Outline Pursuits: Gaze-assisted Selection of Occluded Objects in Virtual Reality

Ludwig Sidenmark[1,*]**, Christopher Clarke**[1,*]**, Xuesong Zhang**[2]**, Jenny Phu**[3]**, Hans Gellersen**[4]

[1]Lancaster University, United Kingdom, {l.sidenmark, chris.clarke}@lancaster.ac.uk
[2]KU Leuven, Belgium, xuesong.zhang@kuleuven.be
[3]LMU Munich, Germany, jennyphu@live.com
[4]Aarhus University, Denmark, hwg@cs.au.dk
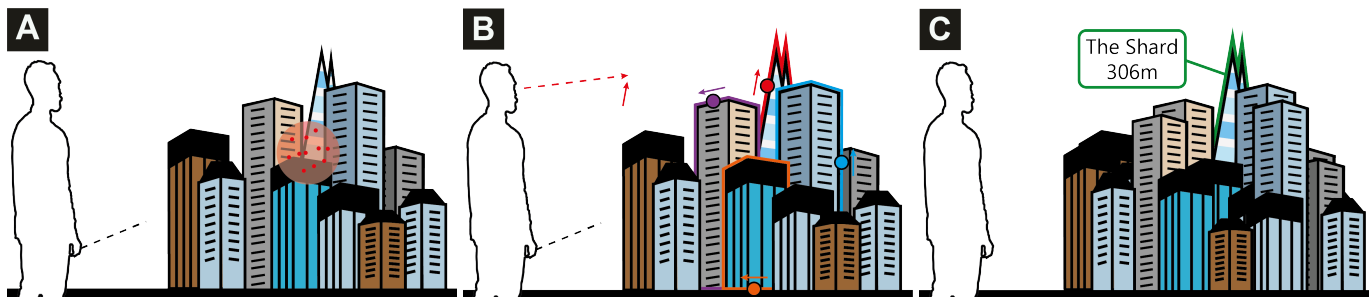***** **contributed equally**

Figure 1. *Outline Pursuits* support selection in occluded 3D scenes. **A: The user points at an object of interest but the selection is ambiguous due to occlusion by other objects. B: Potential targets are outlined, with each outline presenting a moving stimulus that the user can follow with their gaze. C: Matching of the user's smooth pursuit eye movement completes the selection. Note that outline pursuits can augment manual pointing as shown, or support hands-free input using the head or gaze for initial pointing.**

ABSTRACT

In 3D environments, objects can be difficult to select when they overlap, as this affects available target area and increases selection ambiguity. We introduce *Outline Pursuits* which extends a primary pointing modality for gaze-assisted selection of occluded objects. Candidate targets within a pointing cone are presented with an outline that is traversed by a moving stimulus. This affords completion of the selection by gaze attention to the intended target's outline motion, detected by matching the user's smooth pursuit eye movement. We demonstrate two techniques implemented based on the concept, one with a controller as the primary pointer, and one in which Outline Pursuits are combined with head pointing for hands-free selection. Compared with conventional raycasting, the techniques require less movement for selection as users do not need to reposition themselves for a better line of sight, and selection time and accuracy are less affected when targets become highly occluded.

Author Keywords

Virtual reality, Occlusion, Eye tracking, Smooth pursuits

CCS Concepts

•**Human-centered computing** → **Interaction techniques; Virtual reality;**

INTRODUCTION

Pointing is a fundamental task for selection of objects in virtual reality (VR). Users are commonly supported with manual controllers and raycasting [2, 3, 7] but hands-free alternatives are available based on head or gaze tracking [20, 30]. Irrespective of modality, occlusion presents a principal problem for selection in VR environments. Objects that are rendered at different depths can appear overlapping, or even completely occluded, depending on the user's point of view. This reduces the target area for pointing and increases selection ambiguity, as input accuracy is limited by users' motor skills (e.g., hand tremor [47], natural jitter in eye fixations [26]) as well as fidelity and precision of sensing devices. A mitigation strategy is for users to change their position to improve the view they have of an intended target. However, this increases effort, is a limited option when users are seated, and is not possible with portable VR systems that track only rotational head movement but not translation in space (e.g. Oculus Go and FOVE).

In this work we propose to support selection of occluded objects with *Outline Pursuits*. As illustrated in Figure 1, the

concept is to display the outlines of objects that lie in the direction in which the user points, and to generate a distinct motion around each of the outlines. Users can then disambiguate the selection by following the motion around their intended target with their eyes. Outline Pursuits leverage that users look at objects they aim to select, and that it is natural for humans to follow a displayed motion with corresponding smooth pursuit eye movement. Previous work has shown that smooth pursuit is highly effective for selection, as it is a natural closed-loop behaviour that are our eyes only exhibit when they are presented with a moving stimulus [28, 43]. Smooth pursuit can be robustly detected as it is distinct from saccadic eye movement, and the associated target can be inferred by motion correlation [40]. The pursuit technique does not require any calibration of eye gaze to the display coordinates as it matches the user's gaze against targets based on their relative motion [43], and for the same reason is also independent of target size [10].

Outline Pursuits can be combined with any pointing modality. We introduce two specific techniques, one for hands-free selection, and one that augments pointing with a controller. In *Hands-free Outline Pursuits*, potential targets are identified by head pointing followed by smooth pursuit of one of the generated outlines. In this technique, Outline Pursuits not only disambiguates targets but is also used for final selection confirmation. This has the advantage that users are free to use their hands for other tasks or to rest them (e.g. avoiding gorilla arm [13]) while addressing usability issues of hands-free alternatives, such as unnatural dwell times for selection.

In *Controller-based Outline Pursuits*, pointing is initiated with a handheld controller. In this technique, Outline Pursuits is used to provide the user with feedback not only to pre-select candidate targets but also to show which of these matches their eye movement most closely. This enables the user to complete a selection with a 'click' as soon as the intended target is identified. The technique combines advantages of using a controller for familiar point-and-click input with Outline Pursuits to resolve occlusion problems, including the disparity between the user's line of sight to a target and the ray cast from the controller (i.e. when a user has direct line of sight, while the hand controller does not) [1].

We develop the contributions of this work as follows. First, we elaborate Outline Pursuits as a general principle for disambiguation and selection in occluded space. This includes consideration of how candidate targets are identified; how outlines are generated for objects of any shape; and how motion is generated to facilitate selection by pursuit. Next, we describe the hands-free and controller-based techniques each with an example application that we implemented to demonstrate the respective advantages of the techniques. Finally we present two user studies, one that evaluates alternative ways of presenting outlines and movement paths for occluded objects, and one to compare our techniques against baselines of controller- and gaze-based pointing and selection.

RELATED WORK

Outline Pursuits are designed for selection of objects in 3D environments, where they can be beyond manual reach, at different viewing angles and distances, and occluded by other objects. The most commonly used metaphor is *Raycasting* where the user controls a ray via a controller or body part [14]. Raycasting allows for selection of out of reach targets, however targets at greater distances are harder to select due to limitations of human motor control and motion tracking. To compensate for accuracy and occlusion issues, raycasting can be combined with use of volume for selection, such as first demonstrated with the Silk Cursor [46]. In our work, we base Outline Pursuits on a cone cast from the pointing device.

Use of a volume for selection requires disambiguation from among initially selected objects, for which a range of techniques exist. Many of these techniques require an additional manual step for final selection [3, 11, 19], while others apply heuristics or contextual information for implicit disambiguation [8, 11, 33, 37]. In our technique, disambiguation is explicit but based on smooth pursuit eye movement instead of a second pointing step.

A range of work has explored hands-free alternatives to pointing and selection in VR, based on head tracking integral with AR/VR headsets, and eye tracking which is also becoming more widely available in HMDs (e.g., FOVE 0, HTC Vive Pro Eye and HoloLens 2). Gaze has been found to be faster than hand pointing, especially for distant objects [39]. A range of works have compared eye and head pointing showing that eye gaze is faster and less strenuous, while head pointing is often preferred as more stable, controlled and accurate [4, 12, 20, 31]. Eye or head pointing can be combined with fast manual confirmation by click [27, 32, 38], or with dwell time for hands-free selection [18, 25, 30]. It has also been proposed to use gaze for coarse-grained selection followed by head movement for subsequent confirmation [22, 36] or refinement of positional input [20]. Other work has proposed techniques that leverage concurrent eye and head movement for interaction and target depth estimation [21, 23, 35]. The hands-free technique we implemented likewise combines head and eye tracking, however with the head tracked for cone-casting, and eye movement matched against the outline motion presented by candidate targets.

Our work builds on insight from prior work on smooth pursuit for interaction, pioneered by Vidal et al. [43]. Pursuits enable gaze selection of targets without prior calibration and is scale-invariant, as demonstrated for example for walk-up-and-use gaze interaction with public displays [43], gaze input at a glance on smartwatches [10], and gaze control across ambient devices [42]. For Outline Pursuits, we leverage that pursuits are robust for selection, provided candidate motions against which they are matched are sufficiently different in shape, direction, phase or velocity. In particular, pursuits avoid the Midas Touch problems of fixation-based gaze techniques, as the eyes only exhibit smooth pursuit when the user attends to a moving object. A few prior works have used smooth pursuit for selection in VR, however for selection of objects presented in motion [18, 30]. A distinct novelty of our work is that we instead present motion around static 3D objects to facilitate their selection by pursuit without modification to the object's size or position.
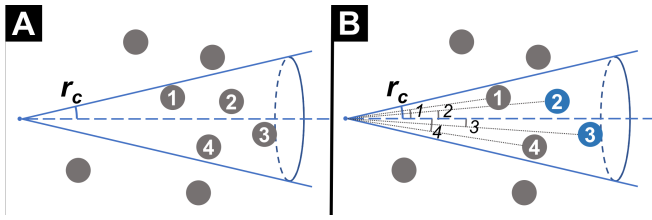
Figure 2. Cone casting with a primary pointer is used to select N_c candidate targets within a visual angle radius r_c. In this example, with $N_c = 2$, four objects are within the cone (A) and objects 2 and 3 are selected as they are closest to the centre of the cone.
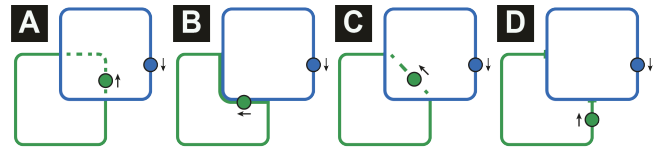

Figure 3. Movement paths for occluded objects. A: *Whole*: the target moves along the whole outline. B: *Shared*: the target moves along the visible part of the object. C: *Cut*: the target moves along the shortest path to the next visible part. D: *Jump*: the target jumps to the next visible part.

In virtual environments, outlining objects has many utilities: it is a common feedback technique to show objects that are interactive or to highlight a current selection. Outlines have also been used in VR environments as an accessibility tool to increase the visual contrast between objects for users with limited vision [48]. The 2D shape of 3D objects has also been used for object selection by using the hand to pick up the 2D silhouette [29], or by using gaze saccades to trace the contour of selectable objects [17]. We add to the utility of outlines by augmenting them with motions for smooth pursuit interaction.

OUTLINE PURSUITS

The interaction principle of Outline Pursuits is defined by the following sequence:

1. A pointing modality is used to pre-select a subset of available objects for interaction;
2. The outline of each object in the subset is extracted;
3. A moving target traverses each outline in such a way that no two motions are exactly the same;
4. A user signals their intent to interact with an object by following the moving target with their gaze;
5. The system detects a user is following a target based on correlation of the target's movement and user's eye positions;
6. The detected object is selected based upon a given criteria (e.g. threshold or duration), or upon confirmation from another modality (e.g. button click).

From a user's perspective, the interaction involves two stages: *candidate selection* where potential targets are identified with the pointing modality, and *target selection* where one of the candidates is selected using smooth pursuits.

Candidate Selection

The first stage is designed to reveal and highlight a limited number of potential targets using coarse-grained pointing.

Cone Casting

Figure 2 shows the cone casting technique. Candidate objects are identified by casting a ray and selecting the nearest N_c targets to the ray direction, within a given visual angle radius, r_c. The ray-casting can be performed by any type of pointing modality, for example a controller, a finger, the head, or gaze. We limit the number of targets to reduce visual clutter in subsequent outlining of candidates, and as smooth pursuit is more robust for selection when the number of selectables is limited [10, 43]. The size of the cone is constrained so that only objects within the user's area of attention are highlighted

for interaction. The choices of N_c and r_c should encourage coarse-grained, lazy pointing for the initial candidate selection.

Outline Generation

Once a subset of candidate objects have been selected we extract the object outlines. There are multiple approaches available to extract an object outline in VR, e.g. silhouette detection [44]. We use an object-space approach where the information stored in the virtual 3D meshes is used for outline generation. Specifically, our outlines are found by comparing the two normals of each mesh edge and looking for cases where one normal faces the user while the other faces away. We further refine this so that any edge that is not part of the outer outline is disregarded.

Displaying the outline to the user acts as a guide, making the motion of the moving target more predictable. Outlines can also be used to show the presence and full shape of occluded objects. An important design aspect of the outlines is their behaviour in the event of object occlusion. We consider four different outline behaviours (see Fig. 3). Only the *whole outline* version shows the outline when the object is fully occluded. As such, the other three techniques require line-of-sight to an object in order to generate an outline. In addition, the choice of outline behaviour may affect the user's ability to follow the target, and/or the performance of the motion correlation algorithm performing the match.

Motion Generation

Once the outlines of the candidate objects have been extracted, we augment them with a moving target for smooth pursuit selection. In the worst case all outlines of the candidate objects assume the same shape, in which case the difference in spatiotemporal properties of the target movements should be maximised to ensure each target's trajectory is unique. In order to ensure maximum variability between the target movements we optimise the direction, phase, and speed of their movement, whilst ensuring that they facilitate comfortable and seamless smooth pursuits eye movements.

When more than one candidate object is selected, they are grouped so that as many candidates as possible move in opposite directions, i.e. half move clockwise, the other anticlockwise. To further distinguish motions moving in the same direction, the difference in the objects' phases are maximised. For phase difference, the targets must have equal number of revolutions per second of their respective outlines so that trajectories do not overlap. The number of revolutions is selected to ensure that the speed, as measured in visual angle, of the target remains within a suitable range for smooth pursuit eye

movements, $[v_{min}, v_{max}]$. For example, if two targets move in the same direction, we start one target at the top of its trajectory and the other at the bottom (i.e. 180° phase separation). If this cannot be achieved (e.g. due to different object sizes) we change the revolutions per second of the targets. If the object size difference is so significant that we are not able to find a revolution suitable for all candidates, the object with the most distinct object size is set to the closest of v_{min} or v_{max}.

Target Selection

Target selection consists of disambiguating one target amongst many, prior to confirming the selection. The fundamental mechanism underlying Outline Pursuits is the smooth pursuit eye movement - a specific eye movement type that occurs when a user looks at a moving target. The system takes as input the user's eye positions and all other target positions as spatiotemporal trajectories, and calculates a similarity score to determine which target the user is looking at. There are many different approaches to determining whether a match exists [9, 41, 43]. In these previous works, smooth pursuits is used as a selection technique in which target disambiguation and confirmation occur in a seamless fashion once given criteria of the matching process are satisfied. In the next section we discuss how modalities afforded by VR setups can be used in combination with the smooth pursuits technique for selection.

TECHNIQUES

We demonstrate Outline Pursuits with implementation of two concrete techniques, one for hands-free selection, and one where it extends controller-based raycasting.

Hands-free Outline Pursuits

Outline Pursuits can be used in a hands-free manner, similar to traditional smooth pursuit selection techniques. VR commonly enables direct manipulation of virtual objects with either the hands or a controller. By providing a hands-free technique users can interact with out-of-reach objects without having to put down or stop interacting with virtual objects already in-hand. It is also compelling for users who do not have the use of their hands, or for simpler hands-free applications. In contrast to other gaze-based techniques, such as gaze-dwell, Hands-free Outline Pursuits enables a user to gaze upon an object for an indefinite period without making a selection.

In the hands-free version, we use the head as the pointing mechanism for candidate selection. The head ray is readily available in VR HMDs in three degrees of freedom, and in more advanced systems with six degrees of freedom (i.e. plus translational movement in 3D space). Prior research has shown a symbiosis between head and gaze movements, with gaze commonly falling within 25° of the centre of the head making it a compelling proxy for gaze [34]. The gaze-ray itself could be used for candidate selection, however this would necessitate eye tracker calibration and by using the head-ray we leverage the ability of smooth pursuits to match relative movements. In addition, the head-ray does not constrain the outlines to being in the middle of the user's focus at all times. We limit the number of candidate objects to 4 to minimise visual clutter whilst optimising the chance of a user successfully selecting a target [43, 10]. In order to facilitate smooth pursuit eye
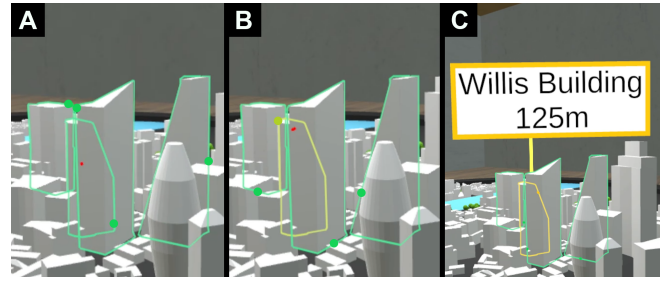

Figure 4. Hands-free Outline Pursuits in an interactive virtual cityscape environment. A: The user selects candidate buildings by pointing with their head; B: The user follows the outline motion with their gaze (red) to select the building. The outline shows feedback of selection progress via colour change. C: The building is selected and contextual feedback is displayed.

movements we restrict the speeds of the targets, measured in visual angle, to a range of $[3°, 15°]$ [9, 15].

For the matching process we use the 2D correlation algorithm introduced by Velloso et al. [41]. As input we use the pupil positions reported by the eye tracker as opposed to the gaze position which requires calibration. We use a sliding window of size N_m frames to calculate the correlation and add the result to a sliding post-hoc buffer of length N_{phoc} as described in [41]. If a given number of values, N_{valid} in the post-hoc buffer are below a given correlation threshold c_t we assume the motions are matched and the corresponding target is selected.

Interactive Virtual Cityscape

We developed a virtual cityscape application to illustrate Hands-free Outline Pursuits (Fig. 4). The application lets users explore the city and individual buildings using only their head and eyes, with a selection triggering a contextual display over the building. The application demonstrates several advantages of Hands-free Outline Pursuits:

- Usage of outlines are two fold: showing which buildings are interactive, and for use in the selection process;
- Buildings may be gazed upon indefinitely without selection;
- Outlines adapt to the user's perspective, allowing selection from all viewpoints;
- Outlines adapt to nearby buildings that occlude individual building, allowing accurate selection of partially occluded buildings.

Controller-based Outline Pursuits

In the second version of Outline Pursuits we demonstrate how a controller can be used to provide selection with minimal physical effort. Similar to HMDs, not all VR controllers support translational movement and may be limited to 3 degrees of freedom. The aim of Controller-based Outline Pursuits is to reduce the amount of physical effort required by utilising a coarse-grained pointing stage that can be performed with much less effort compared to traditional techniques such as ray-casting. This is followed by a smooth pursuits based disambiguation and confirmation stage. In contrast to the hands-free variant, the controller-based version provides greater interaction capabilities once an object is selected, such as object manipulation in 3D space.

The initial candidate selection is performed by pointing with the controller. This affords the user more control compared with head or gaze because they do not have to be continuously pointing towards their field of view and can be more selective as to when to show the outlines. We then separate target disambiguation and confirmation into two distinct steps: a candidate target is highlighted once a user follows its motion; the user then confirms the selection with a button click. The separation of target disambiguation and confirmation brings a number of advantages.

From a system design perspective the criteria for defining a smooth pursuits match can be reduced to a similarity-based classification problem. Conventionally, determining smooth pursuit matching criteria is non-trivial and often necessitates optimisation of a combination of different thresholds and window sizes, all of which could be affected by the properties of the target's trajectory and sensor characteristics [5, 6, 10]. Choice of parameters should ensure that users are able to select objects in a timely fashion, whilst being robust to incorrect selections due to natural eye movements, also known as the Midas touch [16]. However, by combining smooth pursuits with a simple confirmation technique we need only be concerned with the target that gives the best match to the eye movements, thus reducing the parameter space. This also helps reduce the Midas touch problem as the user needs to explicitly confirm the selection. From a user's perspective, separating disambiguation and confirmation stages provides the opportunity to highlight the targets for selection based upon specific timing (e.g. selecting a target at a precise time), or when a user wishes to learn more about an object prior to making a selection.

Similar to the hands-free variant, we use the 2D correlation technique for determining the closest match [41]. We employ the same post-hoc buffer to evaluate the candidate target with the best match to the user's eye movements in order to avoid the highlighted target changing spuriously. The highlighted target is the candidate target with the lowest mean value in the post-hoc window, and is highlighted via the outline changing colour. We do not wait for the post-hoc window to be full before an object was considered for selection because this would add a time delay until the user could confirm their selection.

Room planner
Figure 5 illustrates Controller-based Outline Pursuits in a room planner setting. Controller-based Outline Pursuits allows users to more easily select items that are either partially occluded or that are in close proximity with other objects, for example, books stacked on a shelf. Once selected, users are able to place and move around furniture and decorations, using direct manipulation with the controller. The application demonstrates several advantages of the controller-based variant of Outline Pursuits:

- Highlighting prior to selection allows the user to check additional details prior to confirming selection;
- Outlines can be shown through other objects so there is less need to move around the 3D space to get a clearer viewpoint for selection;
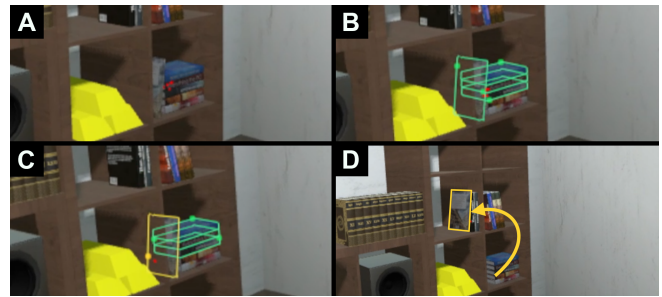

Figure 5. Controller-based Outline Pursuits in a Room Planner setting. A: The user wants to move a partially occluded book to a different shelf but does not know which book is which. B: The user points with the controller to pre-select the book and other nearby items, showing which books are available for interaction via their outline and stimuli movements. C: The user follows the motion of a book with their eyes, which in turn highlights it. D: The user selects the book via a button click and can now manipulate it in 3D space, moving it to another shelf.

- Users can select partially occluded or close proximity objects with minimal adjustment of their hand movement or point of view;
- Users are able to select an object with little hand movement and users can thus plan their hand placement for the following object manipulation.

USER STUDIES
To evaluate the proposed techniques we conducted two user studies. In the first study we investigate how the outline behaviour during occlusion affects users' selection performance and preferences. We then evaluate the Outline Pursuits techniques against traditional selection techniques involving both a controller and eye gaze.

USER STUDY 1 - OUTLINE EVALUATION
One of the most important design decisions for Outline Pursuits is how to display the outlines during occlusion as it highly impacts the movements required for selection. In this study, we explore whether the choice of outline affects user performance, and user preference, using the four different variants of outlines (Fig. 3). We focus our study on partial occlusion as some of the presented outlines require line of sight to display an outline, and only evaluate the outlines using Hands-free Outline Pursuits.

Tasks
The user study was separated into three tasks (Fig. 6). Task one and two consisted of selecting known objects as accurately and quickly as possible at two levels of occlusion. In the first level of occlusion (Fig. 6a) each trial consisted of two cubes which occluded each other. In the second level, each trial consisted of three cubes that were occluding each other (Fig. 6b). Participants performed 16 trials in the first level, of which half were non-occluded and the other half occluded. In the second level, participants performed 12 trials, four trials for each level of occlusion. Trial order was randomised for both levels. The final task (Fig. 6c), required higher cognitive demand and consisted of multiple sequential selections in a memory game setting. We chose this because users may wish to look at objects without selecting them. Participants were
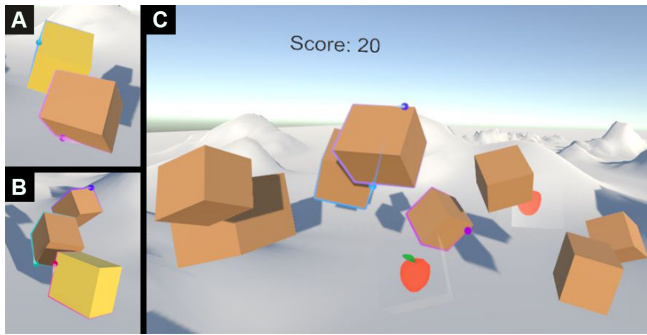
Figure 6. A: Participants were tasked to select the yellow object. B: Participants were tasked to select the yellow object at a higher level of occlusion. C: The user was tasked to finish the memory game.

tasked to complete the memory game as quickly and with as few selections as possible. The memory game consisted of 8 objects (4 pairs). All objects in the study were cubes and had a size of 1.25m across and displayed at distances ranging from 7-9m. No object was occluded more than 50%.

Apparatus

Hands-free Outline Pursuits, the outlines, and tasks were developed using Unity version 2017.3.1. We used an HTC Vive (90Hz) with the Tobii Pro VR Integration eye tracker (120Hz). All gaze data was filtered via a 5-median filter. All outline were visible to the user. We chose the following parameters for hands-free Outline Pursuits: $r_c = 10°$, $N_m = 90$, $N_{phoc} = 45$, $N_{valid} = 30$, $c_t = 0.2$. The target size was set to 0.5°.

Procedure

We recruited 16 participants (8 male, 8 female, 27.3±3.7) for the study. 14 participants reported occasional or no VR experience, while 2 participants had daily to weekly experience. 15 participants reported occasional or no eye tracking experience, while 1 participant reported daily experience. Participants first signed a consent form and answered a demographic questionnaire. Participants were then seated and put on the HMD. The eye tracker was not calibrated during the study. Every participant performed the tasks in the same order for each condition, starting with the abstract tasks and finishing with the memory game. Outline order was counterbalanced with a Latin square. After performing all tasks with an outline, participants were asked to remove the HMD and answer a questionnaire consisting of nine 5-point Likert items based on common usability factors (Exhaustion, Responsiveness, Physical Effort, Mental Effort, Ease, Precision, Comfort, Fun, Confidence). An interview to extract preferences was conducted after all tasks had been completed. The study took 30 minutes to complete.

Results

We compared the average time to select a target from when the outline is first displayed for all tasks, and the error rates for the first two tasks. Repeated measures ANOVA showed no significance differences between the outlines (Table 1). The results indicate that outline type does not have an impact on selection performance for the given levels of occlusion. Friedman tests on the questionnaire items showed no significant differences. When asked about their preferred outline, 8 participants chose

Table 1. Study 1 task performance and repeated measures ANOVA.

	Level 1		Level 2		Memory
	Time (s)	Error (#)	Time (s)	Error (#)	Time (s)
Whole	3.79 ±1.02	0.94 ±0.83	4.14 ±1.28	1.00 ±1.32	4.42 ±1.67
Shared	4.47 ±2.26	1.69 ±1.86	4.42 ±1.55	1.63 ±1.36	4.55 ±1.16
Cut	3.80 ±0.82	0.94 ±0.90	3.84 ±1.24	1.06 ±0.75	4.45 ±1.02
Jump	3.77 ±1.01	1.31 ±1.04	3.85 ±1.06	1.69 ±1.16	4.57 ±1.34
ANOVA	$F_{1.5,22.3}$=1.20, p=.308	$F_{1.9,28.2}$=1.28, p=.294	$F_{3,45}$=1.51, p=.225	$F_{3,45}$=1.91, p=.142	$F_{3,45}$=0.09, p=.966

Whole outline, 5 participants picked Shared outline, 2 participants picked Cut outline, and 1 participant chose Jump outline. Further questioning revealed that each outline type showed certain benefits and weaknesses according to participants:

Whole outline was considered the easiest and most predictable of the outlines. Participants mentioned that it would be especially suitable for beginners due to its consistency. However, participants also mentioned that the outline movement could be straining due to depth issues when the stimulus is shown through other objects.

Shared outline was reported to require less concentration as participants only had to move their gaze along visible parts of the object. However, participants also thought that the common path used by multiple motions made them hard to distinguish, especially when the motions moved along the common path at the same time. Additionally, false selections could happen if two motions move along the common path in the same direction at the same time.

Cut outline was preferred due to spending less time in the occluded or ambiguous part of the outline as it moves along the shortest path to the next non-occluded part. This also simplifies the shape of the motion, which some participants expressed preference for. However, it was also counter-intuitive to participants as it did not follow a visible object edge nor the object shape.

Jump outline proved to be irritating to many participants due to the sudden "teleportation" of the target. Participants commented that the jumping target made them lose focus on the task and it was difficult to predict where the target would jump to. However, participants also expressed that it has the advantage of spending no time in the occluded part, and it was suggested that this could minimise confusion regarding which target belongs to which object.

Discussion

Performance and usability ratings were consistent across outlines, which indicates that designers can freely choose an outline that fits their application without risking major performance penalties. Based on these results we choose to use the *whole outline* in the comparative study as it was the most favoured technique among participants and it does not depend on line of sight. Further work could investigate adaptive outline generation where the outline behaviour can change depending on metrics such as shape simplicity.

USER STUDY 2 - TECHNIQUE EVALUATION

In the second study, we investigate the performance and user reception of the proposed Outline Pursuits techniques for se-

lection in a VR environment. We select baseline techniques from the literature to compare the Outline Pursuits techniques against. For comparison against the controller-based Outline Pursuits (CoOp) technique we chose both controller- (CoRa) and gaze-based (GaRa) ray casting approaches. The former uses a ray from the controller which highlights the first object that the ray intersects. The latter uses the gaze ray for highlighting the first object the gaze intersects. In both cases confirmation is made via button click. For comparison against the hands-free based Outline Pursuits (HfOp) approach we used a gaze-dwell technique (GaDw). Similar to the gaze-based ray-casting approach the dwell technique highlights the first object that the gaze intersects. Confirmation is made once the user has dwelled upon the object for one second, with visual feedback being provided as to the progress.

Task
Participants were tasked with selecting a target object among many as fast as possible at different distances, object densities, and levels of occlusion. Various simple shapes (spheres, cubes, cylinders and capsules) were used as objects, and spread out in pseudo-random positions and rotations in a 2 metre diameter sphere. All objects were visible in the initial field of view and no objects intersected with each other. Objects were of the same size (15 cm across) and each object type was equally used as the highlighted object. All trials for each condition were generated beforehand and used for all techniques for fairness. 16 trials were generated for each target density and each trial was presented at all three distances, resulting in 96 trials per technique and participant. For half of the trials, the highlighted object had less than 50% occlusion while the other half had more than 50% occlusion. Similarly, half the trials had the highlighted object positioned in the front half of the sphere, while the other half had the objects position in the back half of the sphere. Participants could not move on to the next trial until the correct object had been selected or 10 seconds had elapsed. Trials were presented to participants in a random order. Selection time, error rate, and participant movements were recorded for each trial. The independent variables for our study were:

- TECHNIQUE: {CoOp, CoRa, GaRa, HfOp, GaDw}
- DENSITY: {16, 40 objects}
- DISTANCE: {3, 5, 7 metres}
- OCCLUSION: {< 50%, > 50%}

Apparatus
We used the same equipment and setup as in our first user study. We found no significant performance differences between different outline types in the previous study and chose the whole outline for use in this study based on user preferences in the last. Based on feedback from participants we also increased the size of r_c to 15° and c_t to 0.4. In addition, we added a target movement delay of 10 frames for increased correlation performance [9]. For CoOp, we used the same common parameters as HfOp. We used the HTC Vive controller touchpad for selection confirmation for CoOp, CoRa and GaRa. We used a dwell timer of 1s (same as $N_m = 90$) for GaDw. A gaze cursor was used for GaDw and GaRa to show the current gaze position in case participants needed to

further adjust their gaze position due to sensor inaccuracies. Similarly, we used a ray to show the current pointing direction for CoRa. We do not show a ray for the candidate selection stage of either Outline Pursuits technique. The outline was used for object highlighting for all pointing-based techniques.

Procedure
We recruited 20 participants to undertake the study (12 male, 8 female, 28.7±6.3) of which 6 had participated in the previous experiment. The time between experiments was 6 weeks suggesting minimal acquired learning. 18 participants reported none or occasional VR experience, while 2 participants had daily experience. 3 participants reported daily to weekly eye tracking experience, while the rest reported none or occasional experience. 8 participants reported previous experience with smooth pursuit interfaces. Participants first signed a consent form and answered a demographic questionnaire. Participants were then instructed to stand in the designated participant area and put on the HMD. Participants performed a five-point eye tracking calibration for GaRa and GaDw techniques. The eye tracker was not calibrated for CoOp, HfOp and CoRa. Participants had a training session before each test session. Selection technique order was counterbalanced with a Latin square. After completing the task with a technique, participants removed the HMD and filled out a questionnaire consisting of 11 5-scale Likert items based on usability factors. A semi-structured interview was conducted at the end to extract preferences and opinions. In total each participant performed $5 \times 3 \times 2 \times 2 \times 4 \times 2 = 480$ selections. The study took 50-60 minutes to complete.

Results
In this section we analyse and present both performance metrics and user preference data to evaluate how the Outline Pursuits techniques compare against the baseline techniques. Our main focus in the comparison is between the controller-based techniques (CoOp, CoRa, GaRa) and the hands-free techniques (HfOp, GaDw). The eye tracker was able to record gaze data at a mean gaze accuracy of 0.75±0.23°.

Performance Metrics
Our four main dependent variables are selection time, error rate, head movement, and controller movement.

Selection Time
For selection time we assume no penalty for incorrect selections. That is we record the time taken for the user to select the correct target irrespective of whether or not they select the incorrect target prior to the correct. We conducted a four-way repeated measures ANOVA ($5 \times 3 \times 2 \times 2$), Greenhouse-Geiser-corrected in the cases where Mauchly's test indicated a violation of sphericity and with Bonferroni-corrected post-hoc tests where applicable. Only successful times contributed to the average time per trial, and 3 cells (out of 1200) which had missing values were replaced with the maximum value over all participants. Using SPSS, we found seven instances of "extreme" outliers which were removed by performing 90% winsorization. Shapiro-Wilks tests and visual inspection of Q-Q plots were used to validate the assumption of normality.
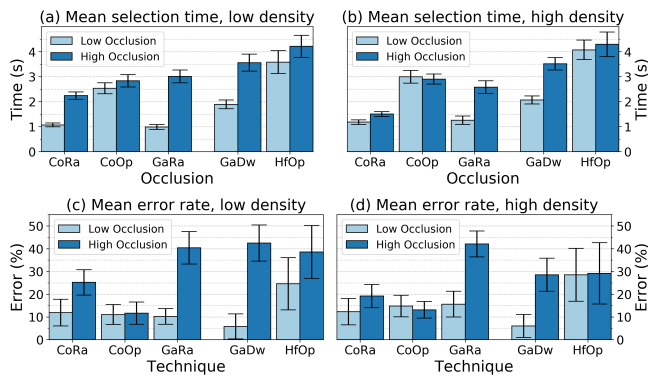
Figure 7. Study 2 task performance at different object densities and levels of occlusion. Error bars represents mean 95% confidence interval.

There was no statistical four-way interaction, however we found a significant three-way interaction between TECH×DENS×OCC ($F_{4,76}$=4.0, p=.005). We also find two-way interactions for TECH×OCC ($F_{3.0,56.6}$=77.5, p<.001) and TECH×DENS ($F_{2.5,47.8}$=12.7, p<.001).

We found a significant main effect for TECH ($F_{2.0,38.2}$=100.9, p<.001) showing significant differences (p=.001) between all TECHNIQUES except for CoOp and GaDw. The results also showed that CoRa was the overall fastest technique (CoRa: 1.51s, GaRa: 1.96s, CoOp: 2.81s, GaDw: 2.80s, HfOp: 4.03s). DIST ($F_{2,38}$=79.0, p<.001) also revealed a significant significant main effect, where greater distances resulted in higher selection times (3m: 2.37s, 5m: 2.59s, 7m: 2.91s). However, we found no significant two-way interaction between TECH×DIST, indicating that selection times for each technique were equally affected by distance. For the Outline Pursuit techniques, the increase in selection time for larger distances could be due to the coarse pointing element involved.

We investigated the three-way repeated measures ANOVA on TECH×DENS×OCC across all levels of DIST. A simple two-way interaction of OCC×TECH (see Fig 7) revealed that for high density targets the time taken to select using Outline Pursuits techniques is independent of occlusion. For all other techniques it takes significantly longer (p<.05) to select targets that are highly occluded.

Error Rate
We define an error as when the participant chooses another target prior to the correct selection, or when no selection is made within 10 seconds. The number of errors was positively skewed and violated the repeated measures ANOVA's assumption of normality after the usual transformations, and the Align Rank Transform technique [45] showed the aligned responses did not sum to ≈0. Using the number of errors as count data we fit an "overdispersed" Poisson regression model [24]. We report the number of errors as the error rate, i.e. the number of trials resulting in an error divided by the total number of trials.

We included all interactions involving TECH in the regression and found the overall model was significant, $\chi^2(39, N=1160)$=552.0, p<.001. Investigation of model effects revealed a significant three-way interaction for

TECH×DIST×DENS ($\chi^2(5)$=12.2, p=.032), a significant two-way interaction for TECH×OCC ($\chi^2(4)$=19.7, p=.001), and significant main effects for TECH ($\chi^2(4)$=20.5, p<.001), DIST ($\chi^2(1)$=76.0, p<.001), and OCC ($\chi^2(1)$=25.2, p<.001).

We ran sequential Šidák pairwise comparisons to assess the effect of the different techniques on the two-way interaction and main effects. The TECH×OCC (Fig. 7) revealed that the error rates for the Outline Pursuits techniques were independent of occlusion, whereas all other techniques had significantly higher error rates for high occlusion targets (CoRa: p=.001, GaRa: p<.001, GaDw: p<.001). Furthermore, CoOp was the most accurate technique for highly occluded targets (12%), significantly outperforming both CoRa (22%, p=.004) and GaRa (41%, p<.001).

The main effect of TECH revealed HfOp (30%) had a significantly higher overall error rate compared to all other techniques, followed by GaRa (27%) which was also significantly different to all other techniques. Results showed no significant difference in overall error rate between the controller-based techniques (CoRa: 17%, CoOp: 13%, GaDw: 21%). We also found that selecting targets at a distance induced more errors, with error rates rising from 15% at 3m, to 21% at 5m, and finally 29% at 7m. The model revealed that every metre added to the distance resulted in 1.305 (95% CI, 1.134 to 1.502) times more errors, a statistically significant result (p<.001).

Body Movements
We compare the average translational head and controller movement per trial by using the 3D coordinates reported by the HTC Vive tracker. For controller movement we only consider CoRa and CoOp. In all cases we low-pass filter the data with a 3rd order Butterworth filter to remove high frequency sensor noise. We performed two sets of four-way repeated measures ANOVA ($5 \times 3 \times 2 \times 2$), Greenhouse-Geisser corrected when Mauchly's test indicated a violation of sphericity, and with Bonferonni-corrected post-hoc tests where applicable. Both modalities violated the assumption of normality, however $\log_{10}$ transformation of the data resulted in normalised data according to Shairo-Wilk's test of normality (p>.05) and visual inspection of Q-Q plots. All further analyses are performed on the transformed data but we report results in the raw format of metres for greater interpretability.

Head Movements
We found a significant four-way interaction ($F_{8,152}$=2.1, p=.039), three-way interactions between TECH×DENS×OCC ($F_{2.6,49.0}$=15.3, p<.001), and TECH×DENS×DIST ($F_{5.0,94.4}$=5.2, p<.001), two-way interactions between TECH×DENS ($F_{3.0,56.3}$=18.9, p<.001), and main effects of DIST ($F_{2,38}$=25.6, p<.001), DENS ($F_{1,19}$=20.0, p<.001), and OCC ($F_{1,19}$=468.5, p<.001).

A two-way TECH×OCC interaction ($F_{4,76}$=138.0, p<.001) (fig. 8a) revealed no significant difference between the three controller-based techniques at low occlusion, however they all require significantly less head movement than the two hands-free techniques (p<.033). HfOp and GaDw in turn, showed no significant difference between each other (p=1.00, CoRa: 0.04m, CoOp: 0.06m, HfOp: 0.13m, GaRa: 0.08m, GaDw:
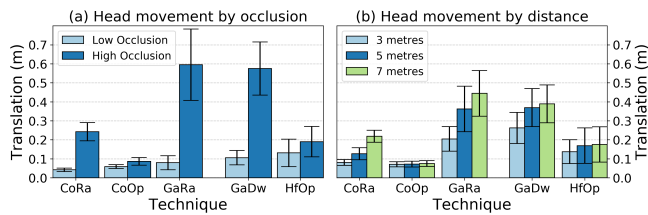
Figure 8. Study 2 average translational head movements during trials. Error bars represents 95% confidence interval for means.

Figure 9. Median scores on a 5-point Likert scale with error bars representing interquartile ranges.

0.11m). For high occlusion targets, CoOp required significantly less head movement than all other techniques ($p<$.008, CoRa: 0.24m, CoOp: 0.09m, HfOp: 0.19m, GaRa: 0.59m, GaDw: 0.57m). In addition, HfOp required significantly less head movement than both gaze-based techniques ($p<$.001). Finally, all techniques required significantly more head movement for highly occluded targets than low occlusion ($p<$.001).

A TECH×DIST ($F_{8,152}$=3.1, p=.003) (fig. 8b) interaction revealed that CoRa requires significantly more head movement as the distance to the target increases at ($p\leq$.008; 3m: 0.08m, 5m: 0.13m, 7m: 0.22m). In contrast, the head movement required when using CoOp for selection is independent of distance (p=1.00; 3m: 0.07m, 5m: 0.07m, 7m: 0.07m). For GaRa we find a significant difference between 3m and all others ($p<$.001), however no significant difference between 5m and 7m (3m: 0.20m, 5m: 0.36m, 7m: 0.44m). A similar pattern was found with GaDw which also showed a significant difference between 3m and all others, yet no significance between 5m and 7m (3m: 0.26m, 5m: 0.37m, 7m: 0.39m). For HfOp, we found a small (4cm) but significant difference in head movement between 3m and 7m, however no significant difference involving 5m (3m: 0.14m, 5m: 0.17m, 7m: 0.18m).

Controller Movements
We found a significant four-way interaction ($F_{2,38}$=3.7, p=.034), three-way interactions between TECH×DENS×OCC ($F_{1,19}$=80.9, $p<$.001), and TECH×DENS×DIST ($F_{2,38}$=11.6, $p<$.001), two-way interactions for TECH×DENS ($F_{1,19}$=98.9, $p<$.001), and main effects for OCC ($F_{1,19}$=339.6, $p<$.001), DIST ($F_{2,38}$=3.5, p=.040), and DENS ($F_{1,19}$=21.5, $p<$.001).

A TECH×OCC ($F_{1,19}$=121.5, $p<$.001) interaction revealed no statistical significance between CoRa (0.13m) and CoOp (0.13m) for low occlusion, yet there is a significant difference between the techniques for high occlusion with CoOp (0.15m) requiring much less controller movement than CoRa (0.49m), at $p<$.001. Both techniques require significantly more controller movement for high occlusion targets over low occlusion (CoRa: $p<$.001, CoOp: p=.006), despite CoOp only requiring an extra 2cm of movement for highly occluded targets compared with low occlusion targets.

The simple-simple main effects of DIST of the three-way interaction of TECH×DIST×OCC ($F_{2,38}$=3.4, p=.044) revealed that at low occlusion CoRa required less movement for targets at 7m (0.11m) compared with 3m (0.15m) (at p=.004). Yet, for highly occluded targets the trend was reversed and 7m (0.66m) required more movement than 3m (0.35m) at $p<$.001, and 5m (0.46m) at p=.041. In contrast, there were no signif-
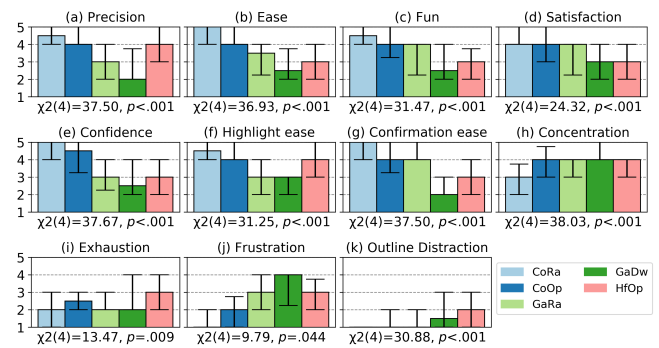
icant simple-simple main effects of DIST for CoOp at either low (3m: 0.14m, 5m: 0.13m, 7m: 0.13m) or high (3m: 0.17m, 5m: 0.15m, 7m: 0.15m) occlusion, indicating that distance has no effect on the amount of controller movement required.

Preference
Friedman tests on usability ratings (Fig. 9) showed significant differences on all metrics. Bonferroni corrected Wilcoxon post-hoc tests showed overall that CoOp and CoRa were more favourably rated than HfOp, GaDw and GaRa, with no significant differences between CoOp and CoRa. Also, HfOp was rated more favourably over both gaze-based techniques. HfOp showed significantly higher precision scores than GaDw (p=.009), and higher ease of highlight scores than both GaDw and GaRa ($p\leq$.004). This indicates that users perceived the coarse pre-selection with the head to be easier than precise gaze pointing. However, CoOp and CoRa showed significantly better usability scores than HfOp in Ease, Fun, Confirmation ease, and Frustration (all $p\leq$.009). In addition CoRa had significantly better usability scores than HfOp in Satisfaction, Confidence, Concentration, and Exhaustion (all $p\leq$.010). These results indicate a preference for controller-based techniques over their hands-free counterparts. CoOp showed significantly higher scores than GaRa and GaDw in Precision, Confidence, and Highlight ease (all $p\leq$.001). CoOp was also considered significantly better than GaDw in Ease, Fun, Satisfaction, Confirmation Ease, and Frustration (all $p\leq$.005).

Interview results showed that technique preferences were split. However, we could again see a general preference for controller-based techniques. The most favoured technique (8 participants) was CoRa. Participants mentioned that CoRa was *"easy"* (P5) and *"accurate"* (P13), and that CoRa gave them a *"wider range of movement"* (P14). Both CoOp and GaRa had 4 participants selecting them as their favourite. Participants thought CoOp was *"clear"* (P10) and *"consistent"* (P17), and participants mentioned that they liked that they could *"see through objects"* (P1). GaRa was praised for being *"direct"* (P4) and *"quick"* (P15). Participants were torn between CoRa and CoOp for preference, however CoRa was usually preferred in comparison to CoOp as it was easier and quicker to select targets at low occlusion. P12 mentioned: "With CoOp and HfOp, it was much easier to follow a motion of an occluded object compared to trying to angle the pointer around a block-

ing object. However, they felt a bit slow when there was no occlusion". HfOp was preferred by 3 participants who liked that they *"only had to follow targets"* (P6) and that it was *"simple"* (P8). Finally, one participant preferred GaDw due to being *"minimal effort"* (P18).

When asked about their preferred pre-selection modality, 9 participants expressed preference for the controller. The participants thought the controller was *"simple"* (P1) and *"effortless"* (P20), while they had to *"consciously think about moving the head"* (P6). On the other hand, 6 participants expressed preference for using the head, stating that it felt *"easy"* (P16) and *"natural"* (P19), while stating that the controller was more *"difficult"* (P2) due to their hands shaking. 5 participants reported that they did not feel a significant difference between the modalities.

DISCUSSION
Outline Pursuits is effective for the discovery and selection of (even fully) occluded objects, without users needing to move to change their perspective or without changing the properties of the objects (e.g. size, motion, shape). The reduction in movement compared with alternative techniques is significant, and highly relevant for contexts where user motion input is constrained by context (e.g. seated usage), technology/sensing, user ability or fatigue.

Outline Pursuits is versatile and lends itself to implementation in different configurations of pre-selection modality, motion feedback, and use of smooth pursuits. Contrasting of HfOp and CoOp, as well as the two studies, gives insight into design choices and their implications. The most significant difference pertains to smooth pursuits: in HfOp used for finding a best match above a threshold (as in prior work on smooth pursuit for selection), versus in CoOp for finding a best match that is confirmed by a different modality with advantages in feedback provision in the process, reduction of the parameter space, and avoidance of Midas Touch and false positive selection. This is a novel aspect and insight of relevance beyond Outline Pursuits, for adoption of smooth pursuit for HCI in general.

In comparison among controller-based techniques, CoRa and GaRa are faster and as accurate as CoOp when occlusion is low, but CoOp showed significantly fewer errors when occlusion is high. CoOp requires less movement and importantly this does not increase with distance, supporting precise selection at a distance where alternative techniques deteriorate due to user precision (e.g. hand shakiness) or sensor error (e.g. calibration). We also conclude that the techniques could be combined, for example using CoRa as default (for fast selection) with extension to CoOp depending on context (occlusion of objects along the ray). Combining CoOp with a ray-casting approach would enable simple selections to be performed quickly, whilst providing reduced movement and error rate for occluded targets. This hybrid approach could be accomplished using a more sophisticated ray-based approach in the candidate selection, for example using lead-by-default based on the ray-cast to initially select a target, for which the user could then override with the smooth pursuits technique. In this work, the disambiguation is performed before a selection, however an alternative future direction could be to use Outline Pursuits for correction instead of disambiguation. Users would then only have to perform the extra interaction steps when a critical error has been performed.

In comparison among hands-free techniques, HfOp induces less head movement compared with GaDw, but was less effective at low occlusion with less difference at higher occlusion. In contrast to GaDw, HfOp does not require line-of-sight to the target and allows users to focus on an object indefinitely. However, HfOp showed larger variance across participants due to some participants not reaching the required pursuit threshold for selection (a problem that CoOp avoids, see above). This suggests significant potential for improvement through personalisation, not only of HfOp but other pursuit selection techniques in the HCI literature.

Interview results from both studies showed no significant evidence that the added outlines or stimulus may cause strain or distract users. With HfOp, there is less control in the display of outlines as the head is generally in a similar direction as the eyes. This highlights another benefit with CoOp in that users have the freedom to decide when to display the outlines with the controller direction. In either case, minimising the amount of selection candidates minimises visual clutter.

Outline Pursuits adds elements (the outline and stimulus) to selectable targets but does not prescribe how these are presented. Our work explored different outline behaviour during occlusion but there are other aspects for designers to consider. Future work could further investigate additional aspects of the outlines such as the visual properties of the outline and stimuli target. Examples of this include blending the stimuli into the surrounding environment for less visual clutter, and to more subtly guide the user to available interactions. An additional benefit of Outline Pursuits is that selection feedback can easily be displayed without the need to change the object. In our work we only used colour changes as user feedback but additional ways of displaying feedback on the target, outline or the actual object itself are compelling to explore.

CONCLUSION
Outline Pursuits is novel in how it addresses challenges of object selection in occluded environments. The technique integrates concepts of cone-casting, outlining, and motion generation for selection by smooth pursuit that each contribute distinct features. Cone-casting enables pre-selection with any pointing modality, and is effective in identifying potential targets irrespective of how occluded they are. Outlining provides visual feedback on candidate objects, and depending on outlining strategy can also be used to reveal hidden shapes. Finally, motion generation around the outline makes the candidates selectable by gaze attention and eye movement correlation, without any change to the object itself.

We demonstrated Outline Pursuits in combination with head pointing for hands-free selection, and in combination with a controller for gaze-assisted manual input. In comparison with respective baselines, Outline Pursuits reduces movement users would otherwise perform to improve line of sight to targets. We also found that selection time and accuracy were less affected by higher occlusion.

REFERENCES

[1] F. Argelaguet and C. Andujar. 2009. Efficient 3D Pointing Selection in Cluttered Virtual Environments. *IEEE Computer Graphics and Applications* 29, 6 (Nov 2009), 34–43. DOI: http://dx.doi.org/10.1109/MCG.2009.117

[2] Ferran Argelaguet and Carlos Andujar. 2013. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics* 37, 3 (2013), 121–136. DOI: http://dx.doi.org/10.1016/j.cag.2012.12.003

[3] Marc Baloup, Thomas Pietrzak, and Géry Casiez. 2019. RayCursor: A 3D Pointing Facilitation Technique Based on Raycasting. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 101, 12 pages. DOI:http://dx.doi.org/10.1145/3290605.3300331

[4] Jonas Blattgerste, Patrick Renner, and Thies Pfeiffer. 2018. Advantages of Eye-gaze over Head-gaze-based Selection in Virtual and Augmented Reality Under Varying Field of Views. In *Proceedings of the Workshop on Communication by Gaze Interaction (COGAIN '18)*. ACM, New York, NY, USA, Article 1, 9 pages. DOI: http://dx.doi.org/10.1145/3206343.3206349

[5] Marcus Carter, Eduardo Velloso, John Downs, Abigail Sellen, Kenton O'Hara, and Frank Vetere. 2016. PathSync: Multi-User Gestural Interaction with Touchless Rhythmic Path Mimicry. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 3415–3427. DOI: http://dx.doi.org/10.1145/2858036.2858284

[6] Christopher Clarke, Alessio Bellino, Augusto Esteves, Eduardo Velloso, and Hans Gellersen. 2016. TraceMatch: A Computer Vision Technique for User Input by Tracing of Animated Controls. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 298–303. DOI: http://dx.doi.org/10.1145/2971648.2971714

[7] Nathan Cournia, John D. Smith, and Andrew T. Duchowski. 2003. Gaze- vs. Hand-based Pointing in Virtual Environments. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems (CHI EA '03)*. ACM, New York, NY, USA, 772–773. DOI: http://dx.doi.org/10.1145/765891.765982

[8] Gerwin de Haan, Michal Koutek, and Frits H. Post. 2005. IntenSelect: Using Dynamic Object Rating for Assisting 3D Object Selection. In *Proceedings of the 11th Eurographics Conference on Virtual Environments (EGVE'05)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 201–209. DOI: http://dx.doi.org/10.2312/EGVE/IPT_EGVE2005/201-209

[9] Heiko Drewes, Mohamed Khamis, and Florian Alt. 2018. Smooth Pursuit Target Speeds and Trajectories. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia (MUM 2018)*. ACM, New York, NY, USA, 139–146. DOI: http://dx.doi.org/10.1145/3282894.3282913

[10] Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015. Orbits: Gaze Interaction for Smart Watches Using Smooth Pursuit Eye Movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 457–466. DOI: http://dx.doi.org/10.1145/2807442.2807499

[11] Tovi Grossman and Ravin Balakrishnan. 2006. The Design and Evaluation of Selection Techniques for 3D Volumetric Displays. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology (UIST '06)*. ACM, New York, NY, USA, 3–12. DOI:http://dx.doi.org/10.1145/1166253.1166257

[12] John Paulin Hansen, Vijay Rajanna, I. Scott MacKenzie, and Per Bækgaard. 2018. A Fitts' Law Study of Click and Dwell Interaction by Gaze, Head and Mouse with a Head-mounted Display. In *Proceedings of the Workshop on Communication by Gaze Interaction (COGAIN '18)*. ACM, New York, NY, USA, Article 7, 5 pages. DOI: http://dx.doi.org/10.1145/3206343.3206344

[13] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-air Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1063–1072. DOI:http://dx.doi.org/10.1145/2556288.2557130

[14] Ken Hinckley, Randy Pausch, John C. Goble, and Neal F. Kassell. 1994. A Survey of Design Issues in Spatial Input. In *Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology (UIST '94)*. ACM, New York, NY, USA, 213–222. DOI: http://dx.doi.org/10.1145/192426.192501

[15] Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Jarodzka Halszka, and Joost van de Weijer. 2011. *Eye Tracking : A Comprehensive Guide to Methods and Measures*. Oxford University Press. 560 pages.

[16] Robert J. K. Jacob. 1990. What You Look at is What You Get: Eye Movement-based Interaction Techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '90)*. ACM, New York, NY, USA, 11–18. DOI: http://dx.doi.org/10.1145/97243.97246

[17] Florian Jungwirth, Michael Haslgrübler, and Alois Ferscha. 2018. Contour-guided Gaze Gestures: Using Object Contours As Visual Guidance for Triggering Interactions. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications (ETRA '18)*. ACM, New York, NY, USA, Article 28, 10 pages. DOI:http://dx.doi.org/10.1145/3204493.3204530

[18] Mohamed Khamis, Carl Oechsner, Florian Alt, and Andreas Bulling. 2018. VRpursuits: Interaction in Virtual Reality Using Smooth Pursuit Eye Movements. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces (AVI '18)*. ACM, New York, NY, USA, Article 18, 8 pages. DOI: http://dx.doi.org/10.1145/3206505.3206522

[19] Regis Kopper, Felipe Bacim, and Doug A. Bowman. 2011. Rapid and accurate 3D selection by progressive refinement. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 67–74. DOI: http://dx.doi.org/10.1109/3DUI.2011.5759219

[20] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 81, 14 pages. DOI:http://dx.doi.org/10.1145/3173574.3173655

[21] Diako Mardanbegi, Christopher Clarke, and Hans Gellersen. 2019. Monocular Gaze Depth Estimation Using the Vestibulo-ocular Reflex. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications (ETRA '19)*. ACM, New York, NY, USA, Article 20, 9 pages. DOI: http://dx.doi.org/10.1145/3314111.3319822

[22] Diako Mardanbegi, Dan Witzner Hansen, and Thomas Pederson. 2012. Eye-based Head Gestures. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12)*. ACM, New York, NY, USA, 139–146. DOI: http://dx.doi.org/10.1145/2168556.2168578

[23] Diako Mardanbegi, Tobias Langlotz, and Hans Gellersen. 2019. Resolving Target Ambiguity in 3D Gaze Interaction Through VOR Depth Estimation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 612, 12 pages. DOI: http://dx.doi.org/10.1145/3290605.3300842

[24] P. McCullagh and J. A. Nelder. 1989. *Generalized Linear Models*. Chapman and Hall/CRC. 532 pages.

[25] Pallavi Mohan, Wooi Boon Goh, Chi-Wing Fu, and Sai-Kit Yeung. 2018. DualGaze: Addressing the Midas Touch Problem in Gaze Mediated VR Interaction. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 79–84. DOI: http://dx.doi.org/10.1109/ISMAR-Adjunct.2018.00039

[26] Brad A. Myers, Rishi Bhatnagar, Jeffrey Nichols, Choon Hong Peck, Dave Kong, Robert Miller, and A. Chris Long. 2002. Interacting at a Distance: Measuring the Performance of Laser Pointers and Other Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*. ACM, New York, NY, USA, 33–40. DOI: http://dx.doi.org/10.1145/503376.503383

[27] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze + Pinch Interaction in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction (SUI '17)*. ACM, New York, NY, USA, 99–108. DOI: http://dx.doi.org/10.1145/3131277.3132180

[28] Ken Pfeuffer, Melodie Vidal, Jayson Turner, Andreas Bulling, and Hans Gellersen. 2013. Pursuit Calibration: Making Gaze Calibration Less Tedious and More Flexible. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 261–270. DOI: http://dx.doi.org/10.1145/2501988.2501998

[29] Jeffrey S. Pierce, Andrew S. Forsberg, Matthew J. Conway, Seung Hong, Robert C. Zeleznik, and Mark R. Mine. 1997. Image Plane Interaction Techniques in 3D Immersive Environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics (I3D '97)*. ACM, New York, NY, USA, 39–ff. DOI: http://dx.doi.org/10.1145/253284.253303

[30] T. Piumsomboon, G. Lee, R. W. Lindeman, and M. Billinghurst. 2017. Exploring natural eye-gaze-based interaction for immersive virtual reality. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 36–39. DOI:http://dx.doi.org/10.1109/3DUI.2017.7893315

[31] Yuan Yuan Qian and Robert J. Teather. 2017. The Eyes Don't Have It: An Empirical Comparison of Head-based and Eye-based Selection in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction (SUI '17)*. ACM, New York, NY, USA, 91–98. DOI:http://dx.doi.org/10.1145/3131277.3132182

[32] Vijay Rajanna and John Paulin Hansen. 2018. Gaze Typing in Virtual Reality: Impact of Keyboard Design, Selection Method, and Motion. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications (ETRA '18)*. ACM, New York, NY, USA, Article 15, 10 pages. DOI: http://dx.doi.org/10.1145/3204493.3204541

[33] Greg Schmidt, Yohan Baillot, Dennis G. Brown, Erik B. Tomlin, and J. Edward Swan. 2006. Toward Disambiguating Multiple Selections for Frustum-Based Pointing. In *3D User Interfaces (3DUI'06)*. IEEE, 87–94. DOI:http://dx.doi.org/10.1109/VR.2006.133

[34] Ludwig Sidenmark and Hans Gellersen. 2019a. Eye, Head and Torso Coordination During Gaze Shifts in Virtual Reality. *ACM Trans. Comput.-Hum. Interact.* 27, 1, Article 4 (Dec. 2019), 40 pages. DOI: http://dx.doi.org/10.1145/3361218

[35] Ludwig Sidenmark and Hans Gellersen. 2019b. Eye&Head: Synergetic Eye and Head Movement for Gaze Pointing and Selection. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19)*. ACM, New York, NY, USA, 1161–1174. DOI: http://dx.doi.org/10.1145/3332165.3347921

[36] Oleg Špakov and Päivi Majaranta. 2012. Enhanced Gaze Interaction Using Simple Head Gestures. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12)*. ACM, New York, NY, USA, 705–710. DOI:http://dx.doi.org/10.1145/2370216.2370369

[37] Frank Steinicke, Timo Ropinski, and Klaus Hinrichs. 2006. *Object Selection in Virtual Environments Using an Improved Virtual Pointer Metaphor*. Springer Netherlands, Dordrecht, 320–326. DOI: http://dx.doi.org/10.1007/1-4020-4179-9_46

[38] Sophie Stellmach and Raimund Dachselt. 2012. Look & Touch: Gaze-supported Target Acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 2981–2990. DOI: http://dx.doi.org/10.1145/2207676.2208709

[39] Vildan Tanriverdi and Robert J. K. Jacob. 2000. Interacting with Eye Movements in Virtual Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '00)*. ACM, New York, NY, USA, 265–272. DOI: http://dx.doi.org/10.1145/332040.332443

[40] Eduardo Velloso, Marcus Carter, Joshua Newn, Augusto Esteves, Christopher Clarke, and Hans Gellersen. 2017. Motion Correlation: Selecting Objects by Matching Their Movement. *ACM Trans. Comput.-Hum. Interact.* 24, 3, Article 22 (April 2017), 35 pages. DOI: http://dx.doi.org/10.1145/3064937

[41] Eduardo Velloso, Flavio Luiz Coutinho, Andrew Kurauchi, and Carlos H. Morimoto. 2018. Circular Orbits Detection for Gaze Interaction Using 2D Correlation and Profile Matching Algorithms. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications (ETRA '18)*. ACM, New York, NY, USA, Article 25, 9 pages. DOI: http://dx.doi.org/10.1145/3204493.3204524

[42] Eduardo Velloso, Markus Wirth, Christian Weichel, Augusto Esteves, and Hans Gellersen. 2016. AmbiGaze: Direct Control of Ambient Devices by Gaze. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. ACM, New York, NY, USA, 812–817. DOI: http://dx.doi.org/10.1145/2901790.2901867

[43] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)*. ACM, New York, NY, USA, 439–448. DOI: http://dx.doi.org/10.1145/2493432.2493477

[44] Wang Ao-yu, Tang Min, and Dong Jin-xiang. 2004. A survey of silhouette detection techniques for non-photorealistic rendering. In *Third International Conference on Image and Graphics (ICIG'04)*. 434–437. DOI:http://dx.doi.org/10.1109/ICIG.2004.28

[45] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 143–146. DOI: http://dx.doi.org/10.1145/1978942.1978963

[46] Shumin Zhai, William Buxton, and Paul Milgram. 1994. The "Silk Cursor": Investigating Transparency for 3D Target Acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94)*. ACM, New York, NY, USA, 459–464. DOI: http://dx.doi.org/10.1145/191666.191822

[47] Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA, 246–253. DOI: http://dx.doi.org/10.1145/302979.303053

[48] Yuhang Zhao, Edward Cutrell, Christian Holz, Meredith Ringel Morris, Eyal Ofek, and Andrew D. Wilson. 2019. SeeingVR: A Set of Tools to Make Virtual Reality More Accessible to People with Low Vision. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 111, 14 pages. DOI: http://dx.doi.org/10.1145/3290605.3300341