



**HAL**  
open science

## Pointing and Beyond: an Operationalization and Preliminary Evaluation of Multi-scale Searching

Emmanuel Pietriga, Caroline Appert, Michel Beaudouin-Lafon

► **To cite this version:**

Emmanuel Pietriga, Caroline Appert, Michel Beaudouin-Lafon. Pointing and Beyond: an Operationalization and Preliminary Evaluation of Multi-scale Searching. CHI '07: Proceedings of the SIGCHI conference on Human Factors in computing systems, Apr 2007, San Jose, CA, United States. 10.1145/1240624.1240808 . inria-00158867

**HAL Id: inria-00158867**

**<https://inria.hal.science/inria-00158867v1>**

Submitted on 30 Jun 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Pointing and Beyond: an Operationalization and Preliminary Evaluation of Multi-scale Searching

Emmanuel Pietriga<sup>1,2</sup>  
emmanuel.pietriga@inria.fr

Caroline Appert<sup>2,1</sup>  
appert@lri.fr

Michel Beaudouin-Lafon<sup>2,1</sup>  
mbl@lri.fr

<sup>1</sup>INRIA  
Bât 490 - Orsay, F-91405, France

<sup>2</sup>LRI - Univ. Paris-Sud & CNRS  
Bât 490 - Orsay, F-91405, France

## ABSTRACT

A number of experimental studies based on domain-specific tasks have evaluated the efficiency of navigation techniques for searching multi-scale worlds. The discrepancies among their results call for a more generic framework similar in spirit to Fitts' reciprocal pointing task, but adapted to a task that significantly differs from pure pointing. We introduce such a framework based on an abstract task and evaluate how four multi-scale navigation techniques perform in one particular multi-scale world configuration. Experimental findings indicate that, in this context, pan & zoom combined with an overview is the most efficient technique of all four, and that focus + context techniques perform better than classical pan & zoom. We relate these findings to more realistic situations, discuss their applicability, and how the framework can be used to cover a broad range of situations.

## Author Keywords

Multi-scale interfaces, Searching task, Controlled Experiment, Focus + Context, Overview + Detail, Zoom

## ACM Classification Keywords

H. Information Systems H.5 Information Interfaces and Presentation H.5.2 User Interfaces (H.1.2, I.3.6)

## INTRODUCTION

Multi-scale interfaces (also called Zoomable User Interfaces or ZUIs) have generated a growing interest over the past decade as a powerful way of representing, navigating and manipulating large sets of data. A number of multi-scale navigation techniques have been designed and implemented, ranging from the original pan & zoom [20] to various focus+context techniques [4, 6, 27]. Up until now, the efficiency of these techniques has been evaluated with two kinds of experimental studies: usability studies based on domain-specific tasks and controlled experiments based on multi-scale versions of Fitts' pointing paradigm.

The usability studies that relied on domain-specific tasks such as searching for items on geographical maps [15], comparing hierarchical data structures [18], or reading textual documents [16] have typically produced inconclusive and sometimes contradictory results. More precisely, experimental findings varied from experiment to experiment, but since application domains varied dramatically, neither these findings can be compared nor generalized. Such results are to be expected since the performance of a given technique is indeed dependent on its context of use [1].

A better understanding of the fundamental aspects of multi-scale navigation could help explain – or even predict – such results, therefore saving valuable time and allowing better exploration of novel techniques. Fitts' pointing paradigm provides such a fundamental tool for exploring and understanding the elementary task of reaching a known target as fast as possible. Originally devised to study pointing in the real world [8], it has been used repeatedly in HCI for evaluating a variety of pointing techniques and devices [3, 2, 23]. Fitts' law has proven remarkably robust, to the point of being used as part of an ISO standard for pointing devices [25]. Fitts' pointing task has also been used with multi-scale interfaces and it has been shown that Fitts' law still applies for pointing targets with pan & zoom [10]. In particular, it has been shown that Fitts' paradigm could address navigation, not just pointing, in interfaces that require scrolling [14] or zooming [10].

While Fitts' pointing paradigm is very powerful, it models a very specific task: that of reaching a target whose location is known to the user. However, this scenario only captures one of several navigation tasks in multi-scale worlds. Users might only have partial information about the target's location and appearance, thus requiring them to search for potential targets and get more details about each one until the actual target is identified. Consider for example a user searching for Brisbane on a multi-scale world map, only knowing that it is a large Australian city. The strategy first consists in zooming towards Australia to then inspect each large city one by one, zooming in to discover that it is not the right one, zooming out, maybe as far as the whole continent, and zooming in to the next potential target until the right city is found. Exploring large spaces in search of a particular target differs from pure pointing, as it requires users to perform additional motor actions to identify the target.

In the same way as Fitts' reciprocal pointing task operationalizes the task of reaching a known target, we propose in this paper to operationalize the above search task in a way that is easily amenable to controlled experiments. We then evaluate how four multi-scale navigation techniques perform in one particular configuration of a multi-scale world: classical pan & zoom, overview + detail, and two focus + context techniques, namely distortion (graphical fisheye) lenses and a variation on the DragMag image magnifier which, to the best of our knowledge, has not yet been evaluated. Our results indicate that in this context overview + detail outperforms the other three, and that the two focus + context techniques outperform classical pan & zoom. However, this multi-scale world configuration is only one particular case in a range of situations. We discuss the limits of this preliminary study, describe the design space that is covered by our abstract search task and present an environment that we have developed to help explore this design space.

## RELATED WORK

A number of experimental studies have compared the performance of different multi-scale navigation techniques and have reported contrasted results. Classical pan & zoom was compared with fisheye and overview + detail on high-level cognitive tasks involving electronic documents [16]: writing an essay after having explored a document, and finding answers to questions within that document. Classical pan & zoom was the least efficient technique; participants read faster using the fisheye; they wrote better essays using the overview + detail, but took more time to answer questions. In North and Shneiderman's experiment [19], participants had to browse the database of U.S. states and counties to answer questions using a detail-only scrollable interface or an overview + detail interface. The overview + detail interface outperformed the detail-only interface by 30-80% depending on the task. However, in another study [18], pan & zoom or overview + detail were not significantly different when participants had to navigate a large node-link representation and make topological comparisons. On the contrary, Hornbaeck et al. [15] have reported that their overview + detail interface was more efficient than pan & zoom. This, however, was true for only one of the two geographical maps that participants had to explore in their experiment.

The findings of these experiments show that the use of domain-dependent tasks makes it difficult to get consistent results that can be generalized, even more so when they require a significant amount of cognitive effort from the participants. Identifying and isolating lower-level, domain-independent tasks can help reach more generalizable results. From a motor perspective, one recurring task performed by users of multi-scale interfaces is to search for targets among sets of objects by navigating through space and scale. This article describes an experimental setup for the controlled evaluation of various interaction techniques considered as appropriate for the task of searching a multi-scale world.

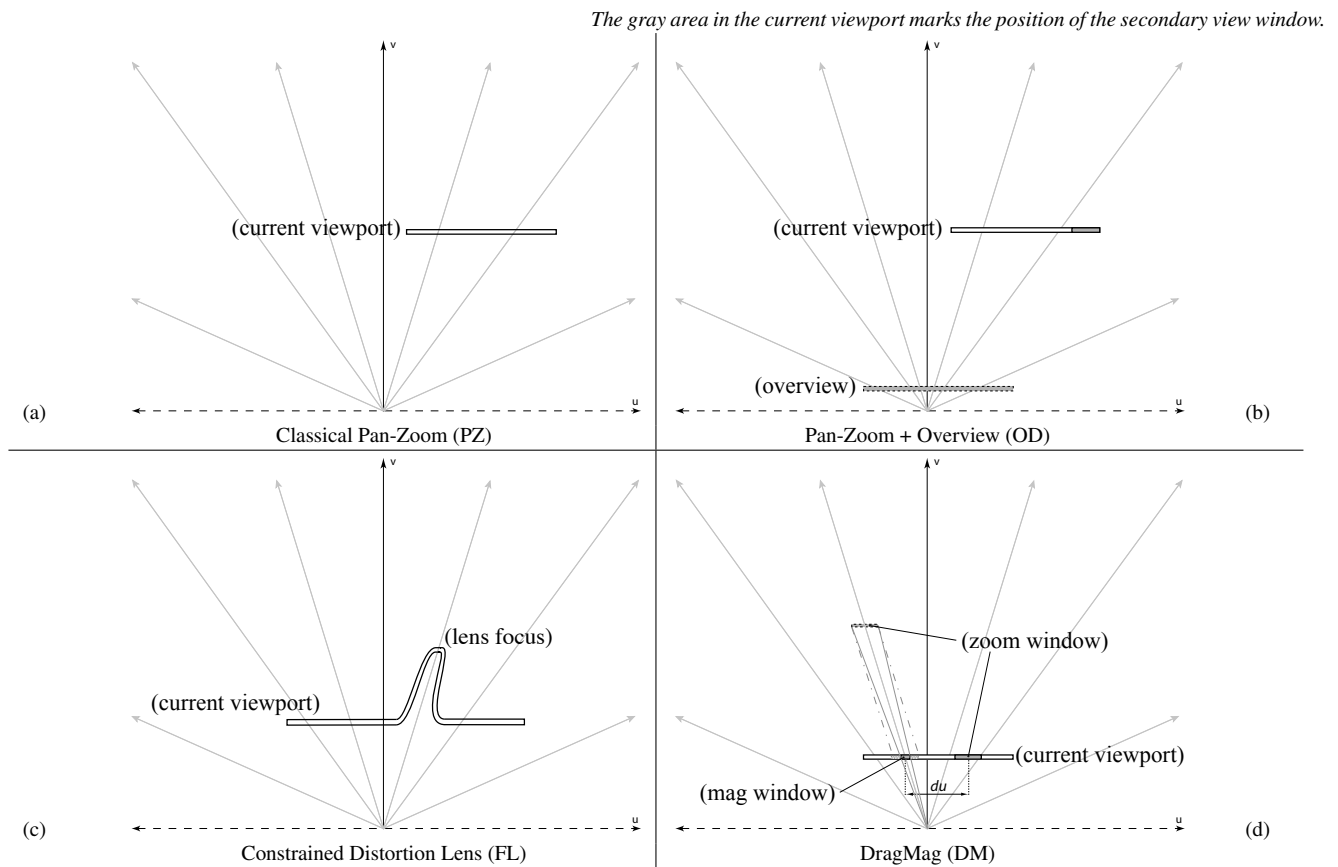
## OPERATIONALIZING MULTI-SCALE SEARCHING

Studying a task through a controlled experiment requires operationalizing it, i.e., defining it as a function of variables

of interest (independent variables) that researchers can act upon to collect measures (dependent variables). The pointing task is a well-known example in the field of HCI, initially operationalized in psychology by Fitts [8]: to study the performance of pointing techniques, researchers act on the index of difficulty (ID) variable and measure the movement time on a reciprocal pointing task. We seek to operationalize multi-scale searching in a similar way. In a multi-scale world, users navigate and look at objects until they find the target. Users have to navigate in both space and scale to a position that reveals enough details about each object, in order to decide whether it is the target or a distractor. Initially, users make a blind choice of a "potential target" at a high scale and navigate to it to acquire enough information. If it is a distractor, they have to navigate to another object, typically by zooming-out, panning, then zooming-in [10].

Since we are interested in studying the performance in time and error rate to find a target according to the required "quantity" of exploration from a purely motor perspective, we abstract the representation from any semantic or topological relationship among objects that could help participants identify the target in an uncontrolled manner (for instance, having reasonable knowledge of the geography of Russia could help locate Saint Petersburg once Moscow has been found on a map, or knowing that Chicago is on the shores of lake Michigan would reduce the area to be explored significantly). To quantify exploration, our experimental setup consists of a multi-scale world containing a set of  $n$  objects, one of them being the target and the others distractors. We define the "quantity" of exploration as the number  $k$  of distractors that users have to visit before finding the target: the larger the number of visited distractors, the larger the quantity of exploration.  $k$  is probabilistically dependent on  $n$ : the larger the number of objects, the higher the probability of having a large number of objects to visit before reaching the target. We control this parameter by forcing participants to visit a predefined number of objects before finding the target; if we chose a priori which object is the target, participants could find it immediately by chance, or on the contrary they could spend a lot of time searching for it, and this uncontrolled factor would have a significant impact on our measurements. We design our experiment to ensure that the target is the  $k^{th}$  object visited, no matter the order of exploration chosen by each participant. The system thus has to know i) when objects are seen by the participant, and ii) whether or not enough detail is displayed about these objects in order to differentiate the target from distractors. Making the system aware of these two pieces of information in a fully reliable manner requires answering the following two questions:

- What minimal scale provides enough information? This depends on visual acuity, which is user-dependent.
- In which region of the screen and for how long should an object be displayed to consider it seen? Assuming that the user visually scans the whole screen is too strong an hypothesis, and probably an unfounded one. Also, if only part of an object is in the viewport, the system cannot know for sure whether or not the user has seen it.



**Figure 1. Representation of multi-scale interaction techniques in space-scale diagrams.**

We address these problems as follows. First, we set a minimum scale (*minScale*) at which the user can collect enough information to detect a target: all objects seem identical except for the target, which reveals a different piece of information when displayed at or above *minScale*. In order to avoid differences among participants, all objects are displayed identically at all scales until the user explicitly asks to reveal the disambiguating piece of information. This explicit “unveiling” action is available only when the scale is *minScale* or more. Second, we make sure that the user cannot reveal several objects simultaneously. Once an object has been revealed, the user has to process the information and, provided that the object is the target, take an additional explicit action to tell the system that this object is the target.

While we cannot be sure that participants actually look at targets when unveiling them, it is in their own interest to do so in order to perform the search task as fast as possible. Therefore we believe that this design operationalizes a realistic search task without having to use more complex devices such as eye trackers. Before presenting a first experiment based on this task, we introduce the multi-scale navigation techniques that we have tested.

### MULTI-SCALE INTERACTION TECHNIQUES

Many representation and navigation techniques have been proposed to interact with multi-scale worlds, some being variations on others. For our study, we narrowed down the possibilities to four techniques, chosen to be representative

of the most widespread and/or efficient techniques currently available. Figure 1 illustrates these techniques using space-scale diagrams [9].

The first technique is the classical *Pan & Zoom* (Figure 1-a). In order to get more detail about specific elements of the representation, users have to move the entire viewport both in space and scale, respectively by panning and zooming. No contextual information is provided; this method is therefore prone to user disorientation.

One way to address disorientation consists in using overview + detail techniques. One such technique, *Pan & Zoom + Overview* (Figure 1-b), enhances classical *Pan & Zoom* by providing users with an inset containing a representation of the region surrounding the area currently seen in the main viewport at a lower scale. The overview is located inside the main viewport, typically in one of the four corners. The goal is to minimize the visual interference caused by occlusion of the elements in focus, but this introduces the problem of divided attention [22].

In overview + detail representations, more screen real-estate is dedicated to the focus than to the context. Conversely, focus + context techniques allocate more screen real-estate to the context than to the focus. We selected two techniques that we consider relevant to the multi-scale searching task: constrained distortion lenses [6] and a variation on the original DragMag Image Magnifier [27].

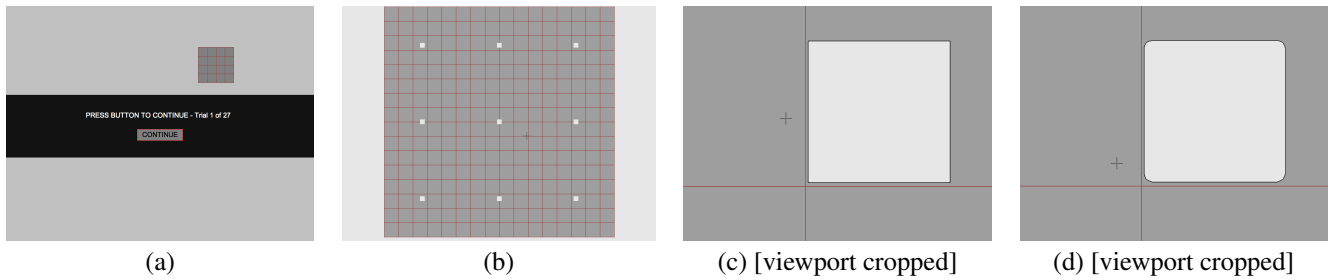


Figure 2. Storyboard: (a) start of trial, (b) navigation to the set of objects, (c) inspection of an object (before unveiling), (d) after unveiling the target

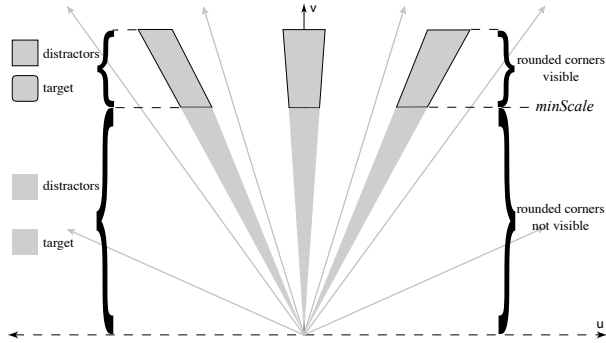


Figure 3. Space-scale diagram of the scene used in the experiment

*Constrained distortion lenses* (Figure 1-c) provide a detail-in-context representation through the local magnification of a region of the screen (the focus of attention). This focus region is integrated in the surrounding context by distorting the representation in the transition region. The distortion is defined by a drop-off function (see [6] for more details). We chose a Gaussian profile as it provides a smooth transition between focus and distortion, and between distortion and context. Our lens also features a flat top because many tasks require the focal region not to be distorted [5]. The *in situ* magnification of these lenses solves the problem of divided attention but introduces a distortion that can cause recognition problems.

The *DragMag* (Figure 1-d) can be considered a special case of fisheye lens often called Manhattan lens, featuring a perpendicular drop-off function. There is no distorted region between the focus and the context, but as a result the region immediately surrounding the focus is occluded. To address this issue, the focus region is translated by a user-controlled offset ( $du$  in Figure 1-d). This results in the occlusion of another region of the context, which is often considered less important than the immediate surroundings of the focus. However, this reintroduces the problem of divided attention encountered with overview + detail representations, and the occlusion can be more cumbersome to handle than with the overview.

## EXPERIMENT

We conducted a 4x9 within-subject controlled experiment to compare the efficiency of these four techniques on one instantiation of the multi-scale search task introduced earlier.

## Task

The task consisted in finding a target among a set of objects as quick as possible while minimizing the number of errors. The virtual scene contained nine light gray squares organized into a 3x3 grid layout and embedded inside a large, darker gray square. We used a grid layout so participants would easily know where the potential targets were. This regular layout also prevented performance to be biased by uneven traveled distances between trials of the same rank  $k$ . A dark red grid was superimposed on the display in order to minimize desert-fog [17] (see Figure 2-b). The grid was adaptive to scale, i.e., new grid lines would fade in when zooming in and some grid lines would fade out when zooming out so that the display would always contain a reasonable number of grid lines. All nine objects had square corners except for the target which had rounded corners. The rounded corners could only be seen when the target was displayed at a large enough scale, called *minScale* (Figure 3). Participants thus had to zoom in onto each square in order to find out whether it was the actual target or not. Zooming in was not sufficient however: once *minScale* was reached, a black border was displayed around the square in focus. Participants could then use the space bar to unveil the object: this would permanently reveal whether the object was the target (round corners) or not. Note that this “unveiling” step does not affect the ecological validity of our task since it penalizes all techniques equally.

Figure 2 shows a storyboard of the task: participants started each trial by pressing a button located at the center of the screen (Figure 2-a). The view was initialized so that the region containing potential targets (dark gray area) was not centered on the screen, requiring participants to reach the region by panning and zooming. The goal was both to better simulate a multi-scale navigation & search task and to avoid a learning effect with respect to the participant’s initial move. Participants had to navigate to that region (Figure 2-b) and then inspect each object more closely by magnifying it using the current navigation technique (Figure 2-c). Participants were allowed to zoom-in further, but zooming in too far would have the object fill the display and make it impossible to find out if it was the target. Once *minScale* was reached for an object, participants could unveil that object by pressing the space bar. The object’s border flashed green for 400 milliseconds, informing the participant that the object had actually been unveiled. If the object’s corners remained square, this meant that the object was not the target and participants had to navigate to the next potential target using the current navigation technique. If, on the contrary, the object’s

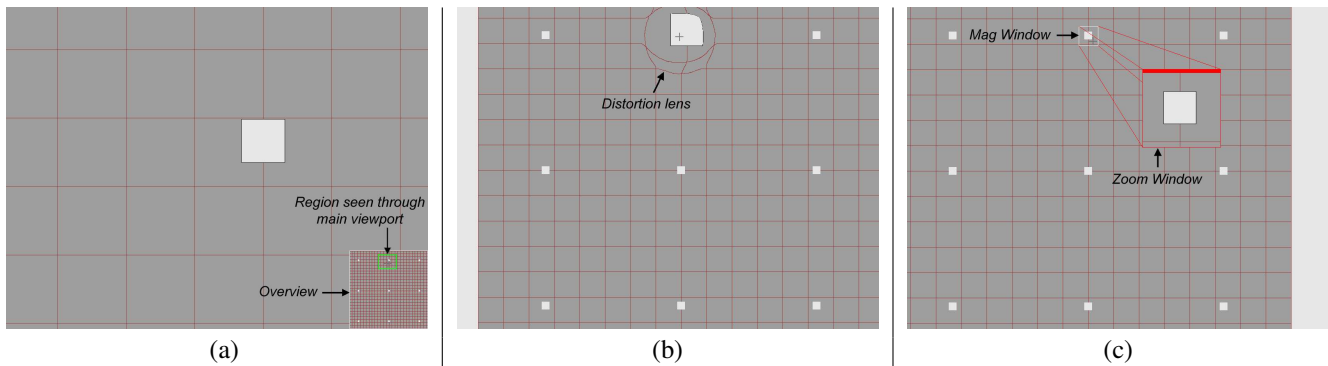


Figure 4. Unveiling an object with (a) pan-zoom + overview [OD], (b) a constrained distortion lens [FL], (c) a DragMag [DM]

corners became rounded (Figure 2-d), participants had to hit the `F1` key to tell the system that they had identified the target and end the trial. Note that figures 2-c and 2-d are cropped versions of the viewport, aimed at illustrating the actual display size of objects at *minScale* on the monitor used for the experiment.

Participants were instructed to go as fast as possible to complete a trial (i.e., between hitting the `Continue` button and hitting `F1`), but they were allowed to rest between successive trials. They were also instructed to minimize the number of visits to the same object and the number of misses, i.e., hitting `F1` when the object was not the target. Such misses terminated the trial and were counted as errors.

### Techniques

The first independent variable we manipulated in our experiment was the technique. The first technique was pan & zoom (PZ). Participants could pan the view by moving the mouse while holding the left mouse button, and zoom in/out by rotating the mouse wheel. These three degrees of freedom could be controlled simultaneously. The magnification factor per wheel step was tuned so as to get an average zooming speed of 8x per second, as advocated in [15]. With this technique, participants panned & zoomed the entire view to get enough details about each object. Each of the other three techniques allowed participants to pan & zoom using the above commands.

The second technique was overview + detail (OD). The region seen through the main viewport was represented by a bright green rectangle in the inset containing the overview (see Figure 4-a). This rectangle could be dragged, resulting in changing the content of the main viewport. With these additional two degrees of freedom, participants could do fine-grain panning in the main viewport and coarse-grain panning in the overview. The representation in the overview was dynamic: it was not necessarily showing all objects in the virtual world, as it followed the camera associated with the detailed view in space and scale when the scale difference between the overview and the detailed view was larger than a factor of 24. The overview implemented by Google Maps<sup>1</sup> demonstrates such a behavior.

<sup>1</sup><http://maps.google.com>

The third technique featured a constrained distortion lens, also called graphical fisheye lens (FL). It allowed for magnification of the region around the mouse cursor (see Figure 4-b). We used a 100-pixel radial lens defined by a Gaussian drop-off function and the  $L(2)$  distance metric [6] with a 60-pixel radius flat top. The lens was not activated at the start of a trial. Participants could activate it by clicking the left mouse button, and deactivate it by clicking the right mouse button. The lens was always centered on the mouse cursor. When the lens was active, participants were still able to pan the context by dragging outside the lens with the left mouse button. The default magnification factor within the flat top was set to 4 times the scale factor of the context (the scale factor in the lens focus is always defined relative to that of the context, since the context can itself be panned and zoomed). Participants could change the lens' magnification by using the mouse wheel, within the limits of twice and twelve times the scale factor of the context. This technique therefore featured five degrees of freedom (2D panning of context, 2D panning of lens focus, and either the lens magnification factor or the context scale factor depending on whether the lens is active or not). Lens activation and deactivation were both animated by smoothly increasing the lens' magnification factor from 1.0 to its default value (4.0) over a period of 300 milliseconds for the sake of perceptual continuity [24]. The lens thus seemed to “emerge” from the flat surface when activated, and flatten itself when deactivated.

The last technique was inspired by the DragMag Image Magnifier (DM), but interaction with the windows differed significantly from the original prototypes [27]. Figure 4-c shows the two windows composing the DragMag: the **mag window** outlines the region magnified in the **zoom window**. Participants could activate and deactivate the DragMag by clicking on the right mouse button. The mag window would then appear centered around the mouse cursor, the zoom window being offset by a default distance of 200 pixels to the southeast of the mag window. As with the previous technique, both DragMag activation and deactivation were smoothly animated over a period of 300 milliseconds, with the zoom window “emerging” from the mag window. Participants could drag the mag region, thus changing the content of the zoom window; they could also drag “through” the zoom window for small scale adjustments, though this feature was not very useful in the context of the experiment.

Participants could also move the zoom window by dragging the thick bar at its top. This feature was useful to reveal objects occluded by the zoom window. The mouse wheel was used to control magnification. Operating the mouse wheel while the cursor was in the zoom window controlled that window’s magnification factor. Operating the mouse wheel anywhere outside this window controlled the scale of the context. The technique therefore featured six degrees of freedom (the context scale factor and the zoom window magnification factor could both be controlled when the DragMag was active). The default magnification factor in the zoom window was 4 times the scale factor of the context, as for the distortion lens. The zoom window was not resizable.

For the purpose of comparing the techniques, the overview of OD, the lens of FL, and the zoom window of DM all used the same amount of screen real-estate: a 200 x 200 pixels region, which represented 4.5% of the total available display area.

### Predictions

Our predictions were as follows:

- **Time is linearly dependent on the rank  $k$  of the target.** We hypothesized that, whatever the technique, the user has to navigate to inspect objects one by one and that each navigation incurs the same cost. Since the cost of revisiting an object is fairly high, we hypothesized that the number of revisits would be very small. Therefore the overall task completion time should be linearly dependent on the “quantity” of exploration, i.e., the target’s rank  $k$  in the sequence of visited objects.
- **Focus + Context (FL, DM) and Overview + Detail (OD) outperform classical Pan & Zoom (PZ).** With PZ, navigating from one object to the next typically consists in zooming out to acquire the next object then zooming in and panning to magnify it. With DM, FL and OD, it simply consists in moving the focus onto the next object. Since the position of the focus can be controlled from the context, we hypothesized that the zoom-out/pan/zoom-in sequence of PZ would take more time than relocating the focus within the context.
- **Overview + Detail (OD) outperforms Focus + Context (DM and FL).** With OD, DM and FL, navigating from object  $a$  to object  $b$  consists in moving the focus from  $a$  to  $b$ . This movement can be seen as a pointing task. With DM and FL, pointing is achieved by relocating the focus (i.e., the DragMag window or the lens’ focus region) while with OD, pointing is achieved by relocating the detailed view. According to Guiard et al. [11], such pointing tasks are view pointing tasks whose Index of Difficulty depends on view size. Since the detailed view is significantly larger than the lens’ focus and the DragMag’s zoom window, we predicted that OD would outperform the two Focus + Context techniques (FL, DM).

### Participants

Twelve unpaid adult volunteers, 11 males and 1 female, ranging from 23 to 52 years old (28 on average, with a median of

25.5), served in the experiment. Before starting, the experimenter checked that they could perceive the rounded corners at *minScale*, showing them squares with squared and rounded corners successively. The experiment was divided into four blocks, one block per technique. Before each block, participants were shown how to achieve the task using the corresponding technique. They were then asked to practice on randomly-chosen trials until they felt comfortable with the technique. The experimenter observed participants and encouraged them to keep practicing until they were familiar enough with the technique.

### Apparatus

We used a Dell Precision 380 equipped with a 3 GHz Pentium D processor, an NVidia Quadro FX4500 graphics card, a 1280x1024 LCD monitor (19”) and a Dell optical mouse with a scroll wheel. The program was written in Java 1.5 using the open source ZVTM toolkit [21] which features a wide range of multi-scale interaction techniques, thanks to different types of portals [20] and arbitrary distortion lenses<sup>2</sup>. The application was limited to a 1080x824 window with a black padding of 100 pixels in order to accommodate instruction messages and simulate screen real-estate that would usually be taken by control and information widgets.

### Counterbalancing strategy

We used a 9x4 within-subject design: we tested 9 target ranks ( $k \in [1..9]$ ) for the 4 techniques (PZ, FL, DM and OD), i.e.,  $9 * 4 = 36$  conditions. Each condition was replicated 3 times so that each participant performed  $9 * 4 * 3 = 108$  trials ( $\approx 45$  minutes). The initial position of the area containing the objects was different for each of these 3 replications and was counterbalanced among blocks with a Latin square. We grouped the trials into 4 blocks, one block per technique, to minimize negative skill transfers. To minimize ordering effects, we computed four different technique orders using a Latin square and composed 4 groups of 3 participants ( $G_1, G_2, G_3, G_4$ ), one group per ordering.

We also counterbalanced the presentation order of the different values of  $k$  within a block: we used a Latin square to compute 9 possible orders for presenting the values of  $k$  and concatenated 3 orders to compose a block (3 orders of 9 trials = 27 trials per block). Three block compositions ( $bc_1, bc_2, bc_3$ ) were obtained through a Latin square. We mapped one block composition per participant within a group. Table 1 summarizes our counterbalancing strategy among participants. While we told participants that the target was selected randomly by the program, this was not, in fact, the case: instead, the program counted the objects being visited by the participant during the trial, and displayed the target when the  $k^{th}$  object was unveiled by the participant. This allowed us to fully control the rank variable. Note that even if the participants had known (or guessed) the actual working of the program, this would not have given them any advantage.

<sup>2</sup>The content of the lens is not a mere magnification of the original pixels, but an actual high-resolution separate rendering of the region seen through the lens, which provides more details. This mechanism also makes it possible to use semantic zooming inside the lens (see Figure 6).

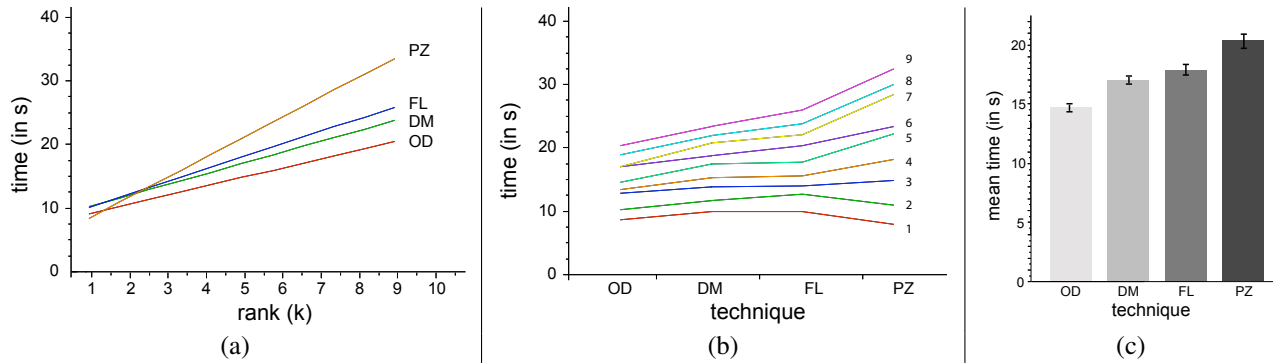


Figure 5. Fit lines for the four techniques (a), Interaction effects on time (in s) for  $rank * technique$  (b), Mean completion time per technique (c)

	$G_1$	$G_2$	$G_3$	$G_4$
$bc_1$	$S_1$	$S_4$	$S_7$	$S_{10}$
$bc_2$	$S_2$	$S_5$	$S_8$	$S_{11}$
$bc_3$	$S_3$	$S_6$	$S_9$	$S_{12}$

Table 1. Counterbalancing strategy for the 12 participants ( $S_i$ ).

	PZ	FL	DM	OD
$r^2$	0.84	0.84	0.81	0.80
$a$	3.2	2.0	1.7	1.3
$b$	5.1	8.2	8.5	7.8

Table 2. Correlation coefficients ( $r^2$ ) and coefficients  $a$  and  $b$  ( $time = a * rank + b$ ) for the four techniques.

## Results

For each trial, the program collected the completion time, whether it was a hit or a miss, the order of visit of each object and the time at which it was unveiled. It also logged cinematic data from the cameras associated with the focus and context viewports. We also collected the participants' preferences among the techniques in a post-hoc test.

For our analyses, we first removed 14 *miss* trials (about 1%) and then 31 outliers (about 2,5% of the *hit* trials). We verified that misses and outliers were randomly distributed across participants, techniques and ranks and that there was no effect of technique presentation order on time. Learning effects were not significant for PZ ( $p = 0.42$ ) and FL ( $p = 0.75$ ), and were significant but moderate for DM ( $p = 0.03$ ) and OD ( $p = 0.02$ ).

We isolated the *rank* variable ( $k$ ) by analyzing it separately for each *technique*. We computed the linear regression of time relative to the rank, treating participants as a random variable. We obtained the high correlation coefficients listed in Table 2. This supports our first prediction: completion time is linearly dependent on the rank (see Figure 5-a).

As expected, the number of revisits was extremely low (less than 1 revisit on average for each technique) and participants optimized the order in which they visited the objects so as to minimize traveled distance. Most participants explored the objects following an S-shaped pattern, some used a spiral; very few made diagonal moves, except for one participant who adopted a very erratic search pattern across all blocks (his results were nevertheless consistent with our overall findings). Table 2 also reports slopes ( $a$ ) and intercepts with the y-axis ( $b$ ) for each linear regression. We note that the value of  $b$  is lower for PZ. The cinematic logs explain this difference: with DM, FL and OD, participants initially spent more time adjusting the scale and position in order to optimize their future interactions. Indeed, with these techniques,

good position of the context allows participants to only pan the detailed view through the overview or the focus from the context without having to adjust the scale.

Since we have evidence that time is linearly dependent on *rank*, we now analyze *rank* as a continuous factor. Analysis of variance with the REML method for repeated measures revealed a significant simple effect on time for *rank*, i.e.  $k$ , ( $F_{1,411} = 1500.5$ ,  $p < 0.0001$ ), a significant simple effect on time for *technique* ( $F_{3,411} = 91.6$ ,  $p < 0.0001$ ) and a significant interaction effect on time for  $rank * technique$  ( $F_{3,411} = 54.2$ ,  $p < 0.0001$ ). Figure 5-b illustrates these results: the larger the rank, the larger the differences among techniques. Tukey post-hoc tests reveal that each technique is significantly different from the others: OD is the most efficient technique, followed by FL, then DM and finally PZ ( $OD > DM > FL > PZ$ ). This supports our second and third predictions: the Overview + Detail technique outperforms the two Focus + Context techniques, which themselves outperform classical Pan & Zoom. We believe the lower performance of FL, compared with DM, could be due to the visual distortion introduced by the lens [13]. We note however that the difference between the means of these two techniques ( $FL_{mean} = 18$  s.,  $DM_{mean} = 17$  s.) is much smaller than with the other two ( $OD_{mean} = 14.8$  s.,  $PZ_{mean} = 21.2$  s.), as shown in Figure 5-c.

The subjective preferences we collected in the post-hoc questionnaire match these results. At the end of the experiment, participants were asked to rank the techniques according to their preference: 11 ranked PZ as the worst technique, and 9 ranked OD as the best technique.

## DISCUSSION AND FUTURE WORK

The search task introduced in this article covers a range of situations where the user has to explore each potential target in a multi-scale environment. Unlike the tasks tested in us-



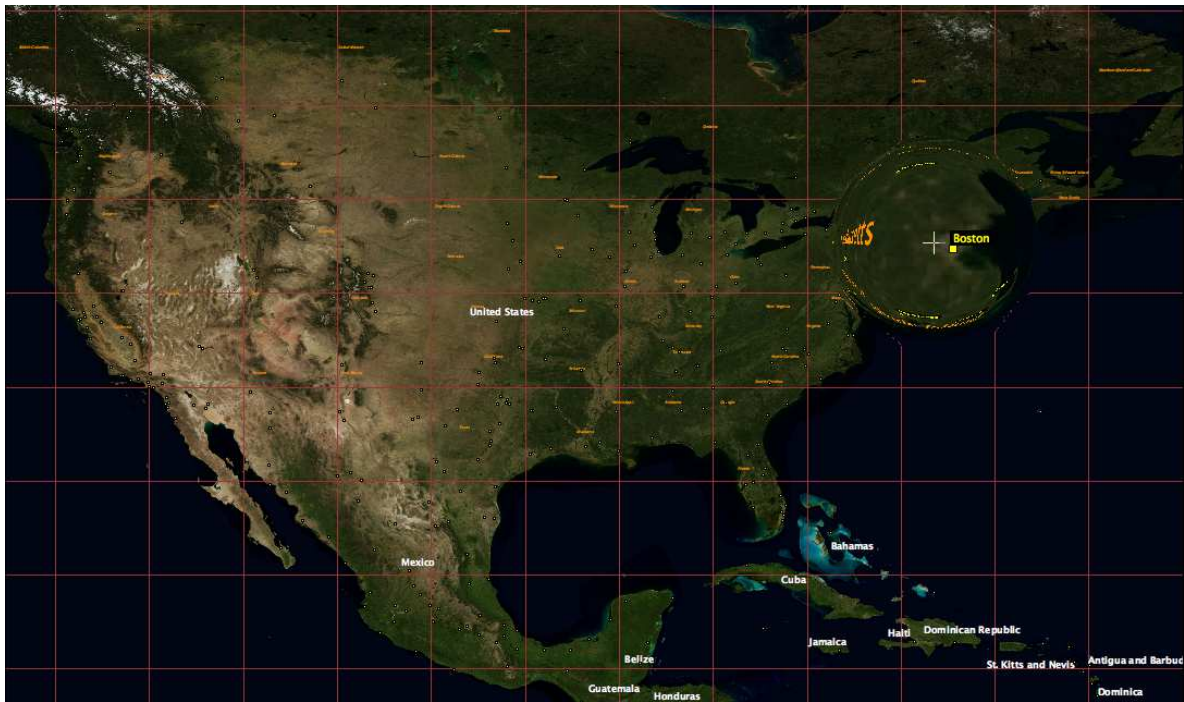


Figure 6. Looking for Boston, Massachusetts, USA with a distortion lens on a multi-scale world map

ability studies, we focus on the motor and perceptual skills and try to exclude the cognitive skills involved in searching. The goal is similar to that of Fitts' pointing paradigm and its use in HCI: to assess the limit performance of searching multi-scale worlds and to come up with predictive performance models and novel navigation techniques that improve multi-scale searching.

Our search task covers a large design space whose main dimensions are the amount of information the user has to acquire in order to decide which object is the target and the structure of the multi-scale world. Our experiment tested an extreme situation in this design space. First, the user had to look in detail at each target by navigating to it, therefore excluding the kind of visual search that occurs, e.g., in a Fitts' pointing task with distractors. Second, we used a specific configuration of the multi-scale world: a "small-world", i.e. an environment in which there exists at least one viewpoint from which all objects can be seen, that contained objects of the same relative size, i.e. same *minScale*, laid out uniformly on a grid. Therefore, the results reported in the previous section cannot be generalized to all search tasks and we need to devise a strategy to explore the design space and operationalize other situations.

Unfortunately, few theoretical models are available to help us structure this design space. While Guiard et al.'s degree of goal-directedness [12] could help quantify the amount of information that users need to recognize a target and Furnas & Bederson's space-scale diagrams [9] could help explore the structure of multi-scale worlds, neither approach is readily applicable to identify relevant points in this design space. Therefore we have developed an environment for testing re-

alistic multi-scale navigation and searching tasks in order to inform our design process.

This environment (see Figure 6) displays a multi-scale version of NASA's Blue Marble Next Generation world map [26] overlaid with geographical features such as countries, states, cities, parks and lakes. The map is 80000x40000 pixels at full resolution, or about 80x40 regular-size screens. The geographical features can be any set of localized items found in the Geo-Names<sup>3</sup> on-line database. The environment provides a set of navigation techniques, including those tested in the study reported in this article. A variant of this environment was used to run the experiment reported in the previous section. Both versions are implemented with the ZVTM toolkit [21] and are publicly available<sup>4</sup>.

We conducted several pilot studies with this environment using a set of 1825 cities, 63 states and provinces, and 192 countries. Participants were asked to search for geographical features by locating first the country, then possibly the state or province and finally the city. Obviously, this task relies on cognitive skills such as the participant's geographical knowledge or contextual hints such as large water bodies. It was extremely useful however for observing users and collecting quantitative data and subjective evaluations and helped us identify interesting multi-scale world configurations.

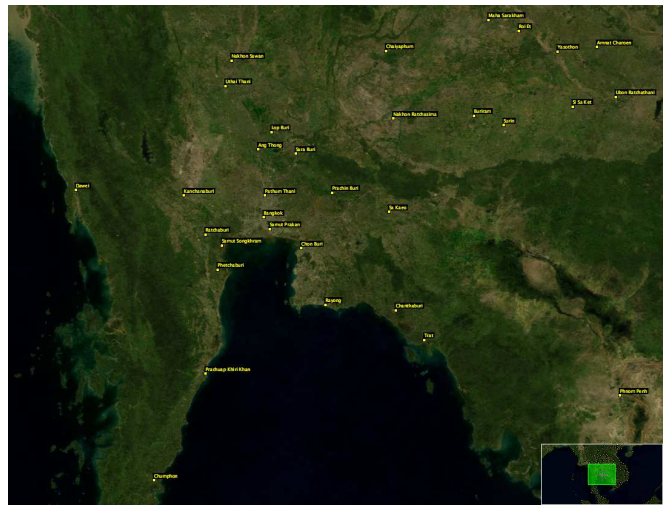
For example, the configuration that we tested in the experiment described in the previous section corresponds to, e.g., finding a large city in Australia. Since there are only eight

<sup>3</sup><http://www.geonames.org>

<sup>4</sup><http://zvtm.sourceforge.net/eval/pb>



(a)



(b)

Figure 7. Exploring a sparse region with a drag-mag (a), and a dense region with an overview (b)

large cities spread over the whole continent (see Figure 7-a), the participant who does not know the geography of Australia has to zoom in onto each city. So the task consists in finding a city among a relatively small, well-identified, set of objects of the same relative size.

In this context, participants found the most useful technique to be Overview + Detail, followed by the constrained distortion lens and the DragMag. For the latter two, the commonly adopted search technique consisted in panning & zooming to make the entire continent fit the viewport (all cities could be seen from this altitude, though their names were not visible), and then activate a lens or DragMag to inspect the potential targets while keeping the context fixed. The same behavior was observed with the abstract task, as reported earlier. It is interesting to note however that the negative effects of distortion were less frequently mentioned for the geographical task than for the abstract task, probably because continuous representations such as world maps withstand distortion better than other types of representations, at least for searching.

Other observations of the participants' behavior with the geographical task have helped us identify situations that seem interesting for subsequent experiments. For instance, densely populated regions such as mainland Southeast Asia (see Figure 7-b), which feature many cities, were most commonly explored with the Overview + Detail technique because the main viewport can accommodate more cities at the scale where their names become readable (the equivalent of *minScale* defined in the abstract task), thus facilitating visual scanning.

These behavior patterns lead us to hypothesize that Overview + Detail techniques work better when exploring dense regions while Focus + Context techniques are also efficient when searching for a target among a sparse set. This may be due to the fact that visual scanning plays an important role in the former case while motor actions take precedence

over visual scanning in the latter, at least within the limits of the magnification factor of graphical fisheye lenses (usually 4 and rarely more than 8 [7]). Providing empirical evidence for this claim requires running more experiments within the framework by varying parameters such as density. Another area for future work is to test configurations in which objects have different *minScale* values, corresponding to situations where users have very limited information about the target, including the scale at which it is visible. Since such situations presumably prompt for more zooming actions than the one we tested, it is possible that the best navigation technique would be different.

## SUMMARY

This paper has introduced a new framework based on an abstract searching task for multi-scale interfaces that operationalizes the situation where one has to look for the target before selecting it. We have used this framework to compare four multi-scale navigation techniques in the context of one specific multi-scale world configuration (small world, uniformly dense layout), showing that in this case a fixed overview afforded better performance than Focus + Context techniques and that traditional pan-and-zoom was the worst. These results cannot be immediately generalized to all multi-scale world configurations, and additional evaluations are required to cover a broader range of situations by varying parameters such as density, topology and the relative size of targets. Our framework allows for the systematic exploration of this design space. Moreover, the geographical environment we have developed can help identify interesting situations and formulate hypotheses about them. These situations can then easily be translated into configurations of the abstract task and tested with controlled experiments.

## Acknowledgements

We wish to thank the anonymous reviewers for their insightful reviews, as well as all the volunteers who participated in our experiments.

## REFERENCES

1. C. Appert, M. Beaudouin-Lafon, and W. Mackay. Context matters: Evaluating interaction techniques with the CIS model. In *Proc. of HCI 2004, Leeds, UK*, pages 279–295. Springer Verlag, Sept. 2004.
2. C. Appert and J.-D. Fekete. OrthoZoom scroller: 1D multi-scale navigation. In *CHI '06: Proc. Human Factors in Computing Systems*, pages 21–30. ACM Press, 2006.
3. R. Balakrishnan. "Beating" Fitts' law: virtual enhancements for pointing facilitation. *Int. J. Hum.-Comput. Stud.*, 61(6):857–874, 2004.
4. S. K. Card, J. D. Mackinlay, and B. Shneiderman, editors. *Readings in information visualization: using vision to think*. Morgan Kaufmann Publish. Inc., 1999.
5. M. S. T. Carpendale, D. J. Cowperthwaite, and F. D. Fracchia. Making distortions comprehensible. *IEEE Symposium on Visual Languages*, pages 36–45, 1997.
6. M. S. T. Carpendale and C. Montagnese. A framework for unifying presentation space. In *UIST '01: Proc. ACM Symposium on User Interface Software and Technology*, pages 61–70. ACM Press, 2001.
7. S. Carpendale, J. Ligh, and E. Pattison. Achieving higher magnification in context. In *UIST '04: Proc. ACM symposium on User Interface Software and Technology*, pages 71–80. ACM Press, 2004.
8. P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psych.*, 47:381–391, 1954.
9. G. W. Furnas and B. B. Bederson. Space-scale diagrams: understanding multiscale interfaces. In *CHI '95: Proc. Human Factors in Computing Systems*, pages 234–241. ACM Press/Addison-Wesley, 1995.
10. Y. Guiard and M. Beaudouin-Lafon. Target acquisition in multiscale electronic worlds. *Int. J. Hum.-Comput. Stud.*, 61(6):875–905, Dec. 2004.
11. Y. Guiard, M. Beaudouin-Lafon, J. Bastin, D. Pasveer, and S. Zhai. View size and pointing difficulty in multi-scale navigation. In *AVI '04: Proc. working conference on Advanced Visual Interfaces*, pages 117–124. ACM Press, 2004.
12. Y. Guiard, Y. Du, and O. Chapuis. Quantifying degree of goal directedness in document navigation: Application to the evaluation of the perspective-drag technique. In *CHI '07: Proc. Human Factors in Computing Systems*. ACM Press, 2007. This volume.
13. C. Gutwin. Improving focus targeting in interactive fisheye views. In *CHI '02: Proc. Human Factors in Computing Systems*, pages 267–274. ACM Press, 2002.
14. K. Hinckley, E. Cutrell, S. Bathiche, and T. Muss. Quantitative analysis of scrolling techniques. In *CHI '02: Proc. Human Factors in Computing Systems*, pages 65–72. ACM Press, 2002.
15. K. Hornbæk, B. B. Bederson, and C. Plaisant. Navigation patterns and usability of zoomable user interfaces with and without an overview. *ACM Trans. Comput.-Hum. Interact.*, 9(4):362–389, 2002.
16. K. Hornbæk and E. Frøkjær. Reading of electronic documents: the usability of linear, fisheye, and overview+detail interfaces. In *CHI '01: Proc. Human Factors in Computing Systems*, pages 293–300. ACM Press, 2001.
17. S. Jul and G. W. Furnas. Critical zones in desert fog: Aids to multiscale navigation. In *UIST'98: Proc. ACM Symposium on User Interface Software and Technology*, pages 97–106, 1998.
18. D. Nekrasovski, A. Bodnar, J. McGrenere, F. Guimbretière, and T. Munzner. An evaluation of pan & zoom and rubber sheet navigation with and without an overview. In *CHI '06: Proc. Human Factors in Computing Systems*, pages 11–20. ACM Press, 2006.
19. C. North and B. Shneiderman. Snap-together visualization: a user interface for coordinating visualizations via relational schemata. In *AVI '00: Proc. working conference on Advanced Visual Interfaces*, pages 128–135. ACM Press, 2000.
20. K. Perlin and D. Fox. Pad: an alternative approach to the computer interface. In *SIGGRAPH '93: Proc. Computer Graphics and Interactive Techniques*, pages 57–64. ACM Press, 1993.
21. E. Pietriga. A Toolkit for Addressing HCI Issues in Visual Language Environments. In *IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC'05)*, pages 145–152, Dallas, USA, 2005.
22. M. Posner and S. Petersen. The attention system of the human brain. *Annual Review of Neuroscience*, 13:25–42, 1990.
23. G. Ramos and R. Balakrishnan. Zliding: fluid zooming and sliding for high precision parameter manipulation. In *UIST '05: Proc. ACM symposium on User Interface Software and Technology*, pages 143–152. ACM Press, 2005.
24. G. G. Robertson, S. K. Card, and J. D. Mackinlay. Information visualization using 3d interactive animation. *Comm. ACM*, 36(4):56–71, 1993.
25. R. W. Soukoreff and I. S. MacKenzie. Towards a standard for pointing device evaluation, perspectives on 27 years of fitts' law research in hci. *Int. J. Hum.-Comput. Stud.*, 61(6):751–789, 2004.
26. R. Stockli, E. Vermote, N. Saleous, R. Simmon, and D. Herring. The Blue Marble Next Generation - A true color earth dataset including seasonal dynamics from MODIS. Published by the NASA Earth Observatory, 2005.
27. C. Ware and M. Lewis. The DragMag image magnifier. In *CHI '95 conference companion, Human Factors in Computing Systems*, pages 407–408. ACM Press, 1995.