Distributed Processing Architecture for Virtual Space Teleconferencing

Haruo Takemura, Yasuichi Kitamura, Jun Ohya and Fumio Kishino

ATR Communication Systems Reserch Labolatories
2-2 Hikari-dai, Seika-cho, Soraku, Kyoto 619-02 JAPAN
+81-7749-5-1211 takemura@atr-sw.atr.co.jp

ABSTRACT

Discussed in this paper is a distributed processing architecture for a virtual space teleconferencing system. Virtual space teleconferencing is a promising application of networked virtual environments. People in different places will be able to meet each other in a virtual teleconference room and proceed with various conferencing taska. When such a system creates a realistic virtual environment, it can be referred to as "Teleconferencing with realistic sensations." Furthermore, as the conference environment can be shared by a number of users, it is possible to perform various kinds of cooperative work using the system.

In this paper, the design strategies of the processing architecture for a virtual space teleconferencing system are reviewed and a case study of building a prototype of such a teleconferencing system is reported. The prototype system reproduces a 3-D image of a conference participant in a virtual meeting room. Two persons can communicate with each other through the system. A limited range of facial expressions and body motion of one user are measured in real time, in order to generate a computer graphics image of the meeting partner at the remote site. It is observed that by distributing redundant data over the network, a relatively short response time and fast image update cycle has been achieved.

KEYWORDS

Virtual reality, communication with realistic sensations, virtual space teleconference, shared workspace, networked virtual environment.

INTRODUCTION

"Teleconferencing with realistic sensations" aims at allowing smooth and intricate communications among users by reproducing various aspects of face to face conferencing. One method for achieving this is a virtual conferencing room created by virtual reality technology. Moreover, if we regard this virtual conferencing room as a shared workspace among users, an environment can be formed where users can cooperatively solve various problems.

In order to achieve "Teleconferencing with realistic sensations," the concept of virtual teleconferencing is proposed[3]. As shown in Figure 1, a "virtual teleconferencing" system creates an image of a conference room (virtual conference room) using computer graphics in real time. It also constructs images of the remotely located conference participants. Users of the system can talk to each other or proceed with the conference with the sensation of sharing the same space. There are many problems to be solved however, before this system can be realized. Some of them are now being studied at our laboratories[1,5,6].

A virtual workspace for virtual teleconferencing is to be generated by computer graphics. "Virtual manipulation" is a facility enabling users to interact with the virtual environment. Users can cooperatively work on tasks such as city planning or modeling a new car. In such a case, it is desired that users be able to handle objects in a virtual environment just as they do in a real environment without a sense of incompatibility. If this can be achieved, users can collaborate in various ways without any prior training of teleconferencing operations.

However, there are not many reports on cooperative environments created by virtual reality. Ishii et al. pointed out the importance of shared work environment for cooperative work, although their approach is not based on a virtual environment or computer-generated environment[2]. Codella et al. implemented a demonstration system of a multi-person virtual world[8].

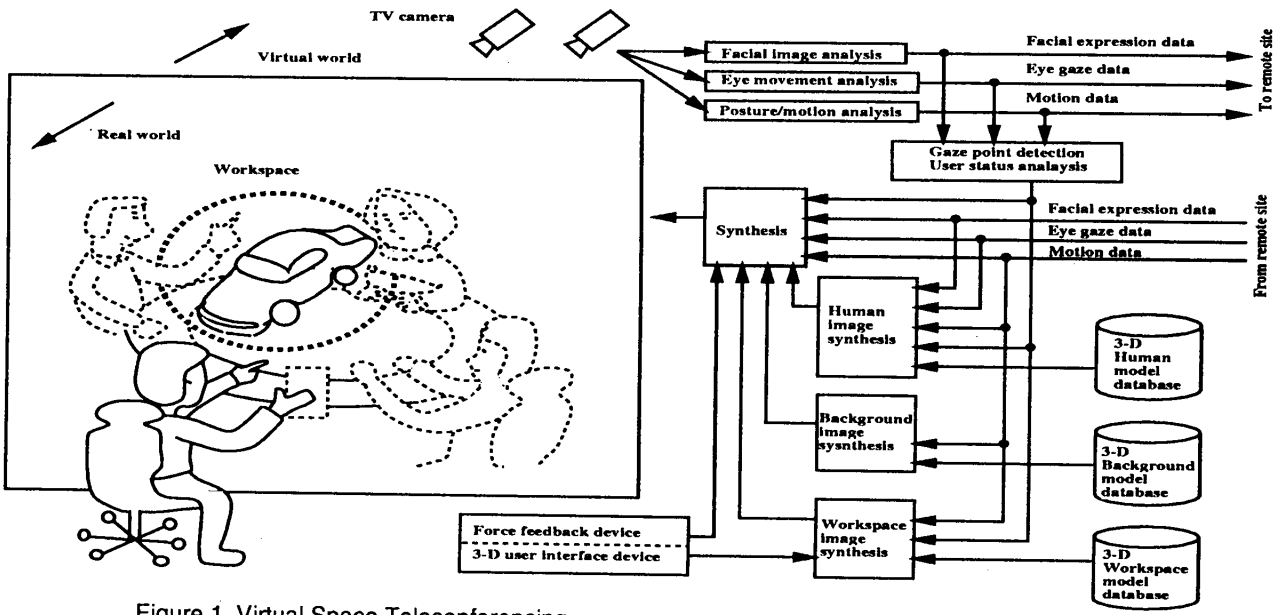

Figure 1. Virtual Space Teleconferencing

This paper first reviews and classifies networked virtual environments according to the level of sharing. Second, implementation issues of the networked virtual environments are discussed. Third, a prototype system of a "Virtual Space Teleconferencing System" is described.

NETWORKED VIRTUAL ENVIRONMENT CLASSIFICATION

Figure 2 shows an overview of a typical non-networked virtual environment. By connecting more than two of these environments, a networked virtual environment is configured as illustrated in figure 3.

A networked virtual environment (NVE) can be classified into 5 categories, as listed in Table 1.

Level 0 This is a non-networked virtual environment and does not share data.

Level 1 Features to be added to level 0 include a facility to access a networked or remote virtual world database (VWDB). The VWDB can be shared among the sites that access the VWDB. Modification of the VWDB is not allowed. The status of other sites is not displayed in VE, i.e., one can not know what the other users are doing in the VE. This type of NVE is equivalent to an on-line database retrieval system Users do not modify the database but merely access and refer to the database. A virtual museum is one possible application of this level. The advantage of this level of NVE is that the communication can be one way from the VWDB to the local site.

Level 2 In addition to the level-1 NVE, the status of each site is distributed to the others. This information can be used to display an image of other users in the same VE. In this case, a bidirectional data path exists between each local site and the VWDB. User can not only share the environment but also know about the other users. This level can be used for communication media, a virtual community etc. However, it can not be used for a cooperative work environment among users, because no local site is not allowed to modify the

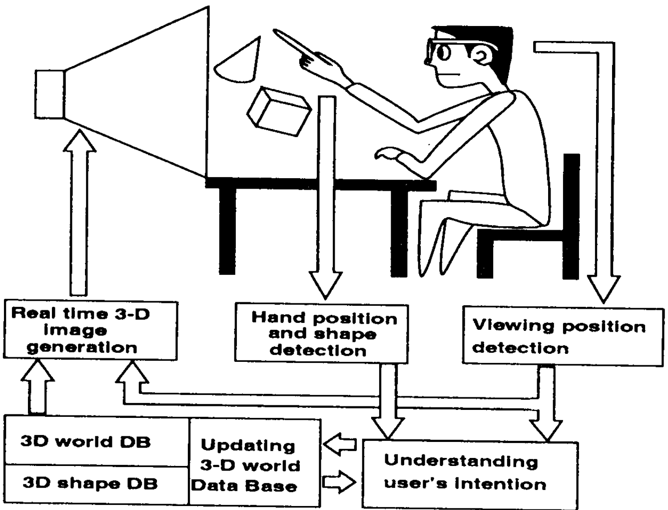

Figure 2. Typical non-networked virtual environment

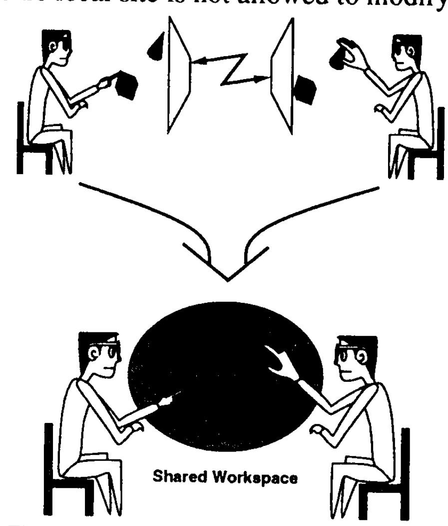

Figure 3. Shared virutal environment

Table. 1 Networked virtual enviroment classification summary

Level	Description
0	Non-networked
1	Share VWDB. No VWDB modification. No Display of other users.
2	Share VWDB. No VWDB modification. Display of other users.
3	Share VWDB. VWDB modification (independent). Display of other users.
4	Share VWDB. VWDB modification (muutal operation). Display of other users.

VWDB or shared environment.

Level 3 In addition to the level-2 NVE, independent VWDB modification is allowed. For example, an operator is allowed to modify the properties of a virtual object, such as position, color, shape. However, VWDB modification which results from operations at multiple sites, such that two people simultatiously try to chage the shape of a single object, is still not allowed. This limits the mutual interaction between two users. For example, at this level, two operators can not hold the same virtual object at the same time. When the system allows independent modification of the VWDB, concurrency control of the VWDB becomes an issue. This level can be used for a cooperative work environment, as each site can share VWDB, can acquire the status of the other site(s) and can modify the VWDB cooperatively. The simplest implementation is that only one client is allowed to perform an modification. The entire database is locked during the operation.

Level 4 In addition to level-3 features, mutual operation is allowed. Mutual operation is an operation on the VWDB that results from control from multiple sites. For example, muttual interaction is the situation that arises when more than two operators cooperatively change the properties of a single object, or when objects controlled by more than two operators collide. In this case, more sophisticated concurrency control over the VWDB is required. However, this is the most flexible structure for NVE.

For a virtual space teleconference system, at least, level-2 NVE is required to enable users to communicate with each other. Thinking of ease of implementation and cooperative work operation, we have chosen to use Level-3 NVE for our Virtual Space Teleconferencing system.

IMPLEMENTATION CONSIDERATIONS

Figure 4 shows a processing outline for a non-networked VE. Basically, it includes componets to hadle three main processes: 1) Acquire user's status, 2) Update virtual world database (VWDB), 3) render images or provide other output (eg. tactile feedback, etc.). Even in this case, when each sensor havs a different sampling rate, the entire processing loop is effected by the slowest input device, as shown in Figure 4. A typical glove type input device such as the DataGlove[8] has an sampling rate of 60Hz. A 3-D position sensor[4] has a sampling rate of 30Hz, when measuring two 3-D positions (usually head position and hand position). To upgrade system performance, using shared memory is effective (Figure 5). The input process continuously acquires data from the sensors. The output process simply reads in the data from shared memory and returns the value. Simple semaphore locking can be used for shared memory control. By using this structure the system can use different cycle times for acquiring data from sensors and reading data from sensors.

When considering a level-3 the NVE following problem must be considered:

Consistency: Each site must share the status of the virtual world, which changes according to operations performed at multiple sites. Therefore, there must be a mechanism to maintain consistency.

The simplest way to maintain the consistency of a distributed virtual world database is to restrict manipulation, such that only one site at a time can perform operations to alter the status of the virtual world, such as grasping, rotating or translating an object. This mechanism can easily be implemented by token passing. The site that has the token is allowed to manipulate objects in the virtual world. When there is no need for manipulation, the token is passed to the next site However, this mechanism obviously narrows the

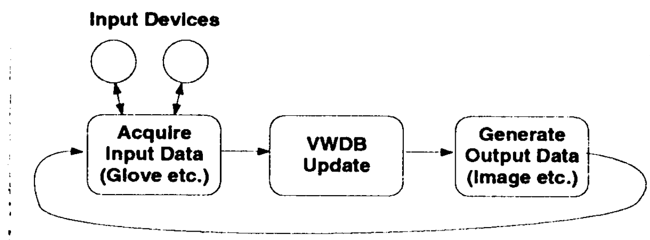

Figure 4. Non-networked VE processing architecture

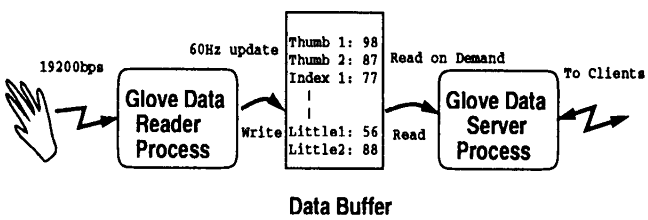

Figure 5. Data acquisition server architecture

Table 2. Measured communication throughput between two workstation

Method	TCP Protocol	UDP Protocol
With RPC	132 two-way / sec (264K byte/sec)	143 two-way /sec (285k byte/sec)
Without RPC	206 two-way / sec (411K byte/sec)	208 two-way /sec (416K byte/sec)

* 1K byte data are send/recived bwtween two IRIS340 VGX workstations using 10Mbps Ethernet. The network were shared and user by 80 workstations during the mesurement.

possibilities of cooperative work. When N sites are used, a single site is the source of the actual operation and the other N-1 sites are present with the results of the operation.

Another way to solve this problem is assuming that an operation on one object does not interfere with an operation on another object. By making this assumption, it is possible to allow N different sites to simultaneously manipulate N different objects. The results of the operations at each site are collected at the master site and then redistributed to all sites. Our prototype system uses this method.

Mutual exclusion: If users simultaneously try to alter the status of the virtual world, contradictions may arise. For example, two users may simultaneously try to grasp the same object and translate it in different directions. These operations must be prohibited to avoid inconsistencies in the virtual world status.

Mutual exclusion to avoid simultaneous manipulation of the same object must also be incorporated. If the system has a master site at which all requests for an operation are concentrated, this can be accomplished by simply assigning a priority to each request. When two sites request an operation at the same time, the master site allows an operation by the site with the higher priority. Objects being manipulated are marked in order to prohibit manipulation by another site.

Interactivity: In order to provide good interaction, the system must respond to users' reactions as quickly as possible. Unlike a single user system, interactivity in a cooperative work environment is affected by the time required for data transmission. Therefore, any processing that can be performed locally must be selected and implemented at each local site. Distributing redundant data to each site may significantly reduce the communication load between two sites and improve system response. In particular, the following strategies are actually used:

1)3-D shape data of objects in the virtual environment are distributed to each local site. A user's head position is also measured locally. Using this head position data and 3-D data, a 3-D scene can be rendered locally. Thus, a user at a local site experiences minimum delay in image generation based on one's head position.

2)When a user grasps an object in the virtual environment, the lag between one's hand motion and object motion is minimized by locally binding the hand position data and object position (Figure 6). At the same time, hand position data is transmitted to the master site to update the master virtual world database, the updated database of which is then distributed to other local sites. This local binding does not cause any inconsistency in the virtual world database provided that each object in the virtual world is manipulated by a single hand at any one time.

VIRTUAL SPACE TELECONFERENCING SYSTEM

Two of the systems in Figure 2 were used to implement a cooperative work environment.

This implementation structure has the following characteristics:

1)Each site can rapidly generate a CG image, because a copy of the data for the image generation is distributed to each site.

2)Gesture understanding and database updating are performed only by VWDBM. Therefore, it is relatively easy to implement a mutual exclusion facility, such as prohibiting two users from simultaneously grasping the same object.

3)Once the "Virtual World Data Base" are distributed to each site, there is relatively little data traffic.

When the communication link between the two sites have a lag, users get slow respons in their their operations. To esamine the communication overhead of the prototype system, several types of measurements were made. First, Table 3 shows the communication

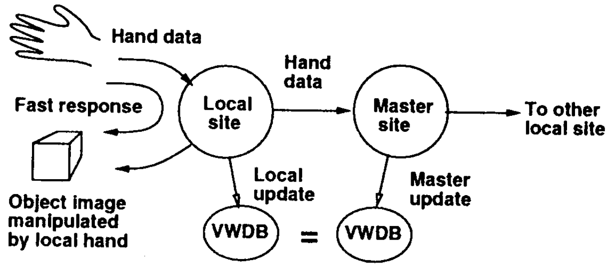

Figure 6. Local loop for fast response

Table 3. System Servers

Name	Description	Data direciton and size
Tracker server	Spools and distributes data from 3-D tracker	c -> s None s -> c 416 byte/request
Glove server	Spools and distrbutes data from DataGlove	c -> s None s -> c 40 byte/request
Gesture server	Handshape-to-Gesture conversion	c -> s 44 byte/request s -> c 4 byte/requiest
VE server	Maintains Master VWDB for update loop	c -> s 76 byte/request s -> c 12 + 326 × (No. of Sites) + 28 × (No. of Moving OBJ)
	when down loading VWDB for a local copy	Size depends on VWDB size

Table 4. Server distribution and system processing cycle speed

System	Servers running
IRIS 4D340VGX	VE server, gesture-server(for-all), glove-server(for site A), tracker-server(for site A) , video-lab(Facial expression detection)
IRIS Crimson/RE(site B)	VE client, glove-server(for site B), tracker-server(for site B) (Rigid face images only)
IRIS Crimson/RE(site A)	VE client (CG human image with facial expression and body motion)

Simple VWDB (Sigle cubic objects displayed)

Site A Processing	min.	max.	ave.	Site B Processing	min.	max.	ave.
VW rendering	0ms	1ms	0.2ms	VW + face rendering	40ms	50ms	17ms
CG human image	60ms	70ms	65ms				
Communication	10ms	40ms	21ms	Communication	10ms	40ms	20ms
Total cycle time	70ms	110ms	88ms	Total cycle time	60ms	110ms	37ms

Complicated VWDB (many objects with aproximately 10,000 polygons in total)

Site A Processing	min.	max.	ave.	Site B Processing	min.	max.	ave.
VW rendering	60ms	120ms	74ms	VW + face rendering	120ms	140ms	129ms
CG human image	60ms	200ms	83ms				
Communication	10ms	80ms	21ms	Communication	10ms	60ms	22ms
Total cycle time	140ms	300ms	181ms	Total cycle time	130ms	180ms	151ms

speed between two workstations used for the prototype system. As is evident from the table, the overall rate is much better when remote procedure call (RPC) is not used. Howerver considering the fact that using RPC makes it easy to support different types of workstations, we have used an RPC-based server-client architecture. Table 3 shows a list of implemented servers. Figure 7 shows the RPC relationship between severs and clients.Table 4 shows which server or client runs on which work station. Also, Table 4 shows the measured cycle time of the prototype system. Regardress of the size of the VWDB, the communication cycle times are quite stable and low enough to maintain the system's quick response time.

Using The System

Figure 8 shows two users facing each other and manipulating the CG image of a space shuttle, located between the two users. An image of the remotely located person is regenerated by computer graphics and displayed on the screen. A 70-inch CRT projector is used as an image output device for a "viewing position tracking stereoscopic display," which is a system for measuring user eye position and generating perspective projection images based on real time measurement. A user who wears a stereo viewer and a DataGlove[8] can directly grasp the virtual object and perform such manipulation as translation and rotation. Some of the objects in the scene are fixed and some can be moved based on the constraints given to the object.

Neither the 3-D cursor nor the computer graphics hand is displayed. The operation, locating one's hand where the object is perceived to be and grasping it, determines the target of the operation. In comparison with the method which uses an indirect pointer such as a cursor, this method is superior because it does not require users to learn how to operate cursor control devices and allows intuitive operation.

A 3-D model of a user at site B is constructed using a wire frame model mapped by color texture and is displayed on the 3-D screen at site A. To realize real-time detection of facial features at the site B, tape marks are attached to the facial muscles, and the marks are tracked visually. To detect movements of the head, body, hands and fingers in real-time, magnetic sensors and a DataGlove are used. When the movements of the participant are reproduced at site B, the detected results are used to drive the nodes in the wire frame model.

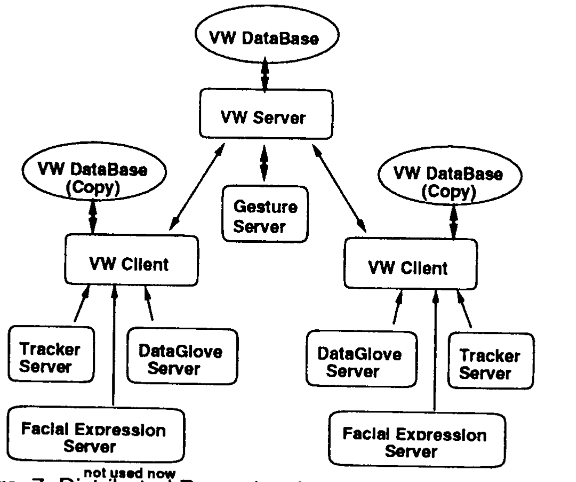

Figure 7 Distributed Procesing Architecture for the system


Figure 8 Virtual Space Teleconferencing System

Although CG generated human images are not rich in expressions, the image of the conference partner is nevertheless very usefull and plays an important role in cooperative work.

SUMMARY

A teleconferencing system using a networked virtual environment is described. Implementation issues of the networked virtual environments are discussed. Specific design criteria of the system, such as concurrency and interactivity control have also been described. A prototype of a "Virtual Space Teleconferencing System" is implemented based on the above considerations. By distributing redundant data to each local site, the data traffic between two sites is relatively light. However, the implementation limits the operation of virtual objects. To give a more free interaction to the user will be the next step of our approach.

ACKNOWLEDGMENTS

The authors wish to thank Dr. N. Terashima - President of ATR Communication Systems Research Laboratories, and Dr. K. Habara - Executive Vice President of ATR International (Chairman of the Board of ATR Communication Systems Research Laboratories), for their thoughtful advice and encouragement on this research. The authors also wish to thank their colleagues who voluntarily took part in discussing this study.

REFERENCES

1. Ishibuchi, K., Takemura, H. and Kishino, F. Real time hand shape recognition and its application for man-machine interface (in Japanese), in *Proceedings of IEICE conference 1991 Autumn* (1991), IEICE.
2. Ishii, H. and Miyake, N. Toward an Open Shared Workspace, *Communications of the ACM,* 34, 12(Dec. 1991).
3. Kishino, F. Communication with Realistic Sensation(in Japanese), *3-D image,* 4, 2(1990).
4. Raab, F.H., Blood, E.B., Steiner, T.O. and Jones H.R Magnetic Position and Orientation Tracking System, *IEEE Trans. on Aerospace and Electronics system,* AES-15,5(1979). pp.709-718.
5. Takemura, H., Tomono, A. and Kobayashi, Y., A Study of Human Computer Interaction via Stereoscopic Display, in *Proceedings of HCI '89*(Boston, 1989), Elsevier Science Publishers, pp. 496-503.
6. Takemura, H., Tomono, A. and Kishino, F. A Usability Study of Virtual Environment, in *Proceedings of the 6th Human Interface Symposium*(Tokyo, 1990), The Society of Instrument and Control Engineers, pp. 577-582.
7. Takemura, H., Tomono, A. and Kishino, F. A Study of Force Feedback in Virtual Workspace(in Japanese), in *Proceedings of IEICE Conference 1991 Spring*(1991), IEICE.
8. Zimmerman, T.G. and Lanier, J. A Hand Gesture Interface Device , in *CHI+GI '87 Proceedings(1987), ACM Press,* pp.189-192.
9. Codella, C. et al. Interactive Simulation in a Multi-Person Virtual World, in CHI '92 Proceedings (Monterlay, 1992), ACM Press, pp. 329-224.