Implementation and Evaluation of "Just Follow Me": An Immersive, VR-Based, Motion-Training System

Article *in* Presence Teleoperators & Virtual Environments · June 2002

DOI: 10.1162/105474602317473240 · Source: DBLP

2 authors:

Ungyeon Yang
Electronics and Telecommunications Research Institute

Gerard Jounghyun Kim
Korea University

Ungyeon Yang
uyyang@postech.ac.kr

Gerard Jounghyun Kim
Virtual Reality Laboratory
vr.postech.ac.kr
Department of Computer Science and Engineering
Pohang University of Science and Technology (POSTECH)
San 31 Hyoja-dong,
Pohang, Kyungbuk,
Korea 790–784

Implementation and Evaluation of "Just Follow Me": An Immersive, VR-Based, Motion-Training System

Abstract

Training is usually regarded as one of the most natural application areas of virtual reality (VR). To date, most VR-based training systems have been situation based, but this paper examines the utility of VR for a different class of training: learning to execute exact motions, which are often required in sports and the arts. In this paper, we propose an interaction method, called *Just Follow Me* (*JFM*), that uses an intuitive "ghost" metaphor and a first-person viewpoint for effective motion training. Using the ghost metaphor (GM), JFM visualizes the motion of the trainer in real time as a ghost (initially superimposed on the trainee) that emerges from one's own body. The trainee who observes the motion from the first-person viewpoint "follows" the ghostly master as closely as possible to learn the motion. Our basic hypothesis is that such a VR system can help a student learn motion effectively and quickly, comparably to the indirect real-world teaching methods. Our evaluation results show that JFM produces training and transfer effects as good as—and, in certain situations, better than—in the real world. We believe that this is due to the more direct and correct transfer of proprioceptive information from the trainer to the trainee.

1 Introduction

Training has been considered to be one of the most natural application areas of virtual reality (VR) (Acchione & Psotka 1993; Badler 1996; Bowman, Wineman, Hodges, & Allison, D., 1999; Cromby, Standen, & Brown, 1996; D'Cruz, Eastgate, & Wilson, 1997; Emerson & Revere, D., 1997; Youngblut, 1998). Most VR-based training systems to date are oriented toward learning a sequence of discrete reactive tasks; that is, "training" occurs first simply by exposing and immersing the user into a virtual environment (with various situation scenarios) that is otherwise difficult to experience in the real world. The goal is to train and test the trainee to select the right type of action in a demanding situation rather than to teach him/her how it is performed kinesthetically (Everett Wauchope, & Perez-Quinones 1998; Hodges et al., 1995; Jayaram, Wang, & Jayaram, 1999; Johnson, Rickel, Stiles, & Munro, 1998; Shawver, 1997; Rickel, & Johnson, 1999; VR Techno, 1998; Wilson, 1994). Even though these types of training systems do not involve the exact following of limb motions, they often require navigation and spatial awareness. Thus, in addition to the effect of trying it out beforehand in a similar environment, im-

Presence, *Vol. 11, No. 3, June 2002,* 304–323

mersive VR is expected to give the trainees an improved frame of reference compared to training with a desktop-based system (Pausch, Proffitt, & Williams, 1997).

This paper discusses the utility of VR for a different class of training: learning limb motion profiles, which is required in sports, dance, and arts (such as for a golf swing, martial arts, calligraphy, sign language, and so on). Our central concept behind VR-based motion training is called *Just Follow Me* (*JFM*), and it uses an intuitive interaction method called the *ghost*. Through the ghost metaphor, the motion of the trainer is visualized in real time as a ghost (initially superimposed on the trainee) moving out of one's body. The trainee, who sees the motion from the first-person viewpoint, is to "follow" the ghostly master as close (and/or as quickly) as possible. Such an interaction is only possible with VR and strives to provide matching sensorimotor feedback especially between the visual and proprioceptive cue. The training process can be facilitated further by showing other guidance cues (such as the master's trail or the third-person view) and performance feedback (indication of how well the trainee is following), and by adjusting the learning requirements (relaxation of accuracy goals on restricting the motion's degrees of freedom).

We hypothesized that such a VR system could help a student learn motion thoroughly and quickly compared to the usual indirect teaching methods (such as watching the master and imitating it).

We conducted the following experiments to evaluate the usability and training effect of the interaction method of JFM. We organized four groups of test subjects according to the type of the learning environment (ghost metaphor-based VR versus real-world indirect training) and by the type of motion characteristic (slow versus fast). All subject groups were asked to follow the same set of motion profiles (in an increasing level of difficulty or degrees of freedom) and were tested using the same tracking devices for measuring the respective accuracy of the learned motion.

Our evaluation results show that JFM, even with non-ideal hardware setup, produced training and transfer effects as good as—and, in certain situations, better than—the real one. We believe that this is due to the more direct transfer of proprioceptive information from the trainer to the trainee; that is, less effort is required with the first-person viewpoint to put oneself in the trainer's shoes. It was also found that, for relatively long-range and high-frequency motion profiles (particularly in the vertical direction), JFM did not perform well, possibly because of the rather heavy HMD that made changing viewpoints uncomfortable. Thus, when reinforced and augmented with presence cues, more-robust tracking, lighter and full-featured HMDs, and rich informative graphics and images, we safely conclude that VR-based training methods will be an attractive alternative to the traditional "trainer-in-residence" or video-based method for learning motor skills.

This paper is organized as follows. First, we review other related research in general and VR-based motion-training systems. We also investigate other similar approaches to interaction for motion guidance. Section 3 explains the central concept of JFM and proposes a general architecture for the VR-based motion-training system and introduces a sample implementation of JFM applied to oriental calligraphy. Section 4 gives details of the usability test conducted to verify the training and transfer effect of the proposed system, and its results. Finally, to conclude the paper, we discuss the probable reasons behind the effectiveness and shortcomings of the VR-based motion-training system, and comment on the on-going extension to current work.

2 Related Work

2.1 General Motion-Training Systems

Motion training can be modeled as a process of transmitting motion information from the trainer to the trainee through a series of interaction by some communication media (See figure 1.) Books and videos have been popular forms for such transmission media, and recently the large storage capability of DVDs and increased computing power have allowed richer and more-organized multimedia content for training and education with text, voice, short video, and images. Despite this increased interactivity, the effect of such indirect training is questionable, especially for motion train-

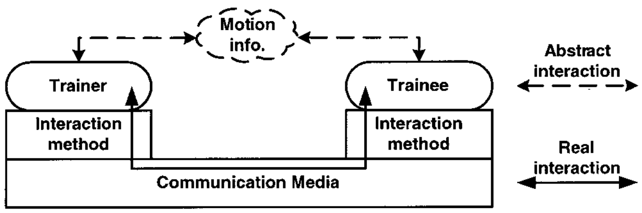

Figure 1. *General model of motion training.*

ing, because the trainee must interpret a large part of the implicit motor control knowledge and evaluate oneself. As in any training or learning process, the interaction should take the form of a two-way communication as far as possible for immediate performance feedback and correction. It is noteworthy that present educational trends are moving toward group and collaborative learning. Thus, it is still quite difficult to surpass the good old direct "trainer-in-residence" mode of teaching.

2.2 VR-Based Motion-Training System

As a technology that can provide real-time two-way communication with a multitude of interaction methods, VR-based training remains a viable alternative to the expensive and difficult "direct" learning methods. To date, most VR-based training systems have been situation based; that is, "training" occurs simply by exposing and immersing the user in a virtual environment that would otherwise be difficult to experience in the real world. Perhaps the most famous example is the NPSNET/SIMNET/Medisim, a network-based simulation for tactical military training (Badler et al., 1996; Macedonia, Zyda, Pratt, Barham, & Zeswitz, 1994). Others include battieship fire escape training (Everett et al., 1998), treatment for fear of flying (Hodges et al., 1995), and machine operation training (Johnson et al., 1998). One can easily realize that these systems are mostly situation-based training for decision-making; the user is expected to make decisions to perform a series of actions, and the particular motion is not very important. For example, in the hostage-situation resolution training system developed by the Sandia National Laboratory (Shawver, 1997), it is important for the trainee to "shoot" the hostage-taker in case the latter does not agree to surrender in a confrontation. In training systems for earthquakes (VR Techno, 1998), it is important for the trainee to "lock" the gas valve before running for cover. Sometimes, the task may not be reactive, as in the case of the "virtual factory" (Rickel & Johnson, 1999). In the VR-based product assembly simulation systems (Jayaram et al., 1999; Wilson, 1994), the system computes and simulates the exact collision-free assembly sequences, the associated paths/orientations of the parts, and even performs reachability analysis. However, they do not address the required human motions for assembly. The VET (Virtual Environment for Training) demonstrates an interesting use of an AI animated agent for step-by-step guidance and training of machine-running procedures (Rickel & Johnson, 1999).

Some motion-training systems using VR have been reported in the rehabilitation domain (Holden, Todorov, Callahan, & Bizzi, 1999; Kuhlen & Dohle, 1995; Todorov, Shadmehr, & Bizzi, 1997; VMW, 1997). For instance, Holden et al. have developed a VE-based motor-training system, similar to JFM (third-person viewpoint, motion trail visualization), to enhance rehabilitation in patients with neurological damage such as stroke or brain injury. In their work, they mainly considered the motion-training effects of the VE and augmented feedback (especially haptic) on injured patients, a situation somewhat different from general motion training (as patients generally knew what to do but rather were physically incapable). Combined with other investigation, their research focus was on constructing a distributed neurological model responsible for learning motor skills. The model suggests, among many things, more-direct stimulations of the spinal modules or muscle activation (for instance, by haptic devices) is a good strategy for reviving the once-disabled motor capability. Jack et al. (2001) have developed and evaluated a force for rehabilitating hand functions in stroke patients. Although demonstrating the training effect, the work concentrated more on faithful reproduction of hand/finger forces and considered the effect of using haptics only (versus the combined use with realistic/immersive visual cue and first-person viewpoint). For VR-based motion training for any VE in general, providing matching sensory modalities would be very important (Graniano, 1999; Magill, 2000; Yokokohji, Hollis, & Kanade, 1999). Yokokohji et al. has addressed this issue using augmented reality systems, and this work concentrated much on the correct registration of the virtual objects and tracking of human body parts in the external world, so that the augmented reality-based training system can fully utilize its strengths in providing the highly matched sensory modalities (for example, objects being in real scale and the correct and natural visual and haptic cues of one's limbs).

2.3 Using a Semitransparent Object as an Interaction Metaphor

VR systems often employ semitransparency effects to avoid occlusion and increase recognizability (and sometimes its relative depth) of important objects in a crowded scene (Zhai, Buxton, & Milgram, 1996). An interface similar to the ghost metaphor introduced in this paper is reported in a system called CAREN (VMW, 1997) that was developed out of a joint European ESPRIT research program. The purpose of CAREN is to train and rehabilitate patients to overcome balance disorders. A patient standing on a moving force plate must practice staying in balance by looking at the avatar (in a third-person viewpoint displayed on a large projection screen in front), which represents the patient. Transparent boxes bounding the avatar's limbs represent the correct posture/motion. Although the idea of using ghostly boxes is similar to our approach, their technical emphasis seems to be in motion capture and real-time computation of remedial postures based on exact biomedical data, rather than in technology for effective interaction. The superimposition of bounding boxes was also used in a system called ARGOS, a system for "tele-programming" a robot (Rastogi, 1996). In ARGOS, a wireframe bounding box is overlaid on a remote manipulator seen through a camera system at the home site. The user can program the remote manipulator by controlling the wireframe robot.

3 The "Just Follow Me" Method

The central concept of Just Follow Me is the use of the first-person viewpoint (egocentric view), which is the main ingredient of the VR systems (See figure 2). Therefore, unlike CAREN or the VR-based rehabilitation systems of Jack et al., the display for JFM requires an immersive display like an HMD; otherwise, the trainee is not able to see the ghost properly (as with a monitor display, for instance, wherein the user sees one's own limb at a distant location and violates the modality consistency requirement).

3.1 The Ghost Metaphor: Concept and Goals

The idea behind the ghost metaphor is straightforward and is illustrated in figure 2. The motion of the trainer is visualized in real time as a ghost (initially superimposed on the trainee) emerging from a trainee's body. The trainee, who sees the motion from the first-person viewpoint, is to "follow" the ghostly master as closely as possible (in regards to both timing and position/orientation). Such an interaction, which takes advantage of the first-person view of the master's motion in real time (see figure 2), is only possible with VR.

The training process can be facilitated further by showing other guidance cues (the master's trail, a third-person viewpoint) and performance feedback (indication of how well the trainee is following), and by adjusting the learning requirements (relaxation of accuracy goals, restricting the motion's degrees of freedom).

In the usual learning based upon the third-person viewpoint, the trainee must cognitively convert the reference frame of the motion to one's own and scale the motion parameter values as well. In a relatively fast motion sequence, the third-person viewpoint has difficulty performing this task on the fly. The student must rely on short-term memory to reproduce the motion. However, this is not to say that a third-person viewpoint is not needed. On the contrary, it should be quite useful for observing the whole body motion, which is otherwise not entirely visible from one's own viewpoint.

The ultimate goal of the ghost metaphor is to provide

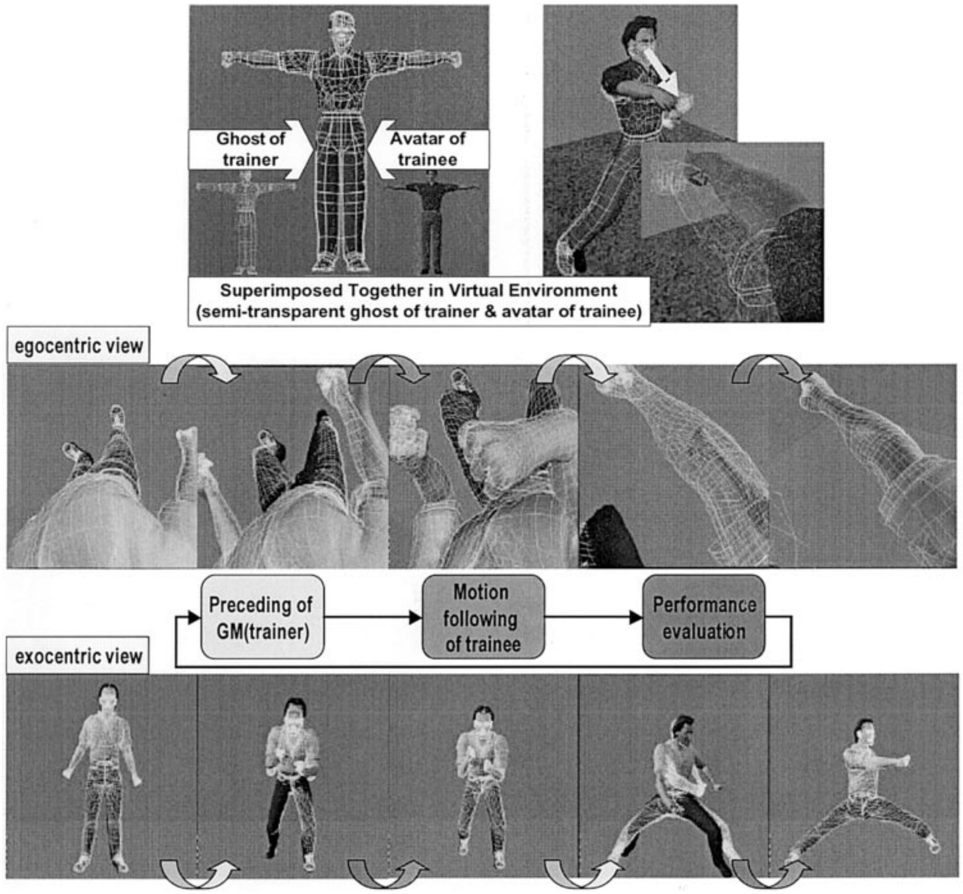

Figure 2. *Interacting with the ghost metaphor in the JFM system.*

matching sensorimotor feedback among different modalities, namely the visual and proprioceptive. For now, we excluded the haptic modality for the following reasons. The type of motions we consider are mostly "free"; that is, there is minimal interaction between the human limbs and the external world except at few contact instants and locations (for instance, swinging a baseball bat, performing a Tae-Kwon-Do maneuver, or learning a dance step); thus, the force feedback plays little role in shaping the motor control knowledge. If we were to consider a motion like rowing, consideration of the haptics would become very important. Even for "free" motions, it is conceivable to use haptic devices to simulate a force field and prevent trainees from making wrong motions, analogs to a trainer physically correcting a trainee's motion in real life. According to the motor-learning literature, too much use of such training methods can actually produce negative transfer effect because such feedback will no longer exist when the motion is applied in actuality (Schmidt, 1991).

3.2 Motion Evaluation and Guidance

In addition to the motion itself, performance data (online information concerning the trainee's performance) is important in effective motor learning (Schmidt, 1991). Many performance measures are possible: accuracy based (such as position/orientation difference, timing difference, and number of oscillations) and speed based (such as task completion time).

To facilitate the learning process, in addition to performance data, other guidance cues and adjustment of

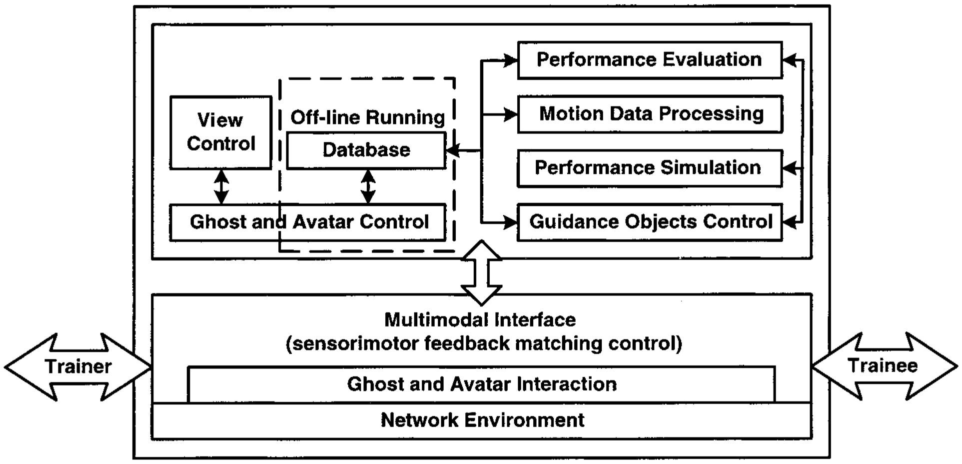

Figure 3. *A possible architecture for a VR-based training system.*

the learning requirements (such as relaxation of accuracy goals and restricting the motion's degrees of freedom) are possible. For instance, some conceivable examples include a curvilinear or volumetric motion trace, the third-person view, colored marks at the critical points of motion, directional arrow (vectors with both directions and magnitudes), textual and voice guidance, and alternative and simultaneous third-person viewpoints. In addition, a very natural extension to providing such guidance features is the use of force feedback for motion guidance. Such a haptic guidance can be both active and passive: an active haptic interface would attempt to correct the trainee's motion, whereas a passive haptic guidance might exist as a virtual wall that physically limits the range of the trainee's motion. Although conceptually intuitive, such physically guided training for types of motion in which the haptic sense would be missing is known to cause a negative transfer effect in actual application. Thus, it is advisable to use such a teaching technique only sparingly (Schmidt, 1991).

3.3 Architecture for VR-Based Motion Training

Based on the features outlined in previous sections, we have devised an architecture for a VR-based motion-training system, as shown in figure 3. The bottom portion of the figure shows the essential part of the system, a virtual environment consisting of a trainer and a trainee (possibly geographically separated but connected through the network), in which the training is conducted using the ghost metaphor and avatar (trainers and trainees).

In addition to this system, modules for online motion evaluation, other auxiliary motion guidance objects, and motion retargeting can be added (shown in the upper right of the figure). The trainer can be replaced by a ghost avatar that is animated with motion-capture data previously retargeted for many different body sizes offline. For online training (for example, motion profile demonstrated by the trainer in real time), an online retargeting module may be required (Baek, 2001; Choi & Ko, 1999).

Even though JFM basically operates in the first-person view mode, views from other angles can still be useful from time to time, for instance, in understanding the overall body posture (versus limited first-person view that can contain only parts of the body) (Blanz, Tarr, Bülthoff, & Vetter, 1996; Bülthoff, Edelman, & Tarr, 1994; Toussaint, 2000; Yang, Lee, Lee, Bok, & Han, 2002). A view control module, thus, is added in the proposed architecture to supplement the basic first-

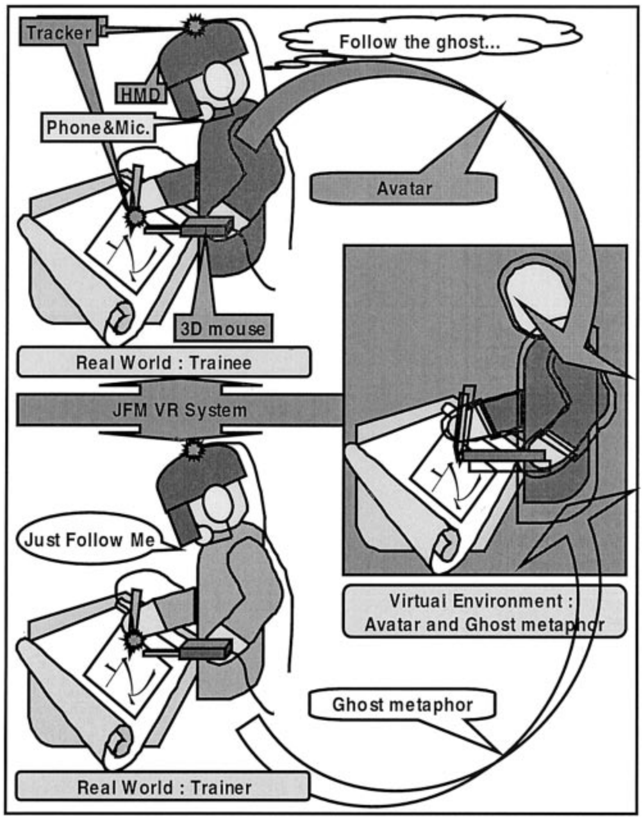

Figure 4. *Illustration of the ghost metaphor for motion training (VR calligraphy education system, first implementation of JFM).*

person view for an effective observation of the motion (or the surrounding environment) to be learned.

3.4 A Sample Implementation: JFM Calligraphy

As a proof of concept, we have implemented an oriental calligraphy training system using the JFM approach. (See figures 4 and 5.) Oriental calligraphy requires specific postures and movements of the brush-holding arm and hand to create aesthetic characters.

JFM calligraphy was implemented using two SGI Indigo2 Impact workstations (one as a rendering client and the other as a sensor server connected via CORBA), four Polhemus FASTRAK six-DOF sensors (for head and calligraphy brush tracking), two Logitech 3D mice (for other system-related input), and two HMDs (Virtual Research System's VR4 and Sony Glasstron PLM-A55). In addition to the master's ghost, a swept volume and acceleration vector were made available as an auxiliary guidance feature, and the online performance evaluation/feedback was not implemented. Figure 5 shows instances of the student attempting to follow the ghostly brush of the calligraphy master.

4 Evaluation of JFM

Our working hypothesis was that a VR-based motion-training system such as JFM would help students learn motion as quickly and efficiently as indirect teaching methods (such as watching the recorded master's motion and imitating it). We conducted a usability test to evaluate and verify the training effect of JFM and assess its usefulness (Helander, Landauer, & Prabhu, 1997; Hix & Hartson, 1993).

4.1 The Basis of Experiment Design

The experiment was designed with help from the human factors group of the industrial engineering department at our university to answer the following three questions related to the utility of a JFM- or VR-based motion-training system.

- Does it indeed provide a better or as good a frame of reference and lighter cognitive load for conveying motion-related information, compared to indirect methods? We compare and observe how well the trainee can follow the trainer's motion in the respective environment with regard to both position (and orientation) and timing.
- Are some types of motion relatively less (or more) suitable for training in VR-based motion-training systems considering the limitation of VR devices? Although some VR features may be useful in a training medium, limitations with the FOV of the HMD, tracking accuracy and range, and such ergonomic aspects as the weight and effect of the cables can also offset such advantages. We compare a trainee's motion following performance in the VR

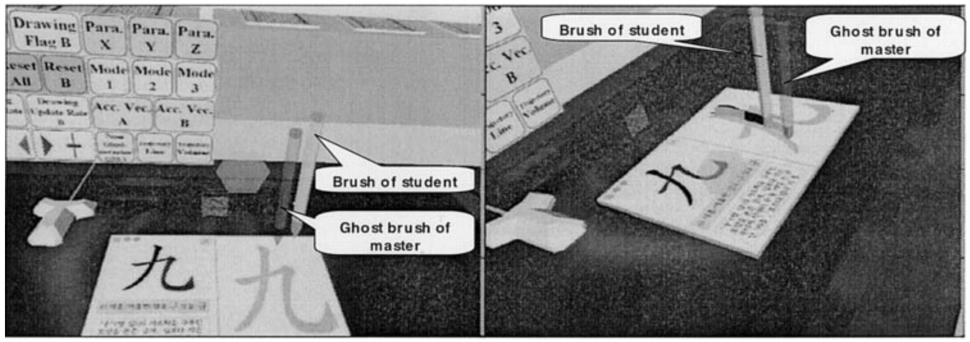

Figure 5. *A trainee's view of the "Just Follow Me" virtual calligraphy.*

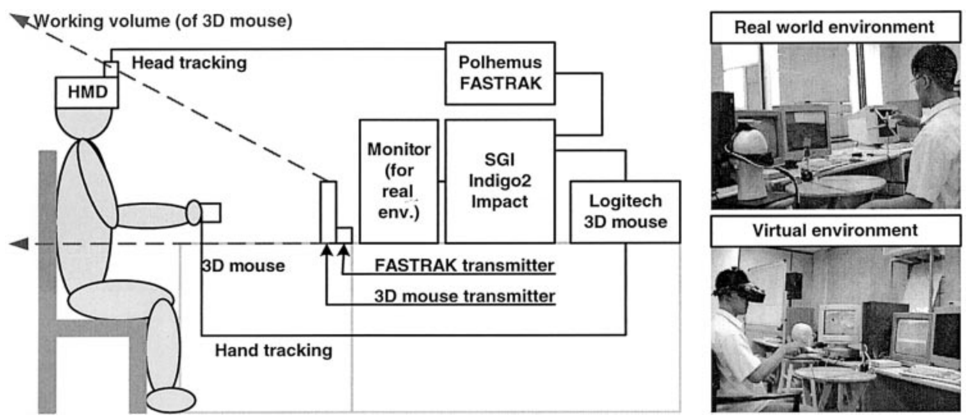

Figure 6. *Experimental environment setup.*

environment using different motion profiles and attempt to link the motion parameters to the device characteristics.

- Does it have a better or as good a transfer effect, compared to the indirect methods; in other words, does the motion learned in the VR environment transfer well when practiced in the real world? It is one matter to produce a system that is easy and natural for the trainees to follow a given motion profile, and another a system from which trained knowledge transfers well in the real world. We measure how well the trainee can reproduce the skill in the real world after a fixed amount of time (for instance, one day), after initially learning the skill in the respective environment.

4.2 Experimental Environment Setup and Task

Figure 6 describes the experimental environment setup used in the evaluation of JFM. We used one SGI Indigo2 Impact graphics workstation for rendering a simple scene, one Polhemus FASTRAK six-DOF tracker for head tracking[1] and one Logitech 3D mouse (six-DOF tracker) for hand tracking.[2] For display equipment, we used a Sony Glasstron PLM-A55 mono dis-

1. Update rate of 120 Hz, tracking range of 10 ft. and accuracy of 0.03 in. RMS with a resolution of 0.0002 in./in.

2. Tracking speed of 30 in./sec., tracking range of 5 ft., and position-resolution of 0.04 in./in. and orientation-resolution of 0.1 in./in.

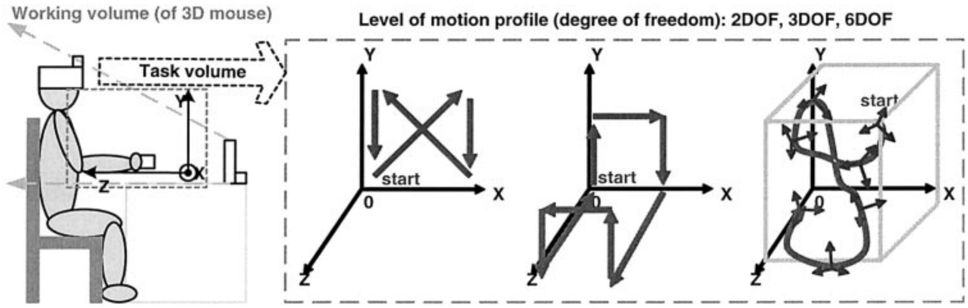

Figure 7. *Task design of motion following.*

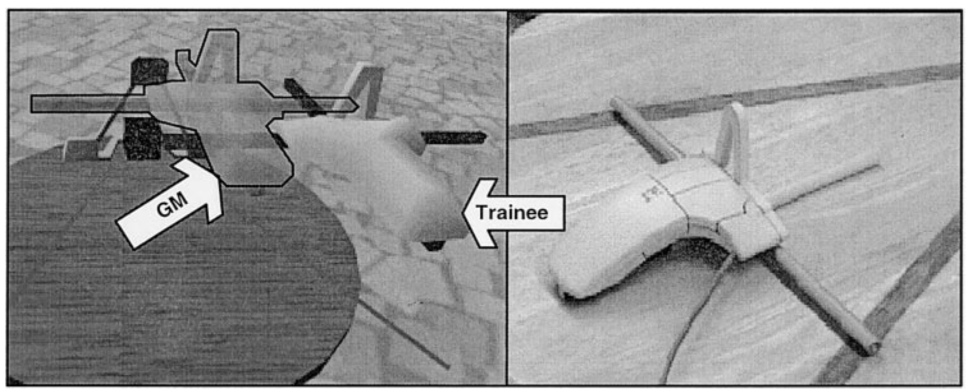

Figure 8. *3-D mouse for hand tracking and captured screen images (GM: semi-transparent ghost of trainer: trainee: 3-D mouse of trainee).*

play HMD[3] in the VE setting and a SGI 20 in. monitor in the real environment. In a less formal pilot study conducted a few months earlier (Yang & Kim, 1999), we used Virtual Research's V8 HMD, and the subject reported inconveniences from its heavy weight and sickness from the stereoscopic image. We thus opted to use a much lighter Sony Glasstron HMD without stereoscopy.

Our strategy was to eliminate as much negative bias as possible in the experiment, and, therefore, tested how JFM would fare against the real video-based training system that would not require stereoscopy (minimal depth perception). Comparing the VE-based JFM to a situation in which a real trainer would demonstrate a motion was also difficult because there would be no way to provide the first-person image of the trainer to the trainee. (The trainee either has to see the back of the trainer or see the trainer from the front.)

The motion task used in the experiment involves test subjects tracing and following a 3-D trajectory using their hands. (See figure 7.) For the sake of convenience, the master's (and likewise the trainee's) trace was rendered as a ghostly three-axis coordinate structure to make the trainee see the changing orientation more clearly. (See figure 8.) For relative accuracy and low jitter, we opted to use an ultrasonic 3-D mouse to track the user's hand. The user grasped a 3-D mouse on which an artificial coordinate structure was mounted (see the left image of figure 8) instead of a magnetic tracker. As indicated in our second experiment design

3. Resolution of 800×225, pixels of 180,000, diagonal FOV of 38° (52 in. virtual screen at 2 m away) and weight of 150 g.

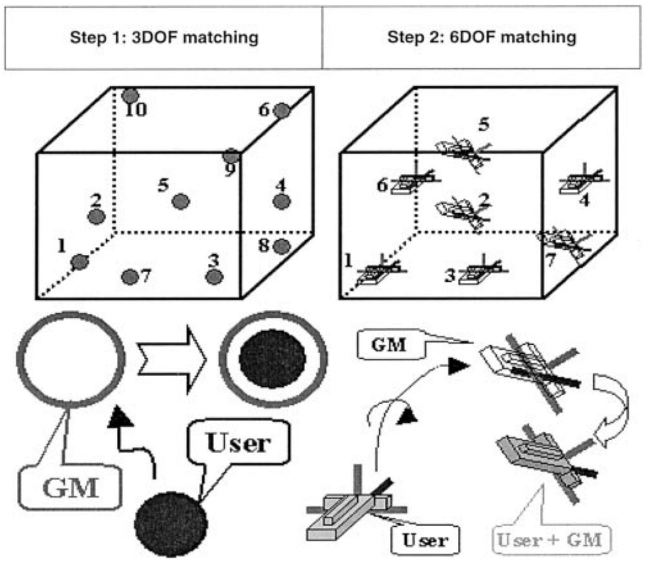

Figure 9. *Skill-normalizing exercise.*

goal (subsection 4.1), we defined three types of motion profiles by their degrees of freedom: a 2-D motion on the *X-Υ* plane, a 3-D (*X-Υ-Z* space) and a 6-D (*X-Υ-Z* and pitch-yaw-roll) motion within the hexahedral volume, defined inside the 3-D mouse's working volume. (See figure 7.) These three motion trajectories represent tasks in an increasing level of difficulty.

4.3 Subject's Skill Normalizing Exercise/Test

Before running the main experiment, we first conducted a VR skill exercise/test to familiarize subjects with the virtual environment and with the VR devices (for example, user's manipulation for moving the 3-D mouse and HMD) and to normalize the required basic skill level across the subject pool. The task involved overlapping the virtual 3-D mouse (with its mounted 3-D coordinate structure) on its ghostly replica that appeared within the task volume at random locations (three DOF) and in random orientations (six DOF). (See figure 9.) Test subjects repeated the exercise ten times, and only those who showed an acceptable performance level were admitted to the main experiment. (As a result, 3 of 39 candidates were excluded.)

4.4 The Day 1 Test

Four subject groups were formed for the Day 1 test: fast and slow motion in the VR, and likewise fast and slow motion in the real training environment. The VR subjects were further differentiated according to their skill levels. Each subject group attempted three different motion profiles (explained in subsection 4.2) in a random order to neutralize and minimize the inter-task influence. (See table 1.)

The VR environment groups used an HMD and were guided by the ghost metaphor from the first-person viewpoint. (See figures 6, 7, 8, and 10.) The virtual space was scaled at 1:1 as the real space.

The real-environment group watched (and followed) the same animated motion in a 20 in. monitor from the third-person viewpoint. (See the right image in figure 10.) The monitor was placed at 1 m from the subject, and the view direction of the virtual camera was adjusted according to the subject's viewing heights.

All subjects were asked to complete a questionnaire to assess some qualitative aspects of the experiment. (See subsection 5.10.1.)

4.5 The Day 2 Test

The Day 2 test was designed to verify the relative learning and transfer effect of the VR learning environment. (See subsection 4.1.) One day after the Day 1 test, the subjects were called back and requested to recall and reproduce the three motions learned during the Day 1 test. Subjects were not told to exercise the motion on their own after the Day 1 test to minimize any experimental bias (Promoim, 1999). This time, all subjects—regardless of whether they initially were tested in the virtual or real environment—were tested in the real environment without any display (that is, no visual guidance) and with only the 3-D mouse.

4.6 Performance Measure

The subject's motion was traced and initially matched with that of the master's using a method illus-

Table 1. *Subject Group Design*

Subject group design		Task Environment: Real World	Task Environment: Virtual Environment	
Speed of Motion	Fast	9 people	VR experience level	Novice: 3 people Normal: 3 people Expert: 3 people
	Slow	9 people	VR experience level	Novice: 3 people Normal: 3 people Expert: 3 people

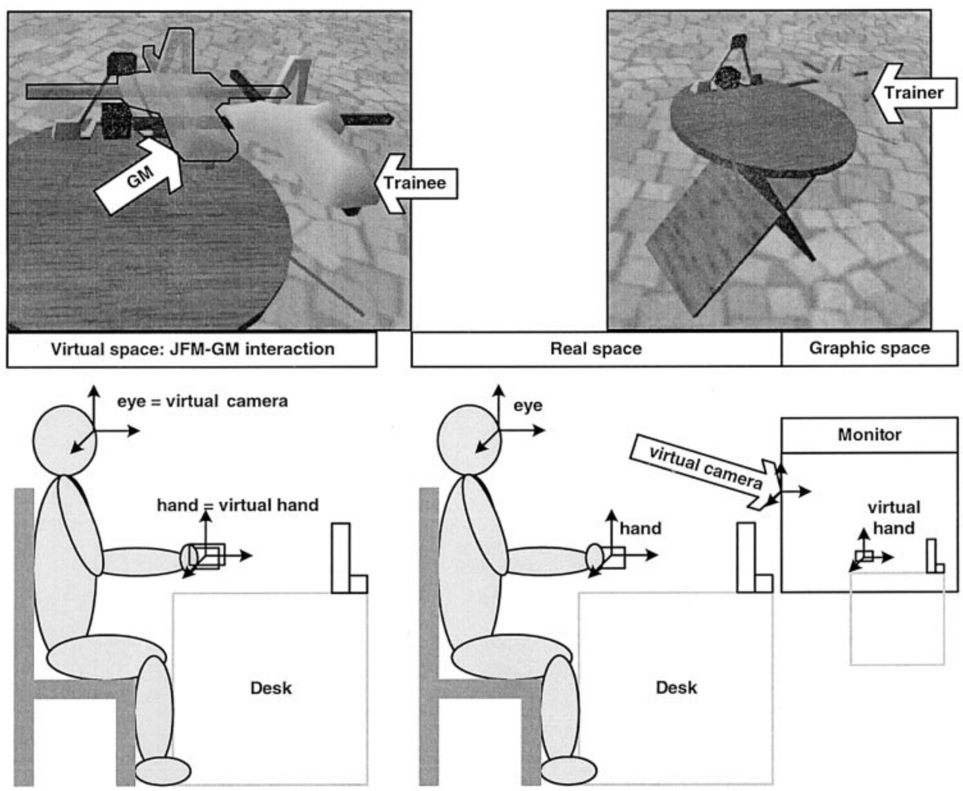

Figure 10. *Motion-following task of the Day 1 test.*

trated in figure 11. This process was required to compute the whole difference between the master and the subject because each subject took a different length of time to complete the task.

A curve-matching process is an optimization task to locate a set of data pairs that minimizes the difference between the two curves. We restricted the search window of the data pairs within 2 sec. of the corresponding target datum. Some experiment management was needed to ensure that the task was completed in a reasonable amount of time. Cross pairings of data were not allowed, although many-to-one mapping was allowed.

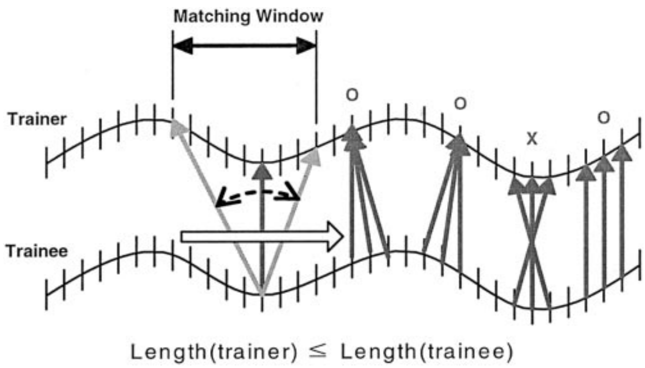

Figure 11. *Motion curve-matching method.*

The simple distance error metrics shown in table 2 were used to evaluate the subject's performance.

Aside from using the distance error, we also attempted a more qualitative similarity analysis between the two motions by segmentizing the motion at major inflection points and considering three different metrics. (See figure 12.) The first was the "regional time difference," a measure of the time taken to reach the inflection points. This measure reflects the amount of delay of the subject in following the master's motion. The second was the ratio between the times taken to complete each segment. The third is the same as the second metric but also considered the total task completion time. These metrics were used to assess the subject's ability to imitate the relative timing and rhythm of the motion in addition to replicating its position and orientation.

4.7 Results and Analysis

All experiment data were analyzed by the ANOVA (analysis of variance) method, and the following results were obtained within a significance level of approximately 1% ($p < .01$) to 5% ($p < .05$) (Cortina & Nouri, 1999; Miller, 1997).

4.7.1 Performance: Distance Error. The general interaction analysis result is depicted in figure 13. First, the ANOVA test did not report any significant correlation between the speed of the motion and the type of the training environment. The figure also shows that subjects performed generally better when the motion was slow and when trained in the virtual environment.

4.7.2 Transfer Effect: Distance Error. Figures 14 and 15 measure the change in subject's performance as measured by distance error between the first and second day. Obviously, the error has generally increased in all aspects in the second day. Figures seem to indicate a general trend that slow motion profiles were relatively easier to remember; furthermore, even though the subjects performed slightly better in the Day 1 virtual environment, the day-after performance was approximately equal.

4.7.3 Performance: Time-Dependent Factor. In terms of timing (the absolute duration of time taken to reach important critical points in the motion profile), figure 16 indicates that subjects trained in the virtual environment performed better (that is, the rhythms of motion were better preserved). Further, motion profiles, which had changing orientation, induced a worse performance.

The environment parameter showed an interesting interaction between speed and degrees of freedom of motion. Although the subject's timing was generally much better in slow motion in the virtual environment, it was still somewhat worse than when trained for fast motion. For fast motion, the difference in the test environment was not a factor for better timing performance. (See figure 17.) In general, adding changing orientations to the motion profile resulted in worse timing performance, more so in the real environment. (See figure 17.) A similar trend was also found with the analysis of the ratio between the respective motion segment lengths. (See figure 18.)

In the analysis of distance error, it was found that, although the subjects performed generally better in the VR environment, as for the movement in the *y* direction (vertical, up-and-down movement), the result was otherwise. (See figure 19.) This conclusion was not immediately apparent at first because the distance error was

Table 2. *Performance Measure (by distance error)*

2DOF = (① + ②)/2
3DOF = (① + ② + ③)/3
6DOF = (① + ② + ③ + ④ + ⑤ +)/6

① $= \frac{1}{N}\sum_{i=1}^{n}|M_{x_i} - S_{x_i}|$ ② $= \frac{1}{N}\sum_{i=1}^{n}|M_{y_i} - S_{y_i}|$ ③ $= \frac{1}{N}\sum_{i=1}^{n}|M_{z_i} - S_{z_i}|$

④ $= \frac{1}{N}\sum_{i=1}^{n}|M_{Pitch_i} - S_{Pitch_i}|$ ⑤ $= \frac{1}{N}\sum_{i=1}^{n}|M_{Yaw_i} - S_{Yaw_i}|$ $= \frac{1}{N}\sum_{i=1}^{n}|M_{Roll_i} - S_{Roll_i}|$

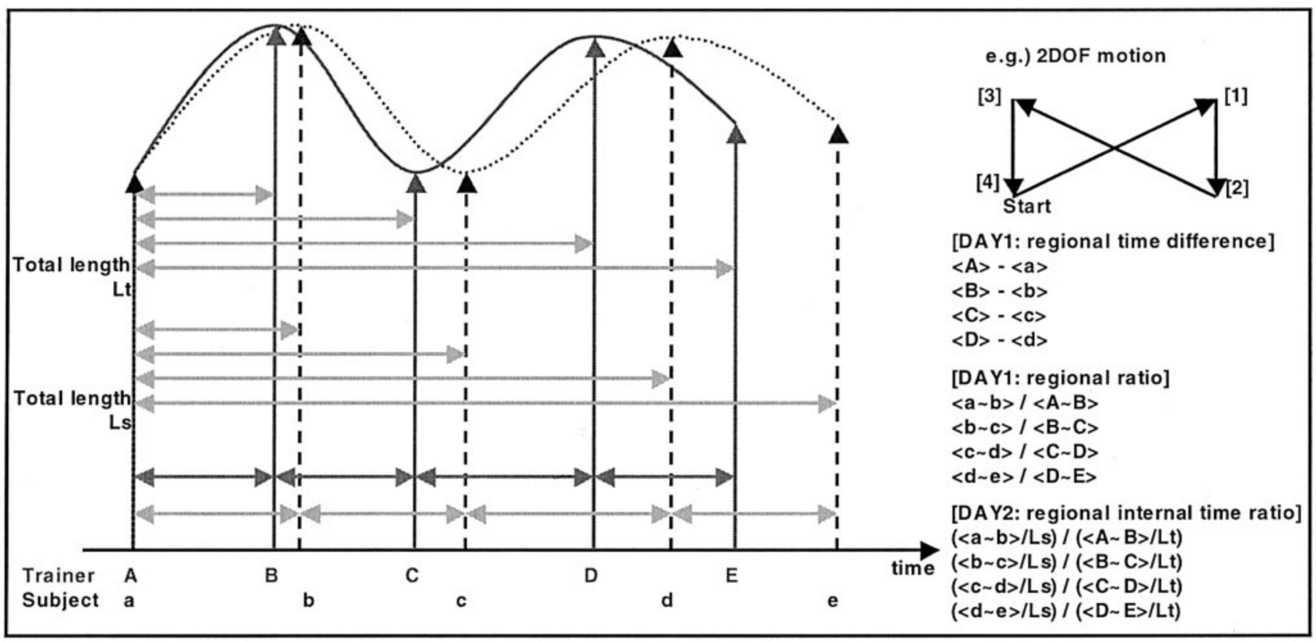

Figure 12. *Motion curve similarity using timing characteristics.*

summed up and averaged for all degrees of freedom. This effect seems to have affected the timing performance as well. For instance, figure 20 show that in the second and fourth segment or region of the motion profile, in which there were relatively less movements in the *y* direction, the overall error was lower. This *y*-movement factor had a strong enough influence to even reverse the general trend of obtaining a better performance in the VR environment. (For example, the real environment produced a small error in the first segment in figure 20.)

4.7.4 Transfer Effect: Time-Dependent Factor. The only result of relatively less difference (between Day 1 and Day 2 tests) we found with the timing behavior in the Day 2 test was that slow motion resulted

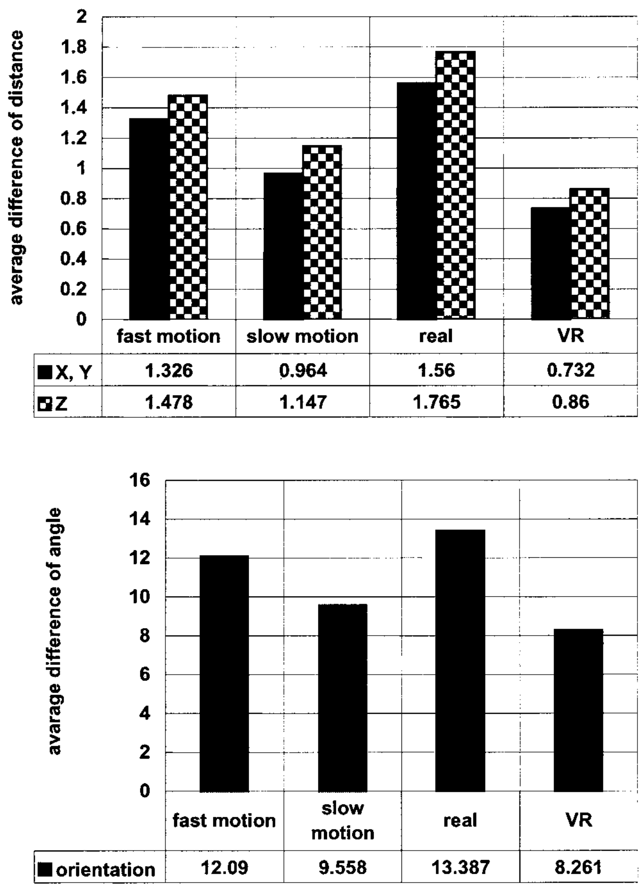

Figure 13. *Subject's performance by types of motion and environment (p < .01).*

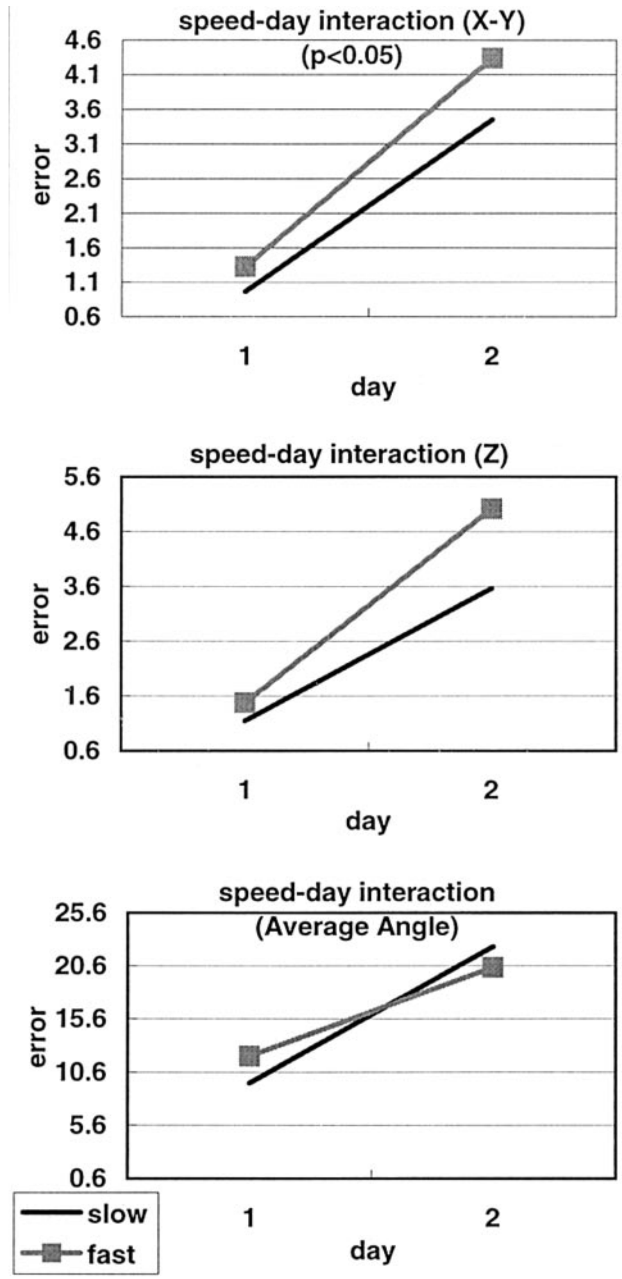

Figure 14. *Day 2 performance by speed (p < .01).*

in less error than the fast motion in the second day, as figure 22 shows.

4.7.5 The Questionnaire. Following is the summary of the main analysis results of the answers to the subjective questionnaire.

Question 1 (Day 1): Subjects were asked how much they felt and recognized the idea of the ghost metaphor. Subjects tested in the virtual environment answered positively with an average value of 5.222 (out of 7). We believe that the ghost metaphor played an important part in helping the user to follow the prescribed motion.

Question 5 and 6 (Day 1): Subjects were asked if they felt that certain types of motion profiles were more difficult to follow than others (See figure 23.)

The results showed that the added degrees of freedom did not explicitly make the subjects "feel" that the motion was more difficult in a significant way, except for pitch control (for which a small correlation was found).

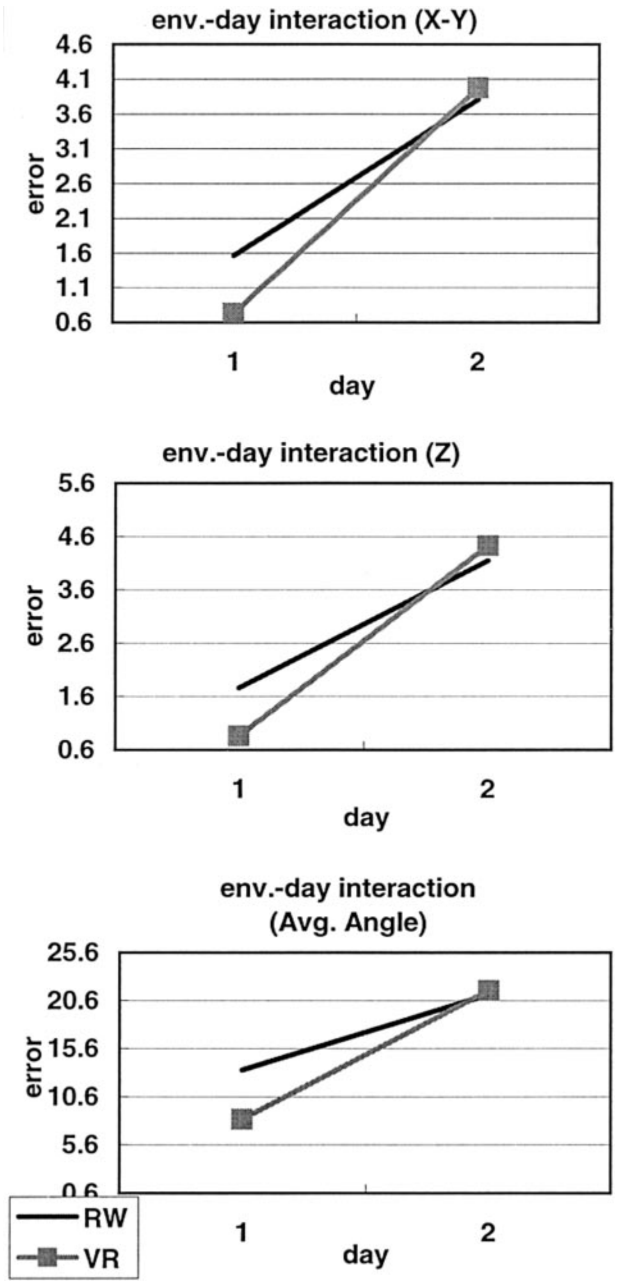

Figure 15. *Day 2 performance by test environment ($p < .01$).*

Question 7 (Day 2): Subjects were asked about their satisfaction with the training method and animation in the real environment or ghost metaphor in the virtual environment. Table 3 shows that the users of the virtual environment responded more affirmatively (average 5.056).

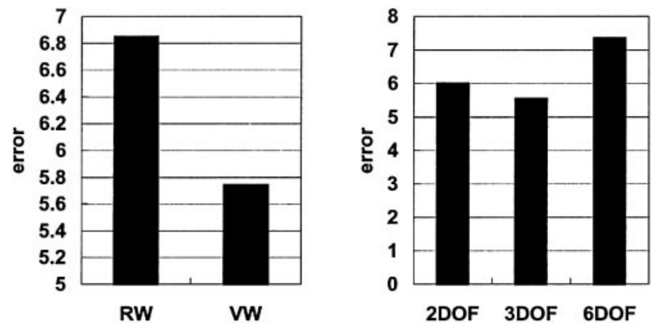

Figure 16. *Timing difference with environment and DOF ($p < .01$).*

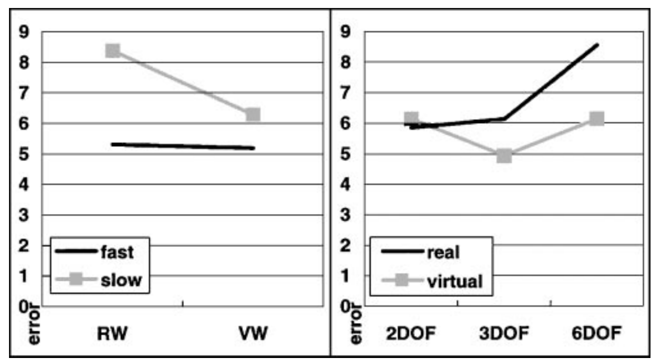

Figure 17. *Interaction among test environment and motion characteristics ($p < .01$).*

4.7.6 Discussion. In this subsection, we summarize some of the major findings regarding the performance of the proposed VR-based motion-training system and offer probable explanations. The first major result is that users followed the reference motion better using the ghost metaphor in the virtual environment, or at least as well as in the real environment. As the major difference between the virtual and the real environment is in the use of the first-person viewpoint and the use of the transparent ghost recognized as moving in the same coordinate space, it is quite natural to conclude that the JFM paradigm provides a more suitable and efficient interaction method, with a better frame of reference for the user to absorb motion-related information. The general finding was that the higher the degree of freedom the lower the performance, which seems to sup-

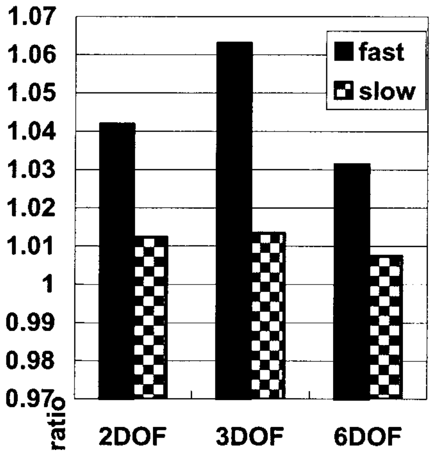

Figure 18. *Proportion of segments length and speed of the motion ($p < .01$).*

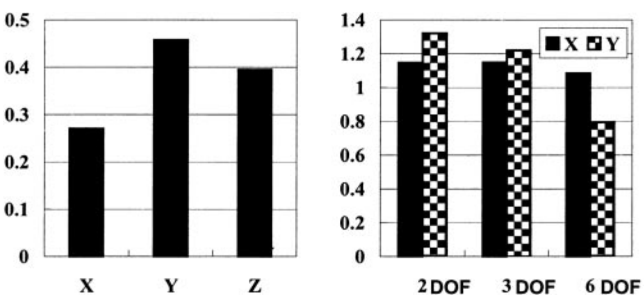

Figure 19. *y-movement factor has relatively large error ($p < .01$) (left: interaction between each parameter and ratio of error-variation to speed-variation; right: interaction between DOF and error of x, y-direction).*

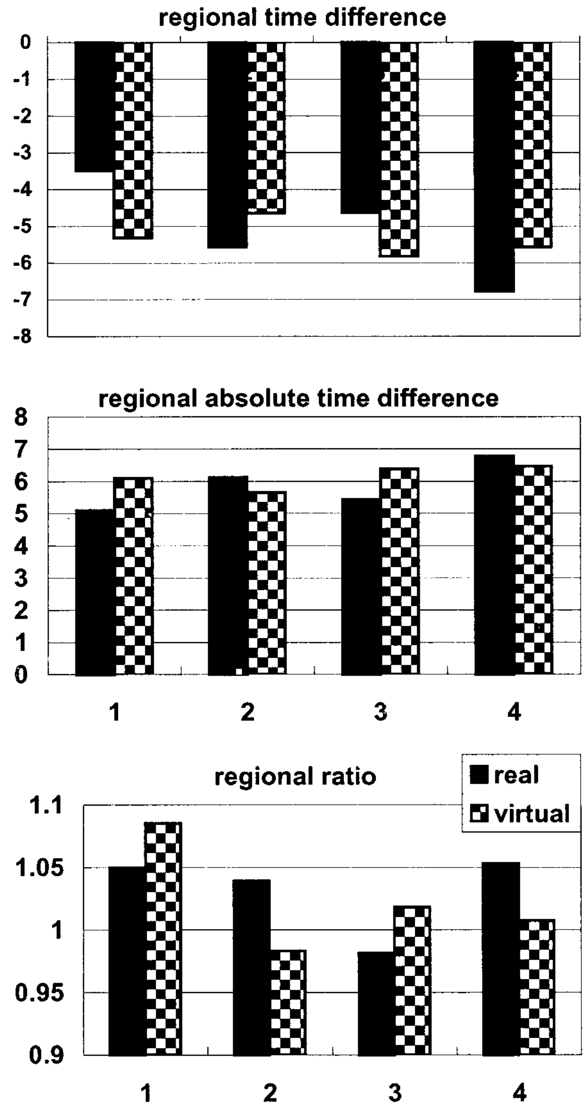

Figure 20. *Segment-wise timing performance and test environments/two-DOF motion ($p < .01$).*

port our hypothesis on the need to reduce the cognitive load for better performance in adapting to and recalling the learned motion.

The trouble with moving correctly in a *y* direction (vertical, up and down) in the virtual environment is probably related to the limitation with the HMD, with its very narrow vertical FOV and its weight (even though we used a relatively light HMD). The narrow FOV made users move their head frequently, whereas this would not be required in the real environment. The weight and wearability factor was particularly problematic when the users had to move their heads in an up-and-down fashion. The general finding that users performed better with the slow motion in the virtual environment is probably related to this factor as well. The faster the motion, the more likely that users have to tolerate higher inertia to move their heads with the HMD. This hypothesis is supported by the fact that the error increased proportionally with the amount of movement in the *y* direction. In the future, the problem

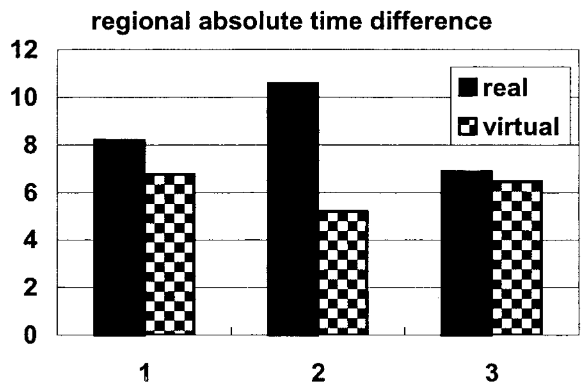

Figure 21. *Segment-wise timing performance and test environment/six-DOF motion (p < .01).*

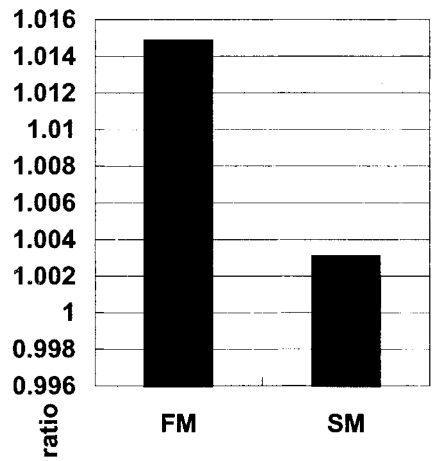

Figure 22. *Timing performance in Day 2 (p < .01).*

may be solved by the development of new lighter and wider-FOV HMDs (Barfield & Caudell, 2001; Microopticalcorp, 2001). However, it is also plausible to think that the ghost metaphor interaction is an inherently time-consuming method for following the motion, having to consciously overlap one's limb correctly on the ghost, thus fit for slow motion training. In general, the literature of motor skill learning states that fast motion learning should first be preceded by repeated practice at slow speed, mentally recounting the steps and postures (more as a guide). Next, when practicing in the fast mode, the motion should be already trained and made almost automatic/reflexive without requiring any cognitive effort (Schmidt, 1991). Thus, the ghost metaphor might be used more as a performance evaluation tool.

Based on the results from the Day 2 test, we found that using the VR devices or VR environment did not produce any significantly negative effect on the performance and its application to the real situation. Our worst-case setup represents the best-case situation with regards to the device dependency of the experimental result, in that the devices can only get better and improve overall training effect and user comfort.

When reinforced and augmented with presence cues, more-robust tracking and lighter HMDs, and rich informative graphics and images, we safely conclude that VR-based training methods will be an attractive alternative to the traditional "trainer-in-residence" or video-based motor skill learning method.

5 Future Work

Many applications of this work are possible (Yang, Ahn, Baek, & Kim, 2001). Conducting long-term (such as Day 3 and Day 4) evaluation tests may be needed to further confirm the transfer and learning effect of the JFM. In the evaluation test of usability of JFM-GM, we manually adjust eye-hand coordinates for the virtual environment to supply the match of sensorimotor feedback between visual and proprioceptive cues as the interaction ghost metaphor. But, to accomplish the full feature of JFM-GM, we need particular studies for synthesizing sensorimotor feedback, especially for visual and haptic interface over human factors and properties of display devices such as stereo HMDs. To further achieve the goal of the ghost metaphor (that is, consistent and complete multimodal sensorimotor feedback), we plan to also consider other cues such as stereoscopy and haptics (Schuemie, Straaten, Krijin, & Mast, 2001). The current version of JFM assumes that the motion data is already retargeted for an appropriate display and employs simplistic similarity measures for comparing two motion profiles. We are currently working on motion data processing techniques for fast, online motion

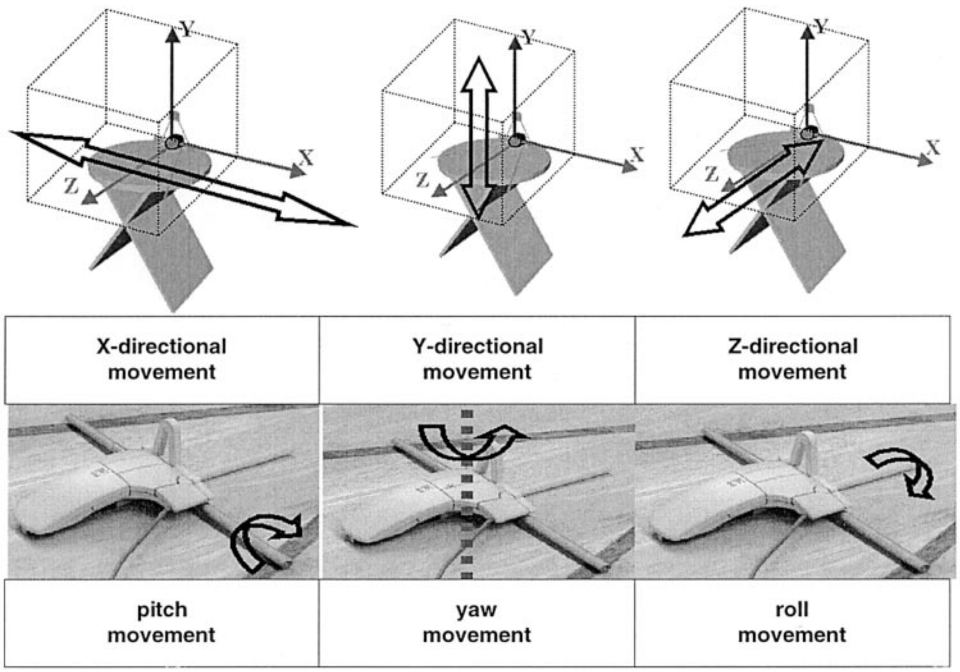

Figure 23. *Questionnaire about the effect of motion factors.*

Table 3. *User Satisfaction with the Test Environment ($p < .06$)*

Environment	N	Mean	SD
Real	18	4.111	1.72
Virtual	18	5.056	1.162

retargeting and more-qualitative analysis of the motion profiles. We need to experiment with other auxiliary guidance cues and assess their role in motor skill learning in the VE. Although the role of the first-person viewpoint has been much emphasized in this paper, the third-person viewpoint role must also be emphasized. As an auxiliary channel of information to observe the overall posture of the master, its relative importance must be evaluated. Presence and copresence are important in leaving a strong impression of the visited virtual environment and probably for increasing the learning effect. In this regard, we are currently investigating ways to use augmented reality and wearable computing equipment.

6 Conclusions

In this paper, we presented a novel interaction method for effectively guiding a trainee to follow and learn exact motion in a VR-based training system. A series of experiments and implementations showed that the system could achieve a transfer and learning effect as effective as traditional learning media, despite relatively low presence and problems with current VR devices. We believe that this is due to the more direct transfer of proprioceptive information from the trainer to the trainee. In other words, less effort is required, using the first-person viewpoint with synthesized sensorimotor feedback, to put oneself in the trainer's shoes.

Thus, when reinforced and augmented with presence cues, more-robust tracking, and rich informative graphics and images, we conclude that VR-based training methods will be an attractive alternative to the traditional "trainer-in-residence," or video-based motor skill learning method.

Acknowledgments

We thank the Statistics and Human Factors Engineering team of the POSTECH Industrial Engineering Department for helping us with the experimental design and analysis. This project has been supported in part by the Korea Ministry of Education BK 21 program and the Korea Science Foundation-supported Virtual Reality Research Center.

References

Acchione, N. S., & Psotka, J. (1993). Mach III: Past and future approaches to intelligent tutoring. *Proceedings of the 1993 Conference on Intelligent Computer-Aided Training and Virtual Environment Technology,* 344–351.

Badler, N., Webber, B., Clarke, J., Chi, D., Hollick, M., Foster, N., Kokkevis, E., Ogunyemi, O., Metaxas, D., Kaye, J., & Bindiganavale, R. (1996). MediSim: Simulated medical corpsmen and casualties for medical forces planning and training. *Proceedings of the National Forum: Military Telemedicine,* On-Line Today Research, Practice, and Opportunities. Los Alamitos, CA: IEEE Computer Society Press, p. 21–28.

Baek, S. (2001). *Posture based motion conversion, evaluation and advice system.* Unpublished master's thesis. POSTECH. [On-line] Available: http://vrlab.postech.ac.kr/vr/gallery/pub/2001/posture.doc.

Barfield, W., & Caudell, T. (2001). *Fundamentals of wearable computers and augmented reality.* Mahwah, NJ: Lawrence Erlabaum.

Blanz, V., Tarr, M. J., Bülthoff, H. H., & Vetter, T. (1996). *What object attributes determine canonical views?* (Tech. Rep. No. 42). Max-Plank-Institut für biologische Kybernetik.

Bowman, D., Wineman, J., Hodges, L., & Allison, D. (1999). The educational value of an information-rich virtual environment. *Presence: Teleoperators and Virtual Environments,* 8(3), 317–331.

Bülthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1994). *How are three-dimensional objects represented in brain?* A. I. Memo No. 1479, Center for Biological and Computational Learning Paper No. 96.

Choi, K. J., & Ko, H. S. (1999). On-line motion retargeting. *Proceedings of the International Pacific Graphics '99,* 32–42.

Cortina, J. M., & Nouri, H. (1999). *Effective size for anova designs.* Sage University Papers Series, Quantitative Applications in the Social Sciences, No. 07–129.

Cromby, J. J., Standen, P. J., & Brown, D. J. (1996). The potentials of virtual environments in the education and training of people with learning disabilities. *Journal of Intellectual Disability Research, 40*(6), 489–501.

D'Cruz, M., Eastgate, R., & Wilson, J. R. (1997). A study into the issues involved when applying virtual environment technology to training applications. *Proceedings of the Virtual Reality Universe '97.*

Emerson, T. C., & Revere, D. (1997). *Virtual reality in training and education: Resource guide to citations and online information.* HITL Technical Publications, B-94-1.

Everett, S., Wauchope, K., & Perez-Quinones, M. (1998). Creating natural language interfaces to VR systems: Experiences, observations, and lessons learned. *Proceedings of the VSMM 98.*

Graniano, M. S. (1999). Where is my arm? The relative role of vision and proprioception in the neuronal representation of limb position. *Proceedings of the National Academy of Science, 96,* 10418–10421.

Helander, M., Landauer, T. K., & Prabhu, P. V., (Eds.). (1997). *Handbook of human-computer interaction.* 2nd ed. New York: Elsvier.

Hix, D., & Hartson, H. R. (1993). *Developing user interfaces: Ensuring usability through product & process.* John Wiley & Sons, Inc.

Hodges, L., Rothbaum, B., Kooper, R., Opdyke, D., Meyer, T., North, M., Graff, J., & Williford, J. (1995), Virtual environments for treating fear of heights. *IEEE Computer, 28*(7), 22–34.

Holden, M., Todorov, E., Callahan, J., & Bizzi, E. (1999). Virtual environment training improves motor performance with stroke: Case report. *Neuro. Report. 29*(9), 57–67.

Jack, D., Boian, R., Merians, A. S., Tremaine, M., Burdea, G. C., Adamovich, S. V., Recce, M., & Poizner, H. (2001). Virtual reality-enhanced stroke rehabilitation. *IEEE Trans-*

actions on Neural Systems and Rehabilitation Engineering, 9(3), 308–318.

Johnson, W. L., Rickel, J., Stiles, R., & Munro, A. (1998). Integrating pedagogical agents into virtual environments. *Presence: Teleoperators and Virtual Environments, 7*(6), 523–546.

Jayaram, S., Wang, Y., & Jayaram, U. (1999). A virtual assembly design environment. *Proceedings of the IEEE VR Conference,* 172–179.

Kuhlen, T., & Dohle, C. (1995). Virtual reality for physically disabled people. *Comput Biol Med., 25*(2), 205–211.

Macedonia, M. R., Zyda, M. J., Pratt, D. R., Barham, P. T., & Zeswitz, S. (1994). NPSNET: A network software architecture for large scale virtual environments. *Presence: Teleoperators and Virtual Environments, 3*(4), 265–287.

Magill, R. A. (2000). *Motor learning: Concepts and applications.* Madison, WI: McGraw Hill.

Microopticalcorp. (2001). The MicroOptical Corporation: HMD technologies. [On-line] Available: http://microopticalcorp.com.

Miller, R. G., Jr. (1997). *Beyond Anova: Basics of applied statistics.* Boca Raton, FL: Champman & Hall.

Pausch, R., Proffitt, D., & Williams, G. (1997). Quantifying immersion. *Proceedings of SIGGRAPH 97,* 13–18.

Promoim. (1999). Promoim consulting Group in Pohang University of Science and Technology. [On-line] Available: http://www.promoim.co.kr.

Rastogi, A. (1996). *Design of an interface for teleoperation in unstructured environments using augmented reality displays.* Unpublished doctoral dissertation. University of Toronto.

Rickel, J., & Johnson, W. L. (1999). Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artificial Intelligence 99*(13), 343–392.

Schmidt, R. (1991). *Motor learning and performance.* Champagne, IL: Human Kinetics Books.

Schuemie, M. J., Straaten, P., Krijin, M., & Mast, C. (2001). Research on presence in VR: A survey. *The Journal of Cyber Psychology and Behavior, 4*(2), 183–202.

Shawver, D. M. (1997). Virtual actors and avatars in a flexible user-determined scenario environment. *Proceedings of IEEE VRAIS 97,* 170–177.

Todorov, E., Shadmehr, R., & Bizzi, E. (1997). Augmented feedback presented in a virtual environment Accelerated learning of a difficult motor task. *Journal of Motor Behavior, 29*(2), 147–158.

Toussaint, G. T. (2000). The complexity of computing nice viewpoints of objects in space. Keynote address at Vision Geometry IX, *Proceedings of SPIE.* [On-line] Available: http://cgm.cs.mcgill.ca/~godfried/publications/viewpoints.pdf.

VMW. (1997). Virtual Medice Worlds. [On-line] Available: http://www.hoise.com/vmw/articles/LV-VM-09-29.html.

VR TechnoCenter (1998). *Brochure of VRES-2: VR Earthquake Simulator.* [On-line] Available: http://www.vrtc.co.jp/english/prod01.htm.

Wilson, R. (1994). Geometric reasoning about mechanical assembly. *Al, 71*(2), 371–396.

Yang, U. Y., & Kim, G. J. (1999). "Just Follow Me": An Immersive VR-based motion training system. *Proceedings of the International Conference on Virtual Systems and Multimedia, VSMM'99,* 435–444.

Yang, U. Y., Ahn, E. J., Baek, S. M., & Kim, G. J. (2001). "Just Follow Me": A VR-based motion training system. *Emerging Technologies 70, Conference Abstracts and Applications, SIGGRAPH 2001,* 126.

Yang, U. Y., Lee, G. A., Lee, J. Y., Bok, I. G., & Han, S. H. (2002). Design and evaluation of selection methodology of canonical view for 3D multi-joint object. *Manuscript submitted to Conf. of Human Computer Interaction 2002, Korea.*

Yokokohji, Y., Hollis, R. L., & Kanade, T. (1999). WYSIWYF display: A visual/haptic interface to virtual environment. *Presence: Teleoperators and Virtual Environments, 8*(4), 412–434.

Youngblut, C. (1998). *Educational uses of virtual reality technology.* (Tech. Rep. IDA Document D-2128). Institute for Defense Analyses, Alexandria, VA.

Zhai, S., Buxton, W., & Milgram, P. (1996). The partial-occlusion effect: Utilizing semitransparency in 3D human-computer interaction. *ACM Transaction on Computer-Human Interaction, 3*(3), 254–284.