

EXHIBIT 2M TO MATSUDA
DECLARATION
(ENGLISH TRANSLATION)

Petition for *Inter Partes* Review of
U.S. Pat. No. 6,665,003

IPR2013-00326

EXHIBIT

Sony-1119

March 28, 2013

Certification

Park IP Translations

I, Christopher Girsch, hereby declare:

I possess advanced knowledge of the Japanese and English languages. The attached translation is, to the best of my knowledge and belief, a true and accurate translation from Japanese into English of the VRSJ Research Report.

I declare under penalty of perjury that the foregoing is true and correct.



Christopher Girsch

Park Case # 38124

VP Gaku Kenpo Vol. 2, No. 1
ISSN 1343-0572

- VRSJ Research Report -
Virtual Reality Society of Japan Research Report

SIG-CyberSpace
Virtual City Research Group

November 27, 1997

Virtual Reality Society of Japan

Table of Contents of the Virtual Reality Society of Japan Research Report
CONTENTS

[Virtual City Research Group]
[SIG-CyberSpace]

November 27, 1997 (Thursday)

VCR 97-11 A Virtual Park on Top of Spline - Bicycle Media Park –
TAKAHASHI Katsuhide, Eric Young-Sang SHIM, MIYAUCHI Nobuhito, SAEKI Toshiaki,
FUKUOKA Hisao (Mitsubishi Denki)

VCR 97-12 Generation of Panoramic Stereo Images from Monocular Moving Images KAWAKITA
Yasuhiro, HAMAGUCHI Yoshitaka, TSUKAMOTO Akitoshi, MIYAZAKI Toshihiko (Oki
Denki)

VCR 97-13 Video Workthrough for Positioning Live Video in a CG Space
KIHARA Tamio, NISHIMURA Tsuyoshi, NAKAKURA Kazuaki (NTT)

VCR 97-14 Methods for Expressing Realistic Action in Virtual Space
HONDA Shinkuro, KIMURA Shoryo, OZAWA Ryuji, OTA Kenji, OKADA Ken'ichi,
MATSUSHITA Yutaka (Keio University)

VCR 97-15 Examination of the Registration of Information Icons in 3-Dimensional Virtual
Space
INOUE Masayuki, KIYOSUE Yasuyuki (NTT)

Virtual City Research Group VCR 97-11
November 27, 1997

A Virtual Park on Top of Spline - Bicycle Media Park -
TAKAHASHI Katsuhide, Eric Young-Sang SHIM, MIYAUCHI Nobuhito, SAEKI Toshiaki,
FUKUOKA Hisao

Information Technology R&D Center, Mitsubishi Electric Corporation

We built the Bicycle Media Park, a virtual part in which multiple users can interact through a network. The Bicycle Media Park is equipped with the facilities of aquarium, space walk building, Ferris wheel and cycling racetrack inside a park that is 1.6 square kilometers, and multiple users who are distributed geographically can meet inside one park, share an experience with these facilities, and talk with one another by audio. This park was built with Spline, a software platform for distributed virtual environments (DVE) that we developed previously. We were able to confirm that it is possible to create distributed virtual environments flexibly and efficiently by employing Spline. In this essay, we describe the characteristics of the implementation of the service facilities inside Bicycle Media Park.

A virtual park on top of Spline — Bicycle Media Park —

Katsuhide Takahashi, Eric Young-Sang Shim,
Nobuhito Miyauchi, Toshiaki Saeki, Hisao Fukuoka

Information Technology R&D Center
Mitsubishi Electric Corporation

We have developed Bicycle Media Park (BMP) which enables multi-user interaction in distributed virtual environment (DVE). BMP has 1.6 km square terrain and provides several amusement facilities. Users of BMP explore the park, visit aquarium, space walk building, ferris wheel and bicycle stadium by riding bike. Geographically distributed users can come across, talk each other, and share their experience in the park. BMP is implemented on top of Spline, a software platform for building DVEs. Spline makes it easy and flexible to provide DVE applications. This paper describes design and implementation of facilities in BMP with Spline.

1. Introduction

In recent years, research on distributed virtual environments (DVE) has been thriving [1], [2], [3]. We built the Bicycle Media Park, a virtual part in which multiple users can interact through a network. This park was implemented on Spline [4], [5], [6], , a software platform for distributed virtual environments (DVE) that we developed previously. In this essay, we describe the characteristics of the implementation of the service facilities inside Bicycle Media Park.

2. Bicycle Media Park

In this section, we provide an overview of Bicycle Media Park and describe the service contents for each facility that is disposed inside the park.

2.1 Overview

This is a system wherein multiple users roam with a bicycle-type input-output device through 1.6 square kilometer park. The park is composed of a rugged topography, and the weight of the pedals of the bicycle-type input-output device change according to its uphill and downhill areas.

A user appears inside the park as an avatar (described below along with the motorbike), which is depicted as being mounted on a bicycle), and two users who meet one another can enjoy a chat with audio. Background noises, such as the chirping of birds, the water sounds of a pond, the fluttering of flags to express the wind or the rustling of the leaves of trees, elegant jazz sounds, etc., have been set for each place inside the park.

Service facilities that can be enjoyed by users like an aquarium and a space walk building, and user participation type service facilities such as a Ferris wheel and bicycle racetrack, has been set up inside the park.

2.2 Facilities and Services inside the Park

● Aquarium

At the huge water tank installed inside the park, users can enjoy the many fish swimming around inside it, as shown in Figure 1. In addition, when one approaches the water tank, the sound of bubbles can be heard.



Figure 1 Fish swimming around the inside of the water tank

- Space walk building

As shown in Figure 2, outer space that is broader than the capacity of the building stretches out inside the building. Users can observe the stars and planets, which revolve around a fixed star, and the astronauts from the observation platform, or walk in outer space.

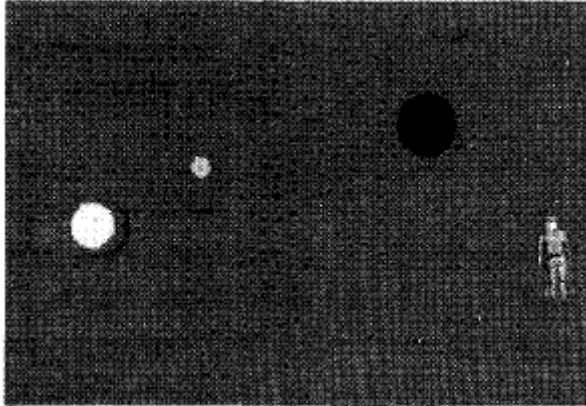


Figure 2 Interior of the space walk building

- Ferris wheel

When the motorbike reaches the boarding spot, it automatically boards the passenger car of the Ferris wheel, and rotates along with the passenger car. Users can look out over the scenery inside the park from a rotating viewpoint. When the passenger car makes one round, the motorbike automatically exits from the passenger car. Figure 3 shows the situation when one more motorbike is observing a motorbike that is riding on the Ferris wheel.

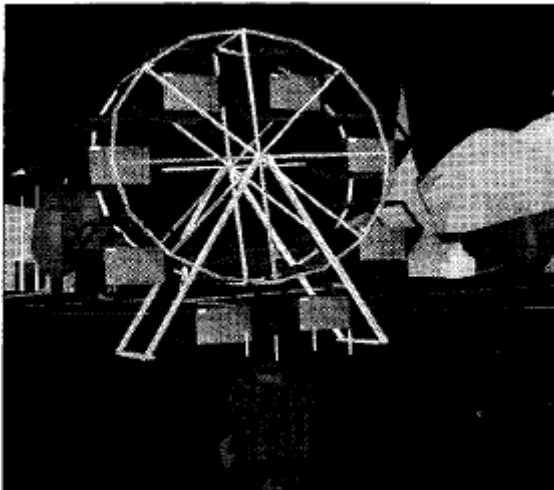


Figure 3 A motorbike watching the Ferris wheel

- Cycling racetrack

This is a facility that was created with a general racetrack with a 400 meter bank as its model. It is possible to race for 1 course with a pair of motorbikes, and in addition users can engage in time trial races, and can also race with a robot bicycle by computer simulation in which they can choose three speeds. An entry area for participating in these 5 kinds of races has been set up inside the racetrack, and a race is started automatically when a motorbike enters this area. Figure 4 shows a situation where a race with a robot bicycle has just started.



Figure 4 Race with a robot bicycle

3. Spline

Spline is a software platform on which DVEs are built, and is composed of a distributed world models that retain the information of the virtual space, a communications layer that conducts communications between each world model, and an application support layer that provides Spline API. The DVE application program is implemented on top the of the application support layer (Figure 5).

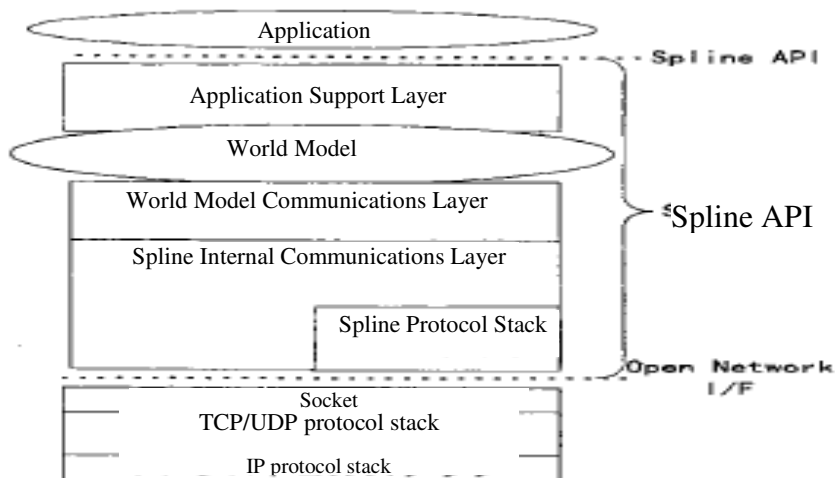


Figure 5 Configuration of Spline

In Spline, one Spline session is composed of the multiple processes composed of such a configuration.

3.1 Distributed World Model

Each application process that is participating in the same session under the Spline environment theoretically shares one world model. Each process acquires information from this world model, and carries out interactions inside the virtual space. In many dispersed virtual environment systems [1], [2], [3], the world model is managed by a concentration-type server, but in the case of Spine attention has been paid to scalability and flexible expandability, and a distributed world model that maintains the reproduction of the world model is adopted for each process.

In each process, the changes carried out for the respective world models are automatically reflected in the world model possessed by the other processes by the Spline communications layer, which has a reliability multicast mechanism. The information held by each world model is not entirely the same at the same instant, and maintains a loose consistency. The world models are composed of the instances of the classes defined by Spline (Spline objects) and their parent-child relationships.

3.2 Application Support Layer

The application support layer provides an API library in order that the application can access one's own world model. Each Spline process can freely add new objects inside the world model with this API, and can make changes to the objects for which the process itself holds ownership. In addition, it is also possible to transfer the ownership of each object to another process.

Spline provides two methods for expressing the movement of objects inside the 3-dimensional virtual space. The first method involves changing the position coordinates of the object. The changes to one's own world model of the position coordinates are updated to the other world models. The other method is the function called smooth motion. In smooth motion, the pathway of the object and the time until it reaches the terminus are specified. This information is transmitted to the other world models, and it is possible to move the objects inside the respective world models while describing a smooth orbit along the specified pathway, without this involving message communications thereafter. In addition, even after the terminus is passed, it is possible to keep making the objects move based on that pathway.

4. Implementation of Services

A description is now provided of the implementation and characteristics of the respective facilities of the space walk building, aquarium, Ferris wheel and bicycle racetrack inside the Bicycle Media Park.

4.1 Flexible Configuration

A motorbike process is present for each user who operates a motorbike the processes comprising the Bicycle Media Park. In addition, the respective facilities and services of the space walk building, aquarium, Ferris wheel and bicycle racetrack are implemented as separate processes. The functions of the respective processes are shown below.

(1) Motorbike process

This creates a motorbike for each user, carries out the input-output to the bicycle-type input-output devices, and displays a 3-dimensional virtual space on the screen.

(2) Space walk building process

This creates the observation platform, hallways, planets and astronauts inside the building, and imparts movement to the planets.

(3) Aquarium process

This creates the fish inside the water tank, and makes them move in circles.

(4) Ferris wheel process

This creates the Ferris wheel facilities, causes the passenger cars and rotating plate of the Ferris wheel to rotate, and causes the motorbike to enter or exit a passenger car of the Ferris wheel.

(5) Bicycle racetrack process

This creates the racetrack, and controls the motorbikes and robot bicycles in each race.

The respective processes are composed in such a manner that they add facilities and services to the virtual park, and they can change the form of the facilities inside the park flexibly. If the facilities are not used, it is all right if the process is not booted up. In addition, in the event that new facilities and services are provided to the Bicycle Media Park, these can be realized by adding processes that manage the facilities, and provide the services.

4.2 Smooth Motion

The fish in the aquarium, the planets inside the space walk building and the passenger cars and rotating plate of the Ferris wheel repeat fixed motions. These motions have been realized as the rotating movement that employs Spline's smooth motion.

● Rotation and orbit of planets

In the space walk building, one fixed star and three planets have been respectively arranged therein, and these are rotated on their own axes. At the same time, the three planets are made to orbit with the fixed star as the center. As shown in Figure 6, the orbiting of the planets is realized by setting the object of the fixed star as the parent object of the three planet objects. The planets, which are the children, are disposed at positions that are removed by only the rotational diameter of the orbit from the position of the fixed star. An orbital movement is imparted to the three planet objects due to the fact that the fixed star object rotates on its own axis.

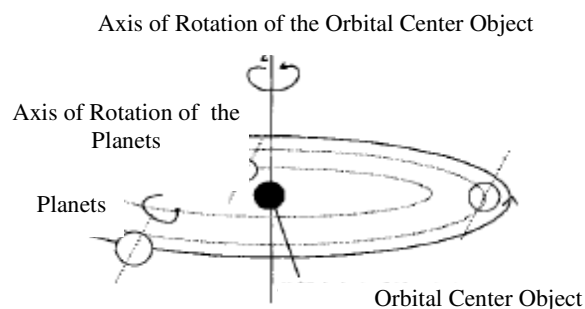


Figure 6 Rotation and orbit of planets

- Passenger cars of the Ferris wheel

Figure 7 shows the use of smooth motion for the Ferris wheel. The rotating plate of the Ferris wheel rotates in a clockwise direction, and the passenger cars of the Ferris wheel rotate in a counter-clockwise direction. By setting the rotational speed of the rotating plate and passenger cars as the same speed, it maintains the orientation in the vertical direction like that of the passenger cars of actual Ferris wheels.

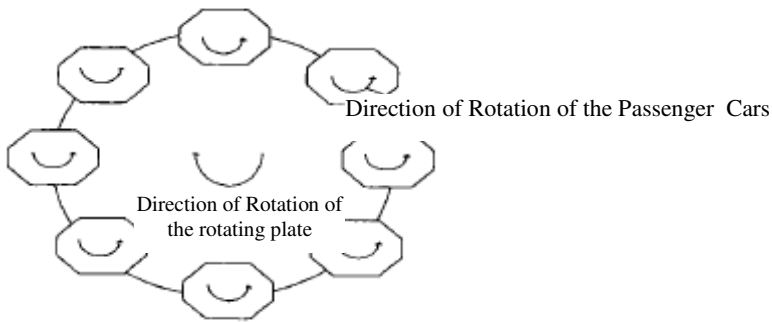


Figure 7 Rotation of the Ferris wheel

- Movement in circles of fish

Figure 8 shows the circular swimming motion of the fish, which employs smooth motion. A rotating object (parent 2 object) that does not have visual data is created at the center of the orbit, and this is set as the parent of the parent 1 object. Moreover, a fish object is set as the child of that parent 1 object. By rotating the 2 parents, the double orbital movement overlaps, and the fish objects are caused to engage in a complicated circular motion.

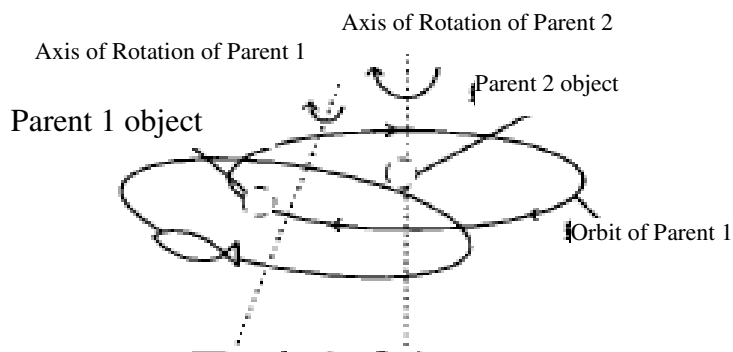


Figure 8 Movement in circles of fish objects

Examples of the use of smooth motion for the three facilities have been shown above. In the event that facilities that have objects that repeat a fixed movement, the implementation can be done simply by employing smooth motion. In addition, a reduction in the communications volume can also be cited as one of the advantages of smooth motion. The reason for this is that in the case of smooth motion the position coordinates of the object to be moved are not updated between world models, compared to the movement processing of ordinary objects.

4.3 Position monitoring and transfer of ownership

The entering and exiting the passenger cars of the Ferris wheel by the motorbike and the racetrack races are realized by position monitoring of the motorbike and the transfer of ownership from the motorbike.

● Entering and exiting the passenger cars of the Ferris wheel

The Ferris wheel process monitors the position of the motorbike, and if the motorbike is at the boarding spot for the Ferris wheel, it acquires ownership from the motorbike. The Ferris wheel process selects the passenger car in which the motorbike will ride, and sets the motorbike as the child of the object of the passenger car. The motorbike rotates together with the passenger car. At the moment when the passenger car for which the motorbike has been set as the child has made one round, the Ferris wheel process sets of the position of the motorbike at the unloading spot, and returns ownership to the motorbike.

● Bicycle race

The bike racetrack process monitors the position of the motorbike, and if the motorbike is at the entry area, it acquires ownership of the motorbikes that will participate in the race and of the robot bikes according to the race. Then, the bike racetrack process returns under the control of the respective processes by returning the ownership to the motorbikes and the robot bicycle, and the race starts. In addition, the bike racetrack process determines the conclusion of each race and monitors whether the motorbikes and the robot bicycle are traveling in the right direction on the bank. These two functions that are performed in order to monitor the race are also realized by position monitoring of the motorbikes and the robot bicycle.

The services of all of the facilities are realized due to the fact that the motorbike process and the processes for managing the facilities act in coordination, with the position monitoring and the function for transferring ownership as the media. The only processing that is required on the motorbike side is the processing for turning over ownership to the processes on the service side and making a request for the return of ownership. In this manner, user participation-type services can be easily provided by employing position monitoring and the function for transferring ownership.

5. Summary

In this essay, we introduced the aquarium, space walk building, Ferris wheel and cycling racetrack, which are the service facilities of the Bicycle Media Park, and described Spline. Moreover, we demonstrated that the services provided by these can be flexibly expanded by employing Spline, and we related that user participation-type services can be easily realized by employing the three functions provided by Spline, namely slow motion, transfer of ownership and position monitoring.

At present, in the means for providing services based on position determination, the motorbikes that enter the entry area are compelled to participate in the bicycle race, and there is no room for choice by the users. As a means for allowing the users who operate the motorbikes to choose, a user interface wherein for example a menu for selecting the races in which to participate is displayed when that user is inside the bicycle racetrack is conceivable, but it is necessary to add a great deal of processing to the motorbike on the side where the services are provided.

In addition, in the event that services such as schools and municipal offices, which will likely be included in virtual cities, are realized, even more processing will be needed, and adding processing to the avatar side in order to receive services is unrealistic. From the standpoint of development of the virtual city, it is necessary to consider a general purpose distribution means or use means for the service processing functions, from the side providing the services (facilities) to the side to which they are provided (avatar), and this is a future issue.

Based on the above-mentioned issues, we are planning to carry out an examination of the unified standards for avatar and classification of the service processing functions, through the addition of new avatars and the creation of a user interface for participation in the bicycle racetrack races.

Bibliography

[1] Pak, "A VRLM Application System That Creates a Virtual Space" FUJITSU.48., 2, March 1997.

[2]S. Ichinose et al. "InterSpace: NTT's Cyberspace Communication Platform", NTT REVIEW,Vol.9, No.3, May,1997.

[3]Honda, Y. et al. Virtual society. Procs. of VRML'95,San Diego. USA. Dec. ACM press 1995 pp.109-116.

[4] <http://www.merl.com/projects/spline>.

[5] FUKUOKA, et al, "Spline, a Software Platform for Distributed Virtual Environments", *Mitsubishi Denki Giho*, February 1997.

[6] SATO et al, "Virtual Tradeshow: One Application System of Spline", First Symposium of the Virtual City Research Group, VR Society, July 1997.

SIG-CyberSpace, VCR 97-12
November 27, 1997

Generation of Panoramic Stereo Images from Monocular Moving Images

Yasuhiro Kawakita, Yoshitaka Hamaguchi, Akitoshi Tsukamoto, Toshihiko Miyazaki
Kansai Laboratory Research & Development Group
OKI Electric Industry Co., Ltd.

This paper presents a technique for generating left-right panoramic images for binocular stereoscopic viewing from images from a single video camera, and a technique for stereoscopic display of panoramic images corresponding to the orientation of the user's line of sight by making use of parallax corresponding to depth. In these techniques, a video camera is attached to a tripod and rotated manually, and the optical flow between the continuously captured images is used to excise the vertical slit images and combine the images to generate left-right panoramic images. In addition, parallax corresponding to the distance to objects in the panoramic image from the video camera position is used for alignment in compositing the left-right panoramic images.

Generation of Panoramic Stereo Images from Movie using Single Video Camera

Yasuhiro Kawakita, Yoshitaka Hamaguchi, Akitoshi Tsukamoto, Toshihiko Miyazaki
Kansai Laboratory Research & Development Group
OKI Electric Industry Co., Ltd.
Crystal Tower 2-27 Shiromi 1-Chome, Chuo-ku, Osaka 540, Japan

This paper presents a technique to create panoramic stereo images easily and to display objects normally in the panoramic stereo images. Significant points of the technique are that we use only one video camera and use depth information on rendering. To create panoramic stereo images, this technique cuts and connects vertical slit images with optical flow between recording original images. Moreover, to display objects rearranged normal position, this technique controls an mixing angle of left and right panoramic images.

1. Introduction

In recent years, a number of research papers have been presented on techniques of building virtual spaces based on photographic images [1][2]. Panoramic images represents one of these techniques.[3][4] In the case of a 360-degree panoramic image, a single panoramic image can be generated from images in all direction with the photographing position at the center. However, most generation of panoramic images requires that the camera be rotated while maintaining a precise angular speed. This limitation exists because a slit width of fixed size is set in advance corresponding to the rotation angle speed of the camera, images are excised as vertical slits having that width, and are sequentially combined to generate the panoramic image. Furthermore, although a standard panoramic image enables one to see in all directions, it remains a 2-dimensional image with no sense of depth reproduced on a flat (cylinder) surface, and there is a problem with respect to the sense of realism.

This paper presents a technique for generating panoramic images with few restrictions on the picture-taking time and capable of being viewed in binocular stereo. Our technique consists generally of the following steps. First, a single video camera is used to capture continuous images. The optical flow between frames of the captured images is detected, and the slit width is determined based on the optical flow detection results. By excising the slit images from the frame images and combining those images, the left and right panoramic images are generated. In addition, a description is provided of stereo viewing of the panoramic image using parallax that corresponds to depth. The following sections describe the technique in detail, following the sequence of generated display.

2. Image Capture

A video camera is attached to a tripod, and the camera is rotated counterclockwise smoothly with the rotating axle of the tripod as the center to record images. At this time, the focal point of the video camera is set to a position a distance L from the rotating axle. In addition, there is no need to keep the rotation speed of the video camera constant, and the camera is rotated manually at an appropriate speed.

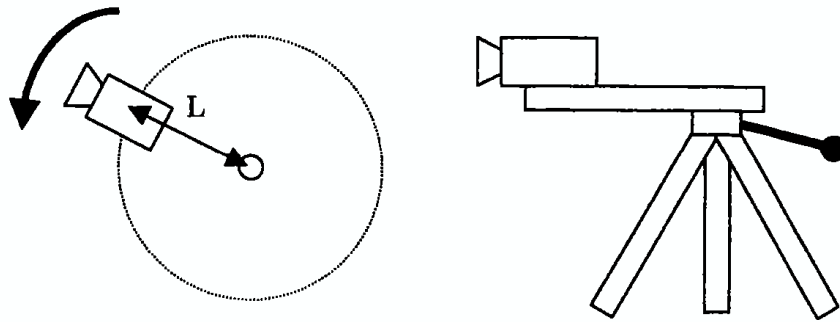


Fig. 1: Image Capture Technique Using a Video Camera

3. Detecting the Optical Flow

Frame images are captured from the recorded video images, and the optical flow between the next frame image and the captured frame image (320x240 (pixels)) (see Fig. 2) is calculated. (b) of the same figure presents the result of detecting optical flow with the next frame image using template matching. The size of the flow vector at each representative point in the template used

for matching is expressed as the length of a line segment with the representative point as the starting point. The search range for template matching extends only horizontally, since the movement of the camera is a rotating motion around a fixed axis. As a result, only horizontal line segments appear in the detection results. Also, areas where no optical flow line segments are present are uncertain areas where the difference between the value of the matching result for the 1st candidate and the value of the matching result for the 2nd candidate was not at least a certain set value.

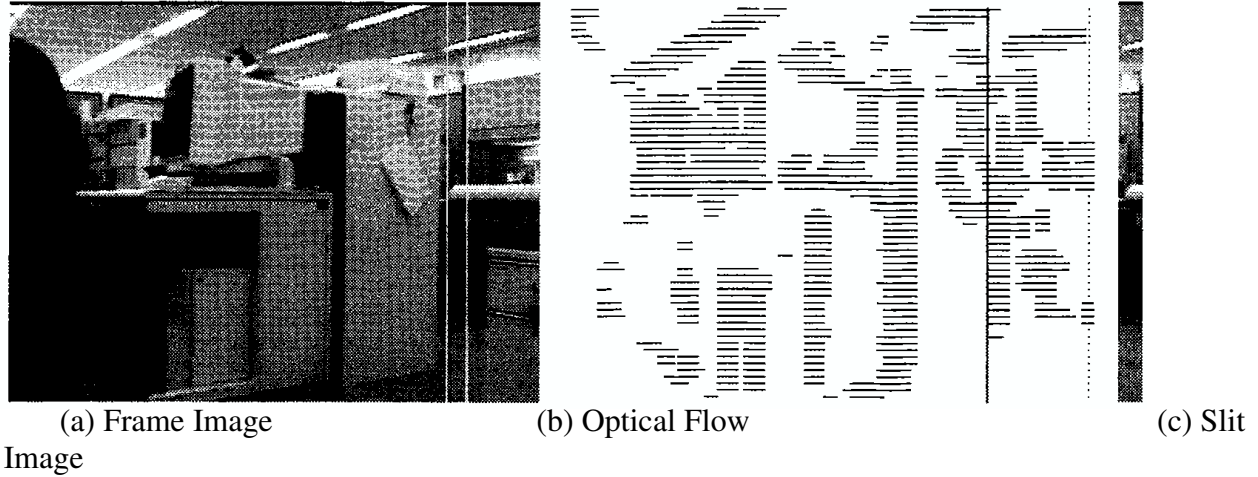


Fig. 2: Optical Flow Detection Result and Slit Image

4. Excising the Slit Images

Once the optical flow is determined between all the continuous frame images, the size of the flow vector is used to set the slit width, and the respective right eye and left eye slit images are excised from the frame images. Here, the flow vector size can be used as the slit width because template matching performs matching in pixel units and the limitation that the search range is in the horizontal direction only.

First, we derive the positions at which the left eye and right eye slit images are to be excised. In Fig. 3, the left and right camera positions correspond to the left and right eye positions when viewing a panoramic image stereoscopically. If the video camera rotation radius is represented as $L(m)$, the sight line (eye) interval as $2e(m)$, and the video camera rotation angle as $2\theta(deg)$, the following relationship obtains.

$$L \sin\theta = e$$

For example if $L = 0.17(m)$ and $e = 0.03(m)$, the result is $\theta = 10.2(deg)$. However, based on Fig. 4, if the distance from the center of the screen to the slit position is represented as $x(pixels)$, the window width as $w(pixels)$, and the horizontal field of vision angle as $fov_x(deg)$, the result is as follows:

$$w/2 : x = \tan(fov_x/2) : \tan\theta$$

For example, if $w = 320(pixels)$, $\theta = 10.2(deg)$, and $fov_x = 31(deg)$, the result is $x = 104(pixels)$. Consequently, an image with a width corresponding to the flow vector is excised from a position 104 (pixels) horizontally from the center position of the image, thus generating the slit image.

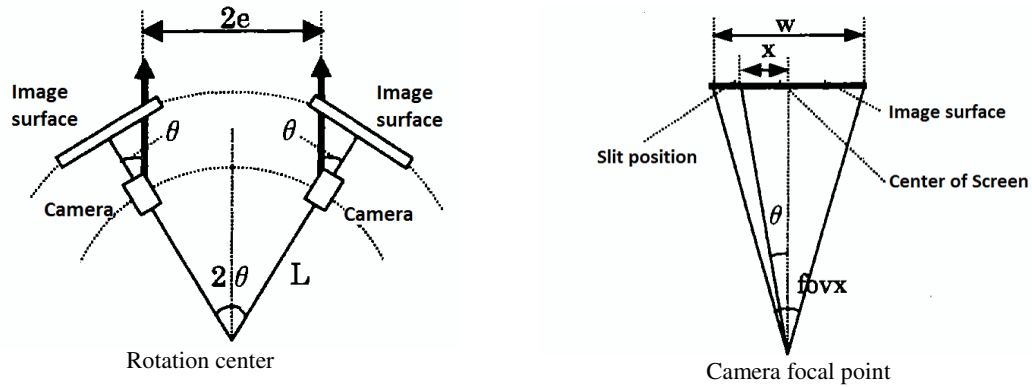


Fig. 3: Relationship Between Rotation Radius and Rotation Angle

Fig. 4: Relationship Between Rotation Angle and Slit Position

Next, we determine the width of the slit image to be excised. The flow vector is detected only as the number of the line of the representative point in the excision position for the slit. We used the mode value for the size of the flow vector as the slit width. Using the mode value signifies that an object at the position corresponding to the size of the flow vector occupies most of the area within the slit, and attention will be focused on that object when the panoramic image is generated. Consequently, we take a majority vote of the valid flow vector values at the previously derived slit position (items representing horizontal line segments in Fig. 2(b)), and the most frequently detected value is set as the slit width sw . The slit image that is ultimate excised is a slit image of width sw from the position $w/2+x$ for the left eye and $w/2-x-sw$ for the right eye. Fig. 2(c) shows the excised left eye slit image.

5. Compositing the Slit Images

All of the slit images excised from the frame images are continuously composited in sequence. Fig. 5 shows left and right panoramic images created from images taken of an elevator hallway. The upper row is the left eye image, and the lower row is the right eye image.

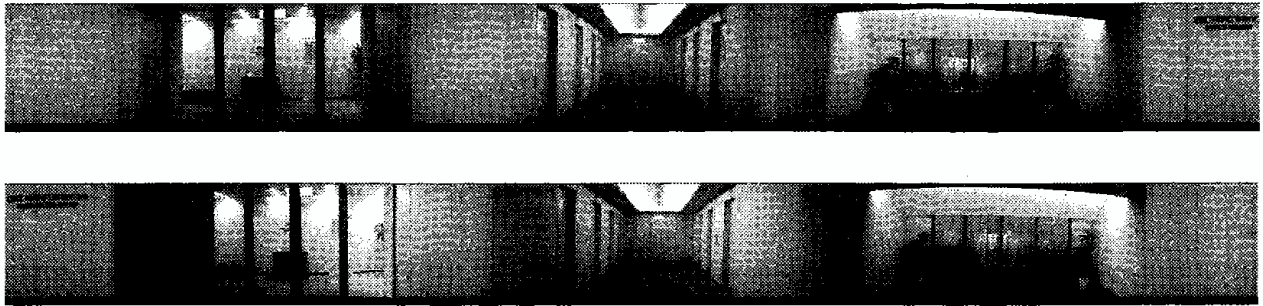


Fig. 5: Panoramic Images of an Elevator Hallway

6. Stereoscopic Viewing Using Depth Parallax Angle

When the left and right panoramic images obtained using the foregoing procedure are viewed binocularly, a stereoscopic view is possible that faithfully reproduces the positional relationships, if the image was captured from a sufficient distance. However, if the

camera was placed at a comparatively close distance, or if the distance from the camera to the objects varies greatly, the positions representing the left and right panoramic images must be adjusted. The reason why this process is necessary can be explained as follows. Fig. 6(a) shows the appearance of the imaging of objects A and B at different distances while rotating on a circle perimeter R with the axis of rotation O at the center. At this time, object A is recorded for the right eye and the left eye, respectively, at positions C and D in the panoramic image. Similarly, object B is recorded at positions E and F in the panoramic image. However, if the panoramic image is viewed stereoscopically as-is, since θ is the parallax angle calculated from parallel light beams, using a captured image in which the objects were placed at a close distance, object A and object B are displayed as infinitely distant. Therefore, objects appear to overlap or some other fault, making faithful stereoscopic viewing impossible.

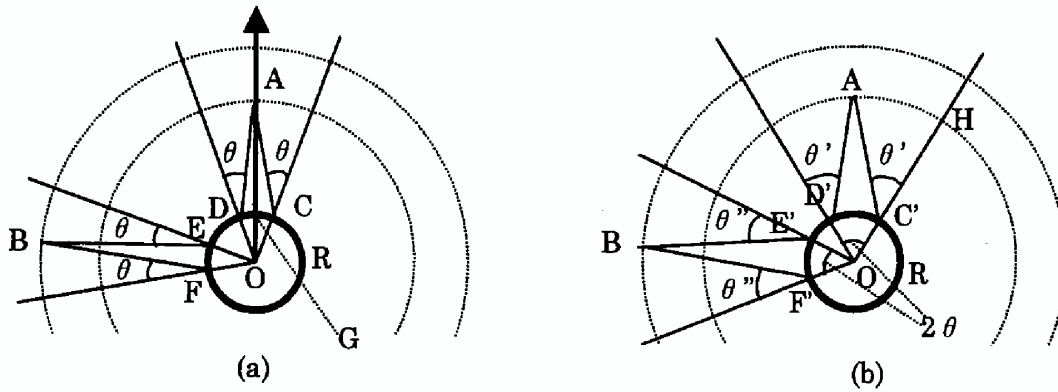


Fig. 6: Explanation of Depth Parallax Angle

Because of this, it is necessary to make corrections so that displayed objects will appear to be at the correct position by using parallax angles corresponding to the distances to object A and object B, and dynamically changing the panoramic image display position.

As shown in Fig. 6(b), when $\angle C'OD'$ and $\angle E'OF'$ are equalized to 2θ , the relationship between θ' and θ'' in the figure is $\theta < \theta'' < \theta'$. In other words, as the position of the object draws nearer in relation to the parallax angle θ calculated based on parallel light beams, the parallax angle increases, and object A and object B must be seen at parallax angles θ' and θ'' . Consequently, rotating the left and right panoramic images by only the parallax angle θ' or θ'' with the respective viewpoint positions at the center enables accurate stereoscopic viewing of object A and object B. In this paper, since the parallax is based on the distance to an object (depth) captured by a camera at this parallax angle, we refer to it as the depth parallax angle.

Fig. 7 presents an example of rotating the left and right panoramic images for display according to the depth parallax angle. In Fig. 6(a), assuming that the camera began rotation for capture from position G (this position is referred to as the reference position for the panoramic image), in the left eye panoramic image, object A is recorded to the left by an angle of only θ from the reference position (position D). For its part, in the right eye panoramic image, object A is recorded to the right by an angle of only θ from the reference position (position C). In addition, by rotating the left eye panoramic image clockwise by θ' and the right eye panoramic image counterclockwise by θ' with the viewpoint as the center, the recorded object A can be viewed as located at the correct position.

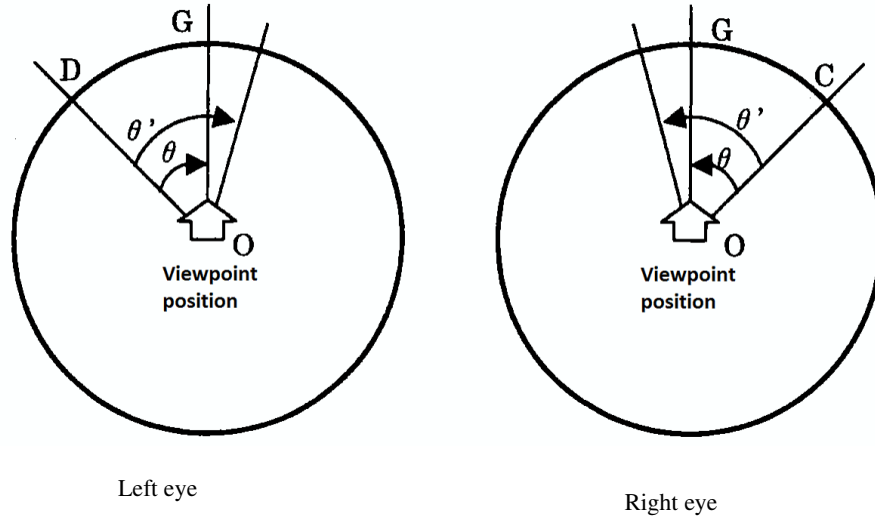


Fig. 7: Rotation of the Left and Right Panoramic Images According to the Depth Parallax Angle

Next we show the relationship between the distance to the object and the depth parallax angle. In Fig. 6(b), if the distance from the rotation axis to the object $AO = a$, the rotation radius $CO = L$, the depth parallax angle in relation to parallel light beams $\angle COA = \theta$, and the depth parallax angle in relation to object A $\angle HC'A = \theta'$, the following relationship obtains.

$$a = L \sin \theta' / \sin (\theta' - \theta)$$

When this equation is modified, it becomes:

$$\tan \theta' = \sin \theta / (\cos \theta - L / a)$$

thus yielding the depth parallax angle θ' corresponding to the distance to the object. By recording this depth parallax angle in correspondence to the panoramic image and applying it to the left and right panoramic images, alignment of the left and right panoramic images can be dynamically controlled, making stereoscopic viewing of the panoramic image possible.

7. Field Test

A field test was conducted applying these techniques to panoramic images of an elevator hallway in which the distance to objects varies greatly. First, while actually looking at the panoramic images, alignment was performed in several sight line directions so faithful stereoscopic viewing would be possible, and the depth parallax angle in each sight line direction was recorded. In addition, the depth parallax angle obtained was continuously varied using linear interpolation, the depth parallax angle for all sight line directions was calculated, and correspondences were made with the panoramic images. As a result of stereoscopic viewing with alignment control of the panoramic images using the calculated depth parallax angles with 10 research personnel, there were no noticeable double images in the objects attended to, and the sense of depth was faithfully reproduced.

8. Summary

This paper has discussed a technique for generating panoramic images for stereoscopic viewing and a technique for displaying panoramic images. Using these techniques, it is possible

to generate panoramic images with a sense of realism and view them stereoscopically with few limitations during image capture and without using special equipment. This time we actually stereoscopically viewed a panoramic image in several slight line directions, made adjustments so that objects at those slight line directions would appear normal, and then observed the corresponding depth parallax angles. In addition, we calculated depth parallax angles for all sight line directions using linear interpolation, based on the obtained representative parallax angles, and made correspondences with the panoramic images. However, we believe it is possible to automatically derive the depth parallax angle using the relationship between the size of the flow vector obtained from optical flow detection and the distance from the camera to the object, and the correlation between the generated left and right panoramic images. Furthermore, we wish to raise as a challenge for the future not only stereoscopic viewing from the point of image capture, but also stereoscopic viewing of panoramic images during horizontal movement.

References

- [1] Uchiyama Shinji, Katayama Akihiro, Tamura Hideyuki, et al., "Realizing Virtual Environments Blending Still Image Data and CG Models Based on Ray Space Theory," 3-Dimensional Imaging Conference '96 Conference Papers, pp. 13-18, Oct. 1996.
- [2] Leonard McMillan and Gray Bishop, "Plenoptic Modeling: An Image-Based Rendering System," SIGGRAPH95, pp.35-46, 1995.
- [3] Hiroshi Ishiguro, Masashi Yamamoto, and Saburo Tsuji, "Omni-Directional Stereo," IEEE Trans.PAMI, Vol.14, No.2, pp.257-262, Feb 1992.
- [4] Tsuji Saburo, "Environment Panorama," Journal of the Institute of Electronics Information and Communication Engineers, Vol. 74, No.4, pp.354-359, Apr 1991.

SIG-CyberSpace, VCR 97-13
November 27, 1997

A Visual Walkthrough System Placing Real World Video on CG Space

Tamio Kihara*, Go Nishimura*, Kazuaki Nakakura*

*NTT Information and Communication Systems Laboratories

Newly developed visual presentation system provides visual walkthrough to users that coupled real world's video with 3D-CG. The system has an authoring subsystem to make the walkthrough-path and its video linkage on a CG space and has a player provides the presentation can switch the video-walkthrough and the realtime 3D-CG walkthrough. We proposed an architecture of the software and a structure of space scenarios.

A Visual Walkthrough System Placing Real World Video on CG Space

Tamio Kihara*, Go Nishimura*, Kazuaki Nakakura*

*NTT Information and Communication Systems Laboratories

Newly developed visual presentation system provides visual walkthrough to users that coupled real world's video with 3D-CG. The system has an authoring subsystem to make the walkthrough-path and its video-linkage on a CG space and has a player provides the presentation can switch the video-walkthrough and the realtime 3D-CG walkthrough. We proposed an architecture of the software and a structure of space scenarios.

1. Introduction

A hypermedia structure of videos has been proposed as a means to provide information based on a real-world video image.^[1,2] A system has been constructed that enables writing as a correlation between the video image and other media while also enabling the writing of a hyperlink structure from display of the video image.

Additionally, a method of integrating images has been proposed in which (1) a video walkthrough can be easily realized by placing the real-world video image into a CG space and (2) the real-world video image and a 3DCG image of real-time production can be selectively exchanged, and an information service system having this method has been constructed.^[3,4]

In this article, (1) a structure of space scenarios by a node-path model and (2) a sequential interpretation of the space scenarios, which serve as the basis for realizing the image walkthrough in the system, are made clear, and a specific implementation is described.

2. System Characteristics

The objects, characteristics and approach of the present system are explained in the following description.

2.1 Video walkthrough

A moving path in a street and in a building is continuously recorded in a real-world video, using the perspective and manner in which a person would observe them while walking or while moving by car. It is recorded so that the real-world video image is replayed as seen when the moving object turned to the left or to the right, following an instruction of turning to the left or the right respectively at an intersection, for example. In the video walkthrough, images are provided in a manner that is the same as if one continuously moves along the above-described path.

A video of real-world footage contains rich information. The video walkthrough can give the user a sense that he or she is actually going to a particular place, and that he or she has seen and heard various things. There is inconvenience in terms of spatial movement, however, as only the path that was filmed can be replayed. However, the usefulness of rich information cannot be ignored, particularly that of real-world video.

2.2 Integration with the 3D-CG image

The provision of the walkthrough image by the 3D-CG of real-time production has been generalized, however. Its power of expression is progressing steadily, and the advantage is that free movement of the space and movement of one's viewpoint are possible. Nevertheless, the expression still lags behind that of the real-world video image when the actual world is considered to be an object of expression. Therefore, a system has been constructed to facilitate the effective provision of information when the visual walkthrough by the real-world video and the visual walkthrough by the 3DCG of real-time production are integrated in a mutually complementary manner. The characteristics of the visual provision of the present system are as follows:

- (1) The video walkthrough is realized, and an exclusive interaction function is provided.
- (2) The video walkthrough and the 3D-CG walkthrough can be selectively switched to each other.
- (3) The spatial coordinates and the view can be matched when they are selectively switched to each other.

Basically, the system provides a video walkthrough that can be selectively branched out at the intersection as if the user is moving from place to place on the street. This is only when the user is moving along the filmed path. The walkthrough image by the 3D-CG is provided for the places where there is no real-world video to enable free spatial movement. That is, the space by the real-world video and the space by the 3D-CG are constructed at the same time as a parallel world, and the paths other than those having the real-world video image are complemented by the 3D-CG image of real-time production.

The user can selectively switch the image of the real-world video and that of the 3D-CG at any given moment within the same window (Fig. 1). When the user goes from the path of the real-world video to the CG space or vice versa, the spatial coordinates and the view (appearance) are matched to each other. That is, the movement when the user moves along the path of the real-world video and the movement in the 3D-CG space are reflected to each other.

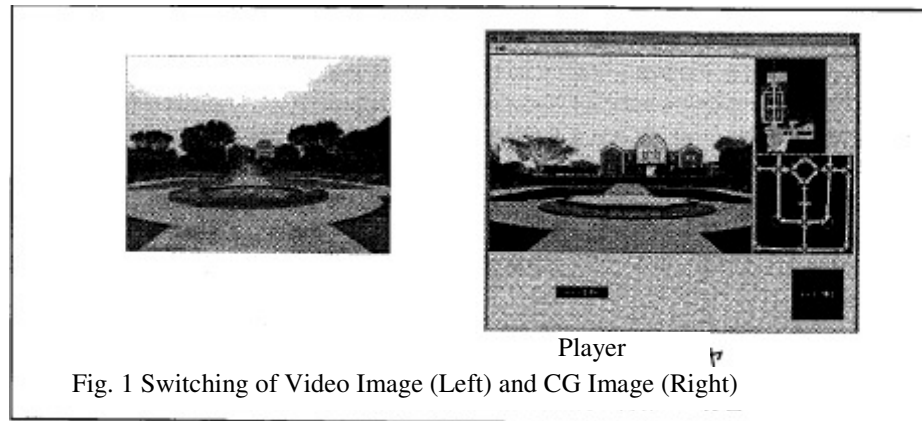


Fig. 1 Switching of Video Image (Left) and CG Image (Right)

2.3 Approach

Research has been performed in which a large number of still images and images of the real-world video were three-dimensionally mapped in the CG space to accurately emulate the actual world. Contrastingly, the approach of the present system is that the images of the path movement that are filmed by a video camera are replayed and provided as a video stream without special treatment.

This method has the following advantages:

(1) The filming of real-world video is simple.

The video recording of the path is performed by moving at nearly equal speed, but the data collection of camera information per frame, positional information, etc., is not performed. A small video camera for home use is sufficient.

(2) Easy authoring is realized.

Handling is also easy because the large amount of calculation processing, etc., to the real-world video is unnecessary. Authoring in this case does not require special knowledge.

(3) Production such as for a TV program is permitted.

Because the real-world video is not mapped as a still image, production such as for a TV program is permitted for the content of individual video images. A person can talk in the video, and the change of things can be captured in the video.

3. System Configuration

The system configuration is shown in Fig. 2. As a whole, the system is configured with a player that provides the image to the user along with an authoring tool and content. The content is configured with the following files:

(1) 3D-CG model file

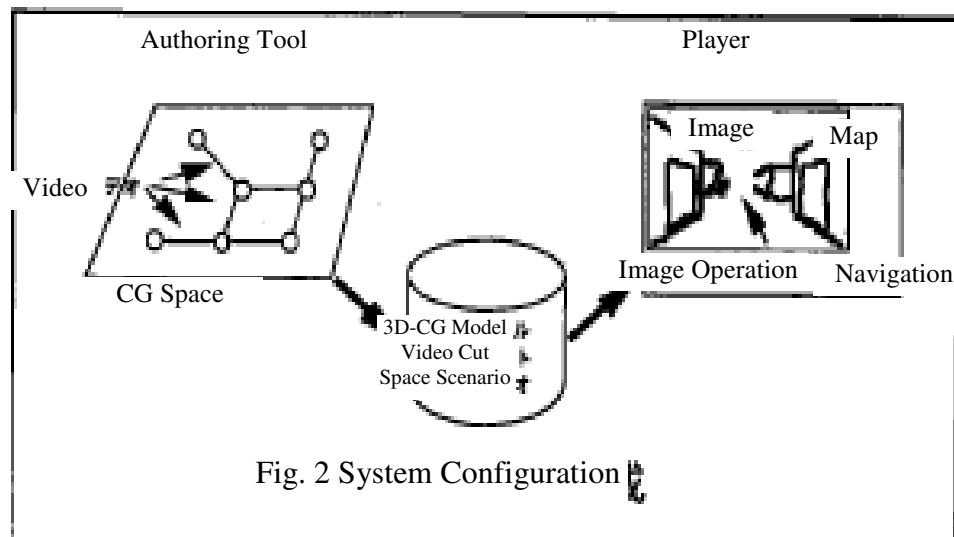
This is a 3DCG model file in which the CG space is written. It is produced with a general construction tool. An OpenInventor form is used.

(2) Video cut file

This is a group of files of the video cut in which the real-world video image is fragmentized. It is produced with a general editing tool.

(3) Space scenario file

This is a file consisting of the structural information peculiar to the present system. The authoring tool produces only this file but does not revise the 3D-CG model file or the video cut file.

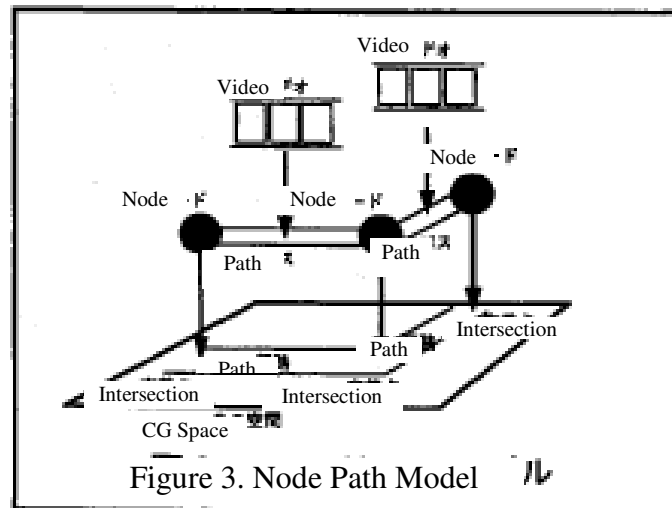


4. Placement of Real-World Video to the CG Space

4.1 Realization of the video walkthrough

It is necessary, in realizing the video walkthrough, to control (1) the connection and order relationship of the video cut in which each path was filmed as well as (2) the direction of each of multiple video cuts that follow at the branch. A large amount of video cuts have a lattice-like relationship. When the user's instruction is done by a word such as "right" and "left," a word attribute can be given to the connection and order relationship. That is, it can be said that geographical position information based on specific values is not necessary when the relationship between paths can be distinguished. However, the method of producing the CG space and then pasting the video cut into it is introduced in the present system. The real-world video is placed on

the CG space to derive the necessary relationship one by one from it. The video walkthrough is considered to be a movement in the path that is fixed to the CG space.



4.2 Node-Path Model

The path of the video walkthrough in the present system is defined by arranging the path, which is an object showing the path, and a node, which is an object showing the branch in the CG space. The video cut is then coupled to it. This structure is called a node-path model (Fig. 3). These node-path models and supplementary information are stored as the space scenario. The CG space is used as a map for placement as it is. It is handled normally as a plane looking from above.

The following order of the authoring procedure is used to produce the space scenario (Fig. 4):

- (1) The nodes are arranged on the plane according to the materials when the real-world video is filmed.
- (2) The paths are arranged between the nodes.
- (3) The video cuts that correspond to the node and path are coupled.

Expressions of the node and path are shown as follows:

(1) Node

This expresses the intersection. It contains short video cuts of all possible paths such as going straight, making right and left turns, and rotating in an intersection (Fig. 5). The coordinate of the node in the CG space is restricted by the initial frame and the final frame of the video cut. That is, the video cuts are associated to each other when the end point is restricted by the node in the CG space.

(2) Path

This defines a path that interconnects the nodes. The video cuts of the walkthrough image along this path are associated. It contains the video cuts of a round trip. In the present system, the continuous coordinates of the path in the CG space are not controlled. The coordinates in the middle of the path can be derived by interpolating the coordinates of the node at both ends.

The node and path are objects of the 3D-CG during authoring. They make the structure of the path of the video walkthrough easy to understand, and they serve to enhance visibility (Fig. 6). They are not recognized as objects in the player. With this method easy, intuitive authoring can be realized without the need for special knowledge.

4.3 Structure of the Space Scenario

The space scenario is basically configured with the following:

- (1) Geometrical information as the 3DCG model of the node and the path,
- (2) A recognition name of the node at both ends of each path in the entry, and
- (3) A file name of the video cut that is associated in the entry of each node and each path.

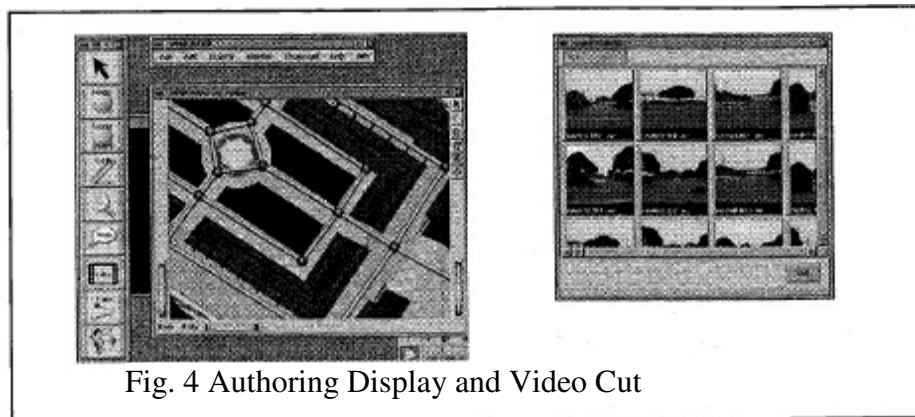


Fig. 4 Authoring Display and Video Cut

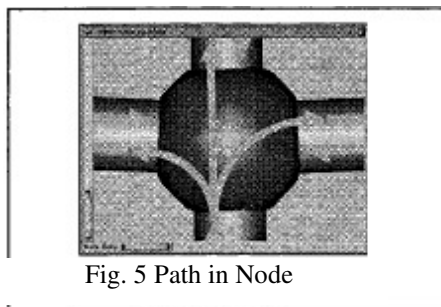


Fig. 5 Path in Node

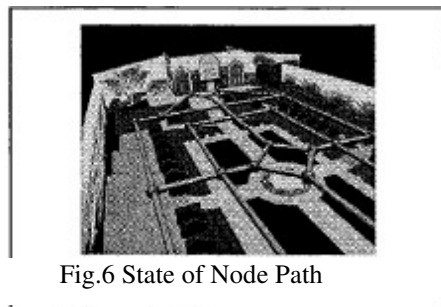


Fig.6 State of Node Path

Illumination information, viewpoint information, link information when the scenario has a hyperlink structure, etc., are also included.

The following processes are performed one by one based on the space scenario when the image is replayed:

- (1) Interaction (branch selection) of the video walkthrough;
- (2) Matching of the spatial coordinate and the view when the user selectively switches between the video walkthrough and the 3D-CG walkthrough.

5. One-by-One Interpretation of the Space Scenario

5.1 Interaction of the video walkthrough

The following procedures are repeated in the video walkthrough:

- (1) When the replay of the video cut associated with a given path is finished, it is stopped at the final end node.
- (2) All nodes in which the position of the node and the geometrical information are matched are searched, and the paths having this node as a beginning point are obtained.
- (3) The user is encouraged to select the path by means of a schematic icon showing the direction of each branch, which is obtained from the geometrical information of an angle with the node as the center.
- (4) The corresponding short video cut in the node is replayed.
- (5) The following video cut is replayed successively.

Thus the mutual relationship of the video cut for the video walkthrough can be indirectly derived by interpreting the space scenario one by one.

During authoring, the node and path are arranged in the CG space, and the video cut is associated to each object. It is not necessary to clearly describe the relationship of interconnection between the video cuts. With the node-path model, the transitional relationship of the connection and the direction indication for selection can be realized by placing the real-world video in the CG space.

5.2 Flip of the real-world video and the 3D-CG

The video walkthrough and the 3D-CG walkthrough can be selectively exchanged as instructed by the user. At this time, the matching of the spatial coordinate and the view is performed. This switching is called an image flip.

- (1) Method of matching the coordinates

An approximation method is used on the assumption that filming has been performed with a video camera moving at nearly equal speed.

The ratio of the path length is used to match the views when going to the CG space when the real-world video is replayed and when entering the path of the real-world video from the CG space. For example, when entering the CG space at such a point that 75% of the total number of frames is passed when the video cut is replayed, it goes to the coordinate at 0.75 of a line segment of the path (Fig. 6). When the coordinate entering from the CG space is at 0.75 of the line segment of the path, the video is replayed from a point at which three-fourths of the total number of frames of the video cut are advanced (Fig. 7). With this, the views nearly match each other. In the present system, matching that is stricter than this is not desired.

(2) Method of adjusting the view

The view is adjusted by parameters to match the height of the viewpoint and the picture angle of the video camera, which differ from the 3D-CG view according to the filming condition of the video, etc.

(3) Selection of the adjacent path

When moving to the CG space, the adjacent position where the real-world video exists is displayed and the user can enter the path by selecting it. At this time, the distance to the path can be calculated so that the user can automatically enter a closer path. The user can enter by traveling around the path, through which contact is smooth in the direction of travel. The direction in which the traveling image is replayed at a particular path is determined by the angle to the path upon entry to the path.

(4) Size of the intersection

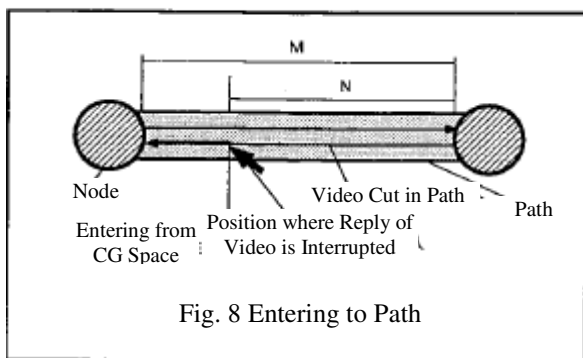
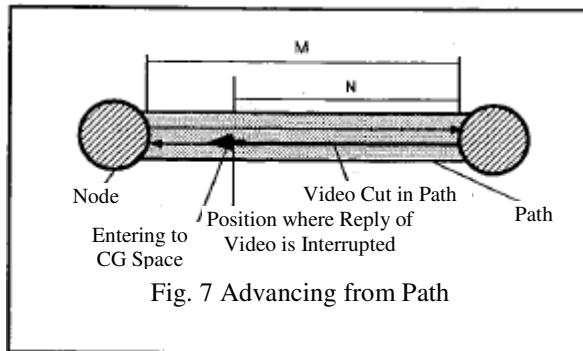
The video cut in the node is a short real-world video of a right turn, a left turn, etc. However, it has a spatial movement as a matter of course. That is, a node has a certain size. Although it depends on the method of filming, the size of the node is determined by the size of the intersection, the width of the road, etc. The size (radius) of the node can be set for every node in the space scenario. It normally has a default value. A calculation is performed by considering the node radius to increase the matching of the spatial coordinate and the view in the vicinity of the node.

6. Summary

A method of integrating images in which a video walkthrough can be easily realized by placing a real-world video image on a CG space, and the real-world video image and the 3D-CG image of real-time production can be selectively switched to each other, and a construction of a system for providing information with this method are described. Particularly, it details the structure of the space scenario by the node-path model and a method of interpreting the structure and a specific method of implementation.

According to the present system, it is possible to realize a spatial accumulation and reconfiguration with emphasis on the real-world video. It is possible to accumulate a fun, interesting space and to create something that does not exist in reality as an object in the 3D-CG space by cutting and pasting according to a theme, making spaces that are not adjacent to each other in reality appear to be adjacent, etc. The present system is intended for application to

objects in which a large amount of information peculiar to the real-world video is effective, such as tourist information and area guidance, as a walkthrough type of information guide system.



References

- [1] Testuo Sakata, Tamio Kihara, Akira Kojima, Tetsushi Sato, "Proposal of Video Hyper Model for Image Walk," Technical Report of IEICE, DE95-35, pp. 65-72, July 1995.
- [2] Yoshikatsu Kishida, Tamio Hihara, Yasuhiro Hirano, Akira Iwabuchi, Katsumi Teranaka, "Application of Video Hyper Media to the 'TakeMe' Tourist Information System," Research Report of IPSJ, 96-DPS, pp. 55-60, May 1996.
- [3] Tamio Kihar, Yoshikatsu Kishida, Tadashi Shibagaki, Katsumi Teranaka, "Multimedia Presentation System Merging Video Walkthrough and Virtual Reality," Research Report of IPSJ, 97-DPS-80, pp. 163, Jan. 1997.
- [4] Kazuaki Nakakura, Yoshikatsu Kishida, Tmio Hihara, "Visual Walkthrough System Integrating Real-world video and VR," Proceeding of the 55th General Conference of IPSJ, 3-448, Sept. 1997.

SIG-CyberSpace, VCR 97-14
November 27, 1997

A Way to Represent the Actual Action in Virtual Space

Shinkuro Honda, Takaaki Kimura, Takaharu Ohsawa,
Kenji Ohta, Ken-ichi Okada, Yutaka Matsushita
Faculty of Science and Technology, Keio University
E-mail: Honda@myo.inst.keio.ac.jp

In this paper, we pay attention to the human action and describe a way to solve the gap problem between the real world and virtual space. We try to support high quality communication among the members and provide triggers for informal communication by letting the avator have gestures for the member's information in a virtual space. We use a virtual office system "Valentine" for 3D virtual space which we have built before, and have built a communication system for office work. First of all, we pay attention to handshake and realize "virtual handshake device" for the way to make users feel virtual action. Secondly, by applying the spotting recognition of gestures from time-varying image, the avator reflects user's state and action.

A Way to Represent the Actual Action in Virtual Space

Shinkuro Honda, Takaaki Kimura, Takaharu Ohsawa,
Kenji Ohta, Ken-ichi Okada, Yutaka Matsushita
Faculty of Science and Technology, Keio University
E-mail: honda@myo.inst.keio.ac.jp

In this paper, we pay attention to the human action and describe a way to solve the gap problem between the real world and the virtual space. We try to support high quality communication among the members and provide triggers for informal communication by letting the avator have gestures for the member's information in a virtual space. We use a virtual office system "Valentine" for 3D virtual space which we have built before, and have built a communication system for office work. First of all, we pay attention to a handshake and realize "virtual handshake device" for the way to make users feel virtual action. Secondly, by applying the spotting recognition of gestures from time-varying image, the avator reflects user's state and action.

1. Introduction

Because of recent improvements in the personal computers in terms of speed and 3D graphic display capabilities, there are many applications available on the market that make use of the virtual space created by the 3 dimensional computer graphics. Among those, attracting attentions are those creating the virtual community that virtually connects people who are physically located apart and supports communication between them^{[1][8]}. These are called avatars and are graphic representation of users in the virtual space.

In order to realize high quality collaboration among users, it is necessary to have close communication enabling mutual understanding. Communication is the process of exchanging opinions and information. There are verbal information that can be expressed by characters and

non-verbal information that is expressed by facial expression, tone of voice, gesture, etc. Some researchers report that verbal information occupies less than 20% of the total communication information. This means that the transmission of non-verbal information is essential in achieving high quality communication.

If parties are facing each other, non-verbal information can be conveyed from one to another, and the location and time are the same for all. If it is a telephone conversation, non-verbal information would decrease considerably but at least the voice is heard. Even in the letter, the stroke of the pen or the writing style would give some non-verbal information. However, the computer communication typified by E-mail or news basically exchanges information by displaying characters and non-verbal information is lacking. Most current applications that utilize virtual space would provide voice and inter-personal distance. In the application such as InterSpace [9] that uses time-varying image, it is possible to obtain limited information such as eye movement but it is difficult to grasp information on body movement, posture or exterior appearance because the avatar is not close to a human shape. On the other hand, in the application (such as still image and CG) that does not use time-varying image, it is easier to get information on body movement, posture, or exterior appearance only because the shape of the avatar is closer to the human shape. In these applications, the avatar can be customized for own liking, and in the Community Place [2], it is supported by body actions, etc. However, it is not an exaggeration to say that currently the avatar is used only for identifying who it is.

Therefore in this paper, in order to make the activities in the virtual space more realistic by providing more non-verbal information, the avatar for each user is provided with gestures that would provide triggers for information communication and would support high quality communication.

2. Virtual Office Valentine

We have been performing a research on providing the virtual office for people working from home [7][3]. The Valentine system, which is created by authors of this paper, creates the virtual office on the network and let the geographically distributed users to report to the office. The system can convey the atmosphere and presence of members to each other and support improved communication.

It is assumed that there is no physical office existing and all users are reporting to the virtual office. In order to convey the atmosphere and presence of other office members located at remote areas, the “surrounding view” and “effective sound” are implemented, and as a result awareness is provided. On the other hand, if excessive awareness information is provided, it would interfere with efficiency of individual performance; therefore the “degree of concentration” of each individual is defined and the awareness information appropriate for the individual degree of concentration is implemented.

Valentine supports the information communication between distributed user members, but its communication is basically based on the voice and image communication. It could not support making physical contacts such as shaking hands or massaging shoulders. Also, the function of

3.1 Sense-of-Touch Information

3.1.1 Handshakes

Among many methods of expression, “Touch” is very first to appear in human history and would be used for many years to come. “Touch” is commonly observed in individual level of behavior as well as in behavior of the entire human beings, and is the most familiar signaling system. The center of the human behavior of touch is a handshake. This is typical as well as most important touch behavior. The handshake is the standard as well as tradition for greetings internationally [4].

Let’s analyze how a body-contact is used in the office. Patting shoulder is often used by a superior to compliment or encourage a subordinate, but not the other way around. A handshake is often exchanged at the beginning of the meeting gathering at one place greeting each other. When meeting a person for the first time, a handshake is used for acknowledgement. A handshake is made at the conclusion of the negotiation.

In this paper, we pay attention to the behavior of a handshake. Let’s suppose that when shaking hands on the virtual space, if the process is to select the box called “handshake” from the menu and click it for letting avatars to automatically shake hands, real users would not feel the action is realistic since users did not shake hands at all. The user does not feel a real action taking place because the action of the avatar on the virtual space does not match the action of the user himself.

The handshake carries out the important role of verifying the mutual trust by actually touching each other while looking at each other in the eyes. In addition to ascertaining mutual trust of people shaking hands, it is also meaningful to those watching the action. For example, if heads of states shake each other’s hands at the international conference, it has a significant meaning to people surrounding the event.

3.1.2 Device

We conceived the idea of providing very realistic virtual handshake using the simple device as shown in Fig. 2.

The device consists of the Hand section and the Face section. The Hand device simulates the hand of the other party, and this is gripped by the user when he makes a handshake. The Face section displays the upper body of the avatar of the other party, and is constructed as part of the virtual office. Sensors are incorporated within the Hand section. These sensors can recognize when the Hand section is gripped and how strong the grip is. When making a handshake, the avatar in the virtual space extends the hand. When the hands of both parties are extended, the handshake can be performed. Both parties can exchange the handshake while looking at each other through the avatars. In this way, both parties feel that they actually made a handshake. People on the network can witness the handshake by looking at avatars shaking hands. Fig. 3 shows the scene in which two avatars are shaking hands in the virtual office.

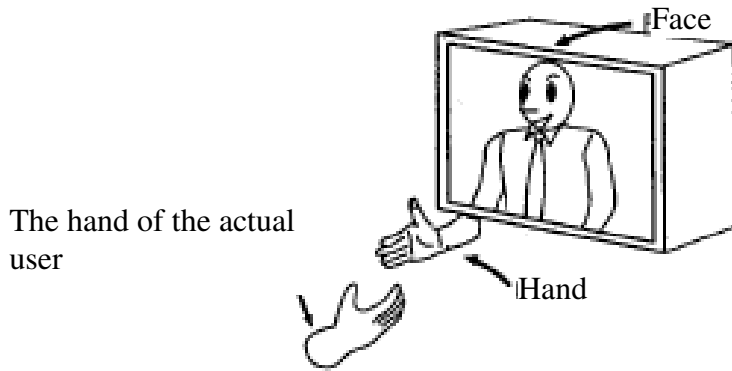


Figure 2 Handshake Device

The features of the handshake device are as follows:

- The virtual action (avatar action) matches the actual action.
- It can give a special impression in greeting between remotely separated locations as if “I shook hands with him before”.
- It can achieve a role of symbolizing the successful completion of the contract.

3.1.3 Implementation

The Hand device is formed with silicon into the hand shape and is embedded with touch sensors. The information of shaking hands is sent from the touch sensors to the workstation operated by the virtual office (SGIO₂), and will be reflected on the avatar behavior. The pressure sensor is also embedded in the Hand section, and the information on how strong the grip is forwarded to the other Hand section at the other remote location for reflecting the grip strength. Each party of the handshake can actually feel how strong the other party is gripping through these Hand sections.

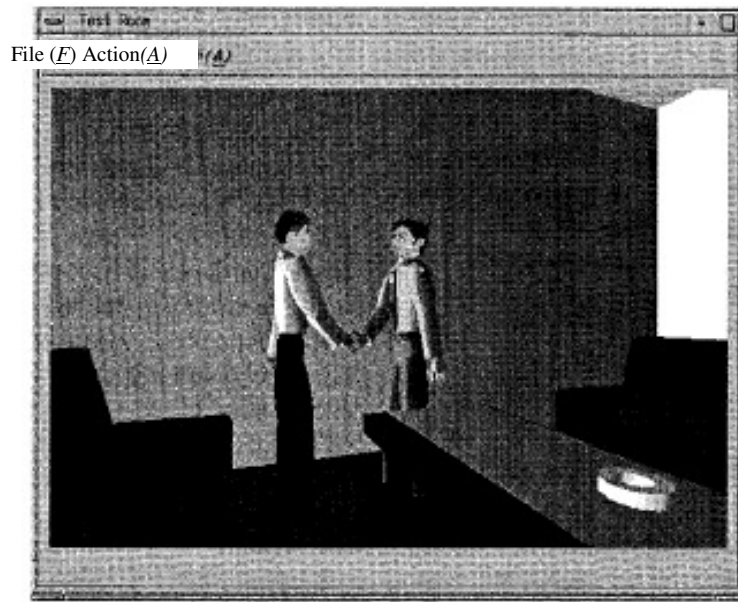


Figure 3 Virtual Handshake

3.2 Non-verbal Information (Gestures)

Let's review how the previously described non-verbal information is conveyed in the current virtual office (Valentine). In the current Valentine, the voice information of each member is multicast and sound localization is set with respect to each member's location. Therefore, each user can recognize inter-personal distance and voice.

The facial expression or the movement of eyes cannot be obtained when the still picture is used displaying the face of a normal user (or users). However, the transmission of the time-varying image is supported when two members face each other, and such non-verbal information can be provided only in this case.

The information on touch can be supported by the Handshake device which is being created in this fiscal year. On the other hand, currently there is not enough information contained for physical movement, posture, or external appearance. But it is possible to convey these information if changes (or movement) are incorporated in the avatar depiction.

Therefore in this research, we pay particular attention to non-verbal information that is feasible such as physical movement, posture, human "gestures" or "manner" even though these are currently not transmitted. By conveying these information through the avatar, we are attempting to provide a trigger for informal communication and higher quality communication in the virtual office.

3.2.1 Implementation Method

As a feature of the non-verbal information, it can be pointed out that this information is spontaneous. That is, it would be unnatural or awkward if the user himself has to manually input the information regarding his state. In order to convey the information without the user feeling unnatural or awkward, it is necessary to detect the user movement and posture automatically by the computer, and reflect the information on the avatar. In order to implement this, we have used the image processing method for the pictures taken from the camera and have used the automatic determination scheme for the degree of concentration which we have already implemented. However, it is probable that as a human nature, some users may want to hide their state or condition. Anticipating this privacy situation, we will support manual input by the user.

In the field of image processing, numerous researches have been done, in which the image from the camera is processed and the human gesture is identified and recognized. Among those researches, we have utilized the spotting recognition method proposed by Nishimura and Oka [6] for the image processing section. The features of this method are as follows:

- It is robust to changes in the background or clothing of the subject person while processing the image
- It is robust to changes in the track of the gesture
- It can recognize the gesture in real time
- It is presupposing the recognition of relatively large gesture.

A certain amount of real time characteristics is necessary for change in state of the user since the change in the user state (gesture) is utilized for supporting the communication in this system. Also, the track of the gesture, clothing or background would be different for each user; it is valid to select the recognition method that would be robust to changes to these items.

The degree of concentration of the user is detected from the movement of the keyboard, mouse or chair [3], and the expression of the avatar is changed depending on the concentrated state or relaxed state.

In terms of expressing the user state, it is necessary to investigate more regarding what behavior or posture a human would take in daily office environment and what impression people around would receive from watching those daily behaviors.

For that purpose, from the researches[5] on the relationship between the avatar behavior and the sensitivity words, we have selected the following feelings words that may be applied to the virtual office:

- enthusiastic
- irritated

- surprised
- angry
- uninteresting
- disappointed
- troubled

Also, the gesture is not only indicating the human state but also in some cases indicating the movement (or action). Reflecting the actual human movement for the avatar on the virtual space would become the important element in carrying out communication smoothly. The following items can be thought of for movements that the user would take daily at the desk in the office:

- beckon
- point a finger
- bow
- wave a hand

It is not certain whether all states or movements can be identified or recognized from those two elements just described above. This is because the gesture recognition method is presupposed to recognize the gesture of relatively large movement, and it would be difficult to differentiate closely resembled gestures. In the future we are planning to select, by a questionnaire survey, effective items for communication from these states and movements just described here, and we will attempt to recognize the limited number of human states and movements that have high recognition rate and are highly effective.

4. Summary

In this research, we have achieved in making the activities in the virtual space more realistic by providing more non-verbal information and in providing the avatar for each user with gestures that would provide triggers for information communication and would support high quality communication. Using the virtual office Valentine which we have developed for the 3-dimensional virtual space, we have constructed the system specialized for intra-office communication. First of all, we paid attention to a “handshake” as a method for providing the realistic feel of the virtual action, and we have implemented the virtual handshake using the handshake device. Also, by applying the gesture recognition for the time-varying image, we have reflected the actual user state and movement into the avatar on the virtual space. By carrying out these achievements, we attempted to narrow the gap between the real world and the virtual space. Our future plan is to carry out further implementation and evaluation of the system.

References

- [1] G. Greenhalagh, and S. Benford. MASSIVE: A Collaborative Virtual Environment for Teleconferencing. *ACM Trans. Computer-Human Interaction*, Vol. 2, No. 3, 1995
- [2] Roger Lea, Yasuaki Honda, Kouichi Matusda. Virtual Society: Collaboration in 3D Space on the Internet. *Computer Supported Cooperative Work The Journal of Collaborative Computing*, Vol. 6, No. 2-3, pp. 227-250, 1997
- [3] S. Honda, H. Tomioka, T. Kimura, T. Ohsawa, K. Okada and Y. Matsushita. A Virtual Office Environment Based on a Shared Room Realizing Awareness Space and Transmitting Awareness Information. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pp. 199-207, Oct. 1997
- [4] Leger Brosnahan, Japanese and English Gesture – Contrastive Nonverbal Communication, Taishukan Press, 1988
- [5] Inoue, Okada, Matsushita, Transmitting Feeling Information in Network using Gesture and Movement, *Journal of Information Processing Society of Japan*, Vol.36, No.8, pp. 2037-2045, 1995
- [6] Nishimura, Mukai, Nozaki, Oka, Spotting Recognition of Gestures of Multiple People from a Single Time-Varying Image using Low Resolution Characteristics, *Journal of Institute of Electronics, Information and Communication Engineers*, Vol. J80-D-II, No.6, pp. 1563-1570, 1997
- [7] Honda, Tomioka, Kimura, Okada, Matsushita, Location Awareness Solving Isolated Feeling of Stay-At-Home Employee/Virtual Office Environment Based on Awareness Space, *Journal of Information Processing Society*, Vol.38, No.7, 1997
- [8] Kazumasa Ochiai, Commercialized Virtual Society State of People Space/Basic Technology/Prospects, *Research Report of Japan Virtual Reality Society*, Vol. 1, pp. 13-18, Jul. 1997
- [9] Suzuki, Sugawara, Nagashima, Watanabe, Evaluation of Cyber-Community, *Research Report of Institute of Electronics, Information and Communication Engineers*, pp. E92-120, Feb. 1993

SIG-CyberSpace, VCR 97-15
November 27, 1997

Considerations Relating to the Registration of Information Icons in a Three Dimensional Virtual Space

Masayuki Inoue Yasuyuki Kiyosue
NTT Human Interface Laboratories

Abstract

Nowadays three-dimensional networked virtual reality system (3D cyber space system) is coming to be popular. Users can communication on this system with others by manipulation his or her own "avatar". We are developping the InterSpace system that is one of 3D cyber space system, and presenting Cyber Society trial constructed by using this system. Getting log data and accepting users' remark in this trial, we are going to upgrade InterSpace system. In this paper we analyze them and propose new architecture about information icon entry and arrangement on three-dimensional cyber space system.

A Study on An Information Icon Entry in the Networked Three Dimensional Virtual Space

Masayuki Inoue Yasuyuki Kiyosue
NTT Human Interface Laboratories

Abstract

Nowadays three-dimensional networked virtual reality system(3D cyber space system) is coming to be popular. Users can communicate on this system with others by manipulating his or her own "avatar". We are developping the InterSpace system that is one of 3D cyber space system, and presenting Cyber Society trial constructed by using this system. Getting log data and accepting users' remark in this trial, we are going to upgrade InterSpace system. In this paper we analyze them and propose new architecture about information icon entry and arrangement on three-dimensional cyber space system.

1. Introduction

In recent years, the popularization of multimedia computers that can handle voice and video, as well as the development of network technology—the primary example being the Internet—has opened up the possibility of communication between people that goes beyond text-based "chats" to incorporating real-time voice and video, typified by the services offered by Microsoft NetMeeting™ [1]

In addition to these developments, virtual reality used for applications, including physical video games, has made it possible to construct virtual spaces that resemble the physical world.

Today, much of the research and development is applied to integrating these technological elements to create next-generation communication tools and three-dimensional cyberspaces.[2, 3]

A three-dimensional cyberspace is a virtual world generated with 3D-CG using on a computer network in which users control avatars that meet and communicate with each other, providing interpersonal interactions.

In July 1997, we began our “InterSpace CyberSociety” public trial,⁴ using “InterSpace” as the “three-dimensional cyberspace” that we had created in our research and development. We opened “InterSpace” to the general public and used the results we obtained in the trial to improve our system.

In this monograph, we provide an analysis of the current state of the InterSpace CyberSociety trial, discuss future developments of InterSpace and introduce the course we are currently pursuing.

2. Public trial[4]

The “InterSpace CyberSociety” public trial runs on the Internet using InterSpace.

A large number of users and information providers were invited to participate in the trial. The objectives of the trial are to explore the possibilities of personal encounters and communication in a virtual town with a large number of users, while pursuing quality and new services.

As residents of the virtual town, users can communicate via face-to-face video and digital voice, can post and read messages on multimedia bulletin boards, play team games, visit a digital aquarium and enjoy other multimedia experiences.



Fig. 1. A screen shot of InterSpace

2.1. Status of participation in the trial

Fig. 2 shows the level of participation in the public trial between July 16 and October 19, 1997. In the graph, the horizontal axis represents the dates of trial, and the vertical axis shows the cumulative number of registered users proportionally, indexing the cumulative number of registered users on October 19 as 1.

The cumulative numbers of registered and logged-in users in were counted starting July 16, and were counted only once regardless of the number of times they logged in. Repeat logins of individual users were tabulated on a weekly basis. Only the second and subsequent logins were tabulated, and for each given week, users were counted only once regardless of how many times they logged in.

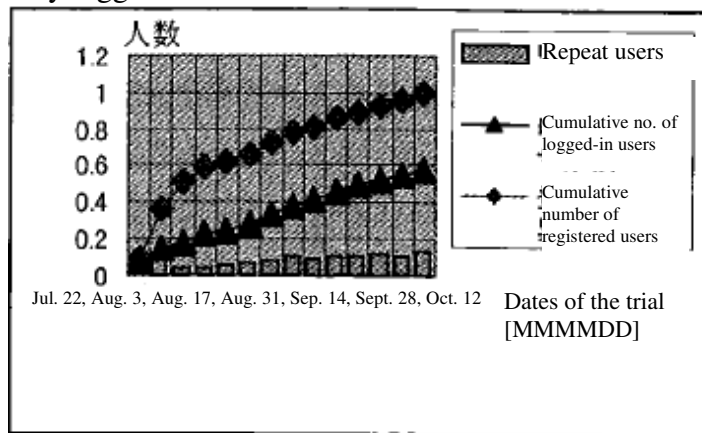


Fig. 2. Status of participation in the public trial

In Fig. 2, there is a difference between the cumulative number of registered users and those who logged in over the same period of time, showing that despite being registered, there were users who could not log in, whom we refer to as L users.

Also, the difference between repeat, or R, users in one week and the cumulative number of users who logged in the week before indicates the number of users who were not repeat users regardless of the ability to log in.

The reasons for having L users may include download and installation problems, which indicates there is a threshold level for using InterSpace.

The reasons for having R users may be a lack of interesting InterSpace content and the fact that returning to InterSpace is not mandatory.

To examine in particular the possibility of InterSpace as a next-generation communication tool, we focused on the R users and conducted study on their use status.

2.2. Use status of participants in the public trial

To provide support to the general user participating in the public trial, as well as to get their unfiltered opinions and impressions, our staff members provide support on a daily rotating basis. A summary of the main reports follows.

1. After logging in, there is no response, so the user logs out
 2. Users who are logged in are a diverse group and cannot find a common topic
 3. A bulletin board that opens the InterSpace website provides easy-to-use guidance
 4. Users meeting for the first time found it difficult to have a conversation
 5. User photos are replaced with bitmaps as a way to generate interest in a shared topic
- Of these, 1 to 3 are reports relating to opportunities to have a conversation, and 4 and 5 relate to maintaining conversations. These reports demonstrate that there are different types of barriers to communication. We therefore turned our focus to human communication behavior in the real world.

3. Communication behavior in the real world

In actual society, people form groups in stages, and communication behaviors are used before and after group formation.

Ishikawa et. Al[5] summarize the group formation process into the following three stages:

Stage 1

A community is formed based on shared elements such as having the same interests, belonging to the same company and living in the same area.

Stage 2

When people in the community meet on a daily basis, their randomly occurring conversations lead to people knowing things about each other. Based on this mutual information, their psychological distance decreases, and groups form.

Stage 3

“Threads” on different topics are created from conversations within the group. Group members vary dynamically according to the topic, which results in the formation of groups within groups. In order for people to learn information about each other as in stage 2, a great deal of communication is required.

The mechanism underlying communication is described by Newcomb (1953) using his A-B-X model. He says, “When two people, A and B, have a relationship through some matter of concern X and their attitudes differ with respect to X, tension arises in the A-B-X system, and communication arises between A and B to resolve the tension and restore equilibrium in the system.”[6] That is to say, some matter of concern X is needed for communication to occur.

This is equivalent to the experience where people at an academic conference or social gathering find themselves unable to strike up a conversation with people they do not know or cannot maintain a conversation if someone else approaches them. This appears to be the reason why people do not find common topics that can serve as the matter of concern X.

4. Future development of InterSpace

For InterSpace to develop going forward, the features of the system need to be closely aligned to the needs of society.

As it is currently, InterSpace has the following two features that have a very close resemblance to the sort of environment found at academic conferences and social gatherings described above:

- 1) Venues are provided for an unspecified large number of users to come together that is not telephoning, video conferencing or groupware.
- 2) Constrained by the real world, people are able to spend only a limited amount of time each day in the virtual world.

As evidenced by personals magazines, it may be noted that people today seek friends who share their interests rather than making friends with neighbors without any interest in common.

It therefore seems that adding the following features to InterSpace would bring it into line with the needs of society:

- 1) The ability to share information that is obtained and created based on personal interests with an unspecified large number of people to find users who share your interests
- 2) The means to communicate non-synchronously such as e-mail so people can arrange to meet, increasing the opportunities to communicate synchronously on InterSpace

With these features, people would be able to find common elements based on shared information, the psychological distance between them would decrease through e-mail and conversation, and groups would form of people with shared interests.

5. Current efforts[7]

Just as in the real world, people have trouble starting up a conversation with people they do not know.

We are currently studying how to create visible information icons on shared virtual space from information that a user has obtained or created based on personal interests. The icon be registered with the system as a shared topic that people could discuss.

5.1. System composition

The InterSpace system is composed of location management servers and clients. The management servers collect data identifying where user avatars and three-dimensional icons are in virtual space from user terminals, and redistribute that data back to the terminals. The clients use the Windows OS, and the use of network computers enables file sharing and connections to the location management servers over the network. The clients then graphically render the virtual space with 3D-CGs, and location data of the avatars and three-dimensional icons from the location management servers is used to depict other avatars and three-dimensional icons in the visible range of the first user's avatar. Avatar operation is carried out with an I/O device such as a mouse, and three-dimensional icons are registered using a drag-and-drop operation.

5.2. Feature overview

Fig. 3 shows how a three-dimensional virtual space could be displayed continuously on the desktop when using a three-dimensional cyberspace. With this format, the user can drag and drop a file icon from the desktop to the three-dimensional virtual space to register a data icon.

There are a number of ways in which three-dimensional icons can be displayed. We took the image captured when data is opened by an application and mapped it onto a graphical object in the form of rectangular solid. Fig. 4 is a screen shot of the three-dimensional virtual space as viewed from an information provider, and Fig. 5 is as seen from an information receiver. This enables anyone in the three-dimensional virtual space to share (give and receive) files, and provides opportunities for even those not exchanging information to talk.



Fig. 3. The screen of an information provider

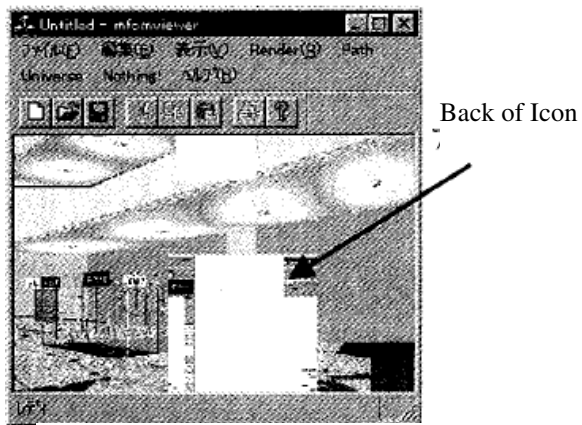


Fig. 4. Information Provider Screen

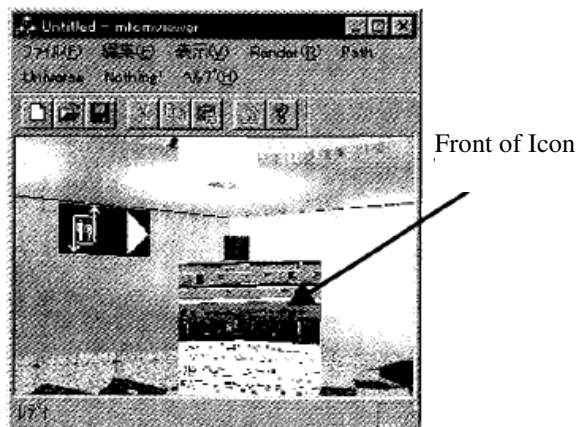


Fig. 5. Information Receiver Screen

5.3. Discussion

Sharing information obtained and created based on personal interests with other users may play a role in new group formation while generating and maintaining conversations.

When the content of InterSpace is considered to be the primary substance of the virtual space, it is difficult to increase frequency of updates. When a user's social activities as manifested through the avatar are included in the content, however, the update frequency varies greatly. The attractive thing about reality is its constant change, and in order to make a virtual three-dimensional space really attractive, a feature is needed whereby the system providers (venue providers) and users (venue participants) can make changes.

To increase the use of InterSpace, considerations are needed not only on the technical front but the psychological one as well.

6. Conclusion

In this paper, we described the current status of our public cybersociety trial using InterSpace and discussed issues that InterSpace will face. As a concrete way to solve these issues, we discussed registering icons in a three-dimensional space.

Going forward, we will next study how to create information icons displayed so that the information can be conveyed. We will evaluate the results by analyzing user behavior, conversations and data use logs. We will also collect opinions and impressions from a large number of people through surveys and reflect the results in our system architecture and design.

References

- [1] http://www.microsoft.com/ie_intl/ja/netmeeting/
- [2] <http://www.lab7.kuis.kyoto-u.ac.jp/free-software/freewalk/freewalk.html/>
- [3] <http://vs.sony.co.jp/>
- [4] <http://cybersociety.elcs.intsp.or.jp/>
- [5] Ichikawa, Haryu: "Study of Hierarchical Community Support Services that Vary Dynamically," *55th Information Processing Society of Japan*, Vol. 4- pp 139, 140, 1997
- [6] Newcomb, T.M.1953: An approach to the Study of communicative acts. *Psychological Review*, 60, 393-404
- [7] Inoue, Kiyomatsu: "*Sanjigen kaso kukan niokeru joho aikon no haichi ni kansuru kento*" [Study of data icon layouts in three-dimensional virtual spaces], *55th Information Processing Society of Japan*, Vol. 4- pp 97, 98, 1997

Membership Request

SIG-CyberSpace of the Virtual Reality Society of Japan

The Virtual Reality Society of Japan (VRSJ) was launched on May 27, 1996. Having launched its SIG-CyberSpace group on March 27, 1997, the VRSJ is currently accepting members from a wide range of fields.

SIG-CyberSpace of the VRSJ creates advanced human interfaces and community features within virtual reality societies, and provides interdisciplinary venues for studying how to use new telecommunication technologies to form and maintain societies over the Internet and the like as a form of integrated social information science. By creating venues for active researchers to exchange information from technical and social perspectives, SIG-CyberSpace furthers academic and artistic progress in virtual communities, aiming to contribute to the improvement and development of culture and the welfare of society both in Japan and around the world. We strongly encourage those with an interest in the engineering or technical aspects of virtual communities in any field such as engineering, the physical sciences, economics, psychology, sociology, philosophy or art, as well as those interested in the psychological, philosophical, social or economic issues to join SIG-CyberSpace.

SIG-CyberSpace is a forum for the interdisciplinary study of network community functions that may have important applications that kindle further interest in the Virtual Reality Society.

[Future activities of SIG-CyberSpace]

SIG-CyberSpace holds regular meetings in which members make presentations in their fields, debates are held, and questions are posed.

SIG-CyberSpace also organizes expert committees, operates SIG-CyberSpace, and provides cooperation to its parent organization the Virtual Reality Society in its national convention, journals and other activities that pertain to cyberspace.

A review is also planned to look into providing information on the Internet.

The fourth society meeting and expert committees are planned in 1997.

[Applications]

Applications are accepted online.

<http://www.myo.inst.keio.ac.jp/~vd/vc.reg.html>

[Deposit information for paying the membership application fee]

Sanwa Bank, Kasuga Branch, regular account: 3584875, name: VR Gakkai Kaso Toshi Kenkyukai

Annual membership fee: 3000 yen

[Inquiries]

Toshihiko Yamaue, Integration Service Laboratory

NTT Multimedia Network Laboratories

(Telephone: 0422-59-3962, fax: 0422-37-97691, e-mail: yam@acm.org)

VR Res. Rep. Vol. 2, No. 1
Issued: November 27, 1997
ISSN 1343-0572

1997 by the Virtual Reality Society of Japan (VRSJ)

Issued by: The Virtual Reality Society of Japan
Laboratory Office, Engineering Building, Department of Mathematical Engineering and
Information Physics, Faculty of Engineering, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo
Telephone: 080-592-3550
Fax: 03-5689-7201
e-mail vrsj@star.t.u-tokyo.ac.jp

The Virtual Reality of Japan c/o Tachi Laboratory, Department of Mathematical Engineering &
Information Physics Faculty of Engineering, The University of Tokyo

7-3-1, Hongo, bunkyo-ku, TOKYO 113 JAPAN

Duplication and reproduction prohibited