Eighth Eurographics Workshop on Virtual Environments (2002)
S. Müller, W. Stürzlinger (Editors)

Scope-Based Interaction – A Technique for Interaction in an Image-Based Virtual Environment

Shunsuke Yoshida[1], Kunio Yamada[1], Kenji Mochizuki[1], Kiyoharu Aizawa[1,2] and Takahiro Saito[1,3]

[1]Telecommunications Advancement Organization of Japan
4F, Hortensia Hakusan, 1-33-16, Bunkyo-ku, Tokyo, Japan
{yoshida, ymdkn, mochi}@hrc4.tao.go.jp
[2]School of Frontier Sciences, the University of Tokyo, Tokyo, Japan
[3]Department of Engineering, Kanagawa University, Kanagawa, Japan

Abstract

Multimedia Ambiance Communication is a means to achieve shared-space communication in an immersive environment constructed of photo-realistic natural images where users can feel they are part of the environment. An image-based virtual environment is generally represented as an extensive field, in scenes showing mainly a landscape, and most objects are beyond the viewer's reach. Additionally, it usually has a single suitable point for observation because of limitations in the capture and representation methods of 3D-image spaces. Therefore, a special technique has to be developed that enables interaction with the environment. This paper describes the concept of a technique to interact with the scene based on a telescope-like virtual tool. The tool enables the user to stereoscopically view a distant object that will appear to be within reach, and to manipulate the object directly by putting a hand in the "scope". Hence, the user can handle objects at any distance, seamlessly and from the best viewpoint, without leaving an immersive environment.

Categories and Subject Descriptors (according to ACM CCS): I.3.6 [Computer Graphics]: Interaction techniques I.3.7 [Computer Graphics]: Virtual reality

1. Introduction

The ability to communicate with a true sense of presence would be desirable in the fields of communications and broadcasting. Sharing an image space with people in distant locations is a method to achieve such an immersive communication environment. Communications using two-dimensional (2D) image media, such as images captured by video camera, are now common. The addition of three-dimensional (3D) information and sounds from the transmitter's surrounding environment would lend a sense of reality to a shared image space. *Multimedia Ambiance Communication* is a means of achieving a shared-space communication environment. One of its goals is to create a photo-realistic 3D shared-space that users can feel a part of [1].

An interaction technique in such a virtual environment is useful not only as a means of communication with distant people or virtual objects, but also as a direct means of constructing and editing the immersive environment. However, an image-based environment created mainly from images of outdoor scenery is generally represented as an extensive field, and usually does not have a perfectly accurate 3D structure because of limitations in the capture and representation methods of 3D-image spaces. For instance, a flat or curved panoramic image is enough for expression of a distant mountain, whereas a detailed 3D structure is necessary for objects close to the observation point.

Various studies on representing an extensive environment have been reported. However, most have focused on artificial objects created entirely from computer graphics (CG). One of the characteristics of an image-based virtual world is that the best viewpoint of the world is only from a specific point and direction, which is usually the capture point of the actual scene, and another is that the expression of the scene objects becomes simplified with their distance from the viewer. In other words, the observed realistic virtual world does not have a perfect 3D structure.

In this paper, we propose a technique for interacting with an extensive image-based world using a telescope-like virtual tool (Figure 1). This window-shaped tool has the effect of a lens, and the manner in which it is used is a metaphor of a telescope or microscope. When a user 'holds' the scope with both hands as a telescope, he/she is able to see stereoscopically a distant object appear within arm's reach. The interface enables him/her to manipulate the distant objects directly, for instance, by putting his or her hand in the window. It is suitable for interaction with an image-based virtual environment, one in which resolution from the capture point is finest.

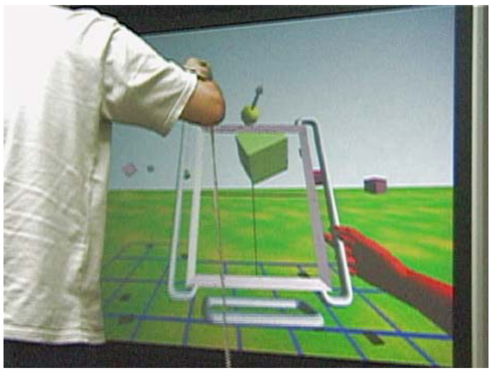


Figure 1: *The Scope-Based Interaction.*

2. Related Work – Interaction with the Virtual Environment

There are many previous works on users interacting with virtual objects [ex. 2]. None, however, are sufficiently general to address the problems of displaying an image-based virtual environment. Most only handle objects within the reach of the user, or use objects whose features are virtually exact.

Research on extensive virtual environments has targeted navigation and manipulation of objects (sometimes remote). Navigation has meant letting the user know the location where he/she is and giving him/her freedom of movement within a scene. Techniques include walk- or fly-through [3] and bird's eye view [4, 5]. The most popular manipulation technique uses a beam pointer [ex. 6]. Others include extending the user's arm [7] or preparing a world in miniature [8] for handling objects intuitively by hand. In some studies, the user is presented with a second viewpoint overlaid on the current scene [4, 5, 8]. This is helpful not only for navigation, but also for manipulations.

Our work assumes that some visible objects may be out of the user's reach. Additionally, it should be recognized that a natural effect is more desirable than models that are unrealistically accurate. We basically precluded options like viewing and manipulation from above a scene, which is map-like versus realistic, preferring instead to present all manipulations of virtual objects from the user's position in the 3D-image space.

The telescopic-sight interaction tool of this paper is useful for interacting with an extensive virtual environment based on images. We consider that a stereoscopic display and two-handed manipulation are important for immersive and intuitive interaction. The tool works as a virtual telescope through which the user is able to see and manipulate a distant object as if it were at close-range.

In this paper, we give an overview of our methods to represent a 3D-image space from real scenery as an example of an interaction target. We then explain the concept of scope-based interaction and describe how it can be implemented. After that, we compare our technique to other existing methods.

3. Construction of a Photo-Realistic 3D Image Space

3.1. Concept of a Three-Layer Image Space

The 3D-image space of our study has a three-layer structure as is found in landscape paintings, and which can be easily treated by computer.

The three-layered image space model expresses a large environment with a reasonably small amount of information without spoiling image quality. It does so by making the level of detail of a representation correspond to the distance from viewer. Table 1 summarizes the characteristics of human visual perception in relation to the distance from the viewer.

For example, using a true 3D structure for a background is too computationally expensive, because human visual perception and the display's resolution truncate depth information characteristics. Therefore, for objects like the sky, mountains, or forests, 2D-like expression should be sufficient.

On the other hand, middle distance objects like trees, buildings, or monuments look slightly different depending on which eye is used to view them. In this case, a pseudo-3D scene expression or a stereoscopic scene expression that enables the viewer to distinguish whether objects are in front of or behind other objects (as well as occlusion) and that provides rough depth information should be sufficient.

For foreground objects, a 3D structure is necessary to provide depth information so that these objects can be viewed from any direction. In communications, however, the priority is to convey visual information about persons in the other party, such as their facial expressions and gestures, and unlike for most objects, the inability to view a communication object from any direction is not a serious handicap.

Furthermore, images of the ground and of water surfaces such as those of lakes, which extend from foreground to middle or background, must appear continuous; in this case, they could be expressed as a polygonal CG objects.

Distance	Condition	Types of images required
Long	- It is possible to perceive that it is a distant landscape. - It can be perceived as an extensive space.	- 2D scene expression - Moving panoramic images
Mid	- Objects (and persons) appear or disappear when the location of a visual point/direction changes. - Whether objects are in front of or behind other objects can be perceived.	- Pseudo-3D scene expression - Stereoscopic images - Setting Representation - Sparse multi-view images
Short	- Objects are visible from the perimeter. - People do not have to be visible from all directions because the objective is communication.	- 3D scene expression - Polygonal Structure based on CG - Dense multi-view images

Table 1: *Three-layered image space: Considerations.*

3.2. Representation of 3D Image Space

3.2.1. Setting Representation using Stereoscopic Image Sequences

Our 3D-image space creation method integrates multiple data capture methods. For representation of long- and mid-range views, planar images are mainly used instead of an exact 3D structure. The space can be represented more simply without too much depth information through an approximate representation using planes. For example, they can be used like the stage settings to describe approximately a distant 3D structure [9]. Figure 2 shows the concept of Setting Representation based on the above considerations.

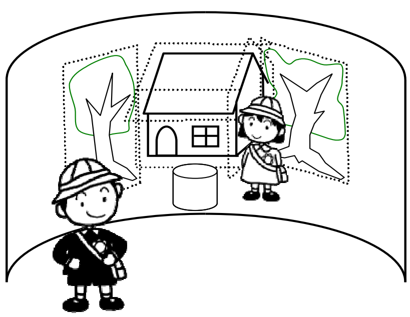


Figure 2: *Conceptual sketch of Setting Representation.*

We have developed a three-camera video system for capturing high-quality panoramic images [10]. Figure 3 shows images generated from images captured by the system. A panoramic image composed of left-, right-, and center-channels is generated from sequential panning shots recorded by the three-camera video system. The Census Transform method and the Region Competition algorithm are used to calculate the disparity map from the stereo images [11]. The disparity map is used for estimation and segmentation of the planar structures in the scene and for the texture mappings on the planes, and gives simple depth information to the images.

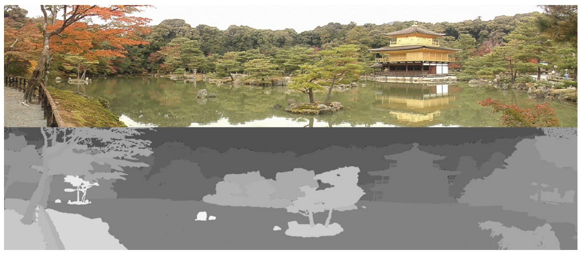


Figure 3: *Generated panoramic image and corresponding disparity map (The Garden of Rokuon-ji Temple).*

3.2.2. Integration of Range Data and Image Data

In the Setting Representation for mid-range objects, the plane parameters cannot be obtained when the tolerance value diverges, thus we also used a rough depth map to obtain the parameters of the estimated planes [1]. Additionally, the details of objects in short-range view must be expressed by using a polygon-based structure, and depending on the capture device this structure could be obtained from range data.

We used the LMS-Z210 3D Laser Range Finder to capture depth information. Although this device is able to obtain the color image of the scene at the same time, its resolution is too coarse to distinguish textures. To solve this problem, we used a digital camera and a video camera to generate high-resolution texture data. The images were captured from the same position as that of the range sensor, and with multiple viewing directions, and then integrated with range data. However, discontinuities become a problem when composing texture data from many captured images. We converted neighboring images into a single image with less distortion by using a Plane Transparent Transformation, with the central image estimated from range data (Figure 4).

3.2.3. Expression of Short-Range View and Life-Like Avatars

Objects in the short-range view must be viewable from any direction; and they must have an accurate appearance. While some objects will be constructed using the above method,

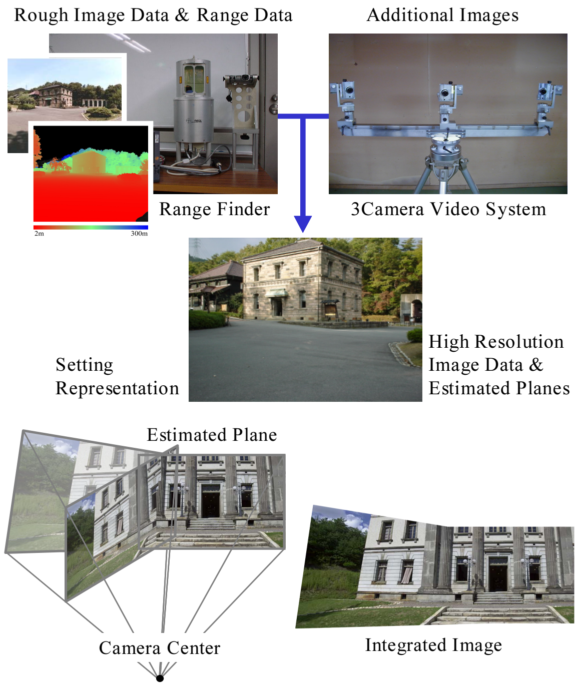


Figure 4: *Range data and image data measured by the range sensor and transformed texture (The Japanese architecture of the Meiji period [1868-1912] in the Meiji Village Museum).*

others, such as the floating menu tool, will be defined using CG.

Although this layer includes a representation for avatars of remote users, an exact representation is not necessary for communication; all that is needed is one that conveys the user's presence, gestures, and facial expression.

The avatar can be generated in two ways. One is based on synchronized video images captured from multiple angles. These camera images (or an image interpolated from two camera images per frame depending on the user's facing direction) are projected onto a cylinder in the shared virtual space. The other way is to use a highly detailed geometric model and high-resolution photographic texture. The avatar is derived from a human body template fitted to the user's body shape (the procedure uses a Cyberware WB4 whole body scanner) [1]. The voice of the user is used to control the lip shape and facial expression of the avatar in real-time (done by using the Media Conversion algorithm) [12, 13].

3.2.4. Improvement of Reality by Combination with CG Techniques

An example of a three-layer image space is shown in Figure 5. Note that in the figure the user's hands and objects such as the ground or the surface of the water are represented as CG objects.

The panoramic image is divided into the background and the ground, and a part of the latter is projected onto the virtual horizontal plane (its plane equation can be estimated from the setting parameters of the camera) [1]. The surface of the water is a CG image, because it is not static and because reflections change dynamically depending on user's viewpoint. Additionally, some effects such as the sunset in the scene are generated by using CG techniques.

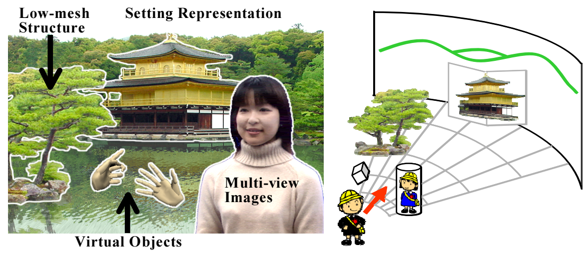


Figure 5: *A scene of three-layered image space.*

4. Scope-Based Interaction

4.1. Basic idea of Telescopic Interaction

Telescopic interaction, which we call Scope-Based Interaction, presents users with a seamless interface to interact with an image-based extensive virtual environment. When in use, the virtual scope tool floats in the 'air' in front of the user's viewpoint (Figure 1). The virtual scope operations are like those of a real telescope. To handle a distant object, the user frames the object in the virtual scope, which magnifies the object with its lens-like effect. Additionally, the scope can be used as an entrance to another scene. To manipulate a distant object, the user puts his/her hands into the scope and grabs the object.

Besides the merit of putting distant objects within easy reach of the user, the telescopic interaction makes selection with a beam pointer unnecessary. In addition, this interaction is good for image-based scenes because it minimizes travel within the imperfect 3D-image scene. An example of implementation of the interface is shown in Figure 6.

4.2. Basic Scope Operations

- **Taking out the Scope**

Some basic operations are related to physical actions to make them easy and intuitive for users. For example, the virtual scope is hung on the user's 'belt' in this implementation. The user touching his or her waist is the key to its activation. This is a good mnemonic to facilitate tool use.

- **Moving and Resizing the Scope**

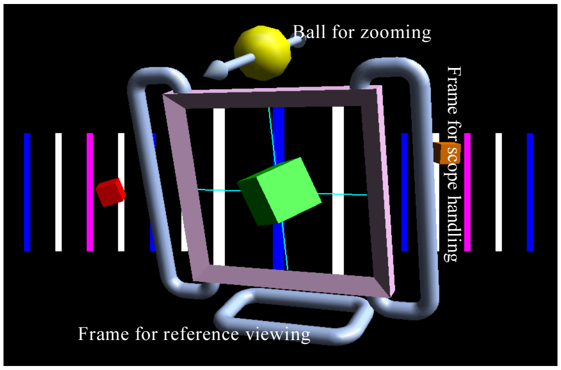


Figure 6: *"Interactive Scope". The user aims the crosshairs at the object. Pulling a ball of the top then zooms in it. The object (displayed stereoscopically) can be handled directly through the scope.*

There are frames at the scope's side for handling the scope. The movement and orientation of the scope are controlled by the user moving his or her arm and the inclination of the wrist, respectively. The activated scope floats in front of the user if the user releases his or her hand from the frame. And, the scope can be translated by the user's arm touching the scope's frame.

Additionally, the user can change the size of the scope by grabbing and enlarging the frames by pulling with both hands. By enlarging the frame to enough big size, the user can effectively work through the frame as if there is another viewpoint.

- **Storing the Scope**

When the scope is not needed, the action of putting it away acts as a command to remove the scope from sight. In our current implementation, a gesture to move the scope back to the waist and release it is the necessary action.

4.3. Interactive Scope as a Virtual Tool

- **Direct Manipulation of Objects within the Scope**

As mentioned before the scope allows the user to manipulate the scene it displays by passing his or her hand through it.

- **Manipulation using Partial Zooming Function**

If the user wants to manipulate large distant objects, full magnification in which objects are displayed as if the user were standing next to them (i.e., life-size) may not be appropriate. The scope, in this case, can display the virtual scene in miniature. When user uses this function, the tool calculates a reduction ratio with respect to the 'actual' size and the depth of the target so that the size of the target does not change because of the manipulation (Figure 7).

- **Reference View**

The scope will block objects behind it that would otherwise be in the user's line of sight. This is sometimes inconvenient. With an immersive display system of large-screen projection, especially, a user's hand blocking his/her sight could cause an incorrect stereoscopic effect. Moreover, the virtual environment is different from the desktop in that it does not have a support for the arms, which tend to be outstretched when operating the virtual tool. This can naturally lead to fatigue. Thus we included a function that suspends the frame crosshairs at the current viewpoint. The scope can then be moved to a more convenient location and the objects in it manipulated. In this operation, the scene as seen through the reference frame is displayed on the scope. (Figure 8).

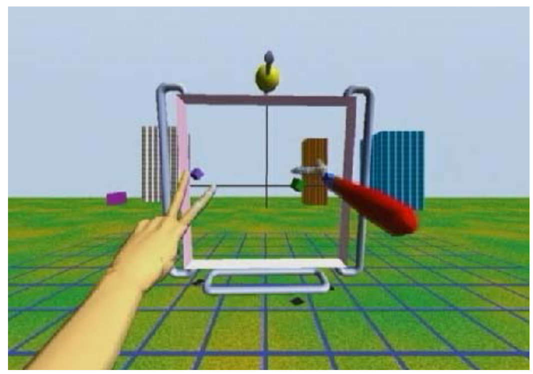


Figure 7: *An example of partial zooming. Hand signs have been assigned as keys for mode selection temporally; this case shows the hand sign 'V'. Note that you cannot see a zooming effect from this printout because of the definition of this function. In fact, the user can feel and grab the distant building in the stereoscopic display.*

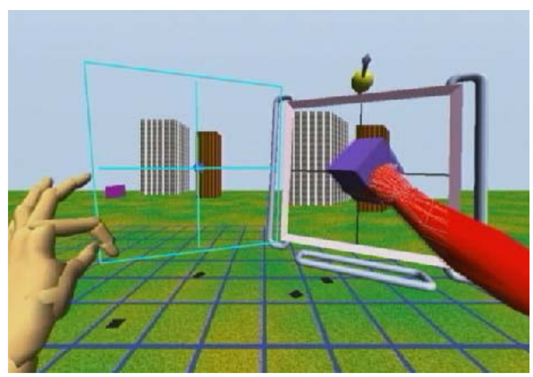


Figure 8: *A snapshot of using the reference viewer. When the user moves the scope by grabbing the bottom frame bar, the crosshairs are left in their current position for reference. The target on the crosshairs is zoomed in automatically by using the hand sign 'OK'.*

5. Implementation

5.1. System Configuration

Figure 9 shows the system configuration of this implementation. The stereoscopic display projects stereo images onto a large planar screen, which is viewed using stereo glasses with a polarizing filter. This method made it comparatively easy to get high luminance and high resolution. The user wears data gloves, and magnetic position sensors on the user's hands and glasses measure their positions and orientation. One managing PC and two rendering PCs create the virtual environment.

For two-handed interactions with the virtual environment, an artificial neural network is used for recognizing actions such as taking hold of a frame or handling objects. It has learned hand gestures for the actions, such as grab an object or release an object. The nearest gesture is obtained as an output from the current input information of the data glove, and a recognition result is used to switch the actions.

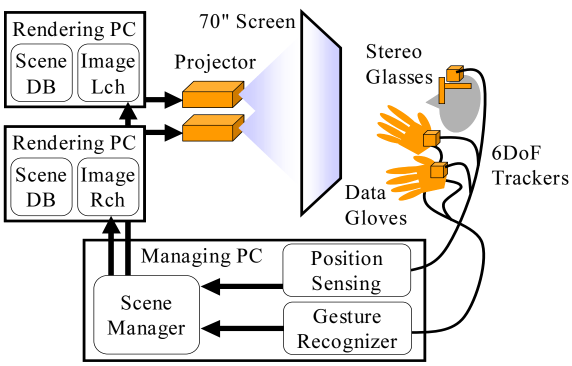


Figure 9: *System Configuration.*

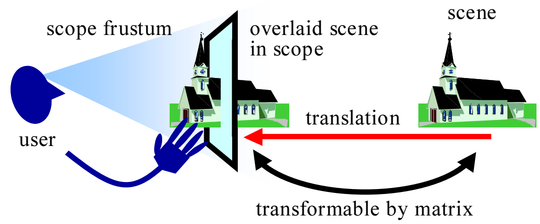
5.2. Rendering of the Scene within the Scope

Two-path rendering generates a lens-like effect inside the scope. The whole scene is represented as a scene graph of objects. In the first rendering path, the scene is drawn using the current viewpoint and the scene graph. In the second path, the drawable area is masked as the inside of a frustum defined by the four corners of the scope. The scene graph is scaled and translated toward the viewpoint, and then the whole scene is drawn. Hence, a user is able to feel the stereoscopic effect even from the scene inside the scope (Figure 10).

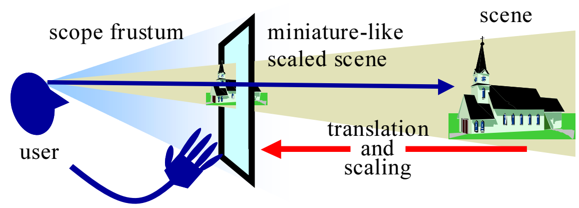
5.3. Interactions with the Scene

If the user is holding an object when a handling gesture is detected, he/she is able to manipulate the object by using the usual manipulation techniques for virtual objects. The measured positions of both hands and the positions of objects in the scene graph are compared in the world coordinates of the scene.

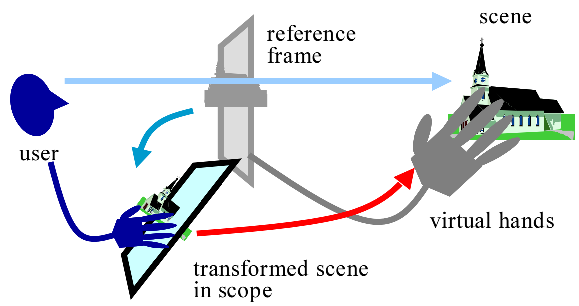
When the user has his/her hand inside the scope, the position and orientation of the hand are transformed into the coordinates inside the scope. Rendering of the second path only refers to the original scene graph of the world, which is not a miniature copy of the world. Hence, the transformation is calculated as an inverse matrix of the transformation used in the rendering process of the overlaid scene. Modification of the scene graph is reflected immediately in the viewed images. The interaction using the scope is like the user having an enormous virtual hand (Figure 10c).



(*a*) *standard zooming effect*



(*b*) *partial zooming effect*



(*c*) *manipulation using reference view*

Figure 10: *Generation of the scene in the scope.*

6. Discussion

6.1. Existing Techniques

Here, we compare our technique to the existing interaction techniques. Our goal in this study is to establish a method for immersively editing the extensive virtual scene, and getting some reactions from the scene through the interaction of users.

There are various approaches to interacting with an extensive environment. For example, walk- or fly-through techniques [3] or a navigation technique using a bird's eye view [4, 5] enable the user to travel through the extensive world. He/she is able to access any object by traveling to a desired place. Although these techniques are also useful for manipulation, we have not focused on them in this paper.

Two-handed direct manipulation and interaction without traveling were considered more important in this implementation. The former is necessary for intuitive interaction, and will become a base technique for a virtual environment editor using virtual tools. The latter saves time by eliminating the need to travel in the extensive world. This is desirable for an image-based scene because being able to do all operations from the best viewpoint is beneficial in that it prevents an observer wandering outside of the intended viewing area in an imperfect 3D scene.

The basic interaction techniques without traveling are roughly categorized as ray-casting, arm-extension, and miniature. They usually consist of two steps, selection and manipulation, and manipulation consists of translation and rotation. Table 2 briefly summarizes these characteristics.

Ray-casting is a widely used method because of its simple concept and implementation. A user is basically able to select any distant object. However, selection of objects at a great distance is difficult because even a delicate movement of the hand is amplified at the tip of the ray. Also, translation through arced movement and rotation at the ray-axis works only crudely. More sophisticated techniques are necessary for forward and backward translation and rotation [6].

Arm-extension is a more intuitive technique because the user's natural arm and hand motions are used to directly move and rotate objects. In the Go-Go interaction technique [7], the virtual hand's movement corresponds to a physical hand's motion according to a non-linear mapping function. This works well in a scene of limited space such as a room. However, a suitable mapping ratio has to be prepared for each scene, and the ratio for an outdoors scene is not easily defined. Techniques based on the button operation or the stretching action have been proposed to avoid this limitation [14], but the same problem as with ray-casting and other drawbacks remain.

The miniature approach is also used for navigation of a scene. Manipulation applied to the miniature scene is immediately reflected in the original scene, or as soon as tasks are completed [8]. This method makes the relationships between the positions of all objects easy to instantly understand. Therefore, selection and rotation is intuitive and rough translation is easy, if the scene is not larger than a room. When an open-air scene is used, a miniature becomes too small to be easily selected and manipulated.

Overall, there is no technique that works well in every situation, and each approach has merits and demerits. Hence, various hybrid techniques have been proposed. Bowman *et al.* [14] compared the ray-casting and arm-extension techniques experimentally, and found that ray-casting was better for selection, and the hand-centered manipulation of arm-extension was more suitable for rotation. Their proposed hybrid method uses ray-casting for selection, and an arm-extension operation for manipulation. The Scaled-World Grab [15] is another hybrid technique. When the user grabs a distant object by using a ray, the scene is automatically scaled down and displayed within the user's reach. The user can manipulate an object as if handling a miniature world, and then the scene returns to its original size when the object is released.

6.2. Comparison with Our Method

A comparison of these techniques and our proposed technique illustrates how our method incorporates part of each basic function. The action of aiming at a distant object with the crosshairs of the scope is similar to selection by ray-casting. The difference is the location of objects to be grabbed. The scope can stereoscopically display distant crowded objects as being close at hand. Hence, the user can easily select the object from an apparently nearby group of objects rather than fish for an object at the tip of a long wand.

Changing the scale factor by using the ball attached to the scope is analogous to the operation of stretching by using buttons or actions in the arm-extension technique. Here, the implemented function automatically zooms in on the object aimed at through the crosshairs, which saves time since the user does not have to adjust the scale factor of the scene. For rotation, the motion of the physical hand is directly reflected by the held object in the same way as with the arm-extension and the miniature techniques. This can be done just in front of the user, but not at a distant place.

There are two types of translation method. In the normal zooming effect (Fig.10a), the motion of the hand and the motion of the objects have a 1:1 correspondence. This is the same situation as in a general virtual-object manipulation method. In the partial zooming effect (Fig.10b), the motion is mapped with a 1:S (S is a scale factor) correspondence. This is equivalent to the miniature method, and the user can manipulate the scene on a macro-scale with an extensive viewpoint. The user can essentially do rough and detailed work at the same time, because the scale is easily changed to correspond to the use.

Technique	Usability	Features / Limitations
Ray-casting	S: good	- A ray is beamed in the user's indicated direction.
	T: fair	- An object that the ray crosses can be selected.
	R: poor	- Selection from crowded objects in a distant place is difficult.
		- Manipulation is done as if using a wand.
Arm-extension	S: poor	- The user's reach is virtually extended through a button operation, a stretching action, or motion mapping.
	T: fair	- Natural manipulation by direct hand motion (hand-centered manipulation).
	R: good	- Distance limitation: reach of virtual hand depends on that of the physical one (by motion mapping), the selection time needed for farther objects is excessive (by button operation or a stretching action).
		- Selecting and manipulating in a distant place is difficult.
Miniature	S: fair	- Miniature copy of the scene is prepared and displayed on a handheld tool.
	T: good	- The user manipulates miniature objects (icons) instead of the remote objects.
	R: good	- High computational cost of preparing the miniature copy.
		- Difficult to select and manipulate a complex and extensive scene.

S: Selection, T: Translation, R: Rotation

Table 2: *Method and limitations of basic interaction techniques.*

The most powerful feature of the scope-based interaction is that the user can handle distant and nearby objects in the same way when working through the scope. In addition, its definition as a virtual tool based on a window metaphor makes it easy to understand the concept of the reference view, and it can be expanded toward a multi-scope tool.

6.3. Application

We are considering applying this technique to a more concrete environment in our future work. For example, it can be used for looking after a virtual garden in an immersive environment. It will be useful not only for arranging the parts, but also for building the virtual garden. Automatic construction of a virtual environment from images has yet to be achieved. This technique might be useful for obtaining suitable plane parameters for Setting Representation.

Figure 11 shows an example of an image-based virtual environment when using the scope, and the scope's functions are summarized in Table 3. In the figure, the Golden Pavilion and trees in the long- and mid-range views are expressed using the Setting Representation technique. The water surface is expressed using a CG technique.

Fig.11a is a snapshot of interaction with the scene – the user has touched the water at a distant point sending ripples through the pond. In Fig.11b, the user moves a tree by using the reference view. By using the virtual scope technique, the user is able to interact with the extensive environment without traveling within it, as if he/she were treating it as a miniature.

For manipulation of the scope	
move	- grab the frame with one hand and move it
resize	- grab the frame with both hands and expand it
zoom	- grab the ball on top of the scope and pull it
automatic zoom	- make a hand gesture 'OK'
partial zoom	- make a hand gesture 'V'
reference view	- grab the bottom of the frame and move it
For manipulation of the scene	
arrange objects	- grab the object directly or put hands 'through' the scope
touch the pond	- touch the surface through the scope

Table 3: *Summary of implemented functions*

7. Conclusions and Future Work

The telescope-type interaction technique lets a user interact with a distant object in an extensive image-based virtual environment. This tool provides the user with a seamless working space; using the interactive scope technique, the user can edit a distant object without leaving an immersive environment. The tool is useful for a perfect virtual environment as well as an image-based environment.

The scope can be regarded as an extension of a panel interface [16]. Thus, previously proposed techniques can be applied to this interface [17, 18]. For example, a menu panel could expand the range of operations available to the user, and editing tools would be helpful for manipulating the virtual environment.

We now plan to test the usability of this interface in an experiment with a group of test subjects. We will also investigate which methods are most suitable for use as activation keys with two-handed manipulation.

Acknowledgements

We would like to thank Rokuon-ji Temple in Kyoto and the Meiji Village Museum in Aichi for letting us capture images of their properties.

(*a*) *interaction with the pond*

(*b*) *arranging the position of the tree*

Figure 11: *Application. The gesture of the left hand activates the automatic zoom functions.*

References

1. T. Ichikawa, S. Iwasawa, K. Yamada, T. Kanamaru, T. Naemura, K. Aizawa, S. Morishima and T. Saito. 3D Image Capturing and Representation for Multi-media Ambiance Communication, *SPIE Stereo-scopic Displays and Applications XII*, **4297A**:132–140, 2001.
2. Z. Huang, R. Boulic, N. Thalmann and D. Thalmann, A Multi-sensor Approach for Grasping and 3D Interaction, *Proc. Computer Graphics International '95*, pp.235–253, 1995.
3. W. Robinett and R. Holloway. Implementation of Flying, Scaling, and Grabbing, *ACM Symposium on Interactive 3D Graphics*, pp.189–192, 1992.
4. S. Fukatsu, Y. Kitamura, T. Masaki and F. Kishino. Intuitive control of "bird's eye" overview images for navigation in an enormous virtual environment, *Proc. of VRST*, pp.67–76, 1998.
5. S. Stoev, D. Schmalstieg and W. Strasser. Two-Handed Through-the-Lens-Techniques for Navigation in Virtual Environments, *Immersive Projection Technology and Virtual Environments 2001*, pp.51–60, 2001.
6. J. Mulder. Remote object translation methods for immersive virtual environments, *1998 Virtual Environments Conference & 4th Eurographics Workshop*, 1998.
7. I. Poupyrev, M. Billinghurst, S. Weghorst and T. Ichikawa. The Go-Go Interaction Technique: Nonlinear Mapping for Direct Manipulation in VR, *ACM Symposium on User Interface Software and Technology*, pp.79–80, 1996.
8. R. Stoakley, M. Conway and R. Pausch. Virtual Reality on a WIM: Interactive Worlds in Miniature, *Proceedings of ACM CHI 95*, pp.265–272, 1995.
9. Y. Ueda, M. Kaneko, T. Saito and H. Harashima. Representation of Moving Images with Skewed Planes and Its Application to the Video Coding, *Proc. of ICIP96*, **II**:505–508, 1996.
10. K. Yamada, T. Ichikawa, T. Naemura, K. Aizawa and T. Saito. High-quality Stereo Panorama Generation Using a Three-Camera System, *SPIE Visual Communications and Image Processing 2000*, **4067**:419–428, 2000.
11. K. Yamada, T. Ichikawa, T. Naemura, K. Aizawa and T. Saito. Structure Analysis of Natural Scenes using Census Transform and Region Competition, *SPIE Visual Communications and Image Processing 2001* **4310**:419–428, 2001.
12. S. Morishima and H. Harashima. A Media Conversion from Speech to Facial Image for Intelligent Man-Machine Interface, *IEEE JSAC*, **9**(4):549–600, 1991.
13. S. Morishima, *et al*. Life-Like, Believable Communication Agents, *ACM SIGGRAPH*, **Course Notes #25**, 1996.
14. D. Bowman and L. Hodges. An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments, *Proceedings of the 1997 Symposium on Interactive 3D Graphics*, pp.35–38, 1997.
15. M. Mine, F. Brooks and C. Sequin. Moving objects in space: Exploiting proprioception in virtual-environment interaction, *SIGGRAPH 97*, pp.19–26, 1997.
16. Z. Szalavari and M. Gervautz. The Personal Interaction Panel – a Two-Handed Interface for Augmented Reality, *Proceedings of EUROGRAPHICS '97*, pp.335–346, 1997.
17. E. Bier, M. Stone, K. Pier, W. Buxton and T. DeRose. Toolglass and magic lenses: The see-through interface, *SIGGRAPH 93*, **27**:73-80, 1993.

18. J. Pierce, A. Forsberg, M. Conway, S. Hong, R. Zeleznik and M. Mine. Image plane interaction techniques in 3D immersive environments, *Proceedings of the 1997 Symposium on Interactive 3D Graphics*, pp.39–44, 1997.