2D Touching of 3D Stereoscopic Objects

Dimitar Valkov, Frank Steinicke, Gerd Bruder and Klaus Hinrichs
Visualization and Computer Graphics Research Group
University of Münster, Einsteinstr. 62, 48149 Münster, Germany
[dimitar.valkov, fsteini, gerd.bruder, khh]@uni-muenster.de

ABSTRACT
Recent developments in the area of touch and display technologies have suggested to combine multi-touch systems and stereoscopic visualization. Stereoscopic perception requires each eye to see a slightly different perspective of the same scene, which results in two distinct projections on the display. Thus, if the user wants to select a 3D stereoscopic object in such a setup, the question arises where she would touch the 2D surface to indicate the selection. A user may apply different strategies, for instance touching the midpoint between the two projections, or touching one of them.

In this paper we analyze the relation between the 3D positions of stereoscopically rendered objects and the *on-surface touch points*, where users touch the surface. We performed an experiment in which we determined the positions of the users' touches for objects, which were displayed with positive, negative or zero parallaxes. We found that users tend to touch between the projections for the two eyes with an offset towards the projection for the dominant eye. Our results give implications for the development of future touch-enabled interfaces, which support 3D stereoscopic visualization.

Author Keywords
Touch screens, multi-touch, stereoscopic displays.

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

General Terms
Experimentation, Human Factors

INTRODUCTION
Two different technologies have dominated recent tech exhibitions and the entertainment market: multi-touch surfaces and 3D stereoscopic displays. These two orthogonal technologies, i. e., multi-touch is about *input*, 3D stereoscopic visualization about *output*, have recently been combined in different setups [3, 27, 29], and first commercial systems that support stereoscopic display and multi-touch interaction are already available [2]. Furthermore, interdisciplinary research projects address the question how users interact with stereoscopic content on a two-dimensional surface [3, 4]. This combination has the potential to provide more intuitive and natural interaction setups with a wide range of applications, e. g., geo-spatial applications, urban planning, architectural design, collaborative tabletop setups, or 3D desktop environments [2, 27]. However, until now these systems are mainly used for navigation purposes whereas the interaction with the stereoscopically displayed objects is supported rather rudimentarily.

The ability to directly touch and manipulate graphical elements without using any unnatural input devices has been shown to be very appealing for novice as well as expert users [7]. In particular, passive haptics supplied by direct-touch devices and multi-touch capabilities have the potential to considerably improve the user experience. Therefore, the FTIR (frustrated total internal reflection) and DI (diffused illumination) technologies and their inexpensive footprints [14, 22] have led to the widespread usage of multi-touch on large displays. These setups detect direct touch contact and thus provide tangible feedback without requiring any further user instrumentation. Since humans in their everyday life usually use multiple fingers and both hands for interaction with their real world surroundings, such techniques have the potential to build intuitive and natural interfaces.

Stereoscopic visualization has been known for decades, but recently it has been reconsidered again due to the rise of 3D motion pictures and the upcoming 3D television. With stereoscopic display, objects might be displayed with different parallaxes, i. e., negative, zero, and positive, resulting in different stereoscopic effects. Objects rendered with *zero parallax* appear attached to the projection screen and are perfectly suited for touch interaction, especially if 2D input is intended. Until recently, multi-touch interaction research was mainly focused on those monoscopically rendered 2D or 3D data sets.

In this paper we present first steps to analyze the users' approaches to touch stereoscopically rendered virtual objects on a multi-touch enabled projection screen. Interaction with stereoscopically displayed content in such a setup opens new research challenges. In contrast to objects rendered with

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Figure 1. Illustration of the main problem for touch interaction with stereoscopic data. The user is either focused on the finger, which makes the selection ambiguous, or on the object, which disturbs the visual perception of the finger.

zero parallax, objects displayed with *positive parallax* appear behind the projection screen and cannot be accessed by direct touch interaction, since the screen surface limits the user's reach [13]. While one can use indirect selection and manipulation techniques for such objects [15, 23, 25], it is difficult to apply these techniques to objects in front of the screen. In fact, objects that appear in front of the projection screen, i. e., objects with *negative parallax*, introduce the major challenge in this context. When the user wants to interact with such an object by touching, she is limited to touch the area behind the object, since most touch sensitive screens capture only direct contacts, or hover gestures close to the screen. Therefore the user has to penetrate the visual objects to reach the touch surface with her finger. If the user penetrates an object while focusing on her finger, the stereoscopic effect for the object would be disturbed, since the user's eyes are not accommodated and converged on the projection screen's surface. Thus the left and right stereoscopic images of the object's projection would appear blurred and could not be merged anymore (Figure 1 (left)). However, focusing on the virtual object would lead to a disturbance of the stereoscopic perception of the user's finger, since her eyes are converged to the object's 3D position (Figure 1 (right)). In both cases touching an object may become ambiguous. However, as suggested by Valkov et. al. [29], users are insensitive to observe discrepancies between visual penetration and touch feedback when they try to touch stereoscopic objects which are displayed close to the surface. In particular, they found that users are less sensitive to discrepancies between visual and tactile feedback, if objects are displayed with negative parallax.

In the monoscopic case the mapping between an *on-surface touch point* and the *intended* object point in the virtual scene is straightforward, but with stereoscopic projection this mapping introduces problems. In particular, since there are different projections for each eye, the question arises where users touch the surface when they try to "touch" a stereoscopic object. In principle, the user may touch anywhere on the surface to select a stereoscopically displayed object. However, according to observations we have made, it appears most reasonable that users try to touch

- the midpoint between the projections for both eyes (so called *middle* eye projection),
- the projection for the *dominant* or *non-dominant* eye, or
- the orthogonal projection of the stereoscopic object onto the touch surface (i. e., the object's "*shadow*").

A precise mapping approach is important to ensure correct selections, in particular in a densely populated virtual scene. In order to allow the user to select arbitrary objects, a certain area of the touch surface, which we refer to as *on-surface target*, must be assigned to each object.

In this paper we present the results of an experiment that we have performed in order to determine the on-surface targets for objects stereoscopically rendered at different 3D positions. The results of this experiment provide guidelines how this mapping can be applied in future applications.

RELATED WORK

Recently many approaches for extending multi-touch interaction techniques to 3D applications with *monoscopic* rendering have been proposed [15, 21, 25, 33]. For instance, Hancock et al. [15] have presented the concept of *shallow-depth 3D*, i. e., 3D interaction with limited depth, in order to extend the interaction with digital 2D surfaces. They have developed intuitive interaction techniques for object selection and manipulation in this context. Extending the interaction space beyond the touch surface has been investigated by Hilliges et al. [16]. They have tested two depth sensing approaches to enrich the multi-touch interaction on a tabletop setup with monoscopic projection. Interaction with objects with negative parallax on a multi-touch tabletop setup is further addressed by Benko et al. [5]. The proposed *balloon selection* metaphor supports precise object selection and manipulation for augmented reality setups.

Nevertheless, direct touch interaction with *stereoscopically* rendered scenes introduces new challenges, as described by Schöning et al. [27]. In their work an anaglyph- or passive polarization-based stereo visualization was combined with FTIR-based touch detection on a multi-touch enabled wall, and approaches based on mobile devices for addressing the formulated parallax problems were discussed. A similar option for direct touch interaction with stereoscopically rendered 3D objects is to separate the interactive surface from the projection screen, as proposed by Schmalstieg et al. [26]. In their approach, the user is provided with a physical *transparent prop*, which can be moved on top of the object of interest. This object can then be manipulated via single- or multi-touch gestures, since it has almost zero parallax with respect to the prop. Recently, multi-touch devices with non-planar touch surfaces, e. g., cubic [8] or spherical [6], were proposed, which could be used to specify 3D axes or points for indirect object manipulation. The parallax problem described in the introduction is known from the two-dimensional representation of the mouse cursor within a stereoscopic image [28, 31]. While the mouse cursor can be displayed stereoscopically on top of objects [28] or monoscopically only for the dominant eye [31], movements of real objects in the physical space, e. g., the user's hands, cannot be constrained such that they appear only on top of virtual objects. Grossman and Wigdor [12] provided an extensive review of the existing work on interactive surfaces and developed a taxonomy for classification of this research. This framework takes into account the perceived and the actual display space, the input space and the physical properties of an interactive surface. As shown in their work 3D volumetric visualizations are rarely being considered in combination with 2D direct surface input.

Even on monosopic touch surfaces, the size of the human fingers and the lack of sensing precision can make precise touch screen interactions difficult [7, 17]. Some approaches have addressed this issue, for example, by providing adjustable [7] or fixed cursor offset [24], by scaling the cursor motion [7] or by extracting the orientation of the user's finger [17].

The kinematics of point and grasp gestures and the underlying cognitive functions have been studied by many research groups [11, 19, 32]. For instance, it has been shown that total arm movement during grasping actually consists of two distinct component phases: (1) an initial, *ballistic phase* during which the user's attention is focused on the object to be grasped (or touched) and the motion is basically controlled by proprioceptive senses, and (2) a *correction phase* that reflects refinement and error-correction of the movement, incorporating visual feedback in order to minimize the error between the hand or finger, respectively, and the target [18]. Furthermore, Mac Kenziea et al. [19] have investigated the real time kinematics of the limb movements in a Fitt's task and have shown that, while Fitt's law holds for the total limb-movement time, humans usually start sooner decelerating the overall motion, if the target seems to require more precision in the end phase. The changes of the kinematics and control of the reaching tasks within virtual environments have been investigated by Dvorkin et al. [9] or Viau et al. [30]. Valkov et al. [29] have shown that users are, within some range, insensitive to small misalignments between visually perceived stereoscopic depth and the sensed haptic feedback when touching the object. They proposed to manipulate the stereoscopically displayed scene in such a way that the objects are moved towards the screen when the user reaches for them. However, the problem is that objects have to be shifted in space, which might lead to a disturbed perception of the virtual scene for larger manipulations.

EXPERIMENT

In this section we describe the experiment in which we have analyzed where users would touch the surface of the projection wall for objects at different 3D positions in space rendered stereoscopically with positive, negative and zero parallax. We have also examined if the stereoscopic parallax impacts users' performance time or the kinematics of the touch gestures.

Experimental Setup

For the experiment we used a multi-touch enabled passive stereoscopic back projection system. The prototype is illustrated in Figure 2. The multi-touch technology of this surface is based on the *Rear-DI* [22] principle. We use a $200cm \times 161cm$ projection screen as touch surface, and a total of six infrared (IR) illuminators (i. e., high power IR-LED lamps) for back-lighting this surface. Since our projection screen is made from a mat, diffusing material, we do not need an additional diffusing layer for it. A digital video camera (PointGrey Dragonfly2) equipped with a wide-angle lens and a matching IR band-pass filter is mounted at a distance of $3m$ from the screen and captures an 8-bit monochrome video stream with a resolution of 1024×768 pixels ($2.81mm^2$ precision on the surface) at 30 frames per second (fps). For visualization we use passive stereoscopic back projection with circular polarization. Two DLP projectors with a resolution of 1280×1024 pixels ($1.56mm$ effective pixel-width, brightness $2800\ ANSIlumen$) provide stereo images for the left and right eye of the user. In order to perceive a stereoscopic image, subjects wear circular polarization glasses. For detection of the touch input we use a modified version of NUI Group's CCV software. The software needed for the experiment runs on a computer with Intel Core i7 @ $2.66GHz$ processor, $6GB$ RAM and nVidia GTX295 graphics card. As illustrated in Figure 2, an optical tracking system tracks the position of the user's head in order to provide view-dependent rendering. However, during the experiments subjects were not moving in front of the projection wall, and therefore head tracking was not used. All participants were recorded with a video camera (640×480 @ $30fps$) during the experiment.

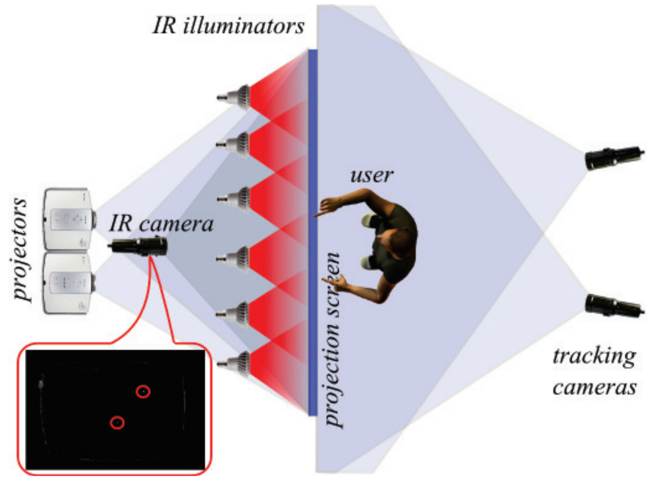


Figure 2. Illustration of the multi-touch-enabled stereoscopic projection wall used in the experiment. The inset shows a screenshot of the touch IR camera for touch recognition.

Materials and Methods

We have used the *Porta test* and the *Dolman test* to determine a subject's *sighting dominant eye* [20]. Subjects exhibiting differing eye dominance in the two tests were excluded from the experiment. Next, subjects judged in a *two-alternative forced choice* [10] task the parallax of four small shapes displayed stereoscopically on the projection wall (two with positive and two with negative parallax) in order to verify their ability of stereoscopic vision.

Subjects were positioned in front of the projection screen in such a way that they could conveniently perform all touch

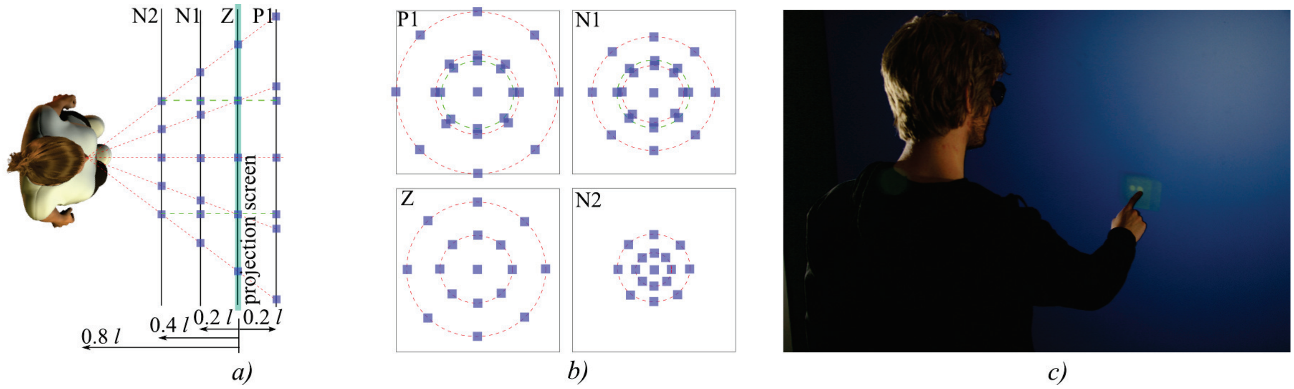


Figure 3. Experiment design; (a) top view of the object arrangement; (b) object arrangement for each parallax plane; (c) photo of a subject while participating in the experiment.

gestures during the experiment with their dominant arm. In a pilot experiment we determined an optimal distance to the projection screen of about 0.8 the subjects arm-length (l). This distance provided an operational radius of $r = 0.6 \cdot l$ around the projection of the subject's head position on the wall (see Fig 3(a)). We marked the corresponding position for each subject on the floor. Subjects were told to remain in this position during all trials of the experiment. If both stereopsis and eye-dominance tests were accomplished successfully, a written task description of the experiment was presented via slides on the projection wall.

For the experiment we have used the method of constant stimuli. In this method the object positions are not related from one trial to the next, but presented randomly and uniformly distributed. For visual stimuli we have used small spheres with a size of $1.5cm$, which ensured a clearly visible target with a reasonable stereoscopic impression; the center of the sphere indicated the exact position subjects should touch. For each trial, the sphere was surrounded by a semi-transparent box to provide additional depth cues (such as perspective distortion, texturing, etc.) to the user. As illustrated in Figure 3(c), we adjusted the color of the box and sphere as well as the background in such a way that stereoscopic crosstalk between the stereoscopic images for the left and right eye was minimized.

In each trial, the subject's task was to touch the center of the sphere, hold her finger at the same position until the object disappeared ($200ms$ after the touch was detected) and then release her finger from the touch wall. $200ms$ after subjects moved their fingers away from the touch surface a new object was displayed, which indicated the beginning of the next trial.

As illustrated in Figures 3(a) and (b), the objects used in the trials were arranged in concentric circles on four different planes parallel to the projection plane at $z = 0$ in a left-handed coordinate system. Since users are more sensitive to discrepancies between visual and tactile feedback if objects are display with positive parallax not accessible for direct touch interaction (cf. [29]), we have focused in this research primarily on objects exhibiting negative parallax. Therefore we have tested two parallax planes, called N1 and N2, with negative parallax at distances $z = 0.2 \cdot l$ and $z = 0.4 \cdot l$, respectively. In addition, we have tested one plane P1 with positive parallax at $z = -0.2 \cdot l$, and the plane Z aligned with the projection plane ($z = 0$), i. e., with zero parallax. The plane P1 was chosen to be relatively close behind the projection surface. If the plane has been chosen to be further behind, it would have been more likely for the subjects to accidentally hit the projection screen while still in the ballistic phase of the motion. The arrangement in concentric circles was used to provide symmetrical view conditions for the objects on the same plane, i. e., with same stereoscopic parallax. As mentioned above, for the $z = 0$ plane we have chosen the radius of the outer circle to match the maximal (convenient) reach distance r of the user, i. e., $0.6 \cdot l$. The inner circle had half the radius ($0.3 \cdot l$). On the planes N2, N1 and P1 the radii of the circles were selected in such a way that corresponding objects across all planes were positioned on a line of sight extending from the user (see Figure 3(a)), thus the users' hand movement distances were the same across all conditions. In addition to these locations, we have added on each plane circles with a constant radius of $0.3 \cdot l$ in order to test also different stimuli that depend only on the stereoscopic parallax, i. e., differ only in their z values.

Participants

13 male and 9 female subjects (age 22-29, $\varnothing : 25.4$; height $163cm{-}196cm$, $\varnothing : 179.9cm$) participated in the experiment. All subjects were right-handed. We determined for 15 subjects that their right eye was dominant (8 male and 7 female), and for 7 subjects (5 male and 2 female) that their left eye was dominant. All had normal or corrected to normal vision. 11 of the subjects wore glasses or contact lenses and none of them reported amblyopia or known stereopsis disruptions. 11 subjects had experience with stereoscopic projections, and 9 had already participated in a study in which stereoscopic projections were used. All subjects were naïve to the experimental conditions. The total time per subject including pre-questionnaire, instructions, training, experiment, breaks, and debriefing took 35 minutes. Subjects could take a break at any time. In addition, after each 45 trials subjects had to take breaks of two minutes in

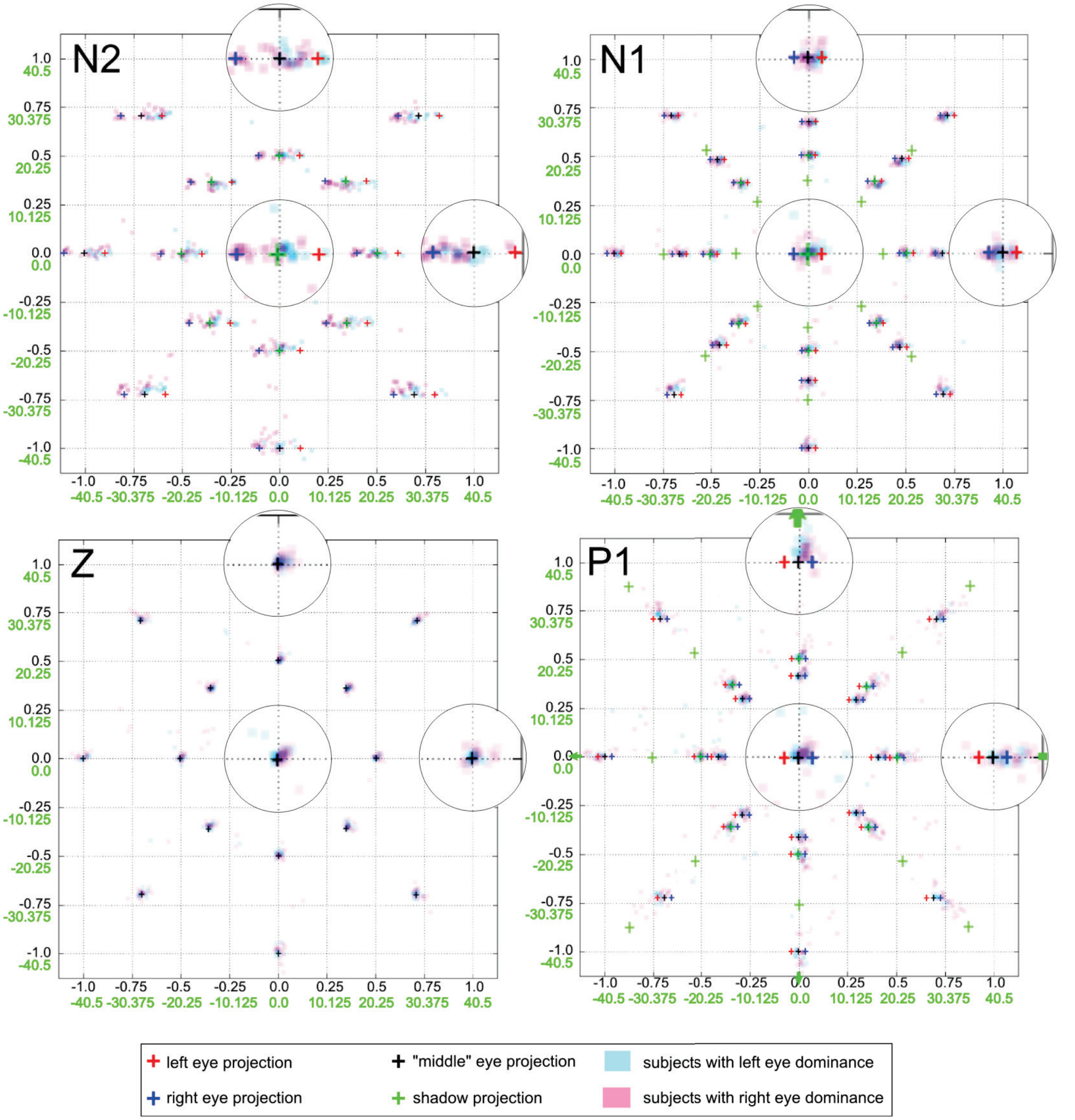
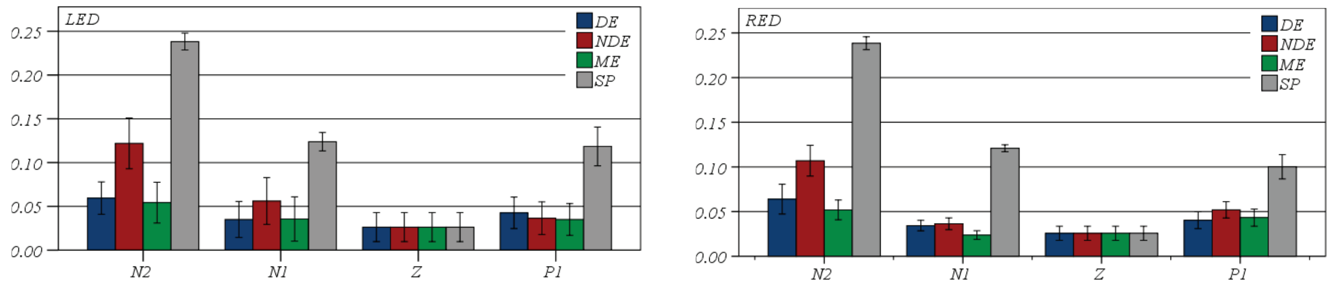


Figure 4. Individual touch results for all trials from the experiment: (top left) shows the touch locations for the subjects in condition N2, (top right) for condition N1, (bottom left) for condition Z and (bottom right) for condition P1. The crosshairs illustrate the position of the different touch targets: (blue) corresponds to dominant eye, (red) corresponds to non-dominant eye, (black) corresponds to the center, and (green) to the shadow projection of the stereoscopic object. The green numbers show the corresponding values in centimeters calculated for subject with $180cm$ body height. The insets show zoomed by factor $\times 2$ views for some of the clusters.

order to minimize errors due to exhaustion or poor concentration. Subjects were students or members of the departments of computer science, mathematics and psychology. Some subjects receive class credits for their participation.

RESULTS

In this section we summarize the results in terms of the touch points and time. For one subject we observed differing eye dominance in the *Porta test* and the *Dolman test*, and therefore excluded this subject from the experiment.



(a) (b)

Figure 5. Mean distances from the target points for different parallax surfaces for subjects from (a) Group LED and (b) Group RED. The vertical bars show the standard error.

Touch Points

Since the objects in our experiment were arranged in concentric circles centered at the subject's eye level, we define the focus point on the projection wall as the origin of a 2D coordinate system, with the y-axis running from bottom to top, and the x-axis running from left to right. As units for both axes we have chosen the subject's maximal (convenient) reach distance $r = 0.6 \cdot l$ (cf. Section *Materials and Methods*). We express the coordinates of all touches performed by the subject in terms of this coordinate system. Since the coordinate systems used for the different subjects take into account the differing arm lengths and body heights of the subjects, the coordinates of the touch points are already normalized and can be compared directly among all subjects.

The individual touch locations for all trials are plotted in Figure 4. The crosshairs illustrate the positions of the different touch targets, i. e., blue corresponds to the projection for the dominant eye (DE), red to the projection for the non-dominant eye (NDE), black to the midpoint point between both projections (middle eye, ME), and green to the orthogonal “shadow” projection (SP) of the stereoscopic object.

We have not found a significant difference in the data for male and female participants (two-sided t-test, $p = 0.932$ for the x-coordinate and $p = 0.637$ for the y-coordinate), so we have pooled the results for all subjects. We have calculated for each tested object the corresponding unified coordinates for the four considered touch targets, i. e., DE, NDE, ME and SP, and determined the distances between the performed touches and the corresponding target points using a 2D Euclidean metric. With a two-sided t-test we found a significant difference between the mean distances for subjects with left eye-dominance and for subjects with right-eye dominance ($t(20) = 2.174, p = 0.042 < 0.05$). Mean distance for the left-dominant subjects was $0.094 \cdot r$ ($SD = 0.0284$), and for the right dominant subjects it was $0.075 \cdot r$ ($SD = 0.0135$). Thus, we split the results for the two groups, i. e., left eye dominance (LED) group and right eye dominance (RED) group, in the subsequent analysis.

Figure 5 (a) shows a bar plot of the distances from the target points for different parallax surfaces for the group LED. We have calculated the mean distances to each target point for the four different parallax planes for each subject of the LED group. Those mean values were then analyzed with a factorial analysis of variance (ANOVA), testing the within-subjects effects of target point and stereoscopic parallax. The analysis revealed a significant main effect for the parallax ($F(3, 96) = 59.61$, $p < 0.001$) as well as for the target point ($F(3, 96) = 69.69$, $p < 0.001$). Post-hoc analysis with the Tukey test showed that subjects touched significantly closer to an object that is displayed on the surface with zero parallax compared to objects displayed with positive or negative parallax ($p < 0.001$ for P1, N1 and N2). Furthermore, there was a significant difference between the touch targets for objects displayed with strong negative parallax N2 and objects displayed with other parallaxes ($p < 0.001$ for P1, Z and N1). We have not found a significant difference between planes P1 and N1 ($p = 0.919$). The post-hoc analysis also showed that the touch points were significantly farther away from the SP target than from all other targets ($p < 0.001$ for DE, NDE and ME). Furthermore, subjects from group LED touched significantly further away from the NDE target in comparison to the targets DE ($p = 0.034 < 0.05$) or ME ($p = 0.01 < 0.05$), but significantly closer than to the target SP ($p < 0.01$). For the LED group, we have not found a significant difference between the two targets, which were closest to the subjects' touch points, i. e., DE and ME ($p = 0.973$).

Figure 5 (b) shows a bar plot of the distances from the target points for different parallax surfaces for the group RED. The mean distances were then analyzed with a factorial analysis of variance (ANOVA), testing the within-subjects effects of target point and stereoscopic parallax.

Again, we calculated the mean distances to each target point for the four different parallax planes for each subject of the RED group and performed a factorial ANOVA to test the within-subjects effects of target point and stereoscopic parallax. The analysis revealed a significant main effect for the parallax ($F(3, 224) = 230.68$, $p < 0.001$) as well as for the target point ($F(3, 224) = 254.19$, $p < 0.001$). Post-hoc analysis with the Tukey test showed that it was also significantly easier for subjects from group RED to touch an object with zero parallax compared to all other parallaxes ($p < 0.001$ for P1, N1 and N2), and there was a significant difference between strong negative parallax N2 and all other

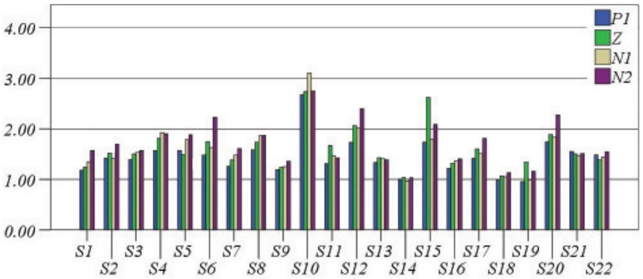


Figure 6. Performance times per subject and parallax.

parallaxes ($p < 0.001$ for P1, Z and N1). As for the LED group, we have not found a significant difference between planes P1 and N1 ($p = 0.463$). Similar to subjects from group LED, subjects from group RED touched significantly farther away from the SP touch target in comparison to all other targets ($p < 0.001$ for DE, NDE and ME). Furthermore, subjects touched significantly farther away from the NDE target in comparison to the targets DE ($p < 0.001$) or ME ($p < 0.001$), but significantly closer than to the target SP ($p < 0.001$). As for the group LED, we found no significant difference between the ME and the DE targets ($p = 0.491$).

Performance Time

Figure 6 shows the mean time elapsed until a subject touched a corresponding object, for each subject and parallax planes P1, Z, N1, N2. We have analyzed the results with a one-way ANOVA, testing the within subject effect of stereoscopic parallax on the mean performance time. We have not found a significant main effect ($F(3, 18) = 1.489,\ p = 0.223$), i. e., the subjects performance time is almost the same for objects on planes P1, Z, N1 and N2. The estimated mean value for the performance time on the parallax plane P1 is $1.446s$ ($SD = 0.3599$), for the Z plane is $1.608s$ ($SD = 0.4287$), for plane N1 $1.575s$ ($SD = 0.4467$) and for plane N2 $1.710s$ ($SD = 0.4359$).

Again, we have not found a significant difference between male and female subjects (two-sided t-test, $p = 0.07$). The mean time for the female subjects was $1.73s$ ($SD = 0.352$) and for the male subjects $1.91s$ ($SD = 0.532$).

DISCUSSION

In general, the results for the LED and RED group show the same qualitative behavior and differ only in quantity. Right-handed subjects with right eye dominance perform significantly more precise touch gestures than right-handed subjects with left eye dominance. However, subjects from both groups tend to choose the same strategy to select a stereoscopic object on a two-dimensional touch surface.

As it can be seen in Figure 4, the touch points for planes N2 (top left), N1 (top right) and P1 (bottom right) are more scattered than the touch points on the Z plane (bottom left), although the size of the projected images for objects on P1 is smaller the size of the projections for objects on the Z plane. Furthermore, the touch points on planes N1 and P1 are comparably scattered, although the projected images for objects on N1 are greater then those on P1. This indicates that touching objects displayed with positive or negative stereoscopic parallax on a 2D surface induces more imprecision than touching objects with zero parallax. The touches on the N2 plane are more scattered compared to those on all other parallax planes and, as described in the section *Results*, the calculated distances to the target points are significantly larger than those for the planes N1, Z and P1. Thus, imprecision increases with stereoscopic parallax, in particular for objects displayed with negative parallax.

As described in the section *Results*, we have not found a significant difference between the per-subject performance times for different parallaxes. Nevertheless, Figure 6 shows that most of the users performed more slowly for objects on the N2 plane than for other objects. The inverse tendency can be seen for objects displayed stereoscopically with positive parallax, i. e., the objects on P1. An analysis of the video records that we made during the experiment revealed that for the objects on N2 and N1 most users perform a “usual” point gesture until they reach the visual representation of the object and then move the finger slowly through it until it reaches the interactive surface, which may be an explanation for the increased performance times. In contrast, some of the users reported that they were “surprised by the surface” while performing some of the touch gestures in order to select objects behind the surface. This also may have had an impact on the decreased performance times and precision, since in these cases, the gesture ended prematurely, without users fully executing the slower and more precise *correction phase* (cf. Section *Introduction*). Furthermore, since the motion of a user’s arm during a touch gesture may differ very much among users and for different object positions, the prematurely ended gestures may have led to the “random touches”, i. e., outliers, on P1, as may be seen in Figure 4.

None of the subjects complained about touch difficulties, for example, accidentally recognized touches, during the experiments. Most of the subjects have observed the parallax problem described in the introduction (see Figure 1) and reported that for objects displayed with negative stereoscopic parallax it was difficult to get a stereoscopic impression when touching the surface behind the object with the finger. None of the subjects evaluated this effect as a strong distraction from the interaction, and some of the subjects, in particular those with lower performance times, have not noticed it at all. Interestingly, some of the subjects reported difficulties to merge the objects on the N2 plane, although they were within the wide accepted maximal distance for positive parallax. This may be due to the fact that the participants were relatively close to the projection wall and thus were more sensitive to small mismatches due to resolution or illumination constraints.

DESIGN IMPLICATIONS

Even though our analyses show that the ME and DE targets are best guesses for the location of the on-surface touch targets for stereoscopic objects, the calculated mean distances to the actual touch points are still rather large. For instance, the mean distance between the DE targets and the corresponding touch points is 0.0656 (for group LED) and 0.0493 (for group RED), which corresponds to $2.65cm$ (group LED) respectively $1.99cm$ (group RED) for a user with a height of $180cm$. Furthermore, the video recordings of the subjects during the experiment reveal that during most of the trials they neither touched the DE nor the ME target, but rather a point "in-between" both touch targets.

We can express this new *intermediate* target point (IMD) as:

$$\text{IMD} = \text{ME} + \alpha \cdot (\text{DE} - \text{ME}) \quad (1)$$

The parameter $\alpha \in [0, 1]$ determines the position of the point IMD according to the segment (DE − ME). For instance, for $\alpha = 1$ the IMD coincides with DE, whereas for $\alpha = 0$ it coincides with ME. One can find the optimal value of α with an optimization algorithm, minimizing the mean distance between the touch points and the IMD target.

Let $P_i \in \mathbb{R}^3$ be the position of the i-th tested object, with $i = 1, \ldots, n$ and $T_{ij} \in \mathbb{R}^2$ is the actual touch point of the j-th trial ($j = 1, \ldots, m$) for the object at P_i. Then the optimization could be expressed as:

$$\min_{\alpha \in [0,1]} \left(\frac{1}{m \cdot n} \sum_{i=1}^{n} \sum_{j=1}^{m} \| T_{ij} - \text{IMD}_i \| \right) \quad (2)$$

Using Equation 2 the optimal α value for the LED group is 0.551 with mean error $\epsilon = 0.0266$, i.e. $1.07cm$ for a subject with $180cm$ body height. For the RED group we have determined $\alpha = 0.165$, $\epsilon = 0.0365$ ($1.47cm$), which suggests that the subjects in the RED group choose a slightly different strategy than the subjects from the LED group.

Apparently, the optimal α may be influenced by several parameters such as the parallax, the users' handiness, performance speeds and preferences. Nevertheless, the reported values could be used to optimize the selection of a stereoscopically rendered virtual object on an interactive surface if the user's eye dominance is known. We expect even greater improvements by using parallax and eyedness dependent α values, which will be addressed in future works.

CONFIRMATORY STUDY

In this section we describe a confirmatory study in which we applied the results of our experiment in a real-world application. The test application has been developed in the scope of the AVIGLE project [1], which explores novel approaches to remote sensing using a swarm of *Miniature Unmanned Aerial Vehicles (MUAVs)*, equipped with different sensing and network technologies. At the end of the pipeline the currently available sensor data from the MUAVs and their status are displayed to a human operator, while new data continuously arrive. The user can interact with this visualization, for instance, by changing the viewpoint to the virtual environment. In addition, she can define new positions for each MUAV moving their visual representation in the virtual environment (see Figure 7). Since MUAVs within a swarm usually fly at different altitudes, stereoscopic visualization is essential to provide additional depth cues. In order to select the correct MUAV from the swarm, it is important to determine the exact touch target for each virtual MUAV as described above. The goal of the confirmatory study was to verify if operators of the AVIGLE system perform better with the touch targets that we determined in the experiments in comparison to the other approaches.

Procedure

8 expert operators of the system participated in this study (6 had right eye dominance and 2 had left eye dominance). In a within-subject design experiment, we placed the operators in front of the stereoscopic multi-touch surface used in the experiment (cf. Section *Experimental Setup*). The visual stimulus was a typical scene of our application showing a view with a swarm of 12 stereoscopically displayed virtual MUAVs (see Figure 7). We tested a subset of 42 locations from our initial experiment for the swarm. The position of each MUAV within the swarm and its altitude was randomized in each trial with respect to the minimal and maximal inter-MUAV distances. In each trial a MUAV was highlighted and the operator's task was to select it. We gave visual feedback about the selection, so that the operator could retry until the highlighted MUAV was selected. We tested two different on-surface targets, i.e., DE and ME, against the IMD (as explained from Section *Design Implications*). In order to simplify the confirmatory study, we averaged the IMD across both groups and rounded the IMD to 0.4. The swarm's position and the on-surface targets were randomized and uniformly distributed. To determine if a MUAV had been selected we constructed a ray with origin at the position of the dominant eye (for DE), or at the camera's position (ME), or shifted by $0.4 \cdot (\text{IOD}/2)$ towards the dominant eye (IMD) and the actual touch point; IOD denotes the interocular distance. A collision test between a cone around the ray with radius $0.03 \cdot l$ ($1.21cm$) at the projection wall and the mesh of each drone was used to determine the selection. The radius $0.03 \cdot l$ ($1.21cm$) was half the standard deviation (cf. Section *Results*). We measured the number of errors in terms of the number of repetitions required to select the correct MUAV, as well as the time required to perform the task.

Results

The mean number of touches the operators required to select the correct MUAV was 2.15 ($SD = 2.291$) for the touch target DE, 2.10 ($SD = 1.951$) for the touch target ME, and 1.73 ($SD = 1.634$) for the touch target IMD. We have analyzed the mean number of touches for each target and subject with a one-way ANOVA over all trials. We have found a significant main effect ($F(2, 1005) = 4.47,\ p = 0.12 < 0.05$) of the touch target on the number of touches required to hit the correct MUAV. Post-hoc analysis with the Tukey test showed that operators required significantly less touches to hit the correct MUAV, when we used the IMD touch target instead of the touch targets DE ($p = 0.018 < 0.05$) or ME ($p = 0.042 < 0.05$). We have not found a significant difference between the number of touches required to hit the MUAV when the touch target was ME or DE. Operators required approximately 1.7 touches to select the highlighted MUAV using our approach. This large value is caused by

Figure 7. Multi-touch interaction with a swarm of virtual MUAVs flying over a virtual city model.

the small radius of the touch target as explained above. However, in a swarm of several MUAVs, touch targets may overlap if their radii are chosen too large.

The mean time the operators required to select the correct MUAV was $2.77s$ ($SD = 2.819$) for the touch target DE, $3.93s$ ($SD = 7.836$) for the touch target ME, and $2.65s$ ($SD = 4.421$) for the touch target IMD derived from our experiments. We have analyzed the mean performance time for each target and subject with a one-way ANOVA over all trials. We have found a significant main effect ($F(2, 1005) = 5.687,\ p = 0.012 < 0.05$) of the touch target on the time required to hit the correct MUAV. Post-hoc analysis with the Tukey test showed that operators required significantly more time to hit the correct MUAV, when we used the ME touch target ($p < 0.05$ for both IMD and DE). We have not found a significant difference between the time required to hit the MUAV when the touch target was DE or IMD ($p = 0.958$).

CONCLUSION AND FUTURE WORK

In this paper we have made some first steps to analyze the relation between 3D positions of stereoscopically rendered objects and the *on-surface touch point*, where the user touches the surface. We performed an experiment in which we determined the positions of the users' touches for objects which are displayed with different parallaxes. The results of the experiment show that users tend to touch between the projections for the two eyes with an offset toward the dominant eye's projection. We gave guidelines to set the on-surface touch points for stereoscopically displayed objects on a multi-touch surface. These guidelines depend on the user's head position as well as eye dominance; we explained how both can be easily determined. We have verified these guidelines in a real-world application and showed the benefits in terms of task performance over other approaches. Our results give novel implications for the development of future touch-enabled interfaces which support stereoscopic visualization.

While these initial findings provide useful insights into how users touch 3D stereoscopically displayed objects on a 2D touch surface, further studies are required to fully understand users' strategies in such setups. First, the scope of the experiment can be expanded to include varying user positions and orientations, as well as objects displayed with larger parallaxes or different projection sizes. Furthermore, we will more deeply consider the impact of handedness as well as eye dominance. The question arises if the IMD point be expressed in a model considering all these factors. We will extend this research and consider also other stereoscopic multi-touch surfaces such as table-top or mobile devices.

The combination of multi-touch technology and stereoscopic display provides an enormous potential not only for simple selection tasks, but also for richer interaction such as 3D manipulations of or collaborative interactions with stereoscopically rendered virtual scenes. These and similar research questions and challenges will be addressed in the future in the scope of the iMUTS project.

ACKNOWLEDGMENTS

The work was partly supported by grants from the Deutsche Forschungsgemeinschaft (DFG) in the scope of the iMUTS project. The AVIGLE project is part of the Hightech.NRW initiative funded by the Ministry of Innovation, Science and Research of the German State of North Rhine-Westphalia. We acknowledge the cooperation of the subjects as well as expert operators, which participated in the experiments and the confirmatory study.

REFERENCES

1. AVIGLE - Avionic Digital Service Platform. http://www.avigle.de/.
2. Immersion SAS. http://www.immersion.fr/.
3. iMUTS - Interscopic Multi-Touch Surfaces. http://imuts.uni-muenster.de/.
4. InSTInCT - Touch-based interfaces for Interaction with 3D Content. http://anr-instinct.cap-sciences.net/.
5. H. Benko and S. Feiner. Balloon selection: A multi-finger technique for accurate low-fatigue 3d selection. *IEEE Symposium on 3D User Interfaces (3DUI)*, pages 79–86, 2007.
6. H. Benko, A. D. Wilson, and R. Balakrishnan. Sphere: multi-touch interactions on a spherical display. In *ACM Symposium on User Interface Software and Technology (UIST)*, pages 77–86, 2008.
7. H. Benko, A. D. Wilson, and P. Baudisch. Precise selection techniques for multi-touch screens. In *ACM Proceedings of Human Factors in Computing Systems (CHI)*, pages 1263–1272, 2006.
8. J.-B. de la Rivière, C. Kervégant, E. Orvain, and N. Dittlo. Cubtile: a multi-touch cubic interface. In *ACM Symposium on Virtual Reality Software and Technology (VRST)*, pages 69–72, 2008.

9. A. Y. Dvorkin, R. V. Kenyon, and E. A. Keshner. Reaching within a dynamic virtual environment. *Journal of NeuroEngineering and Rehabilitation*, 4(23), 2007.

10. J. Ferwerda. Psychophysics 101: How to run perception experiments in computer graphics. In *SIGGRAPH core*, 2008.

11. L. Geniva, R. Chua, and J. T. Enns. Attention for perception and action: task interference for action planning, but not for online control. *Experimental Brain Research*, 185(4):709–717, 2008.

12. T. Grossman and D. Wigdor. Going deeper: a taxonomy of 3d on the tabletop. In *IEEE TABLETOP '07.*, pages 137 – 144, 2007.

13. T. Grossman, D. Wigdor, and R. Balakrishnan. Multi-finger gestural interaction with 3d volumetric displays. In *ACM SIGGRAPH '05*, pages 931–931, 2005.

14. J. Y. Han. Low-cost multi-touch sensing through frustrated total internal reflection. In *ACM Symposium on User Interface Software and Technology (UIST)*, pages 115–118, 2005.

15. M. Hancock, S. Carpendale, and A. Cockburn. Shallow-depth 3d interaction: design and evaluation of one-, two- and three-touch techniques. In *ACM Proceedings of Human Factors in Computing Systems (CHI)*, pages 1147–1156, 2007.

16. O. Hilliges, S. Izadi, A. D. Wilson, S. Hodges, A. Garcia-Mendoza, and A. Butz. Interactions in the Air: Adding Further Depth to Interactive Tabletops. In *ACM Symposium on User Interface Software and Technology (UIST)*, pages 139–148, 2009.

17. C. Holz and P. Baudisch. The generalized perceived input point model and how to double touch accuracy by extracting fingerprints. In *ACM Proceedings of Human Factors in Computing Systems (CHI)*, pages 1165–1174, 2009.

18. G. Liu, R. Chua, and J. T. Enns. Attention for perception and action: task interference for action planning, but not for online control. *Experimental Brain Research*, 185:709–717, 2008.

19. C. L. MacKenziea, R. G. Marteniuka, C. Dugasa, D. Liskea, and B. Eickmeiera. Three-dimensional movement trajectories in fitts' task: Implications for control. *The Quarterly Journal of Experimental Psychology Section A*, 39(4):629–647, 1987.

20. A. P. Mapp, H. Ono, and R. Barbeito. What does the dominant eye dominate? a brief and somewhat contentious review. *Perception & Psychophysics*, 65(2):310–317, 2003.

21. A. Martinet, G. Casiez, and G. Grisoni. The design and evaluation of 3d positioning techniques for multi-touch displays. In *IEEE Symposium on 3D User Interfaces (3DUI)*, 2010.

22. C. Müller-Tomfelde, J. Schöning, J. Hook, T. Bartindale, D. Schmidt, P. Oliver, F. Echtler, N. Motamedi, P. Brandl, and U. Zadow. Building interactive multi-touch surfaces. In C. Mller-Tomfelde, editor, *Tabletops - Horizontal Interactive Displays*, Human-Computer Interaction Series, pages 27–49. Springer London, 2010.

23. J. Pierce, A. Forsberg, M. Conway, S. Hong, R. Zeleznik, and M. Mine. Image Plane Interaction Techniques in 3D Immersive Environments. In *ACM Interactive 3D Graphics*, pages 39–44, 1997.

24. R. L. Potter, L. J. Weldon, and B. Shneiderman. Improving the accuracy of touch screens: an experimental evaluation of three strategies. In *ACM Proceedings of Human Factors in Computing Systems (CHI)*, pages 27–32, 1988.

25. J. L. Reisman, P. L. Davidson, and J. Y. Han. A screen-space formulation for 2d and 3d direct manipulation. In *ACM Symposium on User Interface Software and Technology (UIST)*, pages 69–78, 2009.

26. D. Schmalstieg, L. M. Encarnaç ao, and Z. Szalavári. Using transparent props for interaction with the virtual table. In *ACM Interactive 3D graphics '99*, pages 147–153, 1999.

27. J. Schöning, F. Steinicke, D. Valkov, A. Krüger, and K. H. Hinrichs. Bimanual interaction with interscopic multi-touch surfaces. In *IFIP TC13 Conference on Human-Computer Interaction (INTERACT '09)*, pages 40–53, 2009.

28. F. Steinicke, T. Ropinski, G. Bruder, and K. Hinrichs. Interscopic User Interface Concepts for Fish Tank Virtual Reality Systems. In *IEEE Proceedings of VR2007*, pages 27–34, 2007.

29. D. Valkov, F. Steinicke, G. Bruder, K. Hinrichs, J. Schöning, F. Daiber, and A. Krüger. Touching floating objects in projection-based virtual reality environments. In *Joint Virtual Reality Conference of EuroVR - EGVE - VEC*, pages 17–24, 2010.

30. A. Viau, A. G. Feldman, B. J. McFadyen, and M. F. Levin. Reaching in reality and virtual reality: a comparison of movement kinematics in healthy subjects and in adults with hemiparesis. *Journal of NeuroEngineering and Rehabilitation*, 1(11), 2004.

31. C. Ware and K. Lowther. Selection using a one-eyed cursor in a fish tank vr environment. *ACM Transactions on Computer-Human Interaction*, 4:309–322, 1997.

32. D. Whitney, D. A. Westwood, and M. A. Goodale. The influence of visualmotion on fast reaching movements to a stationary object. *Letters to Nature*, 423:869–873, 2003.

33. A. D. Wilson, S. Izadi, O. Hilliges, A. Garcia-Mendoza, and D. Kirk. Bringing physics to the surface. In *ACM Symposium on User Interface Software and Technology (UIST)*, pages 67–76, 2008.