Still Looking: Investigating Seamless Gaze-supported Selection, Positioning, and Manipulation of Distant Targets

Sophie Stellmach and Raimund Dachselt
Interactive Media Lab
Technische Universität Dresden
Dresden, Germany
{stellmach, dachselt}@acm.org

ABSTRACT
We investigate how to seamlessly bridge the gap between users and distant displays for basic interaction tasks, such as object selection and manipulation. For this, we take advantage of very fast and implicit, yet imprecise gaze- and head-directed input in combination with ubiquitous smartphones for additional manual touch control. We have carefully elaborated two novel and consistent sets of gaze-supported interaction techniques based on touch-enhanced gaze pointers and local magnification lenses. These conflict-free sets allow for fluently selecting and positioning distant targets. Both sets were evaluated in a user study with 16 participants. Overall, users were fastest with a touch-enhanced gaze pointer for selecting and positioning an object after some training. While the positive user feedback for both sets suggests that our proposed gaze- and head-directed interaction techniques are suitable for a convenient and fluent selection and manipulation of distant targets, further improvements are necessary for more precise cursor control.

Author Keywords
Eye tracking; gaze interaction; visual attention; mobile touch input; distant displays; selection; positioning

ACM Classification Keywords
H.5.2 Information interfaces and presentation: User Interfaces: Input devices and strategies

General Terms
Design, Human Factors

INTRODUCTION
In our daily lives, we are surrounded by a growing diversity of display setups, for example, multiple screens at work, large-sized TV sets at home or wall-sized information displays. Regardless of the particular display configuration, fundamental interaction tasks include the selection, positioning, and manipulation of displayed content. One interesting way for seamlessly interacting with such diverse displays is a multimodal combination of a user's gaze as an implicit and coarse pointing modality and ubiquitous smartphones for more explicit and manual fine adjustments [16, 17].

The ongoing developments in mobile eye tracking systems will soon allow for pervasive and unobtrusive gaze interaction in everyday contexts [7]. As frequently pointed out, a user's gaze may serve as a beneficial pointing modality (e.g., [4, 16, 20]). It is fast, it requires low effort, and it is very implicit as a person's eye gaze reaches a target prior to a manual pointer [4]. On the downside, several challenges have to be considered for convenient gaze-based interaction that are summarized in the following:

Inaccuracy: Based on the physiological nature of our eyes and the limitation of eye tracking systems, gaze data is inherently inaccurate. While coarse gaze pointing is practical for selecting large graphical items or roughly indicating a user's region of interest, several approaches for overcoming inherent eye tracking inaccuracies exist [16]. This includes local magnifications of small items or multimodal combinations to enhance pointing precision [16, 23]. Several works investigate gaze-based object selection (e.g., [2, 16]). However, further investigations are required about the fluent execution of a series of interaction tasks.

Double role: If our eye gaze is used for interaction, it assumes a double role for visual observation and control. As a result, we have to carefully consider the nature of interaction tasks for convenient and non-distracting gaze-based controls. As an alternative to gaze input, a user's head direction can be used to roughly indicate a user's region of interest and to resolve this double role (e.g., [11]). However, head-directed pointing is less implicit and requires higher physical effort.

Midas Touch: Since our gaze is an *always-on device* [10], mechanism have to be provided to prevent users to unintentionally issue an action. This can be accomplished by gaze dwelling at a certain location, which however impedes fast and continuous interaction. Alternatively, this issue can be addressed by a multimodal combination of implicit gaze input with explicit manual controls to confirm a user's intention.

In this paper, we investigate how a user's gaze or more roughly head direction may aid in *selecting*, *positioning*, and *manipulating* (e.g., rotating or scaling) an item of interest. On the one hand, we take advantage of fast and implicit gaze input for coarse pointing. On the other hand, we use local target expansions and manual touch input for precise and reliable gaze-supported pointing. For this, we utilize ubiquitous smartphones as flexible and popular interactive multi-touch

CHI'13, April 27–May 2, 2013, Paris, France.

devices. We contribute two practical and consistent sets of gaze-supported interaction techniques (see [16, 17] for an introduction to this style of interaction) that allow for fluent *coarse and fine* object *selection and positioning* of distant targets. The elaborated sets have been evaluated in a user study with 16 participants to help us assess the benefits and limitations of the proposed multimodal techniques. Both sets were very positively assessed by participants, although further improvements are required for more fine-grained object selection and positioning.

The remaining paper is structured as follows: First, we give a brief overview of *Related Work* about mobile multi-touch devices and multimodal gaze-based techniques for selecting and manipulating distant targets. Subsequently, we describe several *Design Considerations* that we took into account for the development of our gaze-supported interaction techniques. In the subsequent section, we specify the two sets of gaze-supported interaction techniques that we have developed and report on their evaluation in the *User Study* section. The paper concludes with a *Discussion* of results and future work.

RELATED WORK

A comprehensive body of work about multimodal interaction with distant displays exists (e.g., [3, 5, 15, 16, 19]). With respect to our intended input combination, we concentrate on the use of mobile multi-touch interfaces and multimodal gaze input for interacting from a distance in the following.

Distant Interaction with Mobile Multi-touch Interfaces

Nancel et al. [15] compare different input modalities for pan-and-zoom operations in a large high-resolution multi-display environment. They show that simple linear touch gestures on a mobile multi-touch device are faster and more reliable compared to freehand gestures. Keefe et al. [12] also use a mobile multi-touch interface in a similar display context for selecting, querying, and visually exploring data visualizations. They highlight the advantage to quickly and accurately select items from various positions, which is vital for observing details. For performing selections they propose an extended gesture set mainly based on flicking touch gestures on the handheld device. Han et al. [9] use multiple handhelds (one in each hand) for remote interaction with large vertical displays. For target selection and translation they use one-handed controls; for rotation and scaling two-handed input, e.g., two-handed pinch gesture (in the air).

Boring et al. [6] present Touch Projector, which enables users to select and position content shown on a distant display through touch input on live video on a smartphone. While this works well for large target sizes on the mobile screen, errors occur due to hand jitter and imprecise touch input, which both hinder fine precision tasks. To provide more stability, Boring et al. [6] propose a combination of zooming in on a target and freezing the video. With respect to gaze-based input, in [16] we propose a similar approach for our so-called semi-fixed gaze-directed local zoom lens that works in combination with a handheld smartphone to interact with distant displays. The lens does not move while the user looks within the magnified region (frozen lens). The user can zoom in further by simply sliding upwards on the mobile touch screen to perform more fine-grained selections.

All these works demonstrate the high flexibility for using mobile multi-touch devices to interact with distant displays. However, investigations about a fluent transition between selecting, positioning, and manipulating targets with respect to a user's visual attention are lacking.

Gaze-supported Selection and Manipulation

Already in the early 1980's, Bolt has envisioned how we could benefit from eye gaze as an additional input modality for fast and natural interactions [5] with large distant displays. Since then, several works have investigated how to facilitate the interaction with distant displays by integrating gaze data with additional modalities, such as keyboard controls (e.g., [13, 14, 20]), mouse input [8, 23], hand gestures (e.g., [22]), or touch input (e.g., [16, 17, 18, 21]). Most of these works investigate how to speed up the selection of graphical items at a distant display. One prominent idea for this is to warp the cursor to the vicinity of the user's point-of-regard and to make manual fine adjustments (by moving the mouse) from there [23]. This idea has been advanced to better indicate when the cursor is supposed to follow the user's gaze using a touch-sensitive mouse [8] and a mobile touch screen [16].

In this context, in previous work we have investigated a combination of gaze and touch input for distant object selection and data exploration [16, 17]. While these investigations do not take a broader interaction process into account (e.g., from selecting, positioning, and manipulating a target), they provide interesting insights and a good foundation for the design of more advanced gaze-supported interaction. Turner et al. [18] also investigate a combination of gaze and touch input for target selection *and* positioning. They distinguish between distinctly selecting and positioning a target (*Eye Cut & Paste*) and seamlessly combining both (*Eye Drag & Drop*). For the latter, a user looks at a target, begins touching a multi-touch surface, looks at a desired destination and releases the touch. Turner et al. do not evaluate the proposed techniques or discuss how these could be applied for the selection of small or closely positioned targets.

Instead of using gaze as a direct pointing modality, Kaiser et al. [11] use a user's head direction to indicate the (approximate) region of interest. This is used in combination with multimodal controls, including speech and finger tracking to allow for decreasing the ambiguity for object selection in a virtual 3D scene. Argelaguet & Andujar [1] also propose combining pointing rays from a user's eyes and hand to select items in cluttered virtual environments more efficiently.

In a nutshell, while several works have dealt with gaze/head input as pointing modality, only little work has investigated how to use it for a fluent combination of interaction tasks.

DESIGN CONSIDERATIONS

In the following, we present key design issues that we have taken into account for our gaze-supported interaction techniques. Some of these considerations are based on our prior

work on gaze-supported interaction with distant displays [16, 17]. In addition, although not considering gaze input, Bezerianos & Balakrishnan [3] describe several design goals for object selection and manipulation in distant locations. This includes, e.g., consistency and minimized physical movement. In the scope of this paper, we focus on single users to build a solid foundation for multimodal gaze-supported interaction before targeting multi-user or collaborative applications.

Interaction Tasks: The intended gaze-supported interaction techniques shall allow for selecting, positioning and manipulating graphical objects shown at a distant display. For this, we aim for convenient selection of large and small or closely positioned objects (i.e., *coarse* vs. *fine* selection). In addition, we distinguish between *coarse* and *fine* target positioning (also see [19]). While roughly moving an item is sufficient to make room for other objects, more precise input is required for exact graphical layouts. Finally, while we mainly focus on object selection and positioning in the scope of this paper, we take additional interaction tasks into account. For this, we consider basic manipulation tasks, such as scaling and rotating a selected target.

Input Combination: On the one hand, we benefit from gaze for direct, fast and coarse pointing. On the other hand, the interaction with a smartphone serves for indirect and more fine-grained input. With this, we follow up on the promising principle *"gaze suggests and touch confirms"* [16].

Eyes-free Interaction: A user's main visual attention should remain at the distant display and not the mobile screen. Thus, it is vital that the smartphone can be controlled without having to look at it. This means, for example, that simple touch events and relative touch gestures should be favored over additional virtual buttons.

Low Effort: The more frequent an action has to be performed, the easier and quicker the interaction should be. For example, to confirm a selection, a tap on the touch screen should be preferred over a complex multi-touch gesture.

Consistency and Seamlessness: A consistent and seamless combination of interaction techniques supports users in performing tasks in an effective and fluent way. This becomes especially vital when taking a longer chain of interaction tasks into account (e.g., selection, positioning, and manipulation). In addition, we need to consider that in some cases users may only want to select a target, but may not want to immediately reposition it. Thus, we distinguish between a *distinct* and *seamless* combination of selection and positioning.

DESIGN OF INTERACTION TECHNIQUES

For the design of gaze-supported interaction techniques, we follow two parallel lines of development differing in how to overcome eye tracking inaccuracies: (1) a touch-enhanced gaze pointer and (2) a gaze-directed zoom lens. With this, we build on our promising gaze-supported target acquisition techniques that we have proposed in [16], which also use a combination of gaze and touch input. However, we extend these investigations by addressing how to fluently select *and* position differently sized objects (from a distance) considering a user's gaze/head direction. For this, we have elaborated two conflict-free sets of interaction techniques combining gaze/head and touch input. We discuss benefits and limitations of gaze- and head-directed input for object selection and for directly controlling a target (for repositioning). As a minor aspect, we also consider how manipulation tasks, such as object scaling and rotation, can be seamlessly integrated. In the following, we first introduce the basic principles of the *touch-enhanced gaze pointer* and *gaze-directed zoom lens* and then describe how the specified interaction tasks can be performed with them.

Touch-enhanced Gaze Pointer

One way to address inaccurate gaze data is to allow for manual fine adjustments of the coarse gaze cursor [8, 16, 23]. Zhai et al. [23] introduce the *Manual And Gaze Input Cascaded* (MAGIC) pointing technique to speed up object selections: The cursor is warped to the vicinity of a currently looked at target and can be repositioned via mouse or touchpad controls. Based on this, our *MAGIC touch* technique [16] focuses on object selections using a combination of gaze data and touch input from a handheld smartphone. To extend on this, we propose an advanced *touch-enhanced gaze pointer* (**TouchGP**) that is applicable for gaze-supported object selection *and* positioning. For this, a user's gaze information serves for roughly indicating an area of interest (*coarse* selection and positioning) and the touch input for more precise adjustments (*fine-grained* selection and positioning). While the cursor movement for the original *MAGIC touch* is limited to a certain boundary around the current gaze point, we lift this restriction to provide higher flexibility. If the user touches the mobile display, the cursor on the distant screen stops following the user's gaze and instead follows the relative touch movement on the mobile display. For this, the initial touch position serves as a reference location from which movements can be performed in all directions. The mapping of touch movements to the movement of the distant cursor can be adapted to different levels of precision. For example, a touch movement of 50 pixels on the mobile display could be mapped to only 5 pixels for very precise cursor positioning.

Gaze-directed Zoom Lens

A common approach for overcoming inherent eye tracking inaccuracies is the use of local magnification lenses to increase apparent target sizes [2, 16]. While an *eye-slaved* zoom lens, a lens that immediately follows a user's gaze, is beneficial for easy gaze-based target selections, it is also more visually distracting and error-prone as the lens is always moving according to the *imprecise* gaze cursor [2, 16]. A *semi-fixed* zoom lens [16], that gradually starts moving towards user's current point-of-regard when looking outside the lens' boundary, provides a higher stability. However, users have mentioned that it is distracting to have to look away from the actual object-of-interest to move the lens.

We propose a hybrid type which we call *Gaze-directed Zoom Lens* that takes advantage of the quickness of the *eye-slaved* zoom lens and the higher stability of the *semi-fixed* one. For this, we distinguish a three-zone design for the zoom lens (also see Figure 1), for which we assume that the cursor is directly mapped to the current gaze position (i.e., gaze-directed

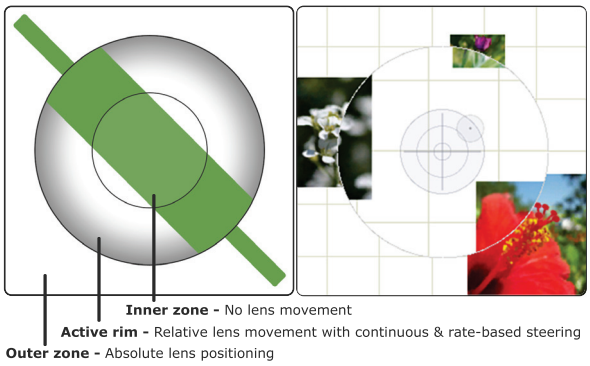

Figure 1. The enhanced semi-fixed zoom lens follows a three-zone design (left). The right image is an actual screenshot from our prototype. The size and magnification level can be adapted via touch controls. An additional crosshair is displayed if a target has been selected (as shown in the right image).

cursor): (1) *Inner zone* – While looking within this zone (i.e., moving the gaze-directed cursor in this region), the lens is not moving. (2) *Active rim* – The further a user looks away from the inner zone towards the outer zone, the faster the lens will follow the user's gaze (*semi-fixed lens behavior*: relative gradient-based lens positioning). (3) *Outer zone* – If a user looks outside the lens, the lens center will be immediately set to the current gaze position (*eye-slaved lens behavior*: absolute lens positioning).

In contrast to the semi-fixed zoom lens, users do not need to look outside the lens to move it and can still benefit from an increased stability if looking close to the lens center. For our prototype, the radius of the *Inner zone* is always half the size of the *Active rim*. Thus, if increasing the *Active rim*/lens size, the *Inner zone* automatically adapts in size as well. The lens can be controlled more smoothly and precisely with a larger *Active rim* as the gradient-based speed values are distributed across a larger region (cf. Figure 1, left). The lens size can be adapted using a horizontal sliding gesture, the zoom level using a vertical one (also see next section and see Figure 2). Finally, the lens is always on and thus no additional input is required for activation. However, the lens is designed in a subtle way to limit visually distracting the user.

Pre-Study: From Gaze- to Head-directed Zoom Lens

We conducted a preliminary interview with three participants to get first insights about the suitability of the gaze-directed zoom lens for selecting and positioning objects. While participants mentioned that the lens has potential for easing gaze-based selections, it was described as awkward and even impractical for positioning a target. This was due to the circumstance that it was very difficult to position an object, while having to look away from it to move the lens. Sometimes a user merely wanted to check whether an object was aligned correctly at all sides and would then accidentally reposition it. This leads back to a challenge of the double role of our eye gaze for convenient gaze-based interaction. Nevertheless, local magnification lenses that are based on a user's visual attention are promising as frequently pointed out [2, 13, 16]. As a conclusion, we decided to refrain from using gaze input to steer the lens, but instead take a user's head direction into account to still take advantage of the rough estimation of a user's visual field. Thus, a user can move the lens by turning his/her head towards a region of interest without any gaze input. In the remaining paper, we refer to this as *Head-directed Zoom Lens* (**HdLens**) instead of *Gaze-directed Zoom Lens*.

Multimodal Gaze-supported Interaction Techniques

We elaborated a set of interaction techniques for **TouchGP** and for **HdLens** that allow for fluently transitioning between roughly and precisely selecting and positioning objects displayed at a distant screen. As a minor aspect, we also consider how manipulation tasks, such as scaling and rotating a target, can be seamlessly appended. For these techniques we assume a single user standing or sitting in front of a large (distant) screen and holding a smartphone in his/her hand. We distinguish two main approaches: The *seamless* combination of selecting and positioning a target (i.e., drag-and-drop) and the dissociation of them in *distinct* phases. This means that a user can first select a target, then explore the displayed content further (e.g., to select additional targets), and finally position the selected items. However, in the context of this paper, we will focus on single targets for now. An overview of the elaborated interaction techniques is illustrated in Figure 2. In the following, these two approaches (*distinct* vs. *seamless*) for object selection and positioning are described in more detail for both **TouchGP** and **HdLens**.

Distinct Selection

For both **TouchGP** and **HdLens**, large targets can be simply selected (*coarse selection*) by looking/turning towards them and briefly tapping the touch screen (single tap). To select small items (*fine selection*), the techniques differ for **TouchGP** and **HdLens**. For **TouchGP**, the user looks at the intended target and slides the finger on the smartphone for finer cursor positioning. For this, the cursor stops following a user's gaze once he/she touches the mobile screen. For **HdLens**, small targets can be selected by increasing the lens' magnification level using a simple upwards sliding gesture on the mobile screen (downwards to zoom out). Once an object is selected, a crosshair is displayed at the center of the lens as a positioning aid (also see Figure 1, right).

Distinct Positioning

After an object has been selected, the user can directly position it to the current gaze location (*coarse positioning*) by briefly tapping the touch screen twice (double tap). To position it more precisely (*fine positioning*), techniques differ for **TouchGP** and **HdLens** again: For **TouchGP**, if a user touches the mobile screen, a preview of the currently selected target is shown at the gaze location and the (distant) cursor stops following the user's gaze. From this location, the object can be manually repositioned more precisely via touch input (slide gesture). For fine positioning with **HdLens**, first the lens can be further zoomed in with an upward touch sliding gesture. To not distract the user while looking around, a preview only appears after touching the mobile screen for a longer time (hold). The target is then attached to the lens center, as this provides a higher stability and control compared to the imprecise head-directed cursor. The lens and attached

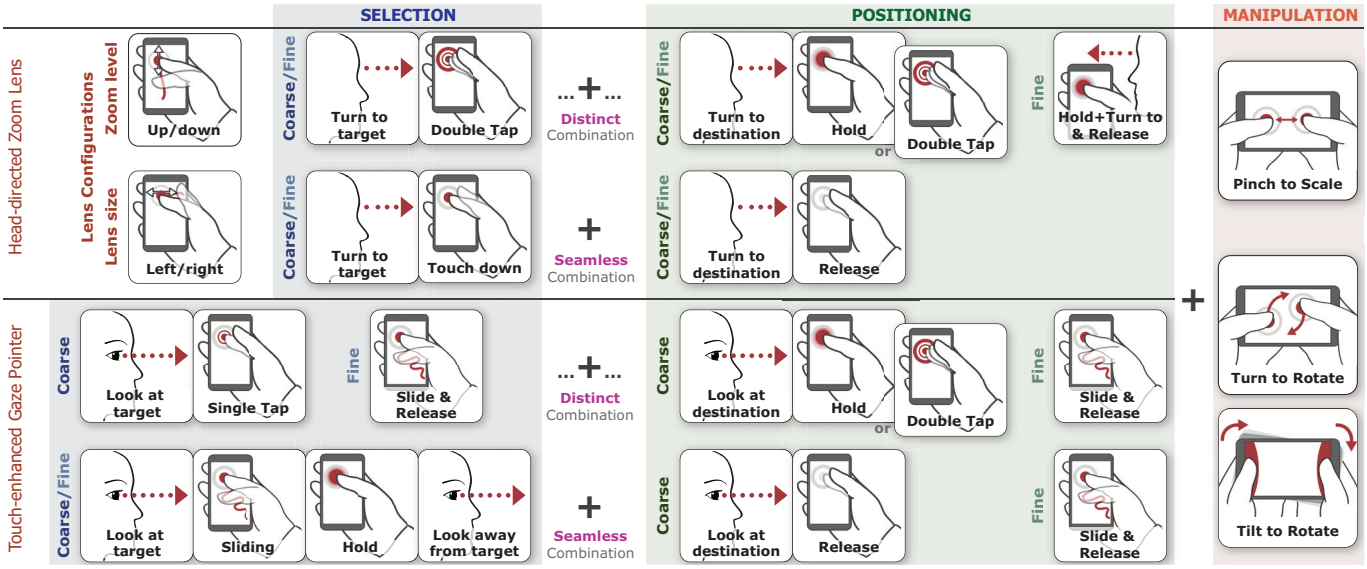

Figure 2. Overview of our proposed gaze-/head-supported interaction techniques for selecting, positioning, and manipulating an object.

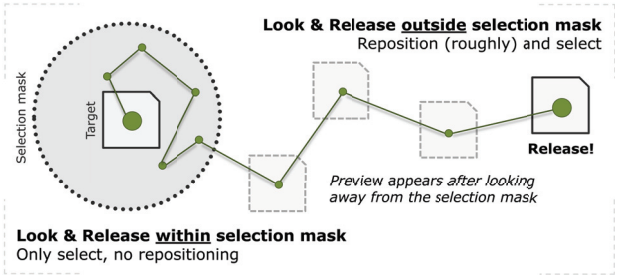

Figure 3. Example of gaze-/head-directed drag-and-drop.

object can be positioned more precisely by moving the head-directed cursor into the *Active rim* or *Outer zone* of the lens (cf. Figure 1, left). Once the user releases the mobile touch screen, the preview disappears, and the object is set to the current lens center. Alternatively, a user can reposition the lens via head-directed input first (without touching the mobile device) and then perform a double tap to warp the object directly to the lens center.

Seamless Selection & Positioning

In the following, we describe how users can *seamlessly* select and position an object using **TouchGP** and **HdLens**. For this, we propose a gaze-/head-directed drag-and-drop approach that is illustrated in Figure 3. The user can simply look at a target and begin touching the mobile screen to mark it for selection (*coarse selection*). For more *precise selections*, the cursor can be repositioned by either moving the finger on the touch screen (**TouchGP**) or performing a vertical slide gesture to zoom in further (**HdLens**). As soon as the mobile screen is touched, the cursor stops following the user's gaze and a *selection mask* is displayed, as illustrated in Figure 3. If releasing the touch while still looking within the boundaries of this selection mask, the marked target is selected but not repositioned. This prevents unintentionally moving an object due to jittery eye movements. If the user looks beyond the selection mask's boundaries and still holds on to the touch screen, a smooth transition from target selection to positioning is achieved (*drag-and-drop*). After crossing the selection mask's boundary, a preview appears and starts following the user's gaze. In case of **HdLens**, a crosshair and a preview of the marked target is displayed at the lens center. The user can simply release the touch for *coarse object positioning* or can continue touching the device for more precise object positioning. As before, *fine positioning* is either achieved moving the finger on the mobile screen (**TouchGP**) or performing a vertical sliding gesture to zoom in further (**HdLens**).

Manipulation Mode

As previously pointed out, we also considered how additional interaction tasks, such as scaling and rotating an object, could be seamlessly combined with the proposed sets of interaction techniques. For simple target manipulations, we wanted to reuse some low-effort touch gestures, such as a one-finger sliding gesture. For this, it is however necessary to distinguish different interaction modes, in our case a *selection & positioning* and a *manipulation* mode. In the manipulation mode, the user can position, rotate, and scale an already selected item. To reach the manipulation mode, we aimed for a quick and reliable mode change that does not require the user to look at the smartphone. Considering common multi-touch manipulation techniques, such as pinch-to-zoom (i.e., moving two fingers together/apart to zoom in/out), we decided that the user can simply flip the smartphone to landscape mode and hold the device in two hands to allow for a *two-thumbs* interaction (see Figure 2, lower right). The shape of the cursor at the distant display will change to provide immediate feedback about the mode change (which again supports eyes-free interaction with the handheld device).

A two-thumbs pinch gesture is used to scale a selected target. To rotate it, the user can perform a two-thumbs turn-to-rotate gesture (i.e., the touch points move into the same clockwise or counterclockwise direction – see Figure 2, right). Alterna-

tively, the user can also tilt the smartphone to rotate an object. For this, the user has to hold the smartphone parallel to the distant screen and can mimic the current orientation of the target. Then, the user needs to touch two active regions at the left and right side of the mobile screen to indicate that the target should start imitating the smartphone's orientation. Thus, this follows the metaphor of holding a large picture at the sides with both hands and rotating it to align it.

USER STUDY

To further investigate the elaborated sets of interaction techniques for **TouchGP** and **HdLens**, we conducted a user study (within-subjects design). Since a seamless combination of gaze-supported object selection and positioning has not been investigated before, we were particularly interested in finding out more about the practicality and suitability of the proposed techniques for these interaction tasks. For this, we varied target sizes and respective destination areas to investigate how the individual techniques would be suited for fine and coarse object selection and positioning. Our main interest was to receive valuable user feedback (both qualitative and quantitative) to gain further insights into gaze-supported interaction. Thus, in this study, we did not aim at beating a given baseline, but instead wanted to find out how people cope with the developed techniques for achieving a given task.

Apparatus. For gathering gaze data we used a head-mounted ViewPoint PC-60[1] eye tracker from Arrington Research. It allows for tracking both eyes (binocular) with an accuracy of about 0.25°-1.0° visual angle, however, without the compensation of head shifts. To track a user's head movements, we used a ceiling-mounted visual tracking system, the OptiTrack V100:R2 IR setup[2], with six cameras at 50 Hz with a 640x480 pixel resolution. For this, IR markers have been attached to the head-mounted eye tracker frame. The gathered gaze/head data was stabilized [16]. For the smartphone interaction, we used a Samsung Galaxy SIII GT-I9300[3] deploying Android 4.0.4. It has a 4.8" display with a resolution of 1280x720. For moving the distant cursor via touch input, we used a 1:1 mapping. Thus, if the touch point moves 50 pixels on the smartphone screen, the distant cursor will also move 50 pixels. Data from the ViewPoint system, OptiTrack server, and smartphone were sent via TCP/IP and an adapted VRPN interface to the application computer, which runs Microsoft's XNA Game Studio 4.0 (based on C#) for the creation of our virtual test environment. Finally, the distant screen projection was 1.5 m wide and 2.5 m high. Participants were seated about 2.2 m away from the screen. An overview of the setup used in this user study is shown in Figure 4.

Participants. Sixteen volunteers (8 male, 8 female) participated in our user study ranging in age from 22 to 33 (Mean (M) = 27.6). None of the participants wore glasses, as we encountered problems with the deployed eye tracker due to reflections. If necessary, participants used contact lenses to correct their vision. All participants indicated that they use computers on a daily basis. Nine participants have used eye trackers before, but only once or twice. Twelve participants own a smartphone and five a multi-touch tablet.

[1]http://www.arringtonresearch.com/scene.html
[2]http://www.naturalpoint.com/optitrack/products/v100-r2/
[3]http://www.samsung.com/global/galaxys3/specifications.html

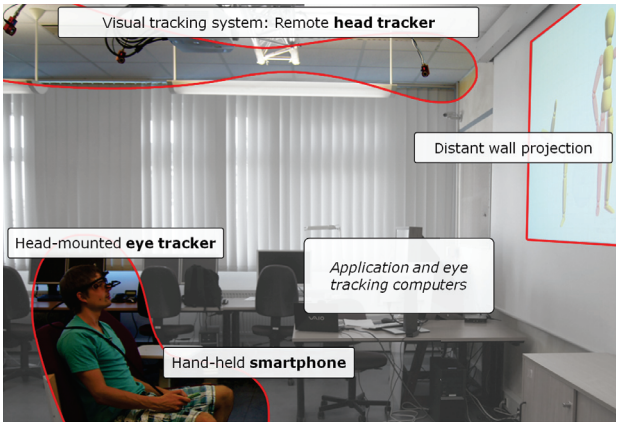

Figure 4. Overview of the hardware setup in the user study.

Procedure. After welcoming participants, a demographic questionnaire was handed out. Participants were seated in front of the projection screen and were instructed to sit fairly still, but without physically restricting their movement. We counter-balanced the order in which the input conditions (**TouchGP** and **HdLens**) were tested.

For **HdLens**, the OptiTrack system has been calibrated beforehand to ensure that the markers that have been attached to the eye tracker frame could be properly tracked. After participants put on the head-mounted gear, we asked them to sit in a comfortable position and to look at the center of the distant screen. This head position and orientation were taken as a reference to account for different user heights. For **TouchGP** we additionally performed a 16-point eye tracker calibration.

After ensuring that the tracking of gaze and head data worked as anticipated, we followed the same procedure for each technique. One input condition (i.e., **TouchGP** or **HdLens**) was tested at a time. First, we explained the interaction principles for **TouchGP/HdLens**. Participants could directly get acquainted with them in a demo application as long as they felt necessary. While this training phase usually did not take longer than 5-10 minutes, we noticed that in particular those users with little experience with smartphones took considerably longer to get used to the indirect touch interaction, as they usually performed rather coarse touch movements which led to cursor overshooting beyond a targeted location. The demo application showed parts of a 3D model that had to be assembled (see Figure 4, right). After the training phase, we proceeded with *Task One* and *Task Two* (cf. Figure 5) that are described in the following.

In *Task One*, we wanted to investigate the suitability of our techniques for *fine & coarse selection & positioning*. For this, a single target had to be selected and positioned as fast as possible. One target at a time appeared at a random screen corner but always with the same distance to the screen center. Targets always had to be positioned in or towards the opposite diagonal corner. Target sizes and corresponding destination sizes were varied (see Figure 5 for an overview). Target sizes

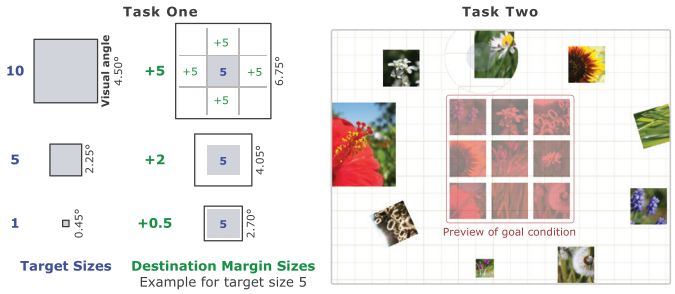

Figure 5. Overview of the two main task scenarios. In *Task One*, target and destination sizes are varied to investigate the suitability of TouchGP and HdLens for fine and coarse selection and positioning. *Task Two* is an open task in which users were asked to select, position, scale, and rotate nine images. For this, the entire set of techniques for TouchGP or HdLens had to be used in concert.

differed from 10 (large) to 5 and 1 (small) based on an internal unit. With respect to our particular setup, the sizes ranged from 4.5° (size 10) to 0.45° (size 1) of visual angle. The size of the destination area was the sum of the current target size plus an additional margin to distinguish between coarse and fine object positioning (see Figure 5 for an example for target size 5). These destination margins ranged from +5 (large), +2 to +0.5 (small). Finally, to add some variation to the tasks, we used two different distances between target and destination. At the beginning of each run, participants had to look at the center of the screen and touch the mobile device to confirm readiness (a circle displayed at the screen's center turned green if looking at it to give a better feedback). This was meant to improve comparability between the task completion times. The described procedure for *Task One* was carried out twice, leading to 36 tasks to complete: three target sizes x three destination margin sizes x two distances x two runs.

In *Task Two*, we added an informal part (no logging of task completion times) in which we wanted to let users freely *play around* with the entire set of techniques for **TouchGP** and **HdLens** to select, position, and manipulate several objects after each other. Here, we wanted to investigate how users would assess the usability of the different selection and positioning techniques in concert. After describing the manipulation techniques to the users, they had to move, scale and/or rotate nine images to match nine preview images shown at the center of the screen (cf. Figure 5, right).

Measures. Since we focused on substantial user feedback about the developed interaction techniques, we handed out several questionnaires at different stages during the study: an initial demographic, two intermediate, and a final questionnaire. The intermediate and final questionnaires contained both quantitative and qualitative questions. All quantitative questions were based on 5-point Likert scales from 1 - *Strongly disagree* to 5 - *Strongly agree*.

The intermediate questionnaire consisted of four parts and was handed out after completing both *Task One* and *Task Two* with a given set of techniques. In the first two parts, users had to rate eight general usability statements for *object selection* (IQ1) and also for *object positioning* (IQ2) (see Figure 7). In the third part (IQ3), we asked individual questions about particular aspects of **TouchGP** (10 questions) and **HdLens** (14 questions). Finally, participants were asked for qualitative feedback about what they particularly liked and disliked about **TouchGP/HdLens** (IQ4).

	1st run	2nd run
TouchGP	$F_T(2,45)=3.38$, **$p<.05$** $F_D(2,45)=3.55$, $p=.06$	$F_T(2,45)=19.65$, **$p<.001$** $F_D(2,45)=6.86$, **$p<.05$**
HdLens	$F_T(2,45)=5.03$, **$p<.05$** $F_D(2,45)=2.71$, $p=.08$	$F_T(2,45)=9.92$, **$p<.001$** $F_D(2,45)=7.85$, **$p=.001$**

Table 1. Overview of the influence of target sizes (T) and destination sizes (D) based on the overall task completion times for a respective input condition within a run. Significant results are printed in bold.

In the *final questionnaire*, we asked six questions about how users liked the manipulation techniques (e.g., the type of scaling and rotation). In addition, we asked participants for a final assessment of and for concluding comments about **TouchGP** and **HdLens**. On average, each participant took about 120 minutes for the described procedure.

RESULTS

For the evaluation of **TouchGP** and **HdLens**, we were particularly interested in user feedback about their assessed usability, usefulness and possible improvements. However, we start the evaluation by briefly assessing the basic practicality of the techniques for coarse and fine object selection and positioning. For this, we evaluate logged data for the time it took users to successfully select and position a target. Please note that only *Task One* was time-measured and that this task was executed twice (run 1 and 2) for each input condition. To investigate statistically significant differences in the logged data, we used a repeated-measures ANOVA (Greenhouse-Geisser corrected) with post-hoc sample t-tests (Bonferroni corrected).

Task Completion Times

An overview of the average times it took users to *select* an appointed target and *position* it within a designated destination area (for *Task One*) is listed in Figure 6. First, we consider how the individual input conditions, **TouchGP** and **HdLens**, are suited for coarse and fine object selection based on varied target sizes. Subsequently, we address their suitability for coarse and fine object positioning based on varied destination margin sizes.

Target Sizes. Although all users were able to select even small targets (size 1) with both input conditions, users needed significantly longer than for larger targets (see Table 1 for statistical results). No significant differences exist for targets of size 5 and 10. Users could select targets significantly faster with **TouchGP** than with **HdLens** for both runs ($F(3,135)=25.11$, $p<.05$). While participants could significantly improve their performance with **TouchGP** between the two runs ($p<.05$), no significant learning improvements exist for **HdLens** in this respect.

Destination Sizes. Users were able to position all given targets in the designated destination areas. However, users were slower with both input conditions for precise target positioning (i.e., destination margin size = +0.5) than for

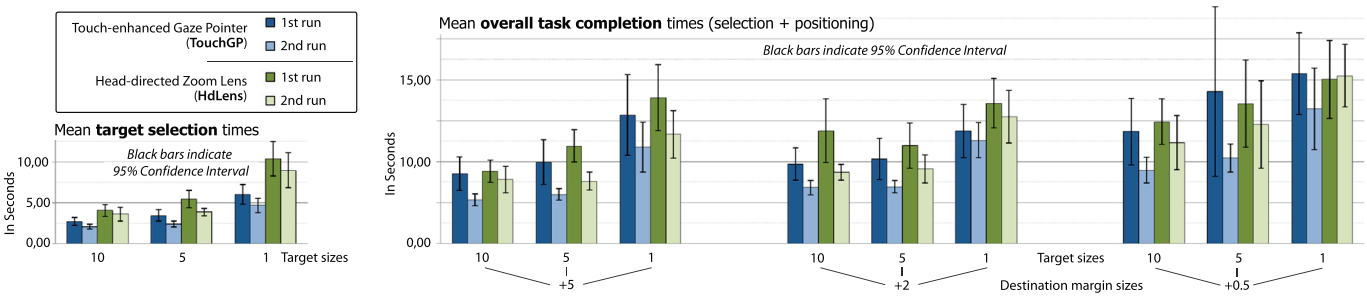

Figure 6. Mean selection and overall task completion times (selection+positioning) across varied target and destination margin sizes for *Task One*.

coarse positioning (significantly in the second run, see Table 1 for statistical results). For **HdLens**, the performance decrease from margin size +5 to +0.5 was even highly significant ($p<.001$). Finally, task completion times significantly differed across the techniques and associated runs ($F(3,360)=14.68$, $p<.001$). Users could significantly improve their performance from the first to the second run for both input conditions (both $p<.05$). In addition, users were significantly faster with **TouchGP** (second run) compared to **HdLens** (in both runs ($p<.001$)).

General Usability Ratings

In the first part of the intermediate questionnaires, we asked participants to rate several statements on the usability of **TouchGP** and **HdLens** for object selection (IQ1) and positioning (IQ2). An overview of the gathered user ratings is shown in Figure 7. The ratings for **TouchGP** and **HdLens** do not differ significantly. Interestingly, although users were able to perform tasks significantly faster with **TouchGP** (see previous section), the perceived *speed* at which actions could be performed was assessed lower than for **HdLens**. Some participants explained that they actually had the feeling that they could be faster with **TouchGP**, but that fine adjustments (both for selection and positioning) were difficult with it due to difficulties with the indirect touch input. In this line, the *accuracy* with which actions could be performed was not perceived as satisfactory, whereby users found that they could be more precise with **HdLens**. Both **TouchGP** and **HdLens** were assessed as *easy to learn*, however, with a bit more effort for using them for target positioning. Concerning the *ease of use*, both **TouchGP** and **HdLens** were assessed highly suitable for coarse selection and positioning. All in all, users found the techniques *intuitive* and suitable for achieving tasks as anticipated (*task-driven use*). In the *overall user rating*, **TouchGP** received similar and even slightly better ratings than **HdLens** (but not significantly).

User Feedback

In the following, we report and discuss the remaining quantitative (IQ3) and qualitative (IQ4) user feedback from the intermediate questionnaires and the feedback from the final questionnaire. For IQ3, participants had to rate how they liked particular aspects of **TouchGP** and **HdLens** based on 5-Point-Likert scales.

Touch-enhanced Gaze Pointer

All in all, users liked **TouchGP** (M=4.19, SD=0.81) and the particular combination of touch and gaze input (M=4.60, SD=0.61). In fact, fourteen participants explicitly mentioned that they enjoyed to roughly position the distant cursor via gaze (M=4.56, SD=0.61) and to precisely position it via touch (M=4.88, SD=0.33). However, the interaction was not found particularly convenient (M=3.56, SD=0.93), as many users explained that they found the touch input too imprecise and tiring for precise selection and positioning (M=3.00, SD=1.00). Users liked the seamless *drag-and-drop* combination for **TouchGP** (M=4.00, SD=0.87). Finally, users liked to double tap (M=3.88, SD=0.99) or hold on to the touch screen (M=3.81, SD=1.07) to position a target. However, they usually preferred performing a sliding gesture to warp the target to the current gaze position and immediately make fine adjustments from there (M=4.38, SD=0.98). In *Task Two* (cf. Figure 5), it became apparent that deselecting a target is not well handled with **TouchGP**, since the sliding gesture is assigned both for fine-grained selection as well as positioning. Thus, with the intention in mind to select a new target while another object has already been selected, users would accidentally reposition the still selected target by performing a sliding gesture.

Head-directed Zoom Lens

Users highly appreciated **HdLens** (M=4.38, SD=0.48). Four participants emphasized that they found it very intuitive to use, especially with respect to the seamless combination of selection and positioning (i.e., drag-and-drop) (M=4.44, SD=0.61). User liked the double tap for quick object positioning (M=4.31, SD=0.98). Users appreciated the three-zone lens design with the inactive inner zone (M=4.69, SD=0.68), the continuous gradient-based lens control (M=4.75, SD=0.43), and the absolute positioning if quickly turning away from the current lens position (M=4.80, SD=0.40). Five people particularly pointed out that they found it very useful that the lens did not move while the cursor was within the inner lens region. Participants also positively assessed that the lens already moved while looking within the magnified region (as in contrast to the semi-fixed zoom lens as proposed in [16]) (M=4.50, SD=0.71). While participants liked the vertical sliding gesture to adapt the magnification level (M=4.38, SD=0.70), the horizontal sliding gesture for changing the lens size was less appreciated (M=4.00, SD=0.71). This is probably due to the fact that many participants rather tried to achieve a task without

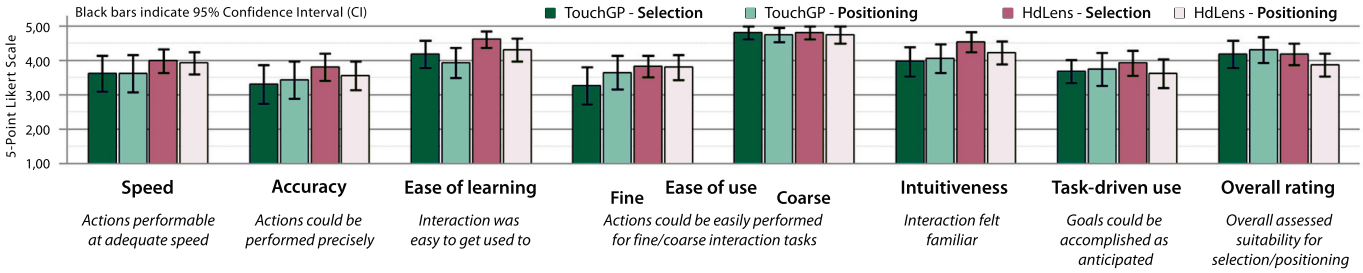

Figure 7. Overview of quantitative user feedback from the intermediate questionnaires IQ1 and IQ2 (with 5 – Strongly agree to 1 – Strongly disagree).

altering the lens size and thus did not assess this feature as important as zooming in, which was vital for selecting tiny targets. In this respect, nine participants particularly mentioned that they appreciated the possibility to adapt lens parameters, especially the magnification level. The provided visual feedback, i.e., target previews, target highlighting, and the crosshair at the lens center were appreciated (M=4.25, SD=1.03). However, additional auditory feedback to confirm a selection was desired (three participants). Three participants also mentioned that the head-directed cursor and crosshair should be more *eye-catching*, as they sometimes had difficulty to immediately detect it. All in all, the head-directed interaction was often described as less distracting, but also as more explicit and straining in the long run compared to **TouchGP** (five participants). Four people particularly mentioned that they preferred steering the lens via their head instead of their gaze, because it gave them more control.

Manipulation. Users did not find the mode change (i.e., turning the smartphone to the side) particularly convenient (M=3.25, SD=1.30). Four participants asked for either integrating all interaction tasks into one single mode or to use an alternative mode switch that would allow for an even quicker mode change. Furthermore, while participants liked the pinch gesture for scaling a target (M=4.44, SD=0.61), they preferred the tilt-to-rotation technique (M=3.94, SD=0.90) over the turn-to-rotate touch gesture (M=3.50, SD=1.17).

DISCUSSION

The positive user feedback suggests that our proposed gaze- and head-directed interaction techniques are suitable for a convenient and fluent selection and positioning of distant targets. Both **TouchGP** and **HdLens** were assessed as intuitive and easy to use in particular for *coarse* object selection and positioning. Users were in general faster with **TouchGP** for which they preferred to look at a destination and slide the finger on the touch screen to precisely position a target. Nevertheless, several participants found that head-directed input with **HdLens** provided a higher feeling of control, as it was less overwhelming than the interaction via eye gaze. However, it was also described as less implicit and more straining in the long run (higher fatigue). In addition, participants appraised the three-zone lens design and well-chosen function mappings (e.g., speed of lens movement). Finally, our gaze-supported drag-and-drop approach for seamlessly selecting and positioning targets was highly appreciated.

While the positive user feedback affirms that our careful design of gaze-supported object selection and positioning is feasible, several improvements are possible. For more precise cursor movements, a more appropriate mapping of touch input to the relative movement of the distant cursor has to be chosen for **TouchGP**. This would allow users to perform coarser touch movements on the smartphone for only slightly moving the distant cursor/target. Further improvements for **TouchGP** include a better way to deselect objects. One simple solution would be to make the respective interaction mode dependent on the current gaze position. Thus, if a user looked at a new target and performed a sliding gesture, the selection mode would be active. Instead if a user looked at a void spot on the screen, the positioning mode would be considered. However, this would pose a problem for overlapping images. As an alternative, an additional control at the back of the device could be used as a mode switch, especially since the remaining four fingers holding the smartphone have no other task so far than to stabilize the device in the user's hand. For **HdLens**, object positioning was found cumbersome at times due to the unusual indirect object positioning as users had to turn their heads to move the lens/the currently selected item. However, this is also closely related to the problem that precise lens movements were sometimes difficult to achieve. For more precise cursor/lens movements with **HdLens**, a smoother distribution of speed values in the *Active lens rim* could help.

The combination of gaze-/head-directed input with a smartphone may benefit diverse user contexts, for example, for the interaction with large-sized or multiple display setups at home or in offices. Virtual objects could be quickly selected and moved across different screens and devices simply by looking at them and touching the smartphone. Another interesting application context is also the integration of the presented techniques with modern gaming consoles and large-sized TV screens for more immersive game experiences.

Further work includes how to integrate additional interaction tasks while still maintaining a fluent interaction. For this, additional investigations are required to find out to what extent we can take advantage of simple manual input that users would naturally prefer without actually having to look at a handheld device (eyes-free). Finally, further studies have to determine how **TouchGP** and **HdLens** would compete against other techniques for interacting with distant displays (e.g., as presented by Nancel et al. [15], for example, in terms of speed, error rate, and comfort.

CONCLUSION

In summary, we have presented two novel sets of multimodal gaze-/head-supported interaction techniques that allow for seamlessly interacting with distant displays. For this, we take advantage of a user's gaze-/head-directed input as a coarse and fast pointing modality and a handheld touch screen from a ubiquitous smartphone for fine-grained input. The sets are based on two basic principles to overcome inaccurate gaze/head pointing: (1) additional manual touch control and (2) local magnifications. With these sets, users can fluently select a small or large target, position it to a desired location, and rotate and scale it. This means that our techniques support both *fine and coarse* target *selection and positioning*. Furthermore, we support both *distinct*, subsequent interactions and a *seamless combination* of target selection and positioning. Both multimodal interaction sets were evaluated in a user study with 16 participants. We received very positive feedback for both sets, which encourages further investigations into how to extend and improve these techniques. Overall, users were fastest with a touch-enhanced gaze pointer after some training. After all, both sets demonstrated a high potential for a practical, low-effort and fluent interaction with distant displays using gaze/head and touch input. However, further improvements are required especially for more precise object selection and positioning.

ACKNOWLEDGMENTS

This research is supported by the German National Merit Foundation.

REFERENCES

1. Argelaguet, F., and Andujar, C. Efficient 3D pointing selection in cluttered virtual environments. *Computer Graphics and Appl., IEEE 29*, 6 (11 2009), 34–43.
2. Ashmore, M., Duchowski, A. T., and Shoemaker, G. Efficient eye pointing with a fisheye lens. In *Proc. of GI '05* (2005), 203–210.
3. Bezerianos, A., and Balakrishnan, R. The vacuum: Facilitating the manipulation of distant objects. In *Proc. of CHI '05*, ACM (2005), 361–370.
4. Bieg, H.-J., Chuang, L. L., Fleming, R. W., Reiterer, H., and Bülthoff, H. H. Eye and pointer coordination in search and selection tasks. In *Proc. of ETRA'10*, ACM (2010), 89–92.
5. Bolt, R. A. Gaze-orchestrated dynamic windows. In *Proc. of SIGGRAPH '81*, ACM (1981), 109–119.
6. Boring, S., Baur, D., Butz, A., Gustafson, S., and Baudisch, P. Touch projector: mobile interaction through video. In *Proc. of CHI '10*, ACM (2010), 2287–2296.
7. Bulling, A., and Gellersen, H. Toward mobile eye-based human-computer interaction. *IEEE Pervasive Computing 9* (2010), 8–12.
8. Drewes, H., and Schmidt, A. The MAGIC touch: Combining MAGIC-pointing with a touch-sensitive mouse. In *Proc. of INTERACT'09*, Springer-Verlag (2009), 415–428.
9. Han, S., Lee, H., Park, J., Chang, W., and Kim, C. Remote interaction for 3D manipulation. In *Proc. of CHI EA '10*, ACM (2010), 4225–4230.
10. Jacob, R. J. K. What you look at is what you get: eye movement-based interaction techniques. In *Proc. of CHI '90*, ACM (1990), 11–18.
11. Kaiser, E., Olwal, A., McGee, D., Benko, H., Corradini, A., Li, X., Cohen, P., and Feiner, S. Mutual disambiguation of 3D multimodal interaction in augmented and virtual reality. In *Proc. of ICMI '03*, ACM (2003), 12–19.
12. Keefe, D. F., Gupta, A., Feldman, D., Carlis, J. V., Krehbiel Keefe, S., and Griffin, T. J. Scaling up multi-touch selection and querying: Interfaces and applications for combining mobile multi-touch input with large-scale visualization displays. *Int. J. Hum.-Comput. Stud. 70*, 10 (Oct. 2012), 703–713.
13. Kumar, M., Paepcke, A., and Winograd, T. EyePoint: practical pointing and selection using gaze and keyboard. In *Proc. of CHI '07*, ACM (2007), 421–430.
14. Monden, A., Matsumoto, K., and Yamato, M. Evaluation of gaze-added target selection methods suitable for general GUIs. *Int. J. Comput. Appl. Technol. 24* (June 2005), 17–24.
15. Nancel, M., Wagner, J., Pietriga, E., Chapuis, O., and Mackay, W. Mid-air pan-and-zoom on wall-sized displays. In *Proc. of CHI '11*, ACM (2011), 177–186.
16. Stellmach, S., and Dachselt, R. Look & touch: gaze-supported target acquisition. In *Proc. of CHI '12*, ACM (2012), 2981–2990.
17. Stellmach, S., Stober, S., Nürnberger, A., and Dachselt, R. Designing gaze-supported multimodal interactions for the exploration of large image collections. In *Proc. of NGCA'11*, ACM (2011), 1–8.
18. Turner, J., Bulling, A., and Gellersen, H. Combining gaze with manual interaction to extend physical reach. In *Proc. of PETMEI '11*, ACM (2011), 33–36.
19. Vogel, D., and Balakrishnan, R. Distant freehand pointing and clicking on very large, high resolution displays. In *Proc. of UIST'05*, ACM (2005), 33–42.
20. Ware, C., and Mikaelian, H. H. An evaluation of an eye tracker as a device for computer input. In *Proc. of CHI'87*, ACM (1987), 183–188.
21. Yamamoto, M., Komeda, M., Nagamatsu, T., and Watanabe, T. Development of eye-tracking tabletop interface for media art works. In *Proc. of ITS '10*, ITS '10, ACM (2010), 295–296.
22. Yoo, B., Han, J.-J., Choi, C., Yi, K., Suh, S., Park, D., and Kim, C. 3D user interface combining gaze and hand gestures for large-scale display. In *Proc. of CHI EA'10*, ACM (2010), 3709–3714.
23. Zhai, S., Morimoto, C., and Ihde, S. Manual and gaze input cascaded (MAGIC) pointing. In *Proc. of CHI'99*, ACM (1999), 246–253.