

Comparing Indirect and Direct Touch in a Stereoscopic Interaction Task

Adalberto L. Simeone*
University of Portsmouth, Lancaster University

Hans Gellersen†
Lancaster University

ABSTRACT

In this paper we studied the impact that the directedness of touch interaction has on a path following task performed on a stereoscopic display. The richness of direct touch interaction comes with the potential risk of occluding parts of the display area, in order to express one’s interaction intent. In scenarios where attention to detail is of critical importance, such as browsing a 3D dataset or navigating a 3D environment, important details might be missed. We designed a user study in which participants were asked to move an object within a 3D environment while avoiding a set of static distractor objects. Participants used an indirect touch interaction technique on a tablet and a direct touch technique on the screen. Results of the study show that in the indirect touch condition, participants made 30% less collisions with the distractor objects.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Input Devices and Strategies;

1 INTRODUCTION

The combination of direct touch input and stereoscopic displays bridges the rich expressivity of multitouch input with the graphical capabilities provided by the addition of the depth perception. The potential of coupling these two technologies has been evidenced in various publications [4, 16]. However, one of the main drawbacks of the combination of these two technologies comes as a consequence of the input choice. In order to express an input, users need to touch the screen. This invariably causes an occlusion of the area used to display the 3D environment or data. This occlusion is not merely caused by one’s own fingertips, but depending on the distance between the user and the target touch point, the user’s arms might also occlude large parts of the screen. It also might lead to miss important details or cause errors due to the inability of seeing the area neighbouring our contact point.

To address this issue, we designed a user study in which participants were asked to move a biplane model in a 3D environment while avoiding distractor objects (see Figure 1). We chose this task as it requires users to perform multiple inputs across the screen. We believe this task to be reminiscent of real-world activities such as 3D modelling or 3D data browsing that require users to perform a sequence of 3D manipulations. Our goal was to evaluate whether using a tablet as an indirect touch input device would incur in better performance as indirect interaction negates the potential for occlusion resulting from direct touch. By indirect touch we mean the use of an external multi-touch device, as opposed to manipulating an object touch inputs not intersecting its screen projection (see [1, 3, 15]). Our results showed that while using the tablet to control the biplane in the 3D environment, participants made 30% less collisions with the distractor objects than in the direct touch condition. This shows that occlusion plays a significant role in direct touch 3D user interfaces.

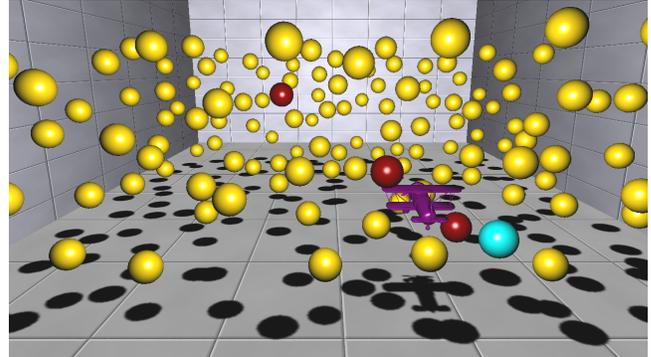


Figure 1: A trial from the user study showing the distractor spheres in gold, collided ones in red and the goal in light blue; the controlled biplane in purple. This trial used 125 spheres having a *large* radius.

While any external device can be considered to be indirect (e.g., gamepads, wands, 3D-mice), we narrowed our focus to the domain of touch devices. Multi-touch based user interfaces in both scientific and consumer contexts are now popular enough (e.g., [17]), that a wide user-base is familiar with its use. Therefore, we believe that in those situations where indirect touch can be advantageous, switching to an indirect technique will be easier than adopting an input device based on an entirely different paradigm. Furthermore, indirect touch can also be used in those scenarios where the 3D display is distant or very large or when the use of the combination of mouse and keyboard is not viable, e.g., while standing.

2 RELATED WORK

The initial exploration of multitouch interaction technique started on monoscopic displays. Hancock et al. [5] proposed the concept of *shallow-depth* interaction. In their work, the authors describe three 6DOF direct touch interaction techniques, using one two or three fingers. A later work, *Sticky Tools* [6], proposed a technique mapping two DOF to each finger. Reisman et al. [14] proposed an adaptation of the well-known Rotate-Scale-Transform 2D metaphor to 3D manipulation. Their technique uses a constraint solver to preserve the original touch points location after each manipulation. The previously described techniques fall in the broader category of “integrated” techniques, where rotation and translation manipulations can happen simultaneously. Interaction techniques where translation is separated from rotation have been proposed as a way to overcome the difficulty of precisely controlling an integrated manipulation [10, 11].

Research on the combination of stereoscopic displays and multitouch input has focused on understanding how touching can be affected while perceiving 3D stereoscopic content [16]. One of the main issues associated with these technologies is the “*Vergence-Accommodation conflict*”, a problem arising due to the disparity between the perceived depth at which the 3D content is displayed and the depth of the screen on which the user is converging her/his gaze [7]. In extreme cases, i.e. while touching the screen to manipulate an object far beyond the screen or in front of it, users may ex-

*e-mail: adalberto.simeone@port.ac.uk

†e-mail: h.gellersen@lancaster.ac.uk

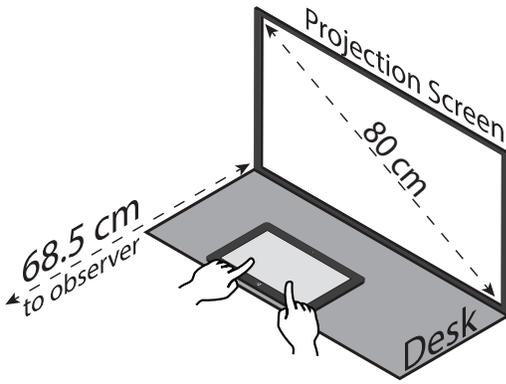


Figure 2: A sketch depicting the experimental apparatus, drawn in scale. The screen was supported by a metal frame.

perience the loss of the stereoscopic effect. To overcome this issue, a number of techniques have been proposed. Some allow users to specify a location in 3D space based on the 2D touch location such as the Balloon technique [1] or the Triangle Cursor [15]. More recently, *Void Shadows* [3] proposed the use of screen-plane shadows as proxies of an associated 3D object. By manipulating the projected shadows, the vergence-accommodation conflict is mitigated.

The use of external touch devices for indirect input has mainly been developed in the context of monoscopic visualization. Knoedel et al. [8] compared an adapted RST technique for an external tablet to its direct touch equivalent. Results from their study showed that the indirect technique is slower but more accurate. Ohnishi et al. [12] used an external tablet to map 2D input to the surface of an arbitrary 3D object. Liang et al. [9] proposed an interaction technique that uses an external tablet to manipulate an object shown on a larger distant monoscopic screen. Daiber et al. [2] described how to use a smartphone as an indirect input device for travelling in an environment shown in a larger 3D display.

The cited works explored indirect interaction as a means to use screen-depth input to *indirectly* specify a 3D location in space [1, 3, 15]. However, since input and display happen on the same device, these techniques can incur in the same risks caused by occlusion. Where input and display are separated, previous research has either focused on monoscopic displays [8] or on different tasks [2]. Our work focuses on the combination of an external touch input device, a stereoscopic 3D display and a path following task.

3 USER STUDY

We hypothesized that, as the number of objects in a 3D environment increases, then so do the chances of occluding them while interacting by direct touch. If users need to repeatedly interact with 3D stereoscopic content while parts of the screen are occluded by their hands or arms, it might then cause the loss of key information or negatively affect user performances. We designed a task representative of such lengthier interactions that can be found within real-world application scenarios. E.g., a 3D modeller applying numerous manipulations to a complex object or during the visualization of scientific or medical data where avoiding missing critical details can be of paramount importance. Our goal was to investigate whether using an indirect touch technique could improve users' performance if compared to direct touch input, in a task involving a cluttered 3D environment.

3.1 Apparatus

We used a rear-projection system based on the diffused illumination technique, measuring $70\text{cm} \times 38.5\text{cm}$ (with a diagonal of

$31.5''$, 80cm , see Figure 2). The 3D projector used was a BenQ w1070 displaying frame-sequential 3D images at a resolution of 1280×720 pixels at 60 fps per eye. In the direct condition, a Point Grey Grasshopper camera running at 60 fps was used to capture touch input. In the indirect condition we used a Microsoft Surface RT tablet where an application intercepted touch events. In both conditions, events were streamed through the network to a different system charged of rendering the 3D scene. The 3D application used for all the tasks implemented in this work was built on a custom 3D engine using SharpDX¹, a managed port of the DirectX API for C#. We used two parallel left and right perspective projections (i.e. not toed-in) using an interocular distance of 6.35cm and a distance of 68.5cm from the screen. We placed the tablet on a desk beside the projection screen and asked participants to remain sitting for the duration of the experiment. This was done to standardise the conditions of the experiment but was not a requirement of the technique.

3.2 Task

The task we designed consisted in controlling a model of a biplane within a densely cluttered environment (see Figure 4). The goal lies in reaching a specific sphere in the environment (the light blue one) while avoiding any impact with the *distractor* spheres (the golden ones). Once the biplane collides with the target sphere, it disappears and another target sphere appears in the environment. Each trial consisted of ten waypoints.

Each target sphere is randomly generated at a fixed distance from the previous one by choosing a random direction within the sphere determined by this distance and ensuring that the angle between these two directions was greater than $\pm 45^\circ$. In this way the user frequently switches the direction of travel, ensuring an ample coverage of the whole 3D environment. The distractor spheres are placed so that none of them completely occludes another one. All spheres are thus visible at the same time from the camera position. This ensures that the user cannot inadvertently collide with a distractor s/he cannot see. Furthermore, the spheres have been laid out so that there is enough space between each other for the biplane to pass through (i.e.: a minimum of $1.5x$ the size of its bounding sphere). The region in which distractors could be placed was scaled inwards by the length of the sphere's radius in all directions to ensure that no sphere could appear in such a way to intersect the room's walls (i.e.: the cluttered zone). A grid-like texture was applied to these walls and a shadow-mapping algorithm provided further depth cues to the users as to the exact whereabouts of the objects.

Collisions, i.e. *intersections* between the plane and the distractor objects, counted as errors and caused the collided sphere to turn red for the duration of the trial. All participants were instructed that movement in the environment was only allowed within the cluttered zone. This limitation was introduced in order to avoid participants moving the manipulable object outside the area delimited by the walls to reach the goal, in such a way to avoid hitting any of the displayed spheres and thus "cheat". Whenever the biplane collided with the boundaries of the cluttered zone, the system forbade further movement along that direction.

Three conditions were part of the study: the interaction *technique* (direct touch or indirect); the *number* of spheres in the environment (either 75 or 125) and the *size* of spheres (either small or large). Every trial was repeated four times, resulting in a total of 32 trials per participant. The biplane was placed always in the same starting location, i.e. at the center of the screen in the negative parallax region. The minimum distance between distractors was also enforced for this initial location in order to ensure that hitting a sphere at the beginning was unavoidable. The order of presentation of the two techniques was counter-balanced, while the order of the trials was randomised. The study lasted approximately 45 minutes.

¹<http://www.sharpx.org>

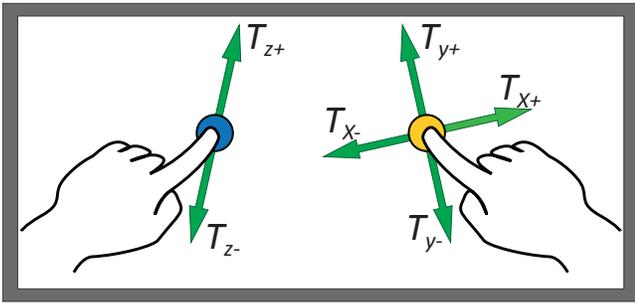


Figure 3: The input mappings of the interaction technique used. The finger intersecting the object (direct), or the first finger touching the tablet (indirect), is assigned to XY-translation on the camera plane. A non-intersecting second finger (direct) or the second finger touching the tablet (indirect) is assigned to Z-translation.

3.3 Interaction Technique

Since the task does not require users to rotate the model, for the direct touch modality we adapted a 3DOF state-of-the-art technique [10], originally intended for monoscopic visualization. Touching the chosen object with one finger allows the user to translate it on the XY camera plane. When another finger touches the screen in a location that does not intersect the object, the user can also translate it along the Z-axis. Moving the finger upwards translates the object further into the screen; conversely, moving it downwards translates the object towards the user. The indirect touch adaptation follows closely its direct touch counterpart (see Figure 3). Input expressed on the tablet is interpreted relative to the initial contact location of the first finger. Any incremental movement is applied to the object being controlled. Thus, regardless of the specific touch contact location on the tablet, dragging one’s finger will move the object in the appropriate direction from the position it currently is. The first finger to touch the tablet is considered to be the one assigned to XY translation. The second one is assigned to Z-translation and works in a similar fashion to the direct touch technique.

3.4 Participants and Procedure

Ten participants (7 Male 3 Female; $M = 26.5, SD = 3.27$) took part to this study. Before beginning the experiment, participants were asked to fill a demographics questionnaire, indicating their familiarity with 3D films and applications; with 3D monoscopic applications; with multi-touch devices in general. Each question ranged on a scale from 1 (very little familiarity) to 7 (very high). The background information reported by the participants indicates that they had little familiarity with 3D stereoscopic films or TV ($M = 1.22, SD = 0.69$) and 3D stereoscopic applications, including games ($M = 0.56, SD = 0.73$). However, they were more familiar with classic monoscopic games or applications ($M = 3.78, SD = 1.86$). They were also all very familiar with multi-touch devices ($M = 6.11, SD = 1.36$). After filling the questionnaire, participants carried out the experiment. Before starting each trial, they were required to rest their arms at their sides. A countdown of three seconds was displayed on screen before starting. A break of at least two minutes was observed before switching to the next technique. In total, 320 trials were collected. After running through all trials, we collected feedback from each participant in the form of semi-scripted interviews on their experience. Participants received £6.

4 RESULTS

In summary, indirect interaction allowed participants to perform the least number of errors across all conditions. We performed a repeated-measures one-way ANOVA on the data we collected. The

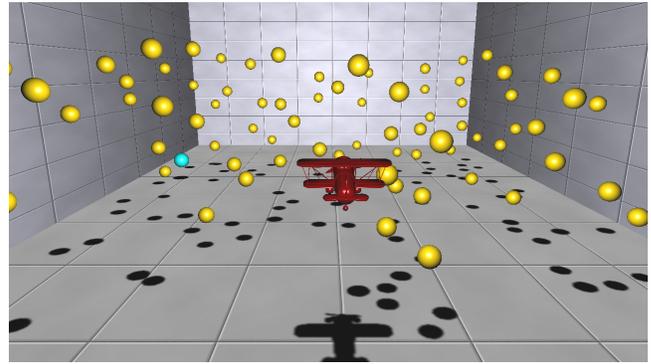


Figure 4: The picture shows one of the condition from the user study, where 75 distractor objects with a *small* radius are displayed.

analysis found a main effect of *technique* on the number of errors made by the participants ($F_{(1,318)} = 23.96, p < .001$). Pairwise comparisons show that the indirect technique led to significantly better results ($p < .001$). The indirect technique led participants to an overall mean of 10.56 errors per trial ($SD = 7.18$), opposed to 14.98 ($SD = 8.86$) in the other condition, a 30% decrease (see Figure 5).

Completion times were not affected in terms of *technique*: a mean result of 67.91s ($SD = 68.51$) in the direct condition and 64.33s ($SD = 24.03$) in the indirect condition. Although an interaction between *technique* and *number* was found ($p < .001$), pairwise comparisons show that direct interaction led to faster times in the condition with the most distractors, while indirect was faster in the other level. This could indicate that participants were frustrated by the inability of completing those trials in a satisfactory manner and proceeded more hurriedly.

We also analysed the coordination (or efficiency [18]) of the translation movements performed by the participants. This measure is defined as the ratio between the total distance travelled by the biplane to the shortest best path (calculated between each pair of waypoints). This *translation coordination* (T_c) can range from 0 to 1 where values close to 1 represent almost perfect routes (i.e. the participant always travelled along the shortest path); conversely, values close to 0 represent very inefficient routes (i.e. the participant travelled a long distance in the environment before eventually reaching the waypoint). An analysis of this T_c shows that it was significantly affected by *technique* ($F_{1,2997} = 39.822, p < .001$). The indirect technique led to better T_c results ($p < .001$): a mean ratio of 0.97 as opposed to 0.88 for the direct condition. This significant difference was further highlighted by an interaction of *technique* \times *number* \times *size* in all conditions but the one with the largest number of distractors and the biggest size ($p < .001$).

5 DISCUSSION

These results support our assumptions that occlusion issues caused by interacting with 3D stereoscopic content via direct touch do cause a substantial increase of the numbers of errors made: an average of 30% more errors when compared to the use of an indirect device. Given the popularity of 3D multi-touch interfaces for 3D scientific exploration (e.g., [17]), this can have implications in those tasks where users need to manipulate a 3D dataset in order to gain insights from its observation. If details tend to become occluded by the user’s hands or arms, this process becomes harder. Likewise, in a 3D multi-touch modelling application (e.g., [13]), occluding parts of the display causes users to lose awareness of the surroundings. Therefore the risk of incurring into unwanted operations are higher, such as moving a vertex or a face into a position already occupied by other elements. Furthermore, an indirect device by virtue

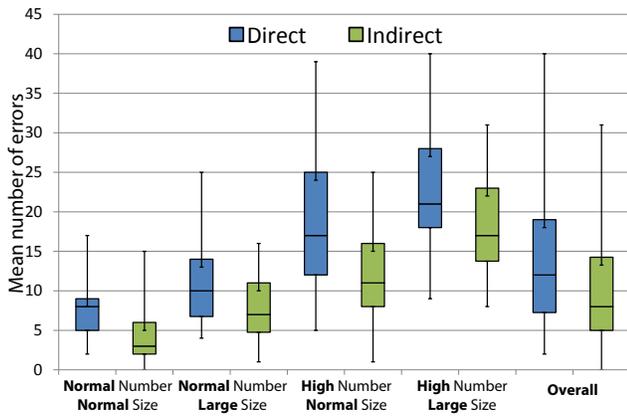


Figure 5: Errors made grouped by condition.

of not needing users to touch the 3D display in order to interact has a lower potential of incurring in the loss of the stereoscopic effect due to the vergence-accommodation conflict, whereas in the direct touch condition it was more likely to intersect objects not at screen-depth. On this regard, participant #10 commented that using the tablet “felt more relaxing”.

Indeed our own observations, the video-recordings and the exit interviews provided support to these conclusions. When we asked whether the participants focused more on the object or on their fingers, some commented (regarding the direct condition) that since they were not able to see the biplane model they had to rely on looking at the shadow it cast on the floor. We noticed how some participants tended to stop interacting to look at the display, now unobstructed, to resume the task after having assessed the situation. This was further confirmed by the significant differences obtained in terms of the *Translation Coordination* of the paths taken by the controlled object. In the indirect condition we obtained a minimal rate of unnecessary translation movements (0.97) as opposed to the direct condition (0.88). We believe the lack of obstructions in the indirect condition allowed participants to better estimate the optimal path from their current position to the goal one. Conversely, the inability of having a complete picture of their surroundings in the direct condition resulted in a higher incidence of frustration as voiced by our participants.

One drawback of the the indirect condition, is that participants commented that they focused more on the object rather than on the actual location of their hands. Indeed, participants were not usually aware of the boundaries of the tablet, thus tending to move their fingers outside. This issue could be mitigated in future indirect 3DUIs by informing the user when their touch contact location is close to the physical boundaries of the interactive surface, for example, by means of an on-screen indicator.

6 CONCLUSION AND FUTURE WORK

We have investigated the role that occlusion plays in touch-based interfaces for 3D stereoscopic displays. We evaluated participants in a study requiring them to manipulate an object through a path within a cluttered environment, while avoiding collision with distractors along the way. Our results showed that the input modality used has a significant impact on their performances. Thus we believe there is the need for further research in the areas of indirect touch interaction: there might be other areas besides occlusion where indirect techniques could provide better results. We focused on a simple task requiring 3DOF input, however most practical applications require 4-6DOF or more. Therefore, future research

should also look into applying indirect interaction to full 3D manipulation. Another direction consists in exploring whether these results are transferable also to smaller form factor touch devices. This could enable the use of walk-up indirect interaction techniques allowing users to manipulate 3D content through their own smartphones with non touch-based displays in public settings.

ACKNOWLEDGEMENTS

The first author was supported by a Marie Curie Intra European Fellowship (3STARS) within the 7th European Community Framework Programme.

REFERENCES

- [1] H. Benko and S. Feiner. Balloon selection: A multi-finger technique for accurate low-fatigue 3d selection. In *Proc. 3DUI '07*, pages 79–86. IEEE.
- [2] F. Daiber, M. Speicher, S. Gehring, M. Löchtefeld, and A. Krüger. Interacting with 3d content on stereoscopic displays. In *Proc. PERDIS '14, article 32*. ACM.
- [3] A. Giesler, D. Valkov, and K. Hinrichs. Void shadows: Multi-touch interaction with stereoscopic objects on the tabletop. In *Proc. SUI '14*, pages 104–112. ACM.
- [4] M. Hachet, B. Bossavit, A. Cohé, and J.-B. de la Rivière. Toucheo: Multitouch and stereo combined in a seamless workspace. In *Proc. UIST '11*, pages 587–592. ACM.
- [5] M. Hancock, S. Carpendale, and A. Cockburn. Shallow-depth 3d interaction: Design and evaluation of one-, two-and three-touch techniques. In *Proc. CHI '07*, pages 1147–1156. ACM.
- [6] M. Hancock, T. Ten Cate, and S. Carpendale. Sticky tools: Full 6dof force-based interaction for multi-touch tables. In *Proc. ITS '09*, pages 133–140. ACM.
- [7] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *J. Vis.*, 8(3), 2008.
- [8] S. Knoedel and M. Hachet. Multi-touch rst in 2d and 3d spaces: Studying the impact of directness on user performance. In *Proc. 3DUI '11*, pages 75–78. IEEE.
- [9] H.-N. Liang, C. Williams, M. Semegen, W. Stuerzlinger, and P. Irani. An investigation of suitable interactions for 3d manipulation of distant objects through a mobile device. *Int. J. Innov. Comput. I.*, 9(12):4737–4752, 2013.
- [10] A. Martinet, G. Casiez, and L. Grisoni. The design and evaluation of 3d positioning techniques for multi-touch displays. In *Proc. 3DUI '10*, pages 115–118. IEEE.
- [11] A. Martinet, G. Casiez, and L. Grisoni. Integrality and separability of multitouch interaction techniques in 3d manipulation tasks. *IEEE TVCG*, 18(3):369–380, 2012.
- [12] T. Ohnishi, N. Katzakis, K. Kiyokawa, and H. Takemura. Virtual interaction surface: Decoupling of interaction and view dimensions for flexible indirect 3d interaction. In *Proc. 3DUI '12*, pages 113–116. IEEE.
- [13] P. Paczkowski, J. Dorsey, H. Rushmeier, and M. H. Kim. Paper3d: Bringing casual 3d modeling to a multi-touch interface. In *Proc. UIST '14*, pages 23–32. ACM.
- [14] J. L. Reisman, P. L. Davidson, and J. Y. Han. A screen-space formulation for 2d and 3d direct manipulation. In *Proceedings of the 22nd Symposium on User interface software and technology, 2009, UIST '09*, pages 69–78. ACM.
- [15] S. Strothoff, D. Valkov, and K. Hinrichs. Triangle cursor: Interactions with objects above the tabletop. In *Proc. ITS '11*, pages 111–119. ACM.
- [16] D. Valkov, F. Steinicke, G. Bruder, and K. Hinrichs. 2d touching of 3d stereoscopic objects. In *Proc. CHI '11*, pages 1353–1362. ACM.
- [17] L. Yu, P. Svetachov, P. Isenberg, M. H. Everts, and T. Isenberg. Fi3d: Direct-touch interaction for the exploration of 3d scientific visualization spaces. *IEEE TVCG*, 16(6):1613–1622, 2010.
- [18] S. Zhai and P. Milgram. Quantifying coordination in multiple dof movement and its application to evaluating 6 dof input devices. In *Proc. CHI '98*, pages 320–327. ACM.