GestureCam: Platform for Augmented Reality Based Collaboration

Hideaki Kuzuoka, Toshio Kosuge**, Masatomo Tanaka***

Institute of Engineering Mechanics
University of Tsukuba
Tsukuba, Ibaraki 305, Japan
Tel: +81-298-53-5258
E-mail: kuzuoka@kz.tsukuba.ac.jp

*Department of Communications and Systems Engineernig
The University of Electro-Communications
1-5-1 Chofugaoka, Chofu, Tokyo 182, Japan
Tel: +81-424-83-2161, ext. 3142, 3151
E-mail: kosuge,masatomo@qr.cas.uec.ac.jp

ABSTRACT
An approach supporting spatial workspace collaboration via a video- mediated communication system is described. Based on experimental results, independency of a field of view, predictability, confidence in transmission, and sympathy to the system were determined to be the system requirements to support spatial workspace collaboration. Additionally, a newly developing camera system, the GestureCam System, is introduced. A camera is mounted on an actuator with three degrees of freedom. It is controlled by master-slave method or by a touch-sensitive CRT. Also, a laser pointer is mounted to assist with remote pointing. As a future work, superimposed computer graphics and image processing are planned to be employed to the system. With these characteristics, the authors think that the GestureCam can be a platform for augmented reality supported collaboration. Preliminary experiments were conducted and the results are described herein.

KEYWORDS: Remote collaboration, CSCW, groupware, field of view, video-mediated communication, confidence in transmission, sympathy, SharedView, GestureCam, augmented reality.

INTRODUCTION
In a previous paper, the author classified collaboration in three-dimensional (3D) environment as spatial workspace collaboration [4]. Generally speaking, spatial workspace collaboration requires 3D expressions and motions which must be explained. To support spatial workspace collaboration, the SharedView video communication system was developed. Using the SharedView, instructions on how to operate machinery were given. As shown in Fig. 1, the operator wore the SharedView. The SharedView consists of the SharedCamera and a head-mounted display (HMD). The SharedCamera is a head-mounted camera which has the capability to follow a person's field of view. The SharedView has 1) portability, 2) ability to share a field of view, 3) capability to use superimposed gestures, and 4) ability to confirm the user's field of view as well as the actual task. However, several experiments

with the SharedView revealed the need for some new system requirements.

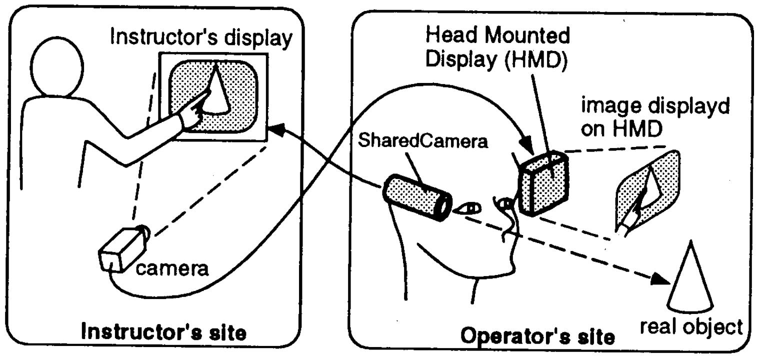

Figure 1: The SharedView system. The operator wears the SharedView. The SharedCamera's image is sent to the display at the instructor's site and the instructor use gestures in front of the display. The display and the gestures are taken by a camera and sent back to the operator's HMD. In this way, the instructor can give instruction with gestures.

Most of the current groupware supports two-dimensional work. However, to support spatial workspace collaboration, groupware should support collaboration in the real 3D environment. Recently some researchers started to use artificial reality system to support collaboration which deals with 3D objects [9]. However, since everything is virtual, such systems cannot deal with real 3D objects, thus users cannot manipulate real objects. On the other hand, an augmented reality systems are expected to be effective for supporting tasks in real world [2, 12]. Therefore the authors are aiming to develop augmented reality supported groupware which can support collaboration in the real 3D environment.

It is ideal for the augmeted reality system that the system can recognize objects in the real world and also the system can give appropriate instruction to the operator using it's knowledge base. However, since the purpose of the system is to support collaboration, the system will be used for unexpected situations which may not be in a knowledge base. Thus the authors think that both remote instructor and computer should be able to work together to augment the real environment. General requirements for a system is as follows.

- The system should have basic function to support collaboration.
- The system should have the ability to acquire information from the real 3D environment and process the information using computer.
- The system should have the ability to merge information provided by a computer with the real world.
- Not only a computer but also a human can show information to be merged with the real world.

The GestureCam video communication system is being developed to support spatial workspace collaboration. In order to determine the basic functions to support communication, the GestureCam was designed based on the experiences with the SharedView. Also, the system is designed so that it can be a platform to augment the real 3D world. In this paper, problems of the SharedView and requirements of the new system are described first. Then, the GestureCam and preliminary experiments using the system are introduced. Finally, some future works to improve the system toward the augmented reality system are described.

PROBLEMS OF THE SHAREDVIEW

The SharedView was used for instruction sessions at the Machining Center (MC). The MC is a numerically controlled machine that performs complex machining functions such as milling, drilling, etc., and can be operated either manually by switches or automatically by a computer. During each instruction session, the instructor remotely taught the operator to manually operate the MC.

Expressing Position

During the experiments, the instructor needed to express positions many times to show various objects to the operator. When SharedView was used, it was difficult to express positions compared with face-to-face communication.

The experiments showed that the task of expressing position consisted of several actions. Video observation of these actions performed in face-to-face instruction sessions were analyzed closely and showed that the instructor acted as follows:

1. The instructor began to turn his head to determine the next object to be explained.
2. After locating the object, he began to change his hand position to a pointing posture while moving his hand toward the object.
3. Actually pointing at the object.

Corresponding to the instructor's actions, the operator turned his head and changed his field of view toward the object. It was noticed that the operator began to change his field of view at an unexpectedly early stage in the instructor's pointing actions. For instance, the operator often began to change his field of view during 1 and 2 of the instructor's actions.

Before the operator began to change his field of view, he was sometimes looking at another object rather than at the instructor's head or hand. It is assumed that the operator detected the instructor's action using peripheral vision. Therefore, for the operator to respond to the instructor's position expressions smoothly, it is desirable for the operator to be able to sense the instructor's field of view and gestures.

For the instructor to find the next object to explain, it should be possible to see the object independently from the operator's field of view, however SharedView always showed the instructor the object that was in the operator's field of view. Under these circumstances, the instructor must ask the operator to turn his head toward the object using words and gestures. After the object was displayed on the screen, the instructor could point to the object. There are two possible ways to solve this problem: 1) another camera which the instructor can control remotely should be prepared, or 2) the camera's field view should be widened so that many objects can be seen in the display at the same time. In fact, instructors complained about the narrow field of view of the SharedCamera.

Confirmation

Since the instructor was apprehensive about damaging the MC, the cost of which exceeds US$100,000, he needed to confirm if (1) the operator understood the instruction, and (2) the MC is working properly. In face-to-face instruction sessions, this is accomplished by the instructor changing his field of view very often, switching between the operator's face, the operator's manipulation, and the MC motion. For example, when the operator turned the handle to move the drill of the MC, coordinate values displayed on the control panel were also changed. During this operation, the operator usually looked at his own manipulation or coordinate values. However, the instructor wanted to look at the drill and workpiece. If the drill strikes the workpiece in the wrong way, the MC can be seriously damaged. When SharedView was utilized, the instructor could not see the drill and workpiece when he wanted. Actually, the instructor complained that he was nervous when he could not see the drill and workpiece area.

SYSTEM REQUIREMENTS

In order to support basic function to support spatial workspace collaboration, initial requirements to the system was determined based on the problems of the SharedView system.

Independency of Field of View

As stated in the previous section, it is not always effective to share the fields of view. For position specification and confirmation, the instructor's field of view at times needs to be independent of the operator's field of view. Therefore, a good system should support both a shared field of view and independent fields of view. Furthermore, it is desirable for the operator to have the ability to know where the instructor is looking. This is important information for the operator to start changing his field of view smoothly.

Predictability

Face-to-face communication is smooth because a people can predict the other person's attitudes from many kinds of visual information [3, 10]. To support position expression of the instructor, a good visual communication system must show the instructor's field of view. Furthermore, this information should be non-verbal and it should be noticed involuntarily. From this information, the operator will be better able to predict where to look next.

Confidence in Transmission

In 1989, the authors developed a camera that was controlled by the instructor remotely. The operator was able to see the motion of the camera while he was getting instructions. The authors expected that the camera would be used as finger pointing, and the instructor would be able to use words like "this" and "that" effectively. In practice, although the camera was pointing toward a certain object (because the instructor was looking at the object through the camera), most of them explained verbally such as "a box on the desk in front of you", thus reducing the effectiveness of the camera. Then, the author became an instructor and used words such as "this" and "that" aggressively. In this case, the operator saw the camera's field of view and was able to find certain objects more efficiently. Based on this experience, the authors noticed that the user interface for instructors is important in addition to simply implementing a function. A system should make instructors feel like using the functions. To accomplish this, a system should make the instructor recognize that the function is effective.

In face to face communication, finger pointing is used effectively since

- the instructor can confirm that his finger was visible to the operator, and
- the instructor is confident that the operator is aware of his finger pointing.

For the remote pointing function to be used effectively, the instructor must be confident that perceptible information is transmitted to the operator's site. Also, the instructor must be able to confirm that the information is perceived by the operator. It is very important for communication system to provide confidence in transmission of information.

User Interface for a Remote Control Camera

If a remote control camera is used to satisfy independency of field of view, the user interface for controlling a camera is very important. The remote control camera mentioned in the last subsection had two degrees of freedom. The camera was controlled by a mouse interface. As the mouse was moved forward, backward, right, and left, the camera turned up, down, right, and left, respectively. However, controlling the camera with the mouse was difficult for many users. For example, while some users expect the camera to turn upward when a mouse was moved forward, other users expected it to turn downward. It seems likely that a joy stick would have the same problem. Furthermore, if a remote controlled camera has more than three degrees of freedom, a mouse or a joystick cannot be used conveniently to control all the freedom. Some other methods are required to control the camera intuitively.

Sympathy to the System

Finally, the system should have a sympathetic interface. Face-to-face communication is important because people can feel sympathy with each other [6]. In the case of remote instruction, if operators and an instructor can feel sympathy with each other, operators will listen to an instructor's instructions carefully and they will be able to ask questions frankly; therefore, an instructor will be able to give instructions more efficiently. On the other hand, if instructions are given impersonally, operators may feel that they are being controlled just like the machine. Such a situation would not be favorable for the system.

Furthermore, when a remote control camera is used to satisfy some of the previously mentioned requirements, it is important for operators to always be aware of the camera (gaze awareness [7]) and for the operators to feel that the camera represents the instructor. Sympathy to the system will help awareness to the camera.

GESTURECAM SYSTEM

The authors are developing a new video communication system, the GestureCam system, to support the abovementioned requirements. For independent field of view, a remotely controlled camera actuator was designed and named the GestureCam. For intuitive interface, the master-slave and touch-sensitive CRT methods are used. For predictability, the GestureCam was made small enough so that it can be placed and operated close to operators, thus it is easier for operators to notice its movement. In this paper, "GestureCam" means the remotely controlled camera actuator, and the term "GestureCam system" includes the user interface which controls the GestureCam. Implementation of the GestureCam system is described in the following subsections.

Master-Slave Controlled Camera

A new remote control camera was developed to support an independent field of view for an instructor. A little finger-sized camera was mounted on an actuator with three degrees of freedom (Fig. 2). This actuator is small enough so that it can be carried around. When the camera is at an inappropriate position to see a certain object, the instructor can ask the operator to move the camera to an appropriate place. Since the finger-sized camera has a small lens, its image is relatively clear across a wide range of distances without adjusting the focus.

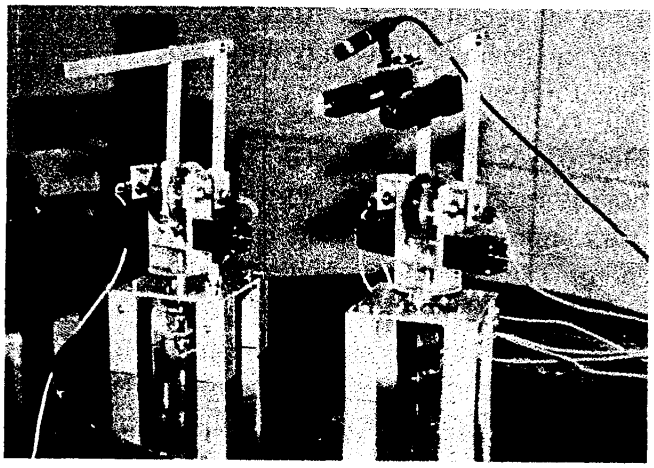

Figure 2: GestureCam system (left: master, right: slave).

It is ideal if a camera can be moved to any place in a space. If we design a bigger mechanism, the actuator will have more degrees of freedom but will be more dangerous because it will be closer to the users. Furthermore, such a big system would not be convenient in various situations.

The master-slave control method was employed because it is difficult to control the actuator with a mouse or a joystick. The master is identical to the slave except that camera is not mounted on the master. With this method, users will rarely get confused about the relationship between a controller and an actuator.

In a 3D environment, objects are placed at random positions and in random directions. If the camera is fixed, operators have to bring objects in front of the camera and turn them to the appropriate direction. If a certain part of an object is too small to see, operators would have to bring the object closer to a camera. Some objects may be heavy and difficult to move. However with the GestureCam, an instructor can control the direction by himself or by asking operators to move the GestureCam to see a certain object.

Because the motion of the actuator can be seen very well, the operator can involuntarily sense its motion in his peripheral view. If the SharedView is used simultaneously, the instructor can confirm whether the operator is looking at a slave, or looking at the proper object. In this regard, the combination of the GestureCam and the SharedView will support confidence in transmission.

Since the slave actuator has three degrees of freedom and it directly reflects the motion of an instructor's hand, its motion is humanized, and this apparently helps operators to feel sympathy to the GestureCam.

Although there are some systems which uses remote controlled camera or master slave robots, the GestureCam is different because it was designed to support human-to-human communication. Thus the GestureCam is designed to be small, to be portable, to support gaze awareness, and some other functions that will be explained in the following subsections.

Laser Pointer

For an instructor to use pronoun "this", a precise means of pointing at an object is required. For this purpose, a small laser pointer is mounted on the actuator so that it points in the same direction as the camera. Thus, an instructor can point at a certain object by controlling the slave actuator. With the laser pointer, an instructor can confirm that perceptible pointing information is transmitted to a remote site. If operators use the SharedView, an instructor can also be confident that the information is perceived by operators. Therefore, laser pointer satisfies the confidence required in transmission.

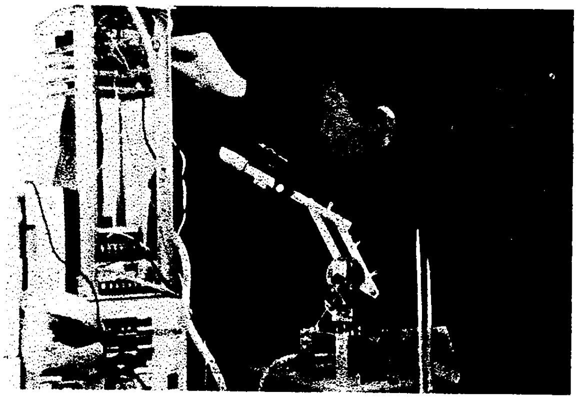

Figure 3: Operators can be aware of the field of view of the instructor.

By employing a laser pointer with the GestureCam, operators can be aware of the field of view of an instructor by the GestureCam's posture. Thus, operators can predict that an instructor may point inside his/her field of view (Fig. 3). In this way, the laser pointer helps more precise pointing. This feature of the GestureCam is expected to support smooth communication in 3D environments.

Touch-Sensitive CRT

Another type of user interface being tested in the touch-sensitive CRT (Fig. 4). Currently, this display is used by an instructor. A touch sensor is preferable to a mouse pointer because most people are accustomed to using fingers to point at an object. Furthermore, the mouse interface is slower because of the time required to reach for the mouse as well as time to move the mouse pointer to a certain object. When a finger is used, only the pointing time is required. Therefore, a touch-sensitive CRT is preferable for faster pointing in terms of both speed and ease of use.

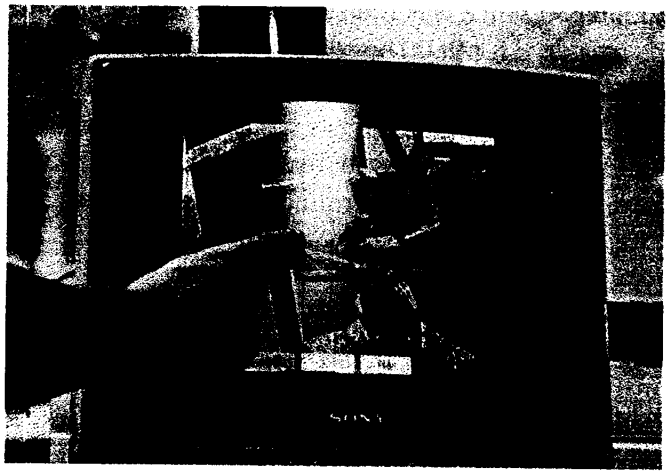

Figure 4: Using touch-sensitive CRT. Controlling camera and drawing annotations with a finger.

Controlling camera A touch-sensitive CRT is used as an interface to control the camera actuator. Three controlling modes (e.g. direction pointing mode, direct pointing mode, and dragging mode) are being tested.

In the direction pointing mode, a user touches the screen and the camera moves toward the touched position. Two degrees of freedom of the actuator are controlled. The distance and direction that the camera moves is defined by the distance and direction from the center of the display to the place where the user point at. Although this method is very simple and intuitive, precise control of the actuator is difficult because of current mechanical and software limitations. Therefore, it is not easy to point at a certain object precisely with the laser pointer.

In the direct pointing mode, a user simply touches a displayed object. Then, the camera is automatically moved so that the object becomes centered in the display.

In the dragging mode, a user changes the image on the display by "dragging" their finger across the display. Currently, this is the best method for controlling the camera precisely. Thus, it is easy to control the laser pointer's position.

Further experiments are required to clarify the advantages and disadvantages of each mode.

Superimpose drawings The touch-sensor is also used to superimpose drawings and a pointer on a displayed image. In the drawing mode, an instructor can use his finger to draw lines and write annotations. In the pointing mode, an arrow-shaped cursor follows an instructor's finger. This superimposed image can be seen both by an instructor and operators. Therefore, an instructor has two ways to specify the remote position.

PRELIMINARY EXPERIMENTS

As a preliminary experiment, a relatively simple task was instructed using the GestureCam system. The purpose of this experiment was to use GestureCam's functions aggressively to test it's basic functions for supporting spatial workspace collaboration. Furthermore, the authors tried to identify as many problems of the GestureCam as possible.

Settings

Communication was carried out between Tohoku University in Japan and the University of Hawaii in the United States. Video and voice was transmitted over a 64-kbps channel of the ETS-V satellite [11]. The telephone line was used for 9600 bps data communication. The instructor was Japanese and he gave instruction from Tohoku University. The operators were at the University of Hawaii and all of them were Americans. The instructor could not speak English fluently. Therefore, it was very hard for him to give technical instruction only verbally. Although the quality of the video transmission was not very good, it was important to determine the problems by conducting experiments with communication setups that were as realistic as possible.

In this experiment, only the master actuator was used to control the slave actuator and the touch-sensitive CRT was not used. The slave actuator, a mouse, TV display, and a computer display were placed in front of the instructor. He used a head set for voice communication. In the operators' site, microphones were placed close to the operators to transmit their voices. A TV display with a speaker was placed in front of the operators. The instructor controlled the master actuator and he could use mouse to control other features of the GestureCam system. The operators had no means of controlling the GestureCam system. Video images from the GestureCam were displayed on both displays at the instructor's and operator's sites. Due to the limited capability of the satellite, the operators usually did not see the image from the instructor's site. Microphones and speakers were prepared for voice communication.

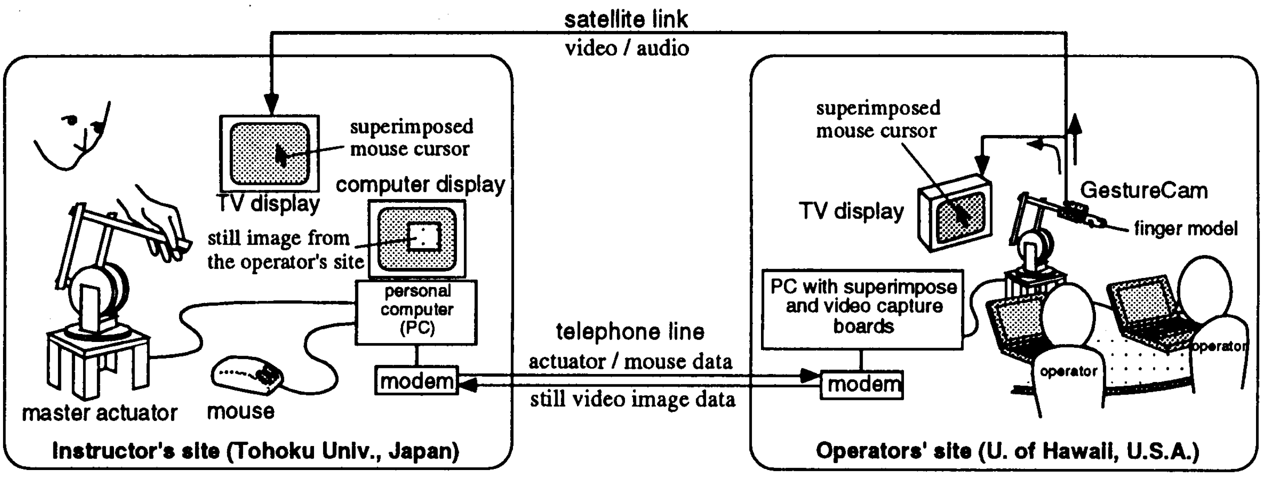

Figure 5: Setting for distance instruction experiment. Microphones and speakers were used for voice conversation, but they are not drawn in this diagram.

The laser pointer was turned on only when the instructor pressed the left mouse button. Since the quality of the video image from the satellite was not very good, a way to transmit still video images from the operators' site to the instructor's site was prepared. By pressing the right mouse button, monochrome still video image was captured in the computer at the operators' site and its data was transmitted to Japan via a telephone line. In this way, the instructor could see clearer images.

The mouse cursor was superimposed on video images so that it was visible to both the instructor and operators. However, only the instructor could control the cursor using the mouse.

It was important to let the operators feel that the GestureCam represented the instructor, that it has the ability to show the instructor's field of view, and that it has the ability to point at objects. It was also important to let the operators feel sympathy to the GestureCam for the previously described reasons. For this purpose, a wooden fake finger was attached to the GestureCam (Fig. 5).

Task

The instructor in Japan gave instructions to the operators in Hawaii. During the experiments, the instructor taught the operators how to connect two portable computers (Apple Macintosh PowerBook) with a network cable and how to share files among two computers (AppleTalk). To connect two computers with cables, the instructor had to show which cable to choose and which port to plug it into. To share files, the instructor had to show how to set up a software e.g. which pull down menu to choose and which buttons to click.

Two experiments were conducted. The same person served as the instructor for both experiments. The operators were using IBM compatible computers with MS Windows regularly.

Initially, the GestureCam was placed close to the operators and portable computers. However, the operators were told that they may place the GestureCam wherever they want during the instruction sessions.

Experiment 1 There were only one female operator. She was not familiar with Macintosh computers at all. A normal-angle camera was mounted on the GestureCam.

Experiment 2 There were two male operators. Only one of them had used Macintosh computers, but he had no experience setting up the networking facility. A wide-angle camera was mounted on the GestureCam.

Results

Generally speaking, the quality of the video image was not good. Especially when the image was moved, it took a few seconds for the image to became clear. Furthermore, due to the video signal problem, the quality of the video image was generally bad. Therefore, the instructor could hardly read the characters on the computer display. Often, it took more than 10 seconds until the image was refreshed.

Because of the language barrier, the instructor took a while to understand what the operators were saying. He also spoke English slowly.

The following results are based on interviews of the subjects.

- Due to the bad quality of the image, it was hard for the instructor to recognize the superimposed mouse cursor. He said he wanted a drawing function so that he could have specified a certain object more clearly.
- Due to the video transmission delay, it was hard for the instructor to point at a certain object with a mouse cursor or the laser pointer.
- The instructor could not recognize the operator's behavior, especially when the normal-angle camera was mounted on the GestureCam. When the wide-angle camera was mounted on the GestureCam, however, he could recognize the operators' behavior better but the results were still unsatisfactory.
- In experiment 1, due to the narrow field of view of the camera, it was hard for the instructor to find objects which were out of the field of view. Thus, he often had to ask the operator questions such as "please show me the cable". Then the operator moved objects or put the GestureCam in the desired position. Furthermore, after the GestureCam was moved by the operators, the instructor often lost track of his position.
- In experiment 2, because of the wide field of view of the camera, the instructor could see many objects at the same time. Thus, he could easily find objects to explain.
- One of the operators said one of the major cause of the long instruction session was the language barrier. However, she said that the laser pointer and other gesturing functions of the GestureCam did help communication.
- Video capture function was tried once. It took about a minute to transfer video data to Japan. However, the image was much clearer than real-time images, and the instructor could read the characters on the PowerBook's display. If video data can be transmitted in 10 seconds, such a video capture function may be useful.
- In this experiment, awareness to the GestureCam was effective. For example, to give instructions to only one of two operators, the instructor turned the camera toward the operator. Then the instructor could start giving instructions to the operator smoothly.

- During instruction, the instructor said hello to the operators and made the GestureCam bow in the Japanese way. This was a simple joke but it turned out to be very effective for instruction. The instructor said he was relaxed after the greetings. This unexpected fact is one of the proofs that the GestureCam has ability to make users feel sympathy. When people from different countries communicate with each other, communication often becomes mechanical because of the language barrier. Therefore, this ability of GestureCam is very important.

Although it is hard to evaluate GestureCam's effect from this experiment alone, the authors had the following impressions.

- The GestureCam's posture helps the operators to be aware of the instructor's field of view, and predict the position that the instructor will refer to next.
- If it takes time to point at a certain object with the laser pointer, the instructor will verbally explain the position. Therefore, it is important to make a good interface so that the instructor will be able to point as quickly as possible.
- The GestureCam has the ability to overcome the language barrier.
- Because the GestureCam has three degrees of freedom and is controlled by a master actuator, its movement is humanized and it helps to make operators feel sympathy to the system.
- Because the GestureCam is portable and it has three degrees of freedom, it has advantages in looking around the 3D environment, looking at an object from the appropriate direction, and getting closer to an object to see it more clearly.
- To know the positions of the GestureCam, other objects, and operators in the environment, an image of the overall environment is required.
- The system should support low bandwidth image transmission.

In the following section, future works to solve the abovementioned problems are described.

FUTURE WORK: TOWARD THE AUGMENTED REALITY BASED COLLABORATION

In this section, future works to extend the GestureCam into augmented reality based collaboration system are described. The GestureCam system already has real image capturing capability (so that the system can acquire image from real 3D environment and process the image), superimposing capability (so that the system can merge information with the real world), and touch-sensitive CRT (so that human can augment the real world). The following proposals that aims to utilize these functions further are currently under study.

Using Both Wide-Angle and Normal-Angle Cameras

During the preliminary experiment, the instructor wanted to see the overall image of the environment to confirm what the operators were doing, where the GestureCam was positioned, and where the other objects were located. For this purpose, the wide-angle camera is preferable. However, Gaver stated that there is a trade-off between field of view and resolution [3]. For instance, if the wide-angle camera is used, it is hard to recognize small objects. To see the overall environment as well as a certain objects clearly, both a wide-angle and normal-angle cameras should be used together. Two kinds of configurations of cameras are being considered.

The first way is to mount both cameras on the GestureCam as shown in Fig. 9. A wide-angle image will be displayed around a normal-angle image.

The second way is to prepare for the wide-angle camera stand and use it together with the GestureCam (Fig. 6).

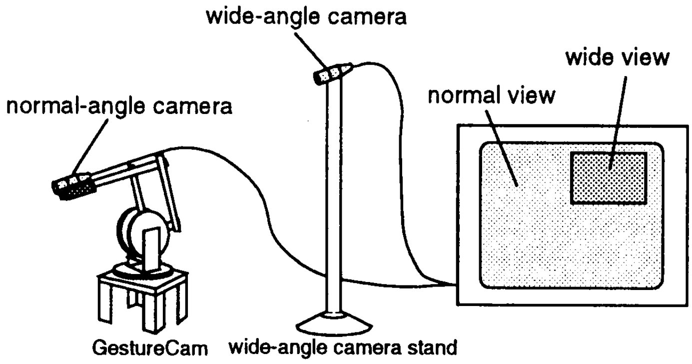

Figure 6: Use GestureCam and wide-angle camera stand together.

In this way, an instructor can confirm both the operator's field of view and his behavior. Moreover, because an instructor can easily find an object to show, position specification will become easier.

Functions for Low Band Width Video Link

Because still images can be sent clearly through a low band width video link, video capturing and superimposing functions are required. These functions have already been implemented in the current system. However, its effect could not be tested in the preliminary experiment because of the video signal problem.

In the preliminary experiment, the image of the mouse cursor was transmitted via satellite. Therefore, its image was blurred and hard to recognize. Touch-sensitive CRT should be effective in preventing this. Actually, the instructor said that he touched the CRT several times and it did not have any effect. Although an interface using the touch-sensitive CRT had been developed, it was not used for these experiments.

To compensate for the video transmission delay, an automatic laser pointing function will be helpful. An instructor will just point at a certain object on a display with a finger. Then a camera will move automatically so that a laser pointer will point at the object.

Mounting Small LCD

If an instructor's display is fixed, the image on the display moves in opposite direction from the master's movement. To solve this problem, small liquid crystal display (LCD) will be mounted on the master actuator (Fig. 7). With the LCD, an instructor can feel as if they are looking in a remote place through a small window.

The LCD can be turned 180 degrees. Then it will be able to either point to the same direction as the camera, or in the opposite direction. Depending on the GestureCam's application, users may change the LCD's direction.

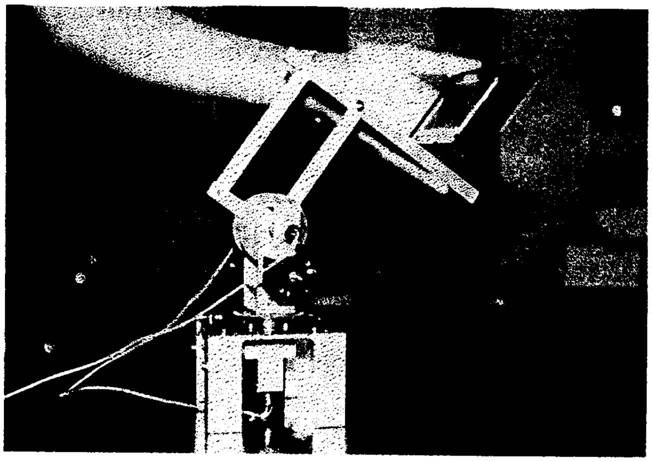

Figure 7: Mounting small LCD.

Laser Drawings
The laser pointer seems to be effective tool to display information on the real world. It will be also effective if the laser can draw some simple drawings on the real world. Even a simple arrow shapes to indicate directions or a circle which surrounds certain objects will be useful enough. However, very small and light mechanism should be developed.

Privacy
Preservation of privacy is an important issue of video-mediated communication systems. In the case of the GestureCam, the camera will always faces downwards as if it is sleeping when nobody is controlling the camera (Fig. 8). When the camera is in use, it wakes up and look around the environment. Although there are a variety of ways to preserve privacy[1], it is adequately preserved using this method.

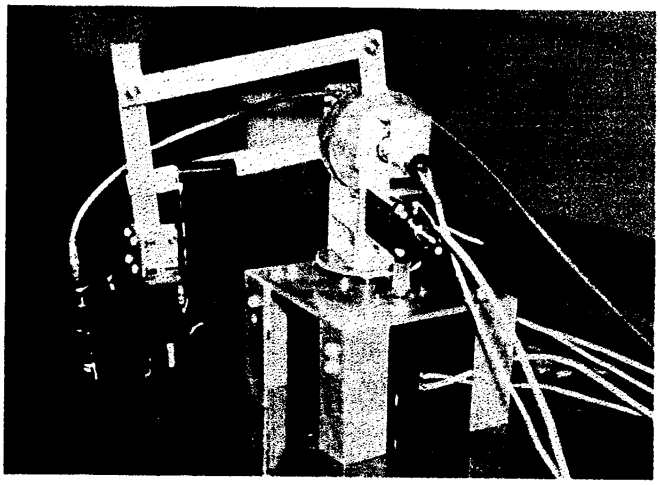

Figure 8: "Sleeping" position.

Bi-directional System
In this paper, only the instructor is able to control the camera. However, to support more interactive communication, the system should support bidirectional communication. Therefore, master and slave will be made identical (Fig. 9). The actuator can be attached anywhere using a clip or a magnet. Each actuator will be equipped with an LCD, normal-angle camera, wide-angle camera, laser pointer,

stereo microphone, and stereo speaker. Users can turn the LCD 180 degrees. When the system is used just for discussions, the LCD will face in the same direction as the camera. If the number of actuators is the same as the number of participants, the system can be used like Hydra [8] (Fig. 10). When the system is used to give instructions, the LCD will face the direction opposite to that of the camera. With this setup, objects which are taken by the camera will be displayed on the LCD and also users can see superimposed instructions.

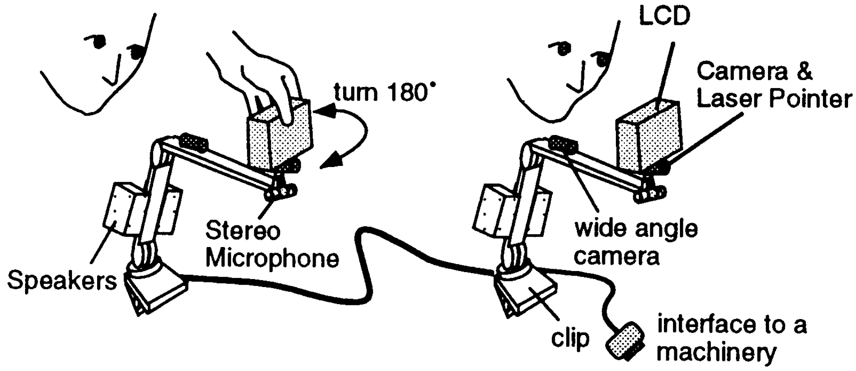

Figure 9: Future system design.

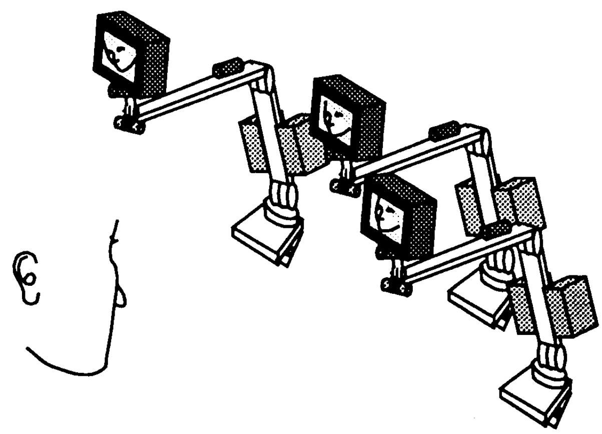

Figure 10: Hydra-like configuration.

Connecting to Machinery

It is desirable that computerized explanations and assistance is provided as well as verbal explanations by the instructor when complex 3D task is instructed [2]. Thus written and graphical assistance will be displayed on an LCD of the slave actuator. In this case, such assistance will be superimposed on the real image that slave's camera takes.

It will be convenient if each machine has assistance information for human and if it has a connector to transmit the information (Actually, some copying machines already display assistance information to recover from paper jam on their LCD display). In that case, a user will be able to connect the GestureCam to the machine, and assistance information will be superimposed on the LCD. Thus both an instructor and an operator can get a machine's conditions. In order to test it's effect, authors are planning to attach tiny CPU boards to some machines and store assistance information to those CPU boards. This function will be useful not only for cooperative work, but also for indevidual work.

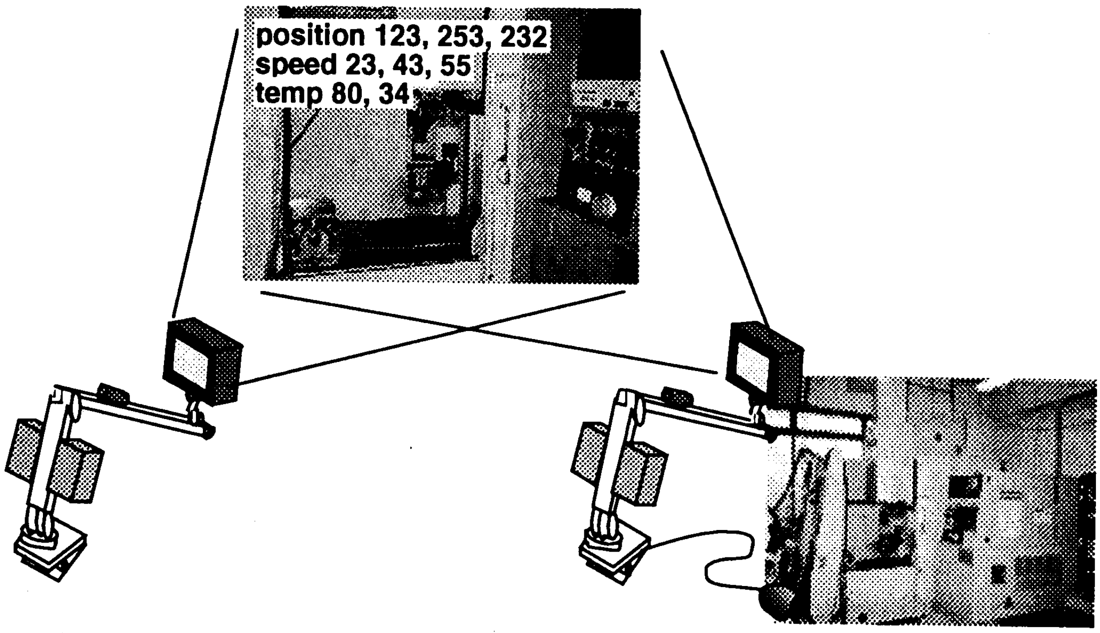

Figure 11: Connecting to machinery and superimpose assistance information.

CONCLUSIONS

In order to augment the real world and support users, more studies on utilizing image processing is required. However, the authors think that it will be quite difficult for the computer to recognize and understand all the objects in the real world. Some other idea to process the real image with relatively simple method will be necessary.

The preliminary experiments showed that the GestureCam's ability to support predictability and precise pointing may help communication between people from different countries, and with different levels of technical knowledge.

Combination of the SharedView and the GestureCam will satisfy many of the requirements of supporting spatial workspace collaboration and to implement augmented reality technology. However, wearing current SharedView will be annoying to operators. Some new device is required to replace the SharedView.

Through studies on various spatial workspace collaboration, the authors are aiming to develop a visual communication system for general purpose such as telephones.

ACKNOWLEDGMENTS

The work was supported by the Inamori Foundation and Japan Ministry of Education. The authors wish to express our thanks to Prof. Iwata and his students of University of Tsukuba for advise in developing the GestureCam. We also thank PEACESAT, University of Hawaii and Dr. Sun of Tohoku University for collaborating on the experiment.

REFERENCES

1. Brittan, D. Being There: The Promise of Multimedia Communications. *MIT Technology Review*. 95(4), May/June 1992, 42–50.
2. Feiner, S., MacIntyre, B., Seligmann, D. Knowledge-Based Augmented Reality. *Communications of the ACM*. 36, 7 (July 1993), 52-62.

3. Gaver, W. The Affordances of Media Spaces for Collaboration. In *Proc. of CSCW'92*, 1992, pp. 17–24.
4. Kuzuoka, H. Spatial Workspace Collaboration: A SharedView Video Support system for Remote Collaboration Capability. In *Proc. of CHI'92*, 1992, pp. 533–540.
5. Kuzuoka, H, Kosuge, T., Tanaka, M. GestureCam: Video Communication System for Sympathetic Remote Collaboration. In *Proc. of CSCW'94*, 1994 (to be presented).
6. Inoue, T., Okada, K., and Matsushita, Y. A Study of Expressions of Emotions from Body Movements for more Effective Computer-Mediated Communication (in Japanese). In *IPSJ SIG NOTE GW*. 94-GW-5, 1994, pp.33–40.
7. Ishii, H., Kobayashi, M. ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact. In *Proc. of CHI'92*, 1994, pp.525–532.
8. Sellen, A. and Buxton, B. Using Spatial Cues to Improve Videoconferencing. In *Proc. of CHI'92*, 1992, pp. 651–652.
9. Takemura, H. Cooperative Work Environment Using Virtual Workspace. in *Proc. of CSCW'92*, 1992, pp. 226–232.
10. Nardi, B., Schwarz, H., Kuchinsky, A., Leichner, R., Whittaker, S., and Sclabassi, R. Turning Away from Talking Heads: The Use of Video-as-Data in Neurosurgery. In *Proc. of INTERCHI'93*, 1993, pp. 327–334.
11. Yamanaka, H., Kobayashi, T. The Contribution of Distance Higher Education to National Development in Pacific Island Countries. In *Proc. of PTC'93*, 1993, pp. 204–211.
12. Newman, W., Wellner, P. A Desk Supporting Computer-based Interaction with Paper Documents. In *Proc. of CHI'92*, 1992, pp. 587–592.