

Overcoming Eye-Hand Visibility Mismatch in 3D Pointing Selection

Ferran Argelaguet*

Carlos Andujar†

Ramon Trueba‡

MOVING Research Group, Universitat Politècnica de Catalunya

Abstract

Most pointing techniques for 3D selection on virtual environments rely on a ray originating at the user’s hand whose direction is controlled by the hand orientation. In this paper we study the potential mismatch between visible objects (those which appear unoccluded from the user’s eye position) and selectable objects (those which appear unoccluded from the user’s hand position). We study the impact of such eye-hand visibility mismatch on selection performance, and propose a new technique for ray control which attempts to overcome this problem. We present an experiment to compare our ray control technique with classic raycasting in selection tasks with complex 3D scenes. Our user studies show promising results of our technique in terms of speed and accuracy.

CR Categories: I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction Techniques

Keywords: raycasting, virtual pointer, 3D selection

1 Introduction

The virtual pointer is arguably one of the most popular metaphors for object selection in virtual environments (VEs). A number of studies have demonstrated that virtual pointing techniques often result on better selection effectiveness than competing metaphors such as the *virtual hand*, allowing the selection of objects beyond the area of reach, and requiring less physical hand movement.

A number of virtual pointing variants have been reported. In this paper we focus on how the pointing direction is defined, i.e. how user’s movements are mapped into the pointing direction. The eye position and direction will be referred to as E and \vec{e} , resp. Likewise, the hand position and direction will be referred to as H and \vec{h} , resp.

Concerning the ray’s origin, pointing techniques can be classified into two groups: those where the ray originates at the user’s hand (including ray-casting [Mine 1995] and variants, see e.g. [Mine et al. 1997]) and those where the ray originates at the user’s eye (see e.g. [Pierce et al. 1997]). We will refer to these two families as *hand-rooted* and *eye-rooted* techniques, respectively.

Whenever a hand-rooted technique is used, the set of objects that are visible from the user’s eye might differ from those which appear unoccluded from the hand position (which often determines which objects are selectable). This issue has been largely ignored in the literature, which has concentrated on pointing facilitation techniques suited for fast and accurate selection of small, distant targets. Although some authors have studied the selection

of partially-occluded objects [Olwal and Feiner 2003; Peter et al. 2006; Vanacken et al. 2007], the occlusion problem has been studied considering solely either the eye position or the hand position but the interaction of both, and its effects in selection performance have not been studied in depth.

In this paper we study the impact of such eye-hand visibility mismatch on selection tasks. We focus on selection in densely-occluded scenes with multiple objects at varying distances to the viewer, since this kind of scene represents the most difficult scenario for selection tasks. We present the results of an empirical evaluation demonstrating the significant negative effect of eye-hand visibility mismatch in user performance. We also propose a new mapping for ray control which attempts to overcome the negative effects of the eye-hand visibility mismatch. In essence, our technique combines the benefits of image-plane techniques (absence of visibility mismatch and continuity of the ray movement in screen-space) with the benefits of ray control through hand rotation (requiring less physical hand movement). We present the result of a user study comparing our technique with default raycasting. We show that our new technique is at least as good as raycasting in all the test cases and even outperforms raycasting in many situations.

2 Previous work

Table 1 summarizes virtual pointing techniques, focusing on how the pointing direction is defined.

Technique	Origin	Direction	Volume
Raycasting [Mine 1995]	H	\vec{h}	ray
Flashlight [Liang and Green 1994]	H	\vec{h}	cone
Two-handed pointing [Mine et al. 1997]	H_n	$H - H_n$	ray
Occlusion selection [Pierce et al. 1997]	E	$H - E$	ray
Aperture [Forsberg et al. 1996]	E	$H - E$	cone
Our approach	E	\vec{h}	ray

Table 1: Summary of virtual pointing techniques. H_n is the position of the non-dominant hand.

Hand-rooted techniques Raycasting [Mine 1995] is the simplest virtual pointing technique used in VEs. The pointing direction is given isomorphically by a 6-DOF sensor attached to the user’s hand, and a ray extending out from the hand is drawn to provide visual feedback. The ray can intersect several objects, but only the first intersected object can be selected (two-handed techniques [Mine et al. 1997; Olwal and Feiner 2003] often circumvent this rule; the depth ray [Vanacken et al. 2007] is another exception). Usually, a button is pressed to signal that the user intends to select the intersected object. Liang and Green [1994] proposed an extension of the raycasting technique where the selection ray is replaced by a conic volume, facilitating the selection of small, isolated targets. The pointing direction can be defined also using a two-handed technique: one hand specifies the ray origin, while the other hand specifies where the ray is pointing to [Mine et al. 1997; Olwal and Feiner 2003]. All the techniques described above use the hand position H as the origin of the ray and thus can potentially suffer from the eye-hand visibility mismatch problem discussed in detail in Section 3.

*e-mail: fargelag@lsi.upc.edu

†e-mail: andujar@lsi.upc.edu

‡e-mail: rtrueba@lsi.upc.edu

Eye-rooted techniques Image-plane selection [Pierce et al. 1997] is accomplished by pointing directly at the screen projection of the object, rather than to the object itself. The selection ray is defined as $E + \lambda(H - E)$. The aperture technique [Forsberg et al. 1996] replaces the selection ray by a conic volume. Eye-rooted techniques proposed so far use the *hand position* H to control the ray direction, and thus require more physical effort from the user than techniques where the ray direction is controlled by the *hand direction* \vec{h} . This limitation is more apparent in immersive displays such as CAVEs, where selection of front objects requires the hand to be roughly aligned with the eye position.

3 Eye-hand visibility mismatch

In this section we study the eye-hand visibility mismatch occurring in hand-rooted selection techniques. We will refer only to the raycasting approach, although the discussion also applies to some extent to all raycasting variants. Two main issues can be identified: the first relates to the solid angle subtended by potential targets with respect to E and H , and it affects all kind of scenes including sparsely-occluded scenes. The second issue considers interobject occlusion, which mainly affects densely-occluded scenes.

3.1 Solid angle mismatch

A distinct feature of raycasting with respect to image-plane techniques is that objects are selected by pointing directly to themselves rather than to their screen projection. As a consequence, *the accuracy required to select objects with raycasting is not given directly by their screen projection*. This situation is depicted in Figure 1(a). Let $\Omega_E(S)$ and $\Omega_H(S)$ be the solid angle subtended by an object S w.r.t the user's eye and the user's hand, respectively. $\Omega_H(S)$ is a good measure of the effective width W of the object and thus a measure of how much accuracy is required to select S . We have analyzed how $\Omega_H(S)$ varies depending on the distance $d = \|S - E\|$ and the eye-hand separation $h = \|E - H\|$ (see Figure 1(b)). For the sake of simplicity we assume that H is vertically aligned with E , and that S is roughly a sphere with radius r horizontally aligned with E . Under the above assumptions, $\Omega_H(S)$ can be computed as $\Omega_H(S) = 2\pi(1 - \cos(\theta_H))$, where $\theta_H = 2 \arcsin(r / \sqrt{d^2 + h^2})$ is the planar angle of the cone defined by S and H .

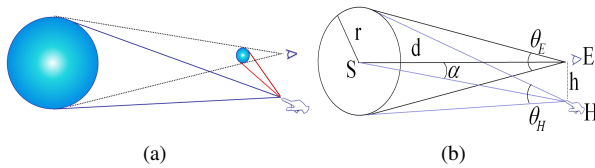


Figure 1: (a) Objects with the same screen projection subtend different solid angles with respect to the hand position; (b) Notation used for measuring the variation of the solid angle.

Figure 2 shows a plot of the ratio $\Omega_H(S)/\Omega_E(S)$ as the sphere moves away from the user while preserving its screen projection (i.e. as the sphere is scaled w.r.t. E). This means that $r = d \sin(\theta_E/2)$, where $\theta_E = \arccos(1 - \Omega_E(S)/(2\pi))$ is the planar angle of the cone defined by S and E . Obviously, $h = 0$ indicates that the hand is collocated with the eye and thus the ratio is 1. As h increases, nearby objects are more difficult to select than distant ones, *although having the same screen projection*. For $h = 80$ cm, which is a typical eye-hand distance when the user is standing up, an object standing at 70 cm from the user will subtend w.r.t the hand only a 50% of its solid angle w.r.t the eye (see Figure 2).

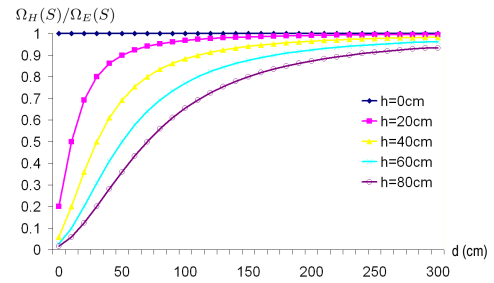


Figure 2: Evolution of the ratio $\Omega_H(S)/\Omega_E(S)$ when the distance to S increases from 10 cm to 3 m, while preserving its screen projection, for several eye-to-hand distances h .

Let us now discuss the potential effects of the above fact on selection performance, assuming $h > 0$. On the one hand, if we scale down a *single object* with respect to E (thus preserving its screen projection), we are clearly decreasing the effective W of the object and thus it will require more accuracy in the control of the pointing direction. However, if we scale down *the whole scene* while preserving its screen projection, we are dividing both W and D by a constant factor, which in essence means we are reducing the C-D ratio while preserving the index of difficulty of the task, $ID = \log_2(D/W + 1)$ [Balakrishnan 2004]. Thus scaling down the scene means that more accuracy will be required to fix the ray on top of the target, but also that the amplitude D of the required movements will be shorter, so at the end the effect of such a change on selection time is quite unpredictable. In any case, we see this as a flaw of raycasting in comparison with image-plane techniques, as the screen projection of a target does not convey appropriate feedback about the difficulty of its selection.

3.2 Occlusion mismatch

So far we have not considered the potential occlusion between targets. Whenever the selection ray originates at a point other than the user's eye, the set of visible objects (those which appear unoccluded from the user's eye) and selectable objects (those which appear unoccluded from the user's hand) might differ.

3.2.1 Occlusion mismatch at object level

Let \mathcal{V}_E be the set of objects visible from the eye position, and let \mathcal{V}_H be the set of objects visible from the hand position. Here we consider partial visibility, i.e. an object S is considered to be visible from a point P if at least one point in ∂S is visible from P . We focus on the subset $\mathcal{V}_E \cup \mathcal{V}_H$, which can be decomposed into three disjoint sets: $\mathcal{V}_E \cap \mathcal{V}_H$, $\mathcal{V}_H - \mathcal{V}_E$ and $\mathcal{V}_E - \mathcal{V}_H$. Objects in $\mathcal{V}_E \cap \mathcal{V}_H$ are both visible and selectable and for now we will suppose that they do not pose any problem.

Let us now discuss the behavior of ray-casting selection with objects in $\mathcal{V}_H - \mathcal{V}_E$. This corresponds to objects which are hidden to the user's eyes but are reachable from a ray emanating from the user's hand. Figure 3(a) illustrates this situation. Object B is occluded from the eye position but not from the hand position. Therefore it is currently selected even though it is not visible. This phenomenon is particularly distracting to the user. The user might think that the currently selected object is A , as there is an apparent intersection of the ray with object A because the screen projection P' of the last visible point of the ray is on the silhouette of the screen projection of A , as shown in Figure 3(a). In the absence of additional feedback, if the user triggers the selection confirmation at this point, the hidden object B would be erroneously selected.

On the other hand, objects in $\mathcal{V}_E - \mathcal{V}_H$ are visible but are not reachable from a ray emanating from the hand. Figure 3(b) illustrates this problem. Object A is visible from the eye position but it is completely obscured from the hand position. If the user starts moving the ray upwards increasing the elevation of the ray trying to bring P' to the screen projection of A , P' would jump from object B to object C . The discontinuous path followed by P' on the screen is shown in the bottom of Figure 3(b). In this situation the only way to select object A is to move the hand to a location from which A appears unoccluded. Again, this situation is a distinct feature of hand-rooted pointing techniques, and does not apply to image plane selection.

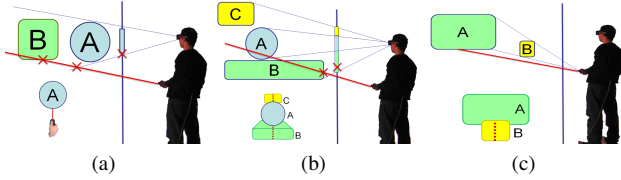


Figure 3: (a) Situation where the user can select an object which is hidden by another object. (b) A visible object cannot be selected because it cannot be reached by a ray emanating from the user’s hand. (c) Object A is visible from both E and H , but no point on its boundary is simultaneously visible from E and H .

3.2.2 Occlusion mismatch at point level

So far in the analysis of the visibility mismatch problem we have supposed that objects in $\mathcal{V}_E \cap \mathcal{V}_H$ do not pose any problem, regardless of the portion of the object that is visible from H and E . An approach that seems to be more accurate to predict potential selection problems is to consider an object S as a potentially difficult target whenever any of $\Omega_E(S)$, $\Omega_H(S)$ is below a certain value. We now show that this approach is still inaccurate, as an object S with large $\Omega_E(S)$ and $\Omega_H(S)$ can still be difficult to select. This situation is depicted in Figure 3(c). Now both $\Omega_E(A)$ and $\Omega_H(A)$ are large, but object A is still difficult to select because no point in its boundary is *simultaneously visible* from E and H . As a consequence, the user can intersect object A with the ray, but the intersection point will be hidden by object B , keeping the user from having visual feedback. Let S' be the set of points of S which are simultaneously visible from E and H . We claim that $\Omega_H(S')$ is an accurate measure of the effective W of the object in terms of selection difficulty.

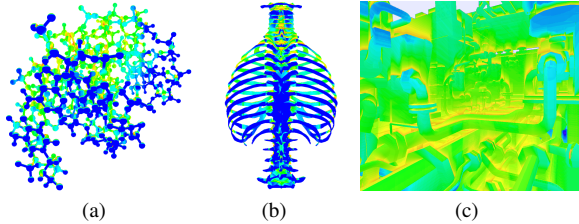


Figure 4: Simultaneous visibility on three test models. Dark blue parts correspond to points whose simultaneous visibility is preserved for all $\alpha = \widehat{EC}H$ in $[0, 60]$, C being the center of the scene. A warmer color indicates that even a small increase in α causes the point to become occluded from H .

The impact on performance of this visibility mismatch depends, among other factors, on the ratio $\Omega_H(S')/\Omega_H(S)$. Our experiments with different complex 3D scenes show that $\Omega_H(S')$ tends to

decrease as h increases, although not uniformly. Figure 4 shows the empirical evaluation of visibility mismatch for several test scenes. The color temperature represents the eye-to-hand distance h at which each point appears occluded from H .

3.3 Evaluation

The goal of this experiment is to evaluate the effects of the eye-to-hand distance h in selection time and error rates when using ray-casting selection. Given that the eye-hand visibility mismatch can substantially reduce the effective size of potential targets, one could expect a negative impact of the eye-to-hand distance h on selection performance. In practice, however, this may not be the case. On the one hand, the effective size reduction can be too small or affect a small set of objects to cause a significant difference. On the other hand, users can develop strategies to compensate the visibility mismatch, by moving quickly their hand to an unoccluded location.

Since we have no direct control over h , instead of increasing h we scaled down the whole scene with respect to the (initial) eye position. Such a scale preserves the screen projection of the objects, but divides both the effective size and the amplitude of the movements.

The experiment was conducted on a four-sided CAVE with stereo projectors at 1280x1280 resolution. The input device was a 6-DOF Wanda and a tracking system providing 60 updates/s with 4 ms latency. Eleven volunteers, aged from 23 to 32, participated in the experiment; 3 participants had no experience with VE applications; 4 had some experience and 4 were experienced users. The task was to select a sequence of objects, where the next object to be selected was clearly highlighted. Users were requested to complete the selection task as accurately as possible. We used three different models with varying density and occlusion levels: a ship model, a DNA molecule and a human thorax. For each model we used two different scale factors while roughly preserving the screen projection of each object. Table 2 shows the resulting angles prior and after scaling down the model, measured from the initial user’s head position, at 2.5m from the front screen.

Model	Avg	Min	Max	Avg	Min	Max
Thorax	10	7	15	27	21	32
DNA	10	7	15	24	17	35
Ship	5	2	6	11	5	20

Table 2: Resulting average angles \widehat{EC}_iH (degrees) on the original model and the scale down version. C_i is the center of object S_i .

A repeated-measures within-subjects design was used. The independent variable was the scale factor. The entire experiment consisted of 6 different trials, 2 for each model. Before the trials users were provided with a short training session. To avoid learning effects the order of the trials was randomized. The dependent measures were total selection time, error rate (number of erroneous clicks) and focus changes (the number of times the target object changed its selection status prior to confirming the selection).

The goal of this experiment was to evaluate the effects of increasing h (controlled indirectly by scaling down the scene) in selection performance. We chose the thorax model to check if the scale factor had a significant impact on selection time in a situation with little inter-occlusion (see Figure 4). We found no significant differences in selection time, error rate and focus changes. This result supports our hypothesis that, although the effective size of the objects is reduced, the required amplitude of the hand movements is also reduced, which in this case leads to similar selection performance. The DNA and ship models were selected as examples of densely occluded scenes subject to high visibility mismatch. We found a

significant effect of the scale factor on selection time ($p < 0.01$) and focus changes ($p < 0.05$ and $p < 0.01$) on both models, users performing much better with low α . For the DNA model, selection was a 12% faster and required a 12% less focus changes, on average. For the ship model, selection was a 17% faster and required a 23% less focus changes. As expected, these results seem to confirm our hypothesis that eye-hand visibility mismatch can be have a substantial negative impact on selection performance.

4 A new technique for ray control

We now propose a new mapping for ray control which attempts to overcome the negative effects of the eye-hand visibility mismatch. Our technique uses conceptually two rays: a *selection ray* and a *display ray*. The selection ray is defined by the parametric equation $E + \lambda \vec{h}$. To the best of our knowledge, this mapping between user movements and pointing direction has never been explicitly proposed nor evaluated (see Table 1). The first intersection of this ray with the scene objects, if any, is called Q , and indicates the selected object. If the ray does not intersect any object, we define Q as the farthest intersection of the selection ray with the scene's bounding sphere. Since the selection ray originates at the eye position, in projective space the ray is orthogonal to the viewing plane. Our solution to the visual feedback is to draw a ray (the display ray) defined by the segment HQ .

Given that our selection technique is not affected by the eye-hand visibility mismatch, in the best case scenario one could expect it to outperform raycasting, at least in complex scenes where visibility mismatch plays an important role. In practice, however, this may not be the case. The behavior of the ray might be counterintuitive for novice users, as the endpoint of the ray is controlled solely with the hand direction, being insensitive to hand translation.

We conducted an informal experiment to evaluate potential advantages of our technique in selection performance. We compared how users performed in the same environment as in Experiment 1, but using our technique for ray control instead of raycasting. In this case only one scale factor per model was chosen since our technique is not subject to eye-hand visibility mismatch. A between-groups design was used, the independent variable being the interaction technique (raycasting or our approach). The dependent measures were the same as in the first experiment.

For the thorax model, no significant differences were found on any variable. This result seems to confirm that, in scenes with low depth complexity, the visibility mismatch does not play an important role and our technique does not provide any advantage. However, for the ship model we found a significant effect of our technique on selection time and focus changes, compared to raycasting selection on the scaled-down model (high impact of visibility mismatch). Users performed much better with our technique (on average, 15% faster and 25% less focus changes). Again, this result confirms our hypothesis, as the effects of the visibility mismatch are more apparent in the scaled down replica (for the unscaled model, the average α value was just 11 degrees, see Table 2). For the DNA model we again found a significant effect of our technique, both with the scaled model and the unscaled one. Selection times were 23% and 32% faster with our technique, resp. The number of focus changes was also significantly smaller 21% and 31% smaller with our technique, resp. Considering the average values for α for the DNA model (see Table 2), this result supports our hypothesis that our technique outperforms raycasting when visibility mismatch plays an important role. Additional results can be found in <http://www.lsi.upc.edu/~virtual/VRST08>.

5 Conclusions and future work

We have studied the problem of eye-hand visibility mismatch affecting pointing selection techniques where the ray originates at the user's hand. We have demonstrated its negative impact on selection performance both analytically and empirically. We have shown that the significance of these effects largely depend on the scene complexity and on the eye-to-hand distance with respect to the scene size. We have proposed a new selection technique for ray control which combines the benefits of image-plane techniques (absence of visibility mismatch and continuity of the ray movement in screen-space) with the benefits of ray control through hand rotation (requiring less physical hand movement from the user). We have shown that our technique is at least as good as raycasting, and that it outperforms raycasting when interacting with complex scenes. Our technique for ray control is orthogonal to most pointing facilitation techniques. We plan to explore it in combination with automatic snapping [Steinicke et al. 2004], bubble cursors [Vanacken et al. 2007] and speed-based adjustment of the C-D ratio. It is also important to acknowledge that our choice of display device, focusing on spatially-immersive displays, has had a significantly impact in the evaluation. It would be useful to explore the effects of eye-hand visibility mismatch with head-mounted displays and volumetric displays.

This work has been partially funded by the Spanish Ministry of Science and Technology under grant TIN2007-67982-C02.

References

- BALAKRISHNAN, R. 2004. "beating" fitts' law: Virtual enhancements for pointing facilitation. *International Journal of Human-Computer Studies* 61, 6, 857–874.
- FORSBERG, A., HERNDON, K., AND ZELEZNIK, R. 1996. Aperture based selection for immersive virtual environments. In *User interface software and technology*, 95–96.
- LIANG, J., AND GREEN, M. 1994. Jdcad: A highly interactive 3d modeling system. *Computers & Graphics* 18, 4, 499–506.
- MINE, M., FREDERICK BROOKS, J., AND SEQUIN, C. 1997. Moving objects in space: exploiting proprioception in virtual-environment interaction. In *Proc. of SIGGRAPH'97*, 19–26.
- MINE, M. 1995. Virtual environments interaction techniques. Tech. Rep. TR95-018, Dept. of Computer Science, Univ. of North Carolina at Chapel Hill.
- OLWAL, A., AND FEINER, S. 2003. The Flexible Pointer: An Interaction Technique for Augmented and Virtual Reality. In *ACM UIST'03*, 81–82.
- PETER, H., ROLAND, W., AND BUES, B. M. 2006. Isith – intersection-based spatial interaction for two hands. *IEEE Symposium on 3D User Interfaces 2006*.
- PIERCE, J., FORSBERG, A., CONWAY, M., HONG, S., ZELEZNIK, R., AND MINE, M. 1997. Image plane interaction techniques in 3D immersive environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, 39–43.
- STEINICKE, F., ROPINSKI, T., AND HINRICHS, K. 2004. Object selection in virtual environments with an improved virtual pointer metaphor. *ICCVG'2004*.
- VANACKEN, L., GROSSMAN, T., AND CONINX, K. 2007. Exploring the effects of environment density and target visibility on object selection in 3d virtual environments. In *IEEE Symposium on 3D User Interfaces*, 115–122.