q_j belongs to the T_i of increased camera c_i and not belongs to the current T which is not refreshed by increasing T_i. n_1 stands for the number of such targets q_j; n_2 stands for the number of total observable targets in the scene. In our research, we set the value of each target 1, so the profit P_i equals to the increased coverage rate. We define the acceptable profit P_r as 30% initially to ensure the cameras which can observe more targets have high computational priority as well as reduce the computational cost. If P_i is larger than P_r, the increased camera i is acceptable. If such P_r does not exist and C_O is less than 90%, the coverage lower limit used in our experience, required P_r will reduce 5 percentage points until it becomes 0. If C_O is more than 90%, P_r will reduce 2.5 percentage points until it becomes 0. The $C_{candidate}$ with the largest value will be outputted as the optimal result.

2.6 Experiment

2.6.1 Camera carrier and Camera

Considering the property of camera carrier, one carrier can only carry one camera. Additionally, carrier and the camera carried is on the line which is parallel to the gravity, which means the position of carrier and camera can be written in the equations (30~31).

$$P_{vehicle} = (x, y, g(x, y)) \tag{30}$$

$$P_{sensor} = (x, y, g(x, y) + z_r) \tag{31}$$

Because carriers here are all vehicles. Positions of these carriers should be different, which makes arbitrary two cameras are not on the same line parallel to the gravity. Besides, they should not be positioned at the same place as required target. Considering the size of camera can be ignored compared with environment size, all cameras are defined invisible this time. Assuming the meaningful distance above the ground is 0 to 6 meters and the interval of two targets on the same vertical line is 1 meter, the position of observable target q_i can be written as (13~14). Here, x, y and h are non-negative integers. In addition, $z \in [0,6]$. In order to observe such targets and simplify the computation, the interval of two neighbor possible position of one camera on the same vertical line is 1 meter, and the range of the vertical position of each camera is 3 to 9 meters. The position equation of the camera c_i should be (2, 4). And x, y, h are all non-negative integers and $h \in [3,9]$.

2.6.2 Environment and Required Targets

We create a terrain map to verify the feasibility of our proposed practical placement strategy (Figure 2.10).
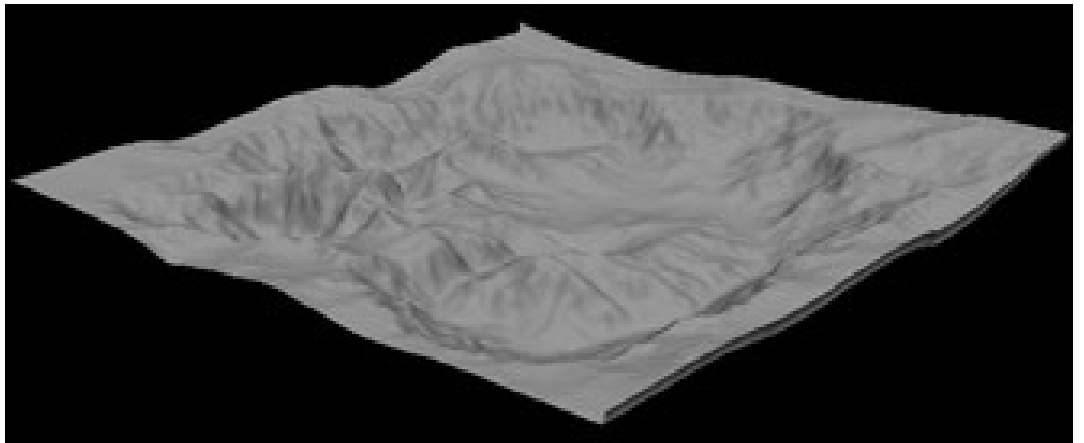
Figure 2.10 Terrain generated.

The 256 x 256 height map used in our environment (http://docs.garagegames.com). Each pixel is thought as a square mesh with the size of 1 meter by 1 meter.

The maximum attitude difference is defined as 50 meters, and the distance between two neighbor pixels is 1 meter. Next, we divide the terrain into meshes 50 by 50. And, demand the tilt angle should be less than 30 degrees and the d in Fig. 3 should be less than 0.3 meters. We put two work centers in the environment. The first center is put at (25, 25) and its range is 7 meters, the other one is at (30, 40) and its range is 15 meters. Namely, in the 51x51x7 environment, the number of required targets is 171, about 0.94 percent of the total vertexes of cubes. Notice, the required targets are set above the vertexes of meshes if needed, and carriers are positioned at the center of meshes if possible. Moreover, the property of each possible camera should be specified according to the tilt angle, pan angle, shooting range and pitch angle requirement. They are all given in Table 1.

Table 1 Parameter value of proposed camera model

Tilt angle	Pan angle	Pitch	Range min	Range max
55.4°	70°	-60°–+60°	0m	30m

2.6.3 Results

First, according to the tilt requirement and smoothness requirement as stated above, the terrain map is rendered as Figure 2.11. Gray meshes mean impossible positions for camera carriers and white meshes mean possible ones.
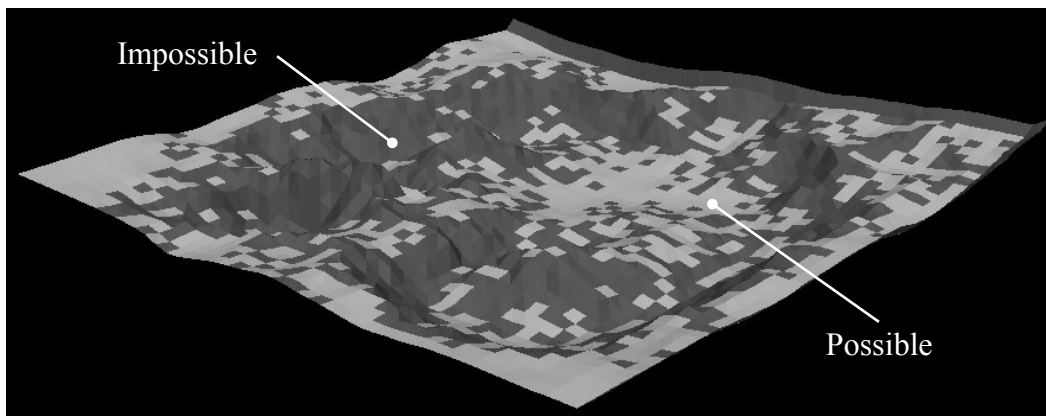
Figure 2.11 Divided terrain

Considering the position conflict of the required targets and cameras, the possible position of carriers are then updated, as shown in Figure 2.12. Possible camera positions are rendered as blue diamonds and surround required targets.
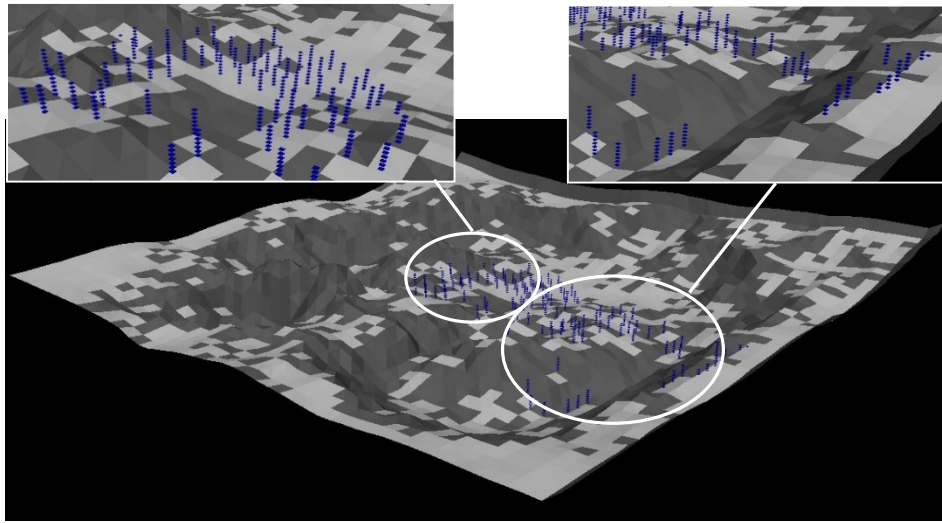
Figure 2.12 Possible camera positions

Then, we can obtain the number of total observable required targets, which is 161, which means that ten points cannot be observed due to installation limitation of camera carriers. Additionally, we can find that 771 camera positions and 117 carrier positions are available according to the camera property and ground property. At the moment, the optimum strategy is used to calculate the possible camera placement patterns. As a calculation result, 144 required targets can be observed as the maximum by 4 cameras. Observable targets are shown in Figure 2.13 and cameras placed in the environment and their coverage areas are shown in Figure 2.14. Red diamonds means the required targets.
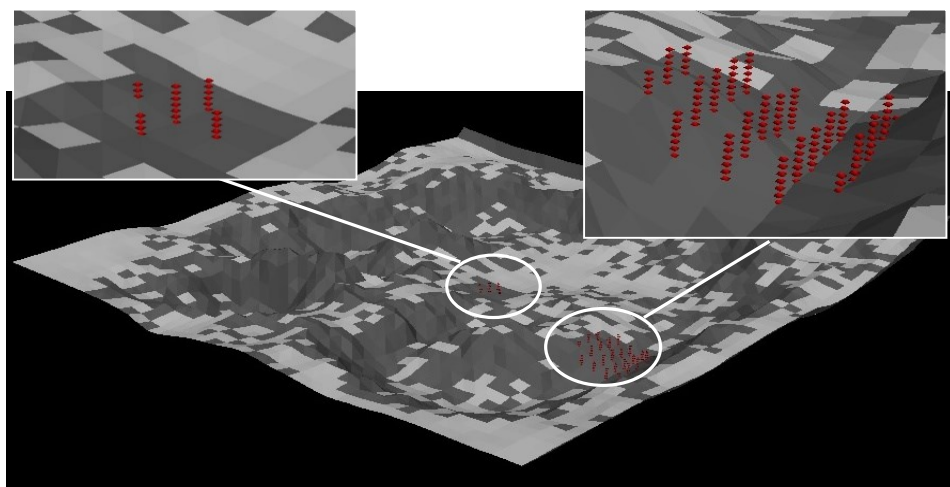
Figure 2.13 Updated possible camera positions and required targets
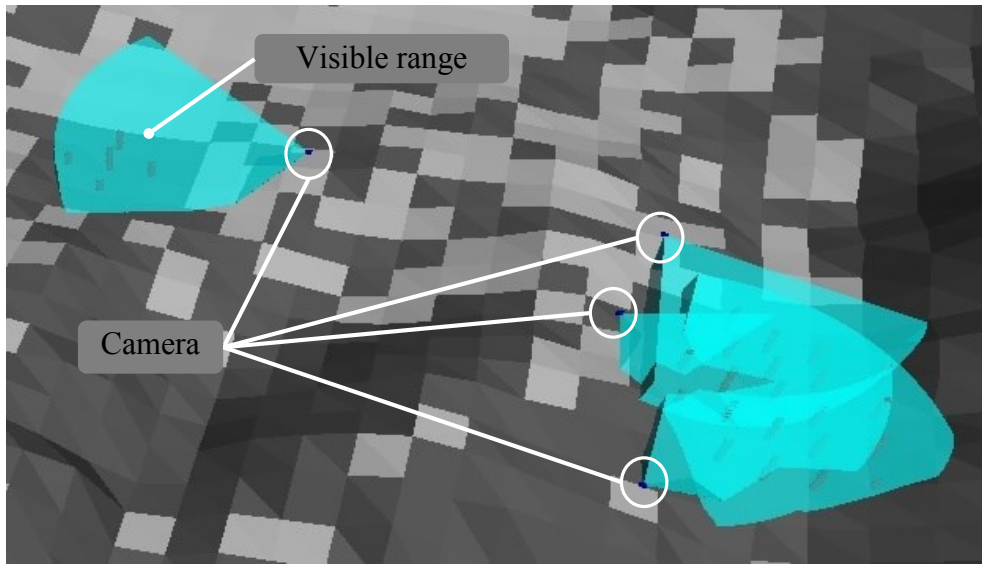
Figure 2.14 Outputted camera placement pattern

4 cameras can observe 144 required targets. The coverage area of each camera is shown as the teal volume and the camera is set at the top of it.

Table 2 Trade-off of profit and cost

Number of cameras	Number of observed targets	Coverage rate
1	51	31.7%
2	100	62.1%
3	123	76.4%
4	144	89.4%

The relationship among the number of cameras, number of observed targets, and coverage rate is listed in Table 2. The total observable targets: 161. We can find that the coverage rate C_O is 89.4% by using just 4 cameras. In other words, the coverage rate does not increase more than 2.5 percentage points (i.e., approximately 4 required targets) even if a camera is increased. The result of the coverage rate is close to the expected one, which means our proposed model and evaluation method are able to describe cameras in 3D space.

2.7 Summary

The 3D camera model which describes the property in 3D environment and practical camera placement strategy which can be combined with other requirements and condi-

tions to realize optimum placement in 3D world, especially for remote control field. Trade-off relationship between coverage rate and cost, i.e., the number of cameras, should be quantified because it can be helpful to make camera placement strategy more practical. The result of camera placement experiment clarified the relationship between the number of cameras, number of observed target, and coverage rate. From the results, it is confirmed that the proposed camera model in 3D environment and evaluation method is effective in 3D space.

Junjie YANG - February 2016

3 SIMULATION PLATFORM

3.1 Requirements

Many technologies have been developed and realized in real world by using physical resources, which is important for practical fields. Compared with the simulation in virtual world, the result in real world is more dependable. However, there are some disadvantages as following, which cannot be ignored.

- High cost

- Time-consuming preparation and experiment

- Irreproducible environment

In order to prepare the unpredictable disaster, operators should practise in different environment repeatedly and novel technologies should be quantitatively experimented. Therefore, developments in real world are not applicable for that purpose. As a common way to address these problems, virtual reality (VR) simulators are often used, such as surgical operation simulators [31], [32] and automobile driving simulators [33]. In contrast to physical method, VR simulators possess the advantages such as lower cost, repeatability, quantitative evaluation possibility and no physical constraints. Thus, VR simulators are widely used for operational skill training, the derivation of comprehensive problems and improvements [31], and the development of advanced technologies, even though physical behavior is difficult to reproduce precisely.

In the construction machinery field, a VR simulator for coaching machine operations has been developed [34]. However, no VR simulators have been developed to solve vi-

sion support problem in the unmanned construction field. Thus I decided to establish a VR simulator to develop the two vision support systems mentioned in chapter 1.

3.2 Development of man-machine interface

Man-machine interface is separated in to parts. One is hardware part which is used to control the construction machine in the established virtual world. The other is software part to establish the virtual world.

3.2.1 Hardware

Hardware is mainly composed by a PC, a monitor, two AD boards (PCI-3168c), a voltage source and machine controllers. In order to make experiment easy to be analysed quantitatively, conventional joysticks are used as those used to control real machine. The connection diagram is as Figure 3.1. One AD board is used for machine control and the other one is used for environmental camera control. A power box connected to AD boards which supplies 5V voltage to the board. In addition, a PC is used to process digital signals and run the simulator, and a monitor (1920 x 1080) displays the output operation interface.

5V
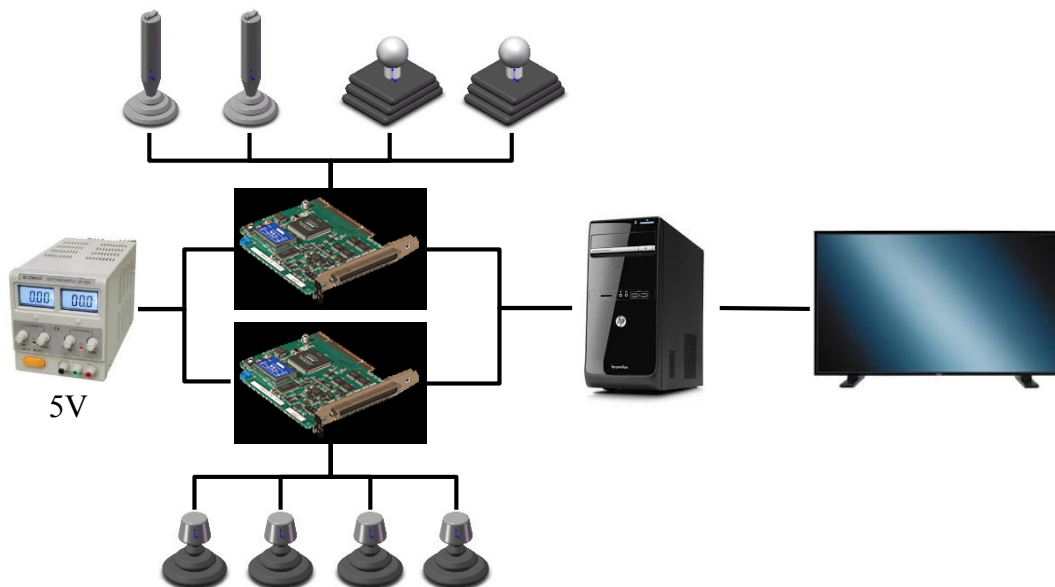
Figure 3.1 Connection of hardware
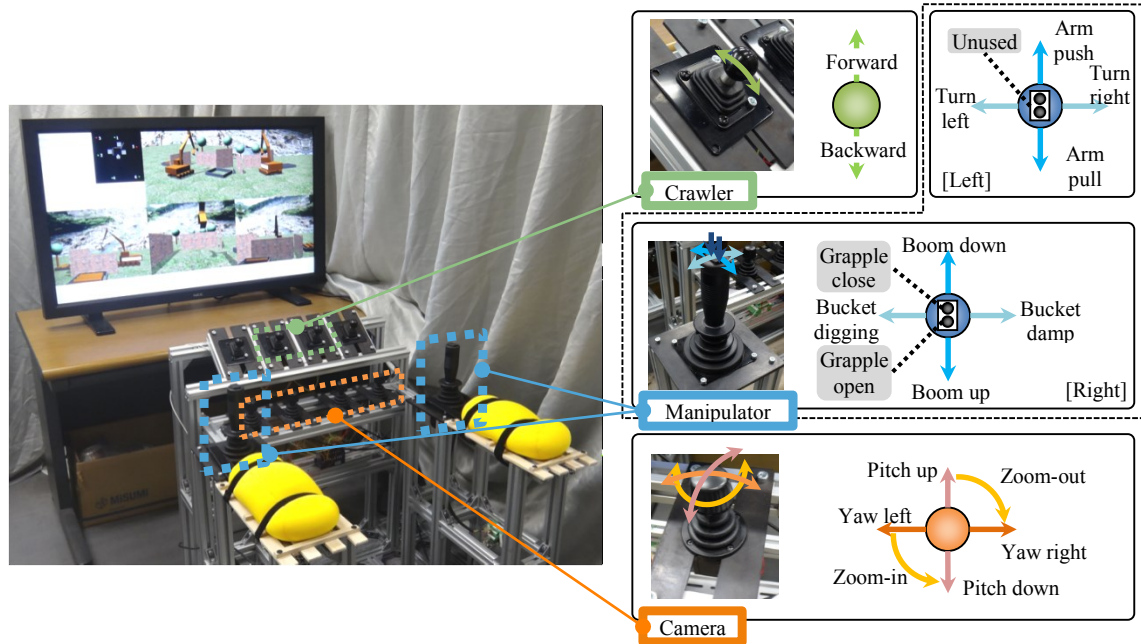
Figure 3.2 Composed hardware system

After assembling these components, the hardware is shown in Figure 3.2.

Two manipulator joysticks are located on each side of operator while four camera joysticks and two crawler joysticks are located in front of the operator. Crawler joysticks are a little higher than camera joysticks. This configuration refers to the real construction machinery one, which makes operators convenient to reach all joysticks. Additionally, JIS standard is referred to define the instruction of machine-control joysticks.

3.2.2 Software

The operation interface is written by visual C++ and divided into two components. One is to conduct dynamic computation to draw out the position and posture of each objects. It is done by Open Dynamic Engine (ODE). Forces, moments and speed of each object in VR environment can be recorded as the evaluation accordance. The other component is rendering, which can render the VR scene by reading the shape, position and posture of objects from dynamic computation component. Here, OpenGL is used to realize it. In the following sub sections, operation interface, controlled machine, cameras and other objects in the scene will be introduced.

3.2.2.1 Operation interface

In the simulator, six 4:3 viewports with the same size are used. And they are arranged as a (2 x 3) pattern [13] [35], which is a common configuration to represent the machine information. The arrangement is symmetric to operator when the bottom middle view-

port is opposite to him/her. This viewport is defined as the main viewport, and others are used to support it. The closer to the main one, the more important it is.

Considering the cab view is the only view used in manned construction, it is set at the main viewport to adapt to the operators' habit. In order to inform operator the machine position and the corresponding status, a 2D mini map will be displayed with machine status and camera connection in another monitor. The other four are used to show views from environmental cameras. A viewport arrangement sample is shown in Figure 3.3.

(i) Status and mini map   (ii) Environment view   (iii) Environment view

(iv) Environment view   (v) Cab view   (vi) Environment view
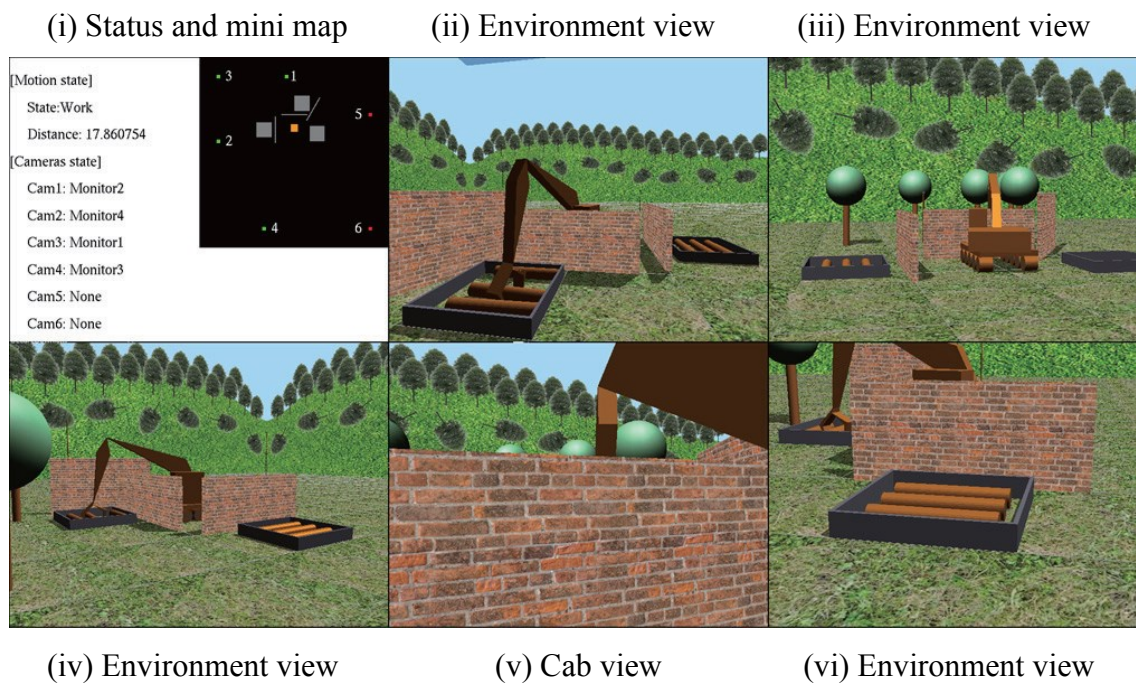
Figure 3.3 Viewports configuration sample

3.2.2.2 Controlled machine

In the VR simulator, construction machine is thought as a dismantling machine. The motions of it can be classified in two kinds, movement of machine itself and manipulation. The movement of machine is based on driving two crawlers, which can realize the differential control. Rotation of each link on manipulator and pivot is to realize the manipulation and grasping. The construction machine image with rotation joints of manipulator is shown in Figure 3.4.
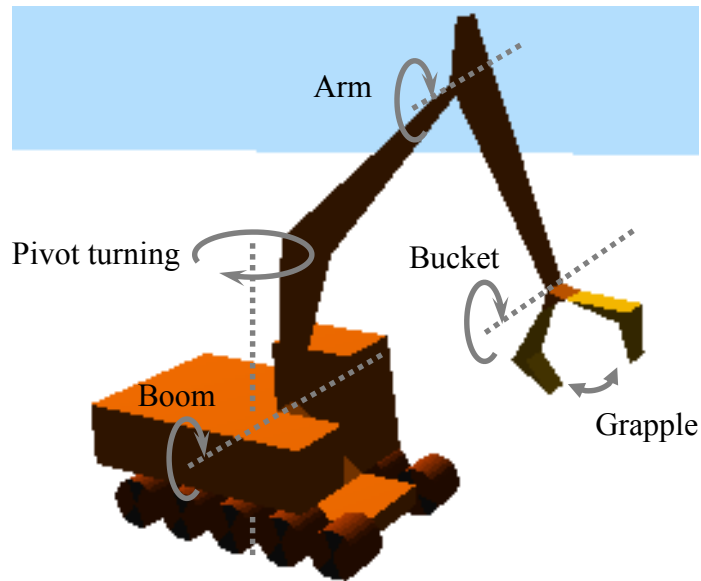
Figure 3.4 Construction machine prototype in virtual world

3.2.2.3 Cameras

Two types of cameras are used. One is adjustable camera, which is shown in Figure 3.5. It is used as environmental camera, which is thought carried on vehicle. The position of each camera is fixed, which is defined in advance. Each camera can rotate around pitch axis or yaw axis to adjust its posture. And it can also change its angle of view to adjust the magnification. The control method of each camera is shown in Figure 3.2.
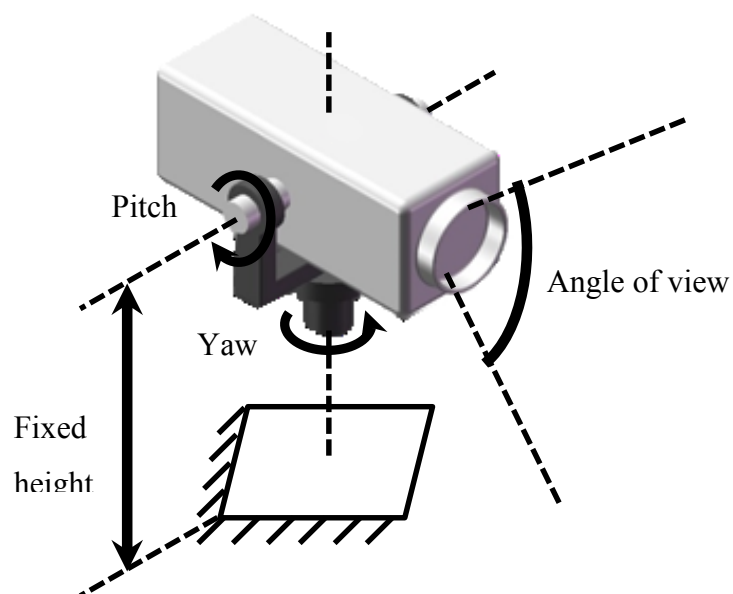
Figure 3.5 Adjustable camera prototype

The other type is camera carried on work machine. It is installed in the cab of work machine with fixed position and posture. And it faces the same direction as the front direc-

tion of the cab, which makes the view from it like that from operator's eye in manned case. And the angle of view is always fixed as 70° to supply a wide vision.

By using these two kinds of cameras, views are shown as those in Figure 3.3.

3.2.2.4 Others

Except the machine and cameras, there also set pipes, boxes, walls and mounds in advance to simulate the work site in virtual world. Pipes are transportation objects, and they are ordered collected to a specified box from others as tasks. Walls and mounds are thought to raise the difficulty. And they are abandoned to contact construction machine. Mounds make the machine moves tougher and make pipes more difficult to be reached. Detailed configuration of these objects in experiment scene will be introduced in the following chapters.

4 AUTONOMOUS CAMERA CONTROL SYSTEM

4.1 Requirements

4.1.1 Targets

At first, the observation targets should be made clear according to the condition of machine. It is used to decide what kind of visual information should be informed of. According to the analysis of motion of machine, there are mainly two patterns:

- Movement, which is realized by driving crawlers. (Figure 4.1 (a))

- Manipulation, which is realized by rotating the joints on manipulator or machine pivot (Figure 4.1(b)).

(a)   (b)
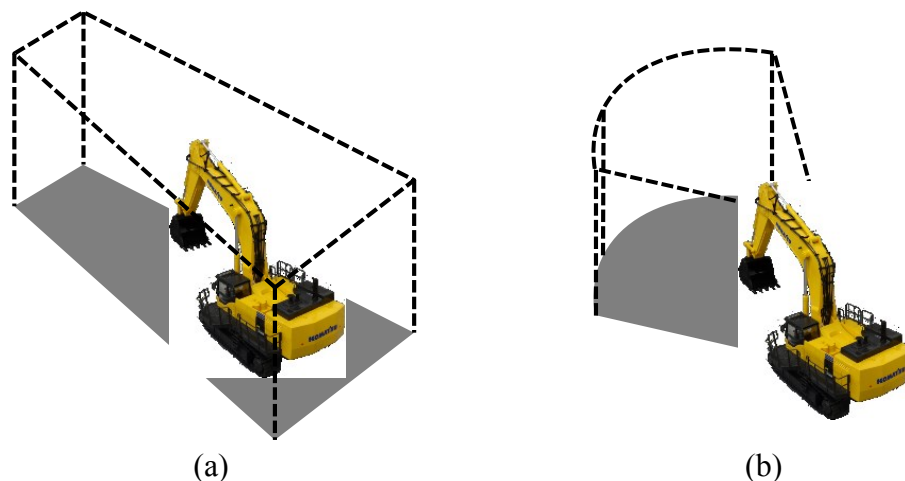
Figure 4.1 Different motion patterns of machine

Except them, the view in front of the cab should be processed as well to make operator feel like sitting in the cab of a real machine as mentioned in Section 1.4.1. By using this view, it can help operator find his/her position and the situation around the machine because it is a first-person view. They are all tactical views. Besides the targets processed for tactical demand, information for strategy demand should be processed as well. According to [14], a mini map by GPS is a common way to inform operator of the current position of machine in the entire work site. Thus, a mini map is used to meet visual strategy demand of unmanned construction. In this case, the entire scene of work site is the observation target.

4.1.2 Display

In order to make the visual information understood by operators, how to process the information should be discussed. In static case, as mentioned in Section 1.4.3, at least 2 views from different directions should be used. Because of the continuous motion of objects in the view, operator can refresh and revise his/her recognition of work site all the time. In this case, 1 view is enough if the motion in it is big and understandable. Contrarily, in some cases, the motion is too slight to be observed clearly, such as manipulation. Thus, 2 views are recommended as that in static condition. Consequently, 2 third-person views each from environmental cameras are used to display the movement and manipulation. Including a cab view and a mini map view, totally six views are needed to express the real-time condition of machine.

- A cab view

- A mini map view

- Two manipulation views or end-effector views

- Two machine views

Except the number of different views, the photography demand should be discussed as well. Commonly, when manipulation is conducted, operator tends to know the position in manipulator planer and that in the planer vertical to manipulator planer. By using these two views, operator can know the detail position of end-effector and conduct manipulation more precisely. In this case, the image of manipulator or end-effector should be displayed in the center of the view and take most space in the view. Considering the height limitation of environmental cameras, the view is on the same height level of machine itself. Thus, view from the back of machine will cover the rear of machine instead

of end-effector, which is not expected. So a side view and a front view of manipulator will be used to describe manipulator.

When machine moves, the relationship between machine and environment around it is important. Thus, machine itself should be observed from different perspectives, such like side, front or back. So the combination of side view and front view, or that of side view and back view will be used. In addition, in this case, the image of machine should cover the center space of the view, because the other space will be used to show environment to describe the relationship with machine.

4.2 Roles of cameras

On the basis of analysing the camera images that needs to be provided for operators in specific situations, the camera roles R_x needed for autonomous control were defined.

4.2.1 Analysis of required images depending on situation

In presently available systems, environmental cameras are not controlled during machinery operations, and they provide basic panoramic views. Even with operators that have superior spatial ability, the back sides of objects cannot be visualized in the present systems. The purpose of the system proposed in this study is to enable operators revert the original scene by observing views from environmental cameras. Toward that end, I did the following. At the beginning, I analyzed which images of each targets would be needed for improving work performance and that are common to various types of work content. Sequences common to disaster response work is used as the basis for choice, such as base movement, reaching, work using an end-effector (for digging, bending, and grasping), transporting, and releasing. The relationship between required operations and camera images for each work situation are listed in Table 3.

Table 3 Relations between required operations and images in each work situation

Situation	Required operation	Required information	Required image
Movement (Operations using crawlers)			
Forward Backward Turning	Safe and efficient positioning	Ground condition Self-position in site Surrounding barricade	Surrounding image involving machine
Manipulation (Operations using manipulator or machine pivot)			
Reaching	Efficient operation without coming into contact with obstacles	●Distance between object and manipulator ●Sense of depth	Manipulator's posture in relation to neighboring obstacles
Work using end- effector	More precise control of position and posture of the end point	●Precise distance between object and end point ●Situation of contact	Enlarged image to show the distance between the end point and an object
Releasing	More precise control of position and posture of the end point	●Precise distance between object and end point ●Situation of contact	Enlarged image to show the distance between the end point and an object
Transport	Careful operation not to drop object and contact to obstacles	●Situation of place where object is transported ●Ground condition	Manipulator image involving spaces displaced in its moving direction

- **Base movement:** Safe and efficient positioning of the machine body is needed. Images of the machine and the surroundings are needed for confirming ground conditions and the relative positions of obstacles and the machine.

- **Reaching:** Efficient operation that avoids contacting obstacles is needed, but accurate depth perception is essentially difficult during teleoperation work. Images that indicate the posture of the manipulator in relation to neighboring obstacles are required for this task.

- **Work using end-effectors:** These activities include digging, bending, pressing, and grasping, where the endpoint applies force to an object. The position and posture of the grapple must be controlled precisely because an excessive application of force in a wrong direction can lead to debris collapse or toppling of machines. Enlarged end-point images are therefore needed to show the distance between the end-effector and an object. Images of this type are also needed during the releasing phase.

- **Transport:** Careful and efficient operation is required during transporting of objects because dropping those objects or contacting obstacles can create extremely dangerous situations. Images of space in the manipulator's direction of motion are needed for pre-entry visualization of the space the manipulator will enter.

4.2.2 Imaging objects and imaging modes

Camera roles correspond conceptually to required images, as listed in the last column of Table 3. Note that Table 3 lists only common required images, so camera roles should be flexible and reusable, with arbitrary definitions permitted. We thus decompose camera roles into two elemental components: an imaging object, which a camera should capture within a specified imaging range; an imaging mode, which represents controls related to the imaging object. Here, the most fundamental imaging objects and modes are defined.

4.2.2.1 Imaging objects

To characterize an imaging object O_x, I need to define the imaging center and the imaging range. The system should provide enough spatial information to enable operators to understand the relations among machines, objects, and the environment. Thus, the imaging center should be based on the state of the machine and be defined uniquely for each object, with common object types available to adapt the various configurations of the machine. The imaging range can be determined from the imaging center and work situation, as mentioned in Section 4.2.1. From the above analysis and the contents of Table 3, I defined three imaging objects O_x, as shown in Figure 4.2.

Thus three detail imaging objects O_x were derived:

- **Machine (O_1):** This is used to provide a machine-centered overview in the work environment. The imaging center is the center of the pivot joint. The imaging range is the machine, including some of the surrounding environment.

- **Manipulator (O_2):** This is used to provide the movement and posture of a manipulator to allow recognition of the spatial relation of the manipulator to obstacles. The imaging center is the geometric center of the manipulator. The imaging range is the whole of the manipulator.

- **End-effector (O_3):** This is used to provide the movement and posture of the end point to allow precise recognition of the spatial relation between the target object and the obstacles. The imaging center is the center of the grapple. The imaging range is the whole of the end-effector.
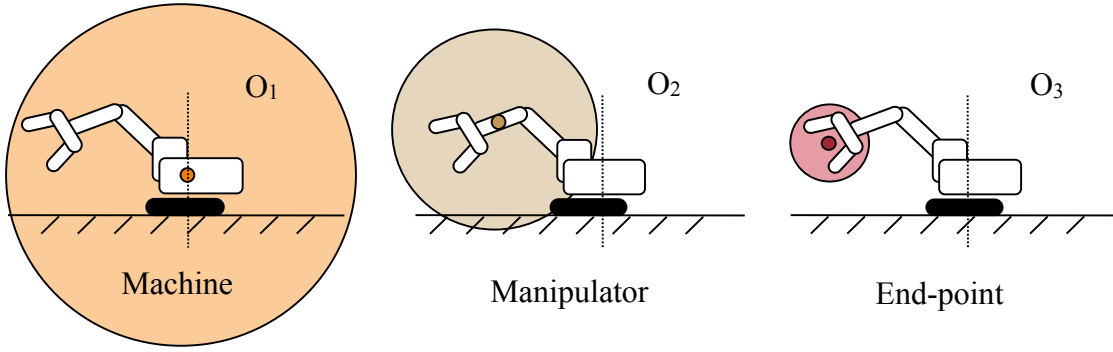
Figure 4.2 Image objects

4.2.2.2 Imaging mode

To define an imaging mode M_x, we need to define ways of controlling the orientation and zoom ratio of cameras. The fundamental mode necessary for camera control is tracking of an imaging center and zooming to the correct imaging range. Advanced modes, but ones that are essential for teleoperation, are the posture mode, which enables operators to visualize depth, and the trajectorymode, which provides information about the environmental that will be entered in the near future. From the above analysis on the basis of control strategies for eye-in-hand robotic manipulators [44] and the information in Table 3, we define four fundamental imaging modes M_x (Figure 4.4):

- **Tracking (M_1):** This provides a continuous view of an imaging object to allow monitoring its state. The tracking mode keeps the camera following an imaging object by controlling the yaw and pitch rotations according to a simple forward kinematics algorithm.

- **Zoom (M_2):** This provides images at a resolution suitable for the situation (from wide to detail). This can be used for viewing the spatial relations between the machine and target objects and obstacles. The zoom mode adjusts the zoom ratio according to the situation as identified by a work-situation identification method (these are explained in Section 4.2.3.2).

- **Posture (M_3):** This provides images from the direction perpendicular to the imaging object, which makes perception of depth easier. The posture mode adjusts the camera orientation within postural constraints, as shown in Figure 4.4 (c). This is quite important for visual feedback in teleoperation systems. This role focuses on recognition of the manipulator's state.

- **Trajectory (M_4):** This provides images in which the imaging center is displaced in a specified movement direction, which helps with effective planning and

operation for improving safety and efficiency. The trajectory mode adjusts the camera orientation according to the direction of the end point's movement, as shown in Figure 4.4 (d). This role focuses on spatial recognition.

Tracking (M_1)
Tracking an imaging object while retaining the imaging range

Zoom (M_2)
Adjusting the zoom ratio according to the work situation

Zoom-in

Zoom-out

Camera

Posture (M_3)
Providing an image from the direction perpendicular to the imag-

Trajectory (M_4)
Leading the imaging object in its movement direction

Camera-1

Caemra-2

Camera

Figure 4.3 Fundamental imaging modes

4.2.3 Definition of camera roles
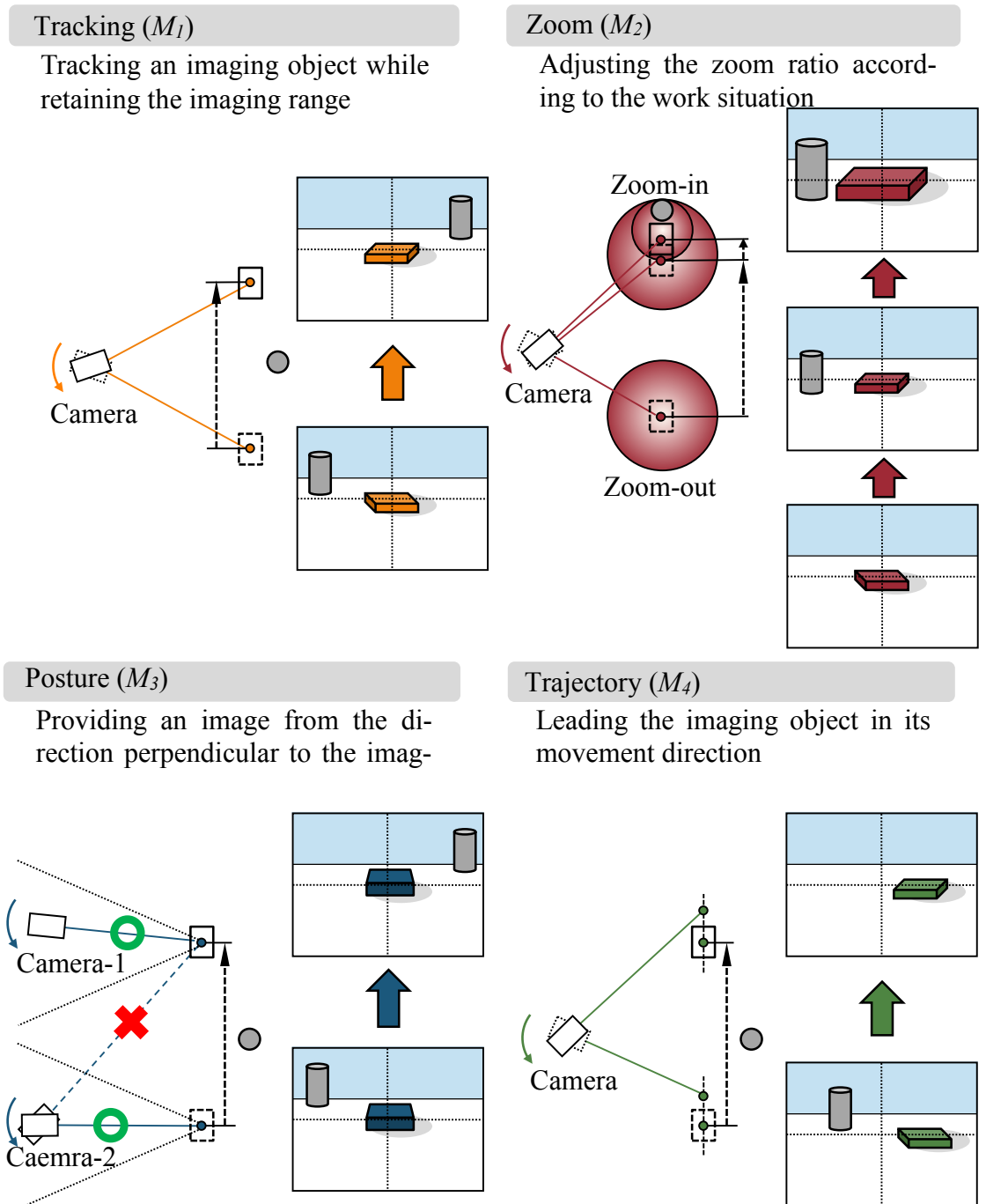
By combining arbitrary imaging objects O_x with imaging modes M_x, we define camera roles R_x, taking effectiveness, independence, and practicality into consideration

4.2.3.1 Camera roles

As mentioned in Section 4.2.1, camera roles can be arbitrarily defined for different purposes and usage conditions. The number of camera roles should be determined in relation to the number of cameras that will be installed in the actual environment. In defining a camera role, we need to consider the effectiveness, independence, and practicality of the role, and specifically the compatibility of the combination of the chosen imaging objects and modes. The end point (O_3) is compatible with all imaging modes. In contrast, the machine (O_1) is not compatible with posture mode (M_3) because the machine is used for perception of the overall environmental situation. On the basis of the above analysis and the contents of Table 3, we define four fundamental camera roles R_x that are commonly useful for arbitrary tasks and environments.

- **Overview-machine R_1 (O_1; M_1, M_2):** In this role, the camera tracks the machine with a panoramic image to provide operators with an understanding of the overall situation, showing hazards such as other machines approaching and nearby objects and terrain, as shown in Figure 4.4 (a). The angle of view is adjusted on the basis of the distance between the machine and the cameras so that the whole of the machine is in the image frame.

- **Enlarged-end-point R_2 (O_3; M_1, M_2):** In this role, the zoom ratio is adjusted depending on the identified work situation and the end point is tracked, as shown in Figure 4.4 (b). The zoom ratio is proportional to the precision necessary for the operations. This role allows operators to be more precise and efficient and avoid dangerous collision with target objects during operations using the end-effectors involving grasping, pressing, or releasing an object.

- **Posture-manipulator R_3 (O_2; M_1, M_3):** In this role, the camera direction is maintained at a right angle to the manipulator, and it tracks the manipulator. This provides operators with a sense of depth and helps them avoid unintended contact with obstacles, as shown in Figure 4.4 (c). The posture of the manipulator can be obtained from angle sensors installed in the machine together with data from a GPS. Perpendicularity is actually an overly strict condition, so the allowable angle is defined as 90°±20° in this study.

- **Trajectory-manipulator R_4 (O_2; M_1, M_3, M_4).** In this role, the camera tracks the imaging center in O_2 offset in the direction of movement of the end point and satisfying role R_3, as shown in Figure 4.4 (d). The offset is proportional to the

end point's velocity. The offset at the maximum velocity was set to 2m in this study. This role provides operators with data about the environmental information that will be entered in the near future, which helps to improve safety and efficiency.
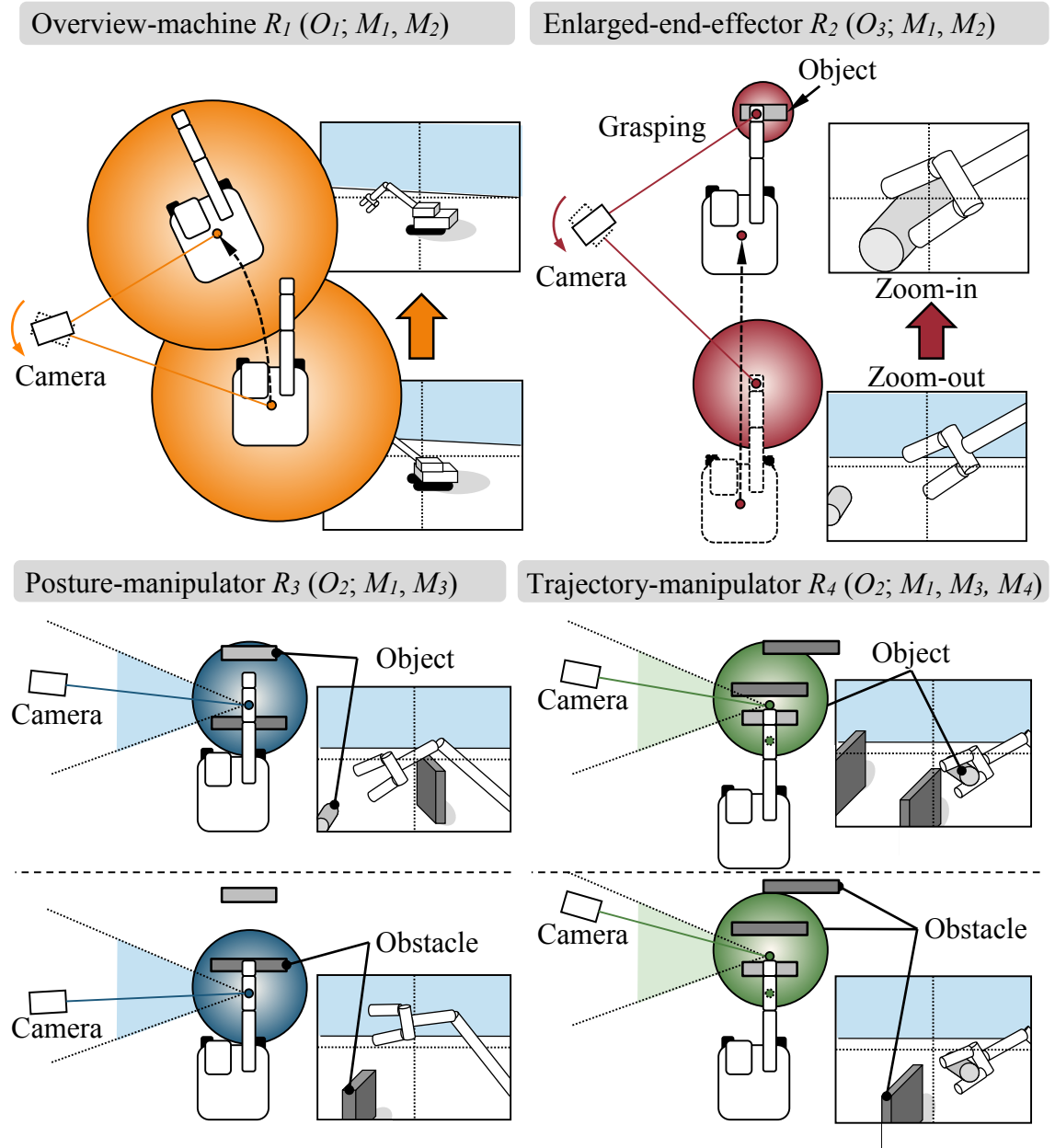
Figure 4.4 Fundamental camera roles

4.2.3.2 Classification of work situations and active camera roles

In previous work, technologies are developed for identifying the state of construction machines [45], [46]. Those technologies were incorporated there. An analysis of common work flows in disaster response work done using construction machines shows that the locations of objects to be transported and the locations where the objects are to be

collected are distinct. Therefore, the positions of the debris yard and the collection yard are adopted as feature values for debris-removal work, this simplification is used when identifying work situations, as shown in the lower part of Figure 4.5. For this paper, a work-situation identification method is defined that classifies the work situation into one of five states S_x.

- **Base movement (S_0)**: This state occurs when there is only operation for base movement during the reaching state (S_1).

- **Reaching (S_1)**: This state occurs when the distance between the end point and the debris yard is larger than 3 m and there is no contact with an object.

- **Grasping (S_2)**: This state occurs when the distance between the end point and the debris yard is smaller than 3m and there is no contact with an object. Once the grapple has grasped the object using the closing operation, the state transitions to S_3.

- **Transport (S_3)**: This state occurs when the distance between the end point and the collection yard is larger than 3m and there is contact with the object.

- **Releasing (S_4)**: This state occurs when the distance between the end point and the collection yard is smaller than 3m and there is contact with the object. Once the grapple has released the object using the opening operation, the state transitions to S_1.

The proximity for state identification (3m) was decided from the result of exploratory trials. The relations between active camera roles, image ranges (zoom ratios), and work situations are shown in the upper part of Figure 4.5. R_1 is always provided because the overview image allows dangerous situations to be noticed, and thus it should be viewable at all times. Additionally, R_2 is always provided because the state of the end point, which physically contacts objects, should be constantly monitored. R_3 and R_4 are chosen conditionally during grasping and releasing (R_3) and during reaching and transport (R_4) because R_3 is mainly for collision avoidance whereas R_4 is mainly for spatial recognition. The variable imaging range for R_2 is defined as 4m (S_0 and S_1), 2m (S_2), 5m (S_3), and 2.5m (S_4), with linear interpolation between situations (see the upper part of Figure 4.5). These parameters were determined during exploratory trials and are chosen to ensure sufficient visibility and comfortable operation. The effectiveness of the parameters used for identifying the work situation and for defining camera roles is discussed as part of the experimental results in section 4.5.
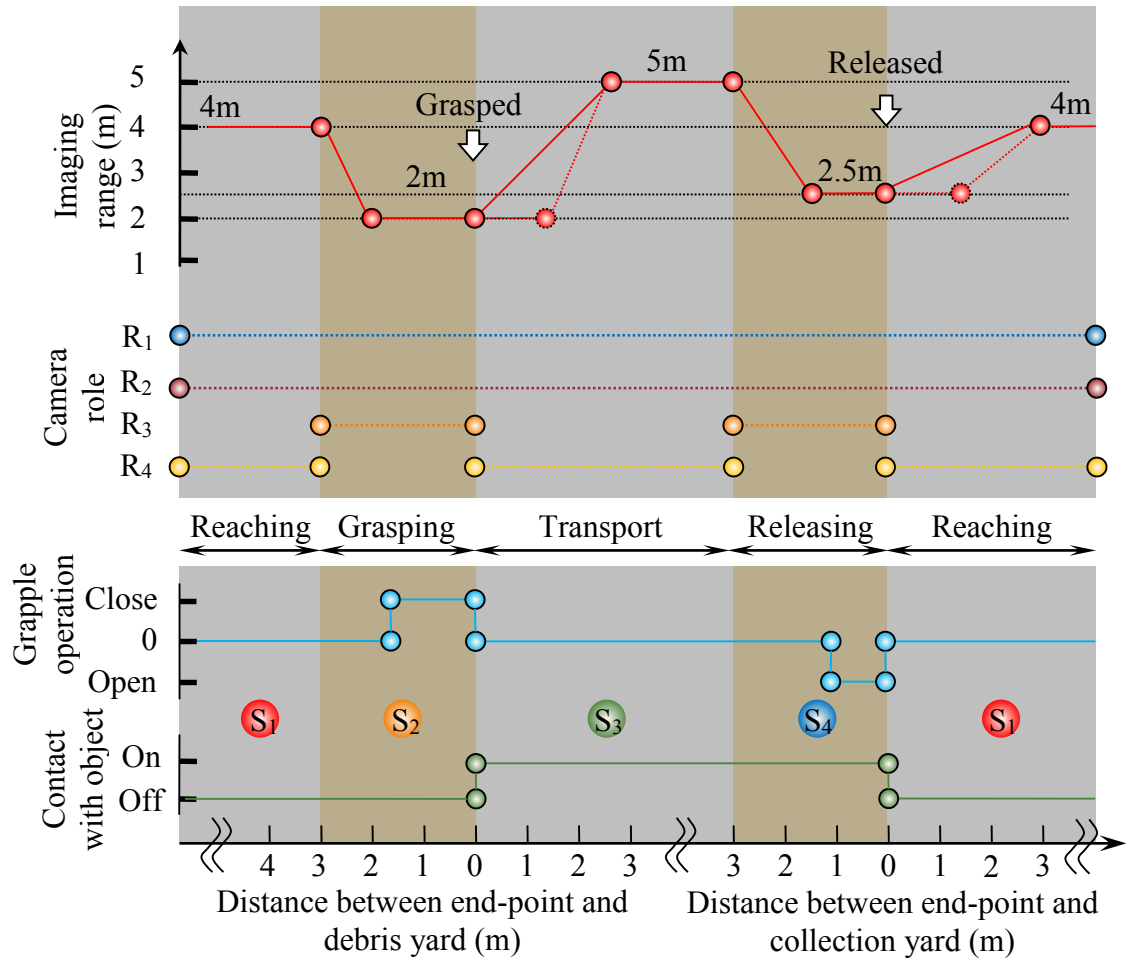
Figure 4.5 Situation identification and adaptive image range control

4.2.3.3 Implementation

Here, the technologies needed are clarified when implementing camera roles and work-situation identification on an actual system in an actual environment. In this study, it is assumed that the information necessary for implementation is obtained by using the technologies described below.

- **Three-dimensional mapping:** This fundamental technology is required to obtain the positions of target objects, obstacles, and collection places in the environment; a digital map of the environment is built from the acquired information. Currently, acquisition systems for such information have not been used in unmanned construction. In practice, the necessary information can be gathered by using GPS data and existing three-dimensional mapping technologies ([47], [48] and [49]).

- **Object recognition and segmentation:** The proposed system relies on technologies that can segment an image into different regions and categorize those regions by content, such as ground, target object, and obstacle. Such systems have not yet been implemented in unmanned construction, but existing image processing technologies ([50], [51] and [52]) can be applied for this purpose. Additionally, the ability to designate objects to be removed and the location of collection yards will be needed. We assume here that this will be realized by a touch interface that will allow designation by operators or managers.

- **Camera control:** Technologies to control the environmental cameras more stably and efficiently, and compensation mechanisms for oscillation from actuation or wind, optical flow, and optical distortion in the actual environment, are required. We plan to implement such systems on the basis of existing image-stabilization and image processing technologies [53].Moreover, we plan to develop a control method that considers the physical limitations of joints and zoom controls [44] as future work.

4.3 Role assignment system

To assign the necessary camera roles to suitable cameras, I developed a role assignment system that checks for possible roles, selects suitable cameras, and assigns the camera roles in real time.

4.3.1 Role Possibility Matrix

4.3.1.1 Check of role possibility

The system first confirms for each camera whether it can play each camera role in real time. A role is possible when there are no obstacles between the camera and the imaging object (visible imaging range, see below) and no limitations from the camera's parameters (within controllable range, see below) by using the positions among cameras, imaging objects, and obstacles, as shown in the left part of Figure 4.6.
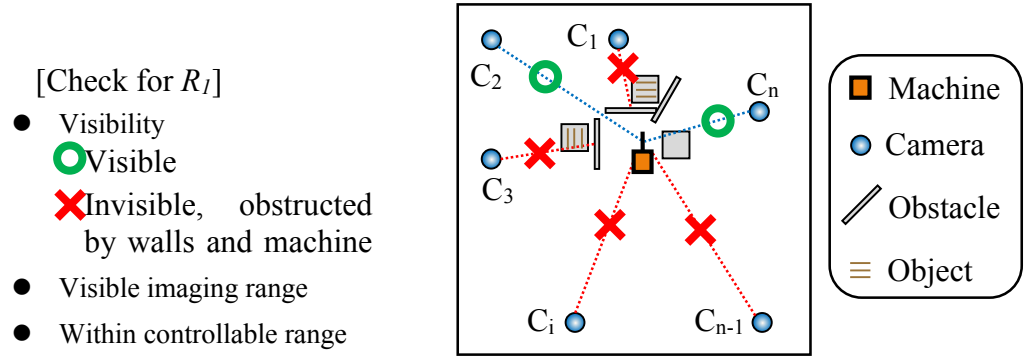
(a) Checking role possibilities

[Check for R_1]
- Visibility
  - ○ Visible
  - ✗ Invisible, obstructed by walls and machine
- Visible imaging range
- Within controllable range

Legend:
- ■ Machine
- ○ Camera
- ╱ Obstacle
- ≡ Object

(b) Role possibility matrix **A**

Higher priority →

	C_1	C_2	C_3	⋯	C_{n-1}	C_n
R_1	0	1	0	⋯	0	1
R_2	0	1	0	⋯	0	0
⋮	⋮	⋮	⋮	⋱	⋮	⋮
R_4	1	1	0	⋯	1	1

Possible: 1
Impossible: 0

Figure 4.6 Check of role possibility and role possibility matrix

- **Visible imaging range:** The system checks whether the camera can project the imaging object in a particular camera role. This clarifies whether there are visual obstructions such as debris, the camera itself, or the ground. In this study, the system judges that the state is possible when there are no obstructions to a line produced by connecting the imaging center to the center of the camera for the sake of simplicity. This means that to be considered visible, more than half the image of an imaging object must be able to be seen. In practice, this decision is made by using a visibility determination method based on a ray-tracing algorithm [54].

- **Within a controllable range:** The system simulates the images that could be obtained by changing the values of the camera's parameters. The system is assumed to know the controllable range and installed positions of the environmental cameras in advance. Then, simulation and judgment are realized by using data from angle sensors installed in the environmental cameras.

Similar to the method mentioned above that we used to identify work situations, we assume that the necessary information about the shapes and positions of target objects and obstacles is already known. The checked possibilities are then used to create an $m \times n$ matrix of the role possibilities $\mathbf{A}$ (where m is the number of roles, and n is the number of cameras), as shown in Figure 4.6 (a). The possibility matrix $\mathbf{A}$ is thus given by equation (32):

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix} \tag{32}$$

When the above judgments have been made, the possibility matrix is constructed by using 1 (0) to indicate a permitted (forbidden) role. For example, $A_{11} = 0$ means that camera C_1 cannot play role R_1, and $A_{23} = 1$ means that camera C_3 can play role R_2.

4.3.1.2 Priority of role assignment

To assign camera roles to cameras unambiguously, an assignment priority must be defined. To facilitate role assignment in using the matrix, the rows and columns are arranged in order of decreasing priority, as shown in Figure 4.6 (b). As a result, the uppermost row and the leftmost column have the highest assignment priority. The priority is determined according to the usage conditions and the purpose of the task. In general, detailed images with a relatively higher zoom ratio are important for executing tasks efficiently and effectively. In contrast, overview images with relatively lower zoom ratios have relatively lower urgency. Thus, higher zoom ratios should be preferentially selected when working. If warning images are adopted to ensure safety, they should be sent as interrupts. In practice, the relative priorities will depend on the situation, and they should be adjustable by an operator or system developer. In the future, a method to automatically prioritize roles and cameras according to the situation should be addressed.

4.3.2 Role assignment system

An assignment system independently assigns camera roles R to cameras C on the basis of the defined priority.

4.3.2.1 Basic assignment rule

Basically, the role assignment system finds the possible cameras for each role and assigns roles to the cameras. As mentioned in Section 4.3.1, the rows and columns of the

matrix are arranged in order of priority, with higher priority cameras at the left and higher priority roles at the top. We further explain the assignment rules here, and Figure 4.7 demonstrates assignment from a specific configuration with four roles and six cameras.
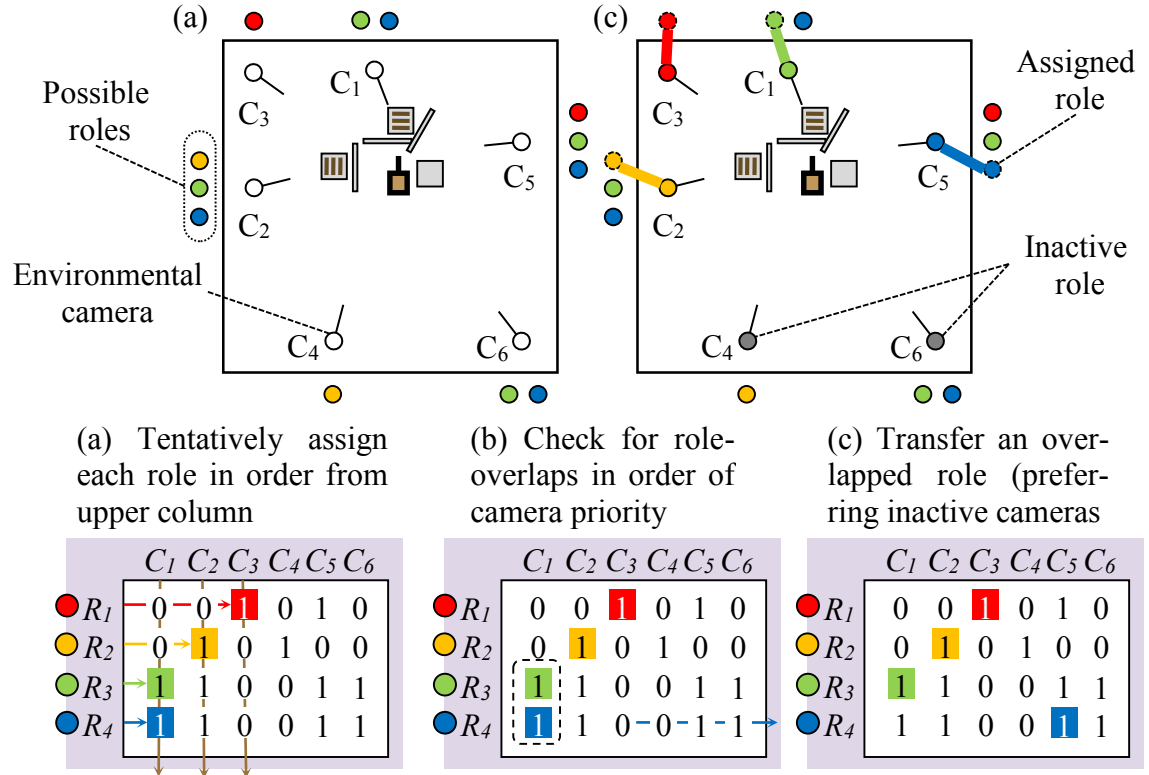

Figure 4.7 Basic camera role assignment rules

During assignment, first the possible cameras for the highest priority role are checked in order of decreasing camera priority, and that role is tentatively assigned to the highest priority camera possible [see Figure 4.7 (a)]. In the same way, the remaining camera roles are assigned to cameras, working in order of decreasing priority. Then, overlaps (in which multiple roles are assigned to one camera) are checked in order of decreasing camera priority. If there are no overlaps, then role assignment is completed for the work situation. When there are overlaps, as shown in Figure 4.7 (b), the roles are reassigned to other cameras on the basis of role priority. First, the role with lower priority is checked to see if it can be transferred to a camera with no role. If no unassigned cameras are found, then the cameras with lower-priority roles are searched. This order of operations reduces the frequency of switching images. In the same way, the tentative assignment of the remaining roles is checked in order of decreasing role and camera priorities. Through this process, all m camera roles (R_1 - R_m) are assigned to m out of the n

environmental cameras (C_1 - C_n) [see Figure 4.7 (c)]. If there is a role that no cameras can play, the camera previously assigned to that role is selected. When the system can find more suitable cameras than it needs, the camera with higher priority is used. In future work, the coverage ratio of the imaging range and the seamlessness of the image before and after switching should be considered.

4.3.2.2 Advanced reassignment rule

Through the basic assignment process described above, the relations between roles and cameras are continuously updated at the specified sampling frequency. As work progresses, the spatial relations among cameras, the machine, and obstacles change. If a camera cannot play its assigned role, then that role must be reassigned to another camera while considering the overall balance. This is done in the following way. First, the possible cameras are searched to find cameras with no role. If no unused cameras can play the role, then the system searches among possible cameras with roles. If a candidate camera is found and that camera's assigned role is a lower priority role than the role to be assigned, the camera is reassigned and the displaced role is assigned to another camera using the same process [see Figure 4.8 (a) and Figure 4.8 (b)]. If no cameras can play the role, then the cameras with higher priority roles than that role are searched. Reassignment proceeds iteratively through roles; a second reassignment triggered by an earlier reassignment is shown in Figure 4.8 (c).

(a) Check possible cameras (preferring inactive)

(b) Check role-overlaps in the camera priority order

(c) Transfer an overlapped role to another camera

	C_1	C_2	C_3	C_4	C_5	C_6
R_1	0	0	0	0	1	0
R_2	0	1	0	1	0	0
R_3	1	1	0	0	1	1
R_4	1	1	0	0	1	1

	C_1	C_2	C_3	C_4	C_5	C_6
R_1	0	0	0	0	1	0
R_2	0	1	0	1	0	0
R_3	1	1	0	0	1	1
R_4	1	1	0	0	1	1

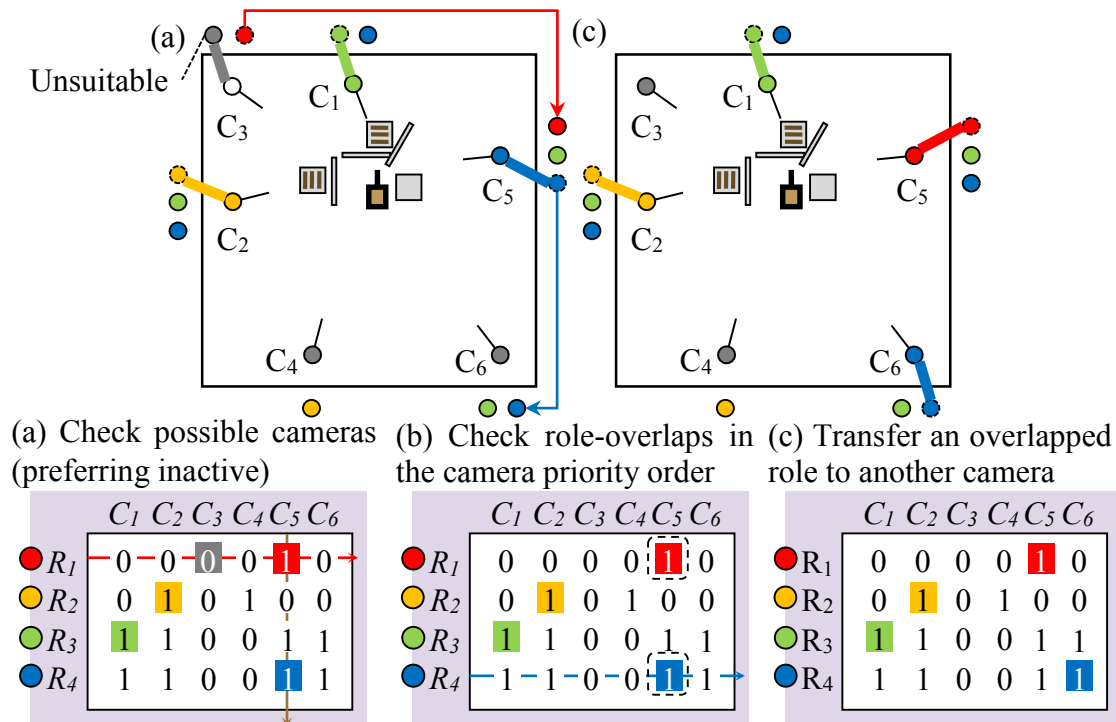|	C_1	C_2	C_3	C_4	C_5	C_6
R_1	0	0	0	0	1	0
R_2	0	1	0	1	0	0
R_3	1	1	0	0	1	1
R_4	1	1	0	0	1	1

Figure 4.8 Advanced camera role assignment rules

The above role assignment system is easily implemented using linear programming [55], and it can be applied to arbitrary configurations. This simple but effective system is important for practical applications such as unmanned construction. The objective function to be used is given by equation (33):

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\left(A_{ij}\cdot X_{ij}\right) = m \tag{33}$$

Here, **A** is the possible matrix from equation (32) and **X** is the assignment matrix (0: unassigned; 1: assigned), given by equation (34):

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{bmatrix} \tag{34}$$

Moreover, each role should be assigned to exactly one camera, and each camera should have no more than one role, so the constraint conditions in equation (35) are enforced,

$$\begin{cases} A_{11}X_{11} + A_{12}X_{12} + \cdots A_{1n}X_{1n} = 1 \\ A_{21}X_{21} + A_{22}X_{22} + \cdots A_{2n}X_{2n} = 1 \\ \qquad\qquad\vdots \\ A_{m1}X_{m1} + A_{m2}X_{m2} + \cdots A_{mn}X_{mn} = 1 \\ A_{11}X_{11} + A_{21}X_{21} + \cdots A_{m1}X_{m1} \le 1 \\ A_{12}X_{12} + A_{22}X_{22} + \cdots A_{m2}X_{m2} \le 1 \\ \qquad\qquad\vdots \\ A_{1n}X_{1n} + A_{2n}X_{2n} + \cdots A_{mn}X_{mn} \le 1 \end{cases} \tag{35}$$

An arbitrary number of camera roles m can be assigned in real time to an equal number of suitable cameras out of the number of all installed cameras n (m ≤ n) by solving the described system with linear programming in real time.

In case that no camera can play some role, an $m \times m$ eye matrix is located to the right of **A**. The expansion matrix **A**ₑₓ from **A** in equation (36) and the expansion matrix **X**ₑₓ from **X** is in equation (37). Thus the boundary condition (35) can be written in (38)

$$\mathbf{A}_{ex} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} & 1 & 0 & \cdots & 0 \\ A_{21} & A_{22} & \cdots & A_{2n} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{36}$$

$$\mathbf{X}_{ex} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1(m+n)} \\ A_{21} & A_{22} & \cdots & A_{2(m+n)} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{m(m+n)} \end{bmatrix} \tag{37}$$

$$\begin{cases} A_{11}X_{11} + A_{12}X_{12} + \cdots A_{1(m+n)}X_{1(m+n)} = 1 \\ A_{21}X_{21} + A_{22}X_{22} + \cdots A_{2(m+n)}X_{2(m+n)} = 1 \\ \qquad\qquad\qquad\vdots \\ A_{m1}X_{m1} + A_{m2}X_{m2} + \cdots A_{m(m+n)}X_{m(m+n)} = 1 \\ A_{11}X_{11} + A_{21}X_{21} + \cdots A_{m1}X_{m1} \leq 1 \\ A_{12}X_{12} + A_{22}X_{22} + \cdots A_{m2}X_{m2} \leq 1 \\ \qquad\qquad\qquad\vdots \\ A_{1(m+n)}X_{1(m+n)} + A_{2(m+n)}X_{2(m+n)} + \cdots A_{m(m+n)}X_{m(m+n)} \leq 1 \end{cases} \qquad (38)$$

4.4 Experiment

To conduct experiments on the effectiveness of the proposed system, VR simulator mentioned in Section 3.2 was used and experimental conditions and procedure were determined.

4.4.1 Parameter settings for role assignment system

Here, I describe how to set the value for each parameter used in the role assignment system as described below.

- **Camera role and priority setting:** It is shown in Figure 4.5 that roles R_3 and R_4 (as described in this paper) will not be active at the same time and there is a seamless transition between them. Because they are disjoint, I facilitate role assignment by unifying these as one role, denoted by R_{3-4}. Moreover, it is useful to have multiple cameras filling the overview-machine role (R_1), because different cameras have different viewpoints, which enhances visibility for the operator. To effect this without violating the constraints described above, I create additional roles with properties identical to R_1, and denote these by R_{1A} and R_{1B}. Consequently, there are four roles (R_{1A}, R_{1B}, R_2, and R_{3-4}). As described in Section 4.3.1.2, the conditions for R_2 and R_{3-4} are stricter than those for R_1, and the grasping is a highly difficult operation, so R_2 should have the highest priority and R_1 should have lower priority. Thus, the rows of the possibility matrix are arranged in the order R_2, R_{3-4}, R_{1A}, and R_{1B} from top to bottom.

- **Monitor and camera:** In determining the number of monitors to be used, both the complexity of the camera monitor system and human limitations should be considered. So that no monitors are left unused, the number of monitors should be no more than the number of camera roles. In this study, the number of camera roles equals to the number of monitors for environmental cameras. As a practical matter, there should be more environmental cameras than monitors because this

provides the system with more options for selecting cameras suited to the situation. In this study, I assume six cameras (C_1-C_6), and so two cameras will be unused at each point in time. The priority for cameras was arbitrarily set to the number of cameras for the sake of simplicity.
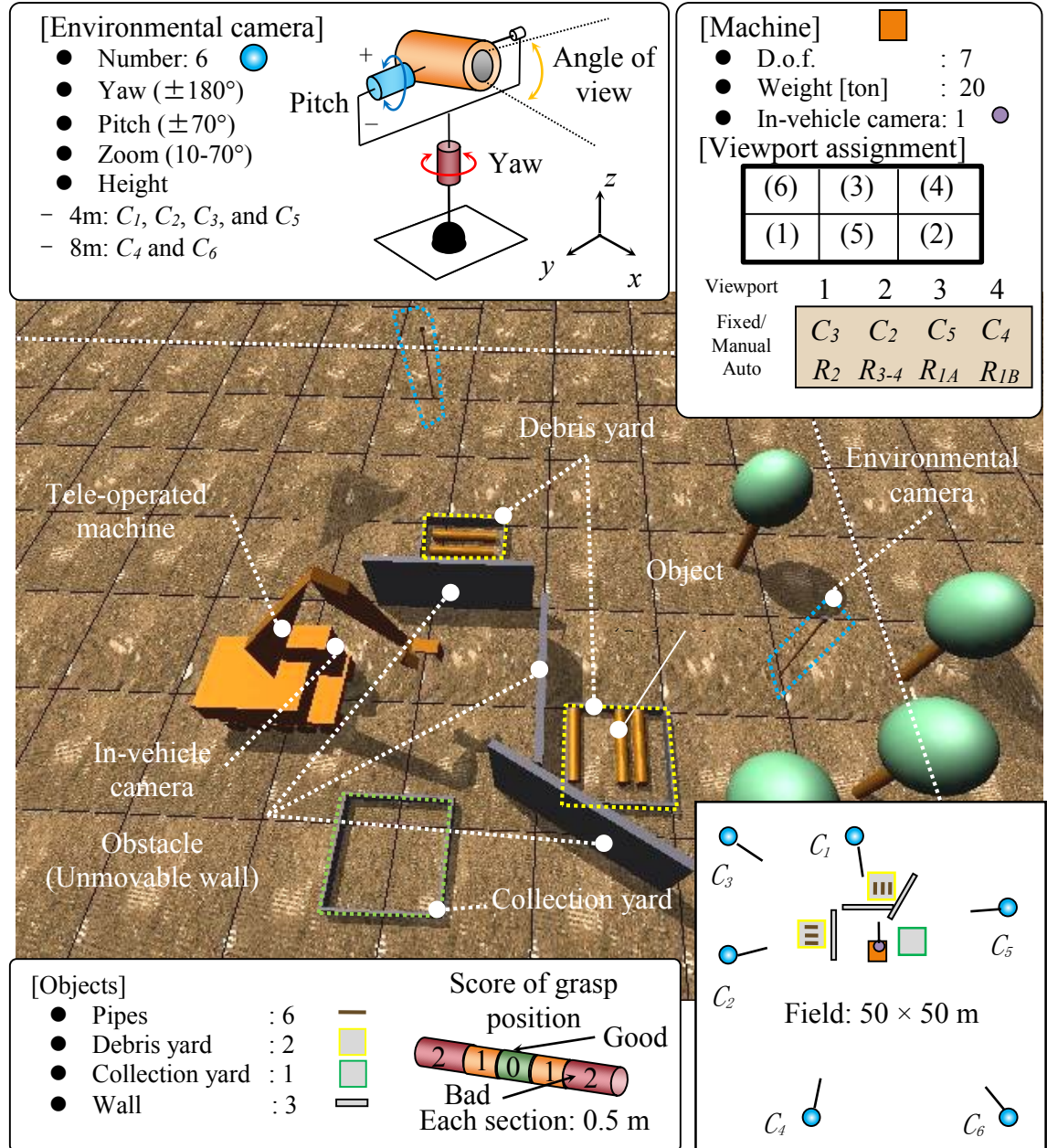

Figure 4.9 Experimental environment

4.4.2 Experimental conditions

The evaluation task was set as debris transport, which is one of the most fundamental disaster response tasks because transporting things is crucial, and blind spots occur frequently during this task. The proposed autonomous system is compared with a fixed

camera system and a manual control system. The sequences of the tasks in each system are shown in Figure 4.10 (autonomous), Figure 4.11 (manual control), and Figure 4.12 (fixed camera).
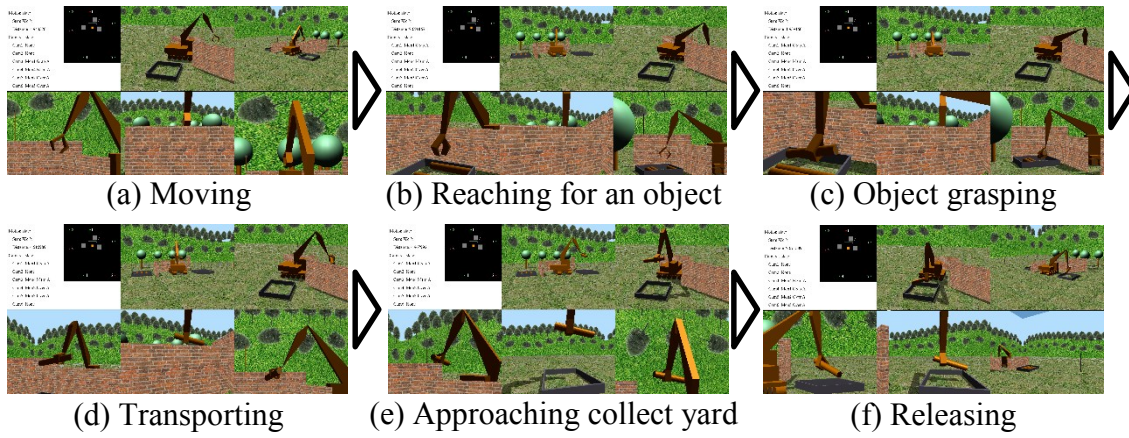

(a) Moving (b) Reaching for an object (c) Object grasping

(d) Transporting (e) Approaching collect yard (f) Releasing

Figure 4.10 Sequence of transport task in autonomously-controlled camera system

(a) Moving (b) Reaching for an object (c) Object grasping

(d) Transporting (e) Approaching collect yard (f) Releasing
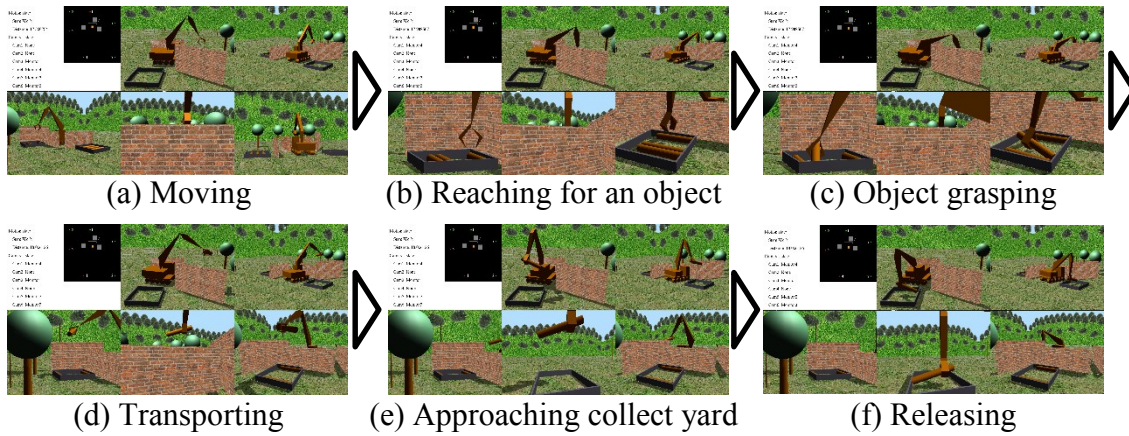
Figure 4.11 Sequence of transport task in manually-controlled camera system

(a) Moving (b) Reaching for an object (c) Object grasping

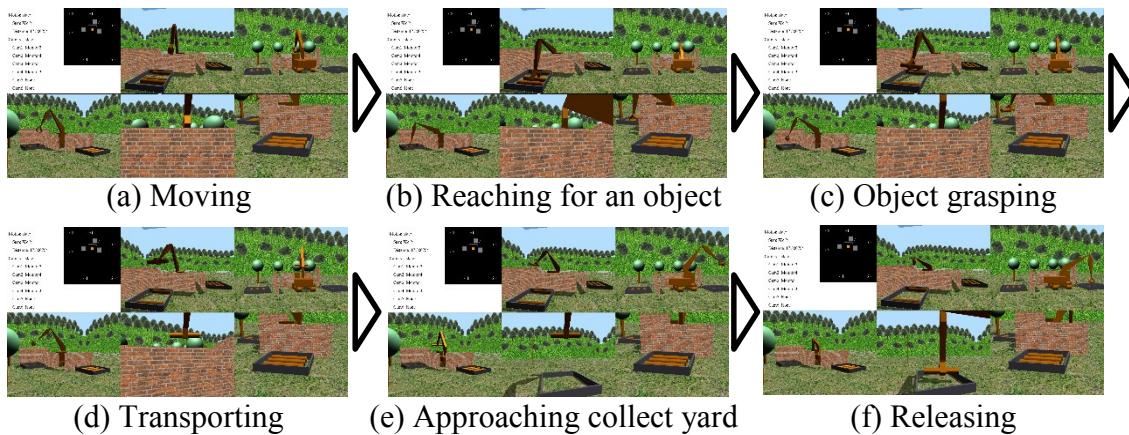(d) Transporting (e) Approaching collect yard (f) Releasing

Figure 4.12 Sequence of debris transport task in fixed camera system

- **Details of the settings:** Six cylindrical objects to be transported were placed in two debris yards. To obscure the operator's vision, two walls and one high wall were placed in front of the debris yards (Figure 4.9). The adjustable ranges of the angle of view (zoom ratio), yaw (z-axis rotation), and pitch (y-axis rotation) were set with reference to typical camera settings (Figure 4.9). In addition, the roll rotation (x-axis rotation), base location (x, y), and installation height (z) were fixed for all the systems (see the left upper part of Figure 4.9). The camera installed at 4 m elevation was used mainly for providing detailed images of the manipulator, while the camera installed at 8melevation was used mainly for providing an overview of the environment. The camera parameters for the fixed and manual (initial state) systems are described in Figure 4.9. Values for these parameters were set according to the results of preliminary trials to best enhance the in-vehicle camera image. For the autonomous system, roles R_2, R_{3-4}, R_{1A}, and R_{1B} were assigned to the viewports 1, 2, 3, and 4, respectively (Figure 4.9).

- **Experimental procedure:** Ten novice operators who were familiar with operating a machine in our VR simulator each performed the task six times with each system. In the task, the operators grasp the middle part of each object without a grasping miss or contact error, then transport the object, as fast as possible without dropping it, to the collection yard. The completion time, number of grasping misses, dropped objects, and overload contacts, as well as the grasped positions of each object, were measured. After the trials, the operators filled out questionnaires about subjective usability. I chose a within-subject design rather than a between-subject design because operator skill levels differ greatly between operators [56]. To compensate for the order effect of selfimprovement due to habituation to operating in the environment, which is a major drawback for a within-subject design, the order of testing was set randomly for each operator (e.g., auto → manual → fixed → manual → ⋯), and no two operators had the same order. To validate the system for actual application, professional operators should be enlisted as participants, and experimental parameters, such as the arrangement of the objects and environmental cameras, should be changed between trials. However, one problem with the current system for unmanned construction is a shortage of skilled operators. Therefore, if novice operators can complete the work, then the system may solve this shortage as well. One of the purposes of this study is to

provide an easy-to-use interface for unmanned construction. Therefore, the experiments have significant value in that they may serve to improve unmanned construction significantly.

4.5 Results

To evaluate the effectiveness of the developed autonomous control system, we conducted debris transport experiments using the VR simulator described in section 4.4.

4.5.1 Confirmation of the developed autonomous control scheme

We first confirmed from experimental data with the autonomous system and for a representative operator whether the camera parameters were adequately adjusted and the camera roles were assigned according to the identified work situation. The distance between the end point and the debris yard, the distance between the end point and the collection yard, the operation input for the grapple, and the on-off state of contact with objects are shown in Figure 4.13 (a). The yaw, pitch, and zoom angles for each camera are shown in Figure 4.13 (b). The variation over time of the relations between camera roles and assigned cameras is shown in Figure 4.13 (c) as a time series.

4.5.1.1 Work-situation identification

As Figure 4.13 (a) shows, each work state was adequately identified and in the correct order, from base movement (S_0) to release (S_4), according to the distances between the end point and the debris (or collection) yard, the operations for the grapple, and the on-off state of object contact. We also found that the moment when the grasping state transitioned to the transporting state and when the releasing state transitioned to the reaching state corresponded to observed phenomena.

4.5.1.2 Camera parameter control

As Figure 4.13 (b) shows, the yaw, pitch, and zoom were dynamically changed as the work progressed. The zoom ratio was adjusted on the basis of the distance between the end point and both yards as well as the assigned camera role, as shown in Figure 4.13 (b-3). For camera C_2, to which role R_2 was assigned, the zoom intensified as the end point approached an object in the grasping state (from 16 to 32s). The pitch angle was dynamically adjusted during states S_1 and S_3, as shown in Figure 4.13 (b-2). This is because the end point moved largely in the vertical direction when picking up an object, lifting it up over the walls, and setting it down in the collection yard. The yaw angle was

also controlled according to the situation. The amount of the adjusted quantity was small because of the spatial relations among the obstacles, cameras, and machine, as shown in Figure 4.13 (b-1).

This is because the conducted tasks did not require large amounts of base movement. From this analysis, we confirmed that camera parameters were controlled according to the work situation.
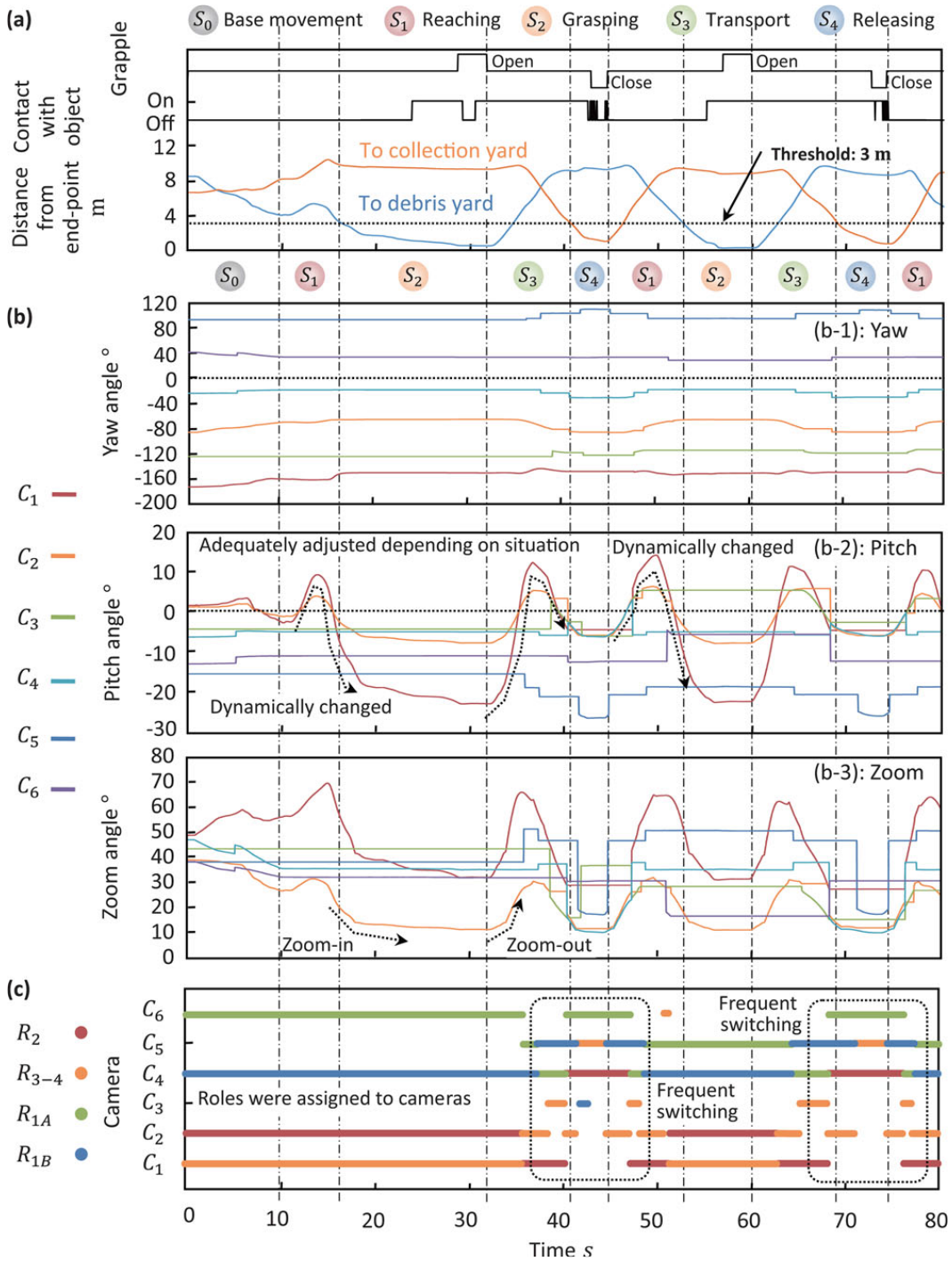

Figure 4.13 Relations among (a) work situation identification using sensor data, (b) each camera parameter, and (c) assigned camera role

4.5.1.3 Camera role assignment

As Figure 4.13 (c) shows, the four camera roles were assigned to four of the six environmental cameras, and the relations between camera role and assigned camera were dynamically changed as the work progressed. The results show that reassignment (i.e., switching) of camera roles happened frequently, particularly during the releasing state and just before and after it. Frequent switching of the camera image may negatively affect operator concentration, and so switching should be reduced to avoid increasing the cognitive load on the operators. Consequently, the results of analysis also indicate that the installation positions of environmental cameras must be carefully considered to enable output to be stable with respect to camera images. This task is left to future work.

From the results of the analysis, we confirmed that the implemented role assignment system assigned roles to cameras according to the identified work situation and the spatial relations between the end point and each yard. The effectiveness of the system is evaluated in the next section.

4.5.2 Effectiveness evaluation of the autonomous camera control system

The averages and standard deviations for task times, number of grasping misses, dropped objects, and overload contacts per trials, and the scores for grasping position and subjective usability, are listed in Table 4 for the three systems. The completion times in the beginning, medium, and final set are shown in Figure 4.14. For use in analyzing the characteristics of operations, Figure 4.15Figure 4.17 show the lever operations and positions of the end point for the fixed camera (Figure 4.15), manual control system (Figure 4.16), and autonomous control system (Figure 4.17).
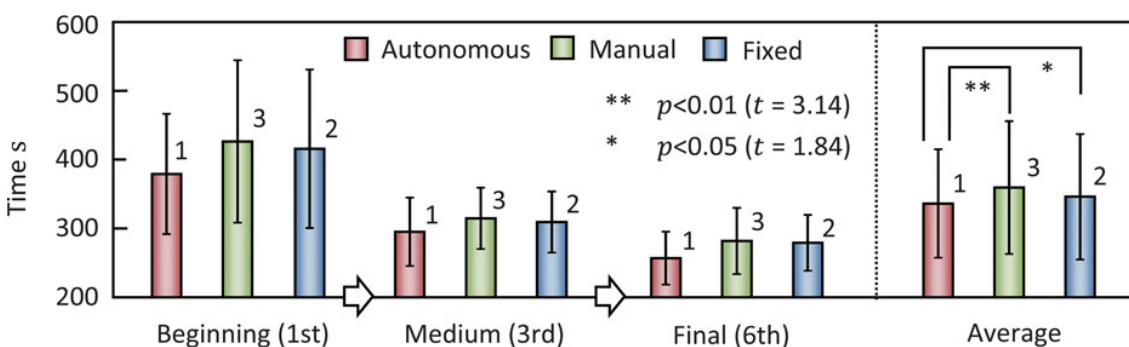

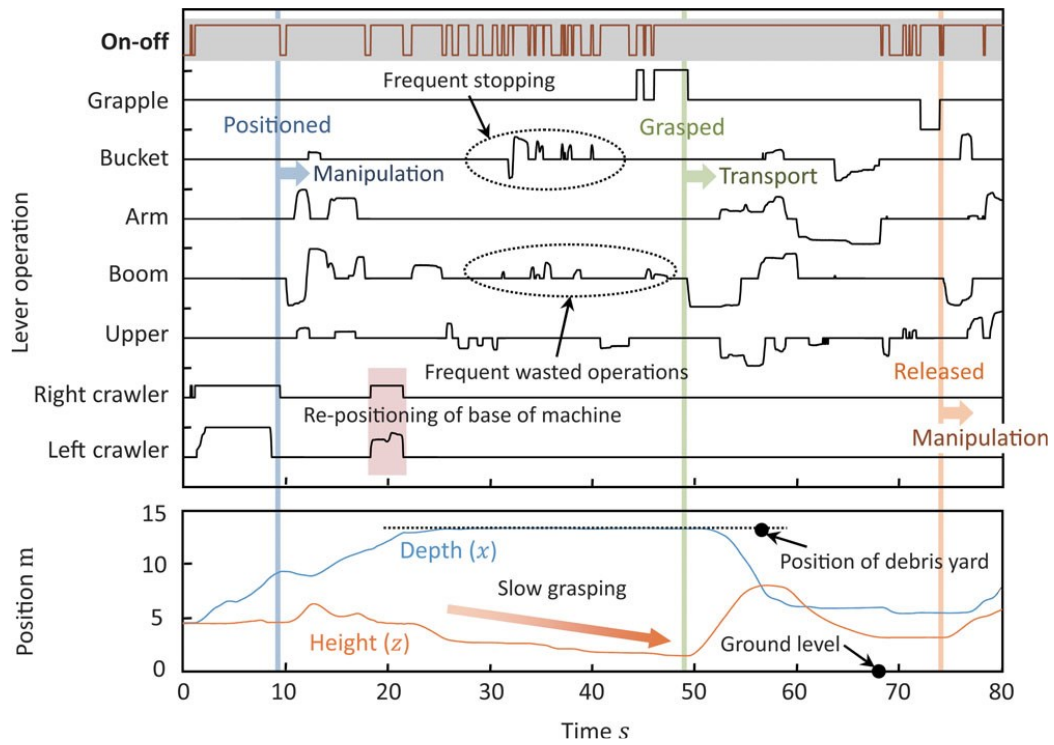**Figure 4.14 Transition and average of completion time for three control systems**

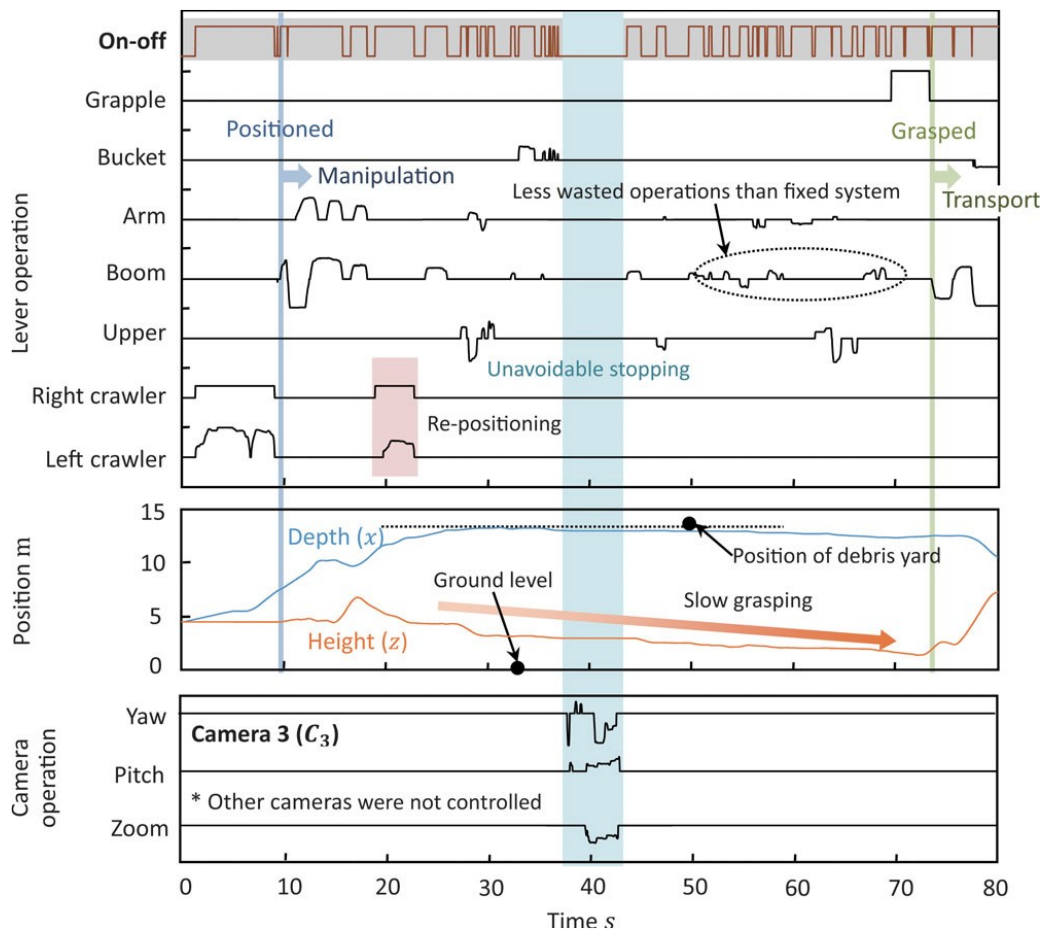**Figure 4.15 Relations between lever input and machine movement in fixed system**

Figure 4.16 Relations between lever input and machine movement in manually-controlled system
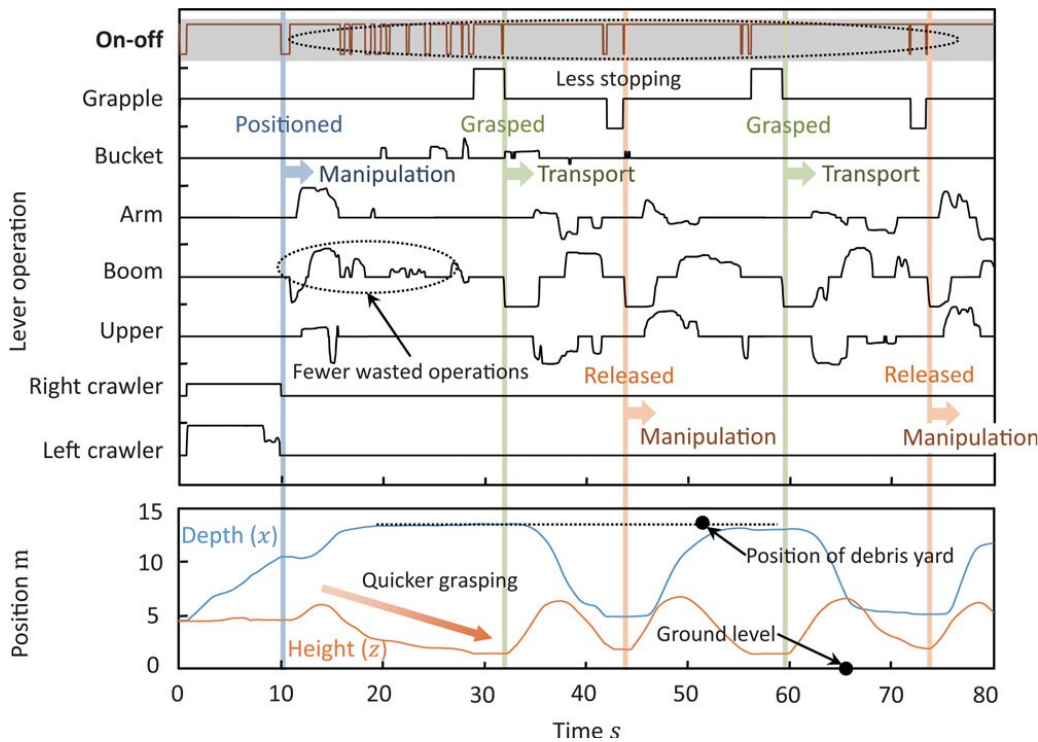
Figure 4.17 Relations between lever input and machine movement in autonomous camera control system

4.5.2.1 Time efficiency

The average task times for the three systems are listed in Table 4 (a). The task time was shortest for the autonomous system and longest for the manual system. At-test indicates that the difference is significant between the autonomous and manual systems (t = 3.14, p<0.01) and between the autonomous and fixed camera systems (t = 1.84, p < 0.05). As Figure 4.14 shows, time decreased in each trial, but the time relations among the three systems did not change between trials. This is likely because even when the operators had habitual ways of operating in a given environment, the improvement plateaued because of insufficient visual information. Moreover, the figure shows that this system would be effective regardless of the degree of proficiency. In the manual system, which necessarily required spending time to control the cameras, the machine operations often stopped as the cameras were controlled. As Figure 4.16 shows, lever operations for the machine were completely stopped while the lever for the cameras (only C_3 in this case) was controlled. In contrast, in the autonomous system, operators could continue to control the machine while the system controlled the cameras. The task time for the autonomous system was shorter than that for the fixed system, meaning that it could provide effective images for smooth control of the machine and reduce wasteful stopping and operations. As shown in Figure 4.15, wasteful stopping occurred often during base

movement and grasping in the fixed system, and a long time was required for the grasping phase. In contrast, in the autonomous system (see Figure 4.17), there were fewer stops, and the grapple could be used to grasp an object smoothly and efficiently.

Table 4 Experimental results of different camera control systems

Average for four operators (standard deviation)	Autonomous	Manual	Fixed
(a) Time efficiency			
Completion time s	336.5 (78.8)	359.6 (96.4)	346.2 (90.9)
(b) Safety (frequency per set)			
Grasping miss (end-point contact)	0.55 (0.81)	0.7 (0.91)	0.60 (1.08)
Dropped object	0.03 (0.17)	0.12 (0.41)	0.09 (0.38)
Overload contact (upper contact)	0 (0)	0.12 (0.41)	0.18 (0.52)
(c) Quality (good: 1↔bad: 0)			
Score of grasping position	0.79 (0.2)	0.79 (0.19)	0.70 (0.27)
(d) Subjective usability (good: 1↔bad: 0)			
Difficulty of overall task	0.87 (0.17)	0.77 (0.27)	0.4 (0.21)
Frequency of blind spots	0.60 (0.21)	0.63 (0.25)	0.33 (0.31)
Difficulty of depth sensing	0.67 (0.22)	0.7 (0.54)	0.4 (0.26)

4.5.2.2 Safety and quality

The average numbers of grasping misses, dropped objects, and overload contacts per trial are listed in Table 4 (b). All such negative items were less frequent in the autonomous system than in other systems. The dropping of objects in the fixed system was typically caused by the grapple grasping the end of the object, while in the manual system it was caused by the pivot joint rotating at an excessive speed, although the object was initially stable. This latter problem occurs because the operator cannot ensure the stability of the grasped object by referring to the fixed images since the images cannot be controlled simultaneously with the machine. As shown in Figure 4.11 (the left lower image display), the grasping of the object is effectively shown, but the view is not useful during transport and release because the image was not adjusted by any operator after the grapple grasped an object. In contrast, the autonomous system can track an imag-

ing object according to the work situation, as shown in Figure 4.10, so the operators were continually able to refer to the images to carefully check spots that were unique to the situation. To evaluate the quality of object grasping, the grasping position was measured. The relation between the position and the score is shown in the right part of Figure 4.9 (d). The results are listed in Table 4 (c). In the table, a score of 1 is the best (grasped at the center of an object) and a score of 0 is the worst (grasped at the end of an object). In the manual and autonomous systems, the scores are better than those in the fixed system. This result confirms that adaptive imaging provides the desired images to enable operators to perform more precise operations.

4.5.2.3 Subjective Usability

Questionnaires were administered to the operators at the end of the experiments. The operators were asked to rate each system on three aspects: the difficulty of the overall tasks; the frequency of blind spots, and the difficulty of depth sensing. The results are listed in Table 4 (d). In the table, a score of 1 is the best and a score of 0 is the worst. The averages (standard deviations) of all the items are 0.71 (0.22) (autonomous), 0.70 (0.25) (manual), and 0.37 (0.26) (fixed). A t-test indicated a significant difference between the autonomous and fixed systems on all items ($p < 0.01$). The manual system was rated as better than the fixed system, meaning that the use of the controllable cameras is sufficient to improve subjective usability. In addition, when the cameras were controlled autonomously, both the difficulty of the task and the frequency of the blind spots were reduced more effectively. From free responses, we found that the preferred timing for changing between states and zoom ratios varied with the operators; moreover, the frequent changes in camera images with the autonomous system, discussed in section 4.5.1.3, negatively impacted the concentration of the operators. These results indicate that a machine-learning scheme to adapt the control parameters to the preference of individual machine operators should be developed, in addition to an alignment method for optimizing the installation positions of environmental cameras.

In the manual system, operators can obtain desired images because the cameras are manually controlled by themselves. This kind of system has both positive and negative effects: the positive effect is improved safety and quality of work; the negative effect is decreased time efficiency of the work. In the autonomous control system, the time efficiency, safety, and quality of the work could all improve if the adaptive imaging system can provide images suitable to the work situation. The analysis results indicate that the autonomous system adequately improved the time efficiency, safety, work quality, and

subjective usability for all operators by providing adaptive images relevant to the situation. These results confirm that the proposed adaptive imaging system could provide images to enhance the operator's effective field of vision.

Junjie YANG - February 2016

5 AUGMENTED REALITY (AR) VISION SUPPORT SYSTEM

In this study, I focus on using augmented reality (AR) vision support elements [37]-[40] in our autonomous camera control system to conduct advanced operation by un-manned construction machinery. AR items are dedicated to induce visual attention to images with adequate information according to work situation, and which can increase the user experience like the work pressure [41].

5.1 Improved operational interface

In this section, I clarify modules to be improved in our VR simulation system and the autonomous camera control system and modify them.

5.1.1 Requirements of improved operation interface

Former autonomous camera control system supplies two enlarged end-effector views when it is near object areas and two overlook views all the time, as shown in Figure 3.3 (i)–(iv). However, I find operators prefer to observe enlarged views in most of the time, because it makes them easy to conduct the detail operation near the end-effector. It easily leads to the ignorance of the relationship between manipulator and obstacles, which may cause unnecessary collision. In addition, according to their questionnaire, they thought that it was hard to estimate the vertical drop point of end-effector and the distance between manipulator and objects because of the weak sense of distance.

5.1.2 Improved control system

I use two cameras (one from the front and another from the side on the same height level of end-effector) to shoot end-effector (Figure 5.1(a) and (d)). I also use two cameras (one from the front and another one from the side of basis of machine at high places) to shoot overlook views (Figure 5.1(c) and (f)). In order to understand the role of each view, we use frames with different colors. Green frame stands for detail view; blue frame stands for overlook view and yellow frame stands for cab view. Considering the operator tends to fix their view sight on detail view and ignore the potential danger, overview view (Figure 5.1(f)) will switch with a potential danger view (Figure 5.2) when manipulator or the upper part of machine is near an obstacle [68].

(a) Detail view   (b) 2D overlook view   (c) Overlook view

(d) Detail view   (e) Cab view   (f) Overlook view /
Danger view
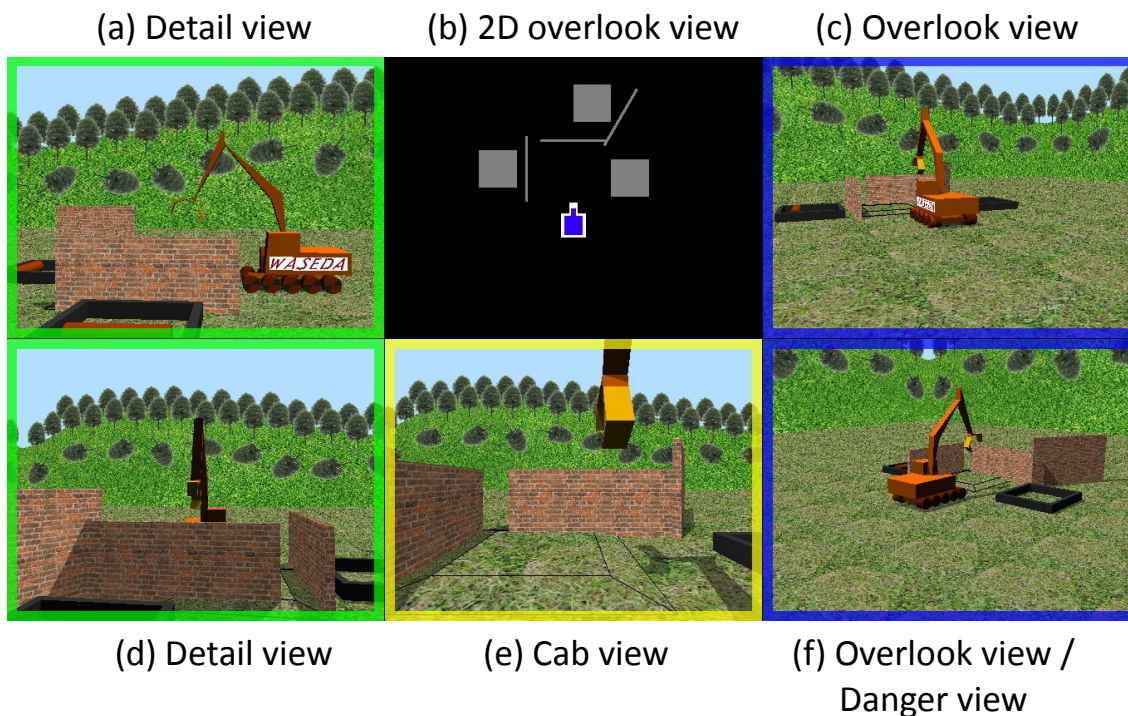
Figure 5.1 Improved operational interface

(a) Area between manipulator and objects   (b) Area between main body and objects
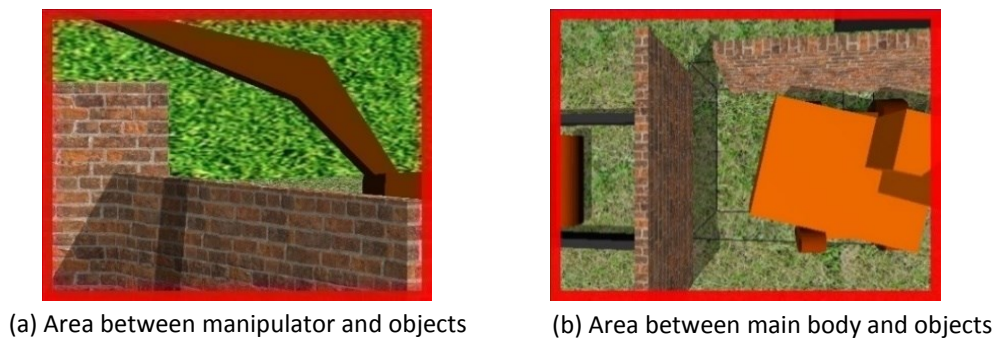
Figure 5.2 Introduced danger views

5.2 AR-based attention inducement system

We develop a visual attention inducement system based on AR prompts for arbitrary work content and environment. Here, I will describe the design of AR items used in our system, such as vertical arrow, guide laser, reachable sphere, distance arrow, distance label and rotation hint [42], as shown in Figure 5.3 (a)–(f). They are all used to supply a better distance sense and help operators to make clear the relationship.
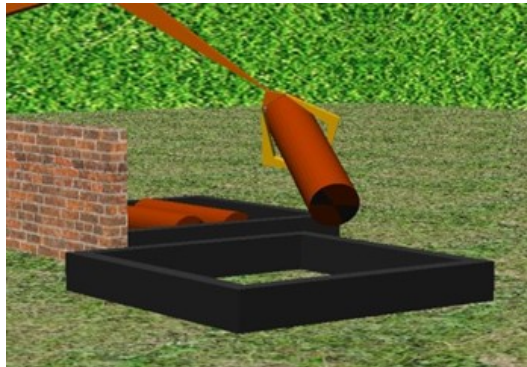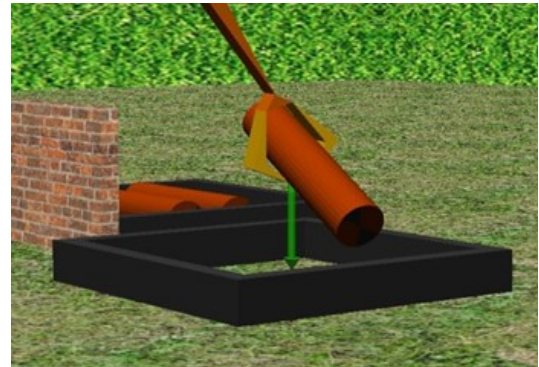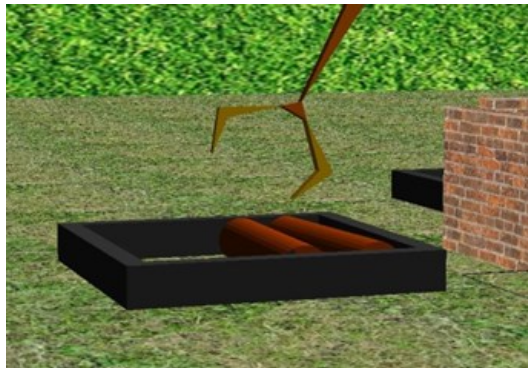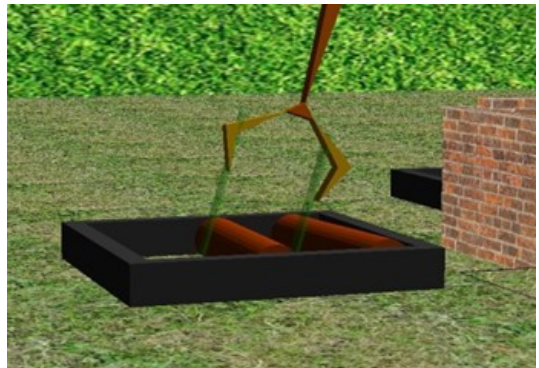

(i) Detail view without support   (ii) Detail view with support
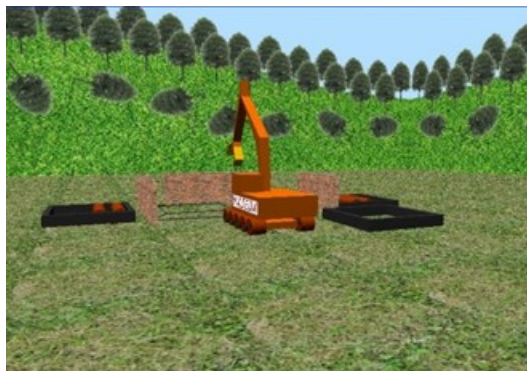
(a) vertical arrow
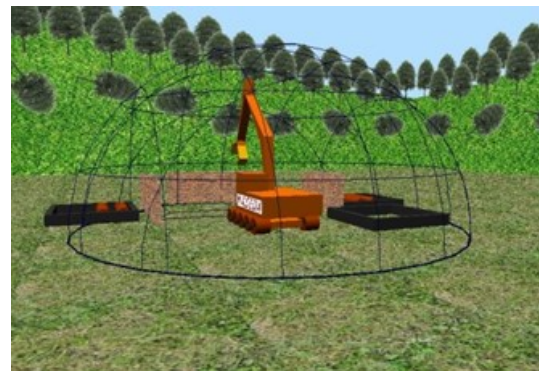

(i) Detail view without support   (ii) Detail view with support
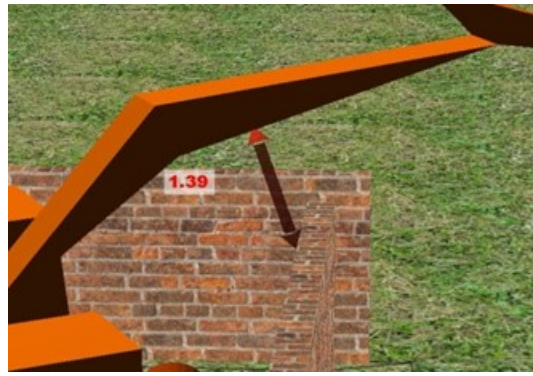
(b) Guide lasers

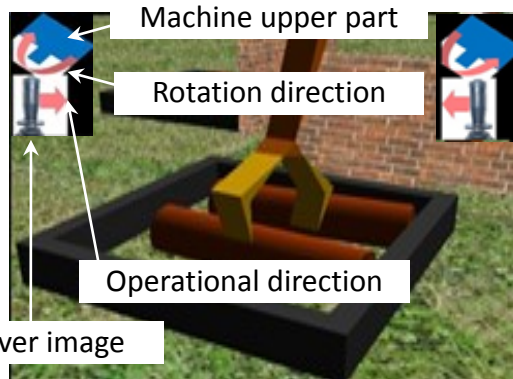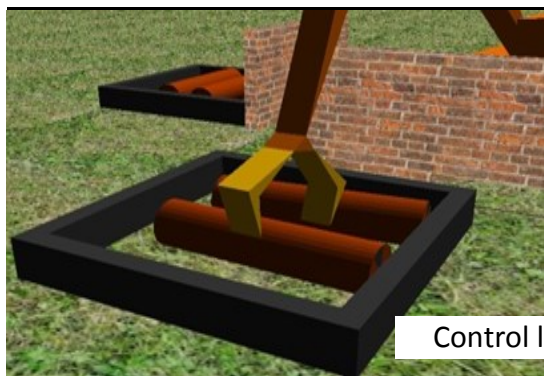(i) Overlook view without support   (ii) Overlook view with support

(c) Reachable sphere

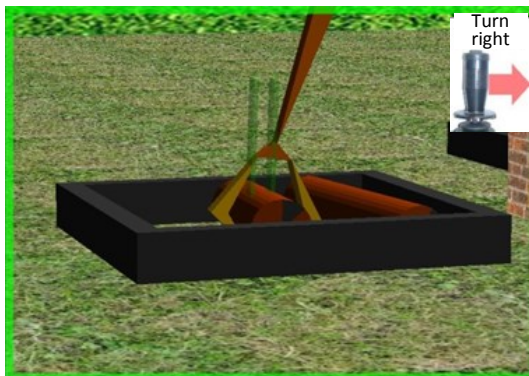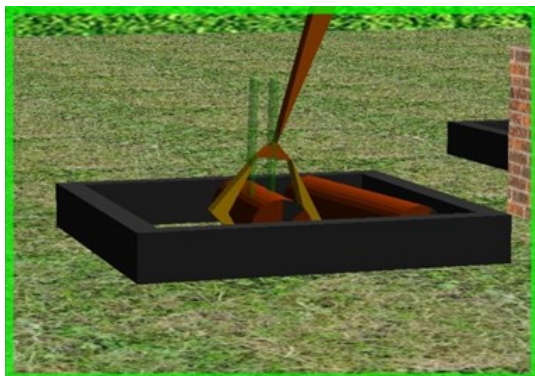(i) Danger view without support | (ii) Danger view with support

(d) distance arrow and distance label

Machine upper part

Rotation direction

Operational direction

Control lever image

(i) Detail view without support | (ii) Detail view with support

(e) Rotation hint

Turn right

(i) Detail view without support | (ii) Detail view with support

(f) Using control lever image with operation direction

Figure 5.3 Difference between without AR support and with AR support

5.2.1 Vertical arrow and guide laser

When operators try to release the object near recycle box, they need to know the drop point of the end-effector. Therefore, the arrow vertical to the ground appears in the left two detail views and middle cab view when the end-effector is near the center of recycle box (5 m). It is green and translucent, as shown in Figure 5.3(a).

5.2.2 Guide laser

The relationship among the two pieces of grapple and obstacle is very important when operators try to grasp a target. In this situation, we use two thin guide lasers in detail views to show the extension lines of two pieces of grapple when end-effector is near a debris box (3.6 m), as shown in Figure 5.3 (b). When the end-effector is near the bottom of box, the guide laser will extend along the opposite direction of end-effector, in order to be observed clearer. By these two lasers, operator can easily know the relationship between end-effector and targets even in a single view.

5.2.3 Reachable sphere

It appears in the two overlook views to show the surface of reachable area of manipulator. In our case, its radius is 9 meters. Weft paralleling to the ground and that paralleling to the machine orientation are used to describe the sphere. They are rendered in black, as shown in Figure 5.3 (c).

5.2.4 Distance arrow and distance label

It appears in the danger view when manipulator is near a wall which is treated as an obstacle (1.5 m). It is translucent and turns from orange to red with the shrinking of itself, as shown in Figure 5.3 (d). Additionally, the shortest arrow appears at the center of view with the label showing distance.

5.2.5 Rotation hint

Operators tend to conduct miss operation if camera shoots machine or end-effector from its front because the tilt direction of lever and the rotation direction of machine upper part are different. In order to inform operators how to rotate as wanted when such situation occurs, we show 2D overlook of machine upper part with rotation orientation and control lever image with the corresponding operation direction as shown in Figure 5.3 (e), when end-effector is near a debris box (1.5 m). In order to inform operator the rotation direction when end-effector is near the debris box or grasps a debris, the system calculate the nearer rotation direction to transport the debris according to the angle between machine upper orientation and the vector from the base to the recycle box (Figure 5.3 (f)).

The entire AR vision support system diagram is shown in Figure 5.4. To identify base movement, grasp, reaching, transport, and releasing as work situation, we apply a work situation identification method proposed in Section 4.4.
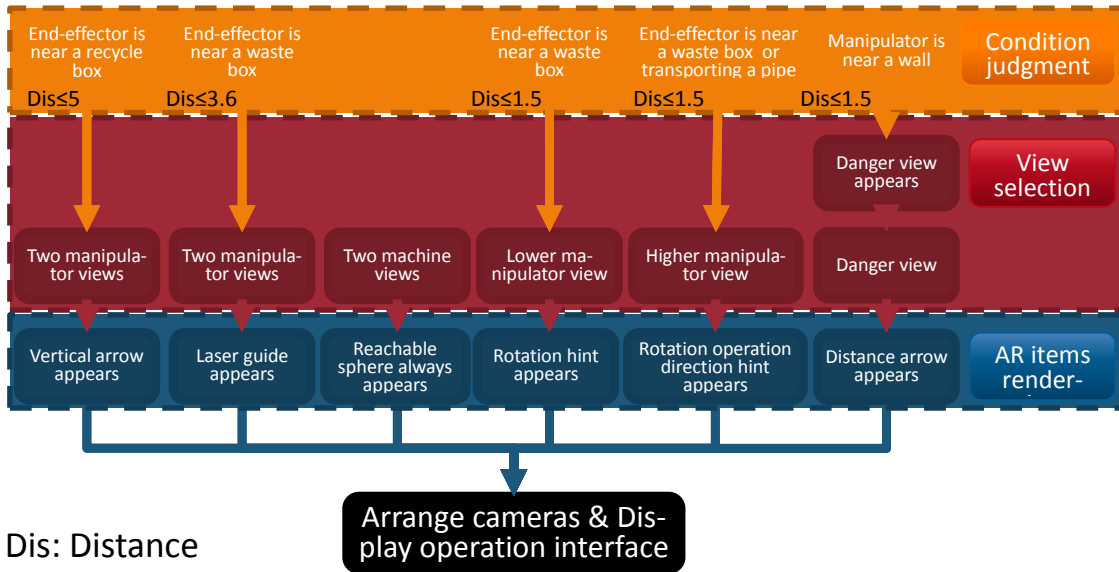


Figure 5.4 AR vision support system diagram

5.3 Experiments

We conducted a transport experiment by using the VR simulator. To derive how the AR items affect in remote control conducted in VR environment and the relationship between user experience and each AR items, comparison experiment is conducted.

### 5.3.1 A. Experimental settings

5.3.1.1 Work site configuration

In our experiment, we have four environments used for training and another three for comparison experiments, as shown in Figure 5.5. Different environments are used to reduce the influence of familiarity. In training environments, we use such four different environment to simulate the different scene after disaster. In each scene, two debris cylinders are placed in one box or separately placed in two boxes which are placed on the ground or on a mound. In the environments used for comparison experiment, three scenes with higher difficulty are used (Figure 5.5). Six debris cylinders are placed in different debris boxes which are placed on the ground or on a mound.
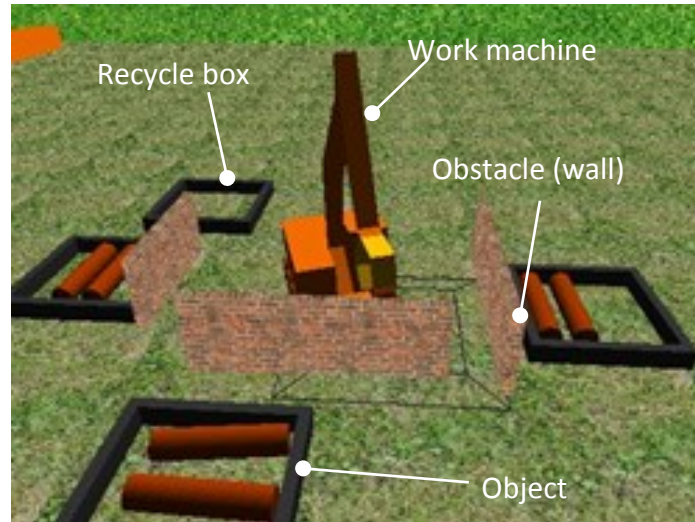
(i) Environment 1
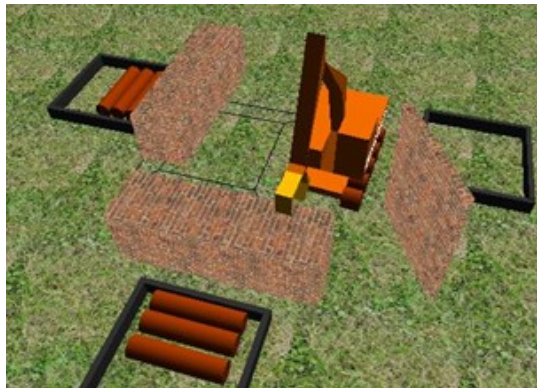
(ii) Environment 2
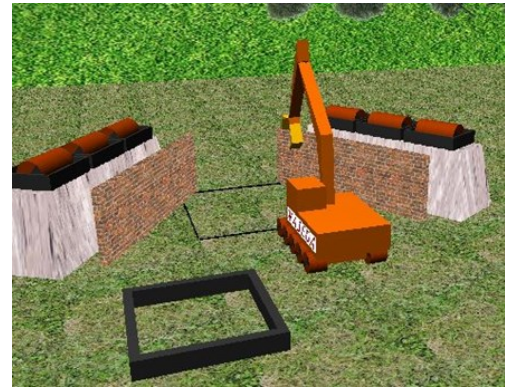
(iii) Environment 3

Figure 5.5 Work site environment used for comparison experiment

5.3.1.2 Camera configuration

In order to make operators have a good sense of the work site, we place six environmental cameras as the current system. Four are set on the height of about 4 meters to have a clear view of debris boxes and recycle box. Another two cameras are set on a high position (8 meters), which enable them take aerial views as well. All camera control are controlled autonomously.

5.3.1.3 Experimental task

In training environment, operators were asked to control the machine to transport all two debris cylindrical sticks one by one from debris box to a recycle box. In the comparison experiment, six cylindrical sticks are asked to transport. In both tasks, operators should avoid unexpected contact with walls or boxes and grasp the center of each stick to keep it balanced. And then, they are expected to conduct as fast as they can.

5.3.2 Experimental conditions

Operational skills are trained and measured in the four training environments by using a simple learning curve. And, we then conducted comparison experiments in the other three environments to analyze whether AR vision prompt supply system is effective and make clear the relationship between each AR item and user experience. Eight operators attended our experiment.

This kind of experiments largely affects the order effect. In this study, we thus adopt between-subjects experiments, not within-subjects design. Operators are separated averagely into two groups according to their task time performance in training. As shown in Figure 5.5, we use three different environments to conduct comparison experiment. In the first day, operators in both group are asked to conduct transport task in each environment 4 times without AR prompt support. In the second day, both groups do the same task in each environment 4 times as the first day. However, one group (Group A) has no support and the other group (Group B) is supported by AR prompt. We recorded the total task time, task time in each work condition, number of collision in the experiment and the questionnaire after a set of tasks.

5.4 Results

We analyze the work time and mental workload for each group in each day to measure the total effect of AR prompts support. Both of them are evaluated by average. We also compared the moving time, grasping time and transporting time, releasing time and frequency of collision to evaluate the effect of each AR item.

5.4.1 Accomplishment time

We compared the average task time in all environments for each group and derived the following graph (Fig. 6 (a)). In the first day, all conditions and tasks of two groups are the same. The average task time of group B is worse than that of group A, which should be nearer as expected. How-ever, in the second day, the average task time of group B is better than that of group A. Additionally, the difference of average task time of group B in two days is extreme significant. So, we can easily know that the AR prompt can raise the time efficiency.
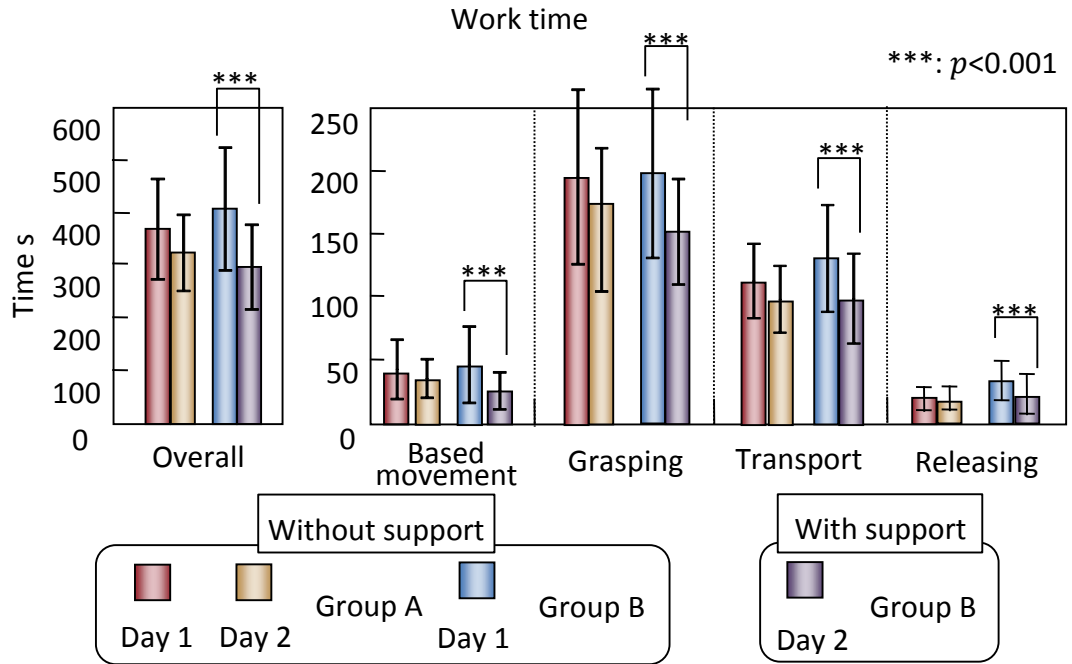
Figure 5.6 Work time in different situations

5.4.1.1 In base movement

Figure 5.6 shows the average time used in base movement condition in all tasks of each group in each day. In the first day, time used of group B is a little more than that of group A. However, in the second day, time used of group B decreased obviously while that of group A decreased only a little. The difference of moving time of group B in 2 days is extreme significant, so that we think the reachable sphere can help operator to find a position where they can conduct the following tasks well.

5.4.1.2 In grasping

In this period, operators conduct the most subtle manipulation to grasp debris cylinder. Any careless manipulation may cause grasping failure and increase the time cost. According to Figure 5.6, average time cost of two groups in the first day are almost the same. However, in the second day, the average grasping time of group B decreases more obviously than that of group A. The difference of average grasping time of group B in two days are extreme significant. So, we can easily know that the guide laser can reduce the grasping time by informing operators the relationship between two pieces of end-effector and other objects.

5.4.1.3 In transport

After grasping debris, operator should transport it to recycle box. During this period, operators tend to rotate in the opposite direction as expected at the beginning. They also

tend to rotate the machine upper part in a worse direction, which may cause the total rotation time much more than expected. Through the rotation hints, it seems operators are easily understand the relationship be-tween control lever and detail view of manipulator, which is indicated in Figure 5.6. Group B performed worse than group A under the same condition in the first day. However, they did almost the same as group A did in the second day. The difference of the transporting time of group B in two days are extreme significant, so that we can derive that the rotation hint can reduce the time cost in transporting period.

5.4.1.4 In releasing

According to Figure 5.6, group B cost much more time when releasing debris than group A in the first day. In the second day, group B performs improves much more than group A. The time cost of two groups in releasing period are almost the same in the second day. Time difference of group B in two days is extreme significant. It means that vertical arrow can increase the time efficiency in re-leasing by judging whether the arrow is in the recycle box instead of imagining relationship between debris and recycle box.

5.4.2 Mental workload

Here, we used NASA-TLX method in our questionnaire to measure the mental workload [43] [73] [74]. Mental demand (MD), physical demand (PD), temporal demand (TD), own performance (OP), frustration (FR) and effort (EF) are scored from 0 to 100 by operators. The higher the heavier. They are also asked to compare the importance between each 2 items from the above 6 to measure the weight of each item, which are scored from 0 to 5 and the total weight is 15. So the WWL score is given by $[\sum(S \cdot W)]/15$. Here, S stands for score of each item and W stands for weight of each item. Such WWL score is used to measure each operator's mental work directly. The higher the score is, the heavier the mental workload is.

According to the calculation, the average mental workload of each group in each day is shown in Figure 5.7. It is easy to find, there is almost no difference between two groups in the first day. However, in the second day, the average mental work of group B decrease obviously while that of group A keeps the same level as the first day. So, we think the AR vision prompts can reduce the operators' mental workload. It also means that the AR prompts can supply a better user experience.
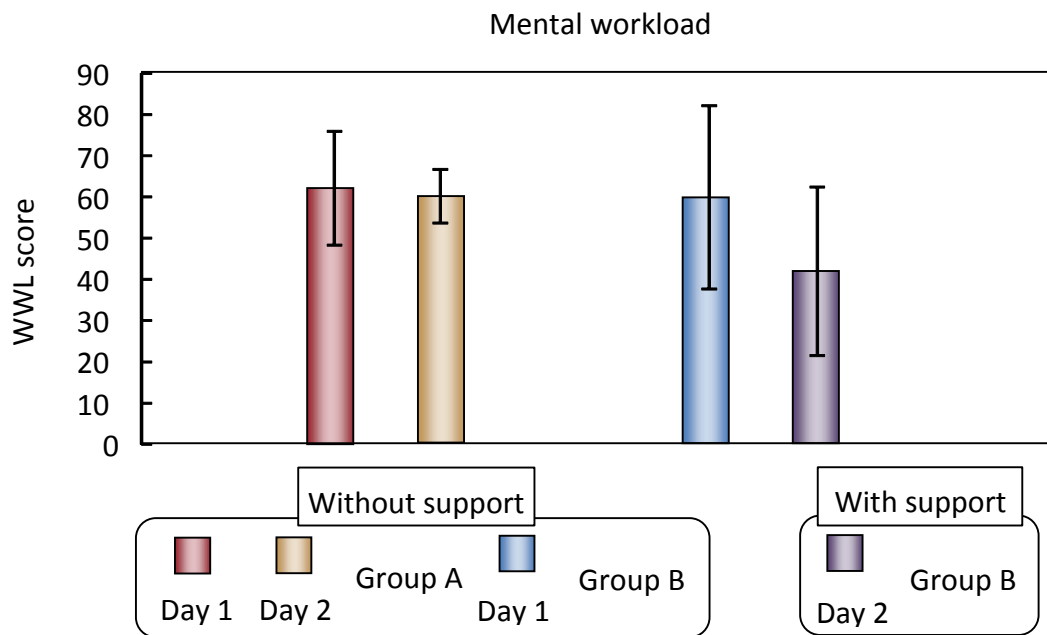
Figure 5.7 Average mental workload of each group

5.4.3 Frequency of collision

Figure 5.8 show the average number of collision times in all tasks of each group in each day. In the first day, the collision frequency of group B is twice than that of group A. However, in the second day, collision frequency of group B decreases obviously while that of group A keeps at a stable level as the first day. That means the distance arrow in our AR vision prompts can reduce the possibility of collision to some extent.
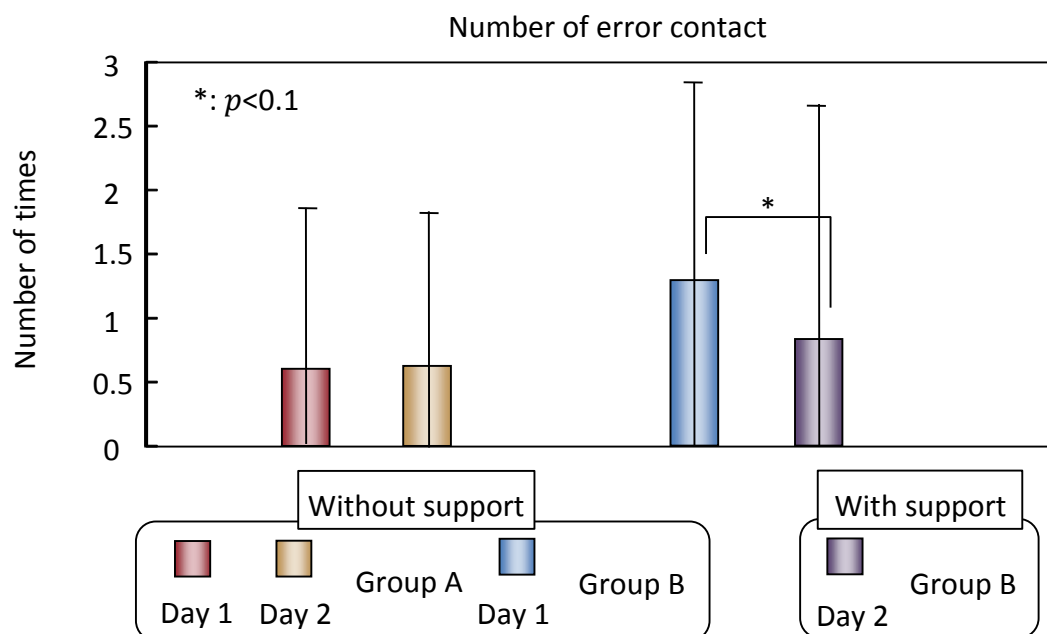

Figure 5.8 Average error contact per task of each group

5.5 Discussion

In the other part in our questionnaires, we survey the sense of distance when AR vision prompt support was used. And, we also find out the most useful items and the worst items in the questionnaire. They are used to make clear what kind of AR item is useful and easy to understand. Figure 5.9(a) and (b) shows the AR item which operators preferred and watched the most frequent, respectively. Sense of distance of each group in two days are given in Figure 5.9 (c). Moreover, Figure 5.9 (d) shows the best and worst understandability in AR prompts.

5.5.1 Questionnaires

We can easily find from Figure 5.9 (c) that AR prompt system can supply a better sense of distance, which can increase the user experience. According to questionnaire of the operators who used AR prompt system, as shown in Figure 5.9 (d), we find that most of them prefer the guide laser and vertical arrow, and some of them prefer the reachable sphere. In addition, most of them think rotation hint is hard to understand and some of them think distance arrow is hard to understand. So, the increased sense of distance almost caused by the above preferred items. After our analysis, we find, each preferred items have the following properties.

5.5.1.1 Vertical arrow

Because the direction of arrow and that of gravity are the same, 3D position of end-effector can be easily converted into a 2D position. After that, the relation-ship among the two borders of recycle box and drop point is easy to understand.

5.5.1.2 Guide laser

It materialize the extended line of two pieces of end-effector and have the point of inter-section on ground plane. So the relationship among two pieces of end-effector and debris can be converted to the relationship among three points on ground plane. It is understandable.

5.5.1.3 Reachable sphere

The reachable sphere have the intersection circle on the ground, and most debris boxes and recycle box are placed on the ground. So, the relationship between boxes and reachable sphere can be easily judge on the ground plane.
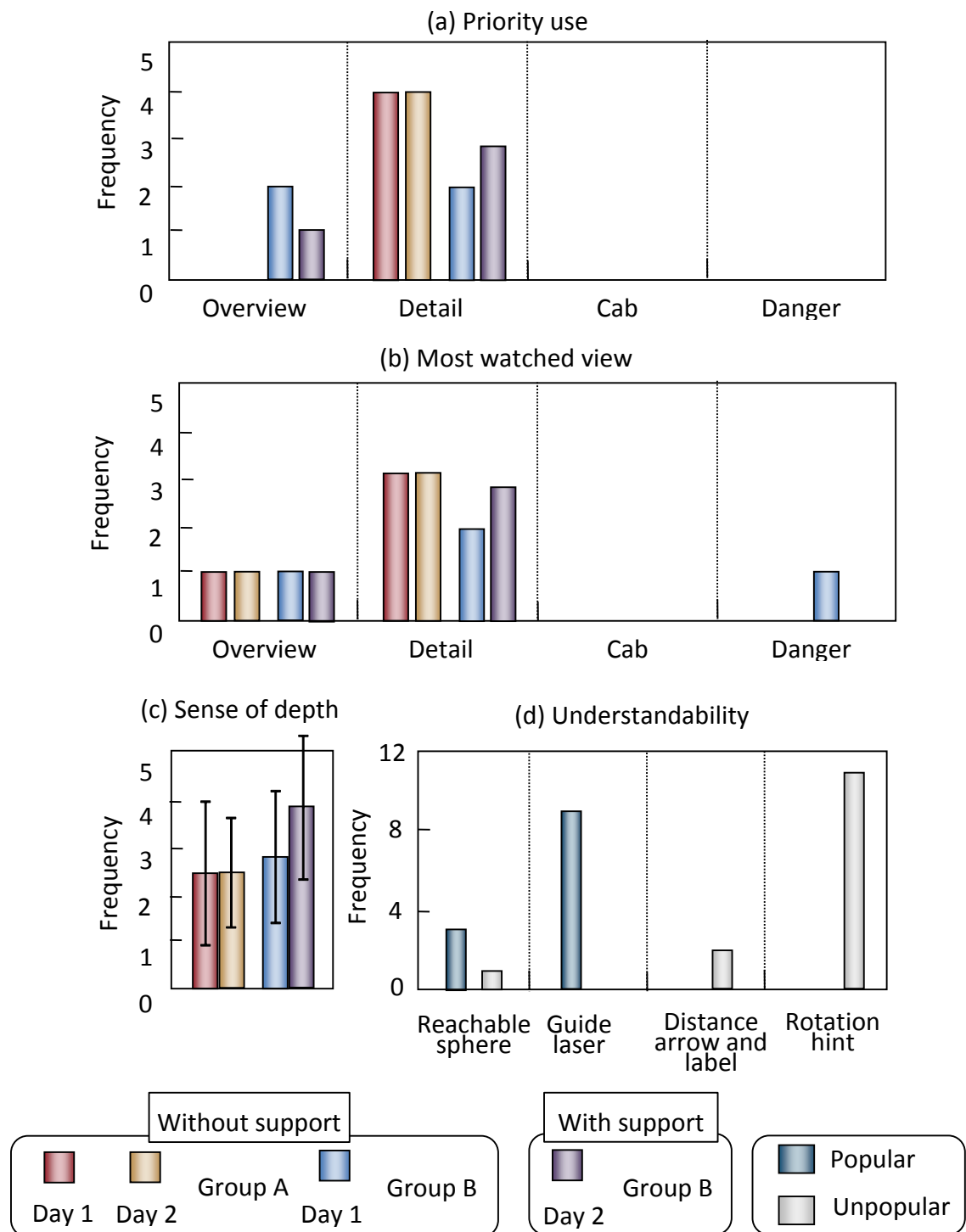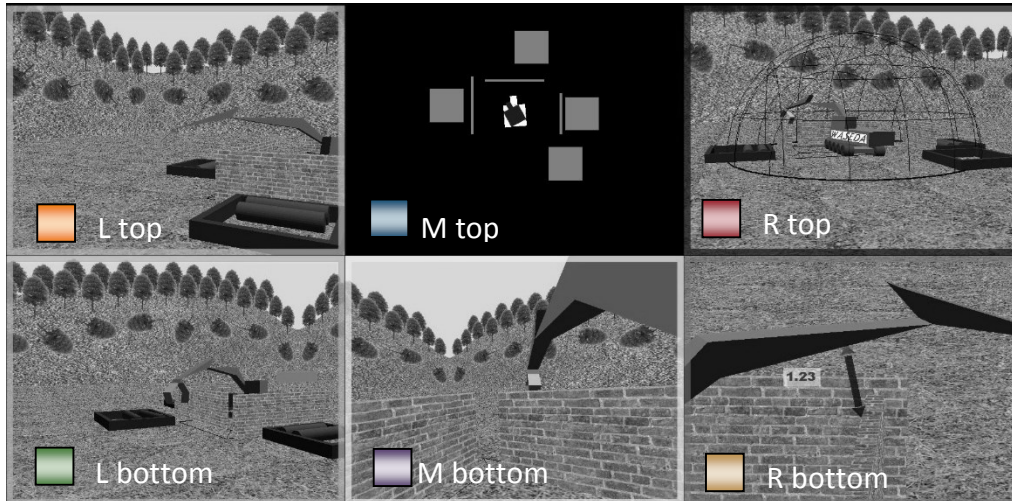
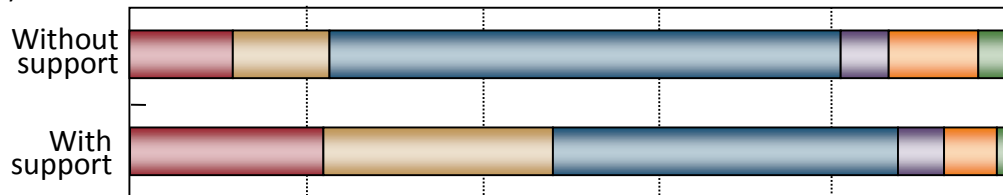**Figure 5.9 Questionnaires about AR vision prompts**

All of these three AR items can convert a 3D relationship into a 2D one, which is easy for operators to understand without a skilled sense of space. However, the distance arrow in 3D world has no such effect, because the direction of itself is not clear without taking the thickness change into consideration. In fact, in many cases, the thickness change is too slight to recognize. Rotation hint has the similar problem, which converts

the 2D rotation direction into a 3D one, because operators have to understand the rotation in front view by understanding that in a top view. These two AR items can hardly convert a 3D judgment problem into a 2D one, so they are not preferred by operators.
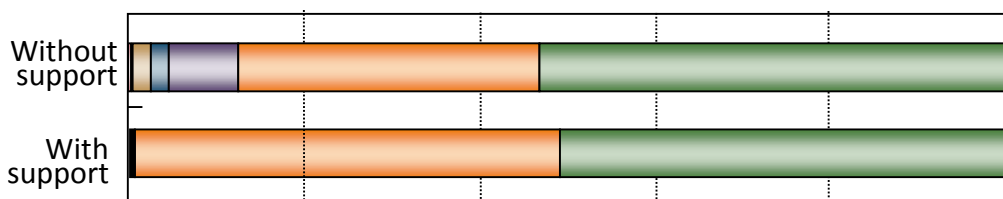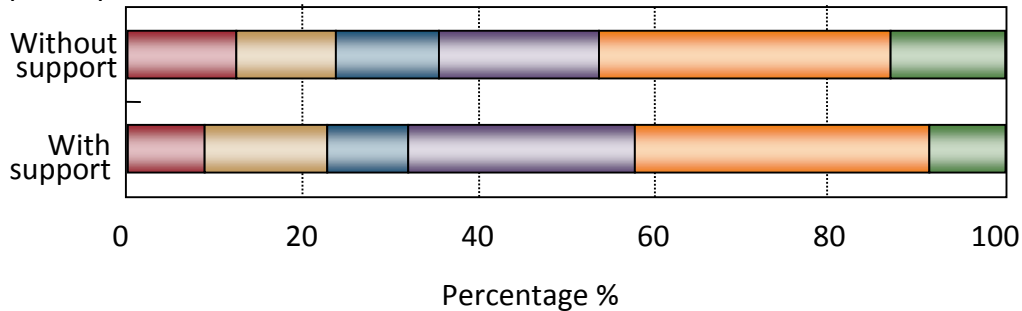
5.5.2 Visual attention inducement

Figure 5.10 Visual attention of each display

We also use eye tracker to measure the view point in each task condition, and the area viewed are divided into 6 according to each sub monitor area [70], [71]. Combined with Figure 5.9 (a), we find operators with AR prompt support prefer and watch 3D overlook views about twice more often than those without support in base movement condition. We think the reason should be the increase of useful information in overlook views attract operators' attention. In grasping condition, operators with AR prompt support can pay more attention to the left detail views. It means they can get more necessary information from detail views which make them almost need no glance at other views for further information. They confirm the usefulness of our AR prompt system as well. Nevertheless, the attention frequency of overlook views (including 2D), detail views and cab views changes not so much according to Figure 5.9 (b) and Figure 5.10. Viewpoints in transport condition are not fixed in few areas as expected.

Junjie YANG - February 2016

6 CONCLUSION

6.1 Conclusion

Three main problems in visual support system has been solved, which means the objective has almost been achieved. The effect of each system works well, and operator's performance in time efficiency, safety and accuracy increase significantly.

In camera placement system, different height of camera can be different, which is different from the conventional camera placement method. By using the system, required area can be videoed by at least two cameras from different aspects, which can supply operators with better views to understand work space better. In addition, camera can be changed quantitatively and modified manually according to the result of the system.

In autonomous camera control system, roles of cameras are defined according to the types of machine motion. And environmental cameras are arranged to play these roles. Thus making observation targets be tracked and framed according to the motion of machine. Cameras can also switch with others if some of them cannot play the occupied roles, which can ensure the operators observe reasonable front view and side one of construction machine and its manipulator to handle the situation of work site continuously.

AR vision support system succeeds in informing operator the invisible information which may be related to current operation. AR prompt items (arrows or numbers) are displayed in relative views respectively by calculating the situation around construction machine. A danger view was introduced to inform operator of the potential danger with

AR prompts. By using the system, operators can understand the space better and their attention should be attracted to related views about important area.

As a result, camera placement system succeed in using the least cameras to show most of the necessary areas of work site. In addition, Autonomous camera control system succeeds in processing different visual information according to the demand of photography. It is also able to arrange cameras to ensure visual information can be always given properly. In AR vision support system, danger view is introduced and AR prompts is able to be displayed with related videos. Eye sight is also be attracted to suitable views according to the situation around controlled machine. Thus making the efficiency, accuracy and safety of operation higher and increasing the user experience while decreasing the mental workload.

6.2 Future works

In order to make the whole system can be transplanted to real world, the differences between the VR simulation and real tele-operation system have to be made clear. They include the method of approaching information and that of camera control.

In VR simulation, all environment position and posture can be available easily. However it is almost impossible in real system. For example, the object can be recognized by the data in advance in VR simulation. In real system, extra devices and recognition algorithm become necessary. In this case, the recognition rate is always less than 100%. Therefore, the difficulty of approaching corrected information should be totally different.

By using VR simulation, the volume of each camera is ignored, so each rotation axis should go through the center of camera. In addition, the rotation speed of each camera is unlimited. Thus, in my proposed system, camera can adjust its posture to the object immediately even if it faced to totally different direction just now. In the real world, the rotation axis of camera are always skew lines, and the rotation speed has limitation. Thus, the camera control method should be changed, rotation time has to be thought about.

Thus, adjusting data approaching method and modifying camera control method have to be the improved first.

In addition, some improvements of hardware can be also introduced to my proposed system. For instance, cameras can be carried on drones instead of vehicles, which ena-

bles better views can be always available by a small number of cameras. However, the energy problem is still tough. And obstacle avoidance has to be thought about to keep the safety in work site.

What's more, the camera control system can be improved to serve more than one work machine, which needs the system to treat some work machine in an area as a group. A single video can be segmented into multiple pieces by digital zooming, which calls for higher photography demand. Additionally, high-level artificial intelligence is needed.

Moreover, with the development of wiggle stereoscopy, sense of depth and sense of distance will be more natural. At that time, it will become easier to use single view with AR technology to describe a single scene. The number of views can decrease.

Reference

[1] 日本応用心理学会, 応用心理学事典, 丸善出版, pp. 578-579, 2007.

[2] D. Bruemmer, C.Nielsen, D. Gertman, "How training and experience affect the benefits of autonomy in a dirty-bomb experiment" ACM/IEEE *Int. Conf.* on Human-Robot Interaction (HRI), pp. 161-168, 2008.

[3] R. Tsuzuki, G. Yamauchi, K. Nagatani, K. Yoshida, "Teleoperation of mobile robots using hybrid communication system in unreliable radio communication environments", IEEE *Symp.* on Safety, Security and Rescue Robotics (SSRR), pp. 1-5, 2014.

[4] X. Lin, H. Wang, Y. Wei, J. Duan, "Design of a wireless obstacle-crossable exploration robot", Control and Decision Conference (CCDC), pp. 4860-4864, 2015

[5] 古澤, "無人化施工に見る技術開発の歴史～情報化施工技術・ICT の先駆けとなった各種の取組み～", NETIS プラス, vol. 2, 2011.

[6] 村松, "情報化施工と今後の動向 ～無人化施工を事例として～", 建設機械, 2002.

[7] 山田, "建設・保全分野でのロボティックバーチャルシステム", 計測と制御, vol. 43, no. 2, 2004.

[8] Q. Li, L. Chen, M. Li, S. Shaw and A. Nuchter, "A Sensor-Fusion Drivable-Region and Lane-Detection System for Autonomous Vehicle Navigation in Challenging Road Scenarios", IEEE *Trans.* Vehicular Technology, vol. 63, no. 2, pp. 540-555, 2013.

[9] J. K. Kim, J. W. Kim, J. H. Kim, J. T. Hyung, Y. J. Park, Y. H. Ko and S. Jung, "Experimental studies of autonomous driving of a vehicle on the road using LiDAR and DGPS", *Int. Conf.* Control, Automation and Systems (ICCAS), pp. 1366-1369 2015.

[10] Y. Lin, J. Hyyppa and A. Jaakkola, "Mini-UAV-Borne LIDAR for Fine-Scale Mapping", IEEE Geoscience and Remote Sensing Society, vol. 8, no. 3, pp. 426-430, 2011.

[11] S. Chiang, X. Guo and H. Hu "Real time self-localization of omni-vision robot by pattern match system", *Int. Conf.* Advanced Robotics and Intelligent Systems (ARIS), pp. 46-50, 2014.

[12] "Optical vs. Digital Zoom", http://www.photoxels.com/digital-photography-tutorials/optical-digital-zoom/

[13] 本間, 以倉, 高井, "無人化施工映像技術に関する検討（バックホウ作業を主体として）", 建設機械, vol. 44, no. 10, pp. 57-61, 2008-10.

[14] 新田, "災害復旧に貢献する無人化施工技術", 土木技術, vol. 67, no. 4, pp. 16-23, 2012-04.

[15] 山口, 吉田, 石松, "遠隔操作におけるマンマシンインターフェイスに関する実態調査", 土木学会第 59 回年次学術講演会論文集, pp. 373–384, 2004.

[16] 北陸地方整備局, "無人化施工 設計施工マニュアル", pp. 2-7, 2015

[17] French, Thomas E., Vierck, Charles J., "A manual of engineering drawing for students and draftsmen (8th ed.)", pp. 111-114., 1953

[18] 建設無人化施工協会, "無人化施工による対応事例"

[19] Q. Xiao, B. Xiao, J. Cao and J. Wang, "Multihop range-free localization in anisotropic wireless sensor networks: a pattern-driven scheme", IEEE Trans. Mobile Computing, vol. 9, no. 11, pp. 1592–1607, 2010.

[20] Z. Yan, M.B. Yeary, J.P. Havlicek, and G. Fan "A new centralized sensor fusion-tracking methodology based on particle filtering for power-aware systems", IEEE Trans. Inst. and Measurement, vol. 57, no. 10, pp. 2377–2387, 2008.

[21] V. Akbarzadeh, C. Gagné, M. Parizeau, M. Argany, and M.A. Mostafavi, "Probabilistic sensing model for sensor placement optimization based on line-of-sight coverage", IEEE *Trans*. *Inst*. and Measurement, vol. 62, no. 2, pp. 293–302, 2013.

[22] D. Shreiner, G. Sellers, J. Kessenich and B. Licea-Kane, "OpenGL Programming Guide Eighth Edition", pp. 207-229

[23] M. Langford, "Basic Photography", pp. 83, 2000

[24] S. Ray, "Applied photographic optics: lenses and optical systems for photography", pp. 294, 2002

[25] X. Bai, S. Kumar, D. Xuan, Z. Yun, and T. Lai, "Deploying wireless sensors to achieve both coverage and connectivity", in *Proc. 7th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, pp. 131–142, 2006.

[26] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing", *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[27] J. Nocedal and S. Wright, Numerical Optimization. *New York: Springer-Verlag*, 1999.

[28] C. Zhu, R. Byrd, and J. Nocedal, "L-BFGS-B: Algorithm 778: L-BFGS-B, Fortran routines for large scale bound constrained optimization", *ACM Trans. Math. Softw.*, vol. 23, no. 4, pp. 550–560, 1997.

[29] N. Hansen and A. Ostermeier, "Completely derandomized self-adaptation in evolution strategies", *Evol. Comput.*, vol. 9, no. 2, pp. 159–195, 2001.

[30] S. Garcia, D. Molina, M. Lozano, and F. Herrera, "A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the CEC'2005 special session on real parameter optimization", *J. Heuristics*, vol. 6, no. 15, pp. 617–644, 2009.

[31] S. Haque and S. Srinivasan, "A meta-analysis of the training effectiveness of virtual reality surgical simulators", IEEE Trans. Information Technology in Biomedicine, Vol.10, No.1, pp. 51-58, 2006.

[32] C. Basdogan, M. Sedef, M. Harders, and S. Wesarg, "VR-based simulators for training in minimally invasive surgery", IEEE Computer Graphics and Applications, Vol.27, No.2, pp. 54-66, 2007.

[33] S. Nagiri, Y. Amano, K. Fukui, and S. Doi, "The study of driving support system for individual driver", Proc. IEEE Int. Conf. Systems, Man and Cybernetics, pp. 563-568, 2002.

[34] S. Strassburger, T. Schulze, M. Lemessi, and G. D. Rehn, "Temporally parallel coupling of discrete simulation systems with virtual reality systems", Proc. Conf. Winter Simulation, pp. 1949-1957, 2005.

[35] 関川, 本間, "無人化施工機械の操作性向上の検討"

[36] T. Ibaraki and S. Muroga, "Adaptive linear classifier by linear programming", IEEE Trans. Syst. Sci. and Cybernetics, vol. 6, no. 1, pp. 53–62, 1970.

[37] M.R. Marner, R.T. Smith, J.A. Walsh, and B.H. Thomas "Spatial user interfaces for large-scale projector-based augmented reality", IEEE Computer Graphics and Applications, vol. 34, no. 6, pp. 74–82, 2014.

[38] S. Zollmann, C. Hoppe, S. Kluckner, C. Poglitsch, and H. Bischof, "Augmented reality for construction site monitoring and documentation", Proceedings of the IEEE, vol. 102, no. 2, pp. 137–154, 2014

[39] W. Broll, I. Lindt, J. Ohlenburg, I. Herbst, and M. Wittkamper, "An infrastructure for realizing custom-tailored augmented reality user interfaces", IEEE Trans. Visualization and Computer Graphics, vol. 11, no. 6, pp. 722–733, 2005

[40] K.F. Hussain, E. Radwan, and G.S. Moussa, "Augmented reality experiment: drivers' behavior at an unsignalized intersection", IEEE Trans. Intelligent Transportation Systems, vol. 14, no. 2, pp. 608–617, 2012.

[41] D. Kalkofen, E. Mendez, and D. Schmalstieg, "Comprehensible visualization for augmented reality", IEEE Trans. Visualization and Computer Graphics, vol. 15, no. 2, pp. 193–204, 2008.

[42] S. Henderson and S. Feiner, "Exploring the benefits of augmented reality documentation for maintenance and repair", IEEE Trans. Visualization and Computer Graphics, vol. 17, no. 10, pp. 1355–1368, 2010.

[43] S. Haga and N. Mizukami, "Japanese version of NASA Task Load Index Sensitivity of its workload score to difficulty of three different laboratory tasks", The Japanese Journal of Ergonomics, vol. 32, no. 2, pp. 71–79, 1996.

[44] B. J. Nelson and P. K. Khosla, "Strategies for increasing the tracking region of an eye-in-hand system by singularity and joint limit avoidance", International Journal of Robotic Research, vol. 14, no. 3, pp. 255–269, 1995.

[45] M. Kamezaki, H. Iwata and S. Sugano, "Operator support system based on primitive static states in intelligent operated-work machines", Advanced Robotics, vol. 23, no. 4, pp. 1281– 1297, 2009.

[46] M. Kamezaki, H. Iwata, and S. Sugano, "A framework to identify task-phase and attentional-condition for supporting complicated dual-arm operations", Journal of Robotics and Mechatronics, vol. 22, no. 4, pp. 447–455, 2010.

[47] F. Kendoul, "Survey of advances in guidance, navigation, and control of unmanned rotorcraft systems", Journal of Field Robotics, vol. 29, no. 2, pp. 315–378, 2012.

[48] K. Kochersberger, K. Kroeger, B. Krawiec, E. Brewer and T. Weber, "Post-disaster remote sensing and sampling via an autonomous helicopter", Journal of Field Robotics, vol. 31, no. 4, pp. 510–521, 2014.

[49] A. Siebert and J. Teizer, "Mobile 3D mapping for surveying earthwork projects using an Unmanned Aerial Vehicle (UAV) system", Automation in Construction, vol. 41, pp. 1–14, 2014.

[50] C. Jung and C. Kim, "A unified spectral-domain approach for saliency detection and its application to automatic object segmentation", IEEE Transactions on Image Processing, vol. 21, no. 3, pp. 1272–1283, 2012.

[51] Z. Liu, R. Shi, L. Shen, X. Xue, K. N. Ngan and Z. Zhang, Z, "Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut." IEEE Transactions on Multimedia, vol. 14, no. 4, pp. 1275–1289, 2012.

[52] Y. Chen, L. Zhu, A. Yuille and H. Zhang, "Unsupervised learning of probabilistic object models (POMs) for object classification, segmentation, and recognition us-

ing knowledge propagation", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 10, pp. 1747–1761, 2009.

[53] R. Kelly, R. Carelli, O. Nasisi, B. Kuchen and F. Reyes, F, "Stable visual servoing of camera-in-hand robotic systems", IEEE/ASME Transactions on Mechatronics, vol. 5, no. 1, pp. 39–48, 2000.

[54] Dachsbacher, "Analyzing visibility configurations. IEEE Transactions on Visualization and Computer Graphics", vol. 17, no. 4, pp. 475–486, 2011.

[55] T. Ibaraki and S. Muroga, "Adaptive linear classifier by linear programming", IEEE Transactions on System, Science, and Cybernetics, vol. 6, no. 1, pp. 53–62, 1970.

[56] G. Charness, U. Gneezy and M. A. Kuhn, "Experimental methods: Between-subject and within-subject design", Journal of Economic Behavior and Organization, vol. 81, no. 1, pp. 1–8, 2012.

[57] 独立行政法人 港湾空港技術研究所, "前例の無い海洋土木の無人化は様々な技術を組合せ、共同研究で実現", Scope Net, vol. 70, pp. 13-15, 2014.

[58] 建設無人化施工協会, "人が入れない所や危険区域での工事を可能にした無人化施工技術", Scope Net, vol. 70, pp. 16-18, 2014.

[59] 淺間, 植木, "災害対応ロボットと運用システムのあり方", 産業競争力懇談会 2012 年プロジェクト最終報告, 2013

[60] 北陸地方整備局, "無人化施工 設計施工マニュアル", 2015.

[61] 大塚, "被災者及び救援者の心の傷と救助活動のあり方について", 2002.

[62] A. Nishiyama and M. Moteki, K. Fujino and T. Hashimoto, "Research on the Comparison of Operator Viewpoints between Manned and Remote Control Operation in Unmanned Construction Systems", *Proc.* 30th *Int. Symp.* Automation and Robotics in Construction and Mining (ISARC), pp. 677-684, 2013.

[63] M. Moteki, K. Fujino, T. Ohtsuki, T. Hashimoto, "Research on Visual Point of Operator in Remote Control of Construction Machinery", *Proc.* 28th *Int. Symp.* Automation and Robotics in Construction and Mining (ISARC), pp. 532-537, 2011.

[64] 気象庁｜活火山とは, URL:
http://www.data.jma.go.jp/svd/vois/data/tokyo/STOCK/kaisetsu/katsukazan_toha/katsukazan_toha.html accessed: 2014/08/01

[65] 防災科研, 6.2 活断層, URL:
http://www.hinet.bosai.go.jp/about_earthquake/sec6.2.html accessed: 2014/08/01

[66] 気象庁｜よくある質問集「地震について」, URL:
http://www.jma.go.jp/jma/kishou/know/faq/faq7.html#9 accessed: 2014/08/01

[67] 建設無人化施工協会, URL: http://www.kenmukyou.gr.jp/index.html accessed: 2012/04/01

[68] 株式会社フジタ, 東京大学, "無人化施工における俯瞰映像提示システムを開発", 2014/3/10.

[69] 古屋弘, 栗生暢雄, 清水千春, "3D画像と体感型操縦を用いた「次世代無人化施工システム」", 大林組技術研究所報, No.76, 2012.

[70] 三浦利章, "注意と眼球運動 その 2. 課題要件と情報の取り込み方", 労研維持会資料, No. 1295~1297, 1991/12.

[71] Fred H. Previc, "Functional specialization in the lower and upper visual fields in humans: Its ecological origins and neurophysiological implications", Behavioral and Brain Sciences, vol. 13, No. 3, pp. 519-575, 1990

[72] 藤野健一, 茂木正晴, 大槻崇, 西山章彦, 橋本毅, "大規模土砂災害等に対する迅速かつ安全な機械施工に関する研究", 2013.

[73] 中山俊一, 坂井龍二, "作業者の作業達成度に関する評価", 日本経営工学会誌, vol. 32, No. 5, 1981.

[74] 芳賀繁, 水元直樹, "日本語版 NASA-TLX によるメンタルワークロード測定−各種室内実験課題の困難度に対するワークロード得点の感度−", 人間工学, Vol. 32, No. 2, 2006.

[75] 舘章, 小森谷清, 澤田一哉, 井床利之, 井上幸三, "HRP 遠隔操作プラットフォーム", 日本ロボット学会誌, Vol. 19, No. 1, p. 16~27.

[76] 総務省消防庁, "消防活動における安全管理に係わる検討会報告書, 第 2 部 安全管理の心理学的分析 第 2 章 心理学の理論を用いた「慣れ」の除去", 2004/11.

ACKNOWLEDGEMENT

Junjie YANG - February 2016

Publication

- Academic papers

    o Inducement of Visual Attention Using Augmented Reality for Multi-Display Systems in Advanced Tele-operation, IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS), Oct 1st 2015, J. Yang, M. Kamezaki, R. Sato, H. Iwata and S. Sugano.

    o Performance Analysis in Advanced Tele-operation System Based on Introduction of Danger-Avoidance View, Int. Conf. Intelligent Robotics and Applications (ICIRA), Aug 26th 2015, J. Yang, M. Kamezaki, R. Sato, H. Iwata and S. Sugano.

    o A 3D Sensing Model and Practical Sensor Placement Based on Coverage and Cost Evaluation, IEEE Int. Conf. CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), Jun 9th 2015, J. Yang, M. Kamezaki, H. Iwata and S. Sugano.

    o Analysis of Effective Environmental-Camera Images Using Virtual Environment for Advanced Unmanned Construction, IEEE/ASME Int. Conf. Advanced Intelligent Mechatronics (AIM), Jul 9th 2014, J. Yang, M. Kamezaki, H. Iwata and S. Sugano.

    o Visibility Enhancement in Unmanned Construction, The 15th Symposium on Construction Robotics in Japan (SCR2015), paper no. O-24, Sep 7th 2015, M. Kamezaki, R. Sato, J. Yang, H. Iwata, and S. Sugano.

    o Visibility Enhancement using Autonomous Multicamera Controls with Situational Role Assignment for Teleoperated Work Machines, Journal of Field Robotics, Apr 9th 2015, M. Kamezaki, J. Yang, H. Iwata and S. Sugano.

    o A Basic Framework of Virtual Reality Simulator for Advancing Disaster Response Work Using Teleoperated Work Machines, Journal of Robotics and Mechatronics, Aug 20th 2014, M. Kamezaki, J. Yang, H. Iwata and S. Sugano.

    o An autonomous multi-camera control system using situation-based role assignment for tele-operated work machines, IEEE Int. Conf. Robotics

and Automation (ICRA), Jun. 5th 2014, M. Kamezaki, J. Yang, H. Iwata and S. Sugano.

o Development of a Tele-Operation Simulator Based on Virtual Reality Environment for Advanced Unmanned Construction, IEEE/SICE Int. Symp. System Integration (SII), Dec 17th 2013, M. Kamezaki, A. Gonzalo, J. Yang, H. Iwata and S. Sugano.

- Lectures

  o 複数画面を使った遠隔操作における注視映像と作業パフォーマンスの関連性分析究, SICE System Integration Division Annual Conf. (SI), Dec 14th 2015, M. Kamezaki, R. Sato, J. Yang, H. Iwata, and S. Sugano.

  o 拡張現実感を利用した複数の視線誘導手法による遠隔操作者の認知負荷軽減に関する研究, SICE System Integration Division Annual Conf. (SI), Dec 14th 2015, R. Sato, M. Kamezaki, J. Yang, H. Iwata, and S. Sugano.

  o Analysis of Spatial Perception Ability Based on Human Eyesight for Teleoperators, Proceedings of 6th International Conference on Advanced Mechatronics 2015 (ICAM2015), Dec 5th 2015, Kui Chen, Junjie Yang, Takahiro Katano, Mitsuhiro Kamezaki, and Shigeki Sugano.

  o Visual Attention to Appropriate Monitors and Parts Using Augmented Reality for Decreasing Cognitive Load in Unmanned Construction, Proceedings of 6th International Conference on Advanced Mechatronics 2015 (ICAM2015), Dec 5th 2015, Ryuya Sato, Mitsuhiro Kamezaki, Junjie Yang, Shigeki Sugano, and Hiroyasu Iwata.

  o 被覆率とコスト評価に基づく 3 次元異方性センサ群の実用的配置手法の提案, 第 33 回日本ロボット学会学術講演会論文集(RSJ2015), paper no. 3H1-06, Sep 3rd 2015, M. Kamezaki, J. Yang, H. Iwata, S. Sugano.

  o Research on Advanced Unmanned Construction Systems, The Robotics and Mechatronics Conference (Robomec2015), paper no. 2P1-P04, May 19th 2015, M. Kamezaki, R. Sato, J. Yang, H. Iwata, S. Sugano.

- o Research on Advanced Unmanned Construction System, Proc. of the 2014 JSME Conf. on Robotics and Mechatronics, May 26th 2014. M. Kamezaki, J. Yang, H. Iwata and S. Sugano.

- o Fundamental Analysis of Time Efficiency in Unmanned Construction by Using a Tele-Operation Simulator, SICE System Integration Division Annual Conf. (SI), Dec 20th 2013, M. Kamezaki, J. Yang, H. Iwata and S. Sugano.