I-NEXT: An Interactive Networked Expression eXperience Testbed*

Dongpyo Hong[1], Wonwoo Lee[1], Jaeheon Jeong[2], Jongwon Kim[2], and Woontack Woo[1]
KJIST U-VR Lab.[1], KJIST NetMedia Lab.[2]
Gwangju 500-712, S. Korea
{dhong, wlee, wwoo}@kjist.ac.kr[1], {jjeong, jongwon}@netmedia.kjist.ac.kr[2]

Abstract. In this paper, we propose "I-NEXT: An Interactive Networked Expression eXperience Testbed" in which users can express their intensions or emotions interactively in the networked virtual environment through a 3D vision-based user interface. The proposed I-NEXT allows multiple users in the distance to interact with not only themselves but also virtual objects in the networked virtual environments. For interactive expression experience over networks, I-NEXT is composed of three key components; SpaceSensor (3D vision based user interface for comfortable expressions), VEManager (networked virtual environment management) and the networking interface (delivery of events and/or messages of interactions for real-time interactions). Detailed explanations and implementation are presented.

1. Introduction

With the rapid progress of technologies in the areas of computers and communications, the future computing environments will support "seamless support from ubiquitous computers and pervasive networking", that is, users could do just-in-time access to any (invisible) computers at anytime and anywhere [1][2]. Consequently, this environment will require users to interact with computers through more natural and comfortable interfaces. In addition, it will be an important issue to deliver the user's intensions or emotions over the network.

Many research activities have been reported on emotional expression. For instance, "Computing Culture" group in MIT Media Lab. has actively reported on the cultural applications for physically embodied media [3]. Franuhofer IGD in German and Media Lab. in Ireland have also actively studied the creative and innovative human computer interfaces (HCI) [4][5]. Furthermore, ATR in Japan has started to build the networked edutainment environments [6]. However, these activities are yet premature for users to express their intensions or emotions with a comfortable user interface, and to interact in real-time with each other in the distance as well as virtual objects or virtual environment (VE). Therefore, it is necessary to build a new type of testbed that provides users with a comfortable user interface for natural expression and real-time interactions over the network [7][8].

In this paper, we propose "I-NEXT: An Interactive Networked Expression eXperience Testbed", which allows users to express their intentions or emotions interactively in the networked VE through a 3D vision-based user interface. For the interactive expressions in I-NEXT, it exploits users' dynamic gestures which are tracked by the 3D vision-based user interface. Based on the dynamic gestures, the interactive expressions are classified into either implicit or explicit ones. In explicit interactions, users directly manipulate virtual objects themselves, e.g., scratching or moving virtual objects. For implicit interactions, however, users make certain gestures using their arms or parts of body. According to the users' gestures, VE is

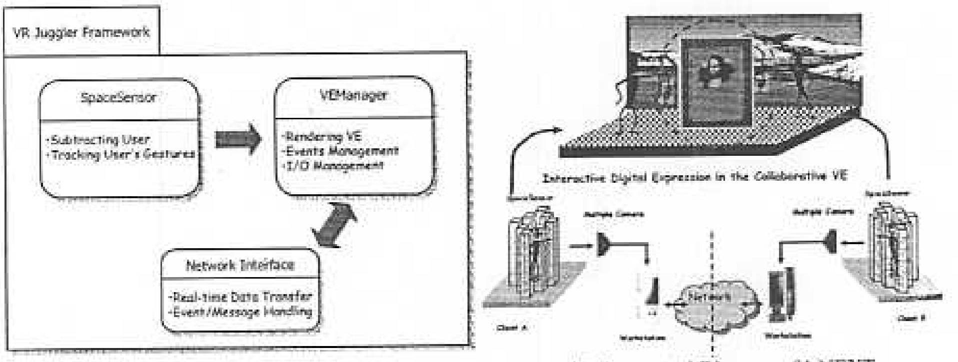
* This work is sponsored by KJIST and ICU Digital Media Lab.

changed interactively, e.g., the color tones of background scene or objects. Furthermore, the explicit and implicit interactions between users in the distance are expressed in real-time over the network. Consequently, users are able to experience the interactive expressions and share the experience over the network in real- time through I-NEXT.

This paper is organized as follows. In section 2, we explain the overall structure and each component of the proposed I-NEXT. In section 3, we show the experimental results and conclude in section 4.

2. I-NEXT

I-NEXT is composed of SpaceSensor (a 3D vision-based user interface), VEManager (networked virtual environment management), and the networking interface (delivery of events and/or messages of users interactions). Figure 1 illustrates the conceptual system architecture of I-NEXT. In SpaceSensor, it tracks users' dynamic gestures and sends the tracked position data into VEManager. At the same time, it sends other data into VEManager, that is, which region of SpaceSensor is collided by the user. In VEManager, it controls the delivered events and messages from a local user and a remote user in order to make them consistent and concurrent. In Networking Interface, it guarantees the real-time transmission of messages and interactions in the networked VE.



(a) Key Components of I-NEXT (b) Conceptual Diagram of I-NEXT

Figure 1. I-NEXT Components and Conceptual Architecture

2.1 SpaceSensor: 3D Vision-based User Interface

SpaceSensor is proposed to overcome the constraints of the traditional user interfaces (keyboard, mouse, data glove, etc.) in VE. It provides users with a natural way to express their intentions or emotions without using any complicated facilities, i.e. the users can express themselves using their dynamic gestures [7]. For the implementation of SpaceSensor, it has to separate the user from the background scene and track the user's gestures with the segmented user by exploiting 3D information from a 3D camera. For the user segmentation, we use the background subtraction algorithm which copes with the difficulties of the separation of a user from his/her natural background scene due to complex backgrounds and the interference of lighting sources such as cast shadows around the user [9][10][11][12]. The used background subtraction technique can replace expensive blue screens and/or chroma-keying devices. Figure 2 illustrates the used background subtraction algorithm along with the segmented result.

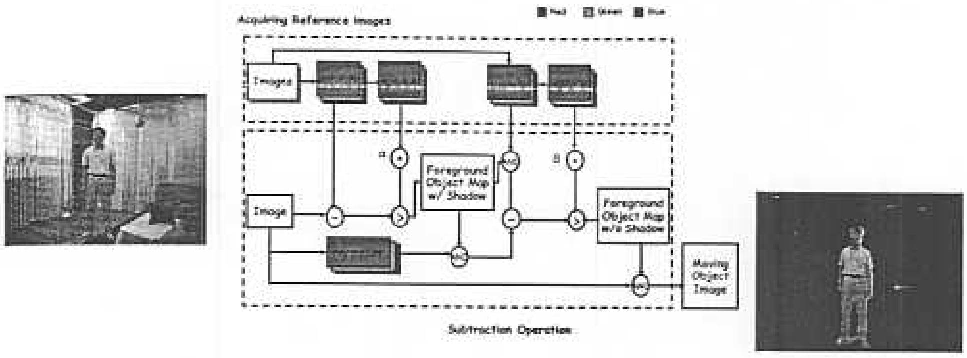


Figure 2. The background subtraction algorithm.

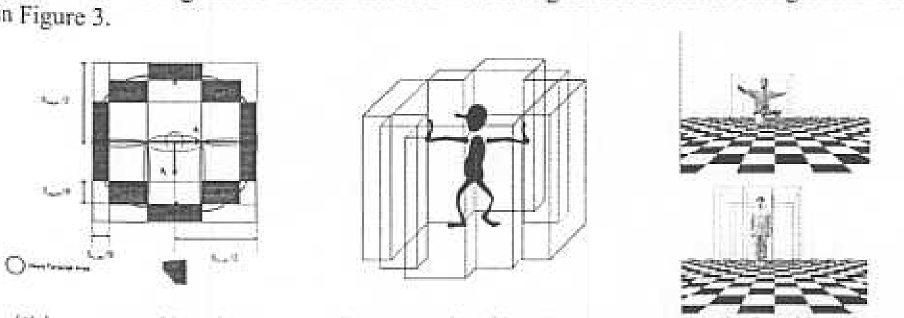
After acquiring the segmented user information from background and 3D position information of the user through a 3D camera, we need to design SpaceSensor in order to track the gestures of the user as accurately as possible while allowing the user to move around more naturally. The design of the proposed SpaceSensor focuses on tracking of natural movements of the user by making it dynamically follow the user, instead of fixing the SpaceSensor at a certain location. Furthermore, it is augmented into a reachable space from the movements of the user in order to recognize the gestures of the user correctly. To implement the proposed SpaceSensor, we first calculate the center position of the user (H_c) as follows.

$$H_c = \frac{1}{N}\sum_{i=0}^{N} D_s(i) \tag{1}$$

where D_s represents the 3D position information of the user which is acquired from the user separation procedure. N is the number of 3D points which corresponds to the number of the segmented user's pixels. From Equation (1), we can compute necessary information to implement SpaceSensor as the following equation.

$$S_{Width} = S_{Height} = S_{Depth} \equiv D_{Y,Top} - D_{Y,Bottom} \tag{2}$$

where S_{Width}, S_{Height}, and S_{Dpeth} represents *Width, Height,* and *Depth* of SpaceSensor which is based on the point of a user's center, respectively. $D_{Y,Top}$ and $D_{Y,Bottom}$ represent the top and bottom points of regions being occupied by the user, respectively. The Equation (2) is based on "*Leonardo da Vinci: The Vitruvian man*" [13]. From Equation (1) and (2), the proposed SpaceSensor is augmented around the user who is segmented from the background as shown in Figure 3.



(a) Arrangement of SpaceSensor (b) Augmentation of SpaceSenosr (c) Visualization of SpaceSensor

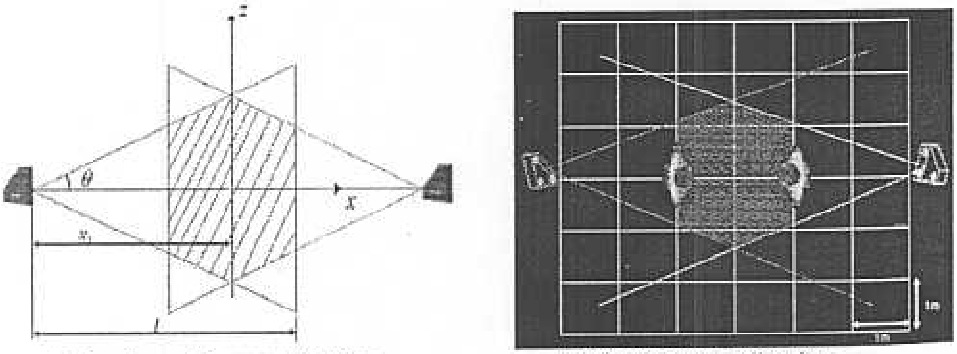
Figure 3. The Arrangment and Augmentation of SpaceSensors around a user.

As shown in Figure 3(a) and 3(b), SS{1,...,8} represents the regions of SpaceSensor. Through SS{1,...,8}, users manipulate virtual objects directly for explicit interactions as well as make gestures for implicit interactions. In the design of SpaceSensor, it has to exploit 3D position information in order to overcome the disadvantages of 2D vision-based systems. As shown in Figure 3(a) and 3(b), we can use not only top, bottom, left, and right position information of the user, but also depth information, i.e., front and back of the user. Using this arrangement of SpaceSensor, it can track users' dynamic gestures because it only examines the touched SS{1,...,8} by users. It can replace complex gesture tracking or recognition algorithms. SpaceSensor is applicable to the new type of a user interface in VE because it enables users to move as natural as possible.

2.2VEManager: An Interactive Virtual Environment Management

VEManager integrates the gesture information of a local user and a remote user through SpaceSensor and controls a virtual environment to make interactions in real-time over the network. There have been many studies about networked VE which supports interactions for multiple users [14][15]. Very few attempts have been made on the management of the layout of VE and the effective analysis of interaction information for multiple users.

In I-NEXT, VEManager takes responsibilities of providing consistent VE to both the local and the remote users as well as reflecting interaction information into VE concurrently and adequately. The proposed virtual environment is designed as shown in Figure 4, where it maximizes the region of the shared VE.



(a) SpaceSensor Cameras Allocations (b) Virtual Cameras Allocations

Figure 4. SpaceSensor Cameras and Virtual Cameras Allocations.

As shown in Figure 4, the collaborative area in VE is calculated as follows, where users interact with virtual objects for explicit interactions.

$$S_{VE} = 4 \cdot \int_{x_1 - l}^{0} \tan\theta\,(x + x_1)\,dx \tag{3}$$

where S_{VE} represents the collaborative area in VE.

From Equation (3), we can compute the maximum collaborative area (S_{max}).

$$S_{max} = S_{VE}\,\Big|_{x_1 = \frac{2}{3}l} \tag{4}$$

From the Equations (3) and (4), the position of camera for maximum collaborative area is 2/3 of the length l, which represents measurable depth range of the camera.

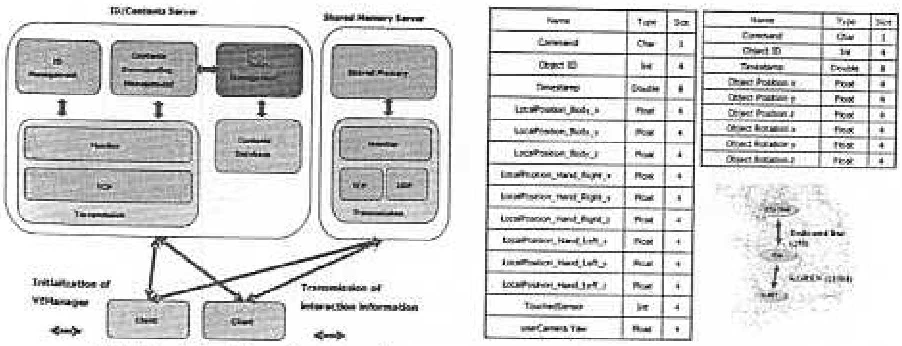
Since we concentrate on the events and messages in the limited area of VE, we can reduce the computation time of processing information as well as exchange information efficiently through the network.

2.3 Networking Interface: Supporting Real-time Interactions

In the networked VE for real-time interactions, Networking Interface enables to share the state information of users and the same VE information. There are many relevant examples to support real-time interactions, such as CAVERN, BrickNet, NPSNET, WAVE, MASSIVE-3, and so on [16][17].

In I-NEXT, the architecture of Networking Interface is composed of two elements: One is distributing the shared data and the other is sharing information of the states of users and a given VE. For supporting two requirements, Networking Interface is implemented by a client/server model which makes it easier to keep consistencies and relatively easier to implement. We exploit the function of shared database in QUNTA library, which is a cross platform adaptive networking toolkit used in the applications requiring intensive bandwidth and/or in delivering interactive data [18].

Figure 5 illustrates the architectures of two servers for supporting real-time interactions: One is ID/Contents server which issues ID and distributes contents to clients. The other is Shared Memory server which transmits shared state information to clients.



(a) Network Interface Configuration (b) Shared Data Formats

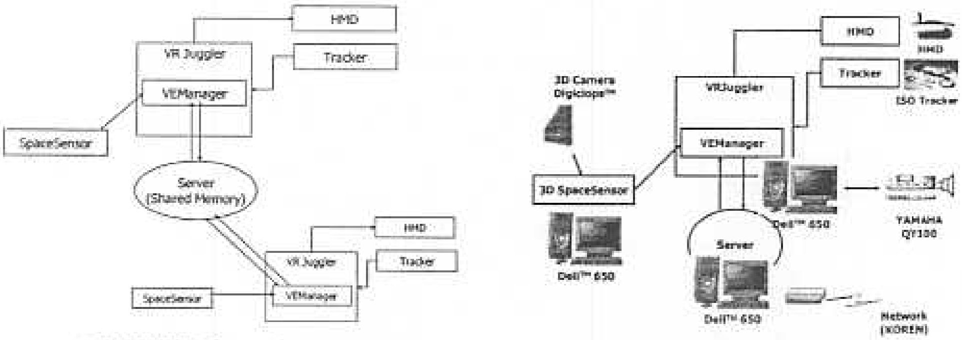
Figure 5. Network Interface Configuration and Shared Data Formats.

As shown in Figure 5(a), ID/Content server manages IDs of virtual objects and file paths through contents database. Index Management stores and controls data read from the database in the memory, Contents Downloading Management allows multiple users to download files, and ID Management delivers ID of contents. Shared Memory server is a kind of centralized file repository model which saves shared state information into a file. This transmits the states information of avatars (57 bytes) and objects (37 bytes) by using frequent state regeneration without writing data into the files.

3. Implementations and Experimental Results

To experience the interactive expressions over the network, we implemented an application of the proposed I-NEXT. The configruations of hardware and software are represented in

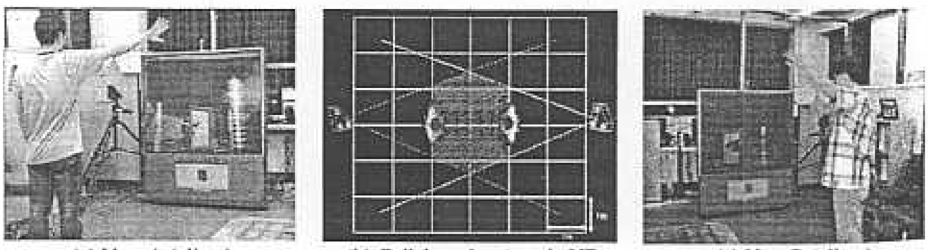
Figure 6. In hardware configuration, we deployed 3 workstations (Dell™ 650) for a server and a client, 3D camera (Digiclops™) for 3D vision based user interface, Ethernet card for networking interface, HMD/Tracker for immersion, and MIDI device for playing sound/music. In software configuration, we designed them as the way of hardware configurations in order to exploit the characteristics of I-NEXT. Moreover, we developed I-NEXT based on VR Juggler toolkit, and we used QUNTA library in the networking interface [19][20]. The relationship of each component is shown in Figure 6(a). We used SpaceSensor for the user interface. In SpaceSensor, it tracks the users' gestures of users and it makes users' gestures information for interactions and/or expressions. After SpaceSensor acquired user's gestures information, it is delivered to VEManager through IPC(Interprocess Communication). In VEManager, after it integrates and analyzes the gestures of both local and remote users, it updates the states of VE with the integrated and analyzed information and finally renders the whole VE scene. In the network configuration, it consists of 155 Mbps (ATM) KOREN (Korea Advanced Research Network) line between KJIST and ICU, a dedicated 2 Mbps (E1) line between ICU and DML. The latency is 20.2 ms, and available bandwidth is 1 Mbps.



(a) I-NEXT Software Components (b) I-NEXT Hardware Component

Figure 6. I-NEXT configurations

Figure 7 shows the result of the application using I-NEXT. We experimented the application in "ubiHome" (ubiquitous computing home testbed in KJIST) because we assumed the proposed I-NEXT could be used in the future home environment in order to help sharing a family love.



(a) User A (client) (b) Collaborative Area in VE (c) User B (client)

Figure 7. An Application of I-NEXT

This application was demonstrated in ICU DML exhibition in June, 2003. Through the exhibitions, the application attracted many media artists and designers. As shown in Figure 7, two users in the distance are able to participate in the same VE over the network. In the VE, they can interact with each other implicitly and/or explicitly through the avatars. For explicit interactions, user A and user B have to be in the colloborative area as shown in Figure 7(b). In this area, they are able to scratch the hidden layer on the virtual object as shown in Figure 7(a) and 7(c). For implicit interactions, user A and user B make the specific gestures. In the case of user A, he tries to change the background scene by moving his right arm from front to back. In the case of user B, he tries to give a special effect in VE by moving his left arm from bottom to top. As shown in Figure 7(b), we designed virtual space as 6m x 6m in order to measure the correspondences between real world and virtual world, where 1 square represents 1m x 1m. According to the Equation (3) and (4), users are located 4m from the virtual and real camera.

In this experiment, the participants were able to express their intentions interactively in real-time over the network. In the explicit interaction, however, they had a difficulty scratching the hidden layer on the virtual object due to the restriction in the resolution of the 3D camera. When users are in 3m~5m from the 3D camera, the minimum and maximum error is 0.04m and 0.07m, repectively. On the contrary, in the implicit interaction, the participants are enough to express their intentions through SpaceSensor. Therefore, we need to improve the accurracy of SpaceSensor to provide interactive expressions.

4. Discussions

In this paper, we proposed I-NEXT where users can express their intensions or emotions interactively in the networked VE through 3D vision-based user interface. I-NEXT enables users to express their intentions or emotions by the comfortable 3D vision-based user interface. Furthermore, it enables users to interact with virtual objects/environment and other users in real-time over the network. I-NEXT is designed to help developers or artists to test new types of expressions or interactions over the network. We hope that I-NEXT will provide a new type of testbed for the networked expressions and interactions through the 3D vision techniques. We are currently working on emotion analysis module of the I-NEXT, VE structure for consistent virtual environment management, etc.

References

[1] M. Weiser, "The Computer for the 21st Century," Scientific American, pp. 94-104, Sep. 1991
[2] S. Jang, S. Lee and W. Woo, "Research activities on Smart Environment," Magazine of the KITE, vol. 28, pp.1359-1371, Dec. 2001
[3] http://compcult.media.mit.edu/ccg/index.html
[4] http://www.fraunhofer.de/english/profile/institute/igd/igd_f_contact_01.html
[5] http://www.medialabeurope.org/hc
[6] http://www.mis.atr.co.jp/index.html
[7] Woontack Woo, Namgyu Kim, Karen Wong and Makoto Tadenuma, "Sketch on Dynamic Gesture Tracking and Analysis Exploiting Vision-based 3D Interface," in Proc. SPIE PW-EI-VCIP'01, vol. 4310, pp. 656-666, Jan. 2001
[8] S. Singhal, M. Zyda, "Networked Virtual Environments: Design and Implementation. Addison Wesley", 1999
[9] T. Horprasert, D. Harwood, and L.S. Davis, "A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection,"Proc. IEEE ICCV'99 FRAME-RATE Workshop, Kerkyra, Greece, September 1999
[10] Ahmed Elgammal, David Harwood, and Larry Davis, "Non-parametric Model for Background Subtraction," 6th European Conference on Computer Vision, Dublin, Ireland, June/July 2000.
[11] A. Elgammal, R. Duraiswami, D. Harwood and L. S. Davis "Background and Foreground Modeling using Non-parametric Kernel Density Estimation for Visual Surveillance", Proceedings of the IEEE, July 2002.
[12] Dongpyo Hong, Woontack Woo, "A Background Subtraction for a Vision based User Interface", Pacific-Rim Conference on Multimedia (PCM2003), *accepted*, 2003.

[13] http://www.aiwaz.net/Leonardo/
[14] Olof Hagsand, "Interactive Multiuser VEs in the DIVE System", IEEE Multimedia, Vol.3, No.1, pp.30-39, 1996
[15] Michael R. Macedonia, Michael J. Zyda, "A Taxonomy for Networked Virtual Environments", IEEE Multimedia, Vol.4, No.1, pp48-56, 1997
[16] Gurminder Singh, Luis Serra, Willie Png, and Hern Ng, "BrickNet:A Software Toolkit for Networked-Based Virtual Worlds", PRESENCE, Vol.3, No.1, pp19-34, Winter 1994
[17] Gurminder Singh, Luis Serra, Willie Png, Audery Wong, Hern Ng, "BrickNet:Sharing Object Behaviors on the Net", Proc. IEEE VRAIS '95, pp19-25, 1995
[18] He, E., Leigh, J., Yu, O., DeFanti, T. A., Reliable Blast UDP : Predictable High Performance Bulk Data Transfer, in *Proc. IEEE Cluster Computing, Sept, Chicago, Illinois*, 2002.
[19] Bierbaum, A., Just, C., Hartling, P., Meinert, K., Baker, A., CruzNeira, C., "VR Juggler: a virtual platform for virtual reality application development", Virtual Reality, 2001. Proceedings. IEEE , 13-17 March 2001, pp89 -96
[20] Dharanikota, S.; Maly, K., "QUANTA: quality of service architecture for native TCP/IP over ATM networks", High Performance Distributed Computing, 1996., Proceedings of 5th IEEE International Symposium on , 6-9 Aug. 1996. pp585 -594