Pointing in Spatial Augmented Reality from 2D Pointing Devices

Renaud Gervais, Jérémy Frey, Martin Hachet

Pointing in Spatial Augmented Reality from 2D Pointing Devices

Renaud Gervais[1,2], Jérémy Frey[1,2,3], and Martin Hachet[1,2]

[1]INRIA, F-33400 Talence, France
[2]Univ. Bordeaux, LaBRI, UMR 5800, F-33400 Talence, France
[3]CNRS, LaBRI, UMR 5800, F-33400 Talence, France
{renaud.gervais,jeremy.frey,martin.hachet}@inria.fr

Abstract. Spatial Augmented Reality (SAR) opens interesting perspectives for new generations of mixed reality applications. Compared to traditional human-computer interaction contexts, there is little work that studies user performance in SAR. In this paper, we present an experiment that compares pointing in SAR versus pointing in front of a screen, from standard pointing devices (mouse and graphics tablet). The results showed that the participants tend to interact in SAR in a way that is similar to the screen condition, without a big loss of performance.

Keywords. Spatial Augmented Reality, Pointing devices

1 Introduction

Spatial Augmented Reality (SAR) consists in projecting digital information directly onto physical objects. Beyond conventional display methods based on monitor screens or planar projections, this approach opens new perspectives in numerous fields including design, education and mediation. Since the pioneering work of Raskar et al. [7], inherent problems of computer vision and computer graphics are being solved today. On the other hand, the problems related to interaction remain largely unexplored. In this work, we have investigated the question of pointing in SAR.

Several strategies to point at augmented objects exist. One is to touch directly the area of interest. This approach is very straightforward and, consequently, it may be valuable in many contexts. However, direct touch suffers from many drawbacks. In particular, anatomical issues including the “fat finger” problem and the fatigue that is linked to mid-air interaction make direct touch little adapted as soon as accurate and prolonged actions are required (e.g. professional object design). In addition, direct touch is not possible when dealing with very fragile objects (e.g. relics in museums) or as soon as the objects are out of reach. For distant interaction, laser pointers or virtual rays can be good alternatives, but they still suffer from similar accuracy and fatigue issues. In our approach, we have explored the use of standard pointing devices (see Figure 1), namely mice and graphics tablets, to point at augmented objects. Years of human-computer interaction (HCI) have shown that these devices are decidedly well suited to point at visual objects displayed on 2D screens. Our assumption is that they can benefit to SAR as well, as soon as precision and prolonged work is required.

Fig. 1. A user moving a cursor (represented in blue) to a target (represented in red) on an augmented object by way of a standard mouse.

As an example, we can imagine an inspection scenario where an engineer points at an augmented circuit board with a mouse to highlight defects on small components. Another example is a design scenario where the artist draws by way of a tablet on a physical object, e.g. a 3D-printed one, to give it a specific appearance. For these two scenarios, it is interesting to note that the user equipped with a standard pointing device is still able to interact efficiently with standard GUI components displayed on a traditional screen, opening the way to true hybrid applications.

Pointing from mice and tablets has been extensively studied in traditional HCI contexts. In particular, Fitts' law [4] is able to predict the speed at which a user will be able to select a target depending on its distance and its size. Other works have been dedicated to pointing in 3D stereoscopic contexts [11,10]. The current work is the first one that studies the question of pointing in SAR, from standard pointing devices. In this work, we are interested in a setup where the user is sitting at a desk (desktop environment) and is interacting with objects located in front of him or her, as illustrated in Figure 1. Our contribution is the evaluation of the performance of pointing in a SAR environment using a standard pointing device compared to a traditional screen-based setup.

2 Related work

Since SAR has been introduced [7,8], there have been some research projects exploring interaction with projected content. Bandyopadhyay *et al.* [1] proposed the first interactive SAR prototype allowing users to "paint" physical objects with projected light using a six degrees of freedom tracked stylus. Physical-virtual tools [5] is a refinement of this concept, introducing more flexible editing tools inspired by real physical tools (e.g. an airbrush). Benko *et al.* [3] interacted with stereoscopic SAR using a mix of tangibles and gestures. These systems aimed for interaction modalities close to real-world metaphors. However, while perhaps more natural, they might prove to be less suited for precise and prolonged work than traditional 2D input devices.

The concept of pointing in SAR is similar to pointing in other contexts, namely multi-display environments (MDEs) and stereoscopic displays. In some ways, SAR can be compared to MDEs in that the physical world acts like a continuous space comprised of small display surfaces. As with MDEs, SAR might have some blind spots where the cursor will disappear because of a lack of projection support. Mouse Ether [2] and Perspective Cursor [6] are both systems that were developed to circumvent problems related to switching from one screen to another. The work of Xiao *et al.* [12] consists in projecting a cursor that can slide on any surface of the environment (which has been modeled in 3D beforehand). However, the system has been designed as a way to give feedback on the cursor's position when transitioning between screens and no targets were located in the environment itself in their evaluation. Pointing on a stereoscopic display has been studied by Teather and Stuerzlinger [11]. They studied different cursor types in what is effectively a "2.5D", or projected pointing task using a 3D Fitts' law pointing task. We also used projected pointing in this study. However, working on a real-world canvas is different from working on a screen since the real-world does not provide any reference frame for the 2D interaction. Moreover, a SAR installation does not suffer from the vergence-accommodation conflict present when using stereoscopic screens. Closer to a SAR setup, Reikimoto and Saitoh [9] proposed a spatially continuous workspace, allowing users to drag and drop content across different surfaces and objects. However, the pointing activity was not studied.

3 Pointing in SAR

Our SAR environment is comprised of a static scene laid out on a table in front of the user. A projector is then used to augment the objects.

On a standard screen configuration, the mouse cursor is generally represented as an arrow moving on the screen plane. When a 3D scene is displayed, the user is able to select any visible part of this virtual scene by picking the rendered result at the cursor location. Since most people are already experienced with this way of pointing, we wanted to know if this technique could be ported to a SAR environment albeit the lack of a physical screen support. Therefore, we used exactly the same metaphor in SAR, with the difference that the 3D scene is physically there, while the screen plane becomes virtual. The user moves the cursor on this virtual plane as he or she would do with a physical screen, as illustrated in Figure 2. A line representing the intersection between the virtual plane and the table is projected onto the table, and an arrow indicates the horizontal position of the cursor (see Figure 1). Contrary to standard screen configurations where the cursor is displayed on the screen plane, our SAR cursor is displayed directly on the physical objects. This cursor is represented as a cross within a 2D circle that is aligned with the underlying surface. Technically, we cast a ray formed by the

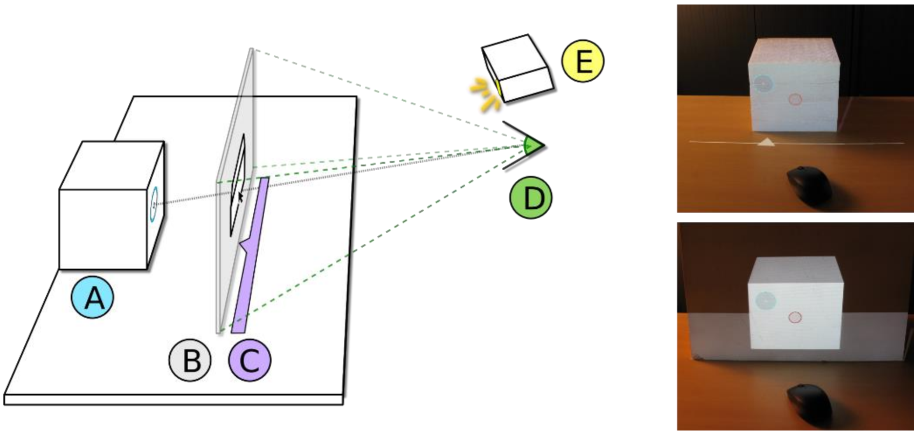

Fig. 2. LEFT: Drawing of the experimental setup. (A) Objects composing the scene to be augmented on which the cursor is displayed (light blue halo). (B) Plane on which the cursor is projected. This plane is either virtual in the SAR condition or physical (white wooden panel) in the Screen condition. (C) Feedback used in the SAR condition indicating the position of the virtual plane with the tip of the triangle indicating the horizontal position of the cursor. (D) The position at which the user is viewing the scene. (E) Projector. RIGHT: Scene in SAR (top) and SCREEN (bottom) conditions, with the same viewing angle.

eye and cursor position on the virtual plane towards the scene. We then position the cursor perpendicularly to the normal of the picked point. The visual feedback (line and arrows) helps to know where the cursor is as soon as the latter does not project onto an object.

4 User study

We conducted a user study to assess the performance of the pointing technique described in the previous section (SAR) in comparison to a screen-based baseline (SCREEN). Our research question was the following: What is the difference in performance of a pointing task realized on a screen compared to one realized with a SAR installation given that all other conditions are constant? Does pointing in SAR follows Fitts' law?

4.1 Participants

Sixteen participants took part in the study (12 males, 4 females, mean age 28.75, SD 4.71). All of them obtained a university degree. Six participants were left-handed (the mouse used during the experiment was adapted to both left- and right-handed users). All the participants were familiar with mice, whereas they had very little experience with tablets. None of them had previous experience with SAR systems.

4.2 Apparatus

The scene to be augmented was laid out on a table in front of the user. Each object of the scene was manually measured and modeled in 3D. A projector was located above and behind the user pointing at the scene. The projector was calibrated using OpenCV's camera calibration functions. We used a 3.6 GHz Core i7 PC with Windows 8 equipped with two GeForce GTX690 graphic boards. The videoprojector was a ViewSonic Pro9000 with a resolution of 1920×1080 pixels. The same setup was used for both SAR and SCREEN conditions. In SAR, the virtual scene was projected directly onto the physical objects whereas a white wooden surface located at the same position was used in the SCREEN condition. This ensured a similar frame rate (50 FPS), colorimetric configuration (color, brightness, contrast) and approximately same pixel size in both conditions. The focus of the videoprojector was set on the screen plane. On this plane, the resolution was effectively of 915×904 pixels.

In the SAR condition, the objects were augmented by reprojecting the virtual scene from the point of view of the projector. In the SCREEN condition, the viewpoint of the user on the scene was virtually reproduced and reprojected on the virtual counterpart of the physical screen. Then, this reprojection was rendered from the point-of-view of the projector, effectively making the viewed scene in both conditions identical (see Figure 2 (right)). We did not use real-time head tracking but the user head's position was measured manually and thus accounted for. The whole installation has been created using the creative coding framework vvvv.

For the input device, we used both a mouse (MOUSE) and a Wacom Cintiq 13HD tablet (TABLET). The screen of the tablet was not used for the experiment and therefore was displaying a black viewport. The button located on the pen was used for the selection action. The mouse was used in a relative mode while an absolute mapping was associated with the tablet. The acceleration transfer function of the mouse was disabled.

The 3D scene was composed of a 21×18×21 cm cube, as well as a more complex shape with comparable dimensions (see Figure 1). The scene onto which the targets to acquire were laid out varied by rotating the cube by an angle of 45º to provide more depth changes between trials. The participants sat at a distance of 1 m from the screen or physical objects, and the height of the chair was set in order for the participants' head to be located at the ideal observer position.

4.3 Procedure

We followed the procedure described in [4]. The participants had first to position the cursor in a home area represented by a red circle. After one second, this circle moved from red to green and a target appeared in the scene. The participants were instructed to select this target as quickly and accurately as possible. The start time was recorded when the cursor left the home area and stopped when the users clicked on the target. The targets were spread on a circle centered on the home area.

4.4 Design

We used a 2×2 within-subjects design. The independent variables were the output modality (SCREEN, SAR) and the input modality (MOUSE, TABLET). The dependent variables were the completion time, the inefficiency defined as $\frac{\mathrm{Path}_{\mathrm{actual}}-\mathrm{Path}_{\mathrm{optimal}}}{\mathrm{Path}_{\mathrm{optimal}}}$ [13] and the number of errors, defined as the number of selections outside the target area. For each condition, the participants had to acquire 40 targets, resulting in 160 target acquisitions by participant, and 2560 records in total. The order for the input and output were counter balanced following a latin square to avoid any learning effects.

5 Results and Discussion

Factors		Time (ms)	Inefficiency	Errors	Throughput (bits/sec)
Input	Mouse / Tablet	ns	0.16 / 0.22 (SD: 0.05 / 0.08) *	ns	ns
Output	Screen / SAR	846 / 959 (SD: 154 / 119) **	0.17 / 0.21 (SD: 0.07 / 0.07) ·	ns	5.75 / 3.84 **
Grand average		902 (SD: 404)	0.19 (SD: 0.30)	0.05 (SD: 0)	–

Table 1. Statistical results. Marks: ** for $p < .01$, * for $p < 0.05$; · for $p < 0.1$; ns: not significant; –: not applicable.

Because the homogeneity of variance couldn't be verified according to Levene's test ($p < 0.001$), we analyzed our data with non-parametric statistics, using multiple Wilcoxon signed-rank tests and false rate discovery correction. We retained trials which did not comprise errors to study time and inefficiency across our factors. Statistical results are reported in Table 1.

Time. There was no significant effect of the input device on completion time. However, output modality had a significant impact. Users were 11% faster in the SCREEN condition compared to the SAR condition. While having higher completion time, the drop in performance is relatively low, especially considering that the cursor reference frame was virtual.

Inefficiency. Inefficiency is a measure of "wasted" cursor movement by the user. Input modality had a significant effect, the tablet being more inefficient than the mouse. This difference can be explained by the lack of experience of almost all participants with such a tablet. Output did not have a clear significant effect on the inefficency of the movements of the users.

Error Rate. There was no significant effect of either input modality or output on the error rate. On average, the error rate was 5%.

Throughput. The target condition is reflected by the Index of Difficulty (ID), which indicates the overall pointing task difficulty. $\mathrm{ID} = \log_2(\frac{\mathrm{D}}{\mathrm{W}} + 1)$ [4]. D is the projected target distance in the virtual screen and W is the perceived target size. W varied according to the location and orientation of the target in the scene. ID was discretised from $[1.91; 4.92]$ to $[2; 5]$ by steps of 0.5. We averaged the completion time across ID and conditions (input $\times$ output). We modeled the movement time (MT) with a linear regression. We obtained an adjusted R^2 value of 0.8479 which shows that the completion time of pointing tasks in SAR using mice and tablets still follows the Fitts' law (see Figure 3), and consequently remains predictible. We also computed associated measures of performance, also known as "throughput", using the slope of the regression lines. $\mathrm{Throughput} = \frac{1}{\mathrm{b}}$ [13].

There was no significant effect of the input device on the throughput, whereas output device did have an effect. The screen condition was significantly more efficient than the SAR condition although, as it was the case for the completion time, the difference is relatively low.

Overall, the participants were slightly less efficient in the SAR condition than the SCREEN one. This difference could be explained by the years of experience of the participants with pointing in front of a screen whereas they were exposed to a SAR setup for the first time. Also, it is interesting that removing the physical reference frame (screen) of the cursor does not prevent users to interact in the same way they are used to, i.e. as if a physical screen was there. We can thus presume that with additional experience, participants may improve their performance with SAR. Another possible cause for the drop of performance is the presence of blind spots where the cursor disappear because of a lack of projection support (such zones were involved in about 1/4 of the trials). It could be interesting to compare the effect of these gaps in MDEs vs SAR to evaluate the impact of the frame of reference provided by the screen. Additionally, possible extensions of this work include studying the performance when moving the viewpoint of the user while using the Perspective Cursor [6] and evaluating if the performance drop observed in the SAR condition can be reproduced with other interaction techniques such as laser-pointer.

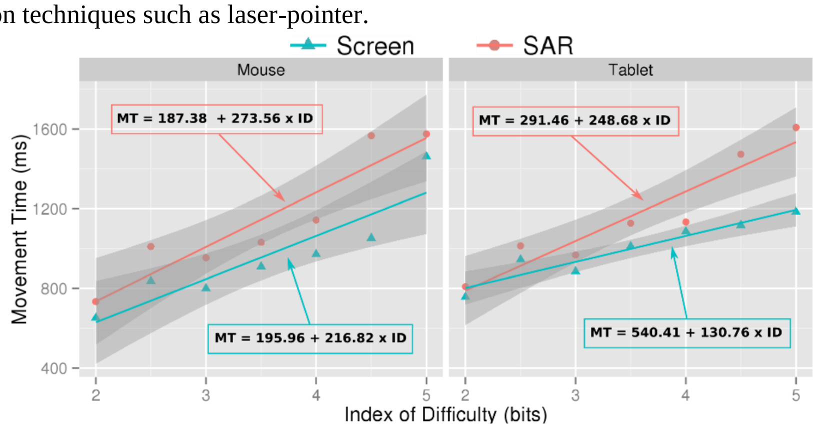

Fig. 3. Fitts' law models. $R^2 = 0.8479$.

6 Conclusion

We presented an approach for interacting with desktop SAR, i.e. when the user interacts with physical objects in front of him/her by way of standard pointing devices. A user study has shown that Fitts' law remains valid even if no physical screen is present. Users are able to point at targets displayed on the augmented objects in a manner that is comparable to what they used to do in front of a standard screen. This finding opens interesting perspectives, allowing desktop SAR applications to be used to extend the current desktop setup with augmented physical objects. Beyond pointing tasks, interaction in SAR is still a domain that has been little explored and, consequently, a large variety of HCI work is still to be conducted.

Acknowledgements
This work was supported by the ISAR project ANR-14-CE24-0013

References

1. Bandyopadhyay D, Raskar R, Fuchs H (2001) Dynamic shader lamps : painting on movable objects. ISAR '01 207–216. doi: 10.1109/ISAR.2001.970539
2. Baudisch P, Cutrell E, Hinckley K, Gruen R (2004) Mouse ether: Accelerating the acquisition of targets across multi-monitor displays. In: CHI'04 eA. pp 1379–1382
3. Benko H, Jota R, Wilson A (2012) Miragetable: Freehand interaction on a projected augmented reality tabletop. In: CHI '12. pp 199–208
4. MacKenzie IS (1992) Movement time prediction in human-computer interfaces. In: Readings human-computer interaction. pp 483–493
5. Marner MR, Thomas BH, Sandor C (2009) Physical-virtual tools for spatial augmented reality user interfaces. ISMAR '09 205–206. doi: 10.1109/ISMAR.2009.5336458
6. Nacenta MA, Sallam S, Champoux B, Subramanian S, Gutwin C (2006) Perspective cursor: Perspective-based interaction for multi-display environments. In: CHI '06. ACM, New York, NY, USA, pp 289–298
7. Raskar R, Welch G, Fuchs H (1998) Spatially augmented reality. In: IWAR '98. Citeseer, pp 11–20
8. Raskar R, Welch G, Low K-L, Bandyopadhyay D (2001) Shader lamps: Animating real objects with image-based illumination. In: EGWR '01. pp 89–102
9. Rekimoto J, Saitoh M (1999) Augmented surfaces: A spatially continuous work space for hybrid computing environments. In: CHI '99. pp 378–385
10. Schemali L, Eisemann E (2014) Design and evaluation of mouse cursors in a stereoscopic desktop environment. In: 3DUI '14. pp 67–70
11. Teather RJ, Stuerzlinger W (2013) Pointing at 3D target projections with one-eyed and stereo cursors. In: CHI '13. Paris, France, pp 159–168
12. Xiao R, Nacenta MA, Mandryk RL, Cockburn A, Gutwin C (2011) Ubiquitous cursor: A comparison of direct and indirect pointing feedback in multi-display environments. In: GI '11. pp 135–142
13. Zhai S (2004) Characterizing computer input with Fitts' law parameters – the information and non-information aspects of pointing. Int J Hum-Comput St 61:1–17.