



**HAL**  
open science

# Efficient 3D Pointing Selection in Cluttered Virtual Environments

Ferran Argelaguet, Carlos Andujar

► **To cite this version:**

Ferran Argelaguet, Carlos Andujar. Efficient 3D Pointing Selection in Cluttered Virtual Environments. IEEE Computer Graphics and Applications, 2009, 29 (6), pp.34-43. 10.1109/MCG.2009.117. hal-02612868

**HAL Id: hal-02612868**

**<https://inria.hal.science/hal-02612868>**

Submitted on 19 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Efficient 3D pointing selection in cluttered virtual environments

Ferran Argelaguet\*

Carlos Andujar†

MOVING Research Group, Universitat Politècnica de Catalunya

## Abstract

Most pointing techniques for 3D selection on spatially-immersive 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 overcomes this problem. We present a controlled experiment to compare our technique with classic raycasting in selection tasks with complex 3D scenes. Our user studies demonstrate that our technique clearly outperforms raycasting selection, both in terms of speed and accuracy.

**Keywords:** raycasting selection, virtual pointer, 3D selection, 3D user interfaces

## 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 in better selection effectiveness than competing 3D interaction metaphors such as the *virtual hand*, allowing the selection of objects beyond the area of reach, and requiring less physical hand movement from the user. In this paper we focus on how the pointing direction is defined, i.e. how user's movements are mapped into the pointing direction. Concerning the ray's origin, pointing techniques can be roughly 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 (also referred to as image-plane techniques, see e.g. [Pierce et al. 1997]). We will refer to these two families as *hand-rooted* and *eye-rooted* techniques, respectively. Visual feedback on hand-rooted techniques is often provided by drawing a ray/cone extending out from the user's hand. Eye-rooted techniques must rely on a different strategy for providing feedback, since no matter how the ray direction is controlled, the projection of the ray onto the viewing plane is a single point.

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; 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 on selection performance have not been studied in depth.

In this article we study the impact of such eye-hand visibility mismatch on selection tasks performed with hand-rooted pointing techniques. We propose a new mapping for ray control, called RCE (Ray Casting from the Eye), which attempts to overcome the negative effects of the eye-hand visibility mismatch. In essence, the proposed mapping 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). This paper builds upon the results of a previous study on the impact of eye-to-hand separation on 3D pointing selection performance [Argelaguet et al. 2008]. In this article we provide empirical evidence that RCE pointing selection technique clearly outperforms classic raycasting selection, both in sparse and cluttered scenes.

## 2 Previous work

A review of all the strategies that have been proposed to facilitate 3D selection in VEs is out of the scope of this article (see [Bowman et al. 2004] for a comprehensive survey). Here we review only virtual pointing techniques, focusing on how the pointing direction is defined (Table 1 shows for a summary of techniques). We use the following notation: the eye position will be referred to as  $E$ . Likewise, the hand position and direction will be referred to as  $H$  and  $\vec{h}$ , resp.

**Hand-rooted techniques** Raycasting [Mine 1995] is the simplest virtual pointing technique used in VEs. In the default implementation, 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). 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 with constant apex angle, facilitating the selection of small, isolated targets. Again, the cone axis direction is given isomorphically by the hand orientation. 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.

**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 most simple approach, occlusion selection, uses a selection ray defined by joining the user's eye position with the user's hand position. The aperture technique [Forsberg et al. 1996] replaces the selection ray by a conic volume whose apex angle is controlled by the user. The user can interactively control the spread angle of the selection volume

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

†e-mail: andujar@lsi.upc.edu

Technique	Origin	Direction	Volume	Aperture $\tan(\theta)$
Raycasting [Mine 1995]	$H$	$\vec{h}$	ray	n/a
Flashlight [Liang and Green 1994]	$H$	$\vec{h}$	cone	$k$
Shadow cone selection [Steed and Parker 2004]	$H$	$\vec{h}$	cone	$k$
Two-handed pointing [Mine et al. 1997]	$H_n$	$H - H_n$	ray	n/a
Direct image plane pointing [Lee et al. 2003]	$H$	$\vec{n}$	ray	n/a
Occlusion selection [Pierce et al. 1997]	$E$	$H - E$	ray	n/a
Aperture [Forsberg et al. 1996]	$E$	$H - E$	cone	$k/  H - E  $
RayCasting from the Eye	$E$	$\vec{h}$	ray	n/a

**Table 1:** Summary of virtual pointing techniques.  $k$  indicates a constant value.  $H_n$  is the position of the non-dominant hand.  $\vec{n}$  is a vector perpendicular to the screen surface for controlling the selection volume. The origin column refers to the ray’s origin or the cone apex.

simply by bringing the hand closer or farther away from the eye. The eye-rooted techniques proposed so far use the *hand position* 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* [Wingrave et al. 2005]. This limitation is more apparent in immersive displays such as CAVEs, where the user is in a stand up position and selection of front objects requires the hand to be roughly aligned with the eye position.

**Pointing facilitation techniques** The application of Fitts’ law to HCI has led to a number of successful techniques to improve pointing performance [Balakrishnan 2004]. Fitts’ law asserts that the time  $T$  to acquire a target of effective width  $W$  which lies at a distance  $D$  is governed by the relationship  $T = a + b \log_2(D/W + 1)$ , where  $a$  and  $b$  are empirically-determined constants and the logarithmic term is called *index of difficulty*(ID).

### 3 Eye-hand visibility mismatch

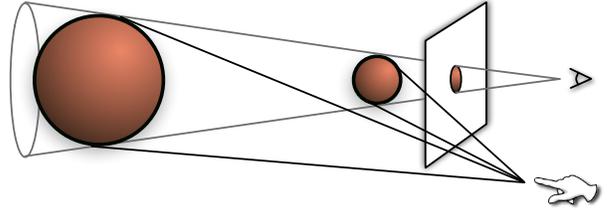
In this section we describe the eye-hand visibility mismatch occurring in hand-rooted selection techniques. We will refer only to the raycasting approach, although the discussion also applies to some extent to all raycasting variants. Two main issues can be identified: the first relates to the solid 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 inter-object occlusion (some objects can appear occluded from the hand but not from the eye and vice versa), 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. A first consequence is that, when using raycasting, objects with exactly the same screen projection might require very different ray orientations (Figure 1). Indeed, as we shall see, *the accuracy required to select objects with raycasting is not given directly by their screen projection*.

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.  $\Omega_E(S)$  is proportional to the screen projection of  $S$ , whereas, in absence of occluding objects,  $\Omega_H(S)$  is a good measure of the effective width ( $W$ ) of the object and thus a measure of how much accuracy is required to select  $S$ . Figure 1 shows two objects with the same screen projection but quite different solid angles. Again this is a distinct feature of raycasting w.r.t image-plane techniques. We analyzed in [Argelaguet et al. 2008] how  $\Omega_H(S)$  varies depending on the distance  $d = ||S - E||$  and the eye-hand separation  $h = ||E - H||$ . 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.



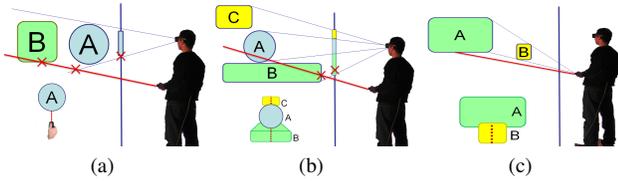
**Figure 1:** Objects with the same screen projection subtend different solid angles with respect to the hand position.

#### 3.2 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  $\partial S$  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 2(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. Object  $A$  might appear to be currently selected, 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 2(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 2(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 2(b). In this situation changing only the hand direction is not enough, and 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.

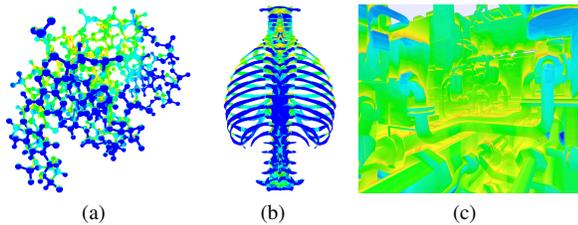


**Figure 2:** (a) Situation where the user can select an object which is hidden by another object. (b) A visible object cannot be selected because it cannot be reached by a ray emanating from the user's hand. (c) Object  $A$  is visible from both  $E$  and  $H$ , but no point on its boundary is simultaneously visible from  $E$  and  $H$ .

### 3.3 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 2(c). Now both  $\Omega_E(A)$  and  $\Omega_H(A)$  are large, but object  $A$  is still difficult to select because no point on 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.

Therefore, a more appropriate measure of the accuracy required to select an object  $S$  can be defined in terms of its simultaneous visibility. 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 3:** Simultaneous visibility on three test models. The color temperature represents the eye-to-hand distance  $h$  at which each point appears occluded from  $H$ .

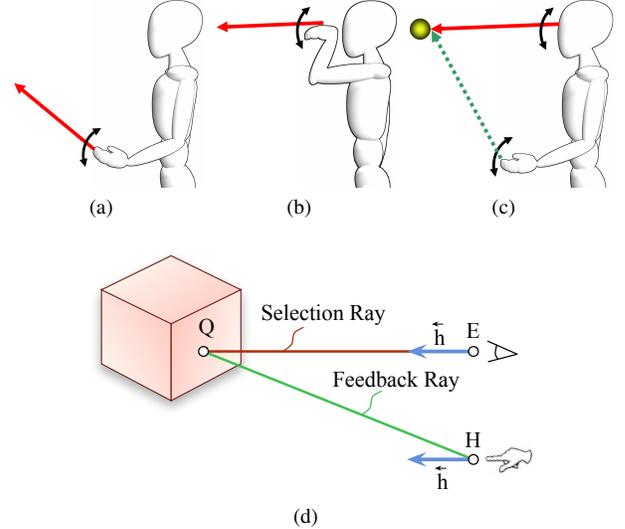
The analysis of the behavior when  $h$  increases requires distinguishing two causes for a point in  $\partial S$  not to be simultaneously visible: self-occlusion, due to faces of the same object, and inter-occlusion, due to other objects. If we increase  $h$ , self-occlusion obviously increases, as more front faces from  $E$  become back faces from  $H$  and vice versa. However, the behavior of the inter-occlusion is more difficult to predict, as it strongly depends on the object layout and the alignment of the  $E$  with the scene objects. Our experiments with different complex 3D scenes show that  $\Omega_H(S')$  tends to decrease as  $h$  increases, although not uniformly [Argelaguet et al. 2008].

Figures 3 show 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  $H$ . Note that visibility mismatch affects a large part of the scene:

around a 25% of the visible pixels correspond to parts difficult to select, for typical eye-hand distances.

## 4 Ray Casting from the Eye

We now propose a new mapping for ray control which attempts to overcome the negative effects of the eye-hand visibility mismatch. In essence, Ray Casting from the Eye (RCE) 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). In this sense, it can be viewed as a hybrid technique between raycasting and image-plane techniques.



**Figure 4:** In classic raycasting (a) the selection ray is cast from the user's hand, thus potentially suffering from eye-hand visibility mismatch. This problem persists unless users align their hand with the pointing direction (b), which results in an uncomfortable position. Our approach (c) uses a selection ray cast from the eye, whose direction is controlled with the hand orientation. Since the selection ray is insensitive to hand position, users can select objects as in (b) but with less physical effort. As the selection ray is not visible to the user, visual feedback is provided by drawing a ray cast from the hand to the first intersection of the selection ray with the scene (d).

Our technique uses conceptually two rays: a *selection ray* and a *feedback ray*. We first explain how we control the selection ray, and then justify the need for a separate 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 preserve the last valid depth.

Since the selection ray is cast from the eye, the ray projects onto a single point in the viewing plane. As a consequence, we must choose another strategy for providing appropriate visual feedback. The first option we considered was to draw a 3D cursor somewhere along the ray. Up to now, we have ignored the fact that displays in VEs are stereoscopic and therefore there is no *single* eye position. Fortunately, the discussion above also applies to stereoscopic displays by simply taking  $E$  as the middle point in the line joining both eyes. However, we can not ignore stereoscopic displays when

deciding at what distance  $\lambda$  along the ray the 3D cursor should be placed.

In a monoscopic display, the value chosen for  $\lambda$  is irrelevant, but in a stereoscopic display this value plays a key role, as the resulting cursor parallax is proportional to  $\lambda$ . A first option which suggests itself is to draw the cursor at  $Q$ , so that the user perceives the cursor as attached to the pointed object. Unfortunately, every time the ray intersects a new object, the cursor's parallax changes abruptly to match the new object's depth. As a consequence, the user's eyes are forced to converge at different distances at a very high rate. Since two objects far apart in 3D space can have arbitrarily close screen projections, the cursor parallax can potentially vary too fast, becoming a seriously distracting factor. This problem is particularly acute in plano-stereoscopic displays, due to the breakdown of the natural relationship between eye accommodation and convergence. Failure to properly converge at the right distance causes diplopia (double vision), which makes this feedback solution inconvenient for our purposes.

We also explored the opposite solution, consisting in using a constant value for  $\lambda$ . For example,  $\lambda$  can be computed as the distance of the user to the screen surface, so that the cursor will appear always at zero parallax condition. Since the cursor and the underlying scene object are displayed with different parallax values, simultaneous binocular fusion of them is impossible and either the cursor or the underlying object will be perceived in diplopia. Near the fovea, the maximum binocular disparity resulting in binocular fusion corresponds to a visual angle of about 10 minutes of arc, which means that the above problem also applies to small scenes with low depth variation. The reader can easily figure out the resulting effect by trying to point at a real object with the thumb finger, without closing her eyes. Our solution to the visual feedback is to draw a ray (the feedback ray) defined by equation  $H + \lambda(Q - H)$ . More precisely, we only draw the segment  $HQ$ . Obviously the parallax of the endpoint of the ray  $Q$  also changes rapidly but the replacement of a 3D cursor by a ray notably alleviates the problem (in this respect RCE behaves like any raycasting variant).

We now state some properties of the feedback ray, which are supported by the empirical evaluation discussed in next section.

- Since the feedback ray originates at the hand position and *responds* to hand orientation, *it feels* like a normal raycasting ray. In fact, both technique tends to be the same as the user aligns her hand with the pointing direction, as in Figure 4(b).
- Unlike classic raycasting, the *endpoint* of the ray is insensitive to the hand position, and depends only on the hand orientation. When the user is standing up in a spatially-immersive display, this allows for a more comfortable posture, requiring only hand rotation, not translation.
- Furthermore, the movement of the endpoint of the feedback ray is continuous ( $Q$  behaves like a 3D cursor) and the screen projection of an object is a good measure of its effective size.

## 5 User evaluation

In [Argelaguet et al. 2008] we conducted a user study to evaluate the impact of eye-hand visibility mismatch on selection performance. Our previous study focused on selection tasks using real world models including a variety of objects. Overall, eye-hand visibility mismatch was found to have a significant effect on selection performance, although we did not study the impact of such mismatch on a per-object, controlled fashion. Therefore, we did not identify which of the intervening factors (solid-angle mismatch, oc-

clusion mismatch, index of difficulty) played a key role in performance loss.

We now describe a new experiment which was designed with two goals in mind. First, we wanted to evaluate in a more controlled setup the impact of eye-hand visibility mismatch on selection tasks. Unlike [Argelaguet et al. 2008], we now control the index of difficulty and the occlusion level on a per-object basis. Eye-hand visibility mismatch is expected to increase the difficulty of selection tasks, but the resulting difference might be nonsignificant. For instance, users can develop strategies to compensate the visibility mismatch, e.g. by moving their hand quickly to a unoccluded location. Our second goal was to evaluate the performance of our technique (RCE) in comparison with raycasting (RC). Given that RCE is not affected by eye-hand visibility mismatch, one could expect it to outperform raycasting, at least in scenes where the eye-hand visibility mismatch plays an important role. However, in practice, this may not be the case. For example, the behavior of the ray might be counterintuitive, as the endpoint of the ray is controlled solely with the hand rotation, being insensitive to hand translation. Furthermore, the feedback ray might intersect occluding objects other than the target object, which might have a distracting effect. An informal study presented in [Argelaguet et al. 2008] suggested that RCE might reduce selection times in cluttered scenes, although we did not study for which objects and under which viewing circumstances objects were easier to select with our technique. The experiment we describe now provides empirical evidence that our technique clearly outperforms classic ray-casting both in cluttered and uncluttered scenes.

We also wanted to explore the correlation between selection time and index of difficulty,  $ID = \log_2(D/W + 1)$ . As the ray is controlled by hand rotation, the distance  $D$  to the target is estimated as the minimum rotation angle required for the selection ray to hit the target from the start position, and the target size  $W$  is estimated as the apex angle of the minimum cone with apex in the user's hand containing the target.

**Apparatus** 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. At the user position (90 cm from the EM emitter), position and orientation RMS errors were below 0.5mm and 0.06 degrees, resp. Users were provided with a wireless mouse to trigger the selection confirmation to avoid the Heisenberg effect and reduce the number of selection errors. The experiment was driven by a cluster of 2.66GHz QuadCore PCs with NVIDIA Quadro FX 5500 cards.

**Participants** Twelve volunteers, aged from 22 to 35, participated in the experiment; 6 participants had no experience with VE applications; 3 had some experience and 2 were experienced users.

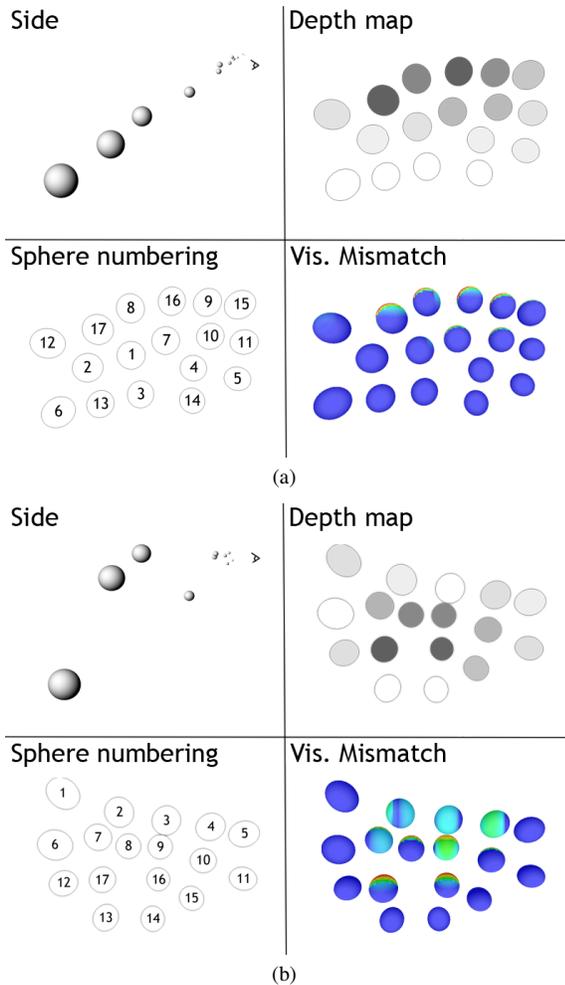
**Procedure** Users were requested to select a sequence of objects as quickly as possible. The next object to be selected was clearly highlighted. To provide additional visual feedback, the object intersected by the selection ray was highlighted in red.

We used two different test models, each one having seventeen spheres with varying size (from 0.2m to 60m) and distances (0.6m to 70m), placed in such a manner that all spheres had approximately<sup>1</sup> the same screen projection. In the first test model, spheres were layout so as to avoid inter-object occlusion from the eye and hand position, with nearby spheres appearing above farther ones

<sup>1</sup>Since users were head-tracked, actual screen projection varied according to user's head movements

from the user's perspective (see Figure 5(a)). We will refer to this model as the *unoccluded* model. In the second test model (see Figure 5(b)), spheres were layout so as to create varying levels of inter-object occlusion from the user's hand position but not from the user's eye position (all spheres were clearly visible so as to keep discovery time from distorting the results). The second model will be referred to as the *occluded* model.

To ensure that the difficulty of all trials remained constant among users, we pre-computed a random sequence of objects the user had to select. The same sequence was used in all the trials, each trial consisting in 300 selections.

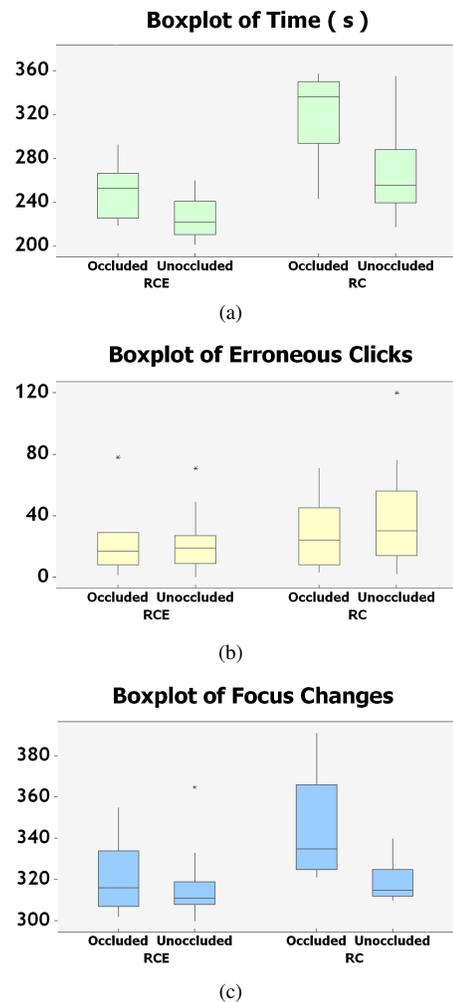


**Figure 5:** Test models used in the study (a) unoccluded, (b) occluded. Color temperature encodes the potential level of mismatch. Only the spheres from 1 to 16 were potential targets; sphere 17 was introduced solely to increase mismatch. While the mismatch in (a) was due to object self-occlusion in (b) it was also due to object inter-occlusion.

**Design** A repeated-measures within-subjects design was used. The independent variables were the test model (occluded, unoccluded) and the selection technique (RC or RCE). Each participant performed the experiment in one session lasting approximately 25 minutes. The experiment was divided into two blocks, one for each technique, and each block had two trials, one for each model. Before each block users were provided with a very short (1 min) train-

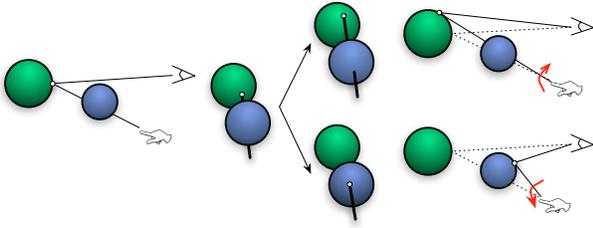
ing session. Surprisingly, some users did not realize they were using a different device-ray mapping for each block. The order of the blocks and the trials was randomized for each user to avoid learning effects. The dependent measures were total selection time, error rate and focus changes. The error rate was measured by counting the number of erroneous clicks. The focus changes is the number of times the target object changed its selection status prior to confirming its selection.

**Results** We first analyze the impact of eye-hand visibility mismatch on raycasting by comparing user performance with the occluded model and the unoccluded model (see right part of time boxplot in Figure 6(a)). Average raycasting selection time for the unoccluded model was 264 s, whereas for the occluded model average time was 319 s. The one-way ANOVA of selection time vs test model confirmed the significant effect of the test model on selection performance ( $p < 0.005$ ). Since the screen-projection of the spheres in both models was roughly the same, this result seems to confirm our hypothesis that, for raycasting selection, the eye-hand visibility mismatch has a significant impact on selection time.



**Figure 6:** Boxplots of time (s), erroneous clicks and focus changes vs technique and test model.

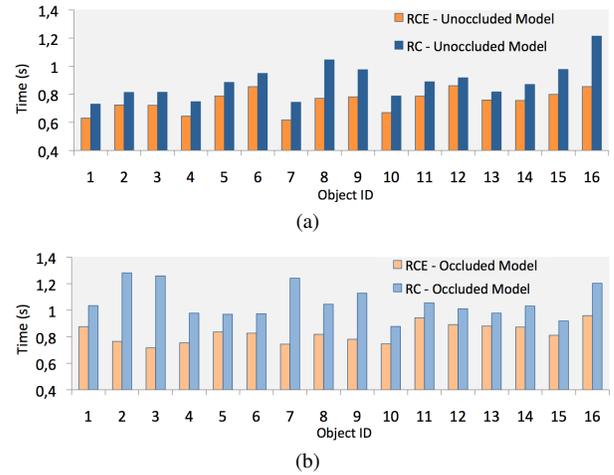
We now study the impact of eye-hand visibility mismatch on RCE. Since selection difficulty with our technique mainly depends on the



**Figure 7:** Worst-case scenario for RCE. Suppose the user wants to select the blue sphere starting from the situation shown in the left. If the user relies only in the feedback ray to aim at the target (like in raycasting), he will rotate his hand upwards, moving the intersection point away from the target (top right). If the user only relies on the intersection point, he will rotate his hand downwards being able to select the target (bottom right).

screen projection of the objects, we expected little or no impact on performance. Selection times are compared in the left part of Figure 6(a). Average selection times were slightly greater with the occluded model. The one-way ANOVA of selection time vs test model confirmed the significance of the difference ( $p < 0.02$ ). During the experiments, we notice that some novice users had difficulties in selecting some objects because of visual feedback conflicts. The worst case scenario for RCE seems to be the selection of a very close object whose screen projection is surrounded by distant objects. This situation is depicted in Figure 7. Suppose the user wants to select the blue sphere. Although both spheres have similar screen projections, the blue sphere is much smaller and closer to the viewer (the picture shows the two spheres with reduced depth disparity for clarity). Starting from the situation shown in the left, an untrained user might think that the action required to select the blue sphere involves intersecting it with the feedback ray. Since the ray is occluded by the blue sphere, the visual feedback indicates clearly that the feedback ray is behind it, so the user might rotate his hand upwards (Figure 7, top). This movement will cause the feedback ray to intersect the object, but this won't change the selection status for the green sphere (remember that the feedback ray is drawn solely for visual feedback, the pointed object being defined by the selection ray). Therefore the correct action to select the blue sphere is to rotate the hand downwards (Figure 7, bottom). This problem only appeared in the occluded test model, as the object layout of the first model avoided this situation. However, users were able to quickly overcome this behavior, as the average selection times only increased a 10%, in contrast with the 21% increment suffered by raycasting selection. Therefore we conclude that inter-object occlusion affects both techniques, being the impact much more noticeable with RC than with RCE. For RC the loss of performance appears to be caused by eye-hand visibility mismatch, whereas for RCE it appears to be caused by depth cue conflicts introduced by the visual feedback.

The second goal of the experiment was to compare the performance of RC versus RCE. We first discuss the selection time box plots in Figure 6(a). Our approach clearly outperforms raycasting with both test models. Users required, on average, 33% less time to complete the trials with the unoccluded model and 15% less with the occluded one. The two-way ANOVA of selection time vs test model and selection technique confirmed a significant effect for technique ( $p < 0.001$ ) (see Figure 6(a)). Indeed, a per-object study revealed that *all objects* were faster to select with our technique. In the case of the unoccluded test model (Figure 8(a)), we observed that selection improvements were particularly large for those objects with higher eye-hand visibility mismatch (objects 8, 9, and 16 in Fig-



**Figure 8:** Comparison of mean selection time per object in (a) Unoccluded model and (b) Occluded Model.

ure 5(a)) due to object self-occlusion. We also observed slight improvements for the rest of the objects. This can be explained by the fact that the amount of rotation required to select an object, and hence its index of difficulty (ID), depends on the selection technique (as discussed in Section 3.1). Indeed, we measured the average index of difficulty of the objects and found it to be slightly higher for raycasting selection (note that accurate control of the index of difficulty is very hard to achieve, as the amount of hand rotation required to select a given object depends on the final hand position after the previous object in the sequence has been selected). Performance gains with the occluded model also applied to all objects (Figure 8(b)). With these results, we can confirm that RCE clearly outperforms raycasting in presence of eye-hand visibility mismatch.

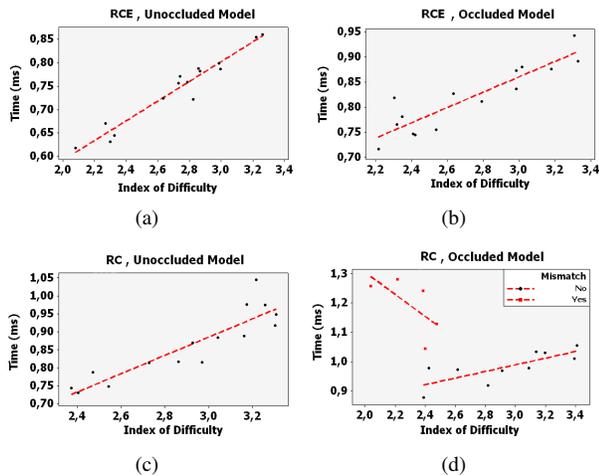
Besides selection time, we also compared the number of focus changes. The two-way ANOVA of focus changes vs selection technique and test model showed a significant effect on the focus changes between techniques ( $p < 0.05$ ), between the test model ( $p < 0.01$ ) and also in the interaction of both ( $p < 0.05$ ) (see Figure 6(c)). The number of focus changes is an indication of how many times the user attempted to point at an object prior to selecting it. Again, our technique outperformed raycasting, requiring a 7% less focus changes on average.

The two-way ANOVA of the number of errors vs technique and test model found no significant effect on technique ( $p = 0.129$ ), test model ( $p = 0.3$ ) and interaction ( $p = 0.5$ ) (see Figure 6(b)). However, we observed that the standard deviation of the number of errors was lower with RCE ( $\sigma = 12.1$ ) than for RC ( $\sigma = 22.6$ ).

Finally, we computed the Pearson correlation between selection time and index of difficulty on a per-object basis (see Figure 9). We observed high correlation values with RCE for both the unoccluded ( $r = 0.947$ ) and the occluded ( $r = 0.902$ ) models. This result provides further evidence of the small sensitivity of our technique to visibility mismatch problems. For RC, we found a correlation value of  $r = 0.898$  for the unoccluded model. However, no correlation was found for the occluded model: as depicted in Figure 9(d), a few objects (e.g. objects 2, 3, 7 and 9 in Figure 5(b)) did not adjust well to the expected positive correlation. We repeated the regression analysis ignoring these objects and then found a relatively high correlation ( $r = 0.755$ ). Interestingly, when we contrasted these objects with those exhibiting a high level of eye-hand

visibility mismatch, we found a perfect match (see Figure 5(b)).

We derive two main conclusions from the above regression analysis. The first one is that RCE adapts quite well to Fitts' law, being almost insensitive to eye-hand visibility mismatch and the screen projection of an object is a good measure of its effective width. The second conclusion is that, when using raycasting selection, the screen projection of an object is not a good measure of its effective width, as it does not account for eye-hand visibility mismatch.



**Figure 9:** Scatterplot of mean object selection time vs index of difficulty. Correlation values are: (a) 0.972, (b) 0.902, (c) 0.898, and (d) 0.755 (this last one has been computed ignoring spheres with large visibility mismatch).

## 6 Conclusions

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 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 provided empirical evidence that our proposed technique clearly outperforms raycasting in selection performance. The performance gain is more evident when interacting with cluttered virtual environments, but it also outperforms raycasting in non-cluttered virtual environments.

A key feature of RCE is that the user's hand does not point exactly to the target, i.e. vectors  $h$  and  $HQ$  are not parallel, as shown in Figure 4(c). It can be argued that the above fact might interfere with the user's proprioception, potentially making pointing tasks more difficult. However, we found that users had no difficulty in controlling the selection ray, even though no explicit explanation about the mapping was given during the experiment. Some users even did not realize that they used a different interaction technique in each block. Note that RCE is comparable to RC when users align their hand with the pointing direction, as in Figure 4(b), but users are free to place their hand in a more comfortable position, allowing e.g. to 'shoot from the hip'.

There are several directions that can be pursued to extend the current work. We plan to explore how this new device-ray mapping can be used in combination with manipulation tasks. We believe

that the best choice would be a hybrid approach such as HOMER. For example, users can use RCE for selection and then switch to a hand-centered manipulation technique. It may be also interesting to devise new visual feedback techniques for the proposed device-ray mapping, minimizing depth cue conflicts. For example, this can be accomplished by placing the origin of the feedback ray closer to the eye position, reducing the mismatch between the selection ray and the feedback ray.

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

## References

- ARGELAGUET, F., ANDUJAR, C., AND TRUEBA, R. 2008. Overcoming eye-hand visibility mismatch in 3d pointing selection. In *VRST '08: Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, ACM, New York, NY, USA, 43–46.
- BALAKRISHNAN, R. 2004. "beating" fitts' law: Virtual enhancements for pointing facilitation. *International Journal of Human-Computer Studies* 61, 6, 857–874.
- BOWMAN, D. A., KRUIFF, E., LAVIOLA, J. J., AND POUPYREV, I. 2004. *3D User Interfaces : Theory and Practice*. Addison Wesley. ISBN: 0-2017-5867-9.
- FORSBERG, A., HERNDON, K., AND ZELEZNIK, R. 1996. Aperture based selection for immersive virtual environments. In *User interface software and technology*, 95–96.
- LEE, S., SEO, J., KIM, G., AND PARK, C. 2003. Evaluation of pointing techniques for ray casting selection in virtual environments. In *3rd International SPIE Conference on Virtual Reality and Its Application in Industry*.
- 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.
- 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.
- STEED, A., AND PARKER, C. 2004. 3d selection strategies for head tracked and non-head tracked operation of spatially immersive displays. In *8th International Immersive Projection Technology Workshop*, 163–170.
- 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.
- WINGRAVE, C. A., TINTNER, R., WALKER, B. N., BOWMAN, D. A., AND HODGES, L. F. 2005. Exploring individual differences in raybased selection: Strategies and traits. *IEEE Virtual Reality 2005*.