A Gesture-Based Interface for Seamless Communication between Real and Virtual Worlds

Masaki Omata, Kentaro Go, Atsumi Imamiya

Department of Computer Science and Media Engineering, Yamanashi University
4-3-11 Takeda Kofu-shi Yamanashi-ken 400-8511, JAPAN
Tel & FAX: +81 55 220 8510
E-mail: {omata, go, imamiya}@metatron.esi.yamanashi.ac.jp

Abstract. This paper proposes a gesture-based direct manipulation interface that can be used for data transfer among informational artifacts. "Grasp and Drop (Throw)" by hand gestures allows a user to grasp an object on a computer screen and drop (throw) it on other artifacts without touching them. Using the interface, a user can operate some artifacts in the mixed reality world in a seamless manner, and learn this interaction style easily. Based on this interaction technique, we developed a prototype of presentation system using Microsoft PowerPoint, a wall size screen, computer screens and a printer. The presentation system with gestures allows a presenter to navigate through PowerPoint slides and transfer a slide from one computer screen to another. We conducted an experiment which evaluate the interaction style of gestures and analyzed the user's satisfaction with a questionnaire. The result shows that the overall mean of successful recognition is 96.9 %, and the learning of the system is easy.

1. INTRODUCTION

Mixed reality is a technology merging real and virtual worlds [Ohta and Tamura 99]. With this technology, users can integrate real world artifacts with virtual world artifacts. In mixed reality studies, one of the research issues is to design human interfaces that allow users to interact with real and virtual artifacts in a seamless manner. In the present human-computer interfaces, however, users must be conscious of a boundary between both worlds [Russell and Weiser 98].

With the use of virtual reality interfaces and other evolving techniques, virtual interfaces are becoming increasing realistic. The transition from virtual to real and vise versa is becoming so smooth that thin wall between these two worlds approaches transparency. We can go from real to virtual and back using simple gestures.

Interacting with artifacts in the mixed reality world requires easy to learn and use, spatially oriented tools. Since we used to use hand gestures to express spatial and temporal content, that is, use them to show three-dimensional relationships between objects and temporal sequences of events, it should be a key reason for using gestures in the mixed reality world to take advantage of this natural, intuitive manipulation and communication mode.

In this paper, we propose a new interaction technique based on hand gesture, which unifies the real and virtual worlds. Also, we present a prototype of a PowerPoint presentation application using hand gesture. Finally, for evaluating effectiveness of the system, we conducted an experiment on its gesture recognition and analyze user's satisfaction through a questionnaire. The result shows that our system is robust for hand gestures and users positively accept the system.

2. RELATED WORKS

The "Pick and Drop" system is a pen-based interaction system that allows a user to exchange information objects among computers [Rekimoto 97]. Using the system, a user can transfer data on his/her screen to another one by picking an icon on his/her own screen, and dropping it onto another screen. This interaction style is similar to our interaction style. "Pick and Drop" corresponds to grasping and dropping gesture, respectively. However, the main difference between [Rekimoto 97] and our style is that our system allows users to manipulate a real world artifact without touching and displaying it.

"FieldMouse" allows users to input a position on any flat surface (e.g., physical paper and wall) and to scan a barcode printed on the flat surface [Siio et al. 99]. Therefore, users can change a mode or a function with the barcode, and input a relative motion. However, it does not allow users to operate on a real world artifact without touching it and change a function without specific media like barcodes. On the other hand, our system allows user to change a mode and a function by hand gestures without specific media.

"Tangible bits" catches users' attention and bridges the gap between a cyberspace and a physical environment by coupling the bits with every day physical objects and architectural surfaces [Ishii and Ullmer 97]. Using the interface, users can manipulate virtual objects physically with graspable objects and ambient media in physical environments. Although this system uses movable bricks as physical handles, a user cannot transfer a virtual object among computer screens with the bricks.

These related studies show us an important issue of designing interface, that is, the need of physical feedback from object manipulation. Since gesture-based interaction may lose the feeling of manipulation, we provide sound-feedback for each gesture.

3. GESTURE INPUT BETWEEN THE REAL AND VIRTUAL WORLD

Our system provides users with a new interaction style, using hand gestures. We think that gesture is even powerful in mixed reality world when combined with other modalities such as direct manipulation and speech or sound. Users often perform hand gestures in the real world, e.g., can gesture toward remote artifacts for pointing without touching them. Accordingly, in our system users can operate remotely either on a virtual artifact on computer screen or on an artifact in real world.

3.1. Concept of the gesture input system

"Grasp and Drop (Throw)" gestures are the main operations in our system (see Figure 1 and 2). Users can grasp objects on a computer screen and move them to another one in a Local-Area Network. As a result, the users transfer objects from the source (computer) to the destination using. "Grasp" is an action of opening and then making a fist-like hand shape toward an artifact. "Drop (Throw)" is an action of opening the grasped hand toward another artifact. The range of real world artifacts that can be controlled by using the 3D gesture input modality is not only limited to the screens, printers and others of interconnected computers, but the system can deal with other real world artifacts or objects. When users want to transfer a document from a computer to another one, they make the grasp gesture for the document object on a computer screen then make the drop gesture toward another one (Figure 1). Likewise, users can transfer the document object from a computer to a printer in the same manner (Figure 2).

Our gesture system may be particularly useful for distributed collaboration such as meetings in a room. For example, if a participant wants to transfer a document object from his/her computer screen to another member’s computer screen for the purpose of sharing the document, he/she can just “grasp” the document object and throw it toward another screen.

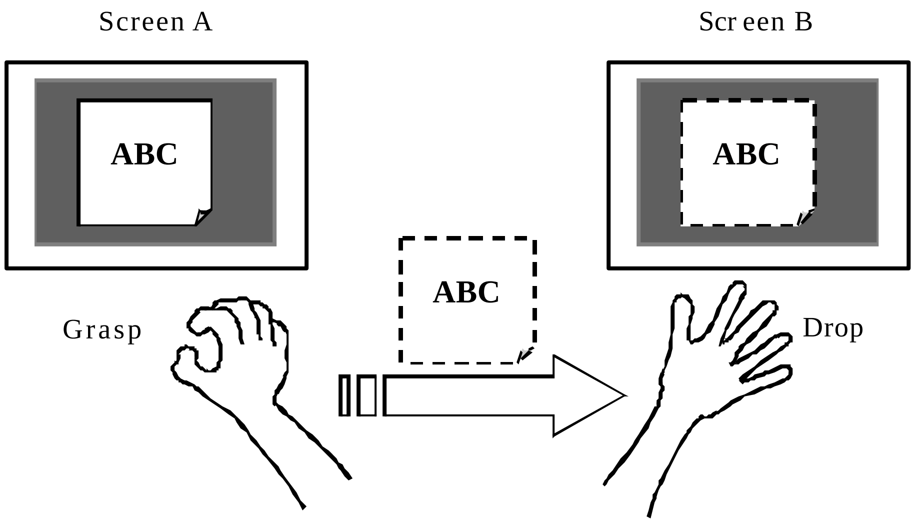

Figure 1. “Grasp and Drop (Throw)” action by hand gestures to transfer a document object to another screen.

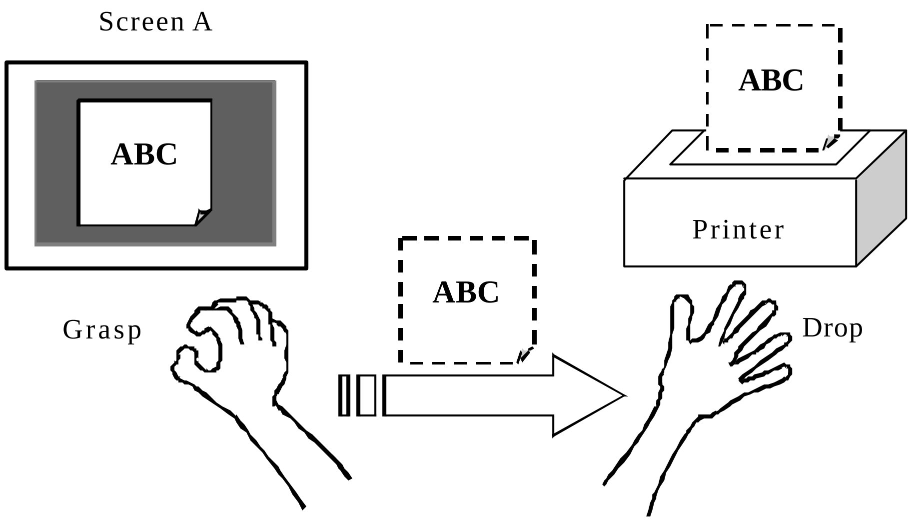

Figure 2. • Grasp and Drop (Throw)” action by hand gestures to transfer a document object to a printer.

3.2. Implementation

In the following subsections, we describe the design and the implementation of our system based on the concept mentioned above. We used the CyberGlove and the FASTRAK for direct and gestural interaction for 3D mixed reality world. The CyberGlove by Virtual Technologies Inc. includes 18 resistive-strip sensors for finger bend and abduction, thumb and pinkie rotation. The POLHEMUS

FASTRAK (3D tracker) permits six degree of freedom localization of hand position (x, y, and z) and orientation (pitch, roll and yow).
A key issue of implementing our gesture interface is how to deal with the position data of real world artifacts. We use the position data to implement the position-recognition of real world artifacts. It is posture-based gesture recognition, which is useful and easy to implement as a pattern classifier for gesture data and position data of artifacts.
In our implementation, a gesture consists of two postures (Figure 3). A posture is a snapshot of the starting or the end point of a gesture and is composed of a hand shape and a hand position. A hand shape consists of CyberGlove's data, and a hand position consists of the value of x, y and z coordinates relative to the origin.
Our recognition system refines and reduces the information from the raw data and facilitates interpretation on the broader context of information. Figure 4 illustrates the architecture of the system consisted of training and recognition components.
The training component extracts the feature of hand shape, i.e., posture, and position from the stream of raw hand data. That is, the raw hand data from the CyberGlove and the FASTRAK is classified into features of posture and orientation at each sample point. Currently, the posture features are opened and closed. The orientation features are extracted by calculating the relative positions between the hand and the artifact.
The recognition component parses the sequence of postures, and finally extracts the context of the gesture sequence. In Figure 5, first, sampled postures and orientations are compared with the training data, and identified with the posture class and orientation in training data, e.g., "open" hand or "close" hand toward an artifact (Figure 3). In the second phase, the sequence of postures and orientations are compared with the pre-segmented sequence of postures and orientation, and identified with a gesture class, e.g., "grasp" or "drop". Finally, in the third phase the sequence of gestures and orientation are compared with the pre-segmented sequence of gestures, and identified with the operation and motion, e.g., transfer operation is identified with "grasping" toward the document object and "dropping" it on the other screen. During the recognition process, the system provides users sound-feedback for each gesture.
We use the formulation of an algorithm of Sawada, et al. [Sawada et al. 98] for both training and recognition of posture and gesture sequence and estimation of the value of parameter. The algorithm calculates a mean and a standard deviation of sample data in the training phase using equation 1 and 2, and calculates the minimum distance between sample data and predefined data using equation 3.

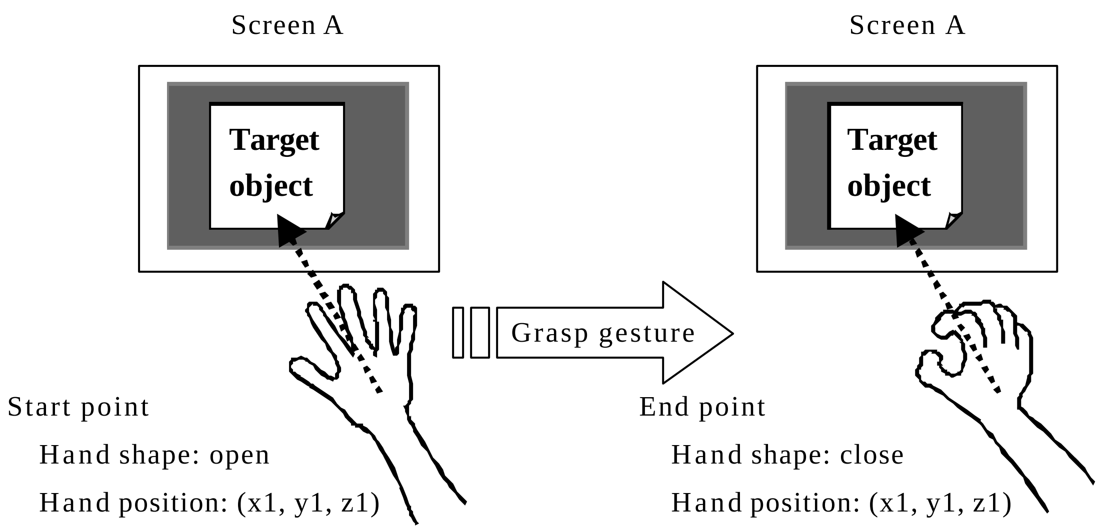

Figure 3. The Posture of grasp gesture.

$$E_a^P = \frac{1}{M}\sum_{i=1}^{M} V_a^{p_i}, \tag{1}$$

$$m_a^P = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(V_a^{p_i} - E_a^P\right)^2}, \tag{2}$$

where

E_a^P : Mean of training data of each item,

M : Times of training,

V : Value of user's input for training,

p_i : i-th sample of posture,

a : One of the CyberGlove data, or one of the FASTRAK data,

m_a^P : Standard deviation of training data of each item.

$${}^eP = \min\left\{\sum_a \frac{\left(V'_a - E_a^P\right)^2}{\left(m_a^P\right)^2}\right\}, \tag{3}$$

where

${}^eP$: Minimum distance between user's input and training data,

V' : User's input value.

4. PROTOTYPE

For evaluating our gesture recognition system, we build a prototype of the gestural input system to control a multiple screens presentation in Microsoft PowerPoint, and conduct experiments. It provides a gestural input means for presenter to navigate through PowerPoint slides and point or draw on multiple screens of PCs in a room.

Our system provides gestural functions for presenters in PowerPoint application to navigate among slides, and draw on the displayed slide. The presenter can navigate forward or backward through a series of slides by a grasp gesture of a slide on a screen and by moving it toward left or right direction. Printing operation is performed by grasping a slide and then dropping (or throwing) it toward a printer. Transferring slide among screens is performed by grasping a slide toward the slide and throwing it at the other screen. Other functions are pointing with the index finger, and marking by the drawing gesture with a pen.

5. EXPERIMENT

In order to evaluate our gesture input system, we conducted an experiment and administered a questionnaire to get subjective information on the subject's satisfaction.

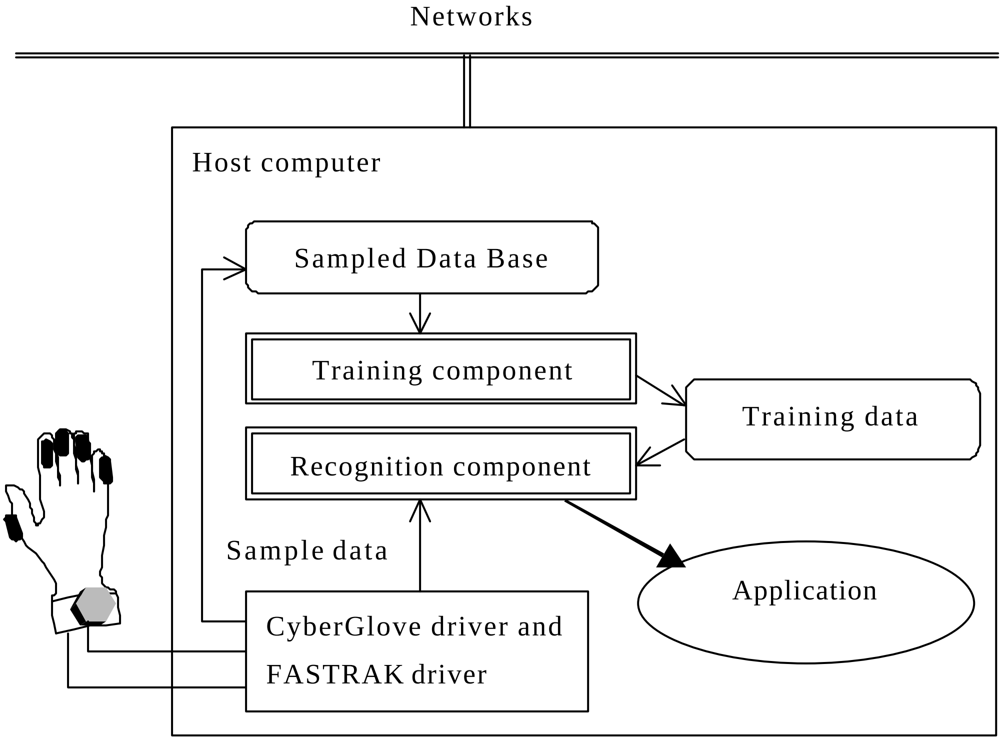

Figure 4. System architecture of gesture interface.

5.1. Procedures

We choose the screen control task of the presentation in PowerPoint with gestures. The subject's four operations are as follows: navigating either forward and backward through the slides on the main screen, printing it on a printer, or transferring it to other screens.

After all subjects have a brief introduction to the gesture system, and practice of presentation with the system, subjects input ten gesture data for each function are entered as training pattern of the system. Then subjects have five practices in each operation, and they became used to controlling the presentation task with gestures.

In each trial, the subjects are instructed to gesture for a specified operation. All trial data are recorded using a PC and videotape. While giving training pattern and trial by gestures, the subjects stood up toward artifacts in real world.

5.2. Apparatus

The experiment was conducted on three PCs connected with the CyberGrove, the FASTRAK, a printer, or a projector (Figure 6). The PCs ran on WindowsNT4.0 connected in a network. The PC-a was used for the gesture recognition system (with CyberGlove and FASTRAK) with sound for feedback of confirmation of grasping, and ran PowerPoint presentation with projector. The PC-b was used for the presentation of PowerPoint, and the printer was connected with PC-c for output of a slide.

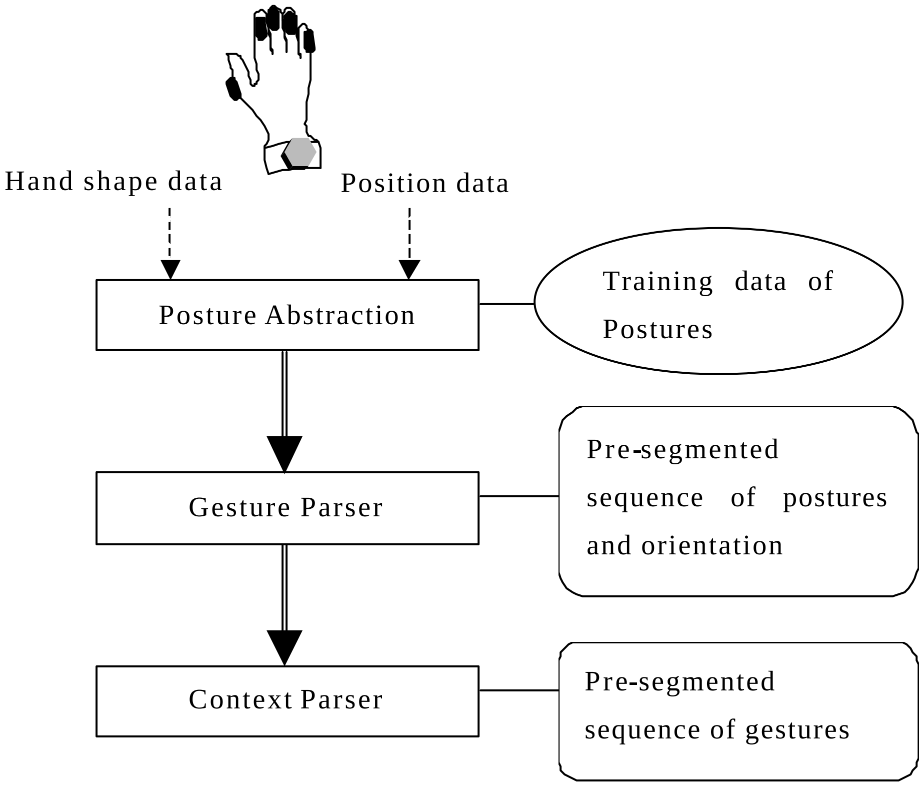

Figure 5. Diagram of gesture recognition.

5.3. Design

A within subjects, repeated measure design was used. All subjects performed experiment of four operations. For each operation, the subjects performed 20 blocks of trials. For each block, the presentation order of four operations was random. Each block consisted of 4x2 trials. The experiment consisted of a total of 160 trials per subject. A questionnaire designed to elicit subjects' preferences of and satisfaction with the system, was completed by subjects at the end of the experiment. We used a part of the QUIS of Shneiderman [Shneiderman 98] (Table 1 and 2). Subjects were asked to rate each question on 1- 9 and NA (Not applicable) scales.

5.4. Subjects

7 subjects participated in the experiment, who were all graduate and undergraduate students of our University. The subjects had more than three times experiences of the PowerPoint presentation.

5.5. Results

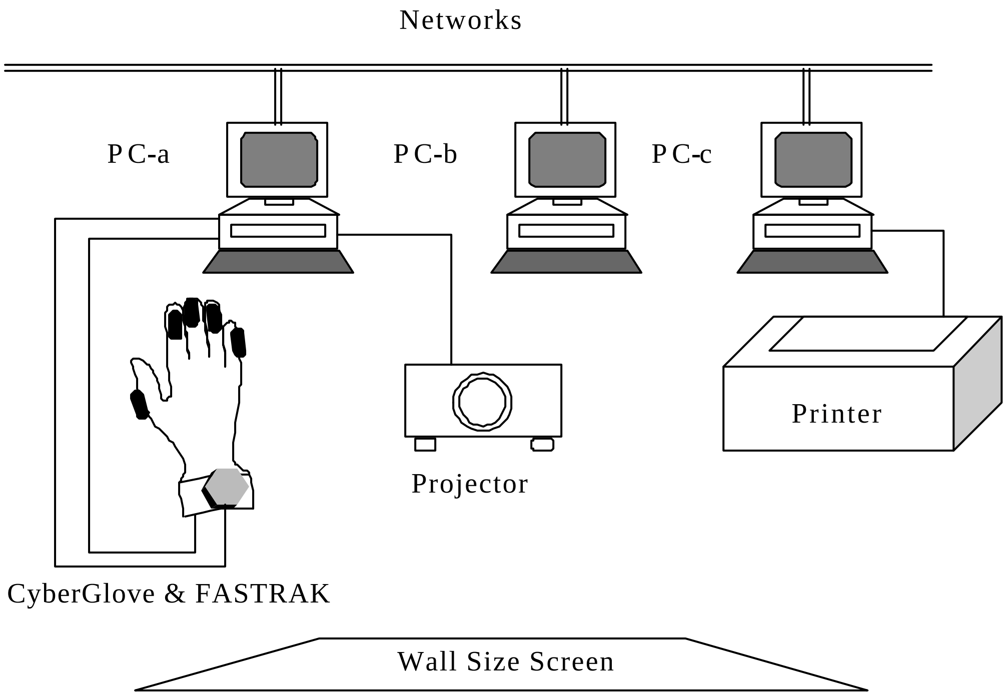

Figure 6. Apparatus of experiment.

Table 3 summarizes the results of the recognition rate for each subject and function. The overall mean of success was 96.9 %. The best case is 100.0 % recognition rate.the subject had no errors at all. The worst case is 93.1 % recognition rate. In the worst case, the errors are slightly different from range of variance of training data. As we can see from the table, our system is robust for hand gestures.

The box plot in figure 7 shows how much individuals satisfy with the gesture system of the presentation. The average score of Q1-1, Q1-3 and Q1-4 is 7.86. This provides some evidence that users positively accept the system and/or become familiar with it.

Q2-1, Q2-2, Q2-4 and Q2-9 have high average scores in the questions (Figure 8). In other words, this suggests that it was easy for subjects to learn the use the system.

Table 1. List of Questions on “Overall User Reactions”.

	Overall reactions to the system	
Q1-1	terrible (1) -wonderful (9)	NA
Q1-2	frustrating (1) - satisfying (9)	NA
Q1-3	dull (1) - stimulating (9)	NA
Q1-4	difficult (1) - easy (9)	NA
Q1-5	inadequate power (1) - adequate power (9)	NA
Q1-6	rigid (1) - flexible (9)	NA

Table 2. List of Questions on "Leaning".

	Learning		
Q2-1	Learning to operate the system		
		difficult (1) - easy (9)	NA
Q2-2	Getting started		
		difficult (1) - easy (9)	NA
Q2-3	Learning advanced features		
		difficult (1) - easy (9)	NA
Q2-4	Time to learn to use the system		
		difficult (1) - easy (9)	NA
Q2-5	Exploration of features by trial and error		
		discouraging (1) - encouraging (9)	NA
Q2-6	Exploration of features		
		risky (1) - safe (9)	NA
Q2-7	Discovering new features		
		difficult (1) - easy (9)	NA
Q2-8	Remembering names and use of commands		
		difficult (1) - easy (9)	NA
Q2-9	Remembering specific rules about entering commands		
		difficult (1) - easy (9)	NA
Q2-10	Tasks can be performed in straightforward manner		
		never (1) - always (9)	NA
Q2-11	Number of steps per task		
		too many (1) - just right (9)	NA
Q2-12	Steps to complete a task follow a logical sequence		
		never (1) - always (9)	NA
Q2-13	Feedback on the completion of sequence of steps		
		unclear (1) - clear (9)	NA

6. CONCLUSION

In summary we described our gesture-based interface that allows a user to transfer data from a computer screen to another artifacts. Using the system, user can operate artifacts in real and virtual worlds without being conscious of boundary between two worlds.

Furthermore, in order to evaluate of its effectiveness, we conduct an experiment to test the effectiveness of our gesture recognition system. We also administered questionnaire for satisfaction

Table 3. Result of recognition rate. [%]

	Next	Previous	Print	Transfer	All
subject 1	95.0	92.5	95.0	90.0	93.1
subject 2	100.0	100.0	90.0	92.5	95.6
subject 3	100.0	100.0	100.0	100.0	100.0
subject 4	95.0	92.5	95.0	95.0	94.4
subject 5	100.0	95.0	95.0	97.5	96.9
subject 6	100.0	100.0	97.5	100.0	99.4
subject 7	100.0	100.0	95.0	100.0	98.8

analysis. High average scores on learning of our system shows that users can use our system easily.

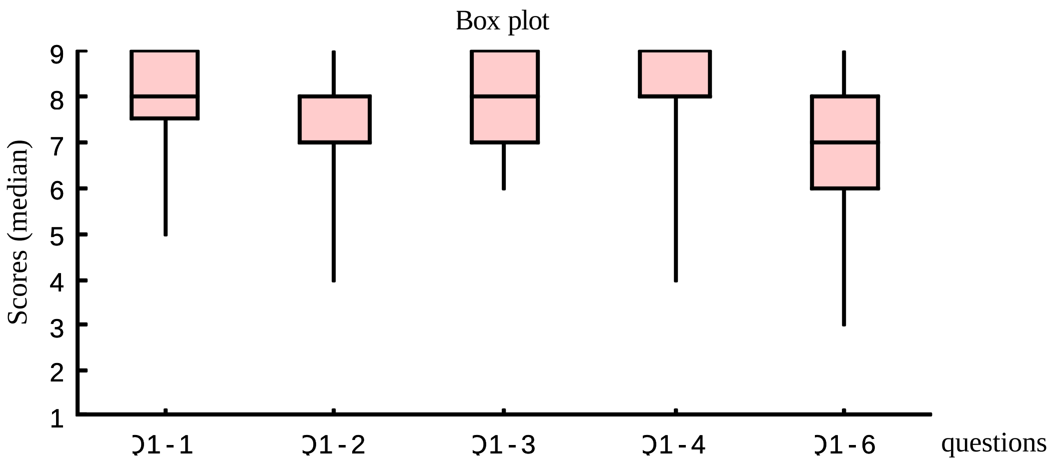

Figure 7. Box plot of questionnaire's score on overall user reactions.

Overall, participants and our experiences with the system have been positive. In our next design stage, we are planning to provide artifacts as icons in desktop metaphor back to real world artifacts. This provides users natural interface that allows the users to see the real world artifacts and to instruct them directly.

We also need to improve the gesture recognition system in order to reduce recognition errors. The recognition system should correct training data of the gesture in real time while a user performs gestures. As a direct result, the system will be able to deal with changes of user's gestures. On the other hand, the system cannot differentiate artifacts from ones that are on the same direction. This is because the system uses the same position data, which are recorded as direction of the user relative to the artifacts. Using acceleration of user's motion, the system can extract start and end point from the user's motion. As a direct result, the system can dump an approximate posture in the way of user's movement.

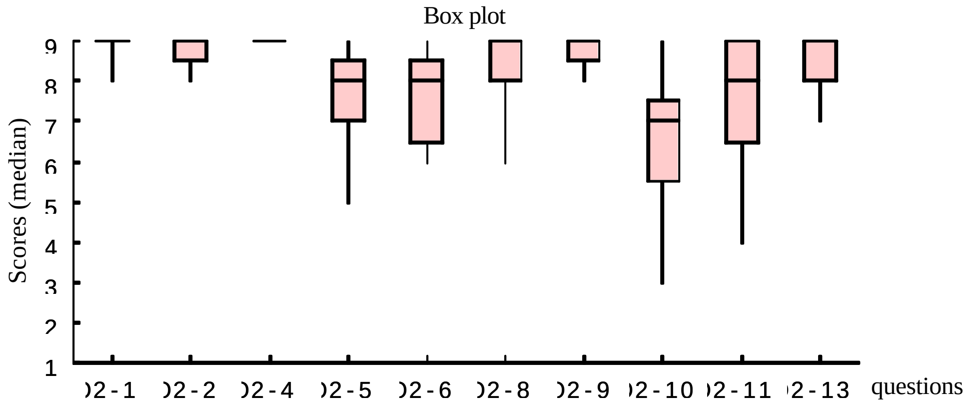

Figure 8. Box plot of questionnaire's score in learning.

We also plan to take the concept of two-handed input [Bolt and Herranz 92, Nishino et al. 97] into our gesture interface. People use two hands in performing tasks in everyday life; such as painting, cutting bread, driving a car, specifying a shape and range, and so on. We believe that study of two-handed input for 3D operations in the mixed reality will result in additional effectiveness and new classes of interactions.

ACKNOWLEDGEMENTS

This research was supported in part by the Telecommunication Advancement Organization of Japan.

REFERENCES

[Ohta and Tamura 99] Y. Ohta, H. Tamura, (eds), *Mixed Reality*, Springer-Verlag, 1999.

[Russell and Weiser 98] D. M. Russell, M. Weiser, The Future of Integrated Design of Ubiquitous Computing in Combined Real & Virtual Worlds, *Proceedings of the conference on CHI '98*, Los Angeles, USA, April 18-23, 1998, pp 275-276.

[Rekimoto 97] J. Rekimoto, Pick-and-drop: a direct manipulation technique for multiple computer environments, *Proceedings of the 10th annual ACM symposium on User interface software and technology*, Banff, Canada, October 14-17, 1997, pp 31-39.

[Siio et al. 99] I. Siio, T. Masui, K. Fukuchi, Real-world interaction using the FieldMouse, *Proceedings of the 12th annual ACM symposium on User interface software and technology*, Asheville, USA, November 7-10, 1999, pp 113-119.

[Ishii and Ullmer 97] H. Ishii, B. Ullmer, Tangible bits: towards seamless interfaces between people, bits and atoms, *conference proceedings on Human factors in computing systems*, Atlanta, USA, March 22-27, 1997, pp 234-241.

[Sawada et al. 98] H. Sawada, S. Hashimoto, T. Matsushima, A Study of Gesture Recognition Based on Motion and Hand Figure Primitives and Its Application to Sign Language Recognition, *Transactions of Information Processing Society of Japan*, 39(5), 1998, pp 1325-1333, (in Japanese).

[Shneiderman 98] B. Shneiderman, *Designing the User Interface Third Edition*, Addison-Wesley, 1998.

[Bolt and Herranz 92] R. A. Bolt, E. Herranz, TWO-HANDED GESTURE IN MULTI-MODAL NATURAL DIALOG, *Proceedings of the fifth annual ACM symposium on User interface software and technology*, Monteray, USA, November 15-18, 1992, pp 7-14.

[Nishino et al. 97] H. Nishino, K. Utsumiya, D. Kuraoka, K. Yoshioka, K. Korida, Interactive two-handed gesture interface in 3D virtual environments, Proceedings of the ACM symposium on Virtual reality software and technology, Lausanne, Switzerland, September 15-17, 1997, pp 1-8.