Computers & Graphics 37 (2013) 121–136

Contents lists available at SciVerse ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on Touching the 3rd Dimension

A survey of 3D object selection techniques for virtual environments

Ferran Argelaguet [a,*], Carlos Andujar [b]

[a] VR4i Research Group, INRIA Rennes, France
[b] MOVING Research Group, Universitat Politècnica de Catalunya, Spain

ARTICLE INFO

Article history:
Received 31 August 2012
Received in revised form
7 December 2012
Accepted 11 December 2012
Available online 20 December 2012

Keywords:
3D interaction
3D selection
Virtual pointing
Virtual reality

ABSTRACT

Computer graphics applications controlled through natural gestures are gaining increasing popularity these days due to recent developments in low-cost tracking systems and gesture recognition technologies. Although interaction techniques through natural gestures have already demonstrated their benefits in manipulation, navigation and avatar-control tasks, effective selection with pointing gestures remains an open problem. In this paper we survey the state-of-the-art in 3D object selection techniques. We review important findings in human control models, analyze major factors influencing selection performance, and classify existing techniques according to a number of criteria. Unlike other components of the application's user interface, pointing techniques need a close coupling with the rendering pipeline, introducing new elements to be drawn, and potentially modifying the object layout and the way the scene is rendered. Conversely, selection performance is affected by rendering issues such as visual feedback, depth perception, and occlusion management. We thus review existing literature paying special attention to those aspects in the boundary between computer graphics and human–computer interaction.

1. Introduction

In the last decades we have witnessed enormous improvements in spatial input devices and motion tracking systems. These advances have motivated the development of a plethora of interaction techniques relying on six DoFs (Degrees of Freedom) input devices and user gestures. Interaction through natural gestures is gaining further popularity since the recent mass commercialization of low-cost solutions for full-body tracking, which is enabling the deployment of natural interfaces outside virtual reality labs. We will use the term 3D interaction to refer to interaction tasks requiring users to make some gestures in free (unconstrained) 3D space. These gestures typically involve one or both hands, and might also involve the user's head and other parts of the body.

The design of appropriate 3D interaction techniques for virtual environments (VEs) is a challenging problem [19,51]. On the positive side, interacting in free space with natural gestures opens a new world of possibilities for exploiting the richness and expressiveness of the interaction, allowing users to control simultaneously more DoFs and exploiting well-known real-world actions. On the negative side, 3D interaction is more physically demanding and might hinder user tasks by increasing the required dexterity. Compare for example the act of selecting an object using a mouse pointer to that of grasping a 3D object in free space. Mouse movement involves small, fast muscles whereas grasping often requires a complex arm movement involving larger and slower muscles [23,48]. Furthermore, current immersive VEs, even the most sophisticated ones, neither fail to provide the same level of cues for understanding the environment, nor reproduce faithfully the physical constraints of the real world [74]. For this reason, although humans are used to perform 3D interaction gestures in the real world, users of IVEs often encounter difficulties in understanding 3D spatial relationships and controlling multiple DoFs simultaneously.

Object selection is one of the fundamental tasks in 3D user interfaces [19] and the initial task for most common user interactions in a VE. Manipulation tasks often depend on (and are preceded by) selection tasks. As a consequence, poorly designed selection techniques often have a significant negative impact on the overall user performance. In this survey, we review major 3D interaction techniques intended for 3D object selection tasks. We do not consider indirect selection techniques, e.g. selecting from a menu or performing semantic queries. A 3D object selection technique requires the user to gesture in 3D space, e.g. grabbing an object or pointing to something (see Fig. 1). Two main 3D selection metaphors can be identified: virtual hand [78] and virtual pointing [63,54]. In the early days, virtual hand techniques were more popular as they map identically virtual tasks with real

* Corresponding author. Tel.: +33 299847487.
E-mail addresses: fernando.argelaguet_sanz@inria.fr (F. Argelaguet), andujar@lsi.upc.edu (C. Andujar).

0097-8493/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cag.2012.12.003

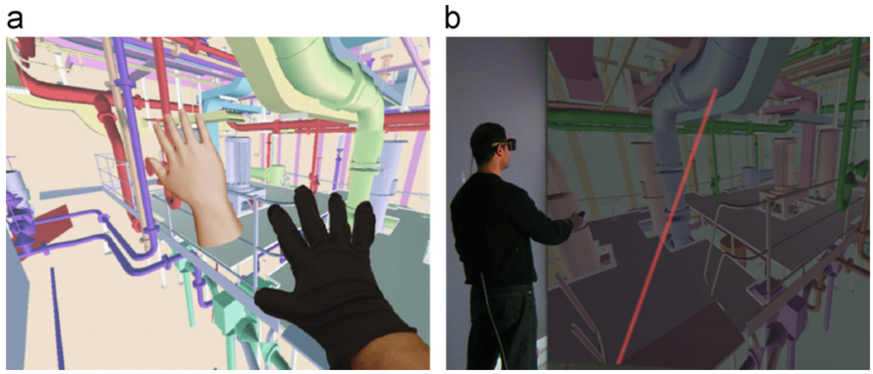

Fig. 1. Object selection using different metaphors and devices: (a) virtual hand and (b) virtual pointing through a hand-held spatial input device.

tasks, resulting in a more natural interaction. Lately, it has been shown that overcoming the physical constraints of the real world provides substantial benefits, e.g. letting the user select objects out of reach by enlarging the user's virtual arm [75], or using virtual pointing techniques such as raycasting [63]. In fact, raycasting selection is one of the most popular techniques for 3D object selection tasks [16]. A number of user studies in the literature have found that virtual pointing techniques often result in better selection effectiveness than competing 3D selection metaphors [19]. Unlike classical virtual hand techniques, virtual pointing techniques allow the user to select objects beyond their area of reach and require relatively less physical movement.

Selection through virtual pointing, though, is not free from difficulties. The selection of small or distant objects through virtual pointing remains to be a difficult task. Some techniques address the selection of small objects by increasing the size of the selection tool [36,73], at the expense of requiring disambiguation mechanisms to guess the object the user aims to select [30]. Noise from tracking devices and the fact that the interaction takes place in free space with no physical support for the hands [55] further hinders the accurate selection of small targets [43]. The user also has to keep the tool orientation steady until the selection confirmation is triggered, for example, by a button press. Such a confirmation action is likely to produce a change in the tool orientation, nicknamed Heisenberg effect [20], potentially causing a wrong selection. Occlusion is another major handicap for accomplishing spatial tasks [33]. Most interaction techniques for 3D selection and manipulation require the involved objects to be visible. A common solution for selecting occluded objects is to navigate to an appropriate location so that the targets become unoccluded. However, this navigate-to-select approach is impractical for selection-intensive applications. Therefore occlusion management techniques are often essential for helping users discover and access potential targets.

A number of approaches have been proposed to improve user performance in terms of task completion times and error counts [15]. A common strategy is to apply human control models such as the optimized initial impulse model [62] and Fitts' Law [34,35]. While the optimized initial impulse model refers to the accuracy a user can achieve given the movement required to perform an action, Fitts' Law estimates the time required to acquire a target. However, as users are bounded by human motor skills, there is a natural trade-off between speed and accuracy. In a typical scenario, high-accuracy rates will produce high task completion times and vice-versa.

In the context of the real usage of 3D interfaces, the subjective impressions of the users about an interaction technique can play a larger role than merely speed. The inability to select objects precisely may prove to be overly annoying and thus frustrate users. A performance increase might not be desirable if it is achieved at the expense of increasing the cognitive load of the task, or using techniques requiring extensive training.

The rest of this paper is organized as follows. Section 2 reviews existing human pointing models. In Section 3 we review major techniques for 3D object selection and extend previously proposed classifications [18,76,29] with a number of additional criteria to further elucidate the potential benefits and drawbacks of existing selection techniques. A comprehensive summary of the reviewed techniques is given in Table 1. Section 4 analyzes major factors influencing selection performance and proposes some usability guidelines. Finally, Section 5 provides some concluding remarks and future research directions.

2. Human pointing models

In order to point to (*acquire*) an object (*the target*), the user is required to perform a set of gestures (*movements*) to position the selection tool (e.g. his finger) over it. For each movement, the final position of the selection tool (*endpoint*) determines whether the acquisition is accomplished (the endpoint is inside the target) or not (the endpoint is outside the target). Once the target is acquired, the user has to trigger some selection mechanism to confirm the acquisition (e.g. pressing a button).

Pointing tasks involving physical interaction are constrained by the human psychomotor behavior. Several human pointing models have been proposed in order to model these aiming movements, to allow a better understanding of the processes involved and provide reliable prediction models of performance. From all the existing human motor models, Fitts' law provides by far the most successful and complete explanation. Fitts' law is one of the few quantitative measures in human–computer interaction and has motivated the development of guidelines for improving 2D and 3D pointing tasks. These guidelines are discussed in Section 4.

2.1. Fitts' law

Fitts' law [34], which emerged from experimental psychology, is a well known human psychomotor behavior model which has been widely adopted in numerous areas, including human factors, ergonomics and human–computer interaction. Fitts' law estimates the time required to perform an aimed movement considering only the physical properties underlying the acquisition task (the size of the target and the amplitude of the movement required to acquire it).

The most common formulation of Fitts' Law was proposed by MacKenzie [58] which asserts that the time T to acquire a target of effective width W which lies at a distance A is governed by the