Toward Compound Navigation Tasks on Mobiles via Spatial Manipulation

Michel Pahud[1], Ken Hinckley[1], Shamsi Iqbal[1], Abigail Sellen[2], and William Buxton[1]
[1]Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA
[2]Microsoft Research Cambridge, 7 J J Thomson Ave, Cambridge CB3 0FB, UK
{mpahud, kenh, shamsi, asellen, bill.buxton}@microsoft.com

ABSTRACT

We contrast the *Chameleon Lens*, which uses 3D movement of a mobile device held in the nonpreferred hand to support panning and zooming, with the *Pinch-Flick-Drag* metaphor of directly manipulating the view using multi-touch gestures. Lens-like approaches have significant potential because they can support navigation-selection, navigation-annotation, and other such compound tasks by off-loading navigation to the nonpreferred hand while the preferred hand annotates, marks a location, or draws a path on the screen. Our experimental results show that the Chameleon Lens is significantly slower than Pinch-Flick-Drag for the navigation subtask in isolation. But our studies also reveal that for navigation between a few known targets the lens performs significantly faster, that differences between the Chameleon Lens and Pinch-Flick-Drag rapidly diminish as users gain experience, and that in the context of a compound navigation-annotation task, the lens performs as well as Pinch-Flick-Drag despite its deficit for the navigation subtask itself.

Keywords

Data navigation; spatial input; pan & zoom; chunking; handheld devices; sensors; 3D interaction; compound tasks

ACM Classification Keywords

H.5.2 Information interfaces and presentation: Input.

INTRODUCTION

Mobile devices demand techniques for navigating through large virtual spaces with a small display. The *Pinch-Flick-Drag* technique, which uses the direct-touch gestures of pinch-to-zoom, drag to pan, and flick for ballistic motion, has rapidly become the predominant idiom. However, it has the drawback that it requires both hands: one hand must hold the device while a second hand (typically the preferred hand) pinches and drags to directly manipulate the view.

In this paper we explore an alternative approach to navigation known as the *Chameleon Lens*, which uses direct spatial movement of the mobile device in the volume of space surrounding the user. It builds on the work of Fitzmaurice and others [8,9,26,27] who have explored the metaphor of moving a lens (that is, a viewport or portal) over a virtual workspace, as opposed to the metaphor of grabbing the underlying document itself. While both of these conceptual models ostensibly support the same interaction goals, we argue that they have important differences which could enable lens-like techniques to support compound navigation-selection, navigation-annotation, and other such tasks with a higher level of semantic complexity.

We believe it is therefore an important research goal to better explore and characterize the strengths and weaknesses of the Chameleon Lens approach. Design and implementation issues, such as which sensing approach and spatial mapping (transfer function) to use, need to be worked out. We also need studies to inform design choices such as zooming direction, use of a clutching button or not, and so forth. Furthermore, to our knowledge, there has been no formal study of generalized navigation tasks for the Chameleon Lens in direct comparison to the Pinch-Flick-Drag approach.

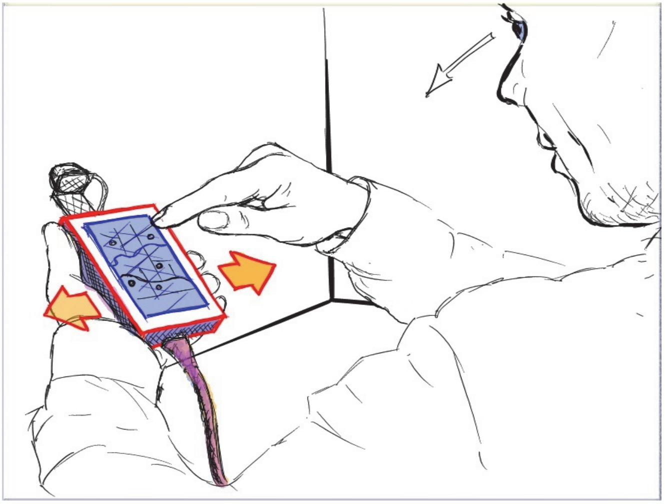

Fig. 1. The Chameleon Lens enables one-handed navigation via moving the device in free-space with the nonpreferred hand, while also affording bimanual touch-screen input with the preferred hand.

This paper contributes the following. First, we provide a rationale which establishes why the Chameleon Lens approach is an interesting alternative to consider, particularly in the context of mobile interaction. Second, we discuss several sensing approaches and spatial mappings to realize the Chameleon Lens. Third, we detail a series of studies which help to narrow down and justify a number of our experimental design choices, leading up to a formal study of map navigation which probes both articulatory (time-motion) and cognitive (spatial memory) aspects of performance. This revealed that, for the basic navigation task

MobileHCI '13, August 27 - 30 2013, Munich, Germany

ACM 978-1-4503-2273-7/13/08…$15.00.

itself, our Chameleon Lens implementation (as currently realized) is significantly slower than the standard Pinch-Flick-Drag technique. Fourth, a follow-on study suggests that differences between the Chameleon Lens and Pinch-Flick-Drag rapidly diminish as users gain experience with the technique. Fifth, a final follow-on study of a compound navigation-annotation task reveals that the Chameleon Lens may offer a performance advantage versus Pinch-Flick-Drag, though this currently does not rise to significance due to the limited number of participants in the final study.

Overall, these contributions provide a firm rationale and theoretical foundation for further research into lens-based navigation techniques and compound tasks for mobiles. Given our current experimental findings, we cannot recommend the Chameleon Lens for general navigation, but we believe that if further insights can improve performance of lens-like approaches, then substantial improvements may be possible. This would enable one-handed pan/zoom for mobile users, while also affording novel interaction designs.

DESIGN PROPERTIES OF THE CHAMELEON LENS

As suggested above, many characteristics of lens-like approaches motivate our interest in their potential, and hence make the Chameleon Lens and related techniques worth refining and studying carefully in this, and future, research.

Direct Input on a Moveable Display. The lens metaphor requires not only sensing the position and orientation of an input device, but also that the display is integrated with the device itself. This distinguishes the technique from virtual reality and desktop virtual environments [3]. Hand-held mobile devices afford such interactions, but larger boom-mounted displays can also employ this approach [26].

Addition of Multi-Touch. Unlike most earlier implementations [8,9,26,27], one notable exception being the Boom Chameleon [8,9,26,27], our Chameleon Lens display combines motion sensing with a soft-touch capacitive (multi)touch screen to offer richer interactive potential.

Integral Pan + Zoom vs. Orthogonal Control. The Chameleon Lens enables the user to simultaneously pan and zoom in *integrated* (or *coordinated*) movements [17,28]. However, this same integration means that without additional mechanisms, it is hard to exercise orthogonal control over zooming vs. panning, which by contrast is an inherent characteristic of Pinch-Flick-Drag.

One-handed Navigation via the Nonpreferred Hand. The pinch gesture is difficult to articulate one-handed [11]. Navigation by moving the hand holding the device frees up a human resource that is otherwise occupied: the user's second hand. This affords real-world tasks such as holding a bag or a child's hand. Furthermore, if the device is held in the nonpreferred hand—as is often the case for mobiles—it opens up the possibility of bimanual interactions where the preferred hand acts upon the results of the reference frame thus established by the non-preferred hand [12].

Touch Screen Freed for Other Inputs. Since motion sensors drive panning and zooming, a key hardware resource on the mobile—the touch screen itself—is not consumed by that task. Hence touch inputs or gestures can be given new interpretations without recourse to an explicit mode switch.

Layering-on of Compound Tasks. These freed-up resources—the second hand and the touch screen—afford layering an additional, possibly complementary, task on top of navigation. For example, the user can perform an integrated compound navigation + inking task by employing one hand for navigation and the other to annotate, mark a location, or draw a path on the screen using the finger or, if available, a stylus [10,16]. Likewise, although beyond the scope of this paper, such annotations can include spoken annotations [26] or commands [7] as well.

Now, having enumerated the above, our results also make it clear that obtaining all of the potential advantages of this approach is challenging. The techniques and studies reported in this paper, we hope, represent a solid first step in the progress that will be necessary to achieve this.

RELATED WORK

Researchers have conducted closely related work in control metaphors, lens-in-hand techniques, layered interactions, body-centric interactions, and spatial input techniques.

Control Metaphors: View-in-Hand vs. Document-in-Hand

Control metaphors have a long history in 2D scrolling as well as for spatial input. For vertical scrolling within documents, for example, grabbing the elevator of the scrollbar moves the *view* in correspondence with the mouse cursor, whereas on a touch-screen tablet dragging the *document* itself via direct manipulation is the preferred mapping. Note that neither the *view-in-hand* nor the *document-in-hand* model is necessarily the "best." Indeed, with the mouse, both the view-in-hand metaphor (scrollbars) and the document-in-hand metaphor (e.g. the "panning hand" used in Adobe Reader) are in common use. In general the input device, the stimulus-response compatibility, and the user's conceptual model of the task all influence the suitability of a mapping [1,2,17,21].

The Lens-in-Hand "Chameleon" Metaphor

The view-in-hand control metaphor (above) typically uses *indirect input*, in a motor space separate from the display itself. By contrast the lens-in-hand metaphor, as typified by the original Chameleon of Fitzmaurice et al. [8,9], is akin to holding a magnifying glass (or camera viewfinder) above a document, scene, or object. This model has been explored for sensor-augmented handheld displays as well as larger displays mounted on armatures or booms [26]. To our knowledge the input mapping (transfer function) and efficacy of the lens-in-hand metaphor has not been explored in detail for integrated (multi-degree-of-freedom) pan/zoom tasks. What is new about the Chameleon Lens reported in this paper, then, is the input mapping we devised for the lens-in-hand metaphor as applied to pan + zoom control, our studies examining its design dimensions and overall efficacy, and our discussion of how this affords layered pan/zoom navigation+selection tasks for mobiles.

Layered Inputs: Peepholes, Lenses, & Compound Tasks

Yee's Peephole display [27] also explores a lens-in-hand metaphor for mobiles. Yee presents a usability study on the 2D peephole, including both a map viewing task and a 2D panning + stylus annotation task. For map viewing, there was no significant difference between the Peephole vs. panning by dragging a stylus. For the annotation task, users were 32% faster and strongly preferred the combination of peephole (for panning) + stylus (for drawing). Yee also describes a 3D peephole using a 6DoF tracker, but does not study it in detail.

The Boom Chameleon [8,9,26,27] offers another example of layering tasks on top of 3D navigation, but employs a subtly different navigation metaphor than the original Chameleon [8,9] or our own Chameleon Lens. A counter-balanced boom supports a 17" display, and enables viewing a 3D virtual object, using a metaphor where one orbits the object with the display. The user can move the display up/down or push/pull to gain additional views—but all are constrained (by the mechanics of a boom) to look in at the central object. The Chameleon Lens has no such boom, of course, and can assume arbitrary orientations. It thus uses the navigation metaphor of looking *outward* from the center. This affords the interactions we considered most relevant for mobiles.

Body-Centric Interaction and Kinesthetic Feedback

The Lens-in-Hand metaphor also offers the potential to support body-centric mobile interaction. For example, Chen et al. illustrate holding a mobile over the hand or wrist to reveal different functions [5], and the Virtual Shelves technique [6] allows users to select particular objects or applications by orienting the device in a body-centric virtual space. Such body-relative orientations allow users to effectively leverage their kinesthetic (also known as proprioceptive) sense of how their body and limbs are positioned in space. This has been shown to foster spatial memory and recall for certain tasks [23,24]. The m+p spaces system [4] shows similar benefits by integrating a handheld with a pico-projector to provide the user with both focus and context for spatially-aware mobile interactions.

Augmented Reality, Spatial Input, and 3D Interaction

A variety of mobile lenses have been realized using marker-based augmented reality techniques [29], where there may (or may not) be an underlying physical object that grounds the spatial interaction. These hybrid techniques point to a wider design space of abstract vs. perceptually-grounded coordinate spaces, of which the Chameleon Lens is but one example, all of which could benefit from experimental study of the underlying motor and perceptual issues.

A number of desktop virtual environments have supported virtual manipulation in an absolute coordinate frame, often with visual feedback on a separate monitor. In this context it has been noted that users find it most natural to move an object in reference to a physical object or landmark. Moving through an imagined 3D volume with fixed Cartesian coordinates is not always as natural as one might expect. Likewise, in our early prototyping we found that a straightforward mapping of the Chameleon Lens viewport based on an abstract and fixed extrinsic coordinate space (with no physical object to perceptually ground the motion) did not feel particularly natural for pan + zoom control.

An analogous issue has been observed even for 2D mouse input. Before the modern generation of optical mice, some devices required a special mouse pad, printed with a fixed dot-pattern. Cursor movement was then mapped to the axes of motion in the coordinate frame of the mouse pad. The resulting motion felt unnatural since users often hold the mouse at an angle, and likewise move the mouse in shallow arcs rather than in perfectly straight lines. The VideoMouse [15] noted this and employed orientation sensing to automatically rotate mouse movement into the local coordinate frame of the mouse (rather than the extrinsic coordinate frame of the mouse pad), resulting in more natural cursor control. We used this insight to devise an appropriate mapping for the Chameleon Lens (described in detail later).

APPARATUS & SOFTWARE

For the Pinch-Flick-Drag technique, we employed a commercial HTC HD7 mobile phone with a 480x800 pixel (4.3 inch diagonal) LCD display with capacitive multi-touch. We used a standard system control (with its default settings) to implement the Pinch, Flick, and Drag gestures; hence the resulting interactions were highly refined and fully comparable to those on commercially available devices.

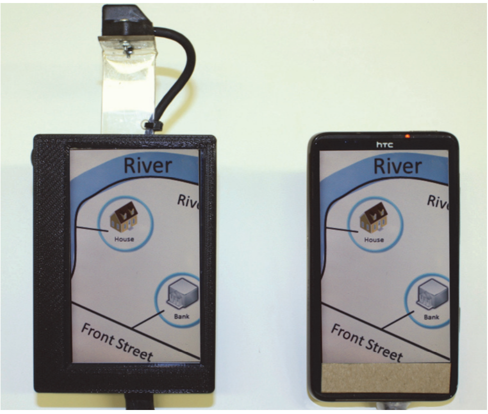

Fig. 2. The Chameleon Lens **(left)** uses a magnetic tracker to sense absolute position and orientation. Pinch-Flick-Drag **(right)** uses a standard multi-touch smartphone with the same resolution and size.

The Chameleon Lens display has the same 480x800 resolution and 4.3 inch diagonal as the mobile phone, but weighs a bit more (230g vs. 160g). Our current sensing approach is a compromise: our goal is to assess the technique's potential, rather than to realize a practical solution that is fully mobile. If the technique fails to perform as we hoped, we do not want it to be due to the current limitations of sensors on commodity mobile devices.

The Chameleon Lens employs the Polhemus FASTRAK 6DOF tracker for sensing absolute position and orientation (this is the same sensing approach that was taken by the original *Chameleon* project of Fitzmaurice et al. [8,9]). The

Polhemus is mounted to the top of the display by a short bracket to minimize magnetic interference. The Polhemus cannot interface directly to a mobile phone; the data is therefore collected by a PC, and the display employed in our prototype is actually a small external monitor of the PC with a capacitive multi-touch overlay. This also means the Chameleon Lens display is tethered by a cable to the PC.

We used an external display for the Chameleon Lens, rather than wirelessly transmitting the Polhemus data to the mobile phone, to avoid exacerbating latency. The importance of low latency for spatial input in has been oft-noted [13]. We 3D-printed a case for the Chameleon Lens display, and added a button for clutching. This makes it somewhat bulkier than the mobile phone, but not markedly so for our purposes.

Chameleon Lens Mapping from Real to Virtual Space

While intuitively a literal 1:1 mapping of the device movement to the virtual space beyond the display would appear to be a natural and obvious choice, our experience with developing a technique to support effective mobile navigation of information spaces with the Chameleon Lens gradually convinced us otherwise. We therefore considered several transformations to map the position and orientation of the device in physical space to a view on the virtual space.

Unlike the scenarios in the original Chameleon [8,9], for general navigation the user does not move the device in reference to a fixed coordinate system. The user may shift body posture over time, there is no external object in the environment that serves as a "prop" to ground the virtual space [14], and even the sloping orientation at which users typically hold a mobile device—not the rigid vertical posture of a physical window, or the horizontal posture of a piece of paper on a desk [27]—requires that the mapping consider the 3D position *and* the 3D orientation of the device.

In our Chameleon Lens prototyping, we explored three different mappings, the third of which we felt worked the best, and therefore settled upon for our formal studies. However, we describe all three here to illustrate what we tried, and why, and how this informed the mapping we chose.

Fixed planar mapping. This mapping ignores the orientation of the device and defines the zoom in terms of how far the device penetrates into the fixed vertical planes of the workspace. So if the user moves the device over the center of the table, this produces a fixed zoom level regardless of how the screen is oriented. However, when the user moves left or right (for horizontal panning) he has to remain within that fixed plane to pan without affecting the zoom level. We found this was difficult for people to do, because users tend to sweep their arm in broad arcs when panning.

Fixed spherical mapping. The tendency we observed for people to move in broad arcs led us to consider a spherical mapping. Here, if the user holds the device steady and pivots their body left/right, the zoom stays unchanged while the map pans left/right. Note also that this means the orientation of the device changes, as seen in the extrinsic reference frame of the workspace. This mapping utilizes the presumptive location of the user's eye-point as the center point about which the pivoting occurs. This mapping also allows the user to pan left/right until they see an object on the screen, and then move the device towards/away from themselves along the ray intersecting their eye-point and the virtual target to zoom in and out, which feels quite natural. The problem with this mapping is that, particularly when the device gets closer to the body, it is quite sensitive to the choice of the center-point, and becomes difficult to control.

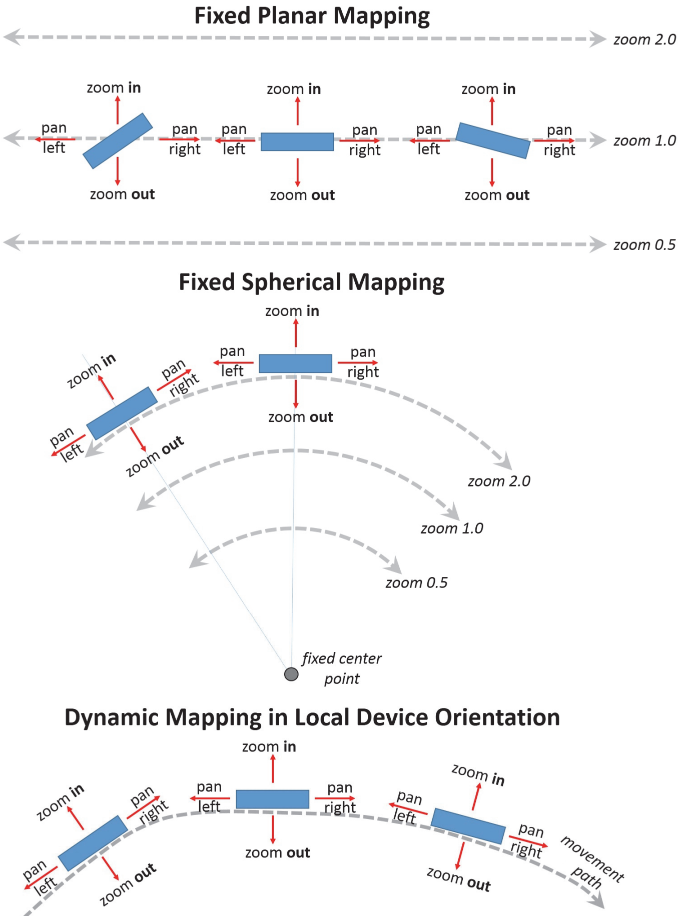

Fig. 3. Illustration of the primary mappings explored for the Chameleon Lens prototype. See text for explanations.

Dynamic mapping in local device orientation. Problems with the above mappings led us to realize that the orientation of the device itself may be the best reference frame in which to compute the user's intended motions through the virtual workspace. This is analogous to the 2D external reference frame imposed by optical mouse pads printed with fixed grids, as noted earlier. We adapted the solution employed by the VideoMouse [15], that of sample-by-sample rotation of movements to the local reference frame of the device, and extended it to the 3D case, so that Chameleon Lens' "zooming" direction is defined as the direction of motion normal to the screen, while panning occurs in the plane defined by the instantaneous left-right and up-down directions as seen by the mobile device. We found that the resulting motion feels natural to users and addresses the major shortcomings of the other mappings noted above.

MAP NAVIGATION EXPERIMENTAL TASK

We chose to study the Chameleon Lens and Pinch-Flick-Drag techniques in the context of map navigation, where the user must pan, zoom, and indicated desired locations on an unfamiliar map. This reflects an ecologically valid usage context where mobiles are employed in everyday life, e.g. to seek out locations to visit in a new city upon arrival.

We used an unfamiliar map so that our results would not be biased by wide variances of users' familiarity with a real locations. Furthermore, we wanted to employ a task which required both physical movement as well as a cognitive aspect in terms of "learning the lay of the land" and, eventually, remembering where oft-visited targets on the map were. We later leveraged this to supplement our primary measures of movement efficiency (time) with a memory test assessing spatial landmark recall (described in further detail below in the section describing our main experiment).

The maps we employed drew on a previous experimental study of knowledge acquisition from maps (including tests of spatial memory of locations), with some street names and landmarks from the "town" map of [25] modified to suit our purposes. The map was large enough such that the entire map could not be viewed in a single screen-full (1440 x 1080 pixels), with zoom levels between 1/3x and 2.5x allowed. The map thus had resolution of 3600x2700 when fully zoomed in, and 480x360 when fully zoomed out, meaning that the full map was just visible on the 480x800 display.

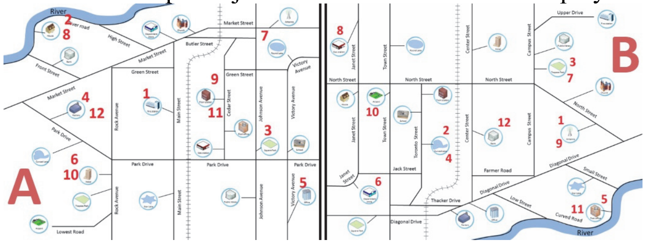

Fig. 4. The mirror-symmetric maps (A / B) used in the study, with order of targets. (The red numbers were not visible to participants).

We limited panning and zooming so that problems such as getting lost or zooming into "desert fog" [18] would not derail the experimental study, while still allowing an adequate range to assess navigational movements in an ecologically valid task scenario. Fine details such as icons and landmark names were not legible when the map was fully zoomed out, and users thus had to both pan and zoom to accomplish the experimental task. The targets were round icons (43 pixel radius at zoom of 1.0) that the user had to tap inside to successfully select the target, which largely forced users to zoom to tap the small targets without errors.

In all pilot studies and in the main experiment, users were initially given 30 seconds to explore a new map as they saw fit. In the experimental trials they visited a prescribed series of named landmarks on the map. We carefully designed all conditions to have an equivalent total movement distance between targets. Furthermore, to avoid building familiarity with a map in one condition and having that influence the follow-on conditions, we devised two variants of the map ("Map A" and "Map B") that were mirror-image symmetric. Thus the complexity of streets and landmarks was identical, and we left the landmarks in the same relative locations but switched the *icons* and landmark *names*, and also changed the street names. That is, in Map B, the landmark names and icons were replaced with a new set—none were the same as in Map A. In this way we know the complexity of the maps remains unchanged, yet (as confirmed by pilots) users are unlikely to transfer experience between Map A and Map B.

REFINEMENT OF DESIGN CHOICES VIA PILOT STUDIES

We ran several pilot studies to guide our design choices for the Chameleon Lens, as well as to refine our experimental design. With the exception of Pilot Study #1, which used pairs of familiar landmarks, the pilot studies employed the map navigation experimental task as described above.

Pilot Study #1: Familiar Landmarks and C:D Ratio

In Pilot Study #1 we presented users with fixed targets at horizontally opposed sides of a map, and subjects repeatedly selected each in a reciprocal pointing task. This was intended, in part, to simulate a condition where users *already know the location of a familiar target*, i.e. without any visual search or navigational behaviors. This task therefore lets us examine Chameleon Lens in a context that can fully leverage the spatial memory of the kinesthetic sense, in terms of moving back and forth largely by "muscle memory" [23,24].

We also used this experimental setting to optimize the Control:Display ratio of the Chameleon Lens. This is expressed below in terms of the total distance of movement necessary to traverse the map at the highest zoom level. The 40 cm level tested below corresponded exactly to the dragging distance that would be required by Pinch-Flick-Drag if no flicking gestures were employed by the user.

Pilot #1, Phase One: C:D Optimization for Panning

Ten users (all right-handed) participated in the pilot, with:

2 conditions (Chameleon Lens, Pinch-Flick-Drag) **X**
{3 C:D ratios (60 cm, 40 cm, 20 cm) for Lens, **or**
1 default C:D ratio for Pinch-Flick-Drag} **X**
10 reciprocal (back-and-forth) pointing motions,
X 10 users = 400 total reciprocal pointing movements.

With Chameleon Lens, per block of 10 reciprocal pointing motions users averaged 49.5s for the 60 cm distance, 42.9s for the 40 cm distance, and 40.1s for the 20 cm distance. Pinch-Flick-Drag, by comparison, was much slower for this task, averaging 72.5s per block of 10 reciprocal pointing movements. Bonferroni-corrected post-hoc comparison of means revealed that Chameleon Lens was significantly faster than Pinch-Flick-Drag for all C:D settings tested ($p<0.0001$). Subjective user responses ranked the 40 cm distance as most fatiguing, with a slight overall preference for the 20 cm mapping. Since this was also the fastest, we therefore selected the 20 cm mapping for all subsequent studies.

It is important to note that the Pinch-Flick-Drag condition in this pilot study used the default system settings, and furthermore that we allowed participants to use a combination of dragging or flicking motions as they saw fit.

We thus tested Pinch-Flick-Drag here in its strongest (idiomatic) form. During the trials, we observed that the Chameleon Lens' advantage for this task appeared to reflect participants' ability to move, via the kinesthetic sense, directly to the targets. The motion was thus a ballistic act, rather than one requiring continuous motor-visual guidance.

Pilot #1, Phase Two: C:D Optimization for Zooming

Six users participated in a second pilot, structured as above, with distances of 20 cm, 40 cm, and 10 cm between the min/max zoom levels for the Chameleon Lens, versus two-finger pinch-to-zoom with Pinch-Flick-Drag. For the Chameleon Lens, participants averaged 26.2s per block of 10 reciprocal zooming movements for the 20 cm zoom mapping, 25.3s for the 40 cm zoom mapping, and 18.2s for the 10 cm zoom mapping. The corresponding time for pinch gestures with Pinch-Flick-Drag was 25.2s. However, even though in the context of the experimental task 10 cm was fastest for the Chameleon Lens, users commented that it felt too sensitive to small hand motions, so subjectively it was not well-received. Based on these results, and to maintain a symmetric coordinate space, we decided to stick with 20cm to traverse the min-to-max zoom level as well.

Discussion of Results for Reciprocal Pointing Task

Our pilot results strongly suggest that the Chameleon Lens approach is advantageous in a context where the location of the target is known beforehand, where user does not have to scan the intervening landmarks on the map. This allows the Chameleon Lens to leverage the kinesthetic sense to navigate directly to an off-screen location. With Pinch-Flick-Drag the user must navigate over—*and visually attend to*—the intervening space to know when to stop moving.

Pilot Study #2: To Clutch or Not to Clutch?

Spatial interaction techniques require some way to reposition the device *without* causing an undesired change in the virtual view on the device's display. This is a tricky design issue [13]: a good clutching mechanism, proper placement and hand posture, and the potential for fatigue if users have to move while holding down a button for long periods of time all mean that including a clutch button in our experimental study had its own challenges.

Furthermore, the presence of a clutch button might interfere with the kinesthetic sense, as found advantageous in **Pilot Study #1** above, or with spatial memory (i.e. the ability to recall landmarks). The clutch, by its very definition, decouples the absolute space around the user from the virtual space, and thus any remembered body posture, or distance moved, loses its significance once clutching occurs.

To scrutinize this experimental design choice we ran a pilot study with 8 participants (all right-handed) comparing Chameleon Lens *With Clutch* versus *No Clutch*. In the *With Clutch* condition a button mounted on the left side of the mobile display, which was designed to be held down by the thumb of the user's nonpreferred hand, was used to engage the lens (that is, the device movement only caused the view shown on the display to change when the button was held down). In the *No Clutch* condition the view always moved when the device moved, regardless of the button state.

The *With Clutch* condition resulted in higher average movement time (9.5s) and a greater total hand movement distance compared to *No Clutch* (8.5s), but the results were not statistically significant. We therefore decided to run the main experiment without a clutch, as it does not appear to be a major factor contributing to performance; if it had any effect at all, our data suggest *No Clutch* may tend towards improved average performance for our experimental task.

Pilot Study #3: Zooming Direction

Chameleon Lens offers two possible zooming metaphors:
1) Zoom Positive: zoom out when the device gets closer to the user, zoom in when the device moves further away; or
2) Zoom Negative: zoom out when the device gets closer to the user, zoom in when the device moves further away.

Since (much like scrolling), choice of metaphor depends on how a user conceives the task, we conducted a pilot with four right-handed users to see if they had a clear preference. All four users strongly preferred *Zoom Negative* and that condition also had the fastest movement time (6.1s, vs. 7.0s for *Zoom Positive*; but with only 4 users, the difference was not statistically significant). We therefore decided to stick with the Zoom Negative mapping for all of our studies.

Pilot Study #4: Left Hand vs. Right Hand Control of Lens

We conducted a final pilot study to assess overall user acceptance of navigation with the nonpreferred hand. We tested three typical mobile device usage postures:

- *Non-Dominant.* Two-handed: user navigates with the nonpreferred hand, and selects targets with the preferred hand, per the principles articulated by Guiard [12].
- *Dominant.* Hand roles reversed to instead navigate with the preferred hand, and select with the nonpreferred hand;
- *Thumb.* Allows fully one-handed input: navigate with preferred hand, select with thumb of preferred hand also.

Twelve right-handed users participated. They ranked *Non-Dominant* as most natural, and were evenly split between the *Dominant* and *Thumb* conditions for runner-up. The *Dominant* condition (4.03s) exhibited the fastest movement time (versus 4.85s for *Non-Dominant* and 4.57s for *Thumb*), but the differences were not statistically significant.

Based on these results and our higher-level goal of investigating the technique as a way to afford compound navigation tasks for mobile devices (such as the aforementioned *navigation+selection* and *navigation+ annotation* tasks), we decided to stick with the *Non-Dominant* hand assignment for our main study. Although our timing results suggest this might not be the fastest technique when considering the navigation portion of the task in isolation, we do believe this represents the most natural idiom of the technique, and the non-significance with 12 users suggest that any performance difference resulting from the hand assignment to task was not a dramatic one.

Summary of Pilot Study Findings

With our careful exploration of the design space of the Chameleon Lens technique, as informed by our pilot studies, we therefore pursued an experimental study with:

- A 20 cm Control:Display mapping for device movement to movement across the entire range of pan and zoom;
- *No Clutch* for the Chameleon Lens to allow full freedom of movement and maximal leveraging of any spatial memory benefits of the kinesthetic sense;
- *Zoom Negative* mapping for the zoom axis, so that moving the device closer to the user zooms out;
- *Non-Dominant* hand assigned to navigation, and the dominant hand assigned to selection (tapping the target).

Furthermore, Pilot Study #1 revealed that Chameleon Lens offers a clear and statistically significant advantage over Pinch-Flick-Drag for moving back and forth between pairs of targets at known locations. Here, the kinesthetic sense affords moving directly to the target location rather, than visually guiding continuous dragging and flick gestures.

Because we conceive of Chameleon Lens as a way to perform general-purpose navigation with the potential to support the layering-on of compound tasks, in our main experiment we continued to focus on map navigation among a series of unfamiliar targets. However, in light of the largely negative result obtained therein, it is worth keeping Pilot Study #1's positive finding for familiar targets in mind. Such a finding suggests the Chameleon Lens approach may be extremely well suited to interactions such as calling up tool palettes [20], or quickly moving between frequently used applications [5,6], because these represent navigation between small sets of familiar target locations. It is also possible that Chameleon Lens could show greater promise for generalize map navigation with more practice, or that it may become advantageous when employed in the context of layered (compound) tasks. We probed these possibilities further, in follow-on studies reported later in this paper.

MAIN EXPERIMENT

Our main experiment employed the map navigation task described above, in the context of moving between a series of landmarks that were initially unfamiliar to the user. Over the course of the experimental trials, we expected users would gain at least some familiarity with the landmarks, and in fact we designed the task with an idea towards testing the facility of landmark recall afforded by the Chameleon Lens compare to the traditional Pinch-Flick-Drag technique.

Spatial Memory: Landmark Recall Test

To evaluate how much peripheral spatial awareness and richness of recall each technique afforded, we developed a landmark recall task [25]. The purpose of this task was to see if the choice had any impact, adverse or otherwise, on user's ability to recall the landmarks they had visited – as well as the surround of other landmarks which they had not visited in the experimental trials. If users are able to correctly recall the locations of targets that they did not explicitly visit, it would suggest that a technique promotes peripheral awareness and better recall of the spatial layout of the map.

This task was administered on a separate multi-touch table, immediately after users finished with a given technique (i.e. Pinch-Flick-Drag or Chameleon Lens). The entire map was visible without any panning or zooming. The table initially showed the map with all of its (named) roads, but stripped of all of the labelled circular target icons (the landmarks). The landmarks appeared on the periphery of the screen and participants were asked to **1)** drag and drop the landmarks that they were absolutely confident they recalled to the corresponding locations on the map, and then **2)** drag any additional landmarks they were somewhat sure about. Users were instructed to leave landmarks untouched if they didn't remember them or know (at all) where to place them.

Note that *we only administered the Landmark Recall test after the first condition*, with Map A, to ensure that users would not change their behavior in the second condition to try to memorize or otherwise scrutinize the targets more carefully in anticipation of an upcoming memory test. Therefore, we also told participants that there would not be a Landmark Recall test administered after the 2nd condition.

Participants

We recruited 32 participants external to our institution (16 female and 16 male, from 20-50 years old with an average age of 36.8) to participate in the study. All users were right handed, owned a touch-screen phone or tablet, and had at least 10 hours experience with the traditional Pinch-Flick-Drag technique prior to the study. All had normal color vision, and none of the participants had previously used Chameleon Lens or participated in any of our pilot studies.

Experimental Design

The main experimental factor was *Technique* (Chameleon Lens vs. Pinch-Flick-Drag). Half of the subjects performed Chameleon Lens first, and the other half performed Pinch-Flick-Drag first, to counter-balance the order of conditions. Subjects were randomly assigned to an order (16 per group). Each subject performed 4 *blocks* of target acquisition tasks with each technique, with each block consisting of a fixed sequence of 12 *trials* of individual target acquisitions. Users had to progress through the targets in order, and were prompted with the target name to find next. Each of the four blocks for a particular technique used the same sequence of targets, so by block 4 users became familiar with the targets.

Successful target acquisition was accomplished by tapping, with a finger of the preferred hand, on the touch-screen within the accuracy constraint specified by the circle around the landmark icons. Users were not allowed to proceed to the next target until they had tapped within the circle. This criteria prevented users from "guessing" or trying to tap "close enough" to where they thought the landmark icons were with the map zoomed very far out; we wanted the task to engage both panning and zooming navigational movements typical of real interactions with mobile maps.

The study employed two maps, Map A for the first technique and Map B for the second technique, but this was not a factor of experimental interest per se. We carefully designed the two maps to have identical complexity and identical total path length between the indicated targets, as noted previously (*Fig. 4*). Indeed, we only used two different maps so that, in whatever technique came second, our assessment of performance would not be influenced by familiarity with Map A gained during the trials for the first technique.

The experiment thus consisted, for each of 32 participants:

2 Techniques (Chameleon Lens, Pinch-Flick-Drag) X
4 Blocks per technique X
12 trials (targets) per block,
= 96 targets per user, and 3072 targets total in the study.

Participants also performed the Landmark Recall Test after completing the first technique (Map A), but not after Map B (to ensure users would not alter their behavior for the second condition in anticipation of a memory test, as noted above).

Procedure

Participants had 30 seconds at the start of each condition to familiarize themselves with the apparatus. Each trial prompted the participant to find a target (specified by name); the map was not visible during this time. Users tapped on the screen to acknowledge the prompt (and bring up the map), which began the next trial. Trials began with the previous target still on-screen, and without any change in zoom factor. The trial ended when users navigated to and tapped successfully on the specified target. Finally, the first trial began fully zoomed-in on the *last* target of the block (since every block moved through the same cycle of targets).

Primary Navigational Task Results

We treated Block 1 as practice to ensure lack of familiarity with the procedure or techniques would not unduly influence our results. Therefore, all statics reported below include blocks 2-4 only, but a separate analyses including all four blocks confirmed this did not influence the findings.

A one-way ANOVA on *Technique* revealed that the mean target traverse time was significantly slower ($F_{(1,30)}$=10.1, $p<0.005$) for the Chameleon Lens (5.60s mean, σ=3.03) than the traditional Pinch-Flick-Drag technique (3.96s, σ=1.1).

We also performed a one-way ANOVA on *Technique* for the mean traversed path length (the sum of total physical displacement of the device in x-y-z), which showed the path length was also significantly longer ($F_{(1,30)}$=8.4, $p<0.01$) for Chameleon Lens (15247 pixels, σ=8325) than for Pinch-Flick-Drag (10965 pixels, σ=2344).

The similar level of differences between techniques for traversal time (Chameleon Lens 28% slower) and path length (Chameleon Lens 29% longer) does suggest, however, that the less direct path which users followed between targets when employing the Chameleon Lens likely accounts for much of the performance difference between the techniques.

However, in contrast to the advantage Pilot Study #1 revealed with Chameleon Lens for Navigation to Familiar Landmarks, the negative result in the main study suggests the mapping we used for simultaneous panning and zooming movements likely requires significant refinement for the Chameleon Lens to perform on-par with Pinch-Flick-Drag for more general navigation tasks. We also probe whether additional experience with the Chameleon Lens reduces this gap in a follow-on probe, reported later in this paper.)

Landmark Memory Test Results

We tallied a target recall metric which awarded 1 point if the user placed the icon for a target at least partially overlapping the correct icon from the full map, and 0.5 points if the icon was placed in the correct region of the map (i.e. within the correct polygon enclosed by the nearest road boundaries). For targets that users reported being "absolutely certain" about, a one-way ANOVA on Technique showed no significant difference ($F_{(1,30)}$=0.75, $p<0.4$, n.s.) between Chameleon Lens (6.38 targets recalled, σ=3.6) and Pinch-Flick-Drag (7.56 targets recalled, σ=4.1).

We also found no significant differences for targets that users were somewhat sure about and placed correctly (Chameleon Lens 4.88 targets, σ=2.9 vs. Pinch-Flick-Drag 3.39 targets, σ=1.7). Likewise we found no difference in the incidental recall of targets that users had not explicitly visited (Chameleon Lens 2.31 incidental targets placed correctly, σ=1.4, vs. Pinch-Flick-Drag 2.19 targets, σ=1.6).

These results confirm the differences in traversal time and path length that we observed in the primary task conferred no clear advantage (or deficit) in terms of user's ability to correctly remember the arrangement of landmarks on the map, whether those landmarks had been explicitly visited or were other landmarks that the user only recalled incidentally.

EXPT. 2: SKILL ACQUISITION WITH CHAMELEON LENS

To help gauge the potential impact that additional practice and experience with the Chameleon Lens might have on our results, we conducted a preliminary follow-up experiment where test users performed 20 blocks of our navigational task with either the Chameleon Lens or the Pinch-Flick-Drag technique. We chose a between-subjects design for this experiment so that we could probe the acquisition of skill with each technique over a substantial number of trials.

Eight subjects participated in the study, and were assigned randomly to either the Chameleon Lens or Pinch-Flick-Drag condition. None had participated in our previous studies. Users performed 20 total blocks with the assigned technique, completing 8 blocks one day and the next 12 blocks on a separate day, all within 72 hours. In the Chameleon Lens condition, mean traversal time improved 33.0% between block 1 (4.5s) and block 20 (3.0s), whereas for the Pinch-Flick-Drag condition, mean traversal time improved just 19.4% (from 3.6s to 2.9s). While this difference was not statistically significant (with the between-subjects design, only 4 subjects performed each condition), it is suggestive that much of the performance gap between the techniques may disappear as the users acquire greater skill with the novel Chameleon Lens technique.

EXPT. 3: A COMPOUND NAVIGATE-ANNOTATE TASK

We conducted a final experiment to investigate the performance of Chameleon Lens and Pinch-Flick-Drag in a compound task consisting of navigating between a series of targets, but then also annotating (circling) each target. This experimental task is meant to be representative of a broad class of tasks where users move between locations in a document, map, or spreadsheet (for example), and then must annotate or otherwise mark-up that location. Note that while employing a pen for such mark-up certainly could be desirable in some tasks scenarios, for the purposes of the present paper we stuck with the predominant mobile idiom of using a finger (touch) for all on-screen interactions.

Experimental Task and Technique Parameters

With the Pinch-Flick-Drag technique, the annotation task requires a mode switch so that the finger can leave a mark instead of panning the display. A touchscreen button at the bottom of the display toggled the mode between navigation and annotation. With the Chameleon Lens, no mode switch was required since the user navigates by moving the device itself; dragging a finger on the display always draws a mark.

Users navigated to the landmark icons in a prescribed sequence. Users then annotated the targets by circling them. The circle that they drew had to follow the outer (circular) boundary of the landmark icons within a prescribed threshold; users therefore had to zoom in close enough to the icon to trace the circular icon boundaries accurately enough.

For this task, we decided to use the clutch button of the Chameleon Lens so that the user could “freeze” motion of the lens while performing the mark-up. We felt this was a more ecologically valid way to present the Chameleon Lens for this task because the clutch allows users to keep the view still and hold the device in a comfortable position while performing the annotation. Furthermore, as established in Pilot Study #2 above, the presence or absence of the clutch does not appear to have a major impact on the core navigational task itself; and in the main experiment, omitting the clutch did not appear to confer any spatial memory advantage for the Chameleon Lens in terms of users’ ability to recall the location of the landmarks.

Participants and Experimental Design

Eight subjects (4 male, 4 female) participated in the study; none participated in any of our previous experiments. Order of presentation of the techniques was fully counterbalanced, with 4 subjects (2 male and 2 female) experiencing each condition first. For each condition, subjects performed a single block consisting of navigation to 15 targets to circle. Navigation to the targets followed a prescribed path which was visible as directional arrows on the map itself.

Results for Compound Task

A one-way ANOVA on technique revealed that, despite the significantly worse performance we observed with the Chameleon Lens on the core navigational task reported in Expt. 1 above, for the compound navigation-annotation task performance was on par between the Chameleon Lens (15.7s, σ=6.9) versus the Pinch-Flick-Drag technique (17.0s, σ=7.8). Although the absolute means appear to favor the Chameleon Lens for this task, the difference was not statistically significant ($F_{(1,7)}=1.98$, $p<0.20$).

Qualitatively, several users noted it was “easy to navigate with one hand and draw with the other” with the Chameleon Lens. However, most (5/8) subjects still preferred Pinch-Flick-Drag for this task. Thus, it appears that unless we can devise a means to improve the efficiency of the Chameleon Lens for the core navigation task, it is unlikely to be well-received by users, even in the context of compound tasks.

DISCUSSION

We learned a great deal in this work – not all of which is reflected in the statistics.

First, we came away from this work with a much deeper appreciation of the three-pronged navigation strategy afforded by Pinch-Flick-Drag—pinch for moving across levels of detail, flick for rapidly moving across long distances, and drag for fine positioning— and in particular the lack of a corresponding ‘Flick’ affordance for the Chameleon Lens. It may be possible for future explorations of the technique to support “flick” via jerking motions [22].

Second, while our implementation did not perform well in the primary 3D navigation/selection study, Pilot Study #1 indicated it could outperform Pinch-Flick-Drag in the 2D case; clearly, it excels for some tasks. Thus, for tasks such as recalling virtual shelves [6], selecting a few commands [5], or glancing into tool-spaces [20], lens-like approaches using device motion may well offer significant advantages.

Third, as the Boom Chameleon [29] demonstrated, the object-centric mapping can work well, suggesting that there is more room for work here as well, both in terms of sensing technologies and experimental studies of device mappings.

Fourth, we missed an important difference between Boom Chameleon and hand-held approaches (as taken by us, and Fitzmaurice et al. [8,9]). In Boom Chameleon, the distance between the eye and the display stays nearly constant; the user moves with the display. With the hand-held Chameleon, zooming involves moving the display towards or away from the eye. Hence, with the boom-Chameleon, like the Pinch-Flick-Drag technique, the zoom factor in display coordinates correlates well to eye coordinates. When zooming by moving the display to/from the eye, the perceived scale is a function of both the scale on the screen *and* the distance from the eye. If we could factor in screen-eye distance in determining the scale on the screen, and perhaps refine other subtleties of the mapping, the results may improve.

Fifth, our struggles with the spatial mapping also suggest opportunities for further technical exploration and experiments—perhaps isolating specific body-centric vs. exocentric coordinate systems and mappings—that could help to zero in on the perceptual and motor considerations that impact behavior with such techniques. The literature lacks a complete understanding of such factors, and teasing them apart in a principled way strikes us as a necessary endeavor if techniques analogous to the Chameleon Lens are going to find wider success, but as of this writing, a deeper understanding along these lines still eludes us.

Nonetheless, in the likelihood that despite our best efforts we may have gotten the subtleties of the spatial mapping wrong, it is important to note that the layered annotation task of our final pilot study (while having too few participants to gain statistical significance) already shows a case where the Chameleon Lens matched the performance of Pinch-Flick-Drag. This suggests improved mappings could lead to corresponding gains for such compound tasks, while also gaining the advantage (as revealed in Pilot Study #1) of rapid movement between a small number of known locations.

Alternate Sensing Approaches & Enhancements

An important pragmatic consideration for lens-like techniques is that body-centric interactions require tracking both the device and the person in a common coordinate frame. Thus a mobile device that tracks only self-motion in its own *local* coordinate frame, using a built-in camera or inertial sensors (for example), cannot readily support body-centric interactions. To achieve a higher-fidelity egocentric mapping, the system should also track both the user (to know the eye-point relative to the screen of the device) as well as the device itself. This could be one source of perceptual misalignment, possibly harming performance, in the mapping that we settled on.

We implemented alternate sensing approaches so that we could try out fully mobile variants of the technique on existing devices. We built a prototype using the phone's motion sensors, but accelerometers cannot track linear distances accurately, and hence the illusion of moving an absolute distance is just that—an illusion. We also implemented a prototype employing the rear-facing camera to sense optical flow (e.g. [19]). We found this problematic due to limited frame rate, difficulty in precisely tracking zooming motions, and high power consumption.

Because these approaches cannot robustly sense the absolute position and orientation of the device at a high frame rate, we found they offered an inferior realization of the Chameleon Lens to our Polhemus-based prototype. However, novel hybrids of these sensors, tagged environments, or new types of sensors may yield future improvements. Indeed, for lens-like techniques to become common, sensor-fusion advances will likely be necessary.

CONCLUSION AND FUTURE WORK

An axiom that drives our work is that everything is best for something and worst for something else. The question is not if one technique is better than another; rather, we want to better understand when to use what, where, why and for whom. Chameleon is an interesting technique, and the concept repeatedly pops up. Yet, our understanding of its subtleties has not grown significantly in the 20 years since it was first shown. Yet in those years, both technologies and tasks have developed to the point where there may be a practical place for this technique in our repertoire. Our hope is that this study advances the field towards such understanding, and provides meaningful guidance as to possible future steps and how (and how not) to pursue them.

REFERENCES

1. Britton, E., Lipscomb, J., Pique, M., Making Nested Rotations Convenient for the User. Computer Graphics, 1978. **12**(3): p. 222-227.
2. Bury, K.F., Boyle, J., Evey, R., Neal, A., Windowing versus scrolling on a visual display terminal. Human Factors, 1982. **24**(4): p. 385-394.
3. Buxton, W., Fitzmaurice, G., HMD's, Caves, & Chameleon: A human-centric analysis of interaction in virtual space. *Computer Graphics, The SIGGRAPH Quarterly*. 1998. p. 64-68.
4. Cauchard, J., et al., m+pSpaces: virtual workspaces in the spatially-aware mobile environment. *MobileHCI 2012*.
5. Chen, X.A., et al., Extending a mobile device's interaction space through body-centric interaction. *MobileHCI 2012*.
6. Chun Yat Li, F., Dearman, D., Truong, K. Virtual Shelves: Interactions with Orientation-Aware Devices. *UIST 2009*.
7. Cohen, P., et al., Synergistic use of direct manipulation and natural language. *CHI 1989*.
8. Fitzmaurice, G., Situated information spaces and spatially aware palmtop computers. Comm. ACM, 1993. **36**(7): p. 38-49.
9. Fitzmaurice, G., Zhai, S., Chignell, M., Virtual reality for palmtop computers. ACM Trans. Inf. Syst., July 1993. **11** (3): p. 197-218.
10. Frisch, M., Heydekorn, J., Dachselt, R. Investigating multi-touch and pen gestures for diagram editing on interactive surfaces. *ITS 2009*.
11. Goel, M., et al., GripSense: Using Built-In Sensors to Detect Hand Posture and Pressure on Commodity Mobile Phones. *UIST 2012*.
12. Guiard, Y., Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. J. Motor Behavior, 1987. **19**(4).
13. Hinckley, K., Pausch, R., Goble, J., Kassell, N., A Survey of Design Issues in Spatial Input. *UIST 1994*.
14. Hinckley, K., Pausch, R., Proffitt, D., Kassell, N., Two-handed virtual manipulation. ACM Trans. Comput.-Hum. Interact., 1998. **5**(3).
15. Hinckley, K., Sinclair, M., Hanson, E., Szeliski, R., Conway, M. The VideoMouse: A Camera-Based Multi-Degree-of-Freedom Input Device. *UIST 1999*.
16. Hinckley, K., et al. Pen + Touch = New Tools. *UIST 2010*.
17. Jacob, R., Sibert, L. The Perceptual Structure of Multidimensional Input Device Selection. *CHI 1992*.
18. Jul, S., Furnas, G. Critical zones in desert fog: aids to multiscale navigation. *UIST 1998*.
19. Liao, C., et al., PACER: Fine-Grained Interactive Paper via Camera-Touch Hybrid Gestures on a Cell Phone. *CHI 2010*.
20. Pierce, J.,. Toolspaces and Glances: Storing, Accessing, and Retrieving Objects in 3D Desktop Applications. *Interactive 3D Graphics*. *1999*.
21. Pique, M.E. Semantics of Interactive Rotation. *Proc. 1986 ACM Workshop on Interactive 3D Graphics*, Chapel Hill, NC.
22. Roudaut, A., et al., TimeTilt: Using Sensor-Based Gestures to Travel through Multiple Applications on a Mobile Device. *Interact 2009*.
23. Tan, D., et al., The Infocockpit: Providing Location and Place to Aid Human Memory. *Workshop on Perceptive User Interfaces*. 2001.
24. Tan, D.S., Stefanucci, J.K., Proffitt, D.R., Pausch, R. Kinesthesis Aids Human Memory. *CHI 2002 Extended Abstracts*.
25. Thorndyke, P.W., Stasz, C., Individual differences in procedures for knowledge acquisition from maps. Cog. Psych., 1980. **12** p. 137-175.
26. Tsang, M., et al., Boom chameleon: simultaneous capture of 3D viewpoint, voice and gesture annotations on a spatially-aware display. *UIST 2002*.
27. Yee, K.-P. Peephole Displays: Pen Interaction on Spatially Aware Handheld Computers. *CHI 2003*.
28. Zhai, S., Milgram, P. Quantifying Coordination in Multiple DOF Movement and Its Application to Evaluating 6 DOF Input. *CHI 1998*.
29. Zhou, F., Duh, H.B.-L., Billinghurst, M. Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. *IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR 2008)*.