Towards a General Model for Selection in Virtual Environments

Anthony Steed*

Department of Computer Science, University College London

ABSTRACT

Selection is one of the fundamental building blocks of all interactive virtual environment systems. Selection is the ability of the user to specify which object, or sub-part of an object in the environment, is the target for subsequent actions. Examples include selecting 3D buttons thus invoking an action or selecting a target upon which an action will occur. Selection is also an implicit or explicit part of manipulation techniques.

In a virtual environment selection can be performed in many different ways. In this paper we develop a generalized model of how interaction is and could be performed in virtual environments using 3D gestures. The purpose of this model is to highlight some potential areas for development and evaluation of novel selection techniques. The model is based on an analysis of the complexity of selection. We develop a model for selection that is based on time-varying scalar fields (TVSFs) that encompasses a very broad range of existing techniques. This model will be abstract, in that a direct implementation will be prohibitively complex, but we show how some standard implementation strategies are good approximations to the formal model.

CR Categories and Subject Descriptors: I.3.6 [Computer Graphics]: Methodology and Techniques - Interaction Techniques. **Additional Keywords**: selection, 3D interaction, virtual environments

1 INTRODUCTION

One of the fundamental building blocks of any interactive virtual environment (VE) system is the user interface for selecting objects. The ability to select an object underpins most interface tasks since a selected object is often the subject for future actions in the interface. This holds in many forms of user interface: the objects of interest are selected and then an action is performed upon those objects. One particular metaphor that exploits selection of targets is direct manipulation, where objects are manipulated as primary elements of the display [27]. This metaphor is taken to its logical extreme within VEs where objects can be manipulated by the user using natural body gestures.

Selection is thus a pre-cursor to subsequent action. Selection would not be necessary in some forms of VE that could perfectly track the human body and simulate force-feedback. Then we could perform manipulations of objects through natural gestures using a simulation of real-world physics. Thus to pick up an object we would not need to select the target, we would model the physical response of the grasp gestures, apply this as forces to the object and compute the resulting behavior of the object and the reaction of the hand. We have previously called this the “virtual reality model” of interaction [28]. Except in demonstrations of grasp gestures (e.g. [36]) this model is un-implementable today, thus even the direct manipulation models that are prevalent in common VE systems have to include selection as an implicit or explicit precursor to most interactions. Selection is the separate stage of identifying which object we want to manipulate. This may be implicit, in that the user may think that they are using their hand to pick up an object, but in the implementation there is a non-physical simulation that determined which object was the intended one, or it may be explicit in that the user points and objects are highlighted so that the user can see which object will be selected. We have called systems with such selection techniques “extended desktop” metaphors [28] to emphasize that many systems have generalized from the conventions of the 2D desktop metaphors. However a better label is perhaps “augmented workspace”, to distinguish the non-desktop nature of these selection techniques.

To generalize slightly: selection is a process of identifying a set of objects or parts of objects that are targets for subsequent action. In 2D selection is almost always done using a 2D cursor controlled by a mouse. In 3D VEs the most common approach is ray selection: a ray is cast from the user's hand in a single direction. This ray strikes a set of objects, from which the one with the closest intersection point is chosen. A completely different approach would use two hands with two conic volumes projecting into a space. Objects that lay in the intersection of the two volumes would be selected. We cover more examples in later sections.

This paper discusses the role of selection to try to understand and point out where opportunities lie for further development in selection techniques. This paper does not try to review all the existing work on selection, rather it uses existing taxonomies and descriptions of illustrative examples. The observation is that although there are very detailed taxonomies, these taxonomies overlook two main aspects of selection: the temporal nature of the user’s action to select an object and the complexities of the spatial task being undertaken. We use the term *selection gesture* to refer the act of the user making a selection: including the indication to start the select action, the subsequent movement of their hand and the termination of the select action. The paper starts by looking at the sometimes non-obvious complexity that underlies some of today's selection techniques. In sections 3 and 4 we take a longer look at the temporal and spatial properties of selection and highlight properties that we would like a generalized model to have. Section 5 considers other properties that selection techniques might have including user roles and multi-stage selection. In section 6 a model is then proposed that satisfies these properties, it uses an integration of time-varying scalar fields (TVSFs). This model, although powerful, is not readily implementable except in simple cases, but its purpose is to illuminate the potential space of potential selection techniques and some of the trade-offs and choices that must be made. We give an example interaction technique derived from this model that covers several difficult cases for selection techniques. In section 7 we discuss implementation strategies and some final observations about selection techniques. We then conclude.

---

*A.Steed@cs.ucl.ac.uk

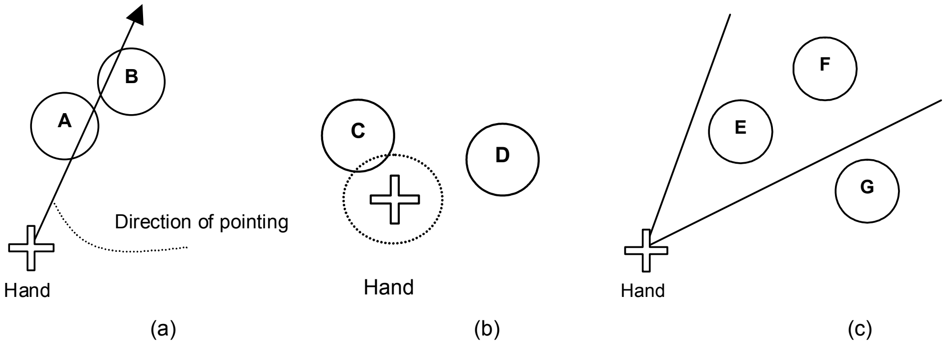


Figure 1. Simple selection techniques (a) Ray-based selection is effected by choosing first object to intersect the ray from the hand (Object A in this case) (b) Small Volume Selection is effected by choosing ever object's volumes intersect the volume attached to the hand (Object C). (c) Cone Selection can be effected by choosing all objects that lie within the cone (Objects E and F).

2 OUTLINE OF CURRENT SELECTION TECHNIQUES

Selection is a well-studied area of VEs because it is a fundamental primitive in interaction. Users need to be able to indicate objects in the environment either for immediate manipulation or to indicate the subject of pending or future actions. The action of selection is usually dependent on a 3D gesture, such as pointing towards the target with hand, head or eyes. Although we will discuss such selection techniques in the context of VEs, we note that prior to their use in VEs 3D gesture had been studied for work in smart-spaces [1] and it continues to be explored for use in ubiquitous computing environment and active surfaces [29].

As highlighted by a number of authors, one problem in designing virtual environments is that there are a wide variety of selection and other interaction techniques to choose from and the users prefer different techniques [5][6][31][38]. Section 2.1 gives an overview of some techniques described in the literature. The efficiency of any particular technique seems to be dependent first on the task and also on user preference [3], and thus some systems allow dynamic customization of interaction techniques (e.g. [30]).

In a VE system we typically have fairly precise 3D tracking of some of the user's limbs, our user has a one to one egocentric view of the environment consistent with the tracking and there is a 3D representation of the simulated environment. Note that none of these is necessarily true for non-immersive or mixed-reality spaces. Never the less, there is a surprising amount of complexity in the selection process. Practically every VE system from early lab prototypes and early commercial systems such VPL's RB2 has had a selection technique, and the variation in implementations is broad. Two techniques do stand out as being much more commonly implemented than most: ray-selection and virtual-hand collision.

2.1 Example Existing Technique

At an implementation level there are two main classes of selection technique: ray selection and volume selection. Ray selection involves casting a ray from one of the user's limbs into the world. It is very simple to implement but it is less immediately natural for the user because it is done at a distance. Selecting small objects at a long distance is obviously difficult due to limb stability, but when tracker jitter and error is included this difficulty is compounded. Typically the ray chosen will project from the hand, but it could be the head or eyes or combinations of these including bi-manual gestures [13][39] and the ray might be controlled indirectly using a scaling mechanism [12]. Other variations in include image plane interaction technique [22].

Volume selection techniques include the virtual hand technique and cone selection. Volume selection can be broken down into two classes: small volume and extended volume. Small volume interaction techniques use a small volume within the hand or surrounding the hand [19]. When this volume intersects other objects those objects are selected. Variations of this technique change the way in which the position of the small volume is changed to get over the problem of only being able to select within arms reach. The Go-Go interaction technique extends the virtual hand technique to support selection at a distance [24]. In [26] Poupyrev et al. note that Go-Go is a superset of the virtual hand and that whenever the virtual hand is used, Go-Go selection is a natural and flexible extension.

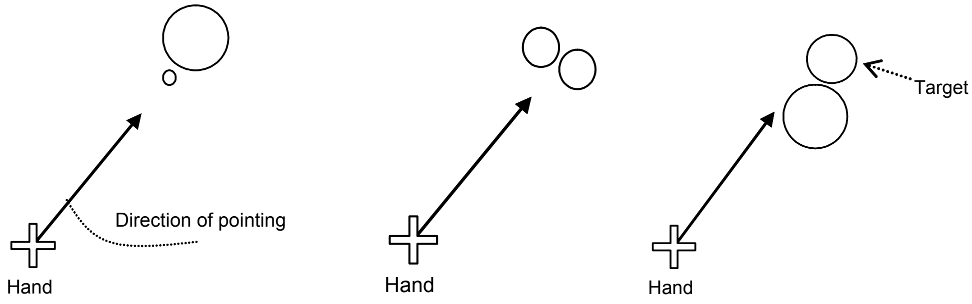
Extended volume selection techniques use a volume projecting in to the world. Objects lying inside the volume are selected. Cone or flashlight selection is a typical technique [18]. In some ways it is more similar to ray selection because it is the volume direction that needs to be controlled by the user rather than the position. This is considered by some to be preferable to ray selection because it is more tolerant of jitter and small errors. Variations include aperture based selection [11] and shadow cone selection [32].

We consider more variations in section 3 and 4 when we consider temporal and spatial aspects of selection. We will use the term *selection geometry* to cover all three types of selection shape.

2.2 Taxonomies

A taxonomy of selection techniques is presented in [4] and discussed in more depth in [8]. It comprises three sub-tasks: feedback, indication of object and indication to select. These have the following possible choices:

- Feedback
  - Graphical
  - Force/tactile
  - Audio
  - Text
- Indication of object
  - Touching
    - 1-1 movement
    - Position maps to position
    - Position maps to velocity
    - Velocity maps to position
    - Position maps to acceleration
    - Indirect control[*1]



(a) (b) (c)

Figure 2. Some difficult cases for selection techniques (a) Small object in front of a much larger object (b) Two objects which are very close together (c) Selecting an object that is partially occluded.

  - Occlusion
  - Pointing
    - 2D
    - 3D hand
    - 3D gaze[*2]
  - Indirect selection
    - List
    - Voice
    - Automatic
    - Iconic objects
- Indication to select
  - Gesture
  - Event
  - Voice
  - No explicit

This was an early taxonomy and there are additions that can be made as other authors have explored options. We make two initial notes (*1 and *2 in the list). Note 1 is that the potential mapping of sensed information (position, velocity, acceleration of the devices) can be mapped in many more ways than listed here. Note 2 is that any limb and even combinations of limbs could potentially be used to point.

The taxonomy is very useful for highlighting that there is a huge variety of possible combinations of feedback, indication of object and indication to select. Choices will depend on the input and output devices used and the task at hand. In the remainder of the paper we will assume that visual feedback may be available, but will say nothing about other types of feedback. However when feedback is discussed it may be possible these other modalities are useful complements or substitutes. The model we propose can support all types of indication to select excluding indirect selection. Indirect selection is somewhat a separate case that arises from specific application demands. For indication to select, we will describe metaphors that require button presses. Other events and voice are similar to a certain extent. No explicit indication usually means dwell time which we will discuss in section 2.4. Gestures to indicate selection can sometimes be incorporated but they fit badly with or constrain pointing indication.

2.3 Role of Selection

In the introduction we distinguished two broad categories of interaction techniques: those can be described as using a "virtual reality" model and those that use an "augmented workspace" model. The virtual realty model suggests that the user is using their own body to interact with the space. It is a pure direct manipulation model that relies on real world metaphors. If a user wants to pick up an object they must reach out and touch it with their hand. Actions and interaction will be based on real-world analogies, thus operations of tools are effected by picking up the tool and applying it by touching the tool to the user interface. This model has not been applied to many application areas though painting of objects is one exception [1]. Applicability is constrained by two main factors: the difficulty of tracking the user's body to a sufficient accuracy and the lack of good physical models of the effect of general tools on the world. For modeling of objects, it is possible to simulate the object as deformable solids and thus use tools in a manner similar as one would in the real world.

In an extended workspace metaphor, almost anything goes as long as it has utility for the user. The power of desktop user interfaces is that there is a fairly well-known and widely used set of conventions involving windows, icons, menus and pull-downs. Icons, menus and pull-downs are commonly transferred into VE toolboxes, though windows doesn't have an equivalent concept and the user typically deals with only one "workspace" at a time. Selection performs a role of indicating objects or parts of objects of interest. An object might be selected and then a menu or other control activated to manipulate intrinsic properties of that object.

It is important for some applications that selection support sub-parts or even individual points on objects. If points are required, then a ray-casting metaphor is almost certainly required. If parts are required, the scene model must either support this directly by supporting application developers identifying parts of an object at the modeling stage, or some form of surface distance or full object selection is required. In the next section we will describe volume techniques that can be used for selection of parts of objects. Alternatives might include multi-stage techniques where a whole object is selected and then a sub-part identified in a separate stage by painting the selection over of the object or using a tool to sub-divide the target objects.

Finally, some selection techniques implicitly or explicitly support groups of objects [15]. This is a critical design choice: single choice techniques are somewhat easier to use whereas obviously a group technique is more efficient when there are actions that can and should be applied to multiple targets. Group selection can be done with a ray-based "hold and release to select" approach by painting over the potential target set and selecting all touched. This can be efficient, but compounds two of the problems: the impact of jitter, error or imprecision when selecting objects and the problem of objects in front of non-target objects The alternative of repeated selects is unappealing. 3D rubber-band style volume description, using, for example an expanding sphere or dragging the corner of a cube can selected a large number of objects quickly. Alternatively using a large cone volume or dynamic cone volume [11] provides good speed when there are only target objects in the volume. Once group selection is available it is important to be able to de-select objects. Here we typically need to fall back on more modes in the interface, such as an extra mode or button for "inverse selection". The mechanisms

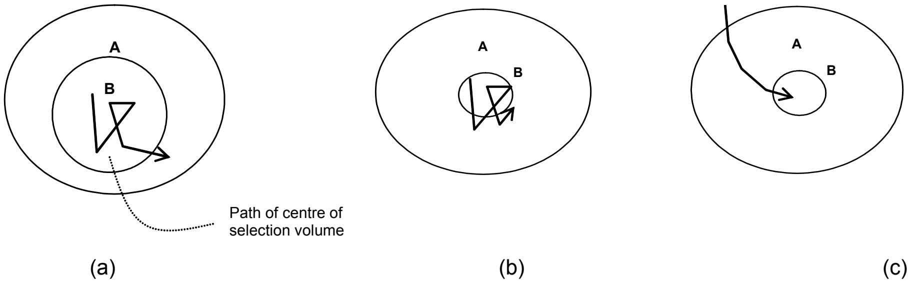


Figure 3. Example failures cases of gesture over time. The paths show the projection of the ray onto a 2D slice through the target (a) The path is stable over the target object until the last sample where it jumps off the target leading to a null or an incorrect selection depending on the selection technique (b) The target is small and the tracking imprecise so that the ray moves on and off the target. The selected object is almost a random choice, or null depending on the selection technique. However it is obvious to a human viewer what is intended (c) A case where the tracking is very precise, but the target is only reached on the last frame.

for these can be analogous to the group select and inverse select in common file explorers for windows systems. Note that that in 3D we have more degrees of freedom so a technique that explicitly works by deselection might be advantageous. Shadow cone is one such technique that we will return to in section 3 [32].

2.4 Properties of Selection

We have discussed existing techniques and in the following sections we will start to discuss more general models that include these techniques as subsets. However first we set up some general properties that we would like our selection techniques to have. These might seem somewhat obvious but existing techniques deal with one or more of these cases in a poor way.

- It should be possible to select objects that subtend a small angle towards the viewer
- Given equal subtended angles, nearby objects are probably more relevant to the user than far away objects.
- It should be possible to select a partially occluded object
- It should be possible to select a small object in front of a large object

Some example cases that are more difficult for selection techniques are shown in Figure 2. We do not claim that these occur particularly frequently, but when they do occur common selection techniques can fail forcing the user to make several attempts or adopt different strategies to make the selection. Figure 2a demonstrates that small objects are hard to select especially if they are in front of other larger objects. This is problematic for ray selection because imprecision will probably lead to mistaken selection of the larger object. With cone or large volume selection care has to be taken in the implementation so that the small object can be selected in preference to the larger object. This means that for example, it is not just the object with the largest intersection with the selection volume that is taken. This case is not a difficult problem with small volume selection as long as the user can position the volume over the small target without hitting the larger object. If they can't then the implementation needs to take care so that, again, not just absolute volume is used. In both these cases relative amount of volume covered is a useful metric; the model in section 6 extends this further.

Figure 2b demonstrates a related case where selection must be made between two objects which are very close together. This is similarly difficult because it involves precision, but note that this case is easier because with a steady selection gesture the intention can be made unambiguously. Imprecision or error might lead to a null selection but it is unlikely that an incorrect object would be selected. With volume techniques, there is a effective strategy of just making sure that the wrong object is not selected by making sure that the target is inside the volume and the other objects is well outside.

Figure 2c shows a very tricky situation of selecting an object that is partially occluded. This is different yet again because imprecision is likely to result in a null or the wrong object being selected. With ray selection this is just difficult and presents a similar problem to the one in Figure 2a, though failure is almost as likely to produce a null response as an incorrect response. With large volume selection, this is a much harder case because it will be very hard to select the target without hitting or accidentally selecting the larger object.

Aside from the feasibility of selection of objects we want to selection to be as easy as possible to support. We could for example, in any ambiguous situation such as those given above, rearrange the current target set of objects to make the selection more straightforward, a technique that might look analogous to the Mac OS X Dock system. However in many situations this may not be possible. Thus we will look at techniques that exploit the temporal and spatial properties of the full gesture of the user as well as the role of the user in the environment.

3 TEMPORAL ASPECTS OF SELECTION

The first set of observations about selection is that several problems surround the actual mechanism of indicating that the selection has been made. Three common techniques include click to select, hold and select on release and dwell on object. The difference between the click to select and hold and select on release appears to be minor, but there are two important differences. The latter affords the possibility of using a pointer feedback that is only visible when the user is making the select gesture whilst the first either uses a permanent pointer, a pointer that only appears when a target is available or relies on the user being accurate in the absence of feedback. Perhaps the more important affordance is that with hold and select on release we can exploit the full gesture of the user, something we discuss below. With the dwell time technique the user must point at it for the order of a second. The problem is identifying inadvertent gestures when, say, their hand is relaxed and pointing at the ground. Dwell time is thus only usable when there are no background objects or when there the developer has identified which objects are selectable. With dwell time there is also the problem of being able to point reliably at a small object for the necessary period of time. However it does remove the problem of errors on selection end; it is unlikely that in the presence of jitter or error that the wrong object would be selected because that object is unlikely to fulfill the dwell time constraint. This situation is illustrated in Figure 3a.

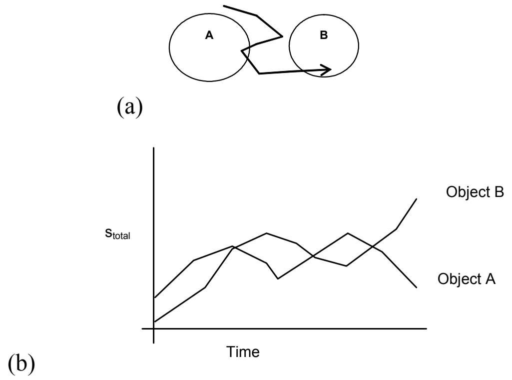


Figure 4. Coping with imprecision by integration over time. (a) Path of the selection volume centre over time. (b) s_{total} for both objects over time

Here the path of the pointer is stable until an error when the button is released.

A key is observation is that we can treat selection as a process of disambiguation. Consider using dwell time and hold and select on release together. In this case it would be straight forward to suggest a technique where the object that had been pointed at most would be selected. This has some immediate advantages: it is less likely to suffer from jitter or error at the end of gesture. Indeed it is extremely robust in the presence of extremely large errors or even tracker failures at the end of the gesture.

There are two minor drawbacks highlighted in the examples in Figure 2. First a small object just may be hard to select and thus it is not easy to dwell on it. Consider the 2D projection of the 3D pointer behavior shown by the path in Figure 3b. The projection roughly represents the view of the user. It is perhaps obvious what the intention of the user is. But the track actually intersects the target object for relatively little of the time. The second problem is that it is now more difficult to make very rapid selections. Consider the path in Figure 3c. It is obvious to use that the user is making a very precise selection but the target is only reached at the end. The initial part of the gesture is the user making their initial point and then relying on the feedback to correct and move to the target. Thus suggests that we need to consider the temporal profile of the gesture and some evaluation of the impact of error.

To a similar end, in [14] de Haan, Koutek & Post suggested using a temporal profile of an object's relevance. An object's relevance was a measure of distance to a ray. However in addition the relevance of objects was integrated over time. That is, an object had an instantaneous score ($s_{contrib}$) due to the configuration of the object, but discrimination between different objects were done on a value (s_{total}) computed as follows:

$$s_{total}(t) = s_{total}(t-1)c_s + s_{contrib}(t)c_g \qquad (1)$$

where c_s is < 1 and is chosen to provide a decay of any existing selection and c_g is a factor indicating rate of increase of an object's score if this is non-zero. Figure 4 shows a hypothetical example of the integration score of the two objects in the problematic situation of Figure 2b. Figure 4a shows two objects and the selection ray's projection. Figure 4b shows the integrations of scores. This technique on its own can't deal with the situations of Figure 2a and Figure 2c when the instantaneous score of the conflicting object could be as high as the target object.

Other observations are that equation 1 can be generalized to support our other requirements for selection techniques. Note first that a selection technique such as shadow cone [32], uses de-selection, that is objects are selected as soon as the button is pressed to activate the cone, but then objects are de-selected as soon as they drop out. This would suggest a variation of equation 1 such as:

$$s_{total}(t) = \begin{cases} s_{total}(t-1)c_s + s_{contrib}(t)c_g & if \quad c_g \geq c_{\min} \\ 0 \quad if \quad c_g > c_{\min} \end{cases} \qquad (2)$$

To preserve more of our requirements we might want to generalize and change c_s and c_g depending on, for example, amount of jitter in the system, the velocity of motion of the selection volume. In fact, the model we propose later covers cases where selection is additive, unique, or subtractive. In the first of these, selection decays slowly so that objects can be added easily to the selection and all objects whose s_{total} value rises above a certain value are selected at the end of the selection gesture. In the second case selection decays rapidly because only a single object need be selected, and thus it is assumed that it is pointed at towards the end of the gesture. The fact that the selection decays at all is to prevent the final frame error discussed in section 2.4. In the third case, selection decays slowly, and any objects whose instantaneous $s_{contrib}$ or s_{total} drops below a threshold are de-selected.

4 SPATIAL ASPECTS OF SELECTION

In most cases, selection is determined by the position and orientation of a tracked part of the user's body, be it the head-gaze, eye-gaze or hand-point. On a desktop system selection is strictly 2D in operation, and selection capability is well understood and modelled using tools such as Fitt's law [10][16]. Fitt's law has been applied to 3D situations, but the situation is much more complex and only limited cases have been studied (e.g. [13]). In 3D selection the input device is typically 6D but the task might have between 2 and 5 dimensions of freedom. It is easy to see that in situations where is no occlusion, selection of an object can be a 2D task in free space, because only two angles are required to specify a ray that joins the position of the hand and the target. Similarly, if the selection volume is compact, it is a 3D task to intersect the volume with the target object. However with both ray selection and extended volume selection, the user has the capability to both translate and rotate their hand over time. Figure 5a shows an example where the selection volume is translated and rotated to deal with the situation of occlusion. Given the discussion of temporal integration, Object B is likely to be the most relevant, and in addition when we have a subtractive selection gesture type, we can drop Object A immediately it falls out of the selection volume.

Another set of spatial issues is based on the observation that once the selection gestures has been started, the task changes in nature from an absolute pointing gesture to a relative one if there is good feedback. For example, the actual ray or volume geometry might be shown, or the objects currently selected or candidates for selection might be highlighted. Selection then becomes a relative task (moving the selection geometry towards the correct target) or a rejection task (move the selection geometry away from undesired objects). Note that the latter might be a technique more suitable for experts; where they can simply gesture in a rough direction that they know from experience will discriminate the selection they want. Figure 5b shows a potential such gesture in a 2D projection where the intersection of the centre of a volume in plotted.

A final note on spatial properties of selection gestures is that users will often "cancel" their select by pointing away at the last

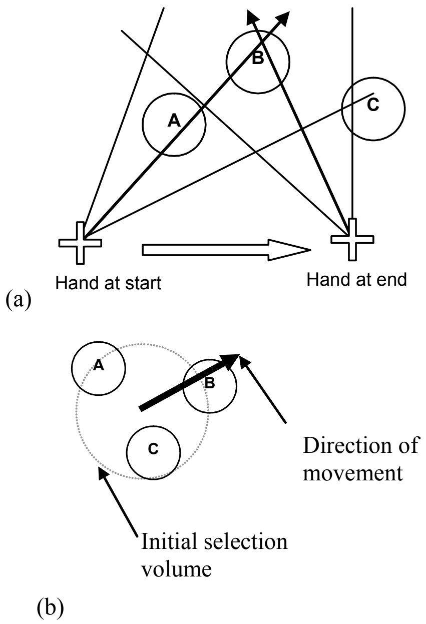


Figure 5. Moving the selection volume over time. (a) User moves the volume to disambiguate the selection. (b) A 2D viewing showing the centre and radius of a selection volume. The arrow indicates the direction the user then moves the volume, with the intention of selecting Object B.

minute. We should try to support this by noting very rapid spatial rotations away from recent dwell regions.

5 OTHER CONSIDERATIONS

Aside from the temporal and spatial characteristics of the gesture we note two other sets of considerations. Role of the user and relevance of objects to the task and multi-stage techniques that exploit partially completed parts of the gesture.

5.1 Role and Relevance

One of the first decisions any designer of an interactive VE simulation has to make is which objects are selectable. Typically there will be a background to the scene which isn't relevant to the task; it serves to orient the user and provide a visual reference to work against. Thus only certain objects will be selectable. As many point and click games players can no doubt testify it can be frustrating to identify which objects are interactive, but this is a necessary part of ensuring usability. To take this one logical step further, the set of selectable objects, or the level of relevance will depend on the role and task that the user is undertaking. Even collaborating users might require different selection capabilities. Furthermore, it may be that relevance or selectability changes dynamically over the time of the task. This is analogous to buttons becoming grayed out in user-interfaces.

Finally on relevance, we note that an object's relevance may depend on the spatial relationship between the object and the user. A button on a menu might not be selectable if it faced away from the user, but also, buttons might not be selectable from an oblique view simple because the designer of the button might enforce that there is a reasonable expectation the user will stand in front of a button panel in order to use it, and otherwise the selection might be inadvertent. In general the relevance might depend on position and orientation of the selection geometry.

5.2 Multiple-Stage Techniques

A number of interaction techniques have more than one stage: perhaps a volume of objects is selected and then the user must select from a list to disambiguate. More common is that if the user makes the same or a very similar selection gesture a second time, the system assumes that whatever arbitrary choice was made the first time was wrong and that a different selection must be presented next time.

A second use of multi-stage techniques is to use the first selected object as an example and make selecting similar objects more rapid. In [35] Stuerzlinger and Smith present a technique for rapid manipulation of groups of objects. In that technique if one object in a closely arranged set is selected, if that object is moved, it automatically "pushes" all the similar and proximate objects that lie in its direction of motion. It is not a strictly physical simulation – the objects do not need to be touching and a half-space of objects is moved. Although presented as a manipulation technique, this transfers easily to a selection technique where first one object is selected and then as the selection ray or volume is moved, close and similarly relevant objects are selected. Conditions need to be set so that the extent of the secondary selection is restricted, but this can perhaps be done by the extent or velocity of this secondary movement.

6 A MORE GENERAL MODEL OF SELECTION

The previous four sections have dealt with a number of properties and features that we would like selection techniques to have. In this section we outline a general abstract model of selection that encompasses all the techniques we have proposed, and then we give an example that has most of the properties and features we desire and might be claimed to be a superset of existing common techniques. We then discuss how these models support the properties and features that we desire.

6.1 Abstract Model

The general model consists of four algorithms: one for computing instantaneous relevance, one for computing aggregate relevance, one for highlighting objects based on aggregate and instantaneous relevance and one for choosing selection set from the final aggregate relevance scores. All four might be inter-dependent.

Instantaneous relevance (r_I) is computed by an integration of the selection geometry's *selection field* (f_S) with an object's *influence field* (f_I). We model these fields as scalar fields [20], thus r_I is a scalar value, f_S and f_I map R^3 to R. The integral is over R^3 of the product of the two fields. Whilst some of the relative position and orientation dependencies that we discussed in section 4 might be captured by representing these as vector or tensor fields, because we are allowing the scalar fields to vary with time, we think it unnecessary to go to higher dimensional fields. If there are multiple selection volumes from different limbs, then the selection field would be the union of several sub-fields. The shape of the selection field can be specified in any number of ways: from small volume or extended volume, through to an extrusion of an occlusion mask.

Aggregate relevance is computed based on a function of previous relevance and instantaneous relevance. We generalize from equations 1 and 2 by the following:

$$r_{total}(t) = g(r_I, r_{total}(t-1), r_{total}(t-2)+\ldots) \quad (3)$$

That is g is a function of the current relevance and the history of previous aggregate relevance.

The third algorithm is a highlighting algorithm that chooses which objects to highlight. This will be a choice based on r_I and the history of r_{total} for each candidate object.

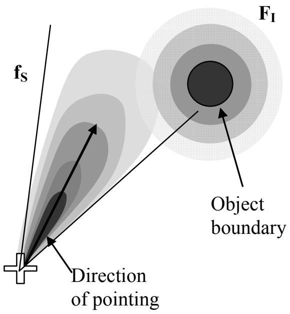


Figure 6. Enhanced cone selection. f_S is the selection field. f_I is the influence field

The final algorithm is a selection algorithm. Typically this will be similar to the highlight algorithm, but with the option of choosing only the most relevant object(s).

The four algorithms might change or have variables that change over time. For example, g in equation 3 might change depending on error and movement characteristics. f_S might change to represent a changing volume (e.g. aperture selection or dynamic zooming in on a target). f_I might change over time depending on task relevancy, selection volume movements and previous selections.

6.2 Enhanced Cone Selection

We believe that with suitable choices of specific algorithms, the abstract model covers the vast majority of current selection techniques. In order to show the generality, we propose a much more specific algorithm called enhanced cone selection. This extends previous cone and shadow cone selection [18] [32].

Firstly f_S is shaped like a cone where the scalar value depends on both angle from the centre ray and the distance from the ray origin. It is restricted to a 30^o angle around the direction of pointing. This still allows us to use the shadow-cone like property of easy discard of objects by moving objects outside the boundary. An object's influence is modeled as distance from the surface outside the object, and a constant inside the object. The constant value for inside the object is scaled by $1/\sqrt{\text{volume}}$ of object, thus small objects have a high constant, large ones a low constant. Figure 6 gives a simplified illustration where the scalar values have been banded. The aggregate relevance function takes in to account velocity, by making the scale of immediate relevance as 1/velocity. This would scaling need to account for jitter and error in the system. Our highlighting functions and final selection functions prioritize objects that are increasing in relevance and above a threshold. Final selection chooses just one object.

Does this technique satisfy all of our requirements? That is a question for a detailed experiment but we can make some observations with respect to Figure 2 and Figure 3. Firstly a technique that biases against distance and for size can help in situations such as Figure 2a. It can't remove problems, but it doesn't make it worse. Figure 2b becomes easy because it will be easy to make a relative change in direction to favor the target. Figure 2c still presents a challenge, but this might only be resolved by a dynamic motion of the hand such as suggested by Figure 5a. By taking an aggregate over time we deal well with the situations of Figure 3a and Figure 3b. By taking a preference for increasing objects and size, we can help with the situation in Figure 3c, especially since we can scaling by 1/velocity. This makes rapid selections much quicker.

7 DISCUSSION

The models in section 6 are difficult to implement directly because they require volumetric integration. This is a fairly straightforward numerical integration, but it would need to be done every frame and it would be necessary to take a very fine volumetric grid to approximate this and still respect the accuracy with which users can point given a sufficiently stable tracker. However we note that there are good approximations to distance functions, such as distance to a bounding sphere, bounding box or simply the object centre. Thus we might approximate the volume integration with the distance of selection geometry axis from the centre of object and the distance of selection geometry axis from the surface of the object. There is also no need to model any volume selection over the whole of R^3. The only candidate objects for which a current relevance needs be computed are those that intersect the selection geometry. This allows for a very rapid broad phase reject, where most objects can be discarded from further treatment.

Keeping a history of aggregate relevance for each object is onerous, but this can easily be truncated depending on the selection technique. For enhanced cone selection it was proposed to keep it so that objects whose relevance is increasing towards the end of the selection gesture are preferred. In practice this would probably not require more than 500ms of records in order to support the discriminatory gestures described in section 4.

There are many possible variants of selection that can be explored using the model we have described. What we have not discussed in much detail is the role of feedback. If group selection is required, then there seems to be little choice other than to highlight each object in the set. If only a single object is required then a useful technique is to indicate the target using a bent ray [14][17][21]. Of these techniques the one from Koutek et al. ([14][17]) has the benefit of indicating with the initial direction of the ray the principal direction of the selection volume.

Although we have generalized quite far, the model we propose does not encompass techniques with multiple representations of the target selection such as WIMs [34] or voodoo dolls [23].

As an aside it is our opinion that ray selection is only a specialization of cone selection and only using a 2D selection ray invites problems of the type we discussed in section 2.4. Thus we suggest avoiding it. Cone selection with a narrow focus (i.e. a small maximum angle and a relevance metric that depended on distance from the cone axis to a power of more than 1.0) and a single highlighted object provides a very similar experience that should be more reliable in almost all situations.

8 CONCLUSIONS

This paper has made an analysis of current selection techniques to build a general model of how selection can and could be done in virtual environments. We have listed several properties that we believe usable techniques should have, and several example situations that are tricky for one or more techniques. By looking at existing techniques and requirements we have proposed a general model of selection. This model captures several aspects of selection that are desirable. The model was greatly inspired by the IntenSelect model [14], but incorporates several novel capabilities.

In previous work we have focused on comparing selection techniques on characteristic tasks [31]. Others have proposed using test-beds [7][25]. We now believe that we should start to evaluate selection in much more complex situations. This is because simple selection tasks are at least feasible with most techniques but our experience of more complex environments is that selection techniques become more problematic and fail more frequently thus forcing the user to change their position to make the selection they want..

REFERENCES

[1] Baxter, W., Scheib, V., Lin, M. & Manocha, D. DAB: Interactive Haptic Painting with 3D Virtual Brushes. *Proceedings of ACM SIGGRAPH*, 2001, pp. 461-468.

[2] Bolt, R. A. Put-that-there: Voice and gesture at the graphics interface. *7th Annual Conference on Computer Graphics and interactive Techniques (SIGGRAPH)*, 1980, pp. 262-270.

[3] Bowman, D. Principles for the Design of Performance-oriented Interaction Techniques. *In Handbook of Virtual Environments*, Stanney, K. (ed.), Lawrence Erlbaum Associates, 2002, pp. 277-300

[4] Bowman, D. & Hodges, L. An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments. *Symp. on Interactive 3D Graphics*, 1997, pp. 35-38

[5] Bowman, D., Gabbard, J. & Hix, D A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. *Presence: Teleoperators and Virtual Environments*, 11(4), 2002, pp. 404-424.

[6] Bowman, D. & Hodges, L. Formalizing the Design, Evaluation, and Application of Interaction Techniques for Immersive Virtual Environments. *Journal of Visual Languages and Computing*, 1999, pp. 37-53.

[7] Bowman, D., Johnson, D. & Hodges, L. Testbed Evaluation of Virtual Environment Interaction Techniques. *Presence: Teleoperators and Virtual Environments*, 10(1), 2001, pp. 75-95.

[8] Bowman, D.A., Kruijff, E., LaViola, J. J., Poupyrev, I. *3D User Interfaces: Theory and Practice*, Pearson Education, 2004.

[9] Dang N.-T. The Selection-By-Volume Approach: Using Geometric Shape and 2D Menu System for 3D Object Selection. *Proc. of the Workshop on New Directions in 3D User Interfaces*, IEEE VR Conference, 2005.

[10] Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6), 1954, pp. 381-391.

[11] Forsberg, A., Herndon, K., Zeleznik, R. Aperture Based Selection for Immersive Virtual Environments. *Proc. UIST'1996*, pp. 95-96.

[12] Frees, S. & Kessler, G. D. Precise and Rapid Interaction through Scaled Manipulation in Immersive Virtual Environments. *IEEE Virtual Reality,* 2005, pp 99-106.

[13] Grossman, T. & Balakrishnan, R. 2004. Pointing at trivariate targets in 3D environments. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Vienna, Austria, April 24 - 29, 2004). CHI '04. ACM Press, New York, NY, 447-454.

[14] de Haan, G., Koutek, M. & Post, F. H IntenSelect: Using Dynamic Object Rating for Assisting 3D Object Selection. *IPT/EGVE 2005: 9th Int. Workshop on Immersive Projection Technology, 11th Eurographics Workshop on Virtual Enviroments*, (E. Kjems and R. Blach, eds.), 2005, pp. 201-209.

[15] Lucas, J. F. (2005) Design and Evaluation of 3D Multiple Object Selection Techniques, MSc Thesis, Virginia Polytechnic Institute and State University.

[16] MacKenzie I. S. & Buxton W. A. S. Extending Fitts' law to two-dimensional tasks. *Proceedings of ACM CHI 1992 Conference on Human Factors in Computing Systems*, pp. 219--226.

[17] Koutek M. & Post F. Spring-Based Manipulation Tools for Virtual Environments. *Proc. Immersive Projection Technology and Eurographics Virtual Environments '01*, 2001, pp. 61–70.

[18] Liang, J., & Green, M. JDCAD: A highly interactive 3D modeling system. *Computers & Graphics*, 18(4), 1994, 499-506.

[19] Mine, M., Brooks Jr., F. P. & Sequin, C. Moving Objects in Space: Exploiting Proprioception in Virtual-Environment Interaction. *Proc. of SIGGRAPH 97*.

[20] Morse, P. M. & Feshbach, H. Scalar Fields. §1.1 *in Methods of Theoretical Physics, Part I*. New York: McGraw-Hill, pp. 4-8, 1953.

[21] Olwal, A. & Feiner S. The Flexible Pointer: An Interaction Technique for Augmented and Virtual Reality. *Conference Supplement of UIST '03 (ACM Symposium on User Interface Software and Technology),* 2003, pp. 81-82,.

[22] Pierce, J., Forsberg, A., Conway, M., Hong, S., Zeleznik, R., and Mine, M. Image Plane Interaction Techniques in 3D Immersive Environments. *Proc. 1997 Symp. Interactive 3D Graphics*, pp 39-44.

[23] Pierce, J. S., Stearns, B. C. & Pausch, R. Voodoo Dolls: Seamless Interaction at Multiple Scales in Virtual Environments. *Proceedings of the 1999 Symposium on Interactive 3D Graphics*, pp. 141-145.

[24] Poupyrev, I., Billinghurst, M., Weghorst, S. & Ichikawa, T. Go-Go interaction technique: Non-linear mapping for direct manipulation in VR. *Proc. of UIST'96*. ACM. pp. 79-80.

[25] Poupyrev, I., Weghorst, S., Billinghurst, M. & Ichikawa, T. A Framework and Testbed for Studying Manipulation Techniques for Immersive VR. *Proc. ACM VRST'97*, pp. 21-28

[26] Poupyrev, I., Weghorst, S., Billinghurst, M. & Ichikawa, T. Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. *Computer Graphics Forum 17(3)*, 1998, pp. 41-52.

[27] Shneiderman, B. Direct manipulation: a step beyond programming languages, *IEEE Computer* 16(8), pp. 57-69.

[28] Slater, M., Steed, A. & Chrysanthou, Y. *Computer Graphics and Virtual Environments: From Realism to Real-Time*, Addison-Wesley, 2001.

[29] Starner, T., Leibe, B., Minnen, D., Westeyn, T., Hurst, A. & Weeks, J. Computer Vision-Based Gesture Tracking, Object Tracking, and 3D Reconstruction for Augmented Desks, *Machine Vision and Applications*, 14(1), pp. 59-71, Springer, 2003.

[30] Steed, A., Slater, M. A dataflow representation for defining interaction within immersive virtual environments, *IEEE Virtual Reality Annual International Symposium,* pp. 163-167, 1996.

[31] Steed, A. & Parker, C. Evaluating 3D Selection Strategies for Head Tracked and Non-Head Tracked Operation of Spatially Immersive Displays. *8th International Immersive Projection Technology Workshop Immersive Projection Technology Workshop,* May 13-14 2004, Ames, IA.

[32] Steed, A. & Parker, C. Evaluating Effectiveness of Interaction Techniques across Immersive Virtual Environment Systems, *Presence: Teleoperators and Virtual Environments*, 14(5), pp. 511-527, 2005.

[33] Steinicke, F., Ropinski, T. & Hinrichs, K. VR and Laser-Based Interaction in Virtual Environments Using a Dual-Purpose Interaction Metaphor. *IEEE VR 2005 Workshop Proceedings on New Directions in 3D User Interfaces*, pp. 61-64, 2005.

[34] Stoakley, R., Conway, M., & Pausch, R. Virtual Reality on a WIM: Interactive Worlds in Miniature. *Proc. SIGCHI '95*, pp. 265-272.

[35] Stuerzlinger, W. & Smith, G. Efficient Manipulation of Object Groups in Virtual Environments. *Proceedings IEEE Virtual Reality 2002*, 2002, pp. 251-258.

[36] Wan, H., Luo, Y., Gao, S., & Peng, Q. 2004. Realistic virtual hand modeling with applications for virtual grasping. *ACM SIGGRAPHInternational Conference on Virtual Reality Continuum and Its Applications in industry,* 2004, pp. 81-87.

[37] Wingrave, C. and Bowman, D. Baseline Factors for Raycasting Selection. *Proceedings of Virtual Reality Internationa*l, 2005.

[38] Wingrave, C., Tintner, R., Walker, B., Bowman, D. & Hodges, L. Exploring Individual Differences in Raybased Selection: Strategies and Traits. *IEEE Virtual Reality*, 2005, pp. 163-170.

[39] Zeleznik, R. C., Forsberg, A. S. & Schulze, J. P. Look-That-There: Exploiting Gaze in Virtual Reality Interactions, Technical Report CS-05-04, Brown University, Department of Computer Science, March 2005.