Are Four Hands better than Two?
Bimanual Interaction for Quadmanual User Interfaces

Paul Lubos
Department of Informatics
University of Hamburg
paul.lubos@uni-hamburg.de

Gerd Bruder
Department of Informatics
University of Hamburg
gerd.bruder@uni-hamburg.de

Frank Steinicke
Department of Informatics
University of Hamburg
frank.steinicke@uni-hamburg.de

ABSTRACT
The design of spatial user interaction for immersive virtual environments (IVEs) is an inherently difficult task. Missing haptic feedback and spatial misperception hinder an efficient direct interaction with virtual objects. Moreover, interaction performance depends on a variety of ergonomics factors, such as the user’s endurance, muscular strength, as well as fitness. However, the potential benefits of direct and natural interaction offered by IVEs encourage research to create more efficient interaction methods.

We suggest a novel way of 3D interaction by utilizing the fact that for many tasks, bimanual interaction shows benefits over one-handed interaction in a confined interaction space. In this paper we push this idea even further and introduce quadmanual user interfaces (QUIs) with two additional, virtual hands. These magic hands allow the user to keep their arms in a comfortable position yet still interact with multiple virtual interaction spaces. To analyze our approach we conducted a performance experiment inspired by a Fitts’ Law selection task, investigating the feasibility of our approach for the natural interaction with 3D objects in virtual space.

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User Interfaces – Input Devices and Strategies, Evaluation / Methodology.

Author Keywords
Spatial user interfaces; virtual environments; 3D interaction.

INTRODUCTION
The advent of affordable head-mounted displays (HMDs) and new interaction devices, such as the Microsoft Kinect or Leap Motion, renewed the interest in immersive virtual environments (IVEs). Such IVEs have the potential to offer natural and direct interaction with objects displayed in the virtual world. In particular, the workspace within the user’s arm reach provides a volume, in which the user can grab virtual 3D objects similar to the real world. Spatial interacting via natural gestures in IVEs allows interaction designers to exploit the richness and expressiveness of the interaction.

However, interaction in 3D mid-air is physically demanding and, therefore, often hinders user satisfaction and performance [4]. The increase in the degrees-of-freedom (DoFs) that have to be controlled simultaneously as well as the absence of passive haptic feedback and resulting interpenetration and occlusion issues when “touching the void” [3, 4] are often responsible for reduced performance. Hence, although significant improvements have been made in 3D input technologies, using tracked human gestures and postures in “mid-air” still introduce challenges to the design of high-performance 3D interaction techniques [4].

In this context, virtual hand techniques are often considered to be the most natural way of directly interacting with virtual objects as they map identically virtual tasks with real tasks. However, direct interaction with a virtual object in an HMD setup significantly differs from a similar task in the real world [1]. In particular, users often cannot see their real body, but at most a virtual representation in form of a virtual hand, marker or 3D point in space. Furthermore, even small inaccuracies and latency of the used tracking system may cause slight mismatches between visual appearance of the virtual hand and the user’s proprioceptive and kinesthetic feedback [3, 4]. Such a decoupling of motor and visual space may degrade performance due to the kinematics of point and grasp gestures in 3D space and the underlying cognitive functions [11].

Despite the well known problems, a large body of literature has shown the benefits of virtual hand interaction, in particular, bimanual interaction, and several promising two-handed interaction techniques have been described [5]. In this paper we propose a novel way of bimanual interaction, which evolves around simulating additional virtual hands for a user resulting in quadmanual user interfaces (QUIs). Using this approach, homing movements of the hands can be effectively reduced by dividing the interaction space into interaction volumes. With this approach we transfer a solution previously devised for large display environments to IVEs, aiming for similar benefits (cf. [8]).

RELATED WORK
Direct interaction arguably provides the most natural type of interaction with virtual objects, but it is often not possible to use direct interaction for objects that are not located within arm’s reach. Performance of constant translational (less so rotational) decoupling of visual and motor
spaces has been found subject to adaptation [7], i.e., performance increased over time, while optimal performance may be achieved when visual and motor spaces are superimposed or coupled closely [9, 14]. A large body of literature has shown the benefits of bimanual interaction and several promising two-handed interaction techniques have been described [5], e.g., for symmetric interaction tasks [2] or touch interaction [13]. However, it is still an open research question, how the position of virtual objects may affect interaction performance. Direct interaction is subject to perceptual limitations, e.g., the vergence-accommodation mismatch, ghosting or double vision, which can result in strong misperception effects [3, 4]. Dependent on the location of virtual objects, users may be unable to discriminate interrelations or perceive distances to objects to be smaller or larger than they are displayed [3]. Such distortions do not appear in the real world and may be related to limitations of current technology to correctly reproduce natural cues from the real world [3]. Moreover, due to varying energy expenditure between users and differences in strength and endurance of arm muscles, interaction performance in mid-air within arm’s reach in IVEs may be affected by different factors related to the ergonomics of direct interaction. In particular, contributing factors may include interaction duration, hand and arm postures, frequency of movements, and comfort [1].

QUI: A QUADMANUAL USER INTERFACE

Figure 1(a) shows a regular bimanual user interface in which movements of a user’s tracked real-world hands are mapped one-to-one to virtual hands. The quadmanual user interface is illustrated in Figure 1(b). With QUIs the user is able to control two pairs of hands. One pair of hands is active, whereas the other pair is inactive and displayed semi-transparently. Using these two virtual hands it is possible to reduce homing movements of the hands as well as the distance between them and target locations by dividing the interaction space into smaller volumes of interaction.

Controlling Four Hands

Since the user has only two real hands available, a mapping strategy is required to map movements of the user’s two real hands to four virtual hands. Two straightforward approaches are possible:

- **simultaneous control**: The user controls both virtual left hands with their real left hand, and both virtual right hands with their real right hand. Although this approach is easy to implement, it has the drawback that all virtual hands are active, even if the user is not focusing on them. Hence, it becomes possible that the hands outside the view interact with the VE, even if not intended to.

- **selective control**: The user controls the virtual left pair of hands with their real left and right hand or they control the virtual right pair accordingly. This approach appears to be more feasible, requiring focus only on the active hands while the inactive hands do not affect the virtual space.

Activating Hands

To determine which two of the virtual four hands should be active, we decided to exploit the gaze direction of the user, approximated by the head position and orientation. Thus, if the user looks to one pair of hands, this pair is activated, which is visually indicated by an opaque, textured rendering, whereas the inactive pair of hands is shown semi-transparently. The active hands stay active until the user focuses on the other hands. In this case, the former active pair of hands is visualized semi-transparently and all virtual hand movements freeze for this pair.

In theory, users should be more efficient using four than two hands, e.g., if we only consider movement distances to spatial targets for selections as predicted by Fitts’ Law [6]. However, it is a challenging task to control four virtual hands with only two tracked hands. Hence, the question arises how much additional perceptual, cognitive and motor effort is required for such an unnatural—or in other words supernatural—way of spatial bimanual user interaction.

EXPERIMENT

In this section we describe a Fitts’ Law inspired experiment, in which we explore how much learning is required until QUIs have the potential to outperform bimanual user interfaces. We analyzed direct 3D selection in the user’s arm reach in an HMD environment using two-handed interaction to control bimanual and QUIs. We evaluate the following two hypotheses:

**H1**: 3D selection performance is initially higher with bimanual compared to quadmanual user interfaces.

**H2**: With training, the performance difference between bimanual and quadmanual user interfaces decreases.

Participants

We recruited 11 participants for our experiment. Nine of them were male and two were female (ages 19 - 45, M = 27.27). One participant was left-handed, the others were right-handed. All of them had normal or corrected vision. We measured the interpupillary distance (IPD) before the experiment (M = 6.54 cm, SD = 0.32 cm), and determined their sighting dominant eye (2 left). All participants were naive to the experimental conditions. The duration of the experiment was one hour.
Figure 2. Illustration of (a) a participant during the experiment, as well as experimental conditions with target spheres for (b) two hands and (c) four hands. The virtual, duplicated hands are dark blue.

Materials
As illustrated in Figure 2(a), users wore an Oculus Rift DK1 HMD tracked in 6 DoF a Naturalpoint TrackIR 5 camera. We used a PrimeSense Carmine sensor and the 3Gear NimbleSDK for skeletal tracking and hand reconstruction. The visual stimulus consisted of a 3D scene (see Figure 1), which was rendered with Unity3D on an Intel computer with a Core i7 3.4GHz CPU and Nvidia GeForce GTX780TI.

Selection targets in the experiment were represented by spheres and colored grey. The current target was colored blue when the user’s finger was outside the target, and green when the user’s finger was inside. Each trial consisted of 8 spheres on a plane in front of the starting positions, four of which were at a close distance and the others at a far distance (see Figure 2). For the two hand condition the targets were divided into a left region for the left hand and the right region for the right hand. The close targets were at a distance of 16.77cm, the others at 21.36cm. In the four hand condition two targets in a region, at distances 15cm and 20cm and IDs of 2 and 3, were assigned to each of the four hands. Taking the different distances into account, we scaled the targets in the two hands condition up to IDs of 2 and 3 to be able to compare the movement time between the two and the four hands conditions. According to Fitts’ Law [6], adapting the target size with respect to the distance between selections results in larger targets for longer selection distances, thus resulting in the same task difficulty between the different interaction volumes.

Methods
To counterbalance, we used a within-subject 2×2×8×8 design. The two conditions (bimanual vs. quadmanual), two indices of difficulty (IDs) and eight target positions, were uniformly and randomly distributed among each repetition. We repeated all conditions eight times to measure learning effects. Each trial consisted of a single selection of a single target, where participants had to keep one finger within the target sphere for one second. To account for perceptual limitations, we implemented an ellipsoid selection volume as suggested by Lubos et al. [10]. To make sure all virtual hands were used, we only accepted selections by the correct finger in the correct sphere. Before each trial, participants held both index fingertips within spheres at the starting position, i.e., the initial positions were the same for each trial.

Results
We analyzed the results with a repeated measure ANOVA and Tukey multiple comparisons at the 5% significance level. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity when Mauchly’s test indicated that the assumption of sphericity had been violated.

The results for movement time, including the selection time of one second, are shown in Figure 3. We found a significant difference between the conditions on movement time ($F(1, 10)=26.88$, $p<.001$, $\eta^2_p=.73$). The mean movement time for the four hands was $M=3.338$, $SD=.464$ and for the two hands $M=2.870$, $SD=.400$. We also found a significant main effect of repetition on movement time ($F(1.79, 17.85)=13.65$, $p<.001$, $\eta^2_p=.58$). We found a significant effect of hand dominance on movement time ($t(10)=−3.96$, $p<.01$). The mean movement time for the dominant hand was $M=2.97$, $SD=.330$ and $M=3.13$, $SD=.411$ for the non-dominant hand.

We analyzed differences in movement time between the co-located hands and the shifted, virtual hands. The mean movement time for the co-located hands was $M=3.046$, $SD=.375$ and $M=3.509$, $SD=.633$ for the non-co-located hands. We found a significant difference ($t(10)=−3.57$, $p<.01$). We didn’t find significant interaction effects.

Additionally, participants answered subjective questions on a scale of 1 = yes, 5 = no. We asked whether having four hands made the task easier ($M=3.55$, $SD=1.21$), whether seeing the transparent hands helped ($M=2.55$, $SD=1.64$), whether it was hard to control four hands ($M=3.82$, $SD=1.25$), and whether they wanted four hands ($M=2.18$, $SD=1.47$).
Discussion
The results show that for four hands the movement time is initially higher than for two hands. Since we adjusted the targets to the same IDs, the main difference between these selections was the time necessary to process the input, which can be explained by the human action cycle \[12\]. Compared to simple bimanual selections, an additional step is needed to turn the head towards the target if the corresponding virtual hands are not activated. Nevertheless, this result confirms our hypothesis H1.

Furthermore, our results show a training effect, as the mean selection times decrease per repetition, which confirms our hypothesis H2. As the experiment was designed to ensure there was no difference in task difficulty between two and four hands, meaning that the target size was artificially increased for the two hands conditions, the difference would be even smaller with targets of the same size, i.e., selections with four hands appear to be feasible. For time critical tasks, bimanual or single hand interaction might be more feasible, but as a training effect is evident in our results, and since it offers reduced homing times, future studies might show that a quadmanual approach is faster for certain time critical tasks. Furthermore, by reducing the need for long homing motions to switch from one interactive element to another in a virtual scene, QUIs reduce the tracking space necessary to properly track the user’s hands. Current hand tracking sensors, such as the Leap Motion, have a very limited tracking volume; this drawback may be overcome with QUIs.

CONCLUSION
In this paper we introduced quadmanual user interfaces, which allow users in IVEs to comfortably interact with a wide virtual interaction volume from a smaller real-world hand tracking volume. Our study investigated the cognitive demand of controlling four hands and compared it to bimanual interaction by measuring the movement time in a Fitts’ Law inspired experiment. While the task is more difficult to complete at first, the result showed that the difference is decreasing in time and, while statistically significant, is negligible. With our approach it is possible to solve tasks where just two hands are not enough, yet by deactivating two hands at a time, we limit the potential. Future research could investigate, whether moving the currently inactive hands might improve the usability of our technique, making it feel more natural. However ways to prevent involuntary manipulation of the virtual world with hands which are not in focus at the time, have to be considered. In our experiment, we used the user’s head direction to determine which set of virtual hands they wanted to control.

Future research could investigate alternative forms of activating the hands, e.g., eye tracking devices could improve the detection of the user’s focus point to better control QUIs. Furthermore, as most user interfaces have a number of specific regions of interest, such as a taskbar, it is possible to define those regions, either through heuristics or manually, and then allow users to place their virtual hands in these regions, eliminating the need to stretch the arms into uncomfortable positions and enabling them to solve tasks quicker due to reduced homing times. Potentially, even more than two sets of hands might be feasible for 3D interaction.

REFERENCES