Interaction and Presentation Techniques for Situated Visualization

Sean Michael White

Technical Report CUCS-046-09

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2009

ABSTRACT

Interaction and Presentation Techniques for Situated Visualization

Sean Michael White

Every day we move through a sea of invisible information. As computation, sensing, and display become more mobile and distributed, interaction shifts from our desktop to our environment. This shift changes how we interact with our surroundings, creating the opportunity for *situated visualization*—visual representation of data presented in its spatial and semantic context. This dissertation contributes novel interaction and presentation techniques for situated visualization in three research areas: mobile visualization, objects as context, and scenes as context. In support of this research, we make further contributions by developing algorithms, architectures, and working artifacts.

For mobile visualization, we present results from a field study and task analysis of botanists performing species identification in the field. We develop an iteratively designed, extensible *Electronic Field Guide* (*EFG*) system and architecture, a conceptual data model, and interface (*LeafView*) for mobile visualization. Field experiments show improved identification speed and interaction efficacy.

Next, we explore spatially- and semantically-driven situated visualization using objects as context in head-worn augmented reality (AR). We develop and evaluate tangible AR and head-movement controlled AR interaction techniques. In interviews following lab experiments, participants reported improved speed for inspection and comparison and a preference for tangible AR. Building on this work, we design, develop, and evaluate visualization and activation techniques for discovering and learning gestures for these user interfaces (*Visual Hints*). Lab experiments show preference for visual hints that combine overlaid graphics and animation. In addition, we investigate menu techniques, activated by shaking an object, for interacting with visualizations (*Shake Menus*). We compare display-, object-, and world-referenced coordinate systems for presentation of menu options. Lab experiments show increased speed and accuracy when using the display-referenced coordinate system.

Finally, we present techniques for spatially-driven visualization in the context of physical scenes. Based on our field study of site visits by urban designers, a hand-held AR visualization tool (*SiteLens*) embodying these techniques enables interaction with invisible aspects of urban sites, such as georeferenced sensor data. Field experiments provide evidence for new insights derived from situated visualization, preference for specific representations, and improved interaction with data using a novel stabilization algorithm.

Dissertation Committee

Chair

Prof. John R. Kender, Ph.D.
Columbia University, New York, NY

Advisor

Prof. Steven K. Feiner, Ph.D.
Columbia University, New York, NY

Internal Member

Prof. Julia B. Hirschberg, Ph.D.
Columbia University, New York, NY

External Member

Prof. Terry A. Winograd, Ph.D.
Stanford University, Stanford, CA

External Member

Prof. Barbara G. Tversky, Ph.D.
Teacher's College, New York, NY
Stanford University, Stanford, CA

TABLE OF CONTENTS

LIST OF CHARTS, GRAPHS AND ILLUSTRATIONS

ACKNOWLEDGEMENTS

I owe a debt of gratitude to many people who helped make this dissertation come to life. The completion of this project would not have been possible without them. In my research and writing, I have been fortunate to have the support of an amazing array of individuals.

First, I thank my advisor, Steven Feiner, for supporting, encouraging, and guiding me throughout the arc of my graduate education. I also thank him for his constant support of my family.

I thank the members of the Computer Graphics and User Interfaces Lab that have provided invaluable discussion, support, and friendship throughout this process: Hrvoje Benko, Eddie Ishak, Steve Henderson, Lauren Wilcox, Ohan Oda, Christian Holz, and Marc Eaddy. I also want to thank Paul Blaer, Matei Ciocarlie, and Gabriela Cretu for their friendship and discussions in the hallway.

I thank the project students who studied with me. This work would have taken twice as long without their participation: Jason Koylec, Dominic Marino, Levi Lister, Randall Li, Charles Macanka, and David Feng.

I thank the many people who helped bring the Electronic Field Guide research to life including Peter Belhumeur, David Jacobs, Ravi Ramamoorthi, John Kress, Ida Lopez, Sameer Sheorey, Daozheng Chen, Yuhao Sun, Rusty Russell, Norm Bourg, Matthew Smith, Michael Butz, Guarav Agarwal, and Matt Brown.

I thank my colleagues in the graduate school of architecture, planning and preservation. In particular, I thank Petia Morozov for helping create a bridge between our departments. Sarah Williams for being so thoughtful. Minna Ninova, Rob Viola, Candy Chang for getting it done.

I thank my committee members. John Kender and Julia Hirschberg provided invaluable feedback in my dissertation proposal. Barbara Tversky introduced me to cognitive psychology and broadened my view of visuospatial reasoning. Terry Winograd introduced me to human-computer interaction and provided me with a role model for the ethical aspects of computer science.

I thank my friends, who have supported me throughout this endeavor: Brenda Laurel for her inspiration, Adam Braid for his spirit, Carrie Weber, Michael Burlingame, and Chris Smith for the time and energy they gave, John Cheary and Dave Barreto for being there.

I thank my parents, David and Andrea White, and my sister, Rebecca White, for believing in me and encouraging me to keep at it.

I thank my children, Rose and Wilder, for being bright lights in my life.

Most importantly, I am forever grateful to my wife, Jennifer, for her love, support, and patience throughout this dissertation process. As she proclaimed in her own dissertation, we did it!

This work was supported in part by National Science Foundation grants IIS-03-25867 and NSF EIA-02-02063, Boston Society of Architect’s Research Grants in Architecture, and generous gifts from Microsoft and NVIDIA.

DEDICATION

To Jennifer, ever my inspiration.

To Rose and Wilder, curiouser and curiouser.

1 Introduction

"Alive in the sea of information" [Snyder 1996]. I've been struck by this phrase ever since I first read the poem that contains it, because it resonates with my own sense of something important: the modern experience of knowledge in our environment. A swimming, dynamic sea of knowledge surrounds us, yet it can be difficult to access, perceive, and interact with when and where we need to. *Visualization*, "the use of computer-based, interactive visual representations of data to amplify cognition" [Card 1997], provides one potential solution to this dilemma. The primary contributions of this dissertation are the design, implementation, and evaluation of several novel user interface techniques for presenting, representing, and interacting with visualizations—techniques that move beyond the computer screen and into the physical world.

Visualization is imaging for insight, to paraphrase Hamming [Hamming 1962]. Card et al. [Card 1997] discuss a variety of ways in which visualization can amplify cognition—for instance, by supporting increased mental resources, reduced search, enhanced recognition of patterns, perceptual inference, perceptual monitoring, and a manipulable medium for exploration. Johnson, in a report [Johnson 2006] to the NSF and NIH on the challenges for visualization, observes that "understanding and ultimately, knowledge, cannot be delivered directly from computation. Visualization is the tool through which computation addresses an end user and allows the user to derive knowledge from the data." Examples of visualization are diverse: reducing search by visualizing a ranked ordering of matching results [Ahlberg 1994], expanding working memory through display of a file hierarchy through a hyperbolic tree [Lamping 1996], or enhancing recognition of patterns by representing abstract multivariate data in 3D [Beshers 1990].

Moving through the world in daily work and life, we often look for this kind of insight. Yet visualization most often appears on the office desktop, away from the objects and spaces where it is most relevant. As computation, sensing, and display become more mobile, the locus of interaction shifts to the environment and objects we encounter in it. This trend towards *mobile computing* supports our ability to bring visualization along with us in an Ultra Mobile PC (UMPC) or smart phone. Such devices can act as both opaque surfaces for user interfaces or transparent panes through which we perceive and interact. Similarly, more immersive displays, such as mass-market head-worn displays for video viewing, are appearing on the market. Both types of mobile displays, hand-held and head-worn, can support mobile *augmented reality* (AR).

Figure 1.1 Milgram's mixed reality continuum [Milgram 1994].

AR overlays virtual sensory information on the physical world, generally in real-time and registered in 3D [Azuma 1997, Feiner 1993b]. Milgram defines *mixed reality* along a "virtuality continuum" with real environments on one end of the spectrum and virtual environments or virtual reality (VR) on the other end [Milgram 1994]. Between the two ends of the continuum are augmented virtuality and augmented reality. *Augmented virtuality* merges the physical world into virtual reality: for example, by displaying a live video feed from the physical world in a virtual world. Augmented reality displays virtual information onto the physical world. AR is distinct from virtual reality (VR) in that the physical environment is still perceived. This makes it possible to display visualizations in the physical environment. While mixed reality and augmented reality have been used interchangeably, we use the term "augmented reality" in this dissertation to specifically mean the augmentation of the physical world with virtual information.

Visualization in AR is a current topic of research and provides interesting challenges and opportunities. Of particular interest in this dissertation are visualizations that are relevant to the semantic or physical context in which they are displayed. Typically, visualizations are shown on a stand-alone display, whether desktop [Ahlberg 1994] or head-worn [Fuhrmann 1998a]. In these examples, the physical background has no meaningful relationship to the visualization. In contrast, we use the term *situated visualization* to describe a visualization that is intrinsically related to its environment; for example, visualizing information about a plant species near a physical plant specimen, based on the shape of the leaf, or mapping relevant urban GIS data directly onto the user's view of the city are two scenarios that we explore in this dissertation. Situated visualizations gain meaning through both the visualization and the relationship between the visualization and environment. Although other researchers have developed visualizations that take into account context (e.g., [Gillet 2004, King 2005, Schall 2009]), there has been no general term or consistent framework applied to this approach. Therefore, we propose the term to represent a set of visualizations and related techniques that have interesting and useful commonalities. We make this distinction because visualization in AR is not necessarily situated visualization. For example, visualization in AR of genetic information, displayed in an empty room, is not related to the physical environment and is not considered situated visualization. We discuss these issues in a conceptual framework for situated visualization in Chapter 2.

To make situated visualization useful and effective, we pursue a course of experimentation, exploring a variety of presentation, representation, and interaction techniques that support common visualization tasks and acknowledge the unique aspects of situated visualization: information overlaid on the physical world combined with meaningful relationships to the physical world.

1.1 Research Questions and Dissertation Goals

This dissertation explores presentation and interaction techniques in situated visualization. Of particular interest are techniques that address real-world visualization tasks. In this investigation, we pose the following questions:

What theoretical framework should we use to classify and characterize situated visualizations? In the human-computer interaction (HCI) research community, it is useful to create a design space and taxonomy for new and existing user interface techniques (e.g., [Bier 1994, Card 1991, Ishii 1997]). This provides a common vocabulary for discussing the research and helps define areas that need to be explored.

What are the best ways to present and display situated visualizations? We hypothesize that gestalt rules [Mullet 1995] apply in terms of learning tasks and that spatial proximity is an important aspect of tasks such as comparison and inspection.

What user interface techniques can we use to best interact with situated visualizations and visualization elements? A variety of techniques have been developed for interacting with visualizations. Here we focus on Shneiderman's visual information-seeking mantra [Shneiderman 1996] and investigate related paradigms for pattern seeking and image identification/comparison in situated visualization. We hypothesize that directly touching and manipulating data, paired with direct presentation of virtual elements in proximity to relevant physical elements, will maximize speed, accuracy, and comprehension of visualization tasks and increase insight gained from visualizations.

What are the benefits of situated visualization and in what tasks and contexts are they most appropriate? We hypothesize that certain tasks—such as inspection/comparison, spatial learning, and in-situ pattern seeking and discovery—benefit from enhanced cognition through situated visualization, compared to the alternatives.

What design principles apply for creating situated visualizations? An important goal of this research is to take results from evaluations and codify them into a set of design principles that can be used when developing situated visualizations for future applications.

1.2 Approach and Process

This dissertation describes an iterative process of ethnographic design research [Laurel 2003], novel user interface invention, system development, evaluation [McGrath 1995], and theoretical construction. To ground our research, we have developed specific prototype applications as examples and test beds for investigating interaction techniques. Evaluation includes objective comparisons of display methods and systems, measuring task performance, and ascertaining user preferences [Ware 2004] through a combination of user studies, structured interviews, and informal feedback. Our approach incorporates collaboration with experts in other fields (botany and urban design/planning), as a way to ground, guide, and validate our work on the one hand, and on the other, to contribute to the computer science community and society at large.

1.3 Contributions

In this research, the chief contributions to human-computer interaction, and computer science more generally, include the development of novel interaction techniques and their evaluation to understand situated visualization and the phenomena surrounding them.

These techniques are based on different types of semantic and spatial context, relating data to physical objects and scenes. In addition, in support of this research, we have made further contributions through the creation of novel algorithms, architectures, and working artifacts. In doing so, we seek to advance the state of the art in computer science, while directly engaging and enabling other disciplines.

In this first chapter, I provide motivation, an overview of the research area and questions addressed by this dissertation, a description of my approach, and an enumeration of the contributions that I make. In Chapter 2, I define situated visualizations and present a framework for characterizing them. Chapter 3 examines and discusses related work and places my research in the context of the broader field of research. The remaining chapters are broken into three related topic areas. In the first topic area (Chapter 4), we address the challenges of mobile visualization, bringing visualization away from the desktop and into the field. In the second topic area (Chapters 5–7), we discuss situated visualization techniques where the context is an object. In the third topic area (Chapter 8), we discuss situated visualization techniques where the context is the scene. These topic areas and the chapters that present them are followed by Chapter 9, which summarizes the results and conclusions of this dissertation.

Beyond looking at interaction techniques across different contexts, there are two additional themes in this research. The first theme involves display systems. The form of display used in a given system strongly affects the interaction techniques and visualizations used with it. To that end, this dissertation investigates several types of hand-held and head-worn displays. Hand-held displays can be opaque surfaces that provide a 2D interface to information. Alternatively, they can be windows through which the world can be viewed, with 3D data mingled with the physical world using AR. Head-worn displays can be *optical see-through*, where the user directly observes the physical world and graphics are overlaid on top of the view. Alternatively, they can be *video see-through*, where one or more cameras acquire imagery of the physical world and the camera imagery and graphics are mixed and displayed to the user through the display. In this research, we investigate techniques using all four types of display.

Figure 1.2 Application domains: Botanists collecting plants on Plummers Island (left) and urban planners evaluating photographs and maps from a site visit (right).

The second theme involves the application domain. User interface techniques do not exist in a vacuum and must apply to specific actions and tasks. In this research, we develop techniques in two distinct application domains, shown in Figure 1.2. The first domain is identification of botanical species in the field, based on collaboration with botanists at the Smithsonian Institution. The second domain is the urban site visit, based on collaboration with urban designers and urban planners in the Graduate School of Architecture, Planning and Preservation at Columbia University.

This dissertation makes the following contributions:

- *Development of a descriptive characterization of situated visualizations*. This classification, which considers characteristics unique to situated visualization, provides a means of organizing current and future research for comparison and discovery of opportunities in the design space.
- *Design, implementation, and evaluation of a robust multi-platform system and user interface techniques for mobile identification and visualization, called* ***LeafView*** [White 2006c, White 2007b]. This combination of an underlying infrastructure and prototype hand-held user interfaces for an Electronic Field Guide (EFG) [Agarwal 2006, Belhumeur 2008] supports image capture, identification, results visualization, and data collection for use in botanical species identification in the field. LeafView is extensible to multiple datasets, visualization techniques, and search algorithms in support of future research by vision and HCI researchers. Prototypes have been evaluated in field experiments at Plummers Island, MD and Wind Forest, MD, informally evaluated by our botanist colleagues, and tried by hundreds of people at several public events. LeafView serves as a platform for learning about use in the field, as an exploration of visualization in context, and as a baseline for comparison with other novel techniques. In particular, we have developed a conceptual model called *Virtual Vouchers*, visual feedback techniques for interacting with vision-based semantic-driven visualizations, incorporated proximal displays for comparison, and explored alternative visualization layouts such as quantum tree maps based on hierarchies created from k-means clustering.
- *Design, implementation, and evaluation of techniques for visualization and comparison and inspection of image matching results presented in AR* [White 2006a, White 2006b]. Visualization that is far from the object of interest still requires constant change in focus and indirect interaction with data. To address this, we investigate augmented reality interfaces for comparison and inspection that bring the visualization and relevant object in closer proximity. Focused on objects as context for situated visualization, these two AR prototypes use the common conceptual abstraction and visual representation of Virtual Vouchers**,** and support visualization, comparison, and inspection of matching and identification results. The first prototype, *Head Movement Controlled AR EFG* (HMCAR-EFG), uses rotation of the head relative to the body to control virtual vouchers. The second prototype, *Tangible Augmented Reality EFG* (TAR-EFG), uses novel gestures made with physical objects and spatial morphing techniques to interact with similar situated visualizations of data. Both systems were evaluated in a qualitative laboratory experiment using structured interviews with botanists at the Smithsonian Institution, and the

TAR-EFG system has been demonstrated at several public events. Among our results, we found that direct manipulation of physical objects associated with virtual objects addresses existing spatial intelligence and provide anchors for direct manipulation of situated visualization data.

- *Design, implementation and evaluation of techniques for dynamic explanations, called **Visual Hints*** [White 2007a]. Users of a gestural user interface may not know what gestures are possible. To address this, these techniques support activating and presenting graphical representations in AR of potential actions and their consequences in the augmented physical world, showing how to make gestures in the user interface through visual representations. In a laboratory experiment, we compare multiple presentation forms and two techniques to activate visual hints. Among our results, we found that hybrid hints such as animations combined with ghosting techniques were the most preferred.
- *Design, implementation, and evaluation of techniques for prop-based menu display and selection coupled with object positioning, called **Shake Menus*** [White 2009b]. This technique directly associates visualization with hand-held objects (physical props) and uses these as tools for interaction. In a laboratory experiment, we compare time to completion and accuracy using display-, object-, and world-referenced coordinate systems for presentation of menu options and investigate additional potential applications. Among our results, we found that display-referenced and object-referenced presentation support fastest completion time, while display-referenced incurred the fewest errors.
- *Design, implementation, and evaluation of a prototype visualization system and techniques using hand-held AR in urban environments, called* **SiteLens** [White 2009a]. The prototype explores interaction and presentation of geo-referenced situated visualization in urban and natural environments for information seeking, pattern seeking, and comparison/inspection tasks. Our prototype system was developed in the context of **Site Visit by Situated Visualization** [White 2007c, White 2007d, White 2008], a joint project with colleagues in the Graduate School of Architecture, Planning and Preservation at Columbia. SiteLens explores hand-held, video see-through AR, and provides a test bed for investigating data curation, data representation, and touch–based interfaces for overview, comparison, and detail-on-demand visualization tasks. We first conducted a field study to understand the tools and process of a site visit. After developing SiteLens, we conducted a field experiment with participants from the Graduate School of Architecture, Planning and Preservation to understand whether the system could be used to gain new insight into the domain and to compare representation techniques. Among our results, we found that concrete representations such as smoke were preferred to abstract representations such as spheres. We also found anecdotal evidence of new forms of insight based on use of the system.

The original motivation for this dissertation arose from our early observations that activities in the field benefit from visualization, yet little attention has been paid to how these visualizations should be provided or how they differ from more typical visualizations found on desktop systems. The ethnographic study described in Chapter 4 provided this kind of insight and guided our early work. This work was followed, in concurrent efforts, by first iterations of the AR techniques discussed in Chapter 5 and the Tablet PC version of the LeafView prototype and associated techniques described in Chapter 4. The LeafView system is discussed first in this dissertation to introduce issues in mobile visualization combined with sensing the physical world, and also to provide a baseline for comparing other situated visualizations. In addition, the form factor of the LeafView prototypes provides a system that could be used in the field without aid; in contrast, the AR interfaces provide more interesting but experimental hardware configurations with ergonomic challenges, such as narrowed field of view and occluded vision. These problems make current AR implementations difficult to use in potentially treacherous physical terrain.

Our Visual Hints techniques, described in Chapter 6, result from addressing visualization issues that arose from use of the TAR-EFG in evaluations. Similarly, the Shake Menus techniques, presented in Chapter 7, followed from continued use of the TAR-EFG in demonstrations and the need to directly interact with data. Finally, SiteLens and associated techniques, discussed in Chapter 8, address the need to support situated visualizations that combine the ergonomics of hand-held systems with the benefits of AR. In the following sections, we describe each contribution in more detail.

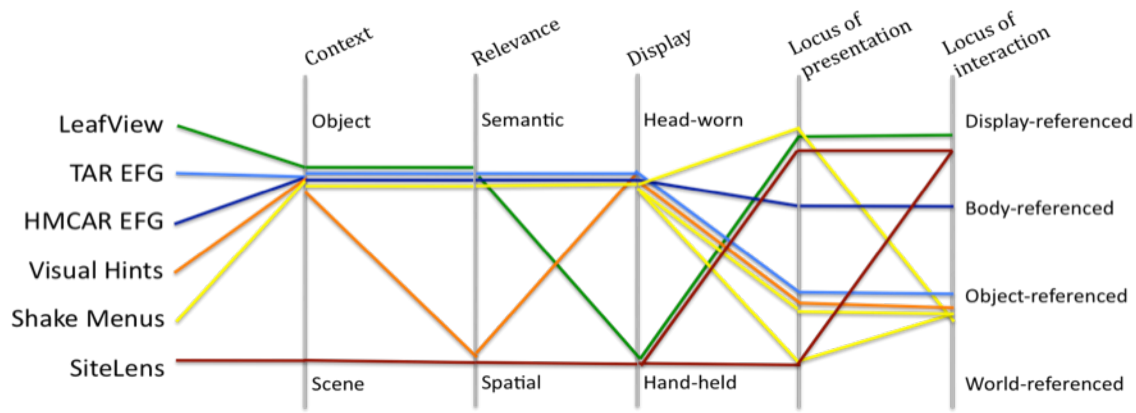


Figure 1.3: Parallel coordinates visualization depicting selected design dimensions of situated visualization.

1.3.1 Characterizing Situated Visualizations

The vocabulary used to describe and characterize computer visualizations does not currently reflect the possibility that the visualization and associated interaction techniques can be related to the underlying physical scene where the visualization is displayed. In Chapter 2, we address this issue by presenting a set of characteristics that can be used to categorize and compare existing and novel situated visualizations. We discuss terminology and review example situated visualizations using the described characterizations. Figure 1.3 depicts situated visualizations systems and techniques we have developed, displayed in parallel coordinates representing this set of categories. Chapter 3 follows with a description and analysis of research and work related to the general concepts of this dissertation. We discuss situated learning, knowledge, and interaction; context-aware computing; AR; visualization; and visualization in AR. Related work more specifically associated with individual techniques is covered in the later chapters that address those techniques.

Figure 1.4: Botanist from the Smithsonian Institution using a second iteration version of the Tablet PC-based LeafView Prototype.

1.3.2 Mobile Identification and Visualization with LeafView

Chapter 4 introduces the first application domain and presents the design, implementation, and evaluation of LeafView, a system for mobile identification, visualization, and collection of botanical species. Here, identification of botanical species in the field represents a general mobile task of object identification, while providing the opportunity to study a concrete application with specific users.

In Section 4.1, we present results from an ethnographic study of botanical identification and collection in the field, including a task analysis of the process. This is followed, in Sections 4.3, 4.4, and 4.8 by the design and implementation of multiple iterations of LeafView, an extensible prototype system that speeds identification and collection using computer vision algorithms (developed by our colleagues) and visualization techniques specific to the mobile context. Our first prototype is based on a Tablet PC with a wireless camera (Figure 1.4), while the second prototype, redesigned in response to feedback from our first field experiment, uses an Ultra Mobile PC with integrated camera and Bluetooth GPS (Figure 1.5). We also discuss subsequent prototypes developed with web-based and mobile phone interfaces.

Figure 1.5 LeafView UMPC displaying matching results.

In Sections 4.5, 4.6, 4.7, 4.9, and 4.10, we present observations and results from a series of formal and informal evaluations, including comments from users about the different prototype systems and associated techniques. We found the following: that system feedback while the user is interacting with the vision algorithm improves the experience and quality of acquired images, that visualization of results in unnumbered groups was preferred, and that proximity of the leaf specimen eases the comparison task.

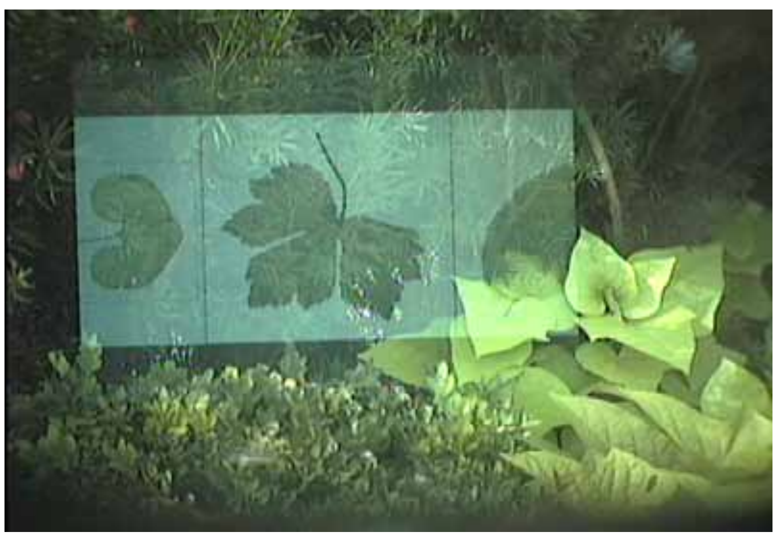


Figure 1.6: View through optical see-through display of situated visualization of matching leaf results in head-movement controlled augmented reality. Visualization is displayed, fixed in space, in reference to the head of the user.

1.3.3 Inspection and Comparison using Head-movement–Controlled and Tangible Augmented Reality

In our ethnographic study, we found that subjects wanted to more closely relate matching results to the physical world and view visualizations within it. This desire motivated us to develop AR user interfaces that support different interactions than those in the LeafView prototypes, which display visualization apart from the physical leaf specimen. We developed two new AR user interfaces, discussed in Chapter 5, to explore the significance of situated visualizations that are both semantically relevant (based on the leaf image) and displayed in physical proximity to the leaf, embedded in the physical world.

The *Head-Movement–Controlled AR EFG* (HMCAR-EFG), shown in Figure 1.6 and described in Section 5.2, provides a hands-free interface and displays the visualization of species fixed in space to the body of the botanist. Visualization tasks such as semantic zooming are accomplished by rotating the head.

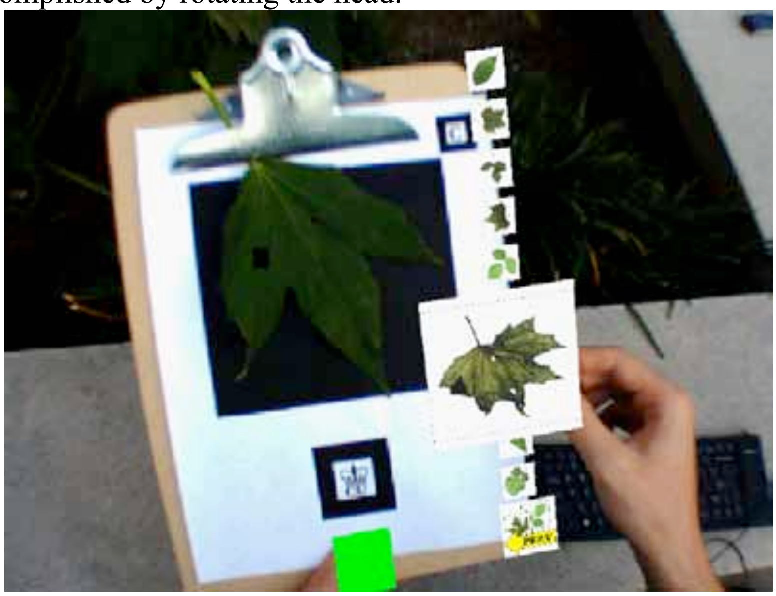


Figure 1.7 Using the Tangible Augmented Reality EFG to compare a leaf with potential matches.

The Tangible AR EFG (TAR-EFG), discussed in Section 5.1, moves the visualization of results from the device to the leaf (Figure 1.7). It therefore allows us to explore manipulation of the data with tangible AR, using visualizations that are displayed as if they are attached to physical objects. We explore techniques such as spatial morphing and tangible gestures for semantic zooming tasks. The implementation for this system is discussed in Section 5.3.

The TAR-EFG and HMCAR-EFG user interface techniques have been implemented and evaluated in a comparative, structured interview study [White 2006b], discussed in Section 5.4, involving four of our botanist colleagues under IRB-AAAB6501. We found that TAR-EFG provides a faster way to compare and inspect virtual data in the visualization and the tangible manipulation is more intuitive for our subjects. However, subjects appreciate having their hands free with the HMCAR-EFG user interface.

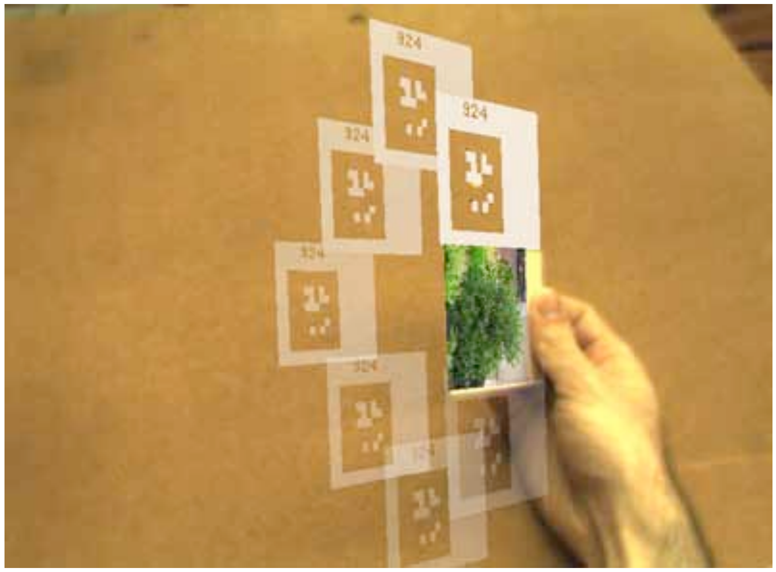


Figure 1.8: View through video see-through display of ghosted visual hint explaining the path and direction of a "reeling" gesture.

1.3.4 Dynamic explanations Using Visual Hints

In Chapter 6, we extend our TAR-EFG system to investigate alternative representation techniques for situated visualizations that directly relate to movement of tangible user interfaces. Tangible AR systems imbue physical objects with the ability to act and respond in new ways. In particular, physical objects and gestures made with them gain meaning that does not exist outside the tangible AR environment. The existence of this new set of possible actions and outcomes is not always apparent, making it necessary to learn new movements or gestures. Addressing this opportunity, we developed visual hints, which are graphical representations in AR of potential actions and their consequences in the augmented physical world. Visual hints enable discovery, learning, and completion of gestures and manipulation in tangible AR, discussed in Section 6.1. We present a variety of representations of visual hints in Section 6.2, and methods for activating them in Section 6.3. We then describe a specific implementation in Section 6.4 that supports gestures developed for a tangible AR user interface to an electronic field guide for botanists, and present results from a pilot study, in Section 6.5, comparing representation techniques. We found that subjects preferred a combination of ghosting and animation for representing visual hints (IRB-AAAC5545).

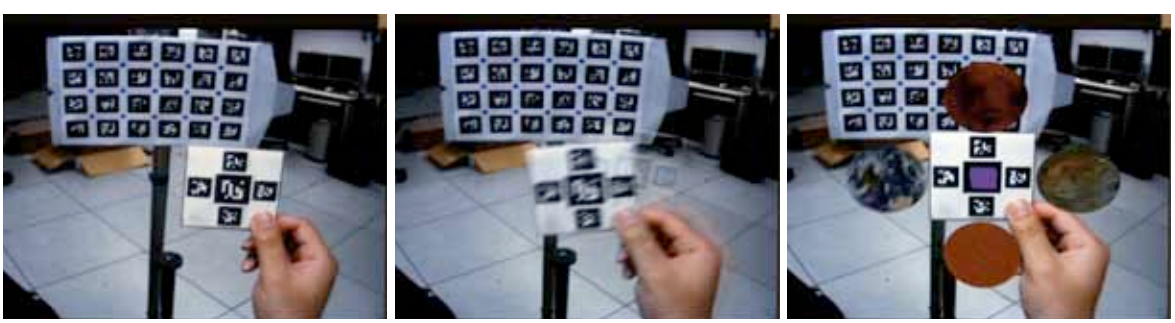


Figure 1.9: A shake menu being used to select and place planets in a 3D environment. (a) User holds an object (in this case, an optically tracked fiducial marker) and (b) shakes the object to (c) display a radial menu of options around the object.

1.3.5 Manipulating Visualizations with Shake Menus

Shake menus, which we discuss in Chapter 7, are a novel method for activating, displaying, and selecting menus or radial information displays presented relative to a tangible object or manipulator in a 3D user interface. They provide ready-to-hand interaction, including facile selection and placement of objects. We present the technique in Section 7.2, and describe a study (IRB-AAAD6617) in Sections 7.3–5 that compares the speed and accuracy of display-, object-, and world-referenced coordinate systems for presentation of menu options. We also present qualitative feedback from use and several illustrative applications, in Section 7.6, of the technique for interacting with visualizations and authoring. Lab experiments show increased speed when using the display- and object-referenced coordinate system and increased accuracy when using the display-referenced coordinate system.

Figure 1.10: Situated visualization using SiteLens (inset), comparing locally-sensed carbon monoxide data (red) and remote EPA sensor reading associated with the site (green).

1.3.6 Georeferenced Data Visualization using SiteLens

Our early techniques of Chapters 4–7 focus on situated visualization where the context is a single object and the relationship between the data and context has primarily been semantic. In Chapter 8, we further explore the design space and consider interface techniques where a physical scene is the context. We begin by introducing a new domain and task focused on urban site visits, in Section 8.2. Urban designers and urban planners often conduct site visits prior to a design activity to search for patterns or better understand existing conditions. In Section 8.4, we present SiteLens, an experimental system and set of techniques for supporting site visits by visualizing relevant virtual data directly in the context of the physical site. We address alternative visualization representations and techniques for data collection, curation, discovery, comparison, manipulation, and provenance. Our implementation is discussed in Section 8.5. A real-use scenario is presented in Section 8.6, and two iterations of evaluation (IRB-AAAD3016) with faculty and students from the Columbia University Graduate School of Architecture, Planning and Preservation are discussed in Section 8.7. The scenario and evaluation together provide directions and insight for further investigation. In particular, we found that the subjects wanted to sense data and visualize at the same time and that more abstract representations such as smoke were preferred over more concrete representations like spheres. We also found that our sensor fusion algorithm and freezing techniques improve stability of interaction and that the system could indeed be used to gain insights not necessarily apparent from existing techniques.

1.4 Reading this Dissertation

The first three chapters of this dissertation provide an overview of the research area (Chapter 1), a conceptual framework and vocabulary for investigating situated visualization (Chapter 2), and discussion of research and work related to the topic (Chapter 3). Chapters 4-8 provide details on specific research and experiments conducted to investigate situated visualization and the phenomena surrounding situated visualization. Chapter 4 focuses on mobile visualization. Chapters 5–7 focus on objects as context. Chapter 8 focuses on scenes as context. The final chapter (Chapter 9) summarizes the dissertation, discusses future work, and ends with a few final words about this research.

For those who wish to quickly skim this dissertation, each chapter begins with an introduction, which outlines the chapter and the contributions in the chapter. Each chapter also ends with a summary, which reviews the main points of the chapter, describes the specific contributions in each chapter, and relates the work to the larger project of situated visualization.

2 Situated Visualization Defined

2.1 Motivation

Visualizations have traditionally been displayed and presented in environments such as desktops or virtual reality systems, where the physical location has no meaningful relationship to the data. While these visualizations have proven useful for enhancing our ability to make sense from information, they are set apart from the physical world. At the same time, a confluence of new technologies, such as mobile computing, displays, and sensing, as well as new techniques, such as augmented reality, tangible user interfaces, and context-aware computing, make it possible to present and interact with virtual information in the physical context in which it is relevant.

To address the opportunity presented by visualization without physical context and this diversity of new technologies and techniques, we have conducted research investigating presentation and interaction with visualization displayed in the relevant spatial and semantic context. We propose the term *situated visualization* to describe this class of visualizations. In our research, we are motivated by our findings that situated visualization provides instrumental benefits for scientists and designers focused on specific tasks. Such visualization also holds the potential to enhance everyday experiences through user interfaces that directly display information in the environment.

As we look to the future of interaction with information in our environment, we see situated visualization playing an important role. While we are not the first to develop visualizations that are situated in their contexts, we see a need to develop a framework and design space to enable discussion within the research community that uses a similar vocabulary and set of characteristics for comparison and analysis. Frameworks for information visualization [Card 1997], context-aware computing [Schilit 1994] and tangible user interfaces [Ullmer 2000] serve as inspiration for our situated visualization framework. We propose this framework as a new lens for viewing and unifying this area of research and as a means of conceptualizing the design space for this dissertation. We consider this a first step in constructing a theoretical understanding of the phenomena surrounding situated visualization.

In this chapter, we discuss the emerging space of interfaces that display and interact with visualization presented in context. We describe important classifying characteristics of these situated visualizations and illustrate the framework using examples from our research and associated literature. Finally, we discuss design implications and challenges for situated visualization.

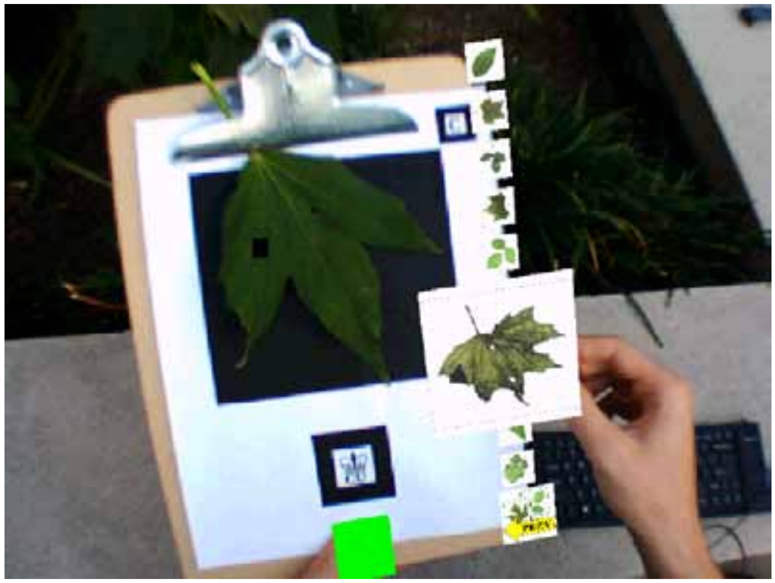


Figure 2.1 Tangible AR EFG interface displaying leaf matches near the physical leaf.

2.1.1 An Example of Situated Visualization

Consider the Tangible Augmented Reality Electronic Field Guide (Figure 2.1), introduced in Chapter 1. The system is used outdoors in a mobile context with an immersive, head-worn display that overlays graphics on the physical scene. A leaf is placed on a clipboard and identified using computer vision. Virtual visual representations of potentially related species and related imagery, called virtual vouchers, are displayed around the physical leaf. A hand-held fiducial marker, used for visually tracking objects, is held in the dominant hand. When the physical marker touches a virtual voucher, the marker in hand visually becomes that virtual voucher. Tangible manipulation of the physical marker, including tangible gestures, supports direct manipulation of the data for visual search, inspection, and comparison.

Here, the physical context is an object: the leaf. The visualization is the proximal display of related species images and different aspects of those species. The relationship between the leaf and data is semantic, based on knowledge about the identity of the leaf. The spatial arrangement of the physical leaf in close proximity to the virtual species images creates the appearance of a relationship between the physical object of interest and the virtual data. Focus shifts between virtual leaves and the physical leaf as the user inspects and compares individual features. By providing situated visualization, focus remains on the task at hand without shifting attention to devices or displays that are not in proximity to the relevant context.

Figure 2.2 Visualization of arterial flow and wall shear stress [Forsberg 2000].

2.1.2 An Example of Non-Situated Visualization

While it is true that all visualizations are, in some form, situated, we are concerned with those that are both connected to the context and content of the scene in which the visualization is presented and displayed in the physical scene. For instance, we do not consider a visualization of arterial flow [Forsberg 2000] (Figure 2.2) displayed in a CAVE, a situated visualization, because the data is not related to the current context. Although one could claim that the CAVE and office provide context and situate the visualization in the workplace, we argue that experiencing that visualization in one office or another would provide little difference to the experience. The visualization is not related to the context in an interesting way.

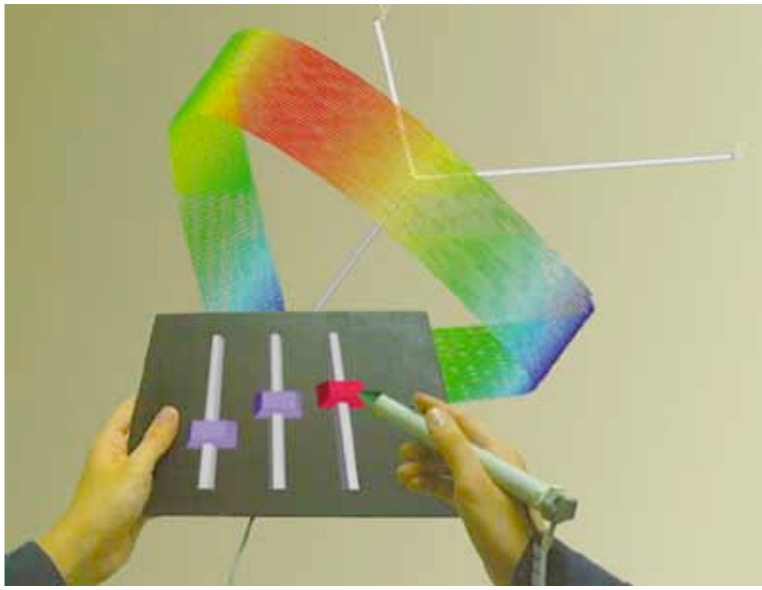


Figure 2.3 Visualization of dynastic cycles in ancient China [Fuhrmann 1998a].

Another example for consideration is an augmented reality visualization of Chinese dynasties floating in space [Fuhrmann 1998a] (Figure 2.3), just as it would be seen through a head-worn display. Here, multiple users can share the visualization and the visualization is displayed directly in the environment, but the data and visualization have no relationship to the environment or any object in it. The virtual visualization is displayed in the physical scene, but there is no inherent semantic or spatial relationship between the two. Therefore, we do not consider this a situated visualization.

2.2 Key Characteristics

Taxonomies, design spaces, and conceptual frameworks provide a common model and vocabulary for research and design. These focus attention on important aspects of a class of user interface and serve as tools for comparing individual instantiations. They also define a design space that can be used for identifying empty or sparsely populated areas requiring additional investigation. Here, we enumerate key characteristics of situated visualizations that are common across all situated visualizations and specify details of those characteristics for comparison. These key characteristics are:

1. Data in the visualization is related to the physical context.
2. Visualization is based on the relevance of the data to the physical context.
3. Display and presentation of the visualization is in the physical context.

2.2.1 Data

Card and MacKinlay [Card 1997] provide an excellent formal treatment of data in visualization. We are particularly interested in whether the data already has spatial or temporal characteristics, so that we take these relationships into account when we visualize the data. For example, the fact that sensor data is already georeferenced implies an existing set of spatial relationships that should be taken into account when displaying such data. In addition, we are interested in values associated with the data that may constrain their layout or relationship from one data point to another. The data may be *nominal*, where instances are either equal or not equal to each other (e.g. categorical information such as married or not married), *ordinal*, where instances are ranked or ordered (e.g. ranked matching results form a text search), or *quantitative*, where instances can be manipulated by arithmetic (e.g. test scores or sensor reading).

2.2.2 Context

One of the most important distinctions in situated visualization is the identification of the context that drives the visualization and the relationship between the data represented by the convergence of visualization and context. We first want to identify whether the context is one or more individual objects or an entire scene. We distinguish these different contexts in several ways. For object as context, the visualization is associated with a single object in the view of the user (e.g., a visualization surrounding a single leaf showing species related to that leaf or a physical model of a molecule with magnetic fields visualized on the surface of the model). This can be extended to multiple objects where the visualization connects the physical objects in some way (e.g., a visualization showing the relationship between two leaves). If the scene is used as context, the entire space of the scene is related to the data and will affect the visualization (e.g., a visualization of carbon monoxide data overlaid on an urban environment, such as Manhattanville in New York City).

2.2.3 Relevance

Once we have identified the context, we can describe the relationship between the physical context and the data represented by the visualization. We use the term *relevance* to refer to this particular relationship between context and data. We focus on relationships that are *semantic* or *spatial*. In a semantic relationship, the relevance between the data and the context comes from knowledge about the context, such as the identity or classification of the object or aspects of the physical scene (e.g., identifying a person's face using computer vision and visualizing a social network around the person based on recognizing the face). We contrast this semantic relationship with a spatial relationship, where the data has a particular location or orientation relative to the context (e.g., visualizing underground infrastructure in a specific location). A temporal relationship could also exist, but we assume that this is in concert with some semantic or spatial relationship.

The scale of the data may affect the relationship to the context. The term scale has different meanings across the diversity of disciplines involved in the use of data. For example, GIS professionals think of data scale as the rough area covered by a piece of data. Computer science professionals often think of data scale as a way of describing the size and quantity of an entire data set (e.g., large-scale data analysis). Because we are interested in the visual display of data, we use scale to signify the semantic or spatial grouping of relevance for a given instance of data. For example, the scale of spatial relevance for a given carbon monoxide sensor reading may be a one millimeter point or an entire city block. The scale of semantic relevance for a specific object such as a leaf may be a specific instance of that leaf (identification) or a genus or species of leaves (classification).

2.2.4 Display

There are many ways to display visualizations relative to the user and the context. In this dissertation, we focus on hand-held and head-worn displays. In particular, we address opaque display surfaces, transparent hand-held displays, and immersive head-worn displays. However, we do not limit our framework to these modes of display and incorporate other alternatives such as projective displays [Hua 2000], where a video image is projected onto the surrounding environment, or non-see-through AR displays such as laptops with USB cameras where the video source is separate from and not fixed to the display.

We are interested in displays for several reasons. First, the display provides the means by which the information is related to the users. Second, the display can affect the immersiveness of the experience. Finally, the display provides a potential locus of presentation and interaction, as discussed in the next few paragraphs.

2.2.5 Presentation

We use the phrase *locus of presentation* to describe the coordinate system used for the spatial presentation of information. We build off the work of Feiner et al. [Feiner 1993a] and use display-referenced, body-referenced, object-referenced, and world-referenced coordinate systems. When we say that visualization is presented *display-referenced*, we mean that the position and orientation of the visualized information are displayed in a reference frame fixed to the display coordinate system. For example, a frames-per-second counter is often shown in the upper right- or left-hand corner of the display. Moving the head or body does not change the location of the counter on the display screen. A *body-referenced* presentation is fixed to some part of the user's body. For example, by wearing a tracker on the waist, a visualization of plant matches can be placed such that it always stays in the same location in front of the body, regardless of how the head or display are moved. Similarly, a tracker attached to the head can be used to place the same plant matches in 3D locations relative to head position such that they stay in the same position relative to the head or are controlled by head movement, regardless of the position and rotation of the torso. An *object-referenced* presentation is fixed to an object in the world. An example would be a visualization of plant species matches that moves with the position of a leaf or clipboard. *World-referenced* means fixed to the reference frame of the earth. For example, sensor data that is placed in a specific longitude and latitude will stay in that physical location no matter how the user moves their body, head, or the display.

These different coordinate references can be combined to use the orientation from one reference frame and the position from another. For example, our Shake Menus technique explores display-referenced position and object-referenced orientation. In this way, the radial menus always stay in the same position relative to the display screen of a head-worn display while orientation changes with changes to the orientation of a hand-held object. Rotating the object forward rotates the visualization forward yet does not change the position of the visualization on the screen.

While not specific to situated visualization, two additional aspects of the visualization are important to consider: representation and interaction.

2.2.6 Representation

Representation has been discussed since the beginnings of information graphics [Bertin 1981, Bertin 1983, Card 1997, Card 1999, Tukey 1977]. Here, we are concerned with several ways in which the visual representation interacts with the physical context. First, the representation maps data onto the visual field. Because we present information in the physical world, cues that might be mapped to values must be carefully considered. For instance, size is often used to represent value, but size can be confused with distance cues in a situated visualization. In addition, we are also concerned with the visual interaction of the virtual and physical, as they affect perception.

In a typical visualization, the author of the visualization chooses the background so that it does not compete with the data representation. In the case of AR-based situated visualization, the representation may need to be aware of the background scene to enforce figure-ground relationships, to closely associate virtual data with the physical surroundings, or to change the representation based on background. Such characteristics may guide or constrain both the visual representation and the spatial arrangement of the visualization. For example, a scene-aware visualization layout might avoid placing data in locations of areas of high frequency texture, because they could have useful information about the context. As another example, a color-aware visualization might avoid using green representations of data when the background is a verdant forest in order to avoid losing the data in the scene.

Also of interest is the mix between the physical and virtual. Rather than considering one or the other exclusively as the figure or the ground, we can consider the spectrum of combinations for mixing the physical world with virtual representations to present a single perceived visualization. In this case, the figure may be all physical, all virtual, or some combination. Similarly, the ground may be all physical, all virtual, or some combination. An example would be a visualization showing new locations of virtual trees to be planted and mixed with physical trees that are already present in a park. The viewer may want to adjust the visualization such that all the trees look the same, so that the mix of virtual and physical is the figure and the background scene is the ground. Alternatively the viewer might want to adjust the visualization so that the new trees visually pop out, making the new trees the figure and both the old trees and background scene the ground. Finally, the viewer might want the old trees visually emphasized, so that the old trees are the figure and both the virtual trees and background scene are the ground.

2.2.7 Interaction

We use the phrase *locus of interaction* to describe the coordinate system in which the user interacts with the visualization. Like loci of presentation, these can be any combination of display-referenced, body-referenced, object-referenced, or world-referenced coordinate systems. While interaction and presentation are often coupled, they can occur in different reference frames. For example, visualization may be world-referenced, but interaction may be display-referenced, as is the case in our SiteLens prototype. Carbon monoxide data is presented in specific locations based on the longitude and latitude of the sample data. However, data is queried by touching the data point on the display screen of the hand-held device. Interaction with a visualization or set of visualizations may also be through multiple loci. We consider this important because it addresses whether data in a visualization may be directly manipulated or whether a tool is used for manipulation.

2.3 Terminology

In addition to the key characteristics described in the previous section, we use certain terms throughout this dissertation, which we define here.

2.3.1 Virtual, Physical, and Real

In discussing the many different combinations of the physical world and that which is computer-generated, the word "real" is often used to refer to the physical world, implying that the virtual world is somehow fake. Here, we prefer a view similar to Ishii and Ullmer [Ishii 1997] and distinguish between the physical—those aspects of the world made of atoms—and the virtual: those aspects made of information (or perhaps simply not made of atoms). Both are perceived by the user and are considered "real" for our purposes. In particular, they may be combined to provide a reality distinct from their individual parts.

2.3.2 Figure-Ground and Focus+Context

Our model of situated visualization borrows the concept of a figure-ground relationship from applications of gestalt theory to visual design [Mullet 1995]. The figure is the primary formal element and the ground is the visual context within which the figure appears. In our model, the figure and the ground can be any mix of physical and virtual; they need not be separate. For instance, a physical object, such as a leaf, may be the figure and the situated visualization, representing ambient information about the plant, may be the ground. Note that this is distinct from focus+context user interfaces [Lamping 1996] in that the figure is not always the focus. Typically, the focus is the subject of the user's attention and requires additional visual detail, while the context provides the global view of the information. The distinction here is that figure-ground tends to be the result of perceptual organization, while focus+context is the result of the user's attention and interaction with the visualization.

2.3.3 Presence, Proximity, and Contiguity

In our research, there is often a need to characterize the type of spatial relationship amongst objects. Here, we suggest a variation on gestalt principles to differentiate spatial relationships. *Presence* implies that one object is simply visible when another is visible. No additional spatial relationship exists. *Proximity* implies nearness, with the adjective "close" understood. One object is proximal to another; it is near another spatially and is often perceived as associated or grouped. *Contiguity* implies a pattern of connectedness or adjacency. The important distinction here is that proximity simply implies a short distance or closeness, while contiguity implies a specific location in proximity--one that often enforces some visual pattern.

Mayer [Mayer 2001] describes several studies suggesting that proximity increases learning in simple 2D learning experiments. Typically, these are considered in a static 2D or 3D scene, projected onto a 2D surface. It is unclear how these generalize across loci of presentation. For instance, if object A is display-referenced and object B is world-referenced, they could be proximal on the 2D projected plane at some point and yet may not be related.

If our goal is to represent associations of virtual data with physical objects and scenes, we may be able to apply gestalt principles of organization to emphasize that relationship. Several attempts have been made to enumerate these principles in the context of information visualization. For instance, Card et al. [Card 1999] provide a table of principles based on work by Tovee [Tovee 1996] and Mullet and Sano [Mullet 1995] discuss a similar set of principles. We describe a combination of them in the following table (Figure 2.4).

Rule	Boundaries
Pragnanz	Every stimulus pattern is seen in such a way that the resulting structure is as simple as possible
Proximity	The tendency of individual elements to more strongly associated with nearby elements than with those farther away.
Similarity	If several stimuli such as shape, size, color, texture, value and orientation are presented together, there is a tendency to see the form in such a way that the similar items are grouped together.
Closure	The tendency to unite contours that are very close to each other and interpret visual stimuli as complete.
Continuity	Neighboring elements are grouped together when they are potentially connected by straight or smoothly curving lines, using the simplest possible physical explanation.
Common fate	Elements that are moving in the same direction seem to be grouped together.
Familiarity	Elements are more likely to form groups if the groups appear familiar or meaningful.
Area	The smaller of two overlapping figures will be interpreted as the figure while the large is interpreted as the ground.
Symmetry	The greater the symmetry of a possible figure, the more likely we are to use it as our interpretation of the whole.

Figure 2.4 Table based on Card et al., based on Tovee and incorporating Mullet

2.4 Examples in the Framework

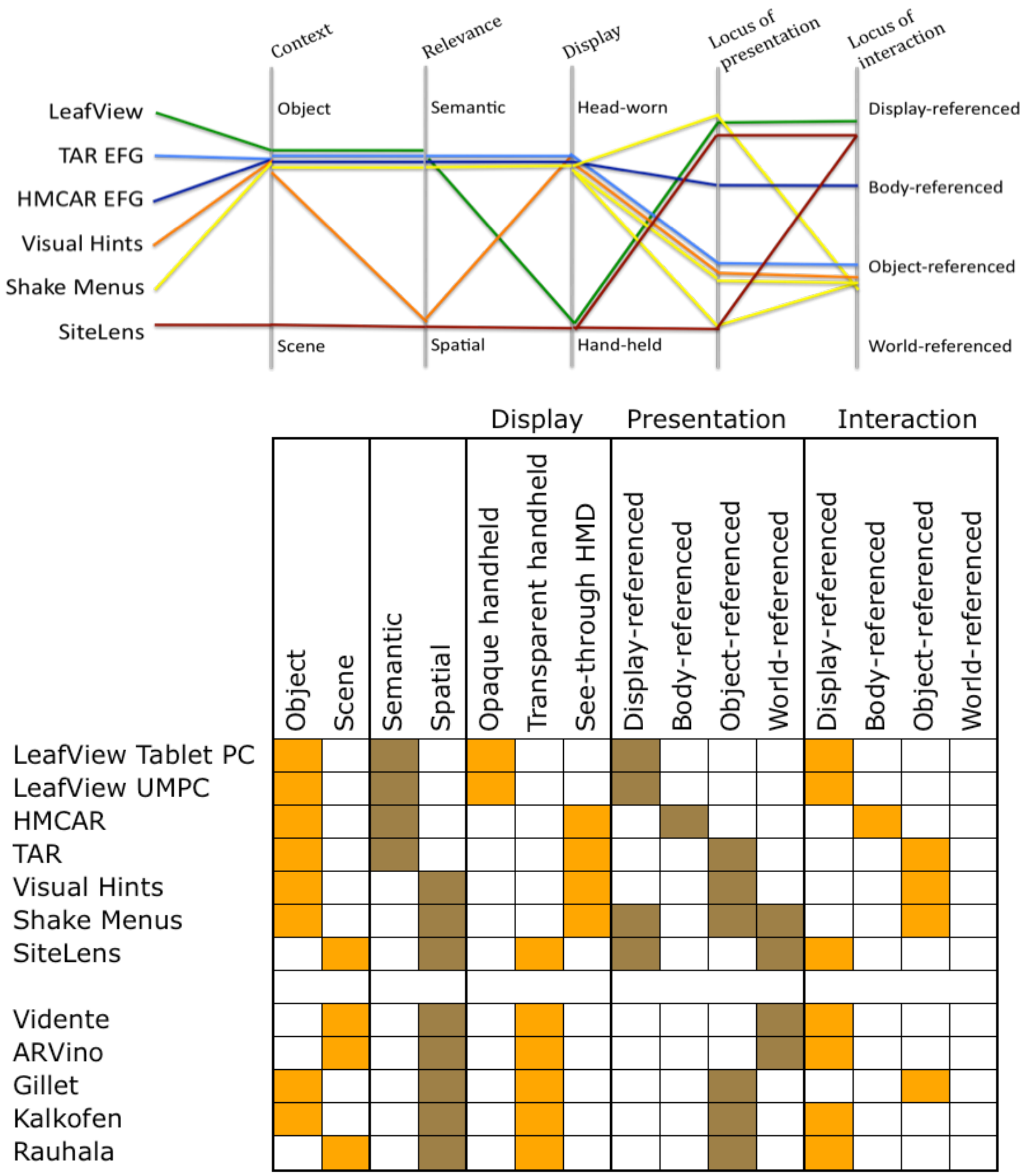


Figure 2.5 Techniques and prototypes from this dissertation placed in the design space using parallel coordinates (top) and compared with related work in a table (bottom).

With our characterization framework, we can now view the similarities and distinctions of the different prototypes in this dissertation across different characteristics.

Figure 2.5 shows the prototypes across characteristics using a parallel coordinate system [Inselberg 1990]. Each vertical axis represents a characteristic, a dimension, in the design space. An individual line through the set of vertical axes represents the set of characteristics for an individual prototype. The locus of interaction is the same as the locus of presentation for each of these, with two exceptions. All shake menus maintain an object-referenced locus of interaction and SiteLens uses a display-referenced locus of interaction.

In the beginning of this chapter, we used the example of the Tangible AR Electronic Field Guide to introduce situated visualization. We can now consider it through the lens of this framework. The physical context is an *object*, the leaf. The relevance of the data to the context is *semantic*. We identify the type of leaf and use that as the context. The spatial relationship is not the main driver of the visualization. Visualization is displayed in an *immersive head-worn display* and the locus of presentation is *object-referenced*. The figure is the combination of leaf and matching results and the ground is the physical scene. The focus shifts between virtual and physical leaves and the context is the remaining leaves. Interaction is also *object-referenced* through direct manipulation of the visualized data.

Similar characterizations for other work in this dissertation are reflected in Figure 2.5. In Chapter 9, we discuss design implications from this framework combined with our research.

3 Related Work

No research exists in a vacuum. A large body of prior and ongoing research has informed and inspired our own work. Here we highlight the major relevant threads and discuss their relationship to situated visualization. Research that is more specific to a particular technique in this dissertation is discussed in the relevant chapter.

The term *situated* has been used in several ways that are relevant to the goals of situated visualization. In section 3.1, we discuss situated learning, situated cognition, and situated knowledge, and situated interaction. Section 3.2 describes a branch of computing, *context-aware computing*, which uses the abstractions of context and situation to change, among other factors, the user experience. *Visualization* frameworks in section 3.3 provides a context for the breadth of work on visualization and subsections on techniques such as semantic zooming and focus+ context address specific task in visualization. Although situated visualization does not require augmented reality, AR plays an important role in this dissertation as the means for displaying visualization. Section 3.4 provides an overview of *augmented reality* research. Specific interaction techniques are particularly relevant for our research and are discussed in subsections on direct manipulation and tangible AR as well as magic lens and magnifying glass techniques. Finally, *visualization in augmented reality* in section 3.5 provides background on work others have done both situated and non-situated to explore visualization techniques in AR.

3.1 Situated Learning, Knowledge, and Interaction

Lave and Wegner [Lave 1991] describe situated learning as a way of learning in authentic context, embedded in the cultural, social, and physical environment that would normally involve that knowledge. This is in contrast to learning in a classroom away from such context. They view learning from a social perspective and argue that learning, such as an apprenticeship, is really about participating in a community of practice and understanding the framework in which the practice is conducted. Situated learning is related to a branch of cognitive psychology called situated cognition, which aims to study human behavior in real situations, where cognition is intimately tied to context. In contrast to a focus on memorizing and retrieving knowledge, situated cognition focuses on a process of perception and action that are coupled with an adapting and evolving environment. Beyond learning, knowledge itself may be situated. Haraway describes situated knowledge as a knowledge that allows the "imaginary and the rational—the visionary and objective vision—[to] hover close to together" [Haraway 1991]. Here, she is arguing for a thoughtful alternative that sits between pure relativism and objectivity. Knowledge and information result from the mix of creation and observation in a given context.

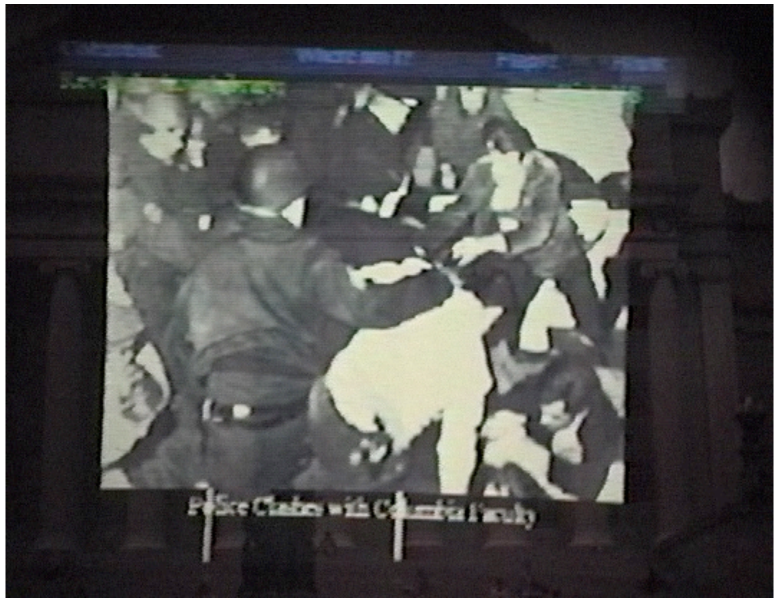


Figure 3.1 Imagery documenting the student revolt in 1968: Still image, overlaid on top of Low Library [Hollerer 1999].

Situated documentaries [Hollerer 1999] provide hypermedia presentations that are linked to specific relevant physical locations. Typically, this means that multimedia documentaries are embedded in the same environment that the events and documentary describe. For instance, a video of a student revolt in 1968 is displayed, using augmented reality, on the steps of Low Library on the Columbia University campus where the riot actually took place.

Figure 3.2 Close orbital view of Georgia Tech and midtown and downtown Atlanta with a single meso-cyclone plus its predicted path over the next 5 minutes.

Endsley [Endsley 1995] defines situational awareness as “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future.” Such awareness is used in tasks performed by military, law enforcement, firefighting and aircraft piloting to stay abreast of a quickly evolving situation. Situational visualization [Krum 2001] addresses this by providing context-aware information to the user. Two example applications discussed in this work are weather reports and 2D map display of locations for wheel chair access.

In a workshop proposal, Schmidt et al. relate the concept of situational awareness to situated interaction [Schmidt 2000] suggesting that adaptation of input and output modalities according to usage situation and recognized requirements take advantage of such awareness. Although Schmidt doesn't define the concept, Dey et al. associates the concept with context-aware computing [Dey 2001b].

3.2 Context-aware Computing

Situated visualization applications are necessarily aware of the context in which visualization is displayed. Perhaps some of the earliest work on context-aware computing comes from Schilit et al.'s work developing applications [Schilit 1994] for the PARCTab, an early small mobile device with a display and location-reporting. They define context-aware computing and go on to describe several applications: proximate selection, automatic contextual reconfiguration, contextual information and commands, and context-triggered actions. Dey later described a context toolkit as a framework for developing context-aware computing applications and introduced the abstraction of a situation [Dey 2001a]. Chen et al. provide a survey of context-aware applications [Chen 2000].

3.3 Augmented Reality

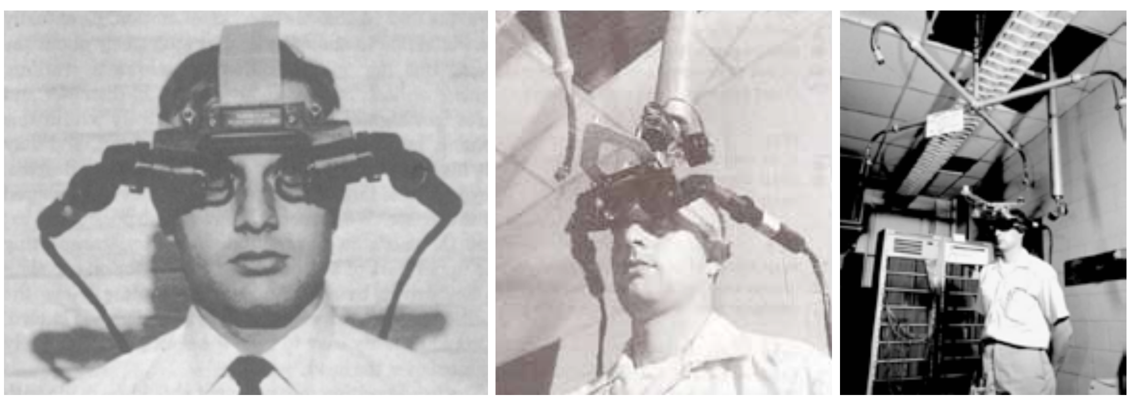


Figure 3.3 The first optical see-through head-worn display [Sutherland 1968]. (a) Display optics with miniature CRTs. (b) Mechanical head position and orientation tracker. (c) Ultrasonic head position and orientation tracker.

As discussed in the first chapter, augmented reality is a type of virtual environment in which computer graphics is overlaid on the physical world, registered in 3D, and interactive in real time. We will not address the entire history of augmented reality here, but highlight early work in the field and more recent relevant research. For two excellent surveys on the topic, we recommend reading [Azuma 2001, Azuma 1997]. We would like to discuss some notable works in AR.

Ivan Sutherland is credited with the earliest stereo, head-tracked, head-worn display [Sutherland 1968]. The display uses a pair of CRTs to provide imagery to the user through half-silvered mirrors so that the "displayed material can be made to either hang disembodied in space or to coincide with maps, desk tops, walls, or the keys of a type-writer. The head tracking used both mechanical and ultrasonic systems to acquire the head position and orientation of the wearer so that the appropriate perspective view of the graphics could be displayed. His two example images were a cyclo-hexane molecule and a four-sided room with labels for North, East, South and West. By moving their head, the wearer of the system could change their view of the three-dimensional model.

3.3.1 Magic Lens and Magnifying Glass Techniques

Bier and colleagues introduced Toolglass widgets and Magic Lens filters as a class of see-through tools and as a means of creating spatial modes in user interface systems [Bier 1993, Bier 1994]. The metaphor of a magic lens is used because the user looks at visual objects through a lens-like visual element, in the same way that one might hold an optical lens up to the eye and look through it. Everything viewed within the bounds of the lens is filtered in some way (e.g. making all elements viewed through the lens rendered in wire-frame or magnified). Viega extended the Magic Lens concept into 3D by considering a volumetric lens [Viega 1996], which acts much like a 2D Magic Lens but filters the view of anything inside a 3D volume. Rekimoto developed a variant for AR [Rekimoto 1995] which he referred to as a magnifying glass technique. Much like a physical magnifying glass, elements in the physical world are visually filtered or transformed when viewed through an AR magnifying glass. Looser and colleagues also developed techniques for 3D Magic Lenses in AR and examined fundamental interactions in magnification, object selection and manipulation, and information filtering [Looser 2004].

3.3.2 Direct Manipulation and Tangible Augmented Reality

Tangible AR provides physical affordances to augmented reality. To some extent, this is a combination of two established user interface paradigms: tangible user interfaces (TUIs) [Ishii 1997] and AR. A TUI in AR creates a physical embodiment or handle to a virtual object, which could be a tool, token, or container [Underkoffler 1999]. Manipulation of these objects takes advantage of our existing spatial skills.

Tangibility has been found to increase sense of presence [Hoffman 1998], enhance realism [Lindeman 1999], and increase visualization understanding [Belcher 2003]. This last study, by Belcher, is of particular interest because it provides initial evidence that tangible augmented reality could improve visualization. The study in question only looks at identification of connectivity in a random 3D graph of connected nodes, and our interest is to explore this beyond visualization in isolation.

3.4 Visualization

3.4.1 Visualization Taxonomies and Frameworks

A variety of taxonomies and frameworks have been developed to model and aid in the design of visualization. Some of the earliest fundamental work is credited to Bertin [Bertin 1981, Bertin 1983] who developed a set of graphic variables and the role of graphics for processing and analyzing information. In more recent years, Shneiderman

[Shneiderman 1996] presented a task by data type taxonomy along with a visual information seeking mantra, while Card and Mackinlay [Card 1997] developed a set of visualization subcategories. We are particularly interested in the work by Shneiderman, as it provides a means of comparing tasks across distinct visualization methods. Chi [Chi 2000] presented a data state model that represents the process of information visualization, which has been expanded and developed into a software toolkit called Prefuse [Heer 2005]. Ware also presents a useful view on visualization from the perspective of perception [Ware 2004]. Our work builds on these taxonomies and takes into account new characteristics of visualization in AR and situated visualization.

3.4.2 Level of Detail and Semantic Zooming

An important aspect of visualization is controlling and visualizing a variety of scales and levels of physical and semantic detail. One of the earliest examples of visualizing a change in levels of detail is Ray and Charles Eames' short film, Powers of Ten [Eames 1968]. As the camera moves away from the Earth, level of detail changes are noted as order of magnitude changes in the width of a square frame. Donelson's Spatial Data Management System [Donelson 1978] made the experience of level of detail changes interactive by giving the user joystick control over the "flight" around a large visual data surface. Zooming into the surface could reveal greater levels of detail or switch to alternative representations. Text could be revealed by zooming into items that represented text. Gurwitz and colleagues' MIDAS supported smooth continuous pan and zoom of an animated microprocessor simulation, changing the level of detail displayed as the user zoomed in and out [Gurwitz 1981]. Herot and colleagues further developed the spatial data management concept, using a hierarchy of icons and graphical data spaces [Herot 1980]. Friedell and colleagues extended this work with the View System, which dynamically generated graphics based on database queries and motion through space [Friedell 1982]. Furnas [Furnas 1986] formalized the idea of generalized fisheye views to use focus and degree of interest to change representations. In more recent years, Perlin and Fox introduced the term semantic zooming with the Pad system [Perlin 1993] which has been expanded by Bederson and colleagues with Pad++ [Bederson 1994] and PhotoMesa [Bederson 2001]. We build on this work to investigate interaction techniques for semantic zooming of search results on 2D displays and in augmented reality.

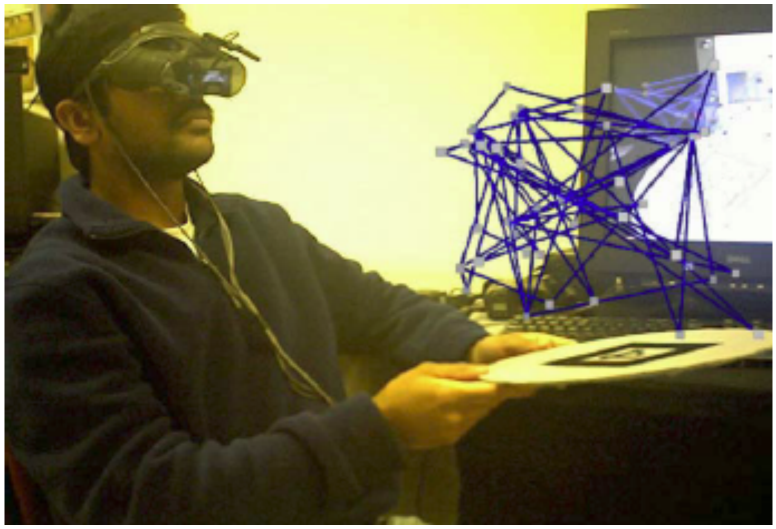


Figure 3.4: Tangible manipulation of connected graph visualization [Belcher 2003].

3.5 Visualization in Augmented Reality

In a sense, visualization is always present in AR. However, display of models and 3D agents are not the emphasis of our research. Several projects have investigated specifically at visualization in Augmented Reality. As mentioned in the previous subsection, Belcher [Belcher 2003] looked at 3D graphs of nodes in AR that were attached to visual markers. Slay [Slay 2001, Slay 2002] similarly displayed 3D graphs of nodes but explored aspects of interaction by using fiducials, marker cards, to turn on and off characteristics of the visualization. Fuhrman [Fuhrmann 1998a] created a Personal Information Panel used sliders and UI widgets to control a data visualization floating in space. DiVerdi and Hollerer [DiVerdi 2004] used distance from the viewer to an object as a means of changing level of detail. Recent work has also included enhancements to scene graphs to change or filter the visualization to represent focus and context [Kalkofen 2007]. Bell et al. [Bell 2001] developed view management systems for AR that focused on labeling and annotation rather than visualization in general.

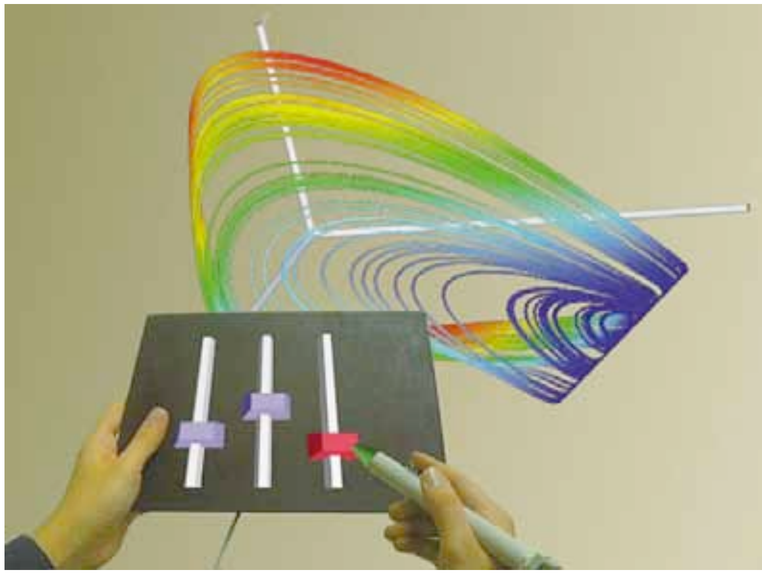


Figure 3.5 Personal Information Panel and data visualization in AR [Fuhrmann 1998a].

In addition to these, there are some types of visualization that we consider situated visualizations because they are related to the environment or an object in the environment. Examples of this can be seen in the AR visualization of magnetic fields around a physical model of a molecule [Gillet 2006] and display of GIS viticulture data on a physical space in the ARVino system [King 2005]. Both systems notably focus on data display and do not provide any direct means of interacting with the visualization—both systems use a keyboard for user input, although Gillet changes visual point of view through manipulation of the molecular model and King changes point of view by moving the tripod, laptop, and camera.

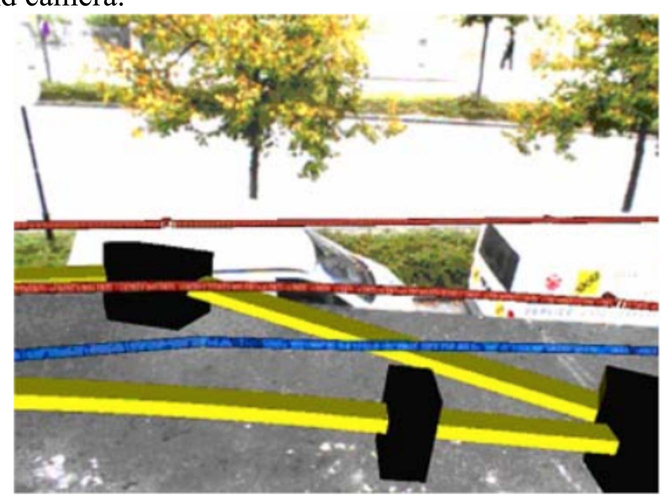


Figure 3.6 Subsurface pipelines and power cables visualized in Vidente [Schall 2009].

The Vidente project [Schall 2009] has been investigating visualization of subsurface features such as pipelines and power cables for utility field workers (Figure 3.6). Their approach takes geographic data models of subsurface features and transcodes them for visualization and filtering.

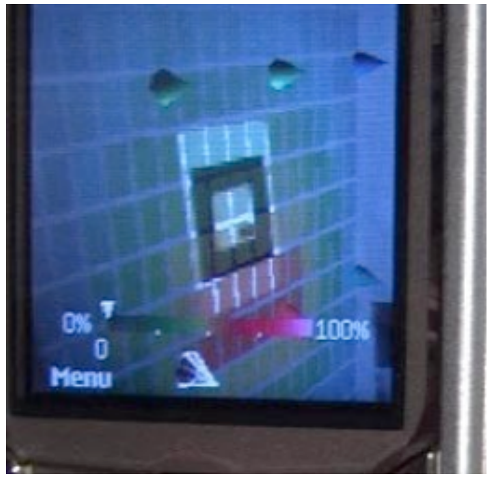


Figure 3.7 View through mobile phone screen of cones representing humidity levels [Rauhala 2006]

Rauhala et al. [Rauhala 2006] developed visualization that represents humidity levels on a 2D perpendicular plane (Figure 3.7). The background in their system is indoors, homogeneous, and static (using tile walls).

4 Improving Mobile Identification and Visualization

Figure 4.1 Smithsonian botanist using Tablet PC version of LeafView to identify plant species.

Mobile visualization brings the tools and techniques of visualization out into the physical environment and provides a variety of new challenges for human computer interaction. We consider this a first step in addressing the issues and phenomena surrounding situated visualization and investigate visualization that is based on physical context and viewed on devices that operate in close proximity to the objects of interest.

As a base line for comparison, we developed a series of mobile prototypes (the first of which is shown in Figure 4.1). These mobile prototypes provide visualization of botanical collections and search results that are based on object recognition and support comparison and inspection. In situated visualization terms, there is relevance between the leaf in the physical world and the visualization on the display, but the presentation is display-referenced on an opaque screen rather than displayed in the same visual coordinate system as the object or the physical world.

In this chapter, we first provide context for the application by describing an ethnographic field study of botanical species identification in the field. Ethnographic field studies in the context of human-computer interaction borrow from sociological techniques to make direct natural observations of the ongoing systems such as roles, processes, and environments, without intruding on or disturbing the natural system [McGrath 1995]. Our study, conducted through a combination of interviews and observing botanists in the field on multiple field trips, seeks to understand how botanists use existing methods to identify and collect plants under natural conditions. From this ethnographic field study, we developed a task analysis to describe the specific tasks and subtasks involved.

After discussing the task analysis, we present a conceptual model for the data and a system architecture derived from the task analysis that is used to support the identification and collection tasks. Next, we describe a series of user interfaces and interaction techniques that are built on top of the system architecture and developed through an iterative process of design, use, and evaluation that enables automated identification and collection of botanical species. We then describe a series of evaluations made through field experiments and interviews with expert users as well as demonstrations of the system and interfaces. In the context of human-computer interaction, field experiments are similar to field studies but actually disturb the existing system by changing a factor such as the tools used for a given task. Our results suggest that the system improves the speed of identification as compared to existing methods and that specific interaction techniques are predominantly responsible for the speed, efficacy, and ease of use of the system.

Figure 4.2 The site of our field study: Plummers Island, Maryland.

4.1 Botanical Species Identification: An Ethnographic Study

As an initial part of our investigation, we joined four botanist colleagues from the Smithsonian Institution on two collection trips (on March 15, 2005 and July 26, 2005) to observe their tools, process, and techniques for gathering plants and data. In both cases, the trips were conducted on Plummers Island (Figure 4.2), an island in the middle of the Potomac River in Maryland that is used for botanical field research because of its relative natural isolation from mainland flora and fauna. The island has been studied for over 100 years by regional botanists.

Motivation for fieldwork varies. The most common goals are collecting specific species for research on that species, discovering new species, creating a census for an area to look at an entire ecosystem, or identification of botanical species as they relate to other biological species (e.g. a biologist studying caterpillars wants to identify the plant that the caterpillar is eating) [John Kress, personal communication, October 5, 2005]. For these trips, the botanists sought to collect and identify species related to a census of Plummers Island.

In preparation for a trip, appropriate tools are gathered and packed in a backpack. Individual botanists brought their own paper field guides, notebooks, pencils, and cameras in separate packs. They shared pruning shears, a GPS receiver, collection bags, plant presses, and a list of species that need to be collected.

Figure 4.3 Paper field guide used for plant identification.

The botanists then travel to the collection site. When a plant on the collection list is observed, they first identify it using knowledge of the local plants and a paper field guide (Figure 4.3). Species identification involves inspection of multiple characteristics and comparison of these characteristics with field guide content. Characteristics for comparison include leaf outlines and venation (vein patterns) as well as plant structure, roots, and bark and fruit (if present). Leaves can be sufficient for identifying many plants, but closer examination of both the leaf and the plant is often required. In some cases, the plant cannot be identified using the field guide and must be compared with physical plant vouchers in a herbarium, often in consultation with an expert taxonomist specializing in the species. This process can take weeks and sometimes months, depending on the remoteness of the plant specimen and expertise of the botanist [Agarwal 2006]. In our field study, although all of the participants were practicing botanists, many of them had difficulty identifying species even with the help of a paper field guide. This is in large part due to the specialization required for academic research, where a large amount of detailed knowledge about a single species is more important than broad knowledge about many species.

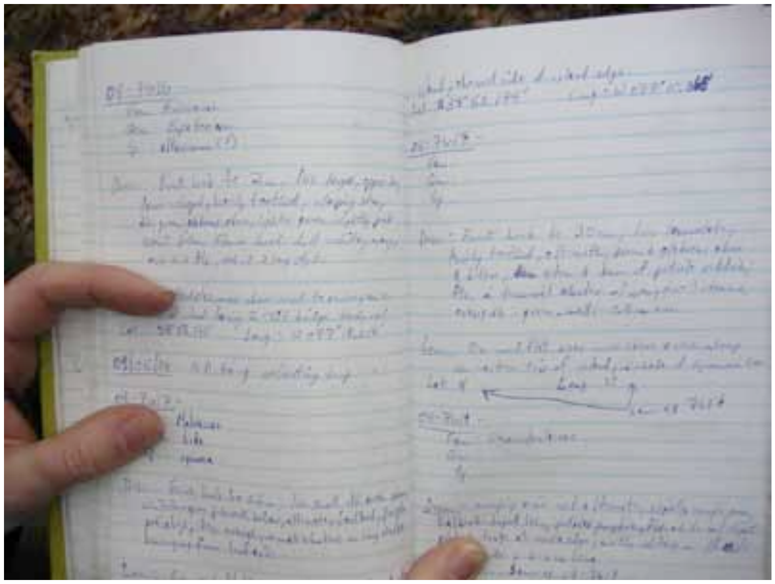


Figure 4.4 Photo of a botanist's field notebook used during the trip to Plummers Island for recording information and observations about collected plant specimens.

Once an initial positive identification has been made, the contextual information from the plant sample is recorded, including the name of the collector, location, date, time, and descriptions of the plant and its local environment. The information is gathered into a paper notebook (Figure 4.4) and saved for later use in labeling the specimen. The sample is then pressed in a plant press (Figure 4.5) and brought back to the herbarium (Figure 4.6a). Once its identification has been verified, it will be added to the herbarium specimen collection and serve as a voucher specimen (Figure 4.6b) for that particular instance of species collection.

Figure 4.5 Botanists pressing plants to be taken back to the Smithsonian Institution and archived in the U.S. National Herbarium.

A *voucher specimen* (often shortened to *voucher*) is a biological sample that has been collected and preserved in a herbarium or specimen collection. A voucher specimen is called a *voucher type specimen* when it represents the first collection of a particular species or is used as the canonical representation of the species. The voucher type specimen is used for identification of species and literally acts as the "voucher" for the existence of a species. It can be used for detailed comparison and may even be used for DNA sampling or similar physical testing. The voucher is labeled with information associated with its origins, such as the location and time of collection, the name of the person who collected the specimen, and the name of the person who made the positive identification. Multiple samples are often collected to create a set of vouchers that represent the diversity of a given species. A given trip may secure just a few samples or as many as 25 different collected species. In cases where the location is unfamiliar, the botanist may collect even more.

Figure 4.6 U.S. National Herbarium at the Smithsonian Herbarium (Photo by Chip Clark) (left). Botanical voucher specimen (right).

We observed that the inspection and comparison tasks were particularly critical in finding the appropriate matching species. The sample leaf was often held in the hand and inspected from multiple angles and distances. Discussion was common among botanists trying to identify the species and, in some cases, the species was left unidentified and taken back to the lab for inspection by other experts in the field and, ultimately, for comparison with a voucher.

The inspection and comparison process involved constantly moving and manipulating objects. At the same time, we observed that the botanists often used their hands to move through terrain and inspect plants near the path. Our observations of tools, process and techniques for gathering plants and data, along with contrasting requirements for physical manipulation and hands-free interaction, motivated our later design choices.

4.1.1 Tasks

In developing our conceptual model and requirements for prototype design, we first considered the task of species identification. We deconstructed the identification and collection process into the following subtasks:

- discovering and acquiring the unidentified specimen,
- inspection of plant characteristics such as venation or serration,
- identifying possible species matches,
- iterative comparison with the potential matches and inspection of details and characteristics,
- selection of the matching species, and
- collection and pressing of the specimen and associated contextual data.

Our architecture and user interface techniques are developed in support of these tasks.

4.1.2 Virtual Vouchers

The Smithsonian Institution maintains a large collection of botanical vouchers. Over the past few years, they have been creating a digital database of high-quality, scanned images based on the specimen voucher collection, which currently contains over 90,000 images [Belhumeur 2008].

We introduce the term *virtual voucher* to describe a digital representation of the botanical reference specimen in conjunction with its contextual and characteristic data. This data includes additional imagery of the whole plant and root systems, location and date of acquisition, name of collector and of identifier, regional information, articles about the specimen, and links to related specimens. More generally, the virtual voucher acts as a holistic virtual representation for any object that exists in the physical world and has multiple aspects or representations. To address this conceptual model and create a platform for learning about the introduction of technological aids to the task, we developed the LeafView system [White 2006c, White 2007b] that we present later in this Chapter.

4.2 Related Work

A number of research projects have developed electronic field guides to aid in species identification or fieldwork. These systems attempt to visualize information on mobile, hand-held displays. The FieldNote system [Ryan 1999] focused on context-aware data collection and Minimal Attention User Interfaces, which minimize the attention required to accomplish a task. This prototype was extended by the same team and used by ecologists in Kenya observing giraffes. While not specifically intended for identifying species, the system was used to support data collection in species observation. Similarly, CyberTracker [Liebenberg 1999] is a PDA-based system that has been used in a number of fieldwork projects for tracking animals. These systems aid in recording observations from the user. While not designed for fieldwork, Cyberguide [Abowd 1997] addresses a similar goal of providing mobile, context-aware information in the form of a tour guide. In this case, context was primarily the location of the user, obtained through GPS. More recently, Yeh and Klemmer developed ButterflyNet [Yeh 2006] as a mobile capture and access system for biologists to share notes and photos with colleagues. While our work has similar goals of data collection found in these systems, in contrast, our system uses computer vision to identify leaf specimens in the field during the collection process, based on Ling and Jacobs' IDSC matching algorithm [Ling 2005], and provides a visualization of related species and access to the entire database.

4.3 Architecture and Implementation

We refer to the underlying system as the Electronic Field Guide (EFG). The architecture of the EFG incorporates aspects from context-aware computing, image identification, and visualization. In particular, our architecture borrows from Chi's Information Visualization Data State Reference Model (or Data State Model) [Chi 2000] and Abowd et al.'s Cyberguide [Abowd 1997] as well as operating system queuing mechanisms [Silberschatz 2004].

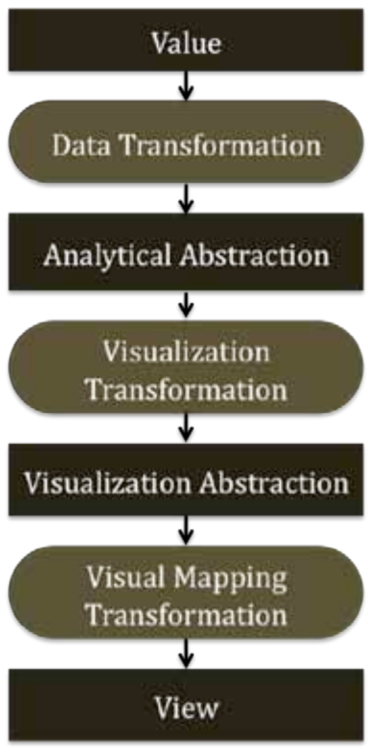


Figure 4.7 Information Visualization Data State Model adapted from [Chi 2000]

In the Data State Model (Figure 4.7), Chi describes four states of data for visualization (value, analytical abstraction visualization abstraction, and view) and the transformations between states (data transformation, visualization transformation, and visual mapping transformation). The data transformation converts data values to an analytical abstraction. For example, data transformation could involve taking a set of ranked matching species data and transforming them into an ordered list. The visualization transformation converts the analytical abstraction into a visualization abstraction. For example, it could transform an ordered list into a set of images displayed in row-major order. The visual mapping transformation then provides a view onto the visual abstraction. For instance, it might render an overview of the entire set of images or a zoomed view of a single image. Heer and Agrawala suggest a design pattern based on this model that they refer to as the Reference Model pattern [Heer 2006].

The advantage of this model is that we can separate out both analytical models and views with view transformations from the original data set. We find this useful in that we can take the original data from both the entire dataset or matching results, convert to hierarchies or lists, and then provide appropriate views such as quantum tree maps [Bederson 2002] or row-major layouts.

The Cyberguide tour guide has four service components: map, librarian, navigator, and messenger. The map component provides a set of maps for the system. The librarian maintains information about a tourist site. The navigator maintains the position of the user in the system. The messenger handles communication between the user and other users or systems. In our case, we are interested in context services that represent location, time, explicit meta-information from the user, and object matching.

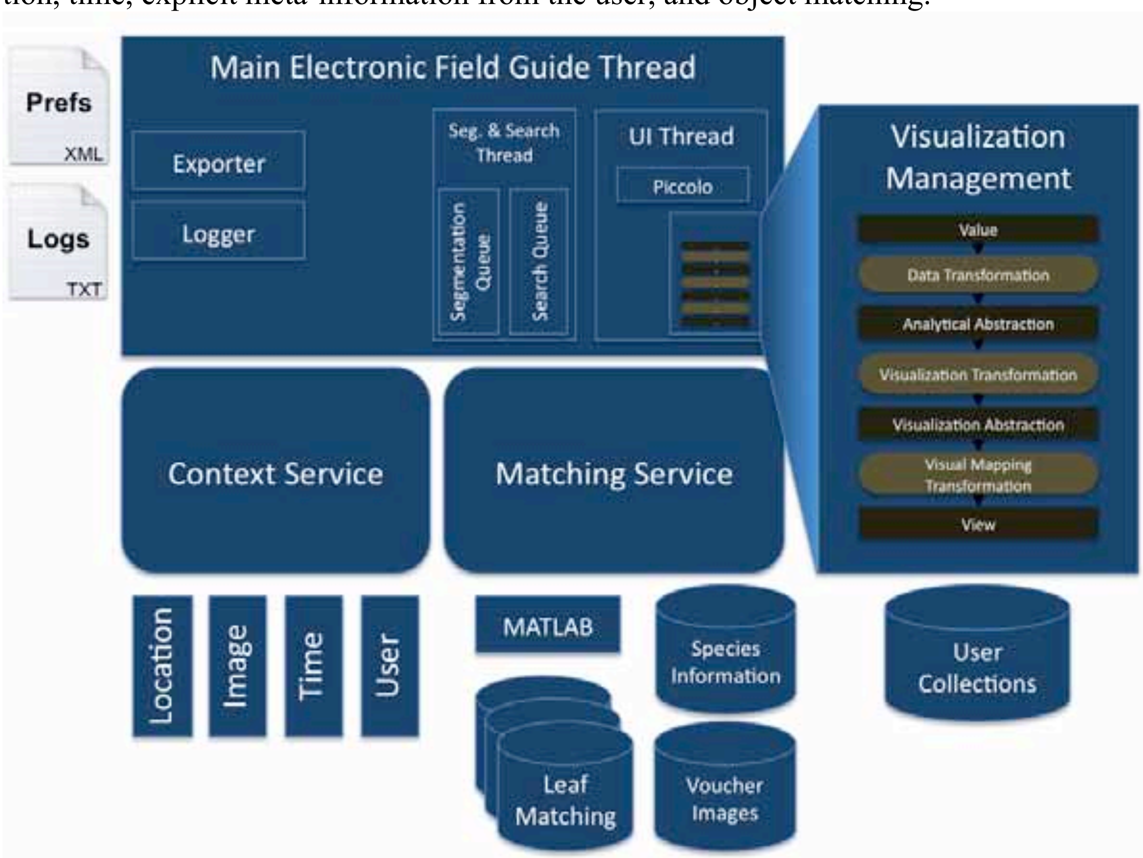


Figure 4.8 Electronic Field Guide Architecture

4.3.1 Electronic Field Guide

We build on the architectures in the previous section for our own Electronic Field Guide architecture (Figure 4.8). In the architecture diagram, blue rounded rectangles represent services, blue rectangles represent individual components, and blue cylinders represent data repositories. The dark brown rectangles represent the four data states and the light brown ovals represent data transformations.

The main EFG thread manages all other threads including asynchronous threads to handle segmentation and search queues, user interface management, and visualization management. We use separate threads here so that the user interface stays responsive while the search is active. The main thread also registers to receive updates from the context and matching services while handling data exporters and logging.

At start-up, thumbnails and species information for the entire data set are loaded into the browse collection, an alphabetically-sorted list for managing matching browsing of the data set. The browse collection is transformed into a visualization abstraction, a list of species sorted alphabetically along with a default overview view for presentation to the user. Next, a previous history of images, matching results, and context data are loaded from the user database into the history collection, a date-ordered and sorted list for managing field trip history. The history collection is then transformed into a horizontal list of images and presented to the user.

When a new image is acquired, the context service notifies the main EFG thread, which then creates a new sample object with the associated sample image and context (such as location, collector, and date). The main EFG thread adds the sample image and context information to the user database and associated image store and places the sample object in the segmentation queue. We save all data to disk as soon as it is acquired to avoid accidental data loss.

A separate, asynchronous segmentation and search thread takes sample objects off the queue and submits the sample image to the matching service in order to run the currently selected segmentation on the image. Once segmentation is complete, the main EFG thread is notified by the matching service while the sample object and associated segmentation are placed on the search queue. The segmentation and search thread removes sample objects from the search queue and submits them to the matching service. Once matching is initiated, the results are associated with the sample object while the main EFG thread is notified. The asynchronous matching service continues in the background.

After notification from the matching service, the results of segmentation and any completed matching results for a given sample object are added to the user database and associated image store. The sample object is placed into the history collection. As results from the matching service arrive, the matching species results from the search are placed in a result collection, a rank-ordered sorted list for managing matching results, and associated with the sample. The result collection is transformed into a visualization abstraction and default overview view that are displayed to the user. As matching results are added to the result collection, the result visualization and view as well as the user database are updated.

A user interface thread handles all user input, updates to the user interface, and rendering the user interface to the screen. We use Piccolo [Bederson 2004], a 2D scenegraph optimized for zoomable user interfaces [Perlin 1993], to display all visualizations. The architecture includes support for logging user activities for UI evaluation, software debugging, and analysis of computer vision algorithms. The main EFG thread writes all logged data to a separate log file. User preferences are saved across sessions in an XML-based preferences file. The EFG architecture is used by two prototypes, one running on a Tablet PC and one running on an Ultra Mobile PC (UMPC), discussed in this chapter.

The architecture also provides the underlying matching service for the two AR prototypes discussed in the next chapter.

4.3.2 Context Service

This service manages image acquisition and maintains a context state for any given moment and for every acquired image. For image acquisition, the service supports the UMPC built-in camera, Bluetooth phone cameras, and WiFi cameras. We discuss this in more detail in the LeafView Tablet PC and LeafView UMPC user interface sections of this chapter. Multiple cameras can be used simultaneously with the same device. A context object for each image sample is generated from location data, time-date stamps, and user name, and can be extended for other characteristics. Location data is acquired through GPS, either via USB or Bluetooth, and can be changed to accept other types of location techniques. New types of context can easily be added to the context service.

4.3.3 Matching Service

Our architecture provides two levels of interconnection in order to support a diversity of matching algorithm choices. A root search class affords flexibility for trying new segmentation and search techniques. Our primary matching technique, implemented in both MATLAB mfiles and compiled MATLAB, uses a color-based EM algorithm for segmentation and the Inner Distance Shape Context algorithm [Ling 2005] from Ling and Jacobs for identification. At the most basic level, a new segmentation or identification algorithm can be tested by replacing existing MATLAB files with new ones that follow the same API. This has enabled the project to test changes to both segmentation and search without recompiling. In addition, new search algorithms can be added by creating a subclass of the search class, a method we use to include a new search algorithm based on a compiled dynamic link library.

The matching service accepts an image for matching and asynchronously returns a segmented image and the set of best possible matches. A given matching process can be terminated at any time if the user decides that a segmented image is poor. The service also accepts parameters, adjustable through a user preference pane in the user interface, to change the matching process. For instance, one matching algorithm tested combines both leaf shape and the amount of serration. The weighting of the two characteristics can be changed through this service.

4.3.4 Datasets

A variety of data sets are required for the system to function, including the active leaf matching data set, the voucher image data set, the species information data set, and a user collection data set. We currently have over 90,000 voucher images in our voucher image data set and three leaf matching and associated species information data sets:

- Flora of Plummers Island. 5,013 leaves from 157 species.
- Woody Plants of Baltimore-Washington, DC. 7,481 leaves from 245 species.
- Trees of Central Park. 4,320 leaves from 144 species.

A data object provides access to a given dataset. For the matching service, a collection of images of all the leaves used to match against the sample (leaf matching data sets) is kept in the file system. Separately, we keep a database of information about every species including all voucher images (voucher image data set) for a given species and species-specific data, such as descriptions and common names (species information data set). On initialization, the active leaf matching data set is used to create a collection of species for browsing. Finally, a user database keeps information about all samples and collections for a given field trip or set of trips. This database maintains the individual sample images, matching results, and context information.

4.3.5 Visualization Management

This component maintains the analytical abstraction, visualization, and views for every visualized collection of data as well as transformations from one state to another, as described in Figure 4.7. This currently includes the entire database for browsing, any results from a specific matching query, and the history collection representing all samples collected. Our primary analytical abstraction is a collection, which is a data structure that can be ordered based on data characteristics such as date, ranking, or species name. We also implement a hierarchy abstraction, which is used for quantum tree map [Bederson 2001] and taxonomy visualization. The primary data objects in these data structures are samples and species. Every sample maintains the acquired image, segmented image, matching results as a collection of species, context information, and user-identified match. Every species includes the set of leaf images used for matching to the given species, voucher images from the herbaria, and all information about the species such as the latin name, common name, physical description, and habitat.

Visualizations supported in this architecture include linear text lists, hierarchical text lists, vertical and horizontal image lists, row- and column major grids, and quantum tree maps. They are used for browsing the full data set, visualizing results, and viewing history. Sample and species are the primary visual nodes or objects used in our 2D scenegraph. New types of collections or visualizations can be created by subclassing primary collection or visualization classes. Examples include a textual list with species name, a row major grid of all species images, and a quantum tree map based on variable numbers of image clusters (Figure 4.10).

A visualization can be shown in one or more views, and views typically default to an overview of the visualization. Through user interaction, described in the next paragraph, the view on the visualization can be changed. For example, the initial view of a visualization of a row-major grid of matching species might start off showing an overview of all the species. By changing the view, the user can zoom into a single species or zoom even further in to inspect a voucher image.

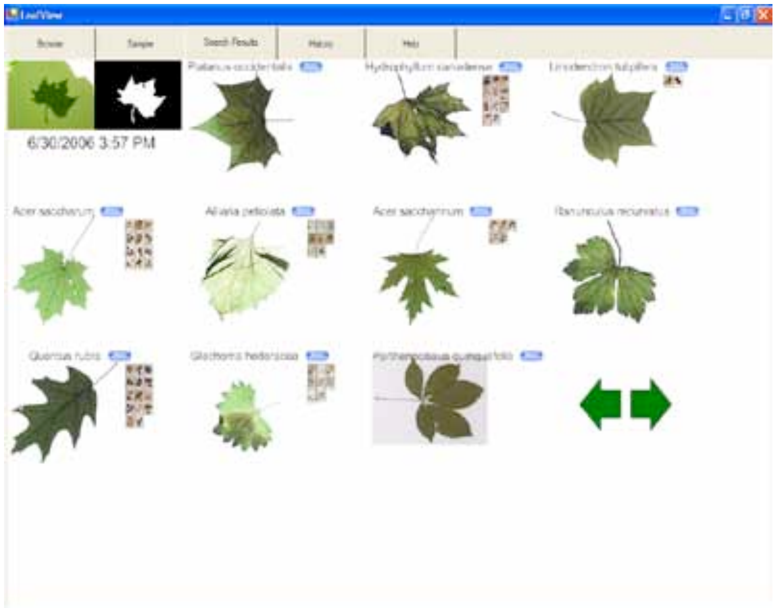
Separate from individual views, we use interaction handlers to change and experiment with user interaction. For instance, one handler maps horizontal and vertical dragging to change the magnification or semantic zoom of a particular view. Another handler supports touching or clicking on different parts of the view to expand or collapse levels of zoom. A rubbing handler supports diagonal rubbing [Olwal 2008] to move in and out of different semantic zoom levels. A list-dragging handler provides a physics-like dragging interface, modeled on the technique used on the iPhone [Apple 2009], for quickly browsing lists.

4.4 LeafView Tablet PC User Interface

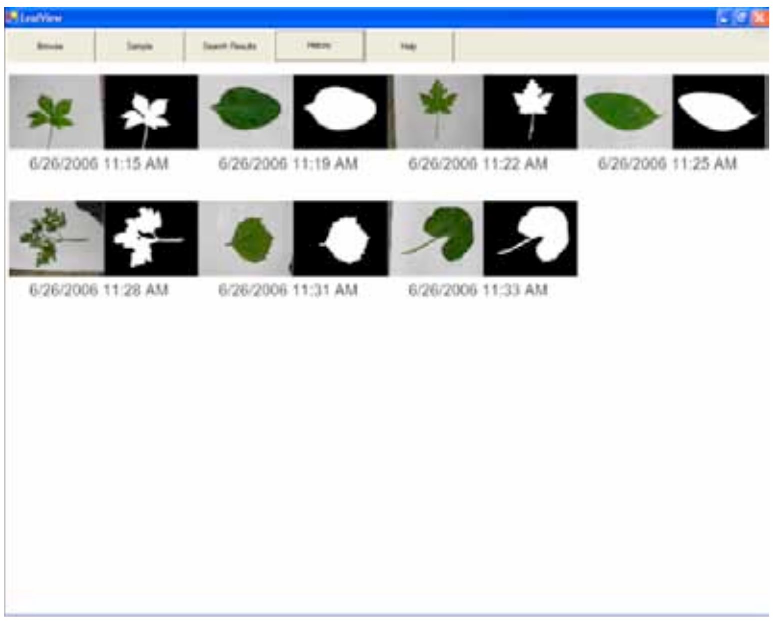
(a)

(b)

Figure 4.9 (a) A photo is taken of the leaf specimen. (b) The image is transferred to the Tablet PC and displayed along with a segmented image to reflect “what the computer sees.”



(c)



(d)

Figure 4.9 (cont.) (c) Results are displayed alongside the original image for comparison. (d) A history of results is displayed for tracking a field trip or series of trips (fourth image).

In this section, we describe the LeafView Tablet PC user interface, our first version, and a set of iterations for a user interface to the Electronic Field Guide. The LeafView user is presented with tabs for browsing samples, search results, history, and help. The collection process starts by taking a photo of the leaf (Figure 4.9a). The image is immediately transferred through an IEEE 802.11g or Bluetooth (we support both) wireless network to the tablet system. On arrival, the image is displayed in the samples tab (Figure 4.9b), and the image is segmented by vision algorithms developed by Ling and Jacobs [Ling 2005]. A thumbnail of the sample is also placed in the history tab (Figure 4.9d) and all contextual data about the sample, such as GPS location, collector, and time/date, are stored in the user database. When segmentation completes (ca. 6–11 seconds from shutter release on the Tablet PC, described in a later section), a search is automatically initiated in the background. The search component uses the inner distance shape context (IDSC) algorithm [Ling 2005] to match plant species, and we provide hooks for integrating other algorithms. Once the search is complete (ca. 35–40 seconds from shutter release), the ranked results are displayed in the results tab (Figure 4.9c) on a zoomable user interface [Perlin 1993] canvas as a set of virtual vouchers, which are initially displayed as individual plant leaves.

Human interaction is required because the vision algorithms are not perfect. The user can pan and zoom to inspect individual virtual vouchers and compare them with the plant sample. Semantic zooming, discussed in our related work chapter, is accomplished by either tapping on a virtual voucher to zoom in a level and reveal sets of voucher images, identification information, and textual descriptions, or by dragging up or down for continuous zooming. Once the identification has been verified by the botanist, a button press associates the identified species with the sample. A zoomable history of samples can be browsed to recall prior samples and search results. The entire dataset can be browsed as a list, as a row major collection like the search results or as a quantum tree map [Bederson 2002] (Figure 4.10).

The software was developed using C#, MATLAB, and Piccolo for the zoomable user interface. The hardware consists of Motion Computing LE1600 (Windows XP, Pentium M 1.5 GHz CPU, 1 GB RAM, 40 GB HD, with daylight-readable display) and Lenovo ThinkPad X41 (Windows XP, Intel Pentium M, 1.6 GHz CPU, 1.5 GB RAM) Tablet PCs, a Delorme Earthmate GPS, a Nikon Coolpix P1 Wi-Fi camera, and a Sony Ericsson T616 Bluetooth camera phone.

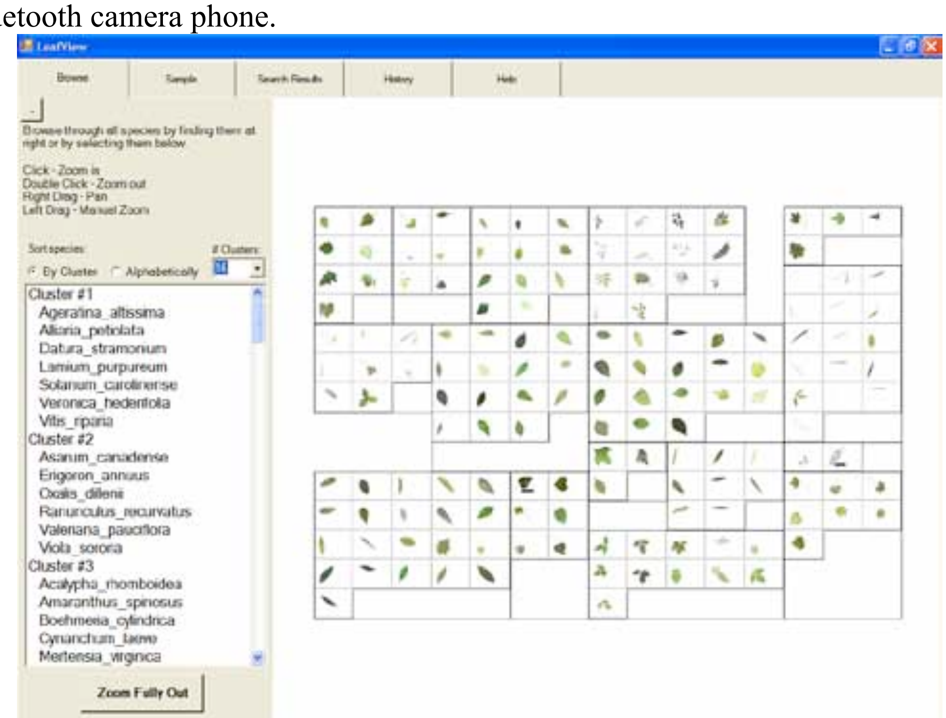


Figure 4.10 Quantum tree map view generated from hierarchy constructed using k-means clustering.

4.5 Evaluation

In order to evaluate the system, we have used multiple methodologies of analysis. These include a situations, task, and user analysis as well as field experiments where the device has been taken out into the field and used for collection. In this section, we describe these evaluations and discuss results.

4.5.1 Situation, Task, and Users Analysis

Olsen suggests [Dan R. Olsen 2007] that the context—expressed in terms of the situation, task, and users (STU)—can be used to evaluate the quality of user interface systems. With that in mind, we consider two different STU contexts. In the first, we have iteratively designed an application for botanists (end users) to identify botanical species (end task) in the field (end situation). Our second context is the architecture we have developed to enable HCI and vision researchers (users) to test different matching algorithms with different data sets (task) for use in real outdoor mobile environments (situation). We evaluate our system in terms of these two STU contexts by considering importance, whether it addresses an unsolved problem, generality and flexibility.

4.5.1.1 Botanical Species Identification

With regard to our first context, we consider the botanist *users* themselves to be important because they help humanity better understand nature. The *task* is certainly considered important by botanists. One Smithsonian botanist colleague, after demonstrating the system to biologists at a conference in Mexico last summer, said: “scientists [at the conference] were exuberant about the everyday use of this system” [W. John Kress, personal communication, April 4, 2008]. The *situation* is important because much of the identification research requires work in the field.

The problem of vision-assisted botanical species identification has not been previously solved. For example, during a census of trees in Barro Colorado Island, Panama (350 species and 50,000 individuals), non-specialists may not be able to identify tree species and even for specialists, our colleagues tell us that our system will “vastly improve the speed and accuracy of those censuses” [W. John Kress, personal communication, April 4, 2008].

We also believe that our system will generalize to both new types of users and new tasks. Other professionals who need to identify plant species include entomologists, ecologists, bio-surveyors, and customs inspectors. Nature enthusiasts of all ages have expressed interest in using the system to identify plants in their surroundings. At a recent demonstration at the Smithsonian, biologists inquired about using the system for fish, butterflies, and birds. While the current matching algorithm [Ling 2005] would not be appropriate for these domains, the system itself generalizes to support these potential alternative data sets and new matching algorithms.

4.5.1.2 User Interface and Computer Vision Research

In the second context, we consider the users to be important as well, because of their contribution to the UI community. The value of the task is in extending our body of knowledge, and the situation is important as it represents the practicalities of testing in real environments.

Our system provides a flexible way to test out new ideas for interacting with vision-based recognition algorithms. We have been able to compare different segmentation and search algorithms and add new data sets. In addition, portions of the underlying architecture for identification have been used in our AR UI [White 2006b]. Data exported from the system has also been used to analyze the efficacy of computer vision algorithms in diverse lighting conditions.

4.6 Plummers Island Field Experiment

To gain insight into the use of the system in the field, four of our six botanist collaborators used the LeafView prototype on Plummers Island (Figure 4.11). We joined botanists on a trip to observe them using the system to identify species and collect data.

Figure 4.11 Smithsonian botanist taking a photo using the LeafView Tablet PC prototype.

4.7 Results and Discussion

In this section, we discuss results from observations of use in the field as well as feedback from botanists about their experience.

4.7.1 Situated Visualization

One of our first questions was whether a visualization that is relevant to the immediate context of an object of interest would provide any benefit. That is, do we improve the task of species identification by visualizing matching species and information about those species in the context of fieldwork? Our initial observation is that bringing the data and visualization in closer proximity to plant species significantly improves the experience. We base this observation on comments from botanists who felt that having the virtual vouchers present in the field, using the system to automatically match the species, and supporting quick inspection of the visualized data improved their ability to identify botanical specimens.

4.7.2 Supporting Comparison and Inspection

As discussed by Ling and Jacobs [Ling 2005], the correct species match is found in the top 10 results 98.5% of the time by the algorithms we are using. However, some inspection and comparison is still required. From our ethnographic study, we found that the inspection and comparison tasks often start at a high level, with general shape, and then focus in on distinguishing details, such as venation or edge serration. Aspects not represented in the voucher images may also be examined, such as plant height.

As part of our design, we support comparing the original leaf with high-resolution species voucher images that can be accessed through semantic zooming on any virtual voucher. Additional information about the plant species and context are also maintained in the virtual voucher, but they are not shown until requested by the user, also through semantic zooming. If uncertainty remains, we support the ability to associate a new sample with multiple matches and save the entire matching results.

We have also found it useful to provide access to the full set of species in the database. When a botanist believes a plant species is present in the data set, but the plant is not matched, we make possible visual and textual browsing of the entire data set used for matching to give closure to questions regarding inclusion in the data set.

Our botanist colleagues verified that the prototype was effective, in place of a physical voucher, for examining detailed characteristics such as venation.

4.7.3 Interacting with Vision Algorithms

Vision algorithms are often treated as black boxes that provide no feedback on success or failure modes. Although some aspects of the species recognition algorithm currently used in LeafView do not directly correlate to visual representations, we can provide feedback on the segmentation of the leaf image. We find that the quality of the segmented image is related to the accuracy of the results and botanists can make minor adjustments based on seeing the segmented image.

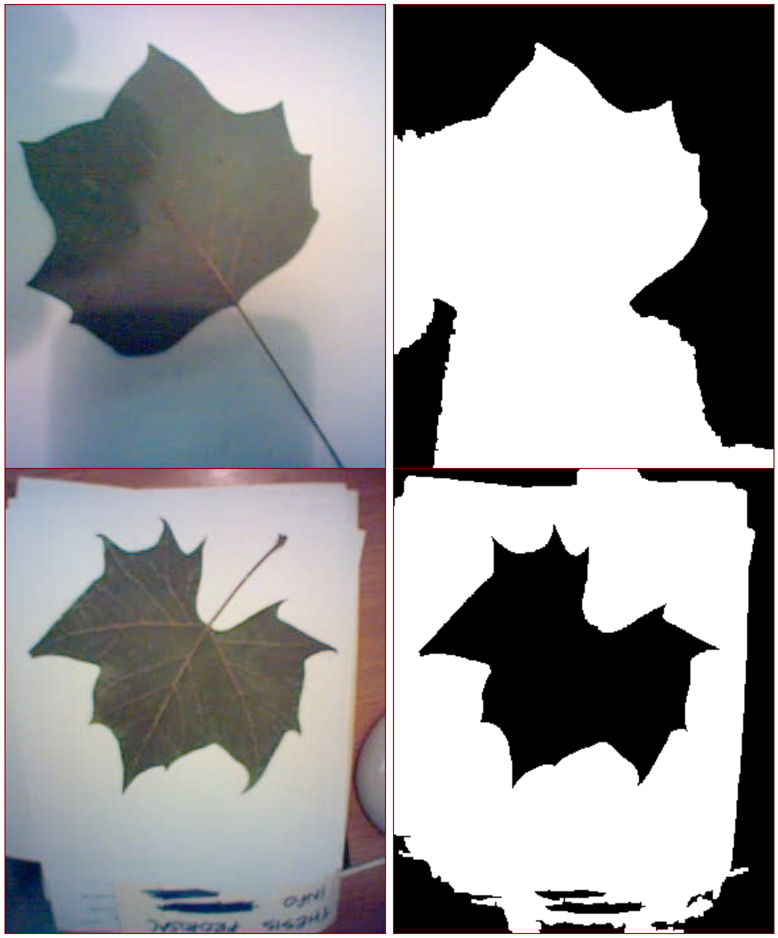


Figure 4.12 Examples of original image (left) and visual feedback to the user (right) of *Liriodendron tulipifera* (Tulip Tree) (top) and *Platanus occidentalis* (Sycamore) (bottom).

To address this, we display the segmented image alongside the sample image while the IDSC algorithm is executed. This provides immediate feedback regarding the quality of the segmentation. For example, if a shadow causes poor matching, the botanist can observe this and retake the photograph to fix the error (Figure 4.12). In observing botanists using the system on Plummers Island, we found that providing feedback by displaying the segmented images enabled them to retake better pictures than those that originally produced bad matches.

4.7.4 Individual and Batch Identification

During our field studies of the LeafView prototype, we observed that botanists performed identification in two very different ways. In the first approach, an image was taken and the system was immediately checked to see the results of the search. The retrieved virtual vouchers were inspected and a match was chosen. The botanist then went on to find another leaf to collect. In the second approach, the botanist took a series of pictures and then used LeafView to review and match the images, in some cases also comparing multiple samples to see if they were the same species.

In the first approach, the prototype successfully fit into the botanists' preexisting collection patterns, as they had originally described and demonstrated to us, and matched a large variety of plant specimens. In contrast, the second approach was not predicted by either the botanists or us, and our earliest prototypes did not support it. As one botanist put it, “the system gets very confused if you send too many [images]”.

We considered this a design opportunity and changed the conceptual model and user interface based on observed use. A segmentation and search queuing model (Section 4.3) was added to the architecture and the history was changed to represent and support this.

4.7.5 Queuing and History

The history display acts as both a queue and as an indication of the stage of progress for each leaf sample, supporting both individual and batch identification as described above. When a leaf is photographed, it immediately appears in the sample tab and is inserted into the history. Multiple photographs can be taken in succession and each new leaf will be placed in temporal order. As segmentation is completed on a particular leaf, the segmented image is shown both in the sample tab and alongside the leaf image in the history (Figure 4.9d). Once the IDSC algorithm and final matching has completed, the results are shown in the results tab and reflected in the history. At each stage—from photo, to segmentation, to matching—the botanist can observe distinctions across images, so that poor quality results can be improved. Images can also be deleted from the history if they are immediately observed to be problematic. This addresses an earlier comment by one of the botanists regarding their desire to see relationships across matches for a collection. Over time, they wanted to “…display the name of the plant selected for a match. That way the user would know what name was selected for something they saw earlier in the day...” (anonymous participant).

4.7.6 Collection with identification

Our initial design was primarily focused on identification. While this was supported by the six botanists who are directly collaborating with us, some other botanists have reacted with some apprehension to the idea. We discovered these reactions at the Smithsonian Science exhibition, discussed in Subsection 4.11.1. When we described the system as a general tool for collection, feedback on the system was very positive; however, we sometimes detected hesitancy when we focused on the identification mechanism. In the course of showing the system to attendees, we found that several botanists were comfortable with the idea of a system that helped them identify specimens in the process of collection, in contrast to a system that simply provided identification.

We suspect that this is due to the perception that a pure identification system is somehow replacing the botanist, while an intelligent collection system or electronic field guide maintains the locus of control with the botanist. In a subsequent conversation with biomedical informaticist Ted Shortliffe [personal communication, July 27, 2006], we learned that he had experienced similar responses from physicians with regard to automated diagnosis systems.

While this may not be an issue with non-experts, it is worth remembering when designing for and presenting to groups with sensitivity towards their own knowledge. A relatively minor (to us) change in system emphasis made a significant difference in perception of the system.

4.7.7 Improvements

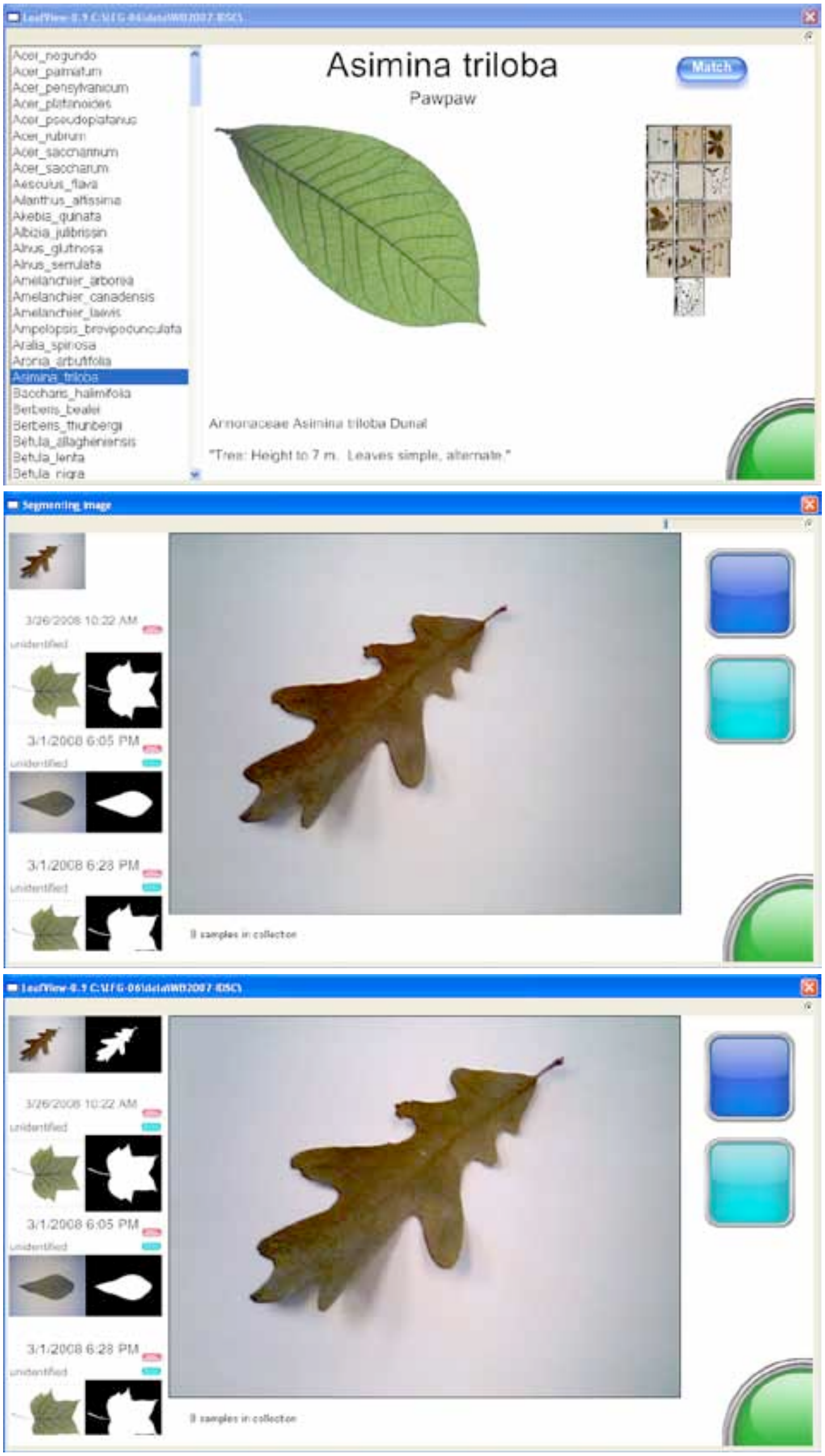
While our initial evaluation of the system provided many positive responses, there were still several opportunities for improvement. First, the device was considered too large and bulky to be carried for long periods during fieldwork. There was also a concern that the stylus might be lost and would be hard to replace in the middle of a trip. Second, while sending an image wirelessly to the Tablet PC from a camera was viewed as “magical,” there were too many devices to carry. In addition, the live leaf images on the camera were still viewed as separate from the system. In other words, we observed that the object of interest, the leaf specimen, was still seen as quite separate from the visualization. Finally, navigation of the visualization using a combination of single and double taps of the stylus for semantic zooming in and out, or dragging the stylus for continuous zoom was difficult.

To address these issues, we worked with our botanist colleagues to design a new prototype that would be smaller, touch-based (using a finger, rather than a stylus), and incorporate live video for capturing individual leaf images in the scene. In addition, our colleagues made improvements to the matching and segmentation algorithms, and we collected new data sets for the woody plants of the Baltimore-Washington, D.C. region. In the next section, we describe the improved user interface and evaluation of that interface.

4.8 LeafView UMPC User Interface

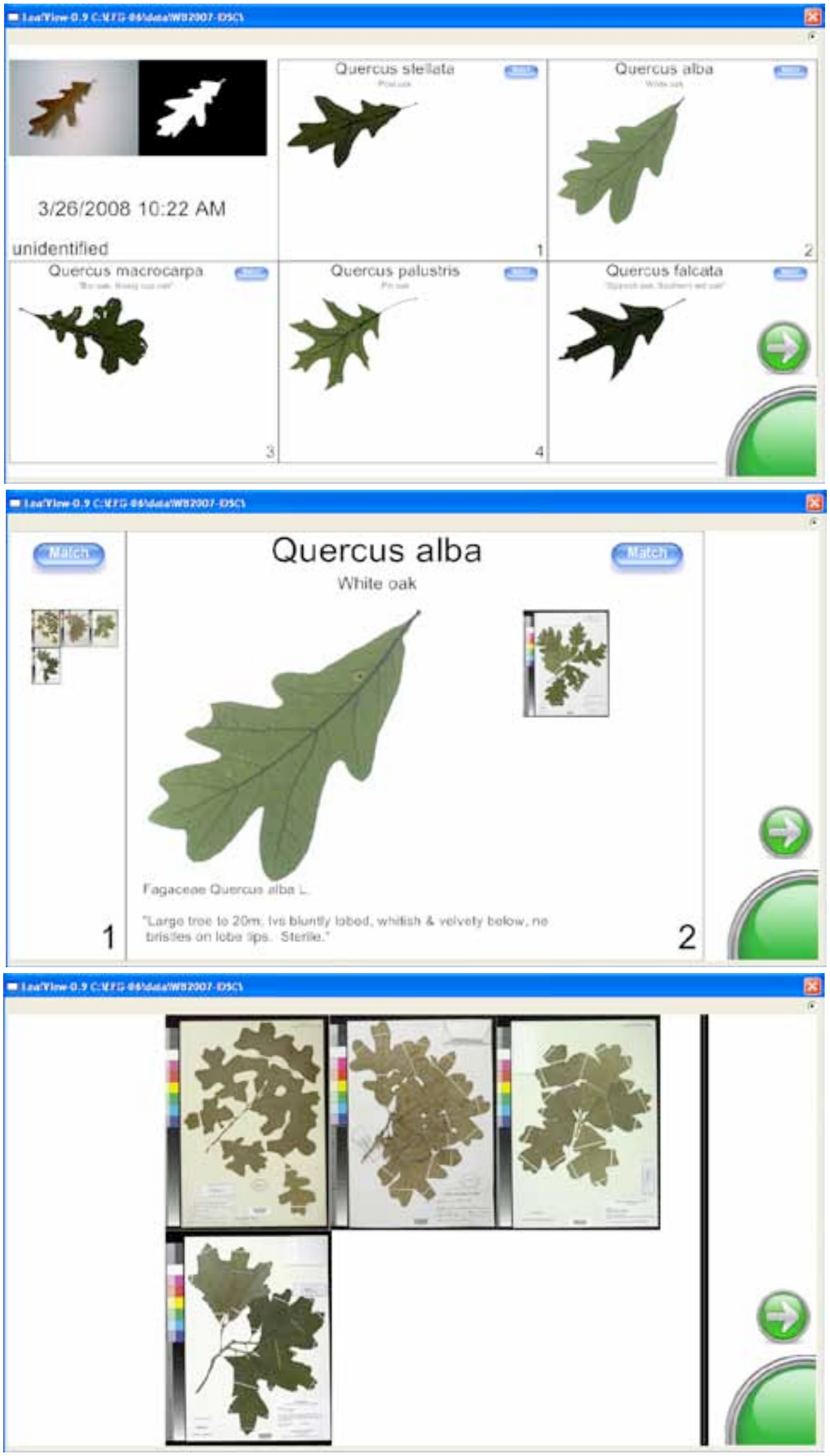
(a)

Figure 4.13 (a) LeafView UMPC.



(b) (c) (d)

Figure 4.13 (cont.) UI screens for (b) browsing, (c) capture, (d) visual feedback.



(e)

(f)

(g)

Figure 4.13 (cont.) UI Screens for (e) matching results, (f) leaf inspection, and (g) voucher inspection.

After several iterations of the interface and software using the Tablet PC, a new hardware platform using a Sony VGN-UX390N Ultra Mobile PC was adopted (Figure 4.13). This platform included a built-in camera in the device and Bluetooth GPS (GlobalSat BT-338). The new platform, based on feedback from the users, traded the larger screen, greater weight, and detached camera of the Tablet PC for the smaller screen, lighter weight, and built-in camera of the UMPC. The result is a hand-held system that provides a live video window and surrounding interface elements for identification. The UI is designed so that all interaction, such as a button press, is on the screen to make the design hardware agnostic.

The main screen is the image capture screen (Figure 4.13). At the center is a live video feed from the camera on the back of the device. Below the live feed is contextual information: the size of the current collection, name of the current user, and GPS coordinates. Pressing the green button in the lower right corner captures the image and initiates segmentation and search. The image immediately appears as a new sample in the history and, once segmentation is complete, the black and white segmented image is displayed. A progress bar shows when the search has finished and tapping on the image in the history shows the first five possible matches in ranked order.

Touching the green match button on a given species associates that species with the sample leaf in the history. If no match is found in the first set, the right arrow button on the last result displays more results. If a match is still not found, a list of all species in the dataset is displayed along with the possibility of marking the leaf as completely unknown and potentially new species. Tapping the green button at the lower right on this screen or any screen other than the capture screen returns to the capture screen. We chose to overload this button because image capture and identification was the main task for our botanist colleagues, and this made it possible for it to be accomplished from any screen with at most two taps. The button is also located conveniently near the user's thumb while they are holding the device.

On the left hand side of the capture screen is a visual history of images that have been sampled. The history provides context for the field trip and can be scrolled by dragging the history with a finger on the touch-sensitive display to see previously sampled images in the collection. Double-tapping any sample in the history brings up the set of matches found for that image.

The top right dark blue button displays a screen for browsing the data set. The browsing screen displays a list of species names on the left side of the screen. Touching any name brings up images of the species, textual descriptions, and a zoomable UI [Bederson 2004] for viewing all the voucher specimens [White 2006b] for that species. Zooming is either continuous or quantized. We use rubbing [Olwal 2008] for continuous zooming, single taps to zoom in one level, double taps to zoom out one level, and dragging to pan the images.

The light blue button displays a preference and configuration screen that allows the user to enter the collector's name or select a matching algorithm. The field collection history can also be exported from this screen. To aid botanists in their data collection, the export puts all images in a single folder and creates a CSV file that includes all data collected about each individual sample.

The software was developed in C#, MATLAB, and Piccolo [Bederson 2004]. The system runs on a Sony VGN-UX280P UMPC with Windows XP and a Sony VGN-UX390N UMPC with Microsoft Vista, and a GlobalSat BT-338 GPS.

4.9 Wind Forest Field Experiment

Figure 4.14 Two Smithsonian botanists use LeafView UMPC prototypes to identify plant species.

In the summer of 2008, we conducted a field experiment to test the new LeafView UMPC user interface, gather additional sample species images for testing the efficacy of the segmentation and matching algorithms, and evaluate improvements to the vision-based matching algorithms. Two botanists, each with their own LeafView UMPC device, set out for a day's worth of collection in Wind Forest, a site located in Maryland (Figure 4.14). The sun was extremely bright, the temperature broke 100 degrees, and the humidity was similar to that found in more tropical locations. The task was to collect and identify as many species as possible using the LeafView UMPC systems. Three computer science researchers accompanied the participants. Botanists carried spare batteries for the devices and additional equipment including blank pieces of paper and cutting shears. The group entered the forest around 9am and completed the field study around 3pm because of the extreme heat. After the trip, we recorded observations and gathered feedback from the two botanists about LeafView UMPC. In the next section, we discuss themes that arose in both observations and feedback.

4.10 Results and Discussion

Improving situated visualization through proximity. In moving from our initial prototype to LeafView UMPC, we used techniques of visual proximity, as discussed in Chapter 2, in two ways. First, we bring live video into the interface to increase the sense that the information about the leaf is closely associated with the leaf. Second, we make sure that the leaf is always present while browsing the results to aid in the comparison and matching task. With screen real estate at a premium, one might question whether it is better to compare a leaf on the screen with an image of the sample on the screen or the physical image of the sample. We observed that shape comparison is preferred when both images are on the screen, but, because of the resolution of the sample image, venation comparison was preferred between the physical leaf and the matching results on the screen.

Identification is an interactive process that involves user input beyond the initial photograph. We emphasize this to contrast it with a model in which identification is a black box that takes an image as input and produces an identification as output. Our images are acquired in uncontrolled environments, and may be corrupted by shadows or poor lighting, so feedback is necessary. The correct identification may not be the first result, so an ordered collection of results is displayed. The current identification may need to be compared with previous identifications. It is even possible that the species may be new to science. All these issues necessitate additional UI support.

We frame our video feed with UI elements. We have found that displaying history and contextual information around the frame of the live video provides relevant interaction such as history browsing, maintaining useful proximity to the visual task at hand.

Light weight and small size are more important than large screen and processing power for this task. However, sufficient screen space for providing visual feedback is still important. Based on initial feedback from our botanist colleagues, we moved from a Tablet PC platform to a UMPC platform, because they need to carry the device for long periods. This might not be the case in a different context, such as a hospital, where Tablet PCs are more common.

A finger is not a stylus. The UI elements we originally designed for a pen on a larger display were inappropriate for touch interaction on a small screen. We enlarged button targets, increased timing and tolerance for tapping, and took into account frequency of use in our earlier prototypes when determining button position and size.

Browsing modes are user dependent. After six months of use, the botanists using our system asked us to simplify the dataset browsing screen in our earlier UI. In our design sessions, they indicated they preferred a simple text list when browsing the entire dataset because they were often looking for a specific name rather than a shape and a screen full of images was overwhelming. This result is in contrast to our initial hypothesis that using a quantum tree map would be better for browsing a large dataset. However, we note that this preference might be different for users who are less familiar with species names and need to browse the data set based on shape.

4.11 Demonstrations

In addition to our field studies, the LeafView Tablet PC and LeafView UMPC prototype were demonstrated at several additional venues. These demonstrations provided an opportunity for additional feedback.

4.11.1 Smithsonian Science Exhibition Event

We provided two LeafView prototypes for attendees to try in an exhibit at the 2006 Smithsonian Institution staff picnic (Figure 4.15). Buckets containing a large variety of plant samples collected on Plummers Island were positioned around the exhibit. Attendees could pick a leaf, take a photo of it that was automatically submitted for identification, and explore the user interface. While this was quite different from actual field experience, it gave us an opportunity to gather feedback from a wide variety of potential users, both professional and non-professional.

Figure 4.15 Visitors learning how to use LeafView at the Science Exhibit for the 2006 Smithsonian Staff Picnic.

4.11.2 National Geographic BioBlitz and Smithsonian Congressional Night

The system was also shown to members of the public at the National Geographic BioBlitz (Figure 4.16) in Rock Creek Park in May 2007 (http://www.nationalgeographic.com/field/projects/bioblitz.html, to members of the U.S. Congress at the 2008 Smithsonian Congressional Night (Figure 4.16) and to researchers in human-computer interaction in demonstrations at ACM UIST 2006 (Symposium on User Interface Software and Technology) and IEEE and ACM ISMAR 2007 (International Symposium on Mixed and Augmented Reality).

Figure 4.16 (left) Students trying out the LeafView system at the National Geographic Rock Creek Park Bioblitz and (right) demonstrations at the 2008 Smithsonian Congressional Night.

4.12 Summary

In this chapter, we first introduce an application domain by describing results from our field study of botanical species identification. From this field study, we develop and present a task analysis, which is used to drive our architecture, user interface techniques, and prototype design. We then describe a system architecture for the Electronic Field Guide that integrates aspects of object matching, context-aware computing, and visualization. This architecture is easily extensible to add new forms of context, matching algorithms, data sets, and visualization views. Building on top of this architecture, we develop a set of techniques to improve botanical species identification embodied in two different prototypes. These techniques include improving comparison and inspection using semantic zooming and reduced spatial proximity of objects for comparison, visual feedback of segmentation from the system for improving image acquisition in real lighting conditions, queuing and batch identification, and inserting identification in the collection process. We develop and validate these techniques through iterative design and two field experiments with botanists, where the prototypes were used for botanical species identification and collection.

This work serves several purposes. First, we believe these techniques, supported by our architecture and application, are applicable to other mobile object identification domains. We have been approached by scientists interested in moths, fish, and insects about applying the system to their domain. While different vision algorithms will be needed, the user interface techniques and infrastructure should support these new domains without change.

Second, the architecture and application will be able to support further experiments by other researchers interested in experimenting with visualization of matching results.

Third, from this work, we found that context and spatial proximity of visual representations play an important role in creating meaning and associations in visualization.

From this observation, we were motivated to investigate techniques where the visualization is moved from the screen to appear in the physical world, using augmented reality techniques. We discuss this research in the next chapter.

5 Improving Matching and Inspection in Mobile Augmented Reality

Figure 5.1 Tangible Augmented Reality Electronic Field Guide

In contrast to LeafView, which presents visualization on an opaque surface, we designed, implemented, and evaluated two AR user interfaces for the Electronic Field Guide. While similar to LeafView in that the data was semantically-driven, here situated visualizations are displayed directly in the physical world. In this chapter, we first present the motivation for these user interfaces. Based on observations of botanists' use of their hands in the field study in the previous chapter, we specifically explore two extremes in user interface: hands-free interaction and tangible (hands-focused) interaction. We then describe these user interfaces in detail, along with their specific techniques for manipulating data. Next we discuss the implementation of each system. Both systems were evaluated in a laboratory study with botanists at the Smithsonian Institution, who used them for three inspection and comparison tasks. In the experiment we used a speak-aloud protocol, and the experiment was followed by interviews and group discussion. We present results from the evaluation and conclude the chapter with a summary of the work.

In this chapter, we emphasize the virtual vouchers (discussed in Chapter 4) as a situated visualization in which the leaf is the context and the visualization is semantically related to the object of interest. We explore object-referenced and body-referenced presentation in these prototypes. In the context of the model-view-controller paradigm [Krasner 1988], each of our prototypes can be considered a different exploration of a control-view combination using the same model of virtual vouchers [White 2006a, White 2006b]. The first prototype uses a see-through, head-worn display to view information in context and provides a tangible user interface for the manipulation of search results. The second prototype also uses a see-through, head-worn display and provides a hands-free user interface that presents search results referenced to the body. It also enables control of the results through head movement and a pair of unobtrusive buttons worn on the body.

In considering different user interfaces, we wanted to explore how quickly and easily a user could see results, select a species for inspection, and inspect the species sample or examine necessary contextual information about the species sample. To that end, a key component of our prototypes is the ability to explore many levels of detail in the virtual voucher, beyond what one might even expect from the physical leaf. Our prototypes build on the tradition of semantic zooming, discussed in Section 3.4.2, by considering the set of virtual vouchers as spatial data in which each individual virtual voucher can be explored in both level of detail and semantic zooming.

5.1 Tangible AR User Interface

Motivated by the way botanists in the field manipulate samples of species, we first prototyped a tangible AR user interface [Ishii 1997, Kato 1999] that provides a physical representation for inspecting visual search results and individual virtual vouchers. As described in Section 3.3.2, tangible user interfaces support interaction with information through the physical environment. Tangible augmented reality extends this concept by incorporating augmented reality techniques such as registered overlay of computer graphics. The visual representation of the virtual voucher is displayed in context and changes based on spatial modalities or gestures performed by moving a tangible marker associated with the virtual voucher. The tangible marker, sometimes referred to simply as a marker or fiducial, is a physical rectangle, made of rigid material such as cardboard or plastic. A pattern on the surface makes it easily identifiable using computer vision techniques for obtaining position and orientation of the marker. An example of a tangible marker is held in the right hand of the person in Figure 5.2a.

The use of spatial modalities to change the user interface has been investigated previously in a variety of ways. Bier and colleagues introduced Magic Lens filters as a class of see-through tools that can be moved around a 2D plane and visually filter elements beneath the tool [Bier 1993, Bier 1994]. Rekimoto developed a variant for AR that uses a magnifying glass as the metaphor [Rekimoto 1995]. A more detailed discussion if found in Section 3.3.1.

We build on this work by inverting the magic lens user interface concept to take advantage of the existing spatial intuition that botanists use for inspecting physical leaf samples. Instead of changing what the user sees through a lens based on the position of the lens, we change the semantics, modality, or magnification of the object based on the 3D spatial location of the object.

In our prototype, called Tangible Augmented Reality EFG (TAR-EFG), a leaf is placed on a clipboard to provide a consistent background for the segmentation algorithm. The clipboard is tracked using fiducial markers attached to the surface. A card containing a visual fiducial (not specific to the particular leaf) is then placed below the leaf to trigger image acquisition and initiate the visual search (Figure 5.2a-b). The results of the search are displayed along the side of the clipboard next to the original leaf sample in ranked relevancy order. The card can then be placed in the same location as one of the search results images so that the card morphs into that image and can be manipulated to inspect the virtual voucher (Figure 5.2c). While this is similar to the picking mechanism described by Kato and colleagues [Kato 2000], in which a physical paddle was used to pick or scoop up virtual objects, our intention is to provide a conceptual model in which the card transforms into a virtual voucher, in contrast to the concept of simply picking and moving a specific virtual object.

(a)

(b)

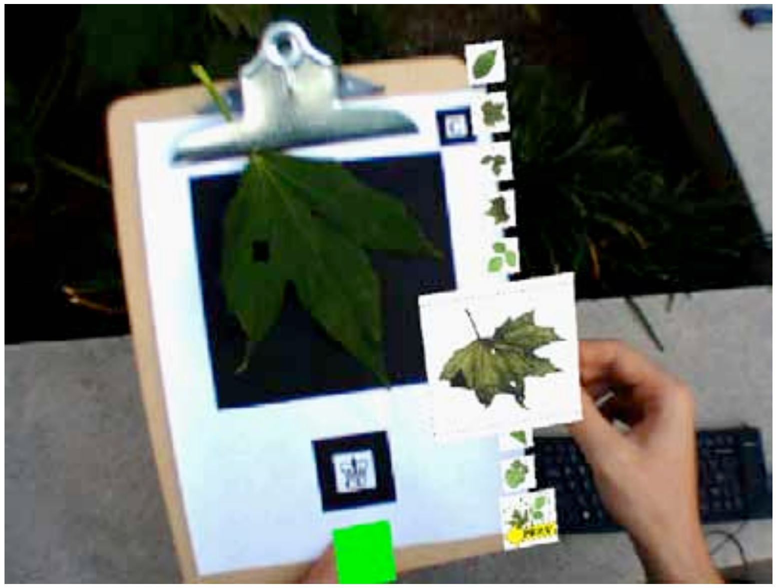

(c)

Figure 5.2 (a) Third-person view of user and fiducials. (b) Initiating a search. (c) View through a video see-through display of results and a virtual voucher in hand.

For the inspection task, the user can magnify the leaf and inspect venation or edge details by moving the card towards the user. The leaf image is magnified disproportionately relative to the actual distance traveled, as if the object was growing in size as it moves towards the user. Semantic changes are based on distance from the user, spatial zones, or orientation of the card. For instance, if a card is held towards the user's left, the image of the full plant is shown, and if the card is held towards the right, the image of the sample leaf is shown. An alternative interaction changes the modality by making gestures with the tangible marker. We discuss these two techniques, spatial morphing and tangible gestures, in the next two subsections.

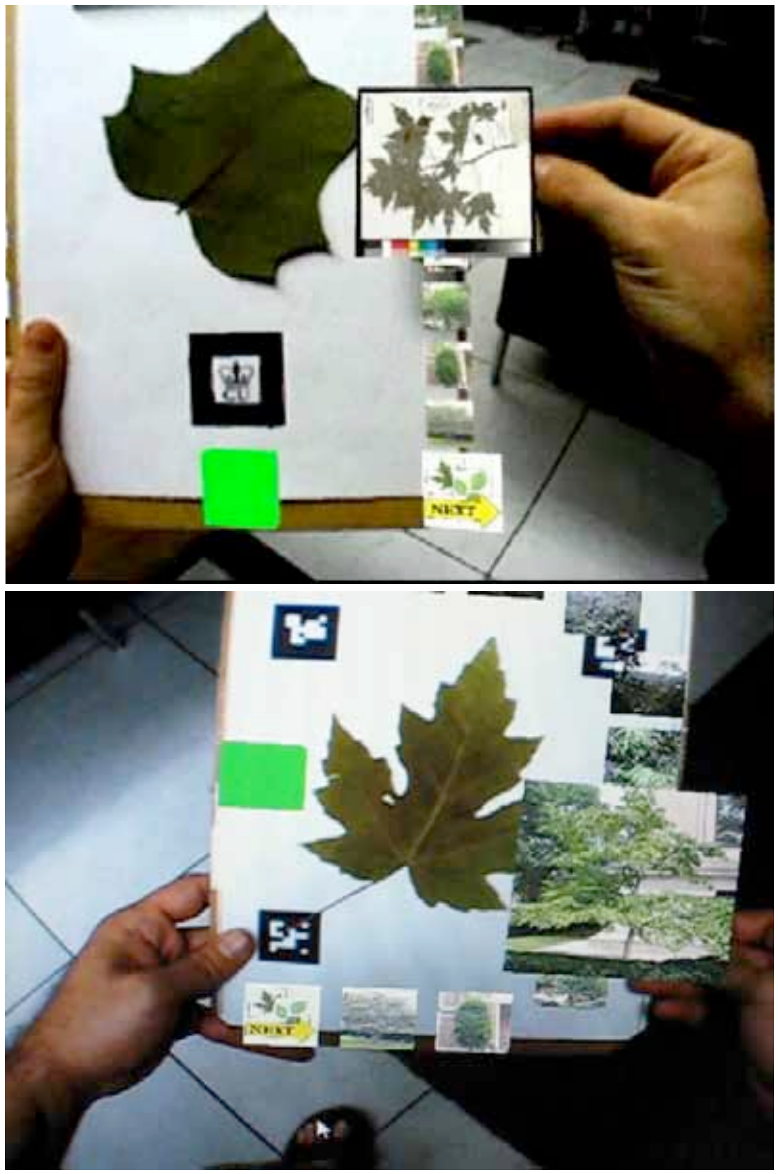
5.1.1 Spatial Morphing

We have considered a number of different spatial mappings for combining level-of-detail and semantic morphing. One approach uses continuous subspaces of magnification or level-of-detail within contiguous zones of modality, semantics similar to spatial modalities found in the n–Vision system [Beshers 1990]. For example, when the virtual voucher is held close to the user, they can magnify and examine a single leaf by pulling it closer or moving it farther away; however, when the virtual voucher is held farthest away, it morphs into the full plant, which also changes in size based on distance. Another example creates zones in the quadrants of space in front of the user. Distance changes level of detail, and the specific zone changes semantics.

In early, informal trials by our botanist colleagues, we found that when the mapping of zones is user-centric, it can be confusing to map too much semantic information along the outward z-axis. Changing to another representation as the object moves away from the user does not provide enough information if the size continues to decrease and the representation is too small to see. The projected view of a 3D object that itself remains constant size in the 3D world coordinate system does not provide enough of a size change, so we exaggerate the increase in scale as the object gets closer. Tracking is lost when the tangible marker is brought close enough to the user that a portion of the fiducial is outside the viewing area. We adjust the magnification so that the leaf image is twice the size of the fiducial when the fiducial fills the entire viewing area.

Once the botanist has decided on the identification, the botanist places the selected virtual voucher below the actual leaf, triggering a match. The new sample is recorded along with contextual data about the sample.

5.1.2 Tangible Gestures



(a)

(b)

Figure 5.3 (a) Virtual voucher representations can be changed to individual voucher specimen images or (b) images of the full plant.

In contrast to moving the virtual voucher into specific spatial zones or quadrants to change the type of information displayed from the voucher, we also explored tangible gestures. These gestures are made by moving the tangible object, in this case the marker fiducial, in specific ways that can be recognized by the system. We explored two types of gestures for semantic zooming: reeling and flipping.

The *reeling* gesture is similar to the gesture one would make when reeling a fishing pole or turning a crank on a jack-in-the-box. The virtual voucher is held in the dominant hand and the object is moved in a circular motion. Each cycle of reeling changes the image of the virtual voucher. For instance, in our prototype the image changes through a series of images that include individual leaves, a voucher specimen, the entire tree, or plant, fruits, bark, and a microscopic view of the species.

The *flipping* gesture is a rotation of the marker along the central horizontal axis of the marker, similar to twirling a coin between fingers. Each flip changes the representation of the virtual voucher in the same way that reeling changes the image.

Figure 5.4 Matching species in the HMCAR, viewed through an optical see-through display.

5.2 Head-Movement–Controlled AR User Interface

Our second design, Head-Movement–Controlled AR (HMCAR), explores providing a hands-free user interface to the specimen collection that can be used while the user's hands are otherwise occupied. Instead of tangible objects, we use head movement to control inspection of the virtual vouchers.

There are many ways in which head movement can be mapped to control of a user interface. Chung's comparison of head-tracked steering modes [Chung 1992] found that "orbital" mode, in which an object rotated in place as the subject's head rotated, produced the highest scores in his rotation task. He attributed this to maintaining the object in the center of vision and muscle memory. Koller and colleagues built on this work to examine additional issues and applications of orbital viewing [Koller 1996]. Hix and colleagues explored concurrently panning and zooming a flat screen display using head position. Moving the head closer to the screen would zoom, while moving the head from side-to-side would pan [Hix 1995]. Fuhrmann and colleagues carried this beyond orbital viewing to develop head-directed navigation in which the pitch moves the subject forward or backward in a virtual world while steering is done through head rotation [Fuhrmann 1998b]. Schmandt [Schmandt 1998] and Brewster and colleagues [Brewster 2003] used head rotation to control channels of audio. We build on this existing work on head-movement control by exploring our own mapping for moving objects and adding the notion of look-and-lock, described later in this section. Mine [Mine 1997a] explored Look-at Menus using gaze for selection, where 2D menus float in space, world-referenced, and are selected based on the head-direction of the user. In contrast, we are changing the object displayed in focus and shifting the visualization of results based on head movement, rather than enabling selection of different menu items. That is, the object is presented head-referenced and moves with head movement rather than just changing the view on a world-referenced object.

In this prototype, we developed two variations for head movement control. In the first variation, which we refer to as *free-moving*, the results of a visual search query are displayed in a row perpendicular to the head direction. They are floating in space and centered in the user's field of view. Rotating the head to the right continuously moves the row of virtual vouchers to the left while maintaining the row perpendicular to the ray extending from the center of the user's vision, as illustrated in Figure 5.5. Rotating the head to the left has the opposite effect. In this way, head rotation is used to quickly move between images in the results. The image that is currently in the center of the user's view is highlighted.

The second variation, which we call *rotating*, is similar, but instead of sliding left or right, the virtual vouchers rotate position. This means the set is always centered and the ordering always stays the same, but the voucher in focus changes with rotation. The second technique is similar to Chung's orbital mode [Chung 1992] and was developed as an alternative to explicitly centering the vouchers. However, we do not rotate a 3D object but rather change the ordering of visual elements.

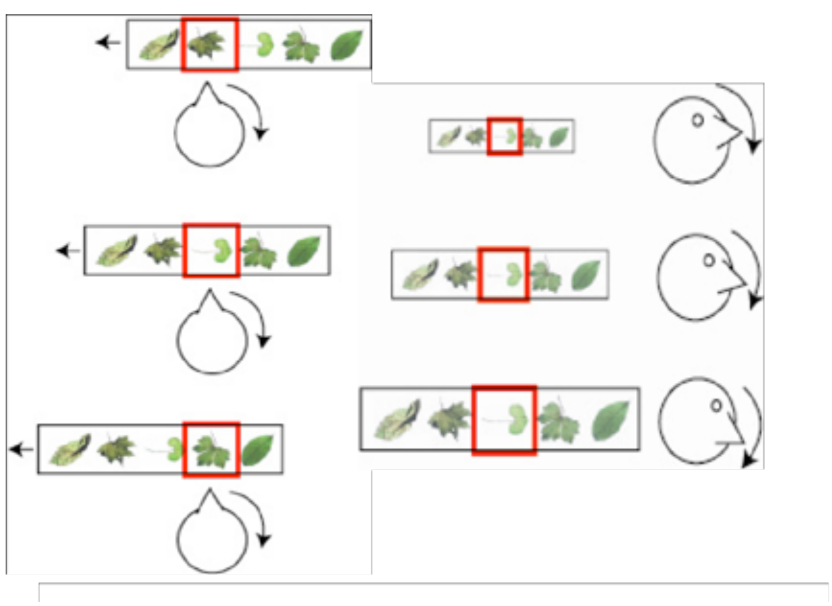


Figure 5.5 (a) Movement of virtual vouchers as the head rotates right. (b) Scale change as the head angles down.

In both variations, magnification is controlled by head pitch (Figure 5.6). Looking down increases the object size while looking up reduces to the overview of all results. This head pitch mechanism is similar to Bell and colleagues' use of head-pitch–controlled scaling for World in Miniature user interfaces [Bell 2002]. We have experimented informally with multiple modalities, but have found thus far that the best mapping technique is to magnify the object and increase detail when angled down, and to provide an overview of all results when looking straight ahead or slightly up.

Figure 5.6 Magnification in the Head-Movement–Controlled AR technique with images enlarged for inspection (left) or reduced for overview (right).

5.2.1 Look-and-Lock

If the user changes head orientation drastically, the visualization can move outside the current view. To address this, visualization can be re-centered on the current head orientation, and subsequent changes are relative to that orientation. The user can also lock the position of search results at any given time, so that the view does not change based on head movement. The lock is currently implemented by pressing a wireless button worn on the body or hand. For example, a user can select a particular leaf with head rotation and then magnify the venation by angling the head down slightly. They can then lock the image in place, so they can move their head without changing the magnified view of the leaf. Figure 5.4 shows a locked voucher viewed through a video see-through display. When the visual results are unlocked, the image is once again controlled by head movement. Pierce and colleagues allowed users to hold objects relative to their head position while using a 3D desktop applications to drop objects in a toolspace [Pierce 1999]. In contrast, we lock the entire view to allow free head movement relative to the world without changing the displayed objects.

We call this clutching mechanism look-and-lock, and we have found the technique useful for quickly finding a point of inspection and then comparing that image with the physical species under consideration. At all times, the visual results are displayed in an egocentric manner, such that a mobile user will always have the results ready at hand.

Once the species has been correctly identified, the lock is held down, triggering a match, and the new sample is recorded along with contextual data about the sample.

5.2.2 Semantic Zooming

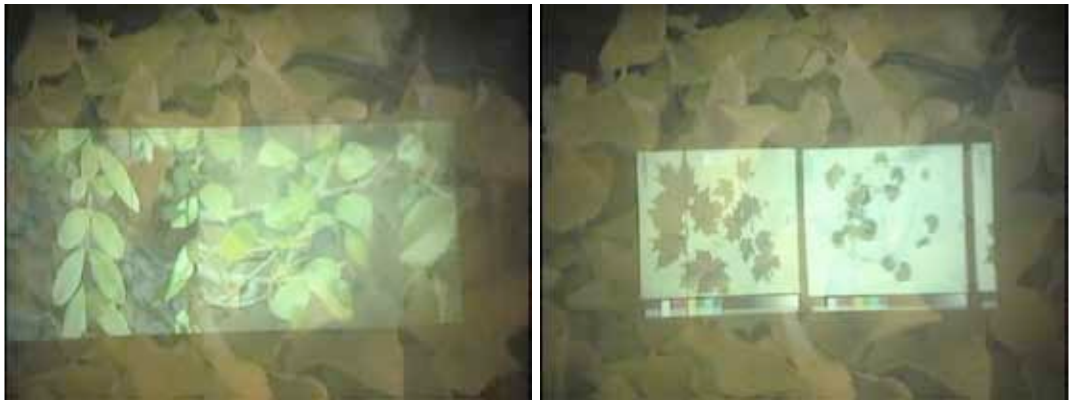


Figure 5.7 Alternative representations, trees (left) and voucher images (right) in the Head-Movement–Controlled AR technique.

A second button was used to shift through different representations in the virtual voucher. Initial matches were visualized using matched leaf images. Pressing the second button changed the virtual voucher to represent tree images, voucher images, and images of fruits and bark (Figure 5.7).

5.3 Hardware and Software Platform

Both systems runs on a Sony U750 hand-held computer under Windows XP, connected to a Liteye-500 800×600 resolution, color, see-through, head-worn display, and mounted on a baseball cap. Our system is intended for use in outdoor environments that require a clear view of the walking path; therefore, we wanted to use an optical see-through display (to transmit the real world at full resolution) with high transparency and brightness and minimal obstruction of the user's view below the display. Although we prefer to use a stereo display, we started with the monocular Liteye-500, because it is significantly brighter and more transparent than the other displays we had available. For comparison, we also used a stereo Sony LDI-D100B 800×600 resolution, color, see-through, head-worn display. The tangible AR user interface uses a clipboard and small cardboard cards with printed fiducials that are tracked using a Creative Labs video camera, also mounted on the baseball cap. We have iterated through multiple versions of the system. In the first version, we use ARToolkit [Kato 1999] to identify and track the fiducials and OpenGL for image display. In the second version, we use ARTag [Fiala 2005] and Goblin XNA [Oda 2008]. The head-movement controlled AR user interface uses an InterSense InertiaCube 2 hybrid inertial orientation tracker to track head orientation along with a Keyspan presentation remote wireless two-button input device that is worn on the body to initiate lock, re-center, and select.

5.4 Evaluation

Although we have tested the functionality of the system outdoors, we required a controlled environment that would allow us to focus on the differences between the prototypes. To evaluate and compare the two AR user interfaces, we conducted a laboratory experiment. The prototypes were tried indoors by four of our colleagues in the Department of Botany at the Smithsonian Institution (two males and two females between the ages of 28 and 55) with both specific and open-ended tasks. The purpose of these trials, together with observations and structured interviews, was to elicit initial feedback on the conceptual model, process, user interface, and hardware configurations of the system. Three of the botanists were comfortable with computer technology, while one had little computer experience. None of the participants had experience with head-worn displays.

5.4.1 Experimental Setup

Figure 5.8 Experimental setup for user study.

For our lab experiment, we had participants use both a monocular optical see-through display (Figure 5.8) and a stereo, video see-through, head-worn display. The full hardware set-up was described in the previous section.

5.4.2 Task and Procedure

We are interested in the ease of use, speed, and efficacy of inspecting an individual virtual voucher and comparison of the virtual voucher with a physical leaf. In addition, we wanted to understand the participants' perceptions of the conceptual model, process, user interface, and hardware configurations.

Prior to the trial, participants were given an opportunity to become familiar with both interface techniques. Once they were comfortable, each participant matched six leaves using each of the techniques. During the task, subjects were encouraged to speak aloud about their experiences, both positive and negative. After matching all leaves, a structured interview was conducted to compare the ease of use and intuitiveness of the user interface techniques and to learn more about specific aspects of the techniques. Two researchers were present—one managing the trial and the other recording notes.

5.5 Results

The botanists represented a wide range of expertise with computers and technology. Each participant used both prototype user interfaces. Here we discuss the results from observations during the trials and structured interviews after the trials.

5.5.1 Conceptual Model

The EFG and virtual voucher conceptual models fit well with a botanist's existing task model and practice. Since botanists already use a paper field guide as a tool for identification, the concept of an EFG that aids in the same task made sense. The virtual voucher extended their existing concept of a specimen voucher. They were aware that the virtual voucher represented a depth of information beyond the image that was on the screen at any given point in time. The model also provided a close association with the purpose of a physical voucher specimen in that it was both a tool for identification and proof that the specimen had been collected. A number of the botanists requested that specific information be displayed to aid in identification. Collectively, they felt that the characteristics that would be most useful for identification were images (leaves, leaves on branches, full tree or plant, trunk, fruit, and bark) as well as locality, region or distribution maps, genus, species, common name, and text description. Although we had included it in the prototype, the botanists did not see a taxonomic hierarchy as being useful to them in the field.

5.5.2 Process

We also discussed the process of identification and collection of specimens. One of the more interesting observations was that identification is as much a process of elimination as it is one of focus. This came up with regard to the tangible AR user interface, when one of our colleagues said he wanted to remove some of the virtual vouchers from the display because he had already eliminated them from the set of possibilities. This decision was based on visual traits, locale, and seasonal traits. For example, a plant that flowers in late August would be excluded when identifying a plant that is flowering in early April. One botanist referred to this as "more a process of elimination than just choosing a winner."

5.5.3 Reactions to Tangible AR

The botanists liked the tangibility of this user interface and the sense that the virtual vouchers were connected to physical objects. The simple action of flipping the card to see a new aspect of the virtual voucher such as the full plant image was also appealing. We observed them making up their own physical language for manipulating the vouchers. One of our colleagues wanted to tap a voucher to make it disappear if he had eliminated it from the possible matches. There were also points when the user interface seemed to confuse them. They all learned to bring the voucher closer to change magnification, but when some of them wanted to change the size of the overview images on the clipboard, none of them realized they could lift the clipboard to bring the set of images closer. After they learned that they could do this, they used that to take a closer look at overview images.

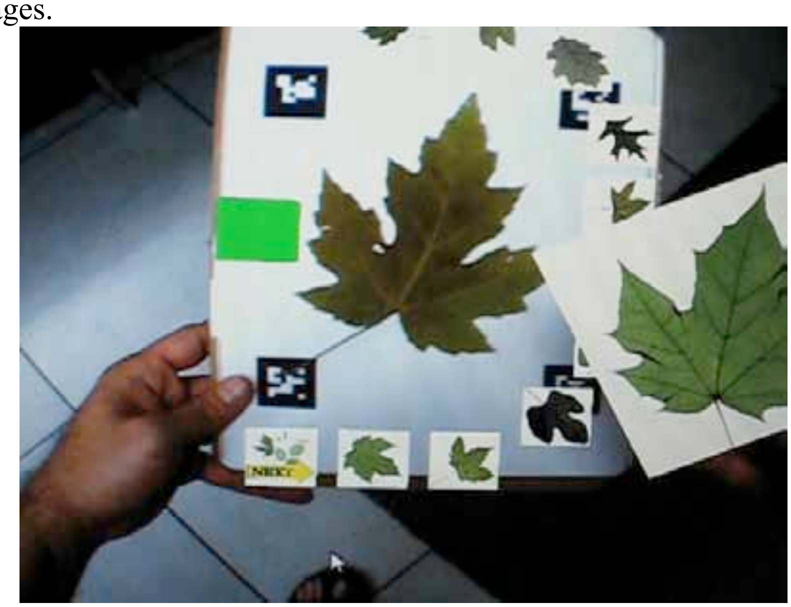


Figure 5.9 Changing the layout of virtual vouchers to surround the sample leaf in an arc.

Layout was also discussed during use of the system. One colleague asked that the results be placed in an arc around the leaf rather than just along the side of the leaf to centralize the comparison focus. We experimented with this layout as well (Figure 5.9) and found that showing results in an arc increases proximity for comparison, but makes it more difficult to select individual virtual vouchers for inspection. This difficulty is due to both occlusion by the arm and the need to reach across the clipboard to select. In contrast, the linear layout along the right side of the clipboard removes some of the close proximity for some leaves but is easier to select.

Another botanist found the ten results we presented overwhelming and remarked that this was a similar problem he had with the large number of results presented in the earlier 2D layout prototype they had tried previously. This botanist would prefer seeing only five results at a time. One botanist was concerned that the fiducial cards would get lost because she "loses things in the field all the time." Another botanist asked if the fiducial could be replaced by a pen, so he could use the tip of his pencil or pen and put the leaf on his field notebook. This was brought up again as a concern that the user interface might keep their hands busy when they wanted them available for recording information.

5.5.4 Reactions to Head-Movement–Controlled AR

Our colleagues used the two versions of the Head-Movement–Controlled AR user interface introduced earlier: free-moving and rotating. Starting with the free-moving variations, all of the botanists were able to learn the head-movement controls very quickly after a minute or two of experimentation. They all remarked positively on the speed with which they could inspect different plants. The look-and-lock mechanism was seen as particularly useful for comparison of details. However, they wanted the locking mechanism either to be head-controlled or to have the button mounted on the head-worn display. Two of the botanists also wanted to be able to lock into a leaf even as they moved around the leaf, so that they would not accidentally slip into another voucher, like picking a direction or “channel” on which to focus. We were surprised to find that they liked the centering mechanism, not for centering, but because it let them remove the user interface from view and then bring it back to center, no matter where they were looking. They did not like the rotating interface because it changed the spatial arrangement of vouchers. This interfered with their comparison process because vouchers were not always in the same location. The system was also found to be too sensitive to small movements when the vouchers were highly magnified.

In comparison to the hand-held interfaces, one botanist commented that he preferred the AR user interfaces because they “make it a part of you.”

5.5.5 Discussion

Based on these initial trials, we have learned a number of lessons that will inform our future work. First, we note that each of the different prototypes provided benefits and costs in usage. We want the benefits of physical affordances provided by tangible user interfaces without the extra requirements of a clipboard and fiducial cards. One way to address this will be to use the notebook and pencil that are already carried around by the botanists on field explorations. Similarly, head-movement control provides a useful means of controlling the interface when the hands are busy, but it is not always necessary and provides less precise control than the tangible interface. We will integrate these modalities together so the botanist has the option of manipulating virtual vouchers with modalities that match different modes of use.

The see-through displays that we used have the benefit of representing the information in context. They also provide the experience of directly manipulating virtual vouchers, which was positively viewed in our trials, but they are still bulky to wear. In our trials, the botanists asked about accessing the same functionality from either the head-worn or hand-held system, depending on the local environment. We believe we can address this and carry the same conceptual model across both head-worn and hand-held user interfaces.

For comparison tasks, we observed that drastically changing the spatial layout of results interfered with the comparison task, because objects of comparison were not positioned in their expected locations. We found this to be problematic in comparing the rotating versus centered head-movement–controlled user interfaces.

While our current application represents visual search results of botanical samples, we believe the techniques for inspection and comparison of visual search results may be able to be generalized to other visual search results in which objects and their many characteristics need to be inspected. In particular, we believe the use of physical manipulation coupled with spatial modalities could provide a quick and easy way to browse through a wide range of visual information.

We also believe the look-and-lock mechanism provides a useful means of supporting hands-free manipulation of objects. Based on our observations thus far, the technique is intuitive to learn and supports a combination of head movement and head-movement control.

5.6 Demonstrations

Figure 5.10 Demonstrations to members of the U.S. Congress, park rangers, and children.

Beyond our evaluation in the lab, the tangible AR system was demonstrated at the National Geographic Rock Creek BioBlitz 2007 and to members of the U.S. Congress, National Park rangers, and children at several events (Figure 5.10). In all cases, the TAR-EFG was shown together with the LeafView system. While these demonstrations were informal, they reinforced our earlier observations that direct manipulation of virtual vouchers provides an easy to learn interaction technique that is faster than the LeafView system for inspection and comparison tasks. Children in particular were able to learn the system with minimal training.

5.7 Summary

In this chapter, we have presented two novel interaction techniques for applying situated visualization to matching and inspection tasks for botanical species identification. In the first technique, virtual vouchers are visualized in the same plane, in close proximity to the sample leaf. We use tangible AR to provide direct physical interaction with virtual vouchers. Spatial zones and tangible gestures are used to change the semantic zoom, while direct manipulation of the virtual voucher supports magnification of the current representation of the virtual voucher. In the second technique, virtual vouchers are displayed in the space in front of the user and referenced to the user's head. Up and down head movement changes the magnification of the visualization, while side-to-side head movement changes the focus of the visualization and selects the current virtual voucher. The visualization can be locked for comparison, and semantic zooming is changed by pressing one of two buttons. We compared these techniques in an evaluation against each other and with the experiences that botanists already had with an early version of the LeafView user interface.

In our evaluation, we found that botanists appreciated the hands-free interaction of the head-movement–controlled augmented reality system but preferred the direct manipulation of the tangible augmented reality system. In addition, we found that inspection of individual leaves and comparison of virtual vouchers with the physical leaf specimen were faster using the tangible augmented reality system. We also observed that comparison and inspection with tangible augmented reality was faster than with LeafView and botanists preferred the tangible augmented reality interaction. However, the current state of head-worn display and tracking is still too bulky and finicky for practical use, while the LeafView system can be used in its current form.

Our contributions in this chapter include the spatial zone and tangible gesture techniques for magnification and semantic zooming, head-movement control techniques for magnification and semantic zooming, and a comparison of these techniques in the context of prior use of the LeafView system.

In situated visualization terms, we have investigated visualizations that use objects as context and are semantically-driven. In both techniques, the display is an immersive, see-through head-worn display. In the case of the tangible AR, the visualization is object-referenced. The head-movement–controlled AR uses a body- or head-referenced coordinate system in which the position and orientation of the visualization remain in the same location while the focus, magnification, and semantic zoom are changed based on head movement. Feedback from evaluation participants suggests that presenting the virtual visualization in close proximity to the physical leaf improves the speed and efficacy of inspection and comparison. We also note that the semantic relationship between the object of interest and the visualization provides significant freedom in the spatial layout of the visualization.

In the next chapter, we build on the tangible AR infrastructure and address situated visualizations where the object is still the context, but the relationship is spatially-driven.

6 Improving Learning and Discovery of Physical Actions

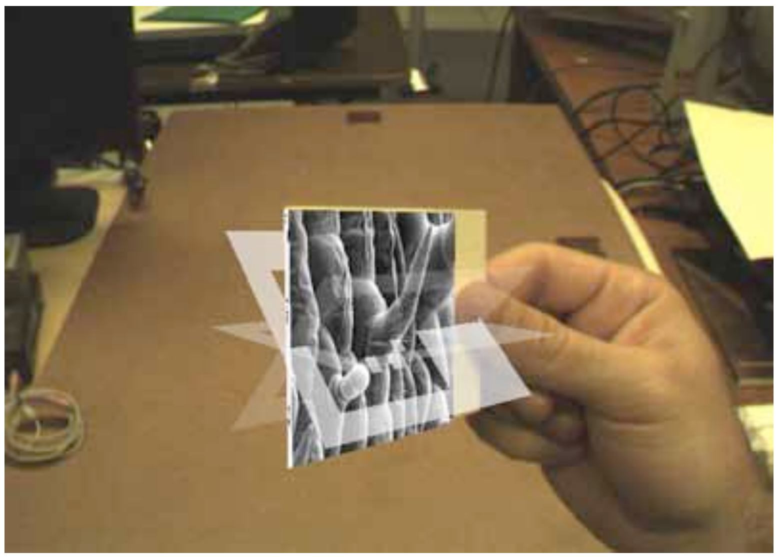


Figure 6.1 Example visual hints showing a twirling gesture (live capture of view through video see-through display).

In contrast to the previous chapter, which investigated situated visualizations with semantic relevance to the object, this chapter investigates presentation of situated visualizations with spatial relevance to the object. Tangible AR, which combines physical input devices with overlaid imagery, allows physical objects and the actions that can be performed on them to be overloaded with new meaning [Kato 2000]. However, the interactions possible in such systems are not always obvious. While menus in a windowing system reveal the set of supported actions, interactive UI documentation that addresses the richer domain of tangible AR has not been well explored. Tangible objects can reveal affordances [Gibson 1986] based on their physical morphology, yet more complex manipulations and gestures are not always readily apparent. Much like manual gestural UIs, the ephemeral nature of tangible gestures makes them difficult to discover and properly learn [Kjeldsen 1996]. Here, we investigate *visual hints* [White 2007a] (Figure 6.1)—situated visualizations in AR of potential UI actions associated with the physical world. Visual hints can potentially improve discovery and learning of gestures and manipulation in tangible AR. Our contribution includes developing and formalizing visual hints as well as presenting results from our implementation, use, and comparison of different kinds of visual hints. In our tangible AR UI to an EFG for botanists [White 2006b], cards with fiducial markings can be manipulated to transform representations of plant species displayed on the card (virtual vouchers) through semantic zooming, magnification, and change of species. One particular gesture involves a twirling or flipping motion that was readily learned through in-person demonstration, but it was not apparent to uninitiated users of our original UI. In this example, the visual hint shows the user the motion needed to execute the gesture.

More generally, visual hints incorporate a variety of methods for representing actions that can be taken, ways to complete them, and their outcomes. Our implementation also supports activating/deactivating hints and cycling through them. Visual hints can be applied to aspects of the environment or specific tangible objects. Here, we investigate visual hints for gestures and manipulation in tangible AR.

6.1 Tangible Gestures

Tangible gestures involve the manipulation of tangible UIs. They provide a simple way to interact with information relevant to a particular object or abstraction. While the space of gesture and tangible UI taxonomies is broad, we focus on gestures that Quek et al. refer to as manipulative or semaphoric [Quek 2002]. *Manipulative* gestures tightly couple the target of manipulation and the gesture. *Semaphoric* gestures are symbolic and typically represent a stylized dictionary of static or dynamic gestures in a system.

Several questions arise in exploring such gestures. How does one discover the gestural affordances of an object or environment? How does one learn the correct movement of a gesture? As a gesture is performed, how does one know it is being completed correctly? Although physical affordances often represent these aspects of a system, they may not always be present or sufficient to reflect the expanded capabilities imbued by AR.

6.2 Representation

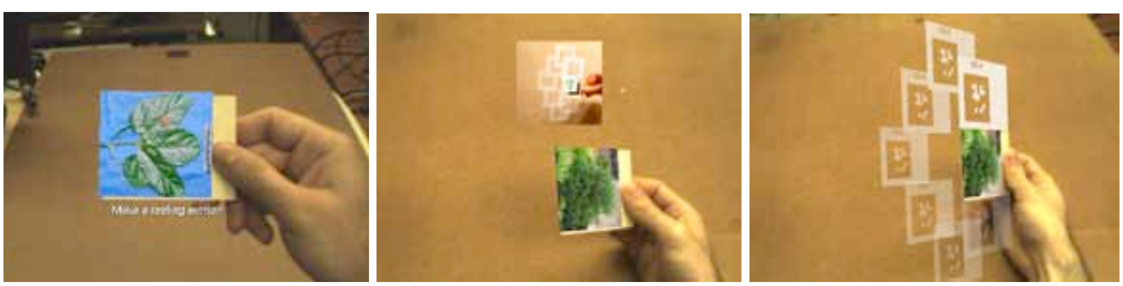


Figure 6.2 A visual hint for a circular "reeling" gesture, represented through (a) text, (b) diagram, and (c) ghosting. (Viewed through a video see-through display.)

There are many ways to represent a visual hint. For example, a textual explanation, a static diagram, or an animated extension of a tangible object can all provide information about the gesture (Figure 6.2). These representations fall into a design space that can be characterized along multiple dimensions, including media type, dynamics, locus of presentation, and proximity. Locus of presentation and proximity are particularly important. Hints can be presented object-referenced or they can be presented display-, body-, or world-referenced, regardless of the position of the object. This is distinct from proximity, where a hint can be displayed close to or distant from an object. Close proximity, which has been shown to improve learning in multimedia [Mayer 2001], can also create the illusion that a hint is part of an object. We consider a variety of representations in our design space, some examples of which follow.

Textual hints (Figure 6.2a) can be read and perceived quickly, but depend strongly on shared meaning and understanding. For example, the English word "flip" means different things in different cultures, has no standard gestural representation, and would need to be translated into other languages.

Diagrammatic hints (Figure 6.2b) provide a spatial image of the appropriate gesture. Although there are some domain-dependent diagramming standards, images typically are designer-dependent. Their location and proximity can also vary. For example, a set of diagrams may be display-referenced to the top of the display or object-referenced and presented in close proximity to the object of interest.

Ghosted hints (Figure 6.2c) represent the action of the gesture in 3D space, starting from the current position of the object and traversing through a series of ghost images that follow the trajectory of the gesture. Ghosting is a well-known illustrative technique in comics [McCloud 1994] and in manual and automated [Feiner 1992] graphic design. In ghosting an object is rendered semitransparent in order to represent its past or future state or to allow other objects that it would obscure to be viewed through it.

Animated hints are similar to ghosted hints, but replace the ghosted image with an animated representation of the movement trajectory. Tversky et al. suggest that animation can improve learning under certain conditions [Tversky 2002].

Composite hints integrate multiple simpler hints. In a later section, we discuss the implications for showing a set of possible actions and results instead of a single possible action.

6.3 Activation and Deactivation

To avoid visual overload, we do not show visual hints all the time, and we limit those that are shown when visual hints are enabled. Thus, there must be ways to activate and deactivate hints (e.g., through an implicit or explicit user action) as well as ways to determine which relevant hints to display. Activation and deactivation can be accomplished using tangible AR interaction or through an additional modality, such as voice. We implemented a variety of activation methods, which we describe here.

Pausing or lack of motion. Inspired by the use of pausing to enable the display of marking menus [Kurtenbach 1991], this can act as either an implicit or explicit activation technique. Pausing works well when it does not normally occur as part of the action, as in skilled selection from a menu. However, it can be problematic when it is part of the action, as is the case in our system. Users often hold the virtual voucher still so that they can study the veins or edges of a leaf, and this could inadvertently trigger display of visual hints when it is not required. Combining pausing with a specific orientation and distance from the user can alleviate some inadvertent activation.

An activation gesture. An example of an activation gesture is shaking the virtual voucher in or out of plane [Kato 2000]. The shaking gesture suggests an attempt to discern the voucher's hidden properties, as in shaking a gift box to hear what is inside.

A key or button press. This conventional activation approach creates a more clearly defined visual hints mode. However, pressing a button outside the tangible AR space of interaction can bring the user away from the task at hand.

We considered but did not implement other methods, such as proximity of the hand and voice activation. Proximity of the hand to a tangible AR object can be used to activate visual hints, but our experience from mock-ups is that once the gesture has been shown, the hint will remain present even if no longer necessary.

Visual hints can also reflect completion of an action. For example, as a user makes the correct circular reeling motion following the path of ghosted images, each ghosted image can be removed.

6.4 Implementation

We implemented the methods described in the Sections 6.1-6.3 in our tangible AR EFG prototype described in Section 5.1. Our system uses a Sony LDI-D100B 800×600 resolution, color, head-worn display, on which we mounted a Point Grey FireFly MV camera to capture the scene for 6DOF fiducial tracking and biocular (non-stereo) video see-through AR. In the mobile system, the display and camera are connected to a Sony U750 hand-held computer running Windows XP, mounted on a fanny pack worn over the shoulder. For our stationary user study, the camera and display were connected to an Apple Macbook with 1GB RAM and a 2Ghz Intel Core Duo CPU. A clipboard and individual rigid paper cards mounted with fiducials provide tangible objects for interaction.

The software is developed in C++, with OpenGL graphics, ARTag [Fiala 2005] 6DOF optical fiducial tracking, and grammar-based gesture recognition. Individual movements such as left, right, up, down, forward, back, and rotation are continuously captured, with those movements below a calibrated threshold omitted to avoid jitter artifacts. Movements are combined in a history that is parsed and fit to state machines representing each of the gestures. If a gesture is matched within a threshold time limit, the gesture is accepted and added into a gesture history. The test system runs at 24 frames per second for both capture and rendering (gesture recognition is less accurate at lower frame rates).

6.5 User Study

Figure 6.3 Test setup for visual hints laboratory experiment.

To gain a better understanding of how to represent and activate visual hints, we ran a pilot user study. Seven participants—three male, four female, aged 19–34 years old—were recruited from our university population through e-mail lists and fliers. Each had prior familiarity with computers and was compensated $10.

The task was to understand, learn, and perform the correct gesture with a fiducial card when given specific stimuli for the gesture. An automated test suite was used to present seven different individual and composite visual hints to the subject for each of three gestures for a total of 21 combinations in a within-subject design. Order of presentation was randomized. The seven visual hint types were text (T), diagram (D), ghost (G), animation (A), ghost+animation (GA), ghost+text (GT), and ghost+text+animation (GTA). Diagrams were display-referenced. All other conditions were object-referenced. Text and diagrams always faced the user. The three gestures were reeling, twirling, and movement into a target area, each of which was used to cycle through 2D images of a plant species presented as if printed on the fiducial card.

Subjects sat at a table and were video recorded to capture both their gestures and the images they viewed. Prior to the study, the experimenter explained the task to the participants, and subjects were given a preliminary trial prior to the actual trial. Subjects wore the head-worn display and held a fiducial in their hand. After each visual hint was presented, the subject was asked to identify and perform the gesture and completion time was logged. At the end of the trial, subjects were introduced to shaking and pausing activation methods for visual hints. After completing the automated tests, the subject was asked to fill out a questionnaire to provide ranked preferences, Likert-scale responses, and qualitative comments on the visual hint and activation approaches.

6.5.1 Results and Discussion

One important observation about our within-subject pilot study is that there were significant learning effects. Once a user knows the gesture, they can apply that knowledge to other hints about the same gesture. This implies that time to completion analysis is only valid for the first instance of a gesture being observed and correctly completed. However, the qualitative results from observations of the subjects and questionnaire results are useful.

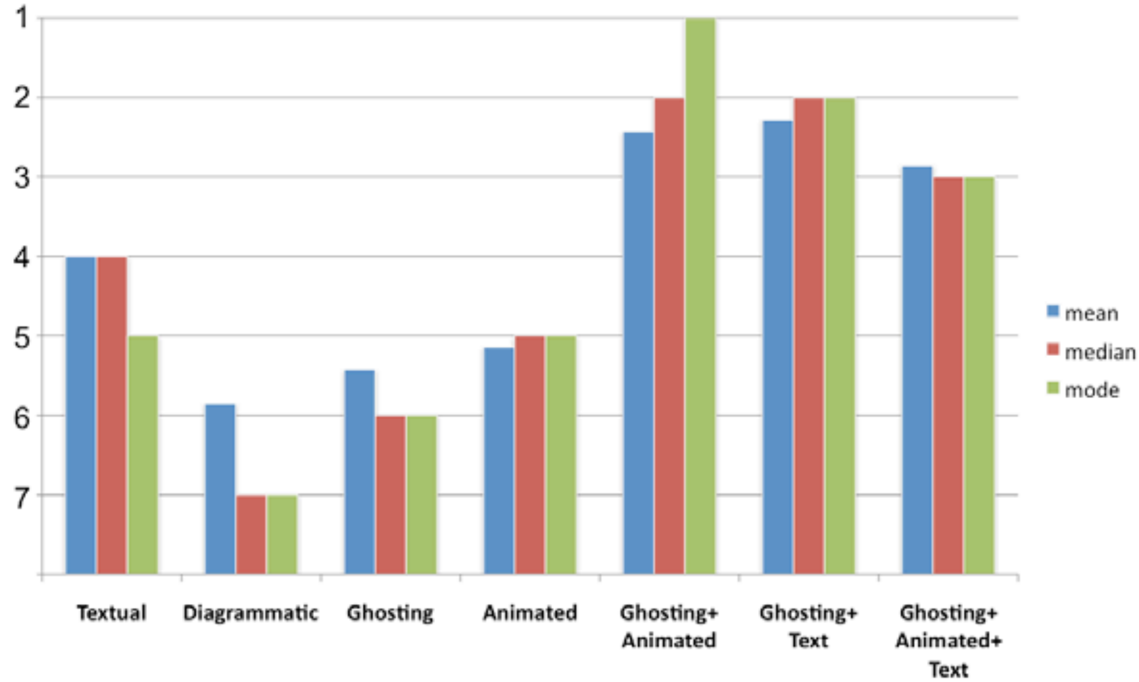


Figure 6.4 Ranked preference for each representation technique. 1 is best.

Subjects were asked to rank the seven techniques in order of preference (Figure 6.4) Looking at a chi-square for all seven categories, X2=16.8, df=6, p=0.01, showing that the distribution of scores between the different options is significant. GA and GT ranked the highest, followed by GTA. This is mirrored in results of ranked comprehension, also in (Figure 6.4). We noted a variety of issues during observation and from participant comments, which we describe below.

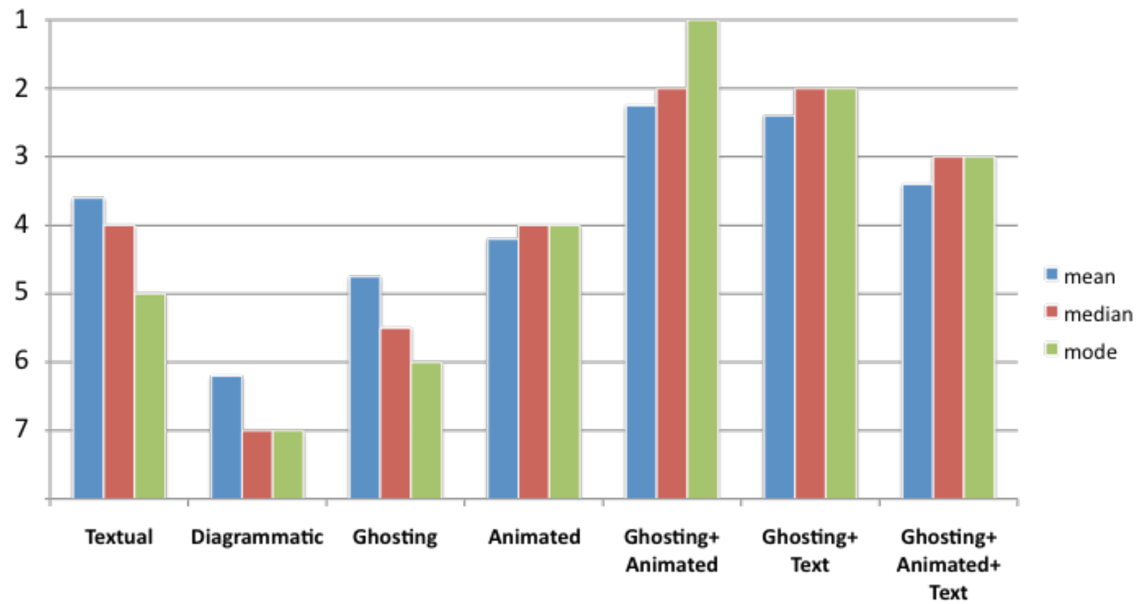


Figure 6.5 Ranked comprehension for each technique. 1 is best.

Text can indeed be ambiguous because of language or culture. As one subject wrote, "I don't know what a 'reeling' motion is." One participant also interpreted twirling as a single motion, moving back and forth, rather than a continuous motion. Even words such as "target" and "left" proved ambiguous. We also found that object-referenced text hints could drop below the edge of the screen when the fiducial was held towards the bottom of the viewable image (which could be fixed by adaptive layout). One subject commented that they would prefer text for known gestures and animation with ghosting for unknown gestures.

Diagrams were less culturally linked, but were misinterpreted by some subjects as showing 2D motion in a plane. This is most likely because of the nature of our diagrams, which were displayed on 2D surfaces. It is possible that different diagrams or stereo imagery would have produced better results.

Ghosting proved to be successful at illustrating the required movement. However, the subjects raised two important issues. First, the ghosted image is fixed in an object-referenced position and orientation, so as subjects tried to perform the gesture, the ghost moved with the fiducial, much like a dog chasing its tail. This could be resolved by keeping the ghost fixed after a test gesture is started or by providing completion feedback. A second issue was directionality. A series of images that vary in transparency can be interpreted in two opposing ways. In comics [McCloud 1994], motion from the past to the present is often shown as a series of images that transform from transparent to opaque, such as those of a fist moving through the air to a punch. The past is transparent, and the present is opaque; that is, time moves forward from the transparent to the opaque. However, when we represent motion from the present to the future, we can either emphasize the forward movement of time and interpolate from transparent to opaque or keep the present as opaque, interpolating opaque to transparent. This ambiguity can cause confusion, although the trajectory of the gesture is clear. (If directionality matters, an arrow can be added, creating a composite hint.)

Animation by itself resolves the issue of directionality, but could also be confusing because the subject is forced to keep the trajectory modeled in their mind. Furthermore, while animation is useful in clarifying the movement of a gesture over time, the speed could be taken either literally or as an abstraction of movement. One subject treated the animation speed literally, trying to match the rate of the animated visual hint, even though the gesture could potentially be made faster.

The combinations of ghosting with text or animation proved to be the most preferred. With animation, the subject saw both the directionality and the trajectory of the gesture; however, the additional movement could be distracting. With text, a potential benefit arose from reinforcing a particular name or label for a gesture.

In their responses to shaking versus pausing for activation, four participants preferred shaking, while three preferred pausing, with one commenting that "it's easier to shake than pause."

Note that the graphic representation used in our visual hints is an image of the fiducial, not the plant. In our experience, the fiducial was less visually complex in terms of texture and color, as compared to the plant, and thus less distracting. Use of the plant image also created confusion between the hint and the virtual object. However, use of other imagery is worth further study.

6.5.2 System Discussion

In implementing and using different types of visual hints and activation mechanisms, we made several observations. Our visual hints primarily represent simple gestures that require a single motion or trajectory to complete. Visual hints for complex or compound gestures might still run into limitations given the necessary reuse of visual space around a tangible object. In ghosting and animated visual hints, we found that the space around the fiducial could support displaying only a single gesture (because of the relatively large spatial extent of the hint and small field of view of our display). In contrast, diagrammatic and textual representations support simultaneous display of hints for multiple gestures by placing multiple diagrams or text strings adjacent to each other.

One might ask why not design the tangible object itself to have visually obvious affordances, so that the set of gestures and manipulations are clear and the physical morphology constrains the user to these gestures? While this is a noble goal, objects may change functionality depending on their context. In a sense, visual hints acknowledge the object's affordances, as defined by Gibson [Gibson 1986], making them more readily perceived. In other words, an object can be manipulated in many different ways, but only certain actions have meaning in the context of the AR system.

6.6 Summary

In this chapter, we have presented visual hint techniques for learning and discovery of potential gestures and manipulation in tangible AR. We motivated the research by discussing the need to learn potential actions in novel tangible and gestural user interfaces. We focus on tangible gestures and discuss them in the context of manipulative and semaphoric gestures. Visual Hints can be represented in many ways. Here, we have described textual, diagrammatic, ghosted, animated, and composite hints. Visual Hints only appear when the user determines that they are necessary and we investigate pausing and shaking techniques for activating and deactivating these visualizations. We presented a user study evaluating seven different types of visual hints and two types of activation/deactivation techniques. Results from our evaluation show that participants preferred composite hints that combine ghosting with animation or text. These composite hints also ranked highest in comprehension. This redundant encoding helps the user perceive the correct action for the gesture. However, composites that included three different types of representation were perceived as visually busy. Our contributions in this chapter include the visual hints techniques, visual representations, activation methods, and the evaluation of these techniques.

In situated visualization terms, we displayed visualizations of physical actions using an object as the context, in this case a fiducial marker. The relationship between the visualization and the context was spatial. Some visualizations, such as text and diagrams, were shown in close proximity to the object of interest while animation and ghosting were displayed in specific positions and orientations contiguous with the fiducial marker. The technique was tested using a head-worn video see-through display, although we believe the technique extends to hand-held AR displays as well. The loci of presentation and interaction were both object-referenced, reinforcing a sense of direct manipulation in the user interface.

In the next chapter, we investigate prop-based gestural activation and presentation techniques for 3D menu options and radial displays of visualization.

7 Visualizing and Interacting with Radial Displays of Information

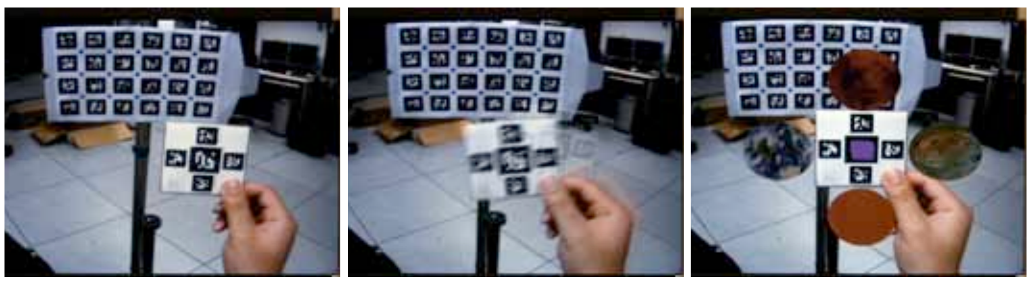


Figure 7.1 A shake menu being used to select and place planets in a 3D space. (a) User holds an object (in this case, an optically tracked fiducial marker) and (b) shakes the object to (c) display a pie menu of options arrayed around the object.

Shake menus (Figure 7.1) are a novel method for activating, displaying, and selecting menus presented relative to a tangible object or manipulator in a 3D user interface. They provide ready-to-hand interaction, including facile selection and placement of objects. In this chapter, we present the technique and a user study comparing the speed and efficacy of several alternative methods for presenting shake menus for activation and radial display (world-referenced, display-referenced, and object-referenced) along with a baseline technique similar to commonly used proximity models for prop-based menus. We also present qualitative feedback from use and several illustrative applications of the technique for interaction with visualizations and authoring.

Menus play an important role in 3D UIs, including AR and VR, for presenting information and system control. Bowman et al. [Bowman 2005] organize 3D interaction techniques by interaction task, separating out selection and manipulation, travel, wayfinding, system control, and symbolic input. Within the system control task, graphical menus are used as a familiar interaction technique, similar to menus in desktop UIs. Complementary to classification by task, Daschelt and Hübner [Dachselt 2007] suggest a taxonomy of 3D AR menu techniques that distinguishes between glove- or hand-based menu selection and physical prop-based menu selection. Across both taxonomies, a wide variety of menu-selection techniques have been developed.

We are particularly interested in the use of menus for information display, system control, and object interaction and authoring in AR. We focus here on prop-based menu selection and menu presentation centered on the prop for several reasons. First, there is evidence that providing a tangible anchor or prop increases the sense of presence [Hoffman 1998], enhances realism [Lindeman 1999], and increases visual understanding [Belcher 2003]. While we recognize that tangible AR is not appropriate for all situations, these benefits have been observed in other research in which users can physically manipulate a physical object that represents a virtual object (Chapters 5). Second, while interfaces fixed to a specific location are useful in stationary situations, we are interested in mobile interactions where the physical environment can potentially change as the user moves. For this, we require menus that present themselves "ready-to-hand" [Heidegger 1962], so that the user does not need to return to a particular palette or a fixed and rooted location. This difference is similar to the difference between selection from a pop-up menu that appears at the tip of a stylus or finger versus a menu bar at the top of a window. The user can continue to focus on the task at hand, rather than system interaction. Finally, AR systems are becoming increasingly popular, in part because simple printable props that rely on marker-based tracking (e.g., [Fiala 2005, Kato 1999]) provide an inexpensive, low barrier tool for interaction. We would like a lightweight menu technique that is actuated by a tangible gesture and does not require dedicated electronics, such as a button, to invoke a menu or confirm a selection.

This chapter presents *shake menus* [White 2009b], a technique that addresses these considerations [White 2009b]. In a shake menu, shaking an object activates a menu that is displayed around the object. A menu option is then selected by moving the object to the option. We are particularly interested in three scenarios of use: menu selection, combined menu selection and positioning, and information query and display. For example, in the first scenario, an optically tracked fiducial marker is used to represent a virtual object, and the object can be manipulated and inspected using spatial gestures. We would like the menu system to support previews of change prior to morphing the marker from one object to another. We would like to use some form of information display such that the information is ephemeral and only present when requested, providing a menu of information about the object. In the second scenario, we want to be able to view a set of optional objects for authoring, and select and place one of the objects using the technique. In the third scenario, we would like to use shake menus to query and manipulate data in AR situated visualizations (Figure 7.2).

Figure 7.2 Example shake menu displaying several aspects of a virtual voucher.

7.1 Related Work

A variety of interaction mechanisms for menus exist in the domains of desktop 2D and 3D, VR, and AR UIs. Here, we place our research in the context of 2D menus, hand-based menus, and prop-based menus. We also discuss work on shaking for actuation.

7.1.1 2D Menus and 3D Spaces

In 2D UIs, radial menus or pie menus [Hopkins 1991] and marking menus [Kurtenbach 1991] let the user stay within the flow of interaction by bringing the interface to the locus of interaction. Rather than shift focus and attention to a separate palette or remote fixed location, menu selection occurs at the current focus of attention. In VR and AR, a variety of techniques have been developed that adapt 2D menus to 3D systems. For examples, Vickers [Vickers 1972] introduced the use of 3D ray intersection in AR to select items on a printed physical menu on the wall of a room. Ring menus [Liang 1994] take advantage of the 1-DOF nature of menu selection to arrange menu items along a 3D ring. Hand and wrist movement rotates the items in the ring and the item within a visually obvious focus location is selected when a trigger action occurs. Mine et al. [Mine 1997b] describe how a virtual toolbelt worn by a VR user can hold needed tools in a known location relative to the user's tracked body. Comparing means of interaction with a 3D widget in a VR system, Mine et al. found that subjects were able to return more easily to a position relative to their own hand than to a position fixed in space. They also found a preference for interacting with a widget fixed to the hand rather than one fixed in space.

7.1.2 Hand-Based Menus

In these techniques, menu selection and presentation is centered on the user's hand. TULIP menus [Bowman 2001] assign menu options to different fingers of a pinch glove and options are displayed on the fingers and palm. Pinching the thumb and finger associated with a given menu option selects that option. Piekarski and Thomas's Tinmith-Hand menu [Piekarski 2002] also maps menu items to individual fingers, but displays menu labels fixed to the bottom of the display. Buchman et al.'s FingARtips [Buchmann 2004] uses the hand for selection, too, but uses gesture recognition to select menu items or objects. Although our technique is hand-oriented, it is prop-based.

7.1.3 Prop-Based Menus

Grosjean et al.'s Command and Control Cube, C^3 [Grosjean 2002], uses a 3×3×3 cubic grid to represent 26 menu options, presented on a rear-projected responsive workbench. The cube is activated by pressing a button and the hand controls an offset spherical cursor that moves within the C^3, which is itself offset from the hand (to avoid being obscured by the hand).

In the Tiles system, Poupyrev et al. use a book of fiducial markers as a menu [Poupyrev 2001]. Each page of the book displays a different object that can be selected and copied by moving a hand-held fiducial near the fiducial in the book. Proximity to a given fiducial marker in the book of options triggers a specific action. The form factor of the book provides a large number of options, but only one option was viewable at a time.

The Personal Interaction Panel [Szalavari 1997] displays sliders and options on the surface of a tablet as a panel of virtual controllers. The tablet is held in the nondominant hand and selections are made using a stylus prop held in the dominant hand. Tuister [Butz 2004] provides a tangible UI to menus using a physical cylinder with a handle. Menu items are displayed on the cylinder and twisting the cylinder changes the menu item. In contrast to these systems, we focus on the use of gestures for activation and specifically compare alternative loci of presentation.

7.1.4 Shaking

Shaking was described by Kato et al. [Kato 2000] as a potential gesture in their VOMAR system. Most of the subsequent work focused on actions such as tilting or proximity. Sinclair et al. [Sinclair 2002] use a shaking gesture to "sprinkle" hypertext links on objects. In this dissertation we have used shaking to activate and deactivate display of Visual Hints (Section 6.3).

Shaking has been used to activate a menu on an LCD screen [Hanumara 2008] embedded in a Chumby digital device. In this case, shaking makes a set of option selections appear on the body of the Chumby LCD screen, much like a mobile phone. Shaking is now commonly used as an actuation method on phones, such as the iPhone, whose accelerometers can detect a shaking gesture.

7.2 Shake Menus Technique

Building on this body of related work, our shake menu technique is also inspired by observing how people shake gifts and other objects to see what is inside; for example, shaking a box and listening closely for some hint of the possibilities hidden behind the wrapping. Here, we take this metaphor of shaking to reveal information and activate a menu. In this way, we provide an ephemeral interface that does not occlude and appears only when necessary. The shaking gesture (described in the following section) is simple and requires very little training.

Initially, no menu is visible. Shake menus are activated by shaking a hand-held fiducial marker or other tracked object. Once activated, a radial menu appears whose items are arranged around the circumference of the hand-held object. A menu element is selected by moving the fiducial into or optionally through the same space as the menu element. If no selection is desired, the fiducial can be shaken again to hide the menu.

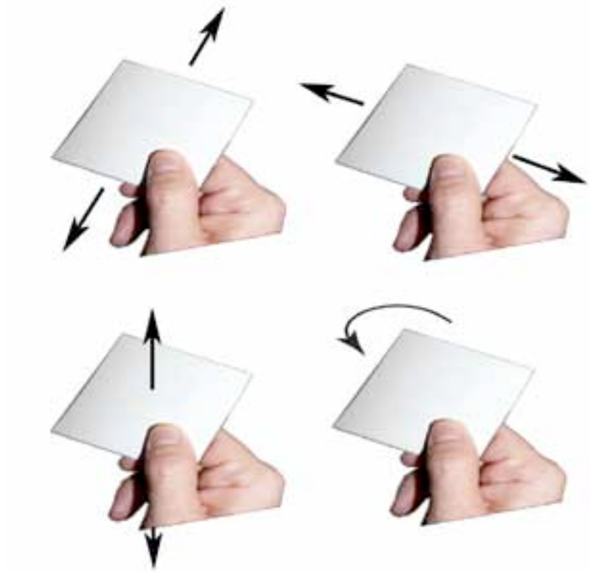


Figure 7.3 Shaking movement can be detected in three directions: (a) horizontal, (b) vertical, (c) back and forth, and (d) rotational.

7.2.1 Shaking

We detect shaking similar to the form suggested in the VOMAR [Kato 2000] system but distinguish horizontal, vertical, back and forth, and rotational shaking (Figure 7.3). We have investigated the use of other gestures, including pausing for activation of menus, but find that pausing is confounded with holding the fiducial still to focus attention on objects displayed relative to the fiducial. While a gesture such as shaking takes more time than a simple button press, gesture recognition has a strong advantage over the use of an active button: it makes it possible to support menus associated with uninstrumented objects, potentially allowing us to shake anything that can be tracked to see what it might reveal.

To detect shaking, we track the positions of our hand-held optical fiducial marker and record them in arrays for horizontal positions (x values), vertical positions (y values), depth positions (z values), and rotational values (about the x, y, or z axis). We detect the movement of the hand-held marker by calculating the difference between the current position and the previous position. For example, if the y value of the current position is greater than the previous position by a specific amount (2 cm for x or y and 4 cm for z in our implementation), we classify it as moving up. Movements are then parsed to generate gestures. Once four continuous movements of opposite sign in any one direction are detected within a time threshold (4 seconds in our implementation), we recognize it as a shake, and provide an auditory cue to let the user know that the shake has been detected. Here, direction is based on movement relative to the plane of the marker.

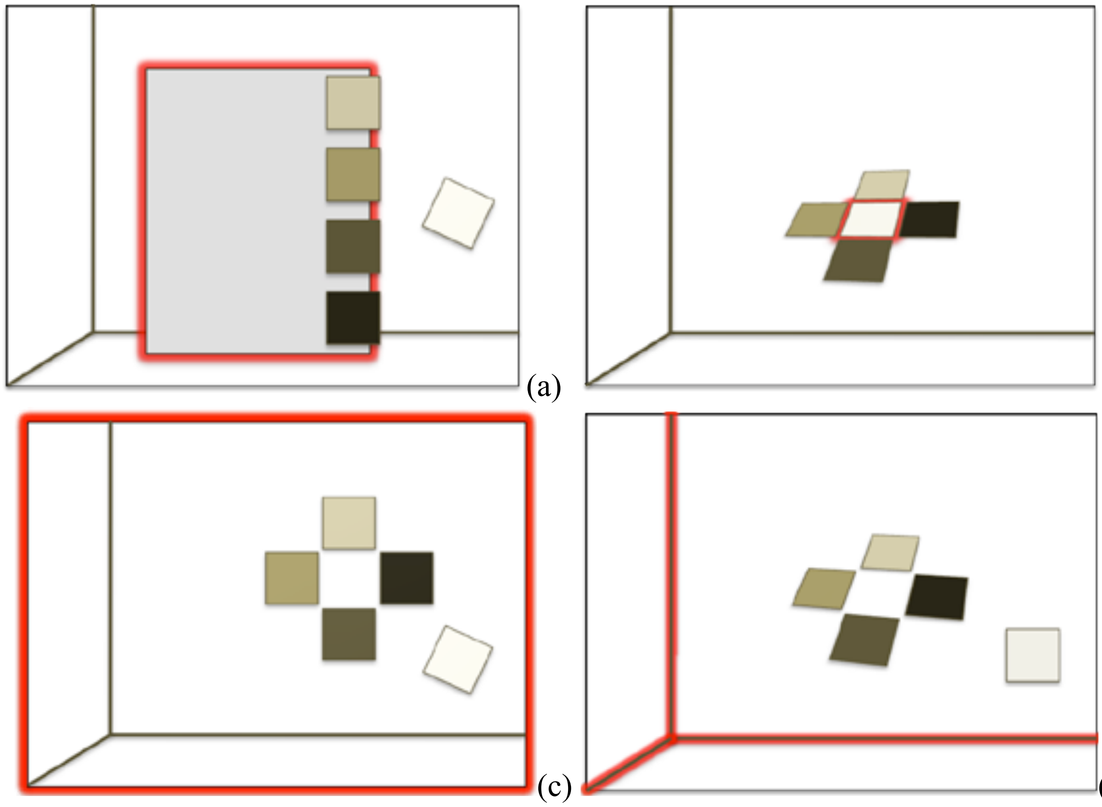


Figure 7.4 Menu placement. Red highlights show the reference for each of the different coordinate system conditions. (a) Clipboard, (b) object-referenced, (c) display-referenced, and (d) world-referenced.

7.2.2 Menu Placement

Once a shake has been detected, the menu appears immediately, and we wait a brief period of time before stabilizing it (1 second in our implementation) and then place the menus relative to the current position of the hand-held marker. We find the brief wait useful because the user normally stops the gesture once they hear the auditory cue that their gesture has been recognized, and this tends to stabilize the hand-held marker. For placement, we consider placement of menu options around the marker that are object-referenced, display-referenced, and world-referenced [Feiner 1993a]. These terms refer to how the menus, once presented to the user, are positioned relative to a specific coordinate system. In *object-referenced* placement (Figure 7.4b), the menus are attached to and move with the hand-held marker. In *display-referenced* placement (Figure 7.4c), menus are frozen, attached to the display in the location where the menus were activated. In *world-referenced* placement (Figure 7.4d), menus stay floating in the world, positioned where the menus were activated. Later in this paper, we discuss evaluation of different placement methods.

In the case of object-referenced placement, we record the position of the hand-held marker, constantly calculate and update the difference between the recorded position and the current position of the hand-held marker, and apply the difference to the menus. We do this so that movement towards a particular option actually moves the marker towards the option, yet the menu options move in the world with the hand-held marker. In this case, the menus stay with the marker even if the marker is held out of sight. This has the advantage that position and orientation of the menu can be changed to provide alternative views on the set of menu choices. This is particularly useful when the elements are 3D objects, supporting a change in point of view by changing orientation of the marker.

In the case of display-referenced placement, menus are frozen, attached to the display in the location where the menus were activated. The menus stay in this position even if the marker is removed. We record the position of the hand-held marker relative to the camera mounted on the head-worn display as a transformation matrix, transfer the menus from the marker node to the scene node, and apply the matrix to the menus, keeping the menus in a fixed position relative to the display.

In the case of world-referenced placement, the menus stay floating in the world, located in the position where the menu was activated. We accomplish this by using a ground plane array of fiducial markers to establish the world coordinate system, in contrast to the hand-held marker. Moving the head or hand-held marker will not move the menus. We record the world-referenced position of the hand-held marker as a matrix, and multiply it with the inverse matrix of the world marker node, transfer the menus from the marker node to the world marker node, and apply the matrix to the menus.

7.2.3 Selection

Our primary means of selection is through alignment of the hand-held marker with one of the menu selections. This provides a simple and intuitive means of selecting menu options. We reduce the degrees of freedom in the selection task by using a ray-casting technique for selection, as suggested by Bowman et al. Once the marker has been aligned with an option, we flash the option to provide feedback that the system recognizes the selection, but do not yet complete selection. This intermediate feedback gives the user the chance to change their mind and avoids errors from accidental alignment. Once the option starts flashing (after N seconds in our implementation) and the user briefly maintains the position, selection is completed and auditory feedback is once again provided so the user knows the selection has been made. We also hide the rest of the menu options to provide redundant cues to the user.

We have also explored crossing [Accot 2002] and marking techniques with shake menus. In these techniques, rather than aiming for a specific target menu item, selection occurs by moving in a particular direction or crossing over a target object. However, these seem more prone to error and require more research. Another option for selection is to use the direction of shaking movement to make the selection. This is similar to marking menus or rubbing [Olwal 2008].

7.2.4 Representation and Structure

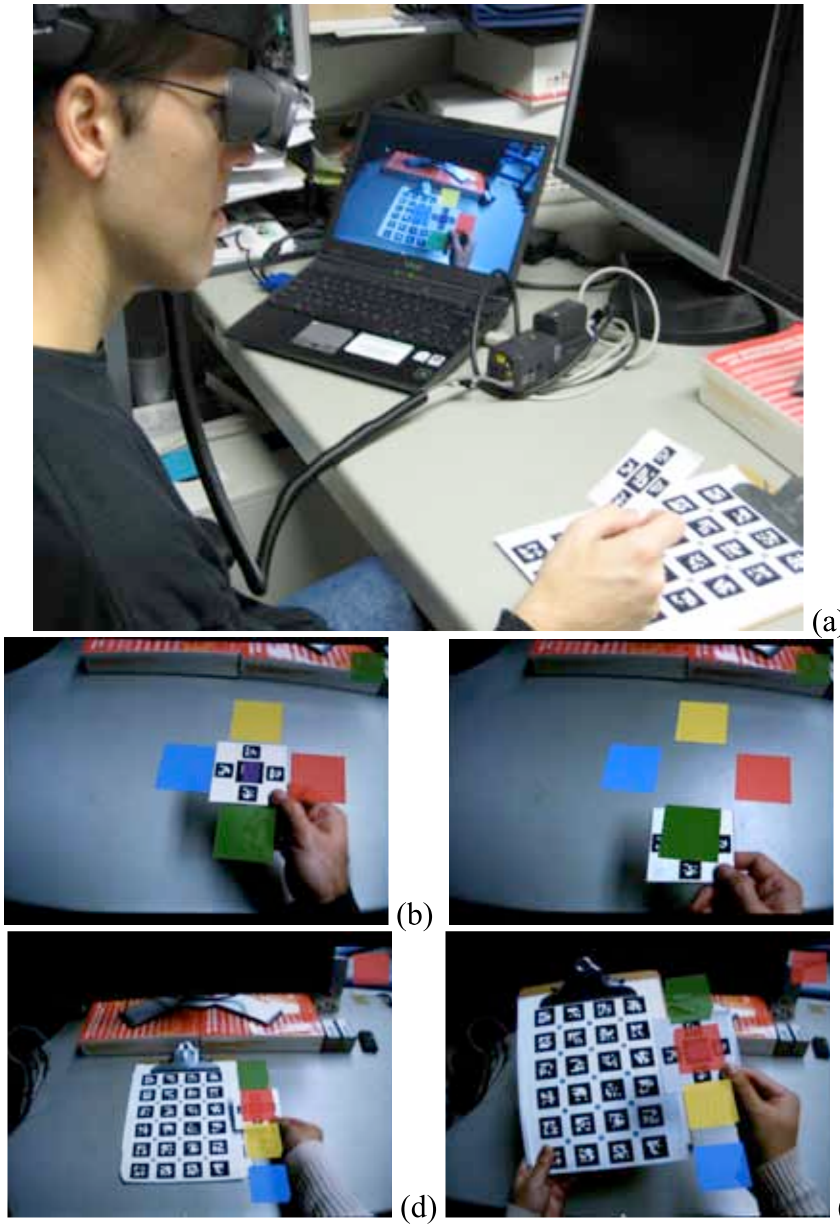
Currently, we present menu choices only in the plane parallel to the hand-held fiducial for several reasons. Grosjean et al. [Grosjean 2002] found that that error rates were higher in the upper and lower planes of the C^3 as compared to the central plane. We also want to keep a single plane of selection to avoid occlusion of 3D objects, allowing the user to tilt the fiducial, if the menus are object-referenced, and see different views of the 3D objects to be selected.

Thus far, we have experimented with only a single level of menu selections, but as with pie menus and marking menus, we believe the system will extend well to hierarchies.

7.2.5 Positioning an Object

We have explored an extension to the technique that supports a combination of selection and object placement, similar to the way that a fiducial marker paddle [Kawashima 2001] has been used for placement. In a normal shake menu interaction, the menu selection is accomplished, and the user continues on with their task. With the addition of positioning, once a menu item has been selected, the hand-held fiducial is then tracked to determine the position of an object to be placed. When the fiducial is quickly removed from view, the model stays in the location where the fiducial was last seen. This is inspired by the quick removal of a lightpen from the display to terminate drawing a line in early 2D graphics systems by causing the system to lose track of the lightpen [Sutherland 1963].

7.3 Experimental Evaluation



(a) (b) (c) (d) (e)

Figure 7.5 (a) Experimental configuration for user study. (b) Object-referenced menu presentation with color prompt in upper right-hand corner. (c) Selection of menu option. (d) Clipboard selection. (e) Bimanual clipboard selection.

In our discussion of presentation in the previous section, we described several different ways in which a menu can be presented to the user. To investigate the differences across presentation methods and to get feedback from users, we conducted a user study. Thirteen paid participants (12 male,1 female), ages 20–37, were recruited by mass email and flyers posted around our university. All participants were frequent computer users.

Our experimental conditions were object-referenced (OBJECT), display-referenced (DISPLAY), and world-referenced (WORLD), as shown in (Figure 7.4b–d). In addition, we included a baseline condition in which the menu is attached to another secondary fiducial, similar to the technique used in the Tiles system. In our study, this was a clipboard condition (CLIPBOARD) (Figure 7.4a). For this condition, we required that the marker come in contact with the menu options. We did this for several reasons. First, we are interested in comparing with proximity-based selection used in systems such as Tiles. Second, we wanted to avoid simple ray casting because some users in our early pilot tests picked up the clipboard and tilted it, causing the ray casting approach we had used to hit two separate menu options. Third, we wanted menu items to be displayed along the edge of the clipboard and selected based on touching the menu item. Note that this is necessary when the hand-held marker is attached to a virtual object that is being inspected because accidental selection can easily occur in simple ray-casting approaches when the marker is held close to the eye.

In this experiment, we compared the participant's performance selecting items from a menu using the four UI conditions described above, as shown in Figure 7.5. Menu item content was represented by colored boxes displayed above, below, to the left, and to the right of the hand-held fiducial marker or attached to a clipboard in the CLIPBOARD condition. For example, a yellow box could appear above the hand-held fiducial (Figure 7.5b). Participants were prompted to select a specific color and asked to make menu selections using each of the conditions. The order in which the conditions were presented was counterbalanced across participants. Time-to-select was recorded along with accuracy of the selection, as well as number of incorrect alignments prior to selection. A post hoc questionnaire was used to assess the participants' qualitative reactions to the different conditions. Total time for the study took approximately one hour and was conducted in a controlled laboratory setting.

We formulate two hypotheses:

H1. Object-referenced presentation will support the fastest menu selection. We believe that this would be true because the menus would always be in a known place relative to the hand.

H2. Object-referenced presentation will result in the fewest errors. Our rationale was the same as for H1.

7.3.1 Experimental Setup

The experiment was performed on an Intel Core 2 Duo 2.33 GHz PC with 2G RAM, running Windows Vista (Figure 7.5a). Video from the PC was output to a Sony LDI-D100B color, stereo, see-through head-worn display running at 800×600 resolution. A Creative Labs VF0070 USB video camera was mounted on the Sony display, capturing video at 640×480 resolution, allowing the display to be run in biocular video-see-through mode. Fiducial markers were mounted to a rigid card for the hand-held marker and a clipboard for the CLIPBOARD condition. The test system was built using Goblin XNA [Oda 2008].

7.3.2 Task

Participants were asked to make a menu selection based on automated prompting. The experimental environment consisted of a set of four colored cubes displayed above, below, left, and right of the marker.

7.3.3 Procedure

A within-subjects, repeated measures design was used consisting of four techniques (OBJECT, DISPLAY, WORLD, and CLIPBOARD). The single-session experiment lasted approximately 45 minutes and was divided into four blocks. Participants could take a break at any time by not activating the menu. Each block consisted of 80 trials of the four techniques (20 trials x 4 techniques) and the order in which the techniques were presented was counterbalanced across participants. Prior to beginning the trials, the participant was shown a video explaining the task and procedure to standardize knowledge about the experiment. The participant was then given a practice session so they could learn and experiment with all the techniques and run through a series of practice trials. The practice blocks were 8 trials of the four techniques (2 trials × 4 techniques) and the participant was allowed to repeat the trial block if they needed more practice. Two participants requested a single additional practice block.

Once the participant was comfortable with the techniques, they began the actual four-block session. Prior to each block, the participant was given an onscreen message telling them that the block was beginning. The participant was then free to shake the hand-held card and activate a menu. A sound was played to tell the participant the system recognized the shaking gesture and the trial menu of color options was displayed. Timing began once the color prompt was displayed in the upper right-hand corner (Figure 7.5b). Once the participant selected a menu option (Figure 7.5c-e), auditory feedback was provided and the menu options were hidden to acknowledge the selection. The next trial began when the participant shook the marker.

7.4 Results

7.4.1 Completion Time Analysis

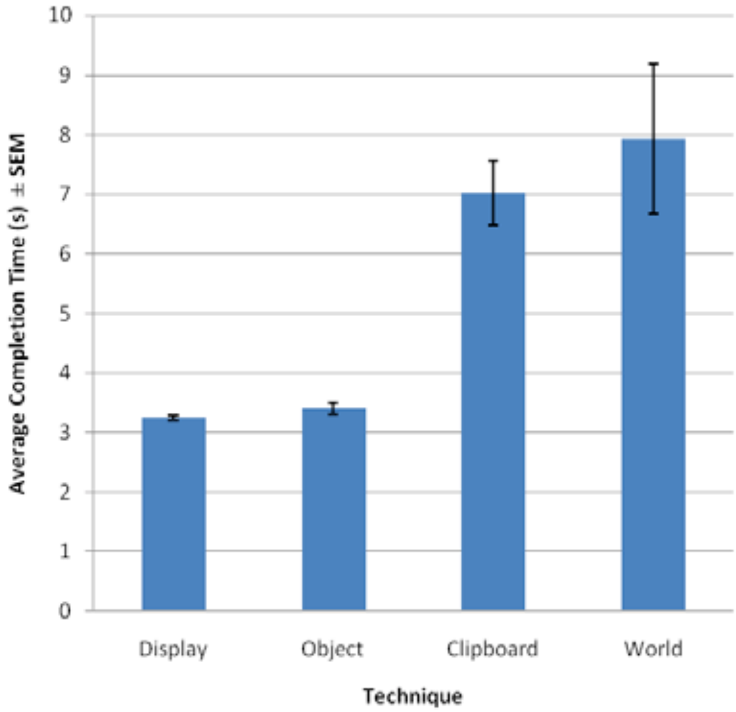


Figure 7.6 Average completion times (seconds) for the four conditions with standard error of the mean (SEM): DISPLAY and OBJECT were significantly FASTER than CLIPBOARD and WORLD.

We performed a one-way ANOVA for repeated measures on mean selection times for the successfully completed trials, with our participants as a random variable. We found significant main effects across several conditions for α=0.05 (Figure 7.6).

Technique had a significant main effect on completion times ($F_{(3,36)}$=8.65, p< 0.001). WORLD was, on average, more than 4.7 seconds slower than DISPLAY ($t_{(12)}$=3.0624, p<0.01) and more than 4.5 seconds slower than OBJECT ($t_{(12)}$=2.6637, p<0.02). With a Bonferroni adjustment (α=0.0125), the difference between WORLD and OBJECT in the paired samples t-test is not significant. However, using a modified Bonferroni procedure [Jaccard 1996] that still retains an overall type 1 error rate of 5%, α=0.05 for this specific test and the results are significant. CLIPBOARD was, on average, more than 3.8 seconds slower than DISPLAY ($t_{(12)}$=5.1290, p<.001) and more than 3.6 seconds slower than OBJECT ($t_{(12)}$=4.2063, p<0.01). There was no statistically significant difference between CLIPBOARD and WORLD, nor was there any statistically significant difference between DISPLAY and OBJECT. Overall, DISPLAY and OBJECT were the fastest for menu selection.

7.4.2 Error Rate Analysis

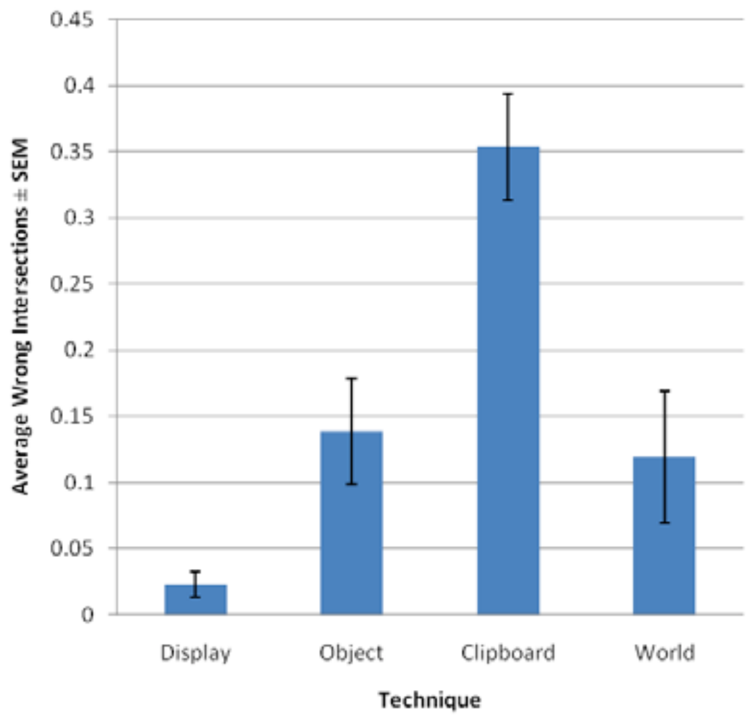


Figure 7.7 Average number of incorrect intersections for the four conditions. DISPLAY was significantly less error prone than OBJECT, CLIPBOARD, or WORLD.

We performed a one-way ANOVA on average incorrect menu item intersection data, with our subjects as a random variable. By incorrect intersections, we mean events where an incorrect option was intersected with a ray or touched with the marker prior to making the final selection. Significant main effects were found across some conditions (Figure 7.7).

Technique had a significant main effect on average number of incorrect intersections ($F_{(3,36)}$=12.22, p<0.0001). On average, CLIPBOARD had 0.33 more incorrect intersections than DISPLAY ($t_{(12)}$=6.1066, p<0.0001), 0.22 more than OBJECT ($t_{(12)}$=3.8751, p<0.01) and 0.23 more than WORLD ($t_{(12)}$=4.1569, p<0.01). On average, OBJECT had 0.12 more incorrect intersections than DISPLAY ($t_{(12)}$=2.8432, p<0.014). Using a Bonferroni adjustment (α=0.0125), the difference between OBJECT and DISPLAY in the paired samples t-test is not significant. However, using a modified Bonferroni procedure [Jaccard 1996] that still retains an overall type 1 error rate of 5%, α=0.025 for this test and the results are significant. There was no statistically significant difference between WORLD and OBJECT or WORLD and DISPLAY.

7.4.3 Subjective Evaluations

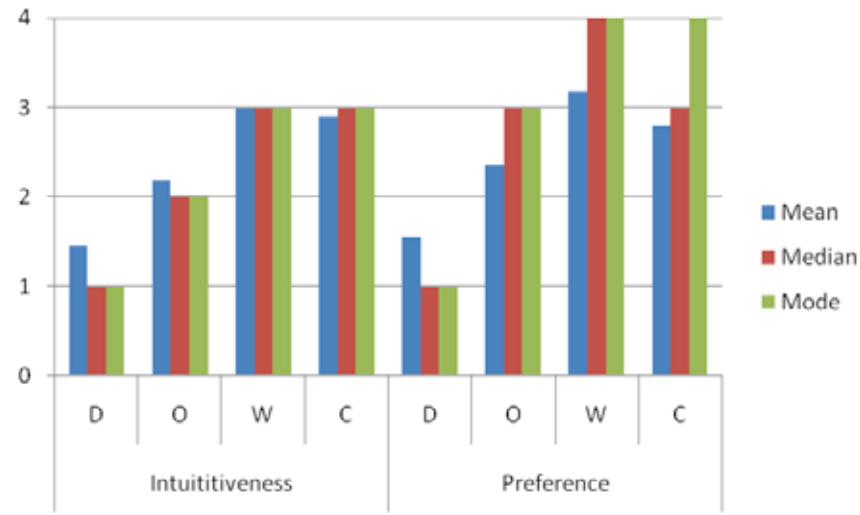


Figure 7.8 Mean, median, and mode for subject ranking of intuitiveness and preference of DISPLAY, OBJECT, WORLD, and CLIPBOARD conditions. Lower is better.

Subjects filled out post-experiment questionnaires rating their experience with four techniques on a five-point Likert scale (1 = most negative, 5 = most positive) for ease of use/difficulty, satisfaction/frustration, and intuitiveness/confusion. Participants were also asked to rank the techniques in order of intuitiveness and preference, from 1 (best) to 4 (worst). Subjects were then asked to respond with qualitative comments on each of the techniques. We present these results as indicators of general trends of user preferences and commentary, rather than conclusive evidence.

In terms of individual responses to ease of use, median response for DISPLAY (5) was highest, followed by OBJECT (4), CLIPBOARD (3), and WORLD (2.5). For satisfaction, median response for DISPLAY (5) was highest, followed by OBJECT (4), CLIPBOARD (3) and WORLD (2.5). For intuitiveness, median response was highest for DISPLAY (5) followed by OBJECT (4), CLIPBOARD (4) and WORLD (3). In terms of ranking intuitiveness comparing all four systems (Figure 7.8), median rankings were DISPLAY (1), OBJECT (2), CLIPBOARD (3), and WORLD (3). This reinforced the individual responses. For ranking preference (Figure 7.8), median rankings were DISPLAY (1), OBJECT (3), CLIPBOARD (3) and WORLD (4).

7.5 Discussion

We expected OBJECT to be faster than all other conditions. While this was significantly true of CLIPBOARD and WORLD, there was no significant difference in speed between DISPLAY and OBJECT. This disproves H1, that OBJECT would be fastest. From feedback and observations, we believe this is primarily due to the stability of the menu options as targets for selection in the DISPLAY condition. This is likely true for the number of incorrect intersections as well. We expected OBJECT to have the fewest incorrect intersections, but DISPLAY was significantly better than OBJECT. This disproves H2, that OBJECT would have the fewest errors. From feedback and observations, we believe two factors contributed to this. The DISPLAY condition was the most stable in the participant's view. In the case of OBJECT, the menu options moved with the marker and in some cases, this meant that the marker was accidentally aligned with a menu choice as both were moving.

We would expect WORLD to be similar to DISPLAY and OBJECT in terms of a Fitts's Law analysis [Fitts 1954] because the distance to the center of the target and width of the target are consistent across conditions. However, we believe additional conscious effort is required to move to a specific location referenced to a coordinate system not associated with the body. Grossman and Balakrishnan survey extensions of Fitts's Law to trivariate (3D) targets and suggest their own extension[Grossman 2004]. These extensions suggest that there may also be perceived distance and size differences as projected on the participants retina in different conditions, which may affect selection time.

We also made several interesting qualitative observations during the trial. Some of the participants did not consciously differentiate amongst world-referenced, object-referenced, and display-referenced conditions. We attribute this, in part, to the nature of the task for evaluation. Had the participants been asked to activate a menu and then move around, the distinctions would have been more apparent. We also found two types of interaction with CLIPBOARD. In some cases, subjects left the clipboard sitting on the table and moved to the location of the clipboard by using only one hand (Figure 7.5d). In other cases, subjects picked up the clipboard in the non-dominant hand and used a bimanual strategy to bring the two together (Figure 7.5e).

Audio feedback, selection flashing feedback, and menu placement timing were adjusted prior to the experiment, based on feedback from pilot usage of the test system. Getting these tuned made a large difference in the usability of the technique.

We also observed that some subjects experimented with orientation angle during the practice sessions prior to the actual timed trials. While this was expected, participants did not continue this experimentation during the actual trials. This is likely because there was no instrumental task that required alternative views of menu selection options. We also observed participants experiment during practice trials with shaking their head instead of their hand to activate the menu or move their head to select a menu item (in the case of DISPLAY) while holding their hand steady.

We initially avoided crossing techniques because of the high error rate reported for C^3 (7.2% for visual feedback) and our own experience. In this evaluation we found a 0.023% error rate for DISPLAY. We note that C^3 took, on average 1.0 second to select an item. Selection with shake menus takes, on average, 1.2 seconds (DISPLAY) and 1.4 seconds (OBJECT) if we do not include the 2-second pause required for selection. This implies that the actual move to a menu option is comparable and suggests that crossing techniques, such as those found in marking menus, should be further explored if we can keep the error rate low.

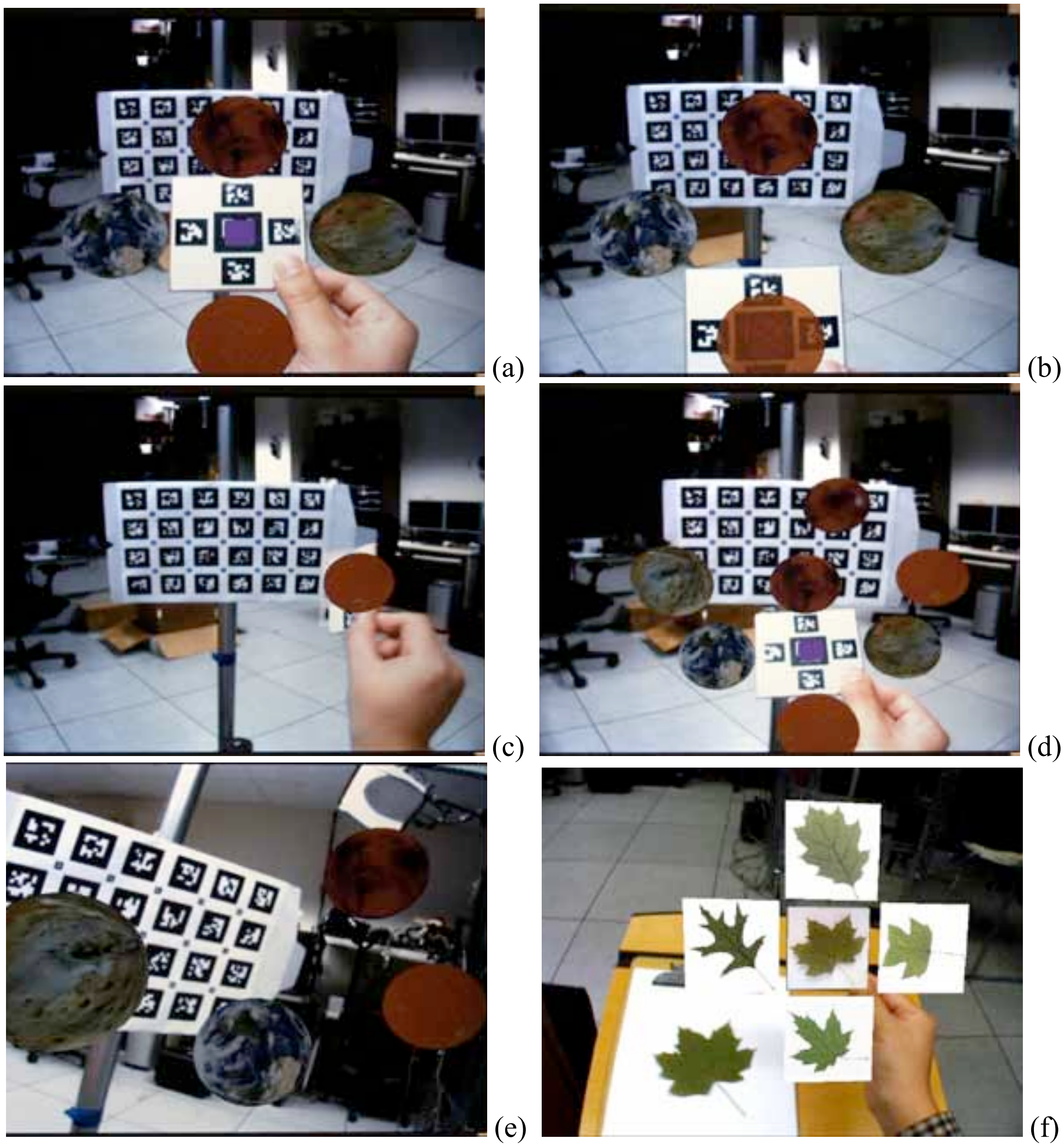


Figure 7.9 Example applications of shake menus, viewed through a tracked head-worn video see-through display. (a) Authoring a planetary system. After an initial shake, the menu appears. (b) A planet is selected. (c) The planet is placed in the appropriate location. (d) The process is repeated to add more planets. (e) Another view of planets. (f) Selecting and viewing potential leaf matches in AR UI to field guide for botanists.

7.6 Applications

Beyond lab evaluation, we wanted to get a sense for the use of shake menus for some of the scenarios introduced at the beginning of this paper. We implemented shake menus for a simple test application that made it possible to author "planetary systems" supporting a single flow of planet selection and placement using the same tool (Figure 7.9a–e). To create a planetary system, the user activates a shake menu and sees a set of choices. She selects the planet to be placed and moves the hand-held marker to the 3D location where she would like to position the planet. She then quickly removes the fiducial from the scene and the planet stays in the last known location of the fiducial. To add a new planet, she activates the shake menu again.

Based on our results from the user study, we implemented display-referenced positioning of the menu selections in this application. While this technique was more accurate, we did find that there is some benefit to object-referenced positioning when the user wants to change their point of view of the menu options, which is often the case when looking at 3D models. We are inspired to address the errors found in crossing styles so that we can remove the pause in the current selection mechanism to maintain a constant flow of action for the user in authoring.

We also tested shake menus for displaying leaf matches in the tangible AR system for use by botanists, described in Section 5.1 (Figure 7.9f) [White 2006b]. As we described in that section, in the current system, leaves are displayed either along the edge of a clipboard (much like condition CLIPBOARD) or in a semicircle around the leaf. Moving a hand-held fiducial marker into a leaf image morphs the object associated with the hand-held fiducial to a representation of information about that particular leaf. We require the user to touch the virtual leaf option because we found that a ray-casting technique caused false selections when the marker is brought close to the user's face to inspect or view the leaf. In using shake menus, we were able to show the sample leaf in the center. This made situated visualization of leaf results and comparison of the leaf results with the sample leaf easier. Selection of information about a virtual leaf was then a matter of choosing a menu option in the shake menu. Both the leaf and planet examples provide examples of potential use cases that inform our interest and direction for future work.

7.7 Summary

We have presented shake menus, a novel 3D menu technique that incorporates gestures for actuation and selection. Our study shows that display-referenced and object-referenced placement of menu selections were faster than world-referenced placement or a clipboard technique. In addition, display-referenced presentation was the most accurate of the techniques for presentation. Although shake menus with object-referenced or display-referenced placement are not appropriate for all situations, we believe they provide a fast, ready-to-hand menu for tangible AR systems and are applicable to a variety of applications. Our contributions include the shake menus technique and the evaluation of different coordinate systems for placement of the visualization.

In situated visualization terms, the technique applies to visualizations where the object is the primary context, and the visualization is semantically driven. In this case, we have used a head-worn, video see-through display and experimented with multiple loci of presentation with the locus of interaction centered on the object.

In the next chapter, we move from objects as context to a scene as context and investigate techniques for spatially driven visualizations.

8 Presentation and Interaction with Visualization Data using Scene as Context

Figure 8.1 SiteLens prototype (inset) and view of locally-sensed, geocoded carbon monoxide data (red) and remotely-sensed, spatialized carbon monoxide data (green) for comparison.

In previous chapters, we focused on techniques that support situated visualizations associated with individual objects of interest as the context. While these techniques serve many purposes, they may not be appropriate when the context is an entire scene or physical location. In this chapter, we present techniques for presentation and interaction with situated visualizations where the entire scene provides the context for situated visualizations. In particular, we are interested in techniques that address situated visualizations where the data is spatially related to the physical scene. We start by introducing a new application domain, *urban site visits*, which grounds our research in specific tasks such as identifying value and location of data in a scene and gaining new insights about a physical site. Urban designers and urban planners conduct site visits prior to a design activity in order to search for patterns or better understand existing conditions. We next introduce SiteLens [White 2009a] (Figure 8.1), an experimental system and set of techniques for supporting site visits by visualizing relevant virtual data directly in the context of the physical site. We address alternative visualization representations and techniques for data collection, curation, discovery, comparison, manipulation, and provenance. A real-use scenario is presented and two iterations of evaluation with faculty and students from the

Columbia University Graduate School of Architecture, Planning and Preservation provide directions and insight for further investigation. This work was done in the context of a larger project investigating tools and techniques to improve urban site visits, which we describe in the next section.

8.1 Site Visit by Situated Visualization (SVxSV)

This work was pursued in the context of a collaborative research project—Site Visit by Situated Visualization—that addresses the proposed site of Columbia's planned expansion, a 17-acre area of Manhattanville near Columbia's current Morningside Heights campus. In collaboration with Professors Sarah Williams and Petia Morozov of the Columbia Graduate School of Architecture, Planning, and Preservation, we have been investigating new ways for urban designers to conduct site visits prior to design interventions [White 2007c, White 2007d]. We take the system infrastructure, representation, and interaction principles from our previous work and apply them to this new domain. Here, our goal is to look at hybrids of manipulation and navigation where the situated visualizations are at environmental scales and can be manipulated through handheld devices or navigated through the physical environment. Our challenge is to both understand the spatial ties across multiple scales and develop principles, presentations, interactions, and algorithms that acknowledge these ties. In the next section, we describe site visits in more detail.

8.2 Field Study: Site Visits

Figure 8.2 Manhattanville area of New York, the focus of our inquiry into site visits.

We conducted a field study in the Manhattanville area of New York (Figure 8.2), engaging our colleagues in urban design and planning to better understand the process and goals of site visits. Urban planners, urban designers, and architects usually visit a site prior to a design activity related to the site. Different professionals use these site visits for different purposes, but the general goals include getting a sense for the physical site, finding patterns, and discovering and recording new insights about the physical location and its characteristics. Site visits are similar to ethnographic study in human-computer interaction research [Laurel 2003] in that they share an interest in the human element; site visits engage the physical place as well as the people and communities in that place.

Figure 8.3 A map of Manhattanvile (center) showing multiple collections of sampled CO data together with photographs (top-left, top-right) that provide context for specific locations on the map.

For example, an urban planner would first create a series of maps about a site that represent its demographics and use. She may then visit the physical site to view and photograph it, closely observing patterns such as congregations of people, traffic flows, and vegetation. On returning to her office, she would record patterns she found onto the maps. Existing tools for this process include geographic information systems such as ArcGIS as well as still and video cameras. A sample map (Figure 8.3) shows geocoded carbon monoxide data (CO) collected in our research along with accompanying photos that provide location context. Once these maps have been generated, discussion about the site focuses around the observed data and maps (Figure 8.4).

Figure 8.4 Urban planners discuss maps and photographs representing CO data in the Manhattanville area.

Several issues arise in the current process. First, there may be aspects of the site that are not visually apparent while visiting the site; for example, air quality and CO levels can be important when considering development, health, and environmental justice issues, but these cannot be seen with the naked eye. Second, the map data representing CO and the physical site are separate, imposing additional cognitive load on the user to place data in the scene or recall the scene when looking at a map offsite. In the photo in Figure 8.4, participants must try to imagine the CO data on the map as it might appear in the photos. Finally, still photos and video may not represent the dynamics of the physical site and its local environmental processes when trying to understand correlations or associations between the data and the site. The photos in Figure 8.4 are not dynamic and do not provide a sense for the variation and activity that may be present at a given location. Video introduces some dynamics but does not represent the current state of the site or significant changes in it at different times. It also doesn't support free exploration of the site.

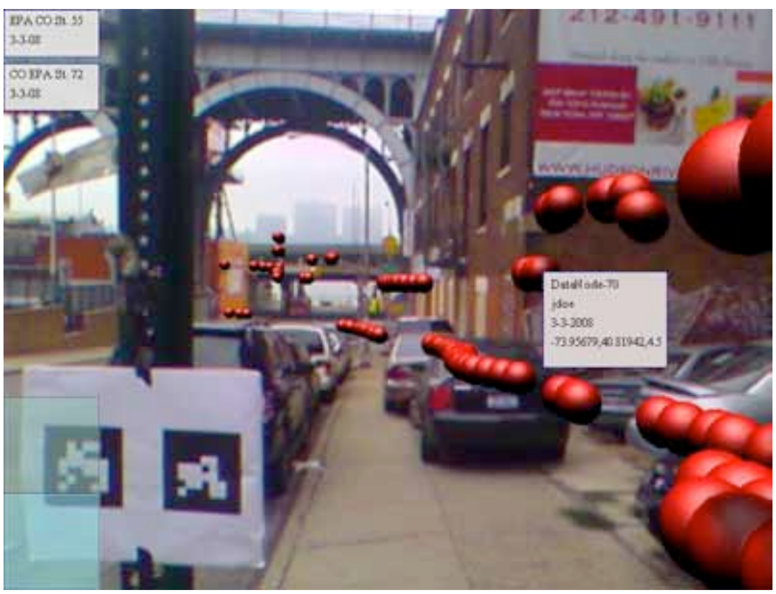
To address these issues, we introduce SiteLens, a prototype hand-held visualization tool to support site visits by enabling interaction with aspects of a physical site that do not have a natural or perceivable visual representation. Our goal was to develop a tool that helps urban planners and designers see hitherto unperceived patterns in a place and gain new insight about it through data visualization and situated visualization techniques that enable exploration of data.

8.3 Related Work

We draw inspiration from several projects. Reitmayr et al. [Reitmayr 2004] developed systems for managing and displaying large scale models and annotations in urban environments. Their focus was on wayfinding and accessing annotation about specific locations, not data visualization. Users could select a portion of a building and get more information about the building through a 2D, screen-referenced presentation. While the user wears a head-worn display, selection is done indirectly by moving a cross-hair in the user's view with a handhheld touchpad. The Vidente project [Schall 2009] has been investigating visualization of subsurface features, such as pipelines and power cables for utility field workers. Their approach takes geographic data models of these subsurface features and transcodes them for visualization and filtering. They use a custom-built, hand-held device for both display and interaction. In contrast, we focus on invisible aspects of a site beyond the built environment that may not have a natural visual or spatial representation and, in addition, on comparing multiple related datasets.

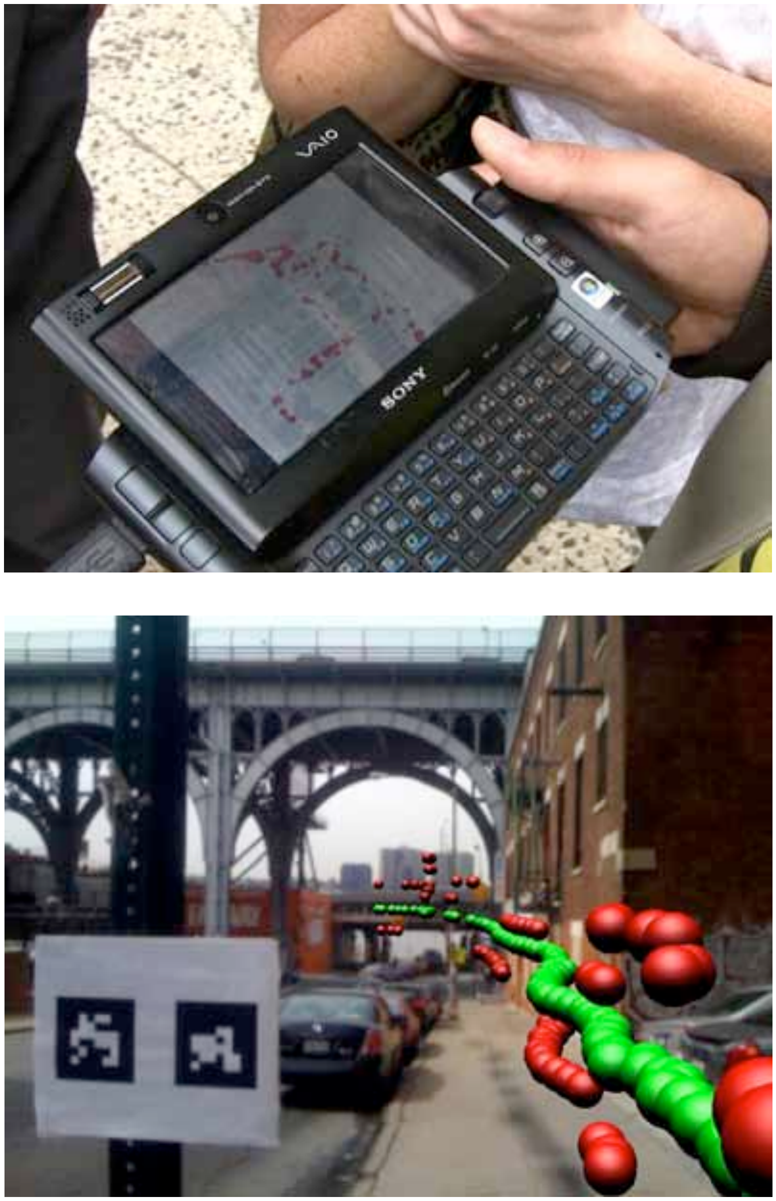
Sensed data has become an important topic in the HCI community as new ways of collecting data evolve and improve, such as mobile environmental sensors [Paulos 2007] and participatory sensing [Burke 2006], where groups of users equipped with sensors collectively gather data. Our work complements these systems by exploring alternative ways to visualize and interact with sensed data.

8.4 Interaction Task



(a)

Figure 8.5 (a) First iteration of SiteLens with data panels (gray boxes in upper-left corner) and detail pane (gray box in middle-right).



(b)

(c)

Figure 8.5 (cont.) (b) Dynamic map view. (c) Comparing locally sensed data (red) and remote EPA sensor reading associated with the site (green).

The following use scenario provides a description of the types of interaction and tasks we support and precedes explanations of specific elements of our prototype.

John is an urban planner. He typically looks for patterns in a physical location when he visits a site, and today he is interested in environmental issues. He arrives at the corner of 133rd Street and Broadway, an area of interest for future design activities, and takes out his SiteLens. The SiteLens shows him that there are several different datasets in the location, so he filters for environmental data. He sees two sets of CO data. He opens one and sees that it is displayed in the world (Figure 8.5a), so he knows that it was collected at the site. He opens the next set and notices that it is displayed referenced to the screen, indicating that it was not collected nearby. He tilts the SiteLens down to get a larger scale map view (Figure 8.5b) to determine where the data was captured. It is quite far away, so he tilts the SiteLens back up. He freezes the scene, queries both sets of data for provenance and notices that someone from the community collected the first dataset and the US Environmental Protection Agency (EPA) collected the second. He wants to compare them, so he drags the EPA data to the local data to spatialize it (Figure 8.5c). He makes sure that both datasets are visualized differently and sees that there is a large difference between the EPA data and local data. He freezes the image again, captures it for later use, and walks to the next street to investigate further.

In the next set of subsections, we describe aspects of the system and techniques in more detail. We focus on data curation, loci of presentation, visual representations, comparing and querying data, freezing to interact, tilt for overview, and sensor fusion for stabilization.

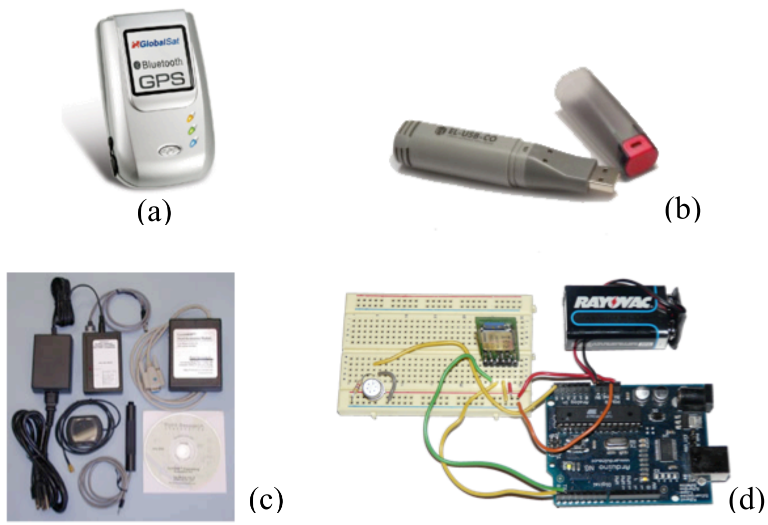


Figure 8.6 Tools for capturing geocoded CO levels. (a) Bluetooth GPS. (b) Lascar CO Datalogger. (c) GyroDRM Ded Reckoning module. (d) Custom-built CO sensor and Bluetooth transceiver.

8.4.1 Data Curation

Although we focus on visualization and interaction, data curation is a necessary and integral component of situated visualizations. As part of this project, we have been collecting and curating a variety of datasets to better understand the tools for collection, aggregation, and distribution. The red dataset in Figure 8.5a encodes CO levels we collected with a Lascar EL-USB-CO data logger and a Honeywell GyroDRM, which combines GPS with a gyro-stabilized dead reckoning module for geocoding when GPS readings are not available (Figure 8.6). Custom software combines data logs and converts the output to KML[KML 2009], an XML-based language schema and file format for representing geographic data maintained by the Open Geospatial Consortium [OSG 2009]. KML is used in Google Earth and Google Maps, making it easy to import datasets from these applications into SiteLens. CO data was also obtained from EPA sites and additional datasets have been curated from georeferenced US Census data and single-location environmental sensing stations. We note that the US Census data is georeferenced on a block or superblock scale, in contrast to our more precisely georeferenced CO data.

Figure 8.7 Census data in the upper left corner is display-referenced, while the red spheres representing CO levels are world-referenced.

8.4.2 Loci of presentation

SiteLens has three primary loci of presentation: a screen-referenced display in the upper left corner, a world-referenced AR display, and when the device is tilted, a world-oriented map display. Our system takes into account the scale and location of the data itself and defaults to displaying it in a locus that is appropriate to the spatial nature of the data. Here we use the term “scale” to describe the physical area that the data represents or is relevant to. We use the following heuristic to decide the appropriate locus: if the scale of the data is larger than 6 meters, we use the screen-referenced locus; otherwise, we use the world-referenced locus. This scale was chosen because we assume data that is on a scale equivalent to the width of a block will likely be difficulty to visualize in a world-referenced presentation. For example, in Figure 8.7, census data is relevant to the site, but it is recorded on a block or superblock (multiple city blocks) scale. Therefore, we present the data screen-referenced in the upper-left corner. In contrast, the locally recorded CO data is presented world-referenced because it is displayed in the locations in which it was recorded. Later, we discuss breaking these boundaries when comparing data.

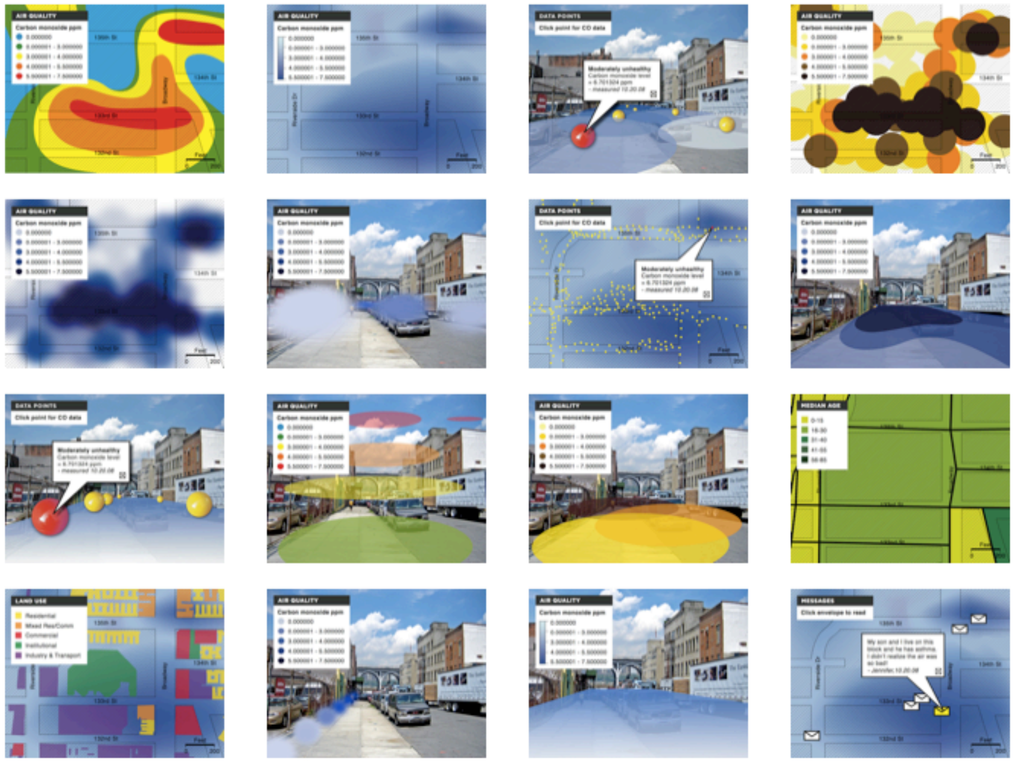
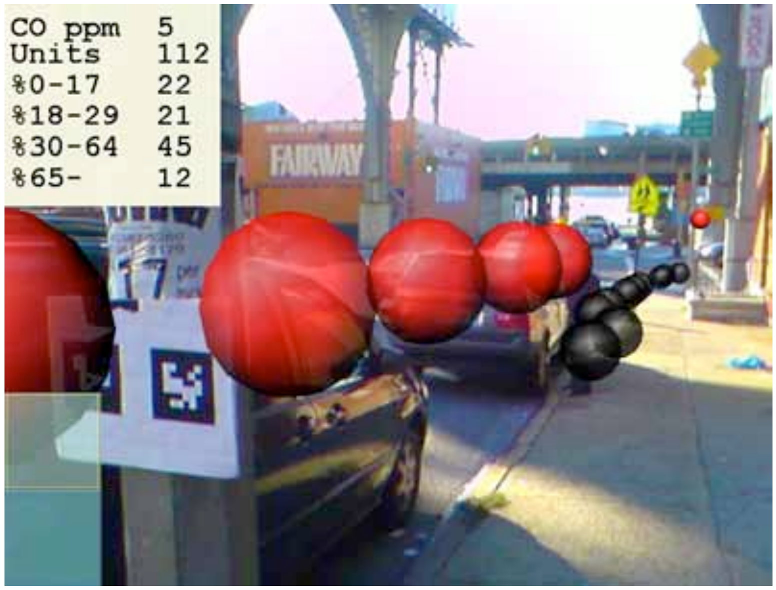


Figure 8.8 Design alternatives for displaying data in map and augmented reality views (developed in collaboration with Sarah Williams, Petia Morozov, and Candy Chang).

8.4.3 Visual Representations

When mapping a non-physical characteristic such as CO level to properties of a visual mark such as the size or altitude of a sphere, we consider the representation both by itself and in the context of the physical scene. To explore different representations, we use three different visual types: spheres, cylinders, and smoke. These three representations were developed in collaboration with our colleagues in urban design and urban planning after exploring a much larger design space of potential representations. A subset of design alternatives is shown in Figure 8.8. We chose these generic representations as a first cut at virtual representation of physical data for several reasons: the dots (spheres) are familiar cartographic representations, the representations were meant to be generic to other sensor data, and the abstractions lend themselves to redundant encoding of values. In each of the representations, the visual mark is displayed in the location where the data was sensed (Figure 8.9). For spheres, the parts per million (ppm) value is mapped to both continuous altitude and bi-level color. We chose 4.5 ppm as the break point because it is half the 9 ppm value considered actionable by the EPA [EPA 2009]. Higher red spheres have higher values, while lower, grey spheres have lower values (the first iteration of the user interface only used altitude and not color). For cylinders, ppm is mapped to both length of the cylinder and color. Taller cylinders have higher values and the color mapping is the same as used for spheres. For smoke, ppm is mapped to density. Denser smoke represents higher ppm values. We also experimented with shadows to reinforce the location of the data (Figure 8.10).



(a)

Figure 8.9 Alternatives for mapping CO level to visual representations. (a) Spheres.

(b)

(c)

Figure 8.9 (cont.) (b) Cylinders. (c) Smoke.

8.4.4 Comparing and Querying Data

Data comparison facilities provide a way to validate existing datasets. If two datasets contain spatial data relevant to a given physical location, they can be compared directly. However, sometimes data intended to represent a physical location is actually collected remotely. In this case, we provide a means to spatialize data to match the georeferencing of a related dataset. For example, in Figure 8.10, the red CO dataset was collected in the locations in which it is represented. However, the green CO dataset, which was collected several miles away, is the closest EPA dataset. Instead of comparing the red data to the single numeric value representing the EPA data by default, we let the user spatialize the EPA data. The user does this in SiteLens by touching the display-referenced data panel on the 2D display for the EPA data and dragging it to the spatialized red data. We hypothesize that this makes visual comparison simpler without losing the relevance of the physical context.

Figure 8.10 Locally-collected, georeferenced data (red spheres) is compared with remotely collected, spaitalized data (green spheres).

As with any medium, additional information about the data being visualized can help the viewer better understand potential issues such as bias or reliability. For example, a visualization of CO data may be perceived differently depending on whether it was created by a community member or a known industrial polluter. To help address this, the user can select any visualization node by touching it to bring up an information pane that provides metadata such as provenance and creation date. An example is shown in Figure 8.5a.

8.4.5 Freezing to Interact

Selection of a particular data point can be difficult when the world and associated visualizations are moving. To address this, we provide a button that freezes the video image but not the dynamics of the AR system, similar to Güven et al.'s use of freezing to author [Guven 2006]. All regular interactions are active in this mode. In addition, a newly visible button activates a scene grabber to save the image for later use.

Figure 8.11 When oriented down (parallel to the ground), SiteLens displays a 2D map view of the local area with data displayed on the map.

8.4.6 Tilt for Overview

Most of the viewing with data is accomplished while holding the device up, like a magic lens, through which the user can see invisible aspects of the site directly overlaid in the scene. However, this does not provide an overview of the larger dataset and physical site together. We address this through a tilt for overview technique. When the device is held upright, data is displayed using augmented reality on top of the scene. When the device is oriented down, parallel to the ground, a 2D map view of the local area is displayed to provide a larger overview (Figure 8.11).

8.4.7 Sensor Fusion Stabilization

Tracking and registration in outdoor augmented reality is still an area of active research. While indoor tracking can rely on highly calibrated placement of sensor arrays such as the Intersense IS-900 or controlled lighting conditions for vision-based recognition, outdoor use must address varying environmental conditions with little reliance on dense sensor systems. Several approaches exist. RTK GPS has been used for highly accurate position and orientation of the mobile system. However, these solutions generally require an active base station, large antenna, and a sizable backpack for carrying the additional hardware for translating the correction signal. More recently, efforts have been made to build models of the surroundings for tracking and registration. These models have been based on aligning features acquired by computer vision with existing models [Reitmayr 2006], or by using simultaneous localization and mapping algorithms (SLAM) to construct position of points in 3D [Klein 2007]. Some of the approaches use sensor fusion [You 2001], combining multiple sensors for tracking such as GPS, orientation sensors, computer vision features, or dead reckoning data. Sources are combined using a variety of different approaches including Kalman Filters [Foxlin 1996] and heuristics for combining individual orientation and position. However, none of these approaches take the data to be viewed into account.

Our goal is to reduce jitter and large jumps that can occur even with small changes in orientation when data is far away. For instance, when a sphere is displayed a foot away from the viewer, an orientation change of a degree produces a minor change. However, when data is displayed 100 feet away, the sphere is much smaller, and although it travels the same distance, a small change in angle that changes the position of the sphere appears amplified and makes the data move much more.

Our approach is a form of sensor fusion that incorporates the distance of viewed data from the viewer. We combine input from a GPS, InterSense InertiaCube3 orientation sensor, and ARTag fiducial markers. Throughout the physical site, we place arrays of fiducial markers with known position and orientation. Because of urban canyon effects, we do not always have access to a GPS signal. The algorithm works as follows: if there is a visible fiducial marker array that is close and steady and if data is near, we use the fiducial array for position and orientation; otherwise, we use the fiducial array for position and use the InertiaCube3 for orientation. Currently, a change from one mode to another causes jumps in the data, but these could be addressed by interpolating between tracking modes. If the fiducial array is far away but visible, we use it for position and use the InertiaCube3 for orientation. If the fiducial array is not visible, we use GPS for position and the InertiaCube3 for orientation. We can express this in pseudocode as:

```
M = set of all visible fiducial marker arrays
O = orientation from orientation sensor
G = position from GPS
D = set of all visible data

if M is not empty
   m = nearest visible fiducial marker array in M
   if (steady(m)) then
      for each data point d in D
            if (distance(d)) < threshold)
                  position = position(m)
                  orientation = orientation(m)
            else
                  position = position(m)
                  orientation = orientation(O)
   else
      position = position(m)
      orientation = orientation(O)
else
   if GPS exists
      orientation = orientation(O)
      position = position (GPS)
   else
      orientation = orientation (O)
      position = last best position
```

The two areas that can be adjusted are the analysis of steadiness of the visible marker observation and the distance threshold. In our approach, we have calibrated this based on observation, but a better approach would likely be to monitor the number of pixels of movement in distant data and set the steadiness and distance threshold based on a pixel distance threshold. In SiteLens, we use a threshold of 20 meters, a number that is based on our experience with the first iteration of the prototype. For steadiness, we considered changes of greater than 10 degrees over a period of one second to be unsteady.

In this example, we apply heuristics to each data point individually. However, when part of the visualization uses the fiducial for orientation and part uses the InertiaCube3, the difference can cause breaks in the apparent contiguity of data. We address this by applying the heuristic to the entire visible set of data as if it was a single point with extent in the farthest position. The advantage of this approach is that it maintains relative stability of data that is distant and generally hard to localize.

8.5 Implementation

Our prototype runs on a 1.2lb Sony VAIO VGN-UX390N Ultra Mobile PC with a built-in camera, GlobalSat BT-338 GPS, and InterSense InertiaCube3 (IC3) inertial orientation tracker. SiteLens is built on top of Goblin XNA [Oda 2008], which supplements Microsoft's XNA infrastructure with AR functionality, including 6DOF optical marker tracking using ARTag [Fiala 2005].

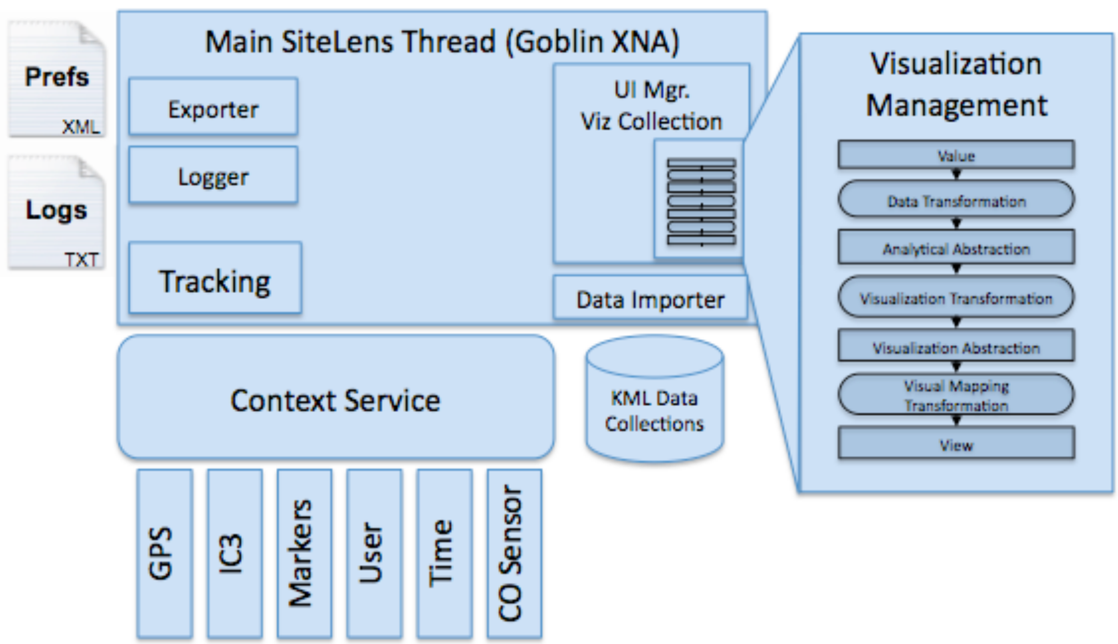


Figure 8.12 SiteLens architecture diagram

The SiteLens architecture borrows from our Electronic Field Guide architecture in Chapter 4, extending it in several areas. First, we now need to know the orientation and location of the display relative to the physical world. A new tracking component, which gathers orientation, GPS, and fiducial marker orientation and position, manages this information to represent spatial context. Second, we use the same concepts for visualization management but maintain a collection of visualizations that are currently active, either display-referenced or world-referenced. Third, we incorporate a component for importing and loading georeferenced data. Fourth, we incorporate a component within the context service for gathering live sensor data via Bluetooth.

8.5.1 Main SiteLens Thread

Goblin XNA maintains the scenegraph and manages user input. On start-up, the main SiteLens thread initializes the context and tracking component service to start gathering spatial context information. The main thread also initializes the UI elements and loads a preselected set of georeferenced data into the visualization collection. UI Management is handled through the main UI thread. This manages UI elements such as buttons for freezing the scene and capturing images.

8.5.2 Context Service and Tracking

The context service gathers data on current location and orientation through a combination of ARTag fiducial markers, GPS, and IC3. (In contrast to Vidente, which was primarily tested in areas with clear visibility to GPS satellites, we use optical markers to address urban areas with limited GPS satellite visibility.) The tracking component implements the algorithm described in Section 8.4.7 to provide a stable presentation to the user.

The context service also receives input from our custom built Bluetooth-based CO sensor and passes this data on to the main thread. This supports gathering and visualizing live sensor data in future research.

8.5.3 Data Importer and Collections

The data importer currently reads in KML files that associate georeferenced locations with values. Our data includes multiple sets of georeferenced CO data gathered for this research, EPA CO data, and georeferenced US Census data. Spatial data is stored in an octree [Samet 1988] to provide quick access to location and distance information from a given data node to the current location.

8.5.4 Visualization Management

The visualization manager is similar to the EFG visualization manager. However, here each visualization is initialized with specific abstract representations for each data element in a data set and these representations can be changed to represent different colors, shapes, and mappings. Each visual representation is a subclass of a data node in our architecture, so we can easily create new visual representations and data mappings.

8.6 Evaluation

Plaisant [Plaisant 2004] argues that evaluations of visualization techniques should incorporate real tasks and field studies. North [North 2006] suggests insight as an indicator for validating visualization techniques. As a first step in evaluating our prototype, we obtained feedback from urban designers and planners in the Columbia University Graduate School of Architecture, Planning, and Preservation (GSAPP) through two iterative field studies. The studies were conducted in Manhattanville. In the first informal study, two colleagues from GSAPP explored the site using the scenario described earlier in this paper. Feedback from this study informed the iterative design of the prototype used in the second study. In the second study, four participants from GSAPP used a revised prototype at the same site and were given a brief post hoc questionnaire that elicited opinions about visual representations and system use. In both cases, researchers were present and observed subjects as they used the system. Additional unstructured discussions with subjects followed both studies. The next section describes the second study.

8.6.1 Experimental Setup, Task, and Procedure

This within-subject, single-session experiment compared each participant's performance on a set of tasks (e.g., identifying data location and value) using the three different representations described earlier. A written post hoc questionnaire assessed the participants' reactions to the different conditions. Total time for the study took approximately one hour.

The study was conducted on the sidewalk of the north side of West 133rd St., west of Broadway, and consisted of two phases. In the first phase, each participant was presented with each of the three user representations and asked to complete a set of tasks that involved identifying specific data values and associating data with a physical location. In addition to measuring time to completion, we asked participants to fill out a questionnaire regarding the perceived ease of use and efficacy of the three representations, the different loci of presentation, comparing and querying data, freezing to interact, and the prototype in general. In the second phase, participants were asked to use SiteLens to observe and compare data found in the visual scene. They then reported on insights, patterns, and structures in the data observed by using the system. In the next subsections, we report on results from this evaluation.

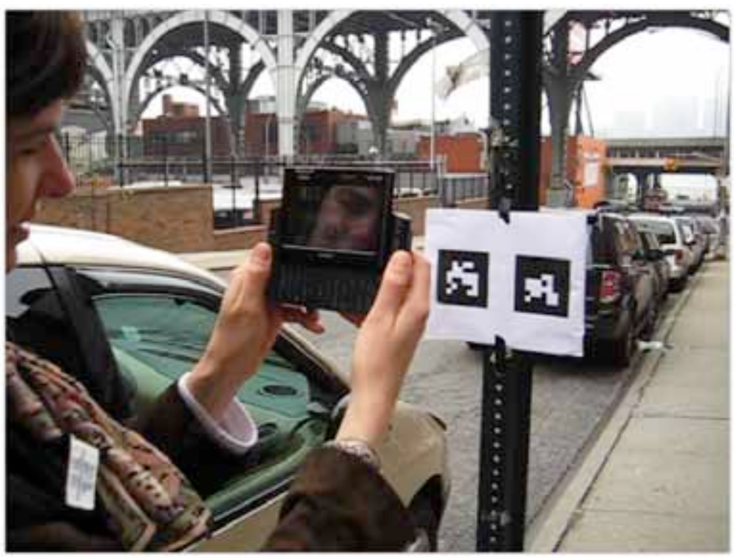


Figure 8.13 Participant looking through SiteLens at visualized data in Manhattanville.

8.7 Observations and discussion

In this section, we discuss our observations and results from the written questionnaire. We focus on the use of situated visualization for generating new insights as well as for presentation, interaction, representation, and distinctions between navigating and manipulating the visualization.

8.7.1 Insight from Situated Visualization

During our field experiment, participants reported on new moments of insight. For example, map data alone could not explain why the locally-recorded CO levels were higher towards the end of one street; however, visual inspection of that street during the field study revealed that cars were idling as they prepared to enter the highway close to where the higher CO levels were recorded. This combination of virtual and physical observation provided insight into potential causality. One participant explains how such insights made her more thoughtful about the site:

> The system raised interesting questions about what I should be sensing or relating to. It prompted me to look for clues about the environment that I ordinarily wouldn't, and it made me curious about what it means. It heightened my awareness, and I thought that was good. (anonymous participant)

As this quotation demonstrates, SiteLens generated new insights about the site in question, but also at a more meta level, it generated insight about the new possibilities for site visits through situated visualization. Participants were inspired to became as engaged with the process of a site visit as with the performance of it.

One frustration with the system was that the data was considered "stale" because it had been collected a month prior to the study. This brought up two issues. First, while there was a closer spatial association between the site and sensed data, the temporal association was unclear. Second, as the following quote indicates, participants expressed a desire to have live or dynamic sensing coupled with existing data:

> It would be cool to think of this alongside social networking tools […] it could still add more information to the analysis in different ways, ways that turn passive observation into overt surveillance […] to generate more qualitative information about a place, like how people perceive the environment, or how people sense pollution without really knowing if it's there. These perceptions could help with the phase of design in which interviews and site surveys are done, but only with a handful of people. (anonymous participant)

This suggests a usage of the system were the tool goes beyond collecting data and enables new ways to perceive and understand the surroundings.

Beyond visualization, some participants wanted to incorporate simulation into the system so they could explore design alternatives in the physical space: "It would also be cool if I could model some conceptual design ideas in the Sitelens, like an overlay, and see how the analysis of the place would change" (anonymous participant). While this activity would typically occur later in the design process, it implies that iteration of study and design intervention could occur in the same tool.

The overview view, which was displayed when the device was tilted down, proved to be useful as well. As one participant wrote: "[I like] the possibility of beginning to see the invisible--to link to complex causality […] that you could move from map view of a larger area to a view where you could scan what was in front of you" (anonymous participant). Using the overview view to quickly moving back and forth between the overview and data in the immediate vicinity enabled participants to view patterns at multiple scales.

8.7.2 Representation and Presentation

Reactions to the different representations were mixed. Spheres were considered better than cylinders for localizing the data. In terms of specific data values, participants were initially confused about whether the CO ppm value was mapped to sphere size or height. Because we did not initially have a model for the altitude of the street, we also found areas where the altitude of data varied because of street grade, not variation in CO level. This was due to the fact that we mapped to altitude above sea level, not altitude above street level. Surprisingly, we found that the psychological impact of the smoke was more important than the more accurate localization and value of the other representations. One participant said "I like the smoke…It's hard to see quantity of things, but… psychologically it helps to represent the idea better" (anonymous participant). Another suggested that, "you just need to know bigger or smaller, but not the actual value" (anonymous participant). In further discussion during a project meeting, smoke with the option of visualizing spheres was suggested, because the initial representation of smoke provided a stronger psychological effect, provoking stronger reactions.

In general, we see the need to provide the user with more control over visual form (geometry, color, size) as well as data mapping. Chuah et al.'s SDM [Chuah 1995] is useful to consider in this regard, because in this system, users had control over the visual representation and could transform the 3D visualization so that data could more easily be compared by lining data up along the 2D surface parallel to the view. For example, while shadows were considered useful for providing distance cues and enhancing the sense of realness, our design choice for mapping CO concentration to height was not considered obvious. Participants wanted to try alternate visual representations to explore changing data perception.

In terms of presentation, there was a difficulty with the screen-referenced display. One participant observed that, "[It] doesn't change when I move around, so it feels less important" (anonymous participant). Thus even though the data was representative of the site (including census data), participants felt that it was insufficiently dynamic.

8.7.3 Interaction

In our first iteration of the system, participants were distracted by the instability of data. Our combination of sensor fusion incorporating data position, discussed in Section 8.4.7, in the second iteration significantly stabilized the visualization. While the actual placement of data was less accurate, the visualization of data was more stable and sufficient for associating with local features of the environment.

Freezing the camera image when desired while keeping the overlaid graphics live supported manipulating the interface and visualization without having to keep SiteLens pointed at the scene being overlaid. As an extension of this, we found that the on-screen user interface controls were best positioned in the lower left and lower right of the screen and along the edges where the user's thumbs could easily access them, which reinforced observations made using the LeafView UMPC prototype. Direct manipulation of the visualizations, such as touching them to show metadata, was useful once the display was frozen. However, selection of specific nodes in dense areas of data was still difficult because of overlapping nodes.

Our participants felt that capturing combined images of the physical and virtual scene to create a single more complete image was useful for documenting the site visit. Using the SiteLens prototype was not felt to be significantly harder than using a video or still camera and could be imagined as a common tool. It was even suggested that SiteLens could be used for an iterative process of data curation: "i wanted to be able to tag, share, map over time, annotate, log. the potential was very exciting…" (anonymous participant). In this data curation scenario, visualization and sensing would be combined with organizational tools to help create new datasets that, in turn, create a portrait of the site.

8.7.4 Navigation and Manipulation

Although we designed the system for direct manipulation of data through the video see-through display, participants also described navigating through the visualization as useful:

> It seems like a very useful tool to move through space engaged with the location at various scales and coming to grips with forms of causality and agency that are simply not visible. (anonymous participant)

> [I could use this] to pose all sorts of questions to as part of moving through place and being part of a place -- long term and short. (anonymous participant)

These user experiences imply that situated visualization, where the scene is the context, requires user interface techniques that address both modes of exploring and interacting with data.

8.8 Summary

In this chapter, we have presented urban site visits, a new application space for grounding situated visualization research. We have developed a set of interaction and presentation techniques for situated visualization in this domain that enable new types of insight. These techniques are embodied in a prototype system, called SiteLens, which we have described and evaluated through iterative field experiments with experts in urban design and planning. We found that situated visualization enables unique insights that would have been difficult to observe using traditional tools for site visits. We also found that our interaction and presentation techniques enabled a new tool for site visits, thereby improving the resulting analysis and increasing understanding of the site.

Our work provides a new tool for urban designs and planners to visualize data during site visits. We also contribute to computer science in the following ways: the creation of new techniques for spatializing data that improve comparison of data sets, analysis of comparing representations for mapping CO values into visualization, the creation of a novel sensor fusion algorithm accounting for data that provides a more stable representation of the visualization, the development of orientation changes for enabling data overview, and the ability to freezing the scene and thus improve stability while querying or interacting with data.

In terms of situated visualization, we investigated the potential for visualizations that use the entire scene as the context, where the data is spatially related to the scene. We found that world-referenced presentation enables association of data with specific aspects of the physical scene. We also found that display-referenced presentation, while useful, was perceived as too static and disconnected from the scene even when the data was relevant to the scene.

9 Conclusions and Future Work

In Section 1.1, we described five dissertation goals: create a theoretical framework, investigate presentation and display techniques, develop interaction techniques, understand benefits, and synthesize design principles. Each of these goals is addressed in this dissertation. Chapter 2 presents our theoretical framework. Chapters 4–8 describe development, implementation, and evaluation of presentation and interaction techniques together with their potential benefits. Section 9.2 concludes with a synthesis of design principles based on our experience investigating situated visualizations.

This dissertation has explored the design, implementation, and evaluation of novel techniques for interacting with and presenting situated visualizations across multiple contexts, relevance relationships, and display modalities. Each of the previous chapters has provided a detailed discussion of the individual contributions. In this final chapter, we first summarize our contributions. We then synthesize our experiences from this research and present a set of design guidelines for creating effective situated visualizations. We conclude with a discussion of potential directions for future work and some final thoughts about situated visualization.

9.1 Summary of Contributions

The research in this dissertation has explored three important aspects of situated visualization: mobile visualization, objects as context, and scenes as context. First, in Chapter 4, we investigated mobile visualization as a means of bringing visualization out of the laboratory and into the world, physically closer to the relevant context. We focused on techniques for inspection and comparison, including feedback loops for understanding the quality of context. This work served as a baseline for comparison with other situated visualization techniques. Next, in Chapters 5–7, we investigated a series of AR visualization techniques that use objects as context. In these chapters, we examined techniques in which the relationships between the visualization and object of interest are semantically- and spatially-driven. We investigated direct manipulation of data for inspection and comparison, representations for discovery and learning tangible gestures, and presentation and selection techniques for radial display of information. Finally, in Chapter 9, we investigated situated visualization that uses the entire physical scene as context. Here, we focused on interaction with the data, specific representations, and presentation across multiple loci of presentation. More specifically, this dissertation has presented the following contributions:

Descriptive characterization of situated visualization (Chapter 2). We presented a framework and vocabulary for describing situated visualizations. This framework serves both as a tool for comparing different forms of situated visualization and as a predictive mechanism for revealing areas that have not been explored.

Mobile visualization techniques (Chapter 4). We initiated a field study with botanists collecting and identifying botanical species to develop a better understanding of the overall task and ground our research with specific tasks. We then created an Electronic Field Guide architecture to provide an extensible infrastructure for experimenting with alternative matching algorithms, data sets, and visualizations. In addition, we developed two hand-held system prototypes, LeafView Tablet PC and LeafView UMPC, through multiple iterations to enable mobile identification of botanical species in the field. Our prototypes, built on our EFG architecture, embody techniques for improving the task at hand, such as contextual feedback, proximity of matching results for comparison, semantic zooming for inspection, and seamless inclusion of identification in the collection process. Field experiments show improved identification speed and efficacy through the provision of feedback on the correct representation of context and through presenting visualizations in close proximity to the object of interest. We also find that bringing visualization into the field further aids identification when matching with additional physical context beyond a single leaf, such as the plant structure with fruit and bark, is supported.

Tangible and Head-Movement–Controlled Augmented Reality techniques (Chapter 5). We developed two mobile augmented reality prototypes, TAR-EFG and HMCAR-EFG, which embody techniques for tangible augmented reality and enable head movement control in support of comparison and inspection through visualization of matching results. Laboratory experiments found a preference for TAR-EFG over HMCAR-EFG and LeafView. We also found improvements in ease of use and speed of inspection and comparison when using direct manipulation techniques for magnification and tangible gestures for semantic zooming.

Visual Hints techniques (Chapter 6). We developed activation, representation, and presentation techniques to enable new ways of learning and discovering tangible gestures through situated visualizations. We investigated two variations for activating visual hints and seven variations for visualizing them, and then we compared them in a laboratory experiment for preference and comprehension. We found that composite hints, such as ghosting with animation, rank highest in preference and comprehension but that combining all techniques together creates too much clutter. We also found that minimal hints such as text can be helpful as a reminder once a gesture has been learned, although text can be ambiguous in some cases.

Shake Menu techniques (Chapter 7). We created prop-based gestural activation and presentation techniques to provide ready-to-hand access to menu options in 3D as well as radial display of visualizations in close proximity to objects of interest. These techniques can be extended for in-situ placement of objects after selection. In a laboratory experiment, we compared the use of display-, object-, and world-referenced coordinate systems for presenting shake menus, and we found that display- and object-referenced coordinate systems support the fastest selection time and display-referenced supports the lowest error rates.

Georeferenced data visualization techniques (Chapter 8). We created visualization techniques for interaction with, presentation of, and representation of georeferenced data in order to enable visualization of invisible data and to support discovery of new insights in urban environments. A field study of urban site visits with urban designers and planners grounds our techniques in real tasks. In a field experiment, we found that spatializing data improves comparison of data sets from distinct geolocations and that ambiguous representations such as smoke were preferred to more accurate representations such as spheres. We also found that the sensor fusion algorithm we developed, which incorporates distance to data, provides a more stable representation of visualization. In particular, through freezing the video of the scene, it improves stability while users interact with data. It also supports visualization of the larger context of the scene through map overviews that are based on the changing orientation of the device.

More generally, we find that situated visualization techniques bring virtual data and the physical world in closer proximity, support direct manipulation of data in the world, and improve visualization tasks such as inspection, comparison, and pattern seeking for insight. In the next section, we synthesize our experiences developing situated visualizations and suggest principles for designing effective situated visualizations.

9.2 Design Guidelines

A good situated visualization provides direct interaction with the visualization, clear representation of semantic and spatial relevance, and presentation and representation that appropriately embed and combine the virtual with the physical world while supporting exploration of both the physical and the virtual world. In the course of our research, we have developed a set of design principles that can be applied to situated visualizations to ensure that they are useful and effective.

9.2.1 Reflect Both Context of the Visualization and Nature of the Relationship

Two key characteristics of situated visualization are the context of the visualization and the relationship between the physical context and the virtual representation. In some cases, the user has explicitly assigned a context, such as in the various user interfaces designed for identifying objects. In other cases, the context may be unclear. To improve the perception of the visualization together with its context, the visualization should reflect the context and, if possible, the type of relationship, whether semantic or spatial. Even in cases where the context is explicit, the user interface should clearly represent what the system views as the context. For example, in the LeafView interface, we provide feedback on the segmentation of what the computer views as the shape context of the leaf. In the SiteLens interface, we found that confusion could occur when the spatial context was unclear. This can be addressed by improving visual association with context through techniques such as highlighting physical landmarks that are objects of interest or relevant parts of a scene.

9.2.2 Make the Locus of Presentation Appropriate to the Semantic and Spatial Nature of the Data

There exist a wide variety of reference coordinate systems. We have experimented with display-, body-, object-, and world-referenced coordinates as well as hybrids that combine position from one coordinate system and orientation from another. Two key observations for design are related to presentation.

First, the locus of presentation should reflect the spatial nature of the data and the relationship between the data and object of interest. For instance, data that is georeferenced at a scale that can be visualized in the scene should be world-referenced (Figure 8.5a). In cases where the spatial scale larger than the user's view, display-referenced presentation (Figure 8.7a) can provide information without falsely associating the data with a specific viewable location.

Second, we note that spatially- and semantically-driven visualizations differ in their flexibility in terms of visual layout. In the case of spatially-driven visualizations, the layout of data in the visualization is dictated by the existing spatial relationship. Representations must often be presented in specific locations and orientations. In these cases, such as ghosted visual hints (Figure 6.1) or visualized georeferenced data (Figure 8.5a), spatial contiguity between the physical and virtual reinforces visual patterns and association. However, semantically-driven visualizations have much more flexibility and require that the designer take care to create spatial relationships, such as presence or proximity for association. For example, we were able to experiment with different linear (Figure 5.2c) and radial (Figure 7.2) layouts with the TAR-EFG, because we were not constrained by the spatial layout.

9.2.3 Make Representations that Acknowledge the Visual Appearance and Geometry of the "Ground"

In a typical desktop visualization, the background is under the control of the designer. This is generally not the case with situated visualizations, where the visualization is presented in the physical world, and this issue should be addressed in several ways.

First, the visualization should take into account the visual appearance of the background. Transparency can be used to avoid hiding pertinent information. Position of data can respond to the texture of the background for proper placement and layout. Rendering of representations can take into account visual appearance of the context to more closely associate or differentiate the data from the context, depending on the visualization needs. For example, virtual objects such as a sphere can be lit and colored based on the lighting and coloration of physically adjacent structures like buildings in order to more closely associate them with the building.

Second, when mapping data to specific representations, it is important to consider redundant encoding. While not specific to situated visualization, redundant encoding [Nowell 2002], such as the use of both color and elevation in SiteLens, reinforces data mapping in visual scenes (Figure 8.9). We also found this beneficial in composite visual hints (Section 6.2). Careful consideration should be applied here to ensure that multiple encodings do not cancel each other or result in a worse encoding [Perlman 1994].

Third, it is important to keep in mind that some visual cues can be mistaken for depth cues and should not be used when mapping data to visual representations. In contrast to visualization that does not need to take the physical world into account, situated visualizations are often forced into perspective projection to align virtual data with the physical world in the same perceptual space. For instance, our initial SiteLens prototype mapped CO values to sphere size. However, sphere size can also be interpreted as a distance cue (small spheres are far away while large spheres are close). Although it is possible to use unconventional projections [Lorenz 2008], in general, designers should avoid using visual mappings that will be interpreted as specific physical cues.

9.2.4 Be Conscious in the Choice of Mix Between Physical and Virtual Used to Create Figure and Ground

In the beginning of this dissertation, we borrowed the concept of figure-ground to describe visual relationships in situated visualization. We noted that the visualization designer can choose whether the physical context should be considered the figure or ground, or whether the mix of both virtual and physical aspects of the scene can be the figure or ground. The focus of the user's attention is unpredictable, but proper design choice of the figure-ground relationship can help focus the user on important aspects of the visualization and contextual relationships. We provide an example in subsection 2.2.6, in which virtual trees are visualized with physical trees. By changing the visual emphasis of the virtual trees or highlighting physical trees, different combinations of figure-ground are created.

9.2.5 Provide Conceptual Models that Bind the Physical and Virtual

The conceptual model is an important and often overlooked aspect of HCI design. Here, it is important that the conceptual model presented to the user bridge the physical and the virtual to create a mental model that applies to both the physical and virtual aspects of the visualization. For instance, shake menus use a conceptual model in which the menus emerge from the object. In this way, we more closely associate the visualization with the object. By using the conceptual model of menus, we also reflect the possible affordances provided by the visualization. Similarly, Virtual Vouchers provide a conceptual model, based on the metaphor of physical type specimen vouchers, that suggests a connection with physical aspects of the world.

9.2.6 Support Direct Manipulation of the Data in the Context of the Physical World

Many factors influence the efficacy of interaction techniques in situated visualizations. The display modality and task initially constrain types of interaction. For instance, a display device that must be held with two hands will be difficult to use along with tangible manipulation of objects in the scene. We recommend implementing the most direct manipulation possible, given a particular display and task. Beyond this, the locus of presentation and locus of interaction should be aligned. If the data is object-referenced, and the object is within physical reach of the user, the interaction should be object-referenced.

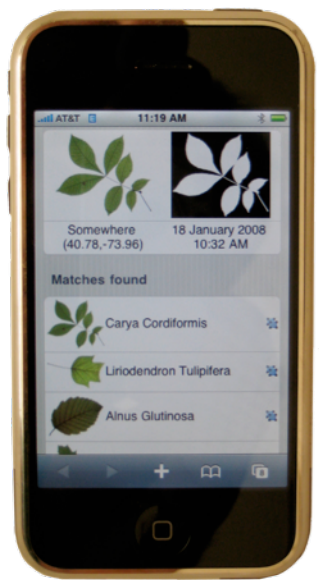
9.2.7 Respect the Physical World, but Break the Rules Consciously

As the old adage suggests, all rules are made to be broken, and these design principles are no exception. For instance, we break the rules in the principle of presentation appropriate to the spatial nature of data in SiteLens by spatializing data from a single location source so that we can compare that data with already georeferenced data. We use color coding to reflect this difference and recommend that situated visualization designers do the same, reflecting to the user that a particular visualization has a different relationship to the context than one might expect.

9.3 Future Work

While this dissertation reflects many years of research, there is still much territory to explore in order to further understand the phenomena surrounding situated visualization, improve user interface techniques, and continue interesting research threads. In this section we discuss work that should be pursued in the future.

9.3.1 Participatory Sensing with the Electronic Field Guide


(a)


(b)

Figure 9.1 Two prototype versions of a client application EFG connecting to a server for processing. (a) Email-based prototype and (b) web-based prototype.

Our LeafView system has met with great interest beyond the research domain. The Smithsonian Institution would like to use the system for censuses of ecological preserves by enabling non-experts to identify plants using the EFG [W. John Kress, personal communication, May 10, 2006]. Non-profit organizations, such as Our City Forest in San Jose, have also contacted the project to use the system for conducting censuses [Matt Jones, personal communication, April 7, 2009]. Educational institutions, such as George Mason University, are interested in using the system for teaching ecology classes [Norm Bourg, personal communication, October 23, 2008]. Beyond making such systems available to a broader audience, there are many interesting research aspects to extending the EFG.

While multiple cameras can be associated with our existing LeafView systems, all acquiring images that are sent to the same device, the prototype still works as an individual system in isolation. We would like to investigate the implications of large-scale participatory sensing [Burke 2006], where groups of users collaboratively collect data in independent devices, as a way of both providing identification functionality to individuals while collecting large, georeferenced data sets of botanical species imagery. Participatory sensing implies that individuals can add to the corpus and share information, thus building the collection. New interfaces and interaction are required to enable this type of sensing.

We are also interested in interfaces that combine dichotomous keys with computer vision to improve identification when computer vision algorithms cannot narrow the selection to a small enough subset of species. Dichotomous keys help identify biological species by presenting the user with a series of choices that differentiate species. By answering a series of questions that represent a hierarchical decision tree, the user eventually comes to a final identification. In addition, we believe that "DNA barcoding" [Savolainen 2005] could potentially improve identification tasks. We believe many of our lessons learned from the LeafView prototypes are applicable here, but such systems will also have new user interface and visualization challenges. For instance, DNA sampling will require a different interface than our computer vision identification technique, but it might still be combined with vision-based registration for overlaying information.

9.3.2 Automating Visual Hints and Authoring Tangible Gestures

We would like to expand our work in visual hints to explore automated authoring of visual hints. We believe that authoring visual hints and associated tangible gestures could be simplified and automated through learning by example. Easy authoring could make visual hints more readily available across systems. We would like to use a Hidden Markov Model based toolkit for gesture recognition (e.g., GT2K [Westeyn 2003]) in order to improve on our own gesture recognition. However, the interesting challenge is in the automated visual representation of visual hints. Once we have a set of examples for a given gesture, we can use one as an example, use spatial or temporal averaging of the set, undersample and interpolate, smooth the signal, or even use canonical movements representative of a single movement. By automating this process, we believe we could include visual hints in many aspects of daily life.

9.3.3 Creating Semantically-driven Visualizations that Use the Scene as Context

In considering the contextual aspects of the design space, we have explored object-focused and semantically-driven visualizations, object-focused and spatially-driven visualizations, and scene-focused and spatially-driven visualizations. However, there remains one quadrant that we have not explored, and to our knowledge, this area has not been explored by others either. Scene-focused and semantically-driven visualizations imply a much greater understanding of an entire scene. The challenge here will be a combination of computer-vision for scene understanding combined with visualization and user interface investigation. Some of our scene-based techniques will likely apply but there is an opportunity here for further exploration.

9.3.3.1 Investigating Perceptual and Cognitive Phenomena

While we have focused on interaction, presentation, and representation techniques in this dissertation, many of the cognitive and perceptual aspects of situated visualization remain unknown. Existing research explores these issues in 3D [Drasic 1996], most often in the context of VR, but the combination of physical and virtual aspects in AR is less understood. Perceptual and cognitive phenemona can inform all aspects of situated visualization research. We note two specific areas that deserve special attention, because they affect the ways in which we understand situated visualizations.

9.3.3.2 Reflecting Virtual-Real Associations and Depth Perception in Situated Visualization

Here, we separate out depth perception from virtual-real association, because the two concepts are distinct, and one may be more important than the other for a given situated visualization. In the case of virtual-real association, we may simply want to know that a virtual object is associated with a physical object. In our research, we have primarily used spatial layout to reflect association, but other cues, such as highlighting context or relevant objects, may increase the perception of association. We distinguish this from depth perception in AR. While a body of work exists in this area, the problem remains unsolved. In particular, visualizations that use the entire scene as context may require that the user comprehend the absolute or relative depth or distance from one data point to another. Previous work has examined color, size, transparency, blur, perspective lines, and visual affordances such as flagpoles to help the user gauge distance [Gabbard 2005, Uratani 2005]. However, none of these have been particularly effective.

9.3.3.3 Presentation, Layout, and Rendering Representations Based on Background

Spatial layout of labels and annotations in AR typically takes into account location and orientation of other objects in the visual field based on models [Bell 2001]. However, we believe a combination of cues that include the underlying content and texture of the scene [Tanaka 2008] together with spatial layout of existing objects will provide a better guide for layout of situated visualization. To achieve this, algorithms for presentation, layout, and even rendering should be guided by perceptual and cognitive principles for combining the physical with the virtual.

9.3.4 Symmetrical Sensing and Visualization

Based on initial usage of the SiteLens prototype, we are interested in increasing the dynamics and symmetry of sensing and visualization by extending the system to live sensor data. In doing this, we want to close the gap between the act of sensing and the act of visualization. One approach to this involves the use of mobile sensors, which can be used for "painting" data as it is sensed, in real time, on the scene. Feedback from our users suggests that this would provide a way for them to further explore unknown regions by guiding them towards areas of interest through their own actions.

9.3.5 Infrastructure for Multiple Situated Visualizations with Discovery and Filtering

In our vision of situated visualization, a user will eventually have easy access to a wealth of information for any given object or scene in the world. This, combined with large data sets, will certainly lead to information overload. To address this, we see a need for algorithms and infrastructure for discovering the presence of hidden information and filtering visualizations so that only the salient and useful information is left. Julier et al. [Julier 2000] have provided some initial work in this area, but we see a need for a comprehensive solution to the problem.

9.4 Closing Remarks

We are indeed surrounded by a sea of information, yet much of it cannot be directly experienced. Visualization provides a powerful tool for extending our perceptual and cognitive abilities beyond their current limits. By applying visualization in the context of the physical world, this dissertation takes steps towards the creation of interaction techniques, artifacts, and a greater understanding of the ways in which we can discover and experience our surroundings through situated visualization.

In his reflections on computer science, Allan Newell describes computer science as "the technology of enchantment" [Newell 1992]. Having observed many people use our situated visualization systems, I have seen the seeds of enchantment, of magic and delight, as they find new ways of viewing the world. While our work is embodied as research prototypes, a future where such experiences are commonplace is not far off. At this writing, new hand-held mobile devices and style-conscious display eyewear that support AR have appeared on the market. My hope and belief is that we have just scratched the surface and that we will eventually match our vision to our imagination.

Appendix A.1: Visual Hints Questionnaire

Note: The questionnaires here are provided to clarify the nature of questions asked during experiments. They have been reformatted to fit the constraints of this dissertation format.

Comparative Study of Interaction Techniques for Visual Hints for Tangible Gestures

Participant ID: ____________
IRB Protocol: IRB-AAAC5545
Principal Investigator: Steven Feiner (skf1)
Co-Investigator: Sean White (sw2061)

User Experience Survey

Date:

Age: Gender: F / M

I use a computer… never
monthly

weekly
daily
multiple times per day

For each question, we would appreciate any additional comments you have in the "Comments" section.

PART I – Activation Study. For the following questions, please circle a number from 1 through 5 to describe your experience using the experimental systems. What did you feel about the different ways to activate visual hints?

Pause:

difficult easy

1	2	3	4	5

frustrating				satisfying
1	2	3	4	5

confusing				intuitive
1	2	3	4	5

Comments:

Shaking:

difficult				easy
1	2	3	4	5

frustrating				satisfying
1	2	3	4	5

confusing				intuitive
1	2	3	4	5

Comments:

PART II – Visual Hints. For the following questions, please circle a number from 1 through 5 to describe your experience using the experimental systems. What did you feel about the following different kinds of visual hints?

None:

difficult				easy
1	2	3	4	5

frustrating				satisfying
1	2	3	4	5

confusing				intuitive
1	2	3	4	5

Comments:

Textual:

difficult					easy
	1	2	3	4	5
frustrating					satisfying
	1	2	3	4	5
confusing					intuitive
	1	2	3	4	5

Comments:

Diagrammatic:

difficult					easy
	1	2	3	4	5
frustrating					satisfying
	1	2	3	4	5
confusing					intuitive
	1	2	3	4	5

Comments:

Ghosting:

difficult					easy
	1	2	3	4	5
frustrating					satisfying
	1	2	3	4	5
confusing					intuitive
	1	2	3	4	5

Comments:

Animated:

difficult				easy
1	2	3	4	5
frustrating				satisfying
1	2	3	4	5
confusing				intuitive
1	2	3	4	5

Comments:

PART III. In the following questions, please place a 1 (best) through 2 (least) next to each choice.

Rank the systems by how intuitive they are, from 1 (most intuitive) to 2 (least intuitive).

___ Pause

___ Shake

Comments:

Rank the systems by your preference for using them, from 1 (most preferred) to 2 (least preferred).

___ Pause

___ Shake

Comments:

Are there other gestures that you would use to activate a visual hint?

PART IV. In the following questions, please place a 1 (best) through 5 (least) next to each choice.

Rank the systems by how intuitive they are, from 1 (most intuitive) to 5 (least intuitive).

___ None

___ Textual

___ Diagrammatic

___ Ghosting

___ Animated

Comments:

Rank the systems by ease of comprehension, from 1 (easiest) to 5 (hardest).

___ None

___ Textual

___ Diagrammatic

___ Ghosting

___ Animated

Comments:

Rank the systems by your preference for using them, from 1 (most preferred) to 5 (least preferred).

___ None

___ Textual

___ Diagrammatic

___ Ghosting

___ Animated

Comments:

Please provide any additional comments about or reactions to any of the techniques:

Appendix A.2: Shake Menus Questionnaire

Comparative Study of User Interface Techniques for 3D Menu Selection

Participant ID: _____________
IRB Protocol: IRB-AAAD6617
Principal Investigator: Steven Feiner (skf1)
Co-Investigator: Sean White (sw2061)

User Experience Survey

Date:

Age: Gender: F / M

I use a computer…

never
monthly
weekly
daily
multiple times per day

For each question, we would appreciate any additional comments you have in the "Comments" section.

PART I – Menu Selection Technique

Screen-Fixed:

difficult				easy
1	2	3	4	5
frustrating				satisfying
1	2	3	4	5
confusing				intuitive
1	2	3	4	5

Comments:

Object-Fixed:

difficult				easy
1	2	3	4	5
frustrating				satisfying
1	2	3	4	5
confusing				intuitive
1	2	3	4	5

Comments:

World-Fixed:

difficult					easy
	1	2	3	4	5
frustrating					satisfying
	1	2	3	4	5
confusing					intuitive
	1	2	3	4	5

Comments:

Clipboard:

difficult					easy
	1	2	3	4	5
frustrating					satisfying
	1	2	3	4	5
confusing					intuitive
	1	2	3	4	5

Comments:

PART II. In the following questions, please place a 1 (best) through 4 (least) next to each choice.

Rank the systems by how intuitive they are, from 1 (most intuitive) to 4 (least intuitive).

___ Screen-Fixed

___ Object-Fixed

___ World-Fixed

___ Clipboard

Comments:

Rank the systems by your preference for using them, from 1 (most preferred) to 4 (least preferred).

___ Screen-Fixed

___ Object-Fixed

___ World-Fixed

___ Clipboard

Comments:

Do you believe that holding an object in your hand helped in the menu selection process or did it interfere? Please explain your answer.

Would you prefer to hold an object in your hand or not hold an object in your hand for selecting menu items? Please explain your answer.

Please provide any additional comments about or reactions to any of the techniques:

Appendix A.3: SiteLens Questionnaire

Comparative Study of Interaction Techniques for Situated Visualization

Participant ID: _____________
IRB Protocol: IRB-AAAD3016
Principal Investigator: Steven Feiner (skf1)
Co-Investigator: Sean White (sw2061)

User Experience Survey

Date:

Age: Gender: F / M

I use a computer…

never
monthly
weekly
daily
multiple times per day

For each question, we would appreciate any additional comments you have in the "Comments" section.

PART I – Representation Study. For the following questions, please circle a number from 1 through 5 to describe your experience using the experimental systems. What did you feel about the different ways to visualize the data value and location of carbon mon-oxide?

Spheres:

	difficult				easy
Data Value	1	2	3	4	5
	confusing				intuitive
	1	2	3	4	5
	difficult				easy
Location	1	2	3	4	5
	confusing				intuitive
	1	2	3	4	5

Comments:

Bars:

	difficult				easy
Data Value	1	2	3	4	5
	confusing				intuitive
	1	2	3	4	5
	difficult				easy
Location	1	2	3	4	5
	confusing				intuitive
	1	2	3	4	5

Comments:

Visible Gas:

	difficult				easy
Data Value	1	2	3	4	5
	confusing				intuitive
	1	2	3	4	5
	difficult				easy
Location	1	2	3	4	5
	confusing				intuitive
	1	2	3	4	5

Comments:

Rank the representations by your preference for using them, from 1 (most preferred) to 3 (least preferred).

___ Spheres

___ Cylinders

___ Visible gas

Comments:

PART II – Pattern Discovery. For the following questions, please compare and contrast your ability to associate visualized data with specific elements of the physical world. Under comments, please report on any new observations or insights made with that particular system.

2D Map:

difficult				easy
1	2	3	4	5
frustrating				satisfying
1	2	3	4	5
confusing				intuitive
1	2	3	4	5

Comments:

Screen-based data levels:

difficult				easy
1	2	3	4	5
frustrating				satisfying
1	2	3	4	5
confusing				intuitive
1	2	3	4	5

Comments:

Visual Data in the Scene:

difficult				easy
1	2	3	4	5
frustrating				satisfying
1	2	3	4	5
confusing				intuitive
1	2	3	4	5

Comments:

Rank the visualization method by preference for using them from 1(mot preferred) to 3 (least preferred).

___ 2D map

___ Screen-based data levels

___ Visual Data in the scene

Please provide any additional comments about or reactions to any of the techniques:

References

[Abowd 1997] Abowd, G.D., Atkeson, C.G., Hong, H., Long, S., Kooper, R., and Pinkerton, M., "Cyberguide: a mobile context-aware tour guide," *Wireless Networks*, 3(5), 1997, pp. 421-433.

[Accot 2002] Accot, J. and Zhai, S., "More than dotting the i's---foundations for crossing-based interfaces," Proc. CHI, April 20-25, 2002, pp. 73-80.

[Agarwal 2006] Agarwal, G., Ling, H., Jacobs, D., Kress, W.J., Russell, R., Belhumeur, P., Feiner, S., Ramamoorthi, R., and White, S., "First steps toward an electronic field guide for plants," *Taxon, Journal of the International Association for Plant Taxonomy*, 55(3), 2006, pp. 597-610.

[Ahlberg 1994] Ahlberg, C. and Shneiderman, B., "Visual information seeking using the FilmFinder," Proc. Conference on Human Factors in Computing Systems, pp. 433-434.

[Apple 2009] Apple. 2009. iPhone Web Page. http://www.apple.com/iphone/ (accessed on January 5, 2009)

[Azuma 2001] Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., and MacIntyre, B., "Recent Advances in Augmented Reality," *IEEE Comp. Graphics and Appl.*, 21(6), 2001, pp. 34-47.

[Azuma 1997] Azuma, R.T., "A survey of augmented reality," *Presence: Teleoperators and Virtual Environments(1054-7460)*, 6(4), 1997, pp. 355-385.

[Bederson 1994] Bederson, B.B. and Hollan, J.D., "Pad++: a zooming graphical interface for exploring alternate interface physics," Proc. Proceedings of the 7th annual ACM symposium on User interface software and technology, Marina del Rey, California, United States, pp. 17-26.

[Bederson 2001] Bederson, B.B., "PhotoMesa: a zoomable image browser using quantum treemaps and bubblemaps," Proc. UIST '01: Proceedings of the 14th annual ACM symposium on User interface software and technology, Orlando, Florida, pp. 71-80.

[Bederson 2002] Bederson, B.B., Shneiderman, B., and Wattenberg, M., "Ordered and quantum treemaps: Making effective use of 2D space to display hierarchies," *ACM Trans. Graph.*, 21(4), 2002, pp. 833-854.

[Bederson 2004] Bederson, B.B., Grosjean, J., and Meyer, J., "Toolkit Design for Interactive Structured Graphics," *IEEE Transactions on Software Engineering*, 30(8), 2004, pp. 535-546.

[Belcher 2003] Belcher, D., Billinghurst, M., Hayes, S., and Stiles, R., "Using Augmented Reality for Visualizing Complex Graphs in Three Dimensions," Proc. Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 03), Oct. 7-10, 2003, pp. 84-93.

[Belhumeur 2008] Belhumeur, P., Chen, D., Feiner, S., Jacobs, D., Kress, W.J., Ling, H., Lopez, I., Ramamoorthi, R., Shirdhonkar, S., White, S., and Zhang, L., "Searching the World's Herbaria: A System for Visual Identification of Plant Species," Proc. European Conference on Computer Vision, Marseille, France, October 12–18, 2008, pp. 116–129.

[Bell 2001] Bell, B., Feiner, S., and Höllerer, T., "View Management for Virtual and Augmented Reality," in *Proc. UIST '01*. Orlando, FL: ACM Press, 2001, pp. 101–110.

[Bell 2002] Bell, B., Hollerer, T., and Feiner, S., "An annotated situation-awareness aid for augmented reality," Proc. UIST '02: Proceedings of the 15th annual ACM symposium on User interface software and technology, pp. 213-216.

[Bertin 1981] Bertin, J., *Graphics and Graphic Information Processing*. Trans. W. J. B. a. P. Scott, Berlin: De Gruyter, 1981.

[Bertin 1983] Bertin, J., *Semiology of Graphics: Diagrams, Networks, Maps*. Trans. W. J. Berg, Madison, WI: University of Wisconsin Press, 1983.

[Beshers 1990] Beshers, C. and Feiner, S., "Worlds within worlds: metaphors for exploring n-dimensional virtual worlds," Proc. UIST '90: Proceedings of the 3rd annual ACM SIGGRAPH symposium on User interface software and technology, Snowbird, Utah, pp. 76-83.

[Bier 1993] Bier, E.A., Stone, M.C., Pier, K., Buxton, W., and DeRose, T., "Toolglass and magic lenses: The see-through interface," Proc. 20th Annual Conference on Computer Graphics and interactive Techniques (SIGGRAPH '93), 1993, pp. 73–80.

[Bier 1994] Bier, E.A., Stone, M.C., Fishkin, K., Buxton, W., and Baudel, T., "A taxonomy of see-through tools," Proc. CHI '94: Proceedings of the SIGCHI conference on Human factors in computing systems, Boston, Massachusetts, pp. 358-364.

[Bowman 2005] Bowman, D., Kruijff, E., Laviola Jr, J.J., and Poupyrev, I., *3D User Interfaces: Theory and Practice*. Boston, MA: Pearson Education, Inc., 2005.

[Bowman 2001] Bowman, D.A. and Wingrave, C.A., "Design and Evaluation of Menu Systems for Immersive Virtual Environments," Proc. VR, pp. 149-156.

[Brewster 2003] Brewster, S., Lumsden, J., Bell, M., Hall, M., and Tasker, S., "Multimodal 'eyes-free' interaction techniques for wearable devices," Proc. CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems, Ft. Lauderdale, Florida, pp. 473-480.

[Buchmann 2004] Buchmann, V., Violich, S., Billinghurst, M., and Cockburn, A., "FingARtips: gesture based direct manipulation in Augmented Reality," *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, 2004, pp. 212-221.

[Burke 2006] Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., and Srivastava, M.B., "Participatory sensing," *ACM Sensys World Sensor Web Workshop*, 2006.

[Butz 2004] Butz, A., Groß, M., and Krüger, A., "TUISTER: a tangible UI for hierarchical structures," Proc. IUI, January 13-16, 2004, pp. 223-225.

[Card 1991] Card, S.K., Mackinlay, J.D., and Robertson, G.G., "A morphological analysis of the design space of input devices," vol. 9: ACM Press New York, NY, USA, 1991, pp. 99-122.

[Card 1997] Card, S.K. and Mackinlay, J., "The structure of the information visualization design space," Proc. IEEE Symposium on Information Visualization (InfoVis '97), Phoenix, AZ, Oct. 20-21, 1997, pp. 92-99.

[Card 1999] Card, S.K., Mackinlay, J.D., and Shneiderman, B., "Information Visualization: Using Vision to Think." San Francisco, CA: Morgan Kaufmann, 1999.

[Chen 2000] Chen, G. and Kotz, D., "A Survey of Context-Aware Mobile Computing Research," Technical Report TR2000-381, Dept. of Computer Science, Dartmouth College, November 2000, 2000.

[Chi 2000] Chi, E.H., "A Taxonomy of Visualization Techniques Using the Data State Reference Model," Proc. IEEE Symposium on Information Visualization (InfoVis 2000), Salt Lake City, UT, Oct. 9-10,2006, pp. 69-75.

[Chuah 1995] Chuah, M.C., Roth, S.F., Mattis, J., and Kolojejchick, J., "SDM: selective dynamic manipulation of visualizations," Proc. Proc. UIST, Pittsburgh, Pennsylvania, United States, pp. 61-70.

[Chung 1992] Chung, J.C., "A comparison of head-tracked and non-head-tracked steering modes in the targeting of radiotherapy treatment beams," Proc. Proceedings of the 1992 symposium on Interactive 3D graphics, Cambridge, Massachusetts, United States pp. 193-196.

[Dachselt 2007] Dachselt, R. and Hübner, A., "Three-dimensional menus: A survey and taxonomy," *Computers & Graphics*, 31(1), 2007, pp. 53-65.

[Dan R. Olsen 2007] Dan R. Olsen, Jr., "Evaluating user interface systems research," Proc. ACM symposium on User interface software and technology, Newport, Rhode Island, USA, pp. 251-258.

[Dey 2001a] Dey, A.K., "Understanding and Using Context," *Personal and Ubiquitous Computing*, 5(1), 2001a, pp. 4-7.

[Dey 2001b] Dey, A.K., Kortuem, G., Morse, D.R., and Schmidt, A., "Situated Interaction and Context-Aware Computing," *Personal and Ubiquitous Computing*, 5(1), 2001b, pp. 1-3.

[DiVerdi 2004] DiVerdi, S., Höllerer, T., and Schreyer, R., " Level of detail interfaces," Proc. Third IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2004) Arlington, VA, Nov. 2-5,2004, pp. 300-301.

[Donelson 1978] Donelson, W.C., "Spatial management of information," Proc. 5th Annual Conference on Computer Graphics and interactive Techniques (SIGGRAPH 78), August 23 - 25, 1978, pp. 203-209.

[Drasic 1996] Drasic, D. and Milgram, P., "Perceptual issues in augmented reality," Proc. SPIE Volume 2653: Stereoscopic Displays and Virtual Reality Systems III, pp. 123-134.

[Eames 1968] Author,"Powers of Ten," Eames Office, 1968, 9 min

[Endsley 1995] Endsley, M.R., "Toward a theory of situation awareness in dynamic systems," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1), 1995, pp. 32-64.

[EPA 2009] EPA. 2009. Carbon Monoxide. http://www.epa.gov/iaq/co.html (accessed on April 20, 2009)

[Feiner 1993a] Feiner, S., MacIntyre, B., Haupt, M., and Solomon, E., "Windows on the world: 2D windows for 3D augmented reality," Proc. Proceedings of the 6th annual ACM symposium on User interface software and technology, Atlanta, Georgia, United States, pp. 145-155.

[Feiner 1993b] Feiner, S., Macintyre, B., and Seligmann, D., "Knowledge-based augmented reality," *Commun. ACM*, 36(7), 1993b, pp. 53-62.

[Feiner 1992] Feiner, S.K. and Seligmann, D.D., "Cutaways and ghosting: satisfying visibility constraints in dynamic 3D illustrations," *The Visual Computer*, 8(5), 1992, pp. 292-302.

[Fiala 2005] Fiala, M., "ARTag, a fiducial marker system using digital techniques," Proc. CVPR 2005, San Diego, CA, pp. 590-596.

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

[Forsberg 2000] Forsberg, A.S., Laidlaw, D.H., van Dam, A., Kirby, R.M., Karniadakis, G.E., and Elion, J.L., "Immersive virtual reality for visualizing flow through an artery," Proc., pp. 457-460.

[Foxlin 1996] Foxlin, E., "Inertial head-tracker sensor fusion by a complementaryseparate-bias Kalman filter," Proc. IEEE Virtual Reality, Santa Clara, CA, March 30 - April 3, 1996, pp. 185-194.

[Friedell 1982] Friedell, M., Barnett, J., and Kramlich, D., "Context-sensitive, graphic presentation of information " *SIGGRAPH Comput. Graph.* , 16(3), 1982, pp. 181-188.

[Fuhrmann 1998a] Fuhrmann, A., Loffelmann, H., Schmalstieg, D., and Gervautz, M., "Collaborative visualization in augmented reality," *Computer Graphics and Applications, IEEE*, 18(4), 1998a, pp. 54-59.

[Fuhrmann 1998b] Fuhrmann, A., Schmalstieg, D., and Gervautz, M., "Strolling Through CyberspaceWith Your Hands In Your Pockets:Head Directed Navigation In Virtual Environments," Proc. Proceedings of the 4th EUROGRAPHICS Workshop on Virtual Environments, pp. 216-227.

[Furnas 1986] Furnas, G.W., "Generalized fisheye views," Proc. CHI '86: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 16-23.

[Gabbard 2005] Gabbard, J.L., Swan, J.E., Hix, D., Schulman, R.S., Lucas, J., and Gupta, D., "An empirical user-based study of text drawing styles and outdoor background textures for augmented reality.," Proc. Proc. IEEE Virtual Reality, March 12-16, 2005, pp. 11-18.

[Gibson 1986] Gibson, J.J., *The Ecological Approach to Visual Perception*. Lawrence Erlbaum Associates, 1986.

[Gillet 2004] Gillet, A., Sanner, M., Stoffler, D., Goodsell, D., and Olson, A., "Augmented Reality with Tangible Auto-Fabricated Models for Molecular Biology Applications," Proc. Visualization '04, October 10-15, 2004, pp. 235-242.

[Gillet 2006] Gillet, A. and Olson, A., "Tangible Augmented Interfaces for Structural Molecular Biology," Proc. ISMAR, Santa Barbara, CA, pp. poster.

[Grosjean 2002] Grosjean, J., Burkhardt, J.M., Coquillart, S., and Richard, P., "Evaluation of the Command and Control Cube," Proc. IEEE Multimodal Interfaces, pp. 473-478.

[Grossman 2004] Grossman, T. and Balakrishnan, R., "Pointing at trivariate targets in 3D environments," Proc. Proceedings of the SIGCHI conference on Human factors in computing systems, Vienna, Austria, April 24-29, 2004, pp. 447-454.

[Gurwitz 1981] Gurwitz, R.F., Fleming, R.T., and Van Dam, A., "MIDAS: A Microprocessor Display and Animation System," *IEEE Transactions on Education*, E-24(2), 1981, pp. 126-133.

[Guven 2006] Guven, S., Feiner, S., and Oda, O., "Mobile augmented reality interaction techniques for authoring situated media on-site," Proc. Mixed and Augmented Reality, 2006. ISMAR 2006. IEEE/ACM International Symposium on, pp. 235-236.

[Hamming 1962] Hamming, R.W., *Numerical methods for scientists and engineers*. New York, NY: McGraw-Hill Book, xvii, 411 p, 1962.

[Hanumara 2008] Hanumara, P. and Coyle, L., "Connecting Families by Sharing the Minutiae of their Lives," Report UCD-CSI-2008-05 University College Dublin, Dublin, Ireland 2008.

[Haraway 1991] Haraway, D.J., *Simians, Cyborgs, and Women: The Reinvention of Nature*. Routledge, 1991.

[Heer 2005] Heer, J., Card, S.K., and Landay, J.A., "prefuse: a toolkit for interactive information visualization," Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI 05), Portland, OR, April 2-7, 2005, pp. 421-430.

[Heer 2006] Heer, J. and Agrawala, M., "Software Design Patterns for Information Visualization," *Visualization and Computer Graphics, IEEE Transactions on*, 12(5), 2006, pp. 853-860.

[Heidegger 1962] Heidegger, M., *Being and Time*. New York: Harper & Row, 1962.

[Herot 1980] Herot, C.F., "Spatial management of data," *ACM Trans. Database Syst.* , 5(4), 1980, pp. 493-513.

[Hix 1995] Hix, D., Templeman, J.N., and Jacob, R.J.K., "Pre-Screen Projection: From Concept to Testing of a New Interaction Technique," Proc. Proceedings of ACM CHI'95 Conference, Denver, 1995, pp. 226 - 233.

[Hoffman 1998] Hoffman, H.G., "Physically touching virtual objects using tactile augmentation enhances the realism of virtual environments," *Proceedings of the IEEE Virtual Reality Annual International Symposium*, 98, 1998, pp. 59-63.

[Hollerer 1999] Hollerer, T., Feiner, S., and Pavlik, J., "Situated Documentaries: Embedding Multimedia Presentations in the Real World," *Proc. ISWC'99 (Third Int. Symp. on Wearable Computers)*, 1999, pp. 79–86.

[Hopkins 1991] Hopkins, D., "The design and implementation of pie menus," *Dr. Dobb's Journal*, 16(12), 1991, pp. 16-26.

[Hua 2000] Hua, H., Girardot, A., Gao, C., and Rolland, J.P., "Engineering of head-mounted projective displays," *Applied Optics*, 39(22), 2000, pp. 3814-3824.

[Inselberg 1990] Inselberg, A. and Dimsdale, B., "Parallel coordinates: a tool for visualizing multi-dimensionalgeometry," Proc. First IEEE Conference on Visualization, October 23-26, 1990, pp. 361-378.

[Ishii 1997] Ishii, H. and Ullmer, B., "Tangible Bits: Towards Seamless Interfaces between People, Bits, and Atoms," Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI 97), Atlanta, GA, Mar. 22-27, 1997, pp. 234-241.

[Jaccard 1996] Jaccard, J. and Wan, C.K., *LISREL approaches to interaction effects in multiple regression*. Thousand Oaks, CA: Sage Publications, 1996.

[Johnson 2006] Johnson, C., Moorhjead, R., Munzner, T., Pfister, H., Rheingans, P., and Yoo, T.S., "NIH/NSF Visualization Research Challenges," IEEE Computer Society, 2006.

[Julier 2000] Julier, S., Lanzagorta, M., Baillot, Y., Rosenblum, L., Feiner, S., Hollerer, T., and Sestito, S., "Information filtering for mobile augmented reality," Proc. Augmented Reality, 2000. (ISAR 2000). Proceedings. IEEE and ACM International Symposium on, pp. 3-11.

[Kalkofen 2007] Kalkofen, D., Mendez, E., and Schmalstieg, D., "Interactive Focus and Context Visualization for Augmented Reality," Proc. ISMAR 2007, Nara, Japan, pp. 191-200.

[Kato 1999] Kato, H. and Billinghurst, M., "Marker tracking and HMD calibration for a video-based augmented reality conferencing system," Proc. 2nd IEEE and ACM International Workshop on Augmented Reality pp. 85-94.

[Kato 2000] Kato, H., Billinghurst, M., Poupyrev, I., Imamoto, K., and Tachibana, K., "Virtual object manipulation on a table-top AR environment," Proc. IEEE and ACM International Symposium on Augmented Reality (ISAR 2000), Munich, Germany, Oct. 5-6, 2000, pp. 111-119.

[Kawashima 2001] Kawashima, T., Imamoto, K., Kato, H., Tachibana, K., and Billinghurst, M., "Magic Paddle: A Tangible Augmented Reality Interface for Object Manipulation," Proc. ISMR, pp. 194-195.

[King 2005] King, G.R., Piekarski, W., and Thomas, B.H., "ARVino—Outdoor augmented reality visualisation of viticulture GIS data," Proc. Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 05), Oct. 5-8, 2005, pp. 52-55.

[Kjeldsen 1996] Kjeldsen, R. and Kender, J., "Toward the use of gesture in traditional user interfaces," Proc. Automatic Face and Gesture Recognition, Killington, VT, pp. 151-156.

[Klein 2007] Klein, G. and Murray, D., "Parallel tracking and mapping for small AR workspaces," Proc. ISMAR 2007, November 13-16, 2007, pp. 225-234.

[KML 2009] KML. 2009. KML. http://www.opengeospatial.org/standards/kml/ (accessed on March 2, 2009)

[Koller 1996] Koller, D.R., Mine, M.R., and Hudson, S.E., "Head-tracked orbital viewing: an interaction technique for immersive virtual environments " Proc. Proceedings of the 9th annual ACM symposium on User interface software and technology, Seattle, Washington, United States pp. 81-82.

[Krasner 1988] Krasner, G.E. and Pope, S.T., "A Description of the Model-View-Controller User Interface Paradigm in the Smalltalk-80 System," *J. Object-Oriented Programming*, 1(3), 1988, pp. 26-49.

[Krum 2001] Krum, D.M., Ribarsky, W., Shaw, C.D., Hodges, L.F., and Faust, N., "Situational visualization," Proc. ACM symposium on Virtual reality software and technology, Baniff, Alberta, Canada, Novermber 15-17, 2001, pp. 143-150.

[Kurtenbach 1991] Kurtenbach, G. and Buxton, W., "Issues in combining marking and direct manipulation techniques," Proc. UIST, pp. 137-144.

[Lamping 1996] Lamping, J. and Rao, R., "The Hyperbolic Browser: A Focus+ Context Technique for Visualizing Large Hierarchies," Proc. UIST 1996, pp. 33-55.

[Laurel 2003] Laurel, B., *Design Research: Methods and Perspectives*. MIT Press, 2003.

[Lave 1991] Lave, J. and Wenger, E., *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press, 1991.

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

[Liebenberg 1999] Liebenberg, L., Steventon, L., Benadie, K., and Minye, J., "Rhino tracking with the CyberTracker field computer," *Pachyderm*, 27(January-December), 1999, pp. 59.

[Lindeman 1999] Lindeman, R.W., Sibert, J.L., and Hahn, J.K., "Towards usable VR: an empirical study of user interfaces for immersive virtual environments," *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, 1999, pp. 64-71.

[Ling 2005] Ling, H. and Jacobs, D.W., "Using the Inner-Distance for Classification of Articulated Shapes," Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 719-726.

[Looser 2004] Looser, J., Billinghurst, M., and Cockburn, A., "Through the looking glass: the use of lenses as an interface tool for Augmented Reality interfaces," Proc. 2nd international Conference on Computer Graphics and interactive Techniques in Australasia and South East Asia (GRAPHITE '04), Singapore, June 15-18, 2004, pp. 204-211.

[Lorenz 2008] Lorenz, H., Trapp, M., Döllner, J., and Jobst, M., "Interactive Multi-Perspective Views of Virtual 3D Landscape and City Models," in *The European Information Society*, 2008, pp. 301-321.

[Mayer 2001] Mayer, R.E., *Multimedia Learning*. Cambridge University Press, 2001.

[McCloud 1994] McCloud, S., *Understanding Comics: The Invisible Art*. New York: Harper Collins, 1994.

[McGrath 1995] McGrath, J.E., "Methodology matters: doing research in the behavioral and social sciences," in *Human-computer interaction: toward the year 2000*: Morgan Kaufmann Publishers Inc., 1995, pp. 152-169.

[Milgram 1994] Milgram, P. and Kishino, F., "A Taxonomy of Mixed Reality Visual Displays," *IEICE Transactions on Information Systems*, 77, 1994, pp. 1321-1329.

[Mine 1997a] Mine, M.R., "Exploiting Proprioception in Virtual-Environment Interaction.," PhD diss., UNC Chapel Hill, 1997a.

[Mine 1997b] Mine, M.R., Brooks Jr, F.P., and Sequin, C.H., "Moving objects in space: exploiting proprioception in virtual-environment interaction," Proc. Siggraph pp. 19-26.

[Mullet 1995] Mullet, K. and Sano, D., *Designing visual interfaces: Communication oriented techniques*. Englewood Cliffs, NJ: Prentice Hall, 1995.

[Newell 1992] Newell, A., "Fairy Tales," in *AI Magazine*, vol. 13, 1992, pp. 46-48.

[North 2006] North, C., "Toward measuring visualization insight," *Computer Graphics and Applications, IEEE*, 26(3), 2006, pp. 6-9.

[Nowell 2002] Nowell, L., Schulman, R., and Hix, D., "Graphical encoding for information visualization: an empirical study," Proc. IEEE Symposium on Information Visualization (INFOVIS), 2002, pp. 43-50.

[Oda 2008] Oda, O. 2008. GoblinXNAv2. http://www1.cs.columbia.edu/graphics/projects/goblin/ (accessed on

[Olwal 2008] Olwal, A., Feiner, S., and Heyman, S., "Rubbing and Tapping for Precise and Rapid Selection on Touch-Screen Displays," Proc. CHI 2008, Florence, Italy, April 5-10, 2008, pp. 295-304.

[OSG 2009] OSG. 2009. Open Geospatial Consortium. http://www.opengeospatial.org/ (accessed on March 20, 2009)

[Paulos 2007] Paulos, E., Honicky, R.J., and Goodman, E., "Sensing Atmosphere," *Workshop on Sensing on Everyday Mobile Phones in Support of Participatory Research*, 2007.

[Perlin 1993] Perlin, K. and Fox, D., "Pad: an alternative approach to the computer interface," Proc. 20th Annual Conference on Computer Graphics and interactive Techniques (SIGGRAPH 93), Marina del Rey, CA, pp. 57-64.

[Perlman 1994] Perlman, G. and Swan, J.E., "Relative Effects of Color, Texture, and Density Coding on Visual Search Performance and Subjective Preference," Proc. 38th Annual Meeting of the Human Factors and Ergonomics Society (HFES), pp. 235–239.

[Piekarski 2002] Piekarski, W. and Thomas, B.H., "Tinmith-Hand: Unified User Interface Technology for Mobile Outdoor Augmented Reality and Indoor Virtual Reality," Proc. VR, pp. 287-288.

[Pierce 1999] Pierce, J.S., Conway, M., Dantzich, M.v., and Robertson, G., "Toolspaces and glances: storing, accessing, and retrieving objects in 3D desktop applications," Proc. SI3D '99: Proceedings of the 1999 symposium on Interactive 3D graphics, pp. 163-168.

[Plaisant 2004] Plaisant, C., "The challenge of information visualization evaluation," Proc. Advanced Visual Interfaces, Gallipoli, Italy, pp. 109-116.

[Poupyrev 2001] Poupyrev, I., Tan, D., Billinghurst, M., Kato, H., Regenbrecht, H., and Tetsutani, N., "Tiles: A Mixed Reality Authoring Interface," Proc. INTERACT 2001, pp. 334-341.

[Quek 2002] Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X., Kirbas, C., McCullough, K.E., and Ansari, R., "Multimodal Human Discourse: Gesture and Speech," *ACM Trans. Comput.-Hum. Interact.* , 9(3), 2002, pp. 171-193.

[Rauhala 2006] Rauhala, M., Gunnarsson, A.S., and Henrysson, A., "A novel interface to sensor networks using handheld augmented reality," Proc. 8th Conference on Human-Computer interaction with Mobile Devices and Services, pp. 145-148.

[Reitmayr 2004] Reitmayr, G. and Schmalstieg, D., "Scalable Techniques for Collaborative Outdoor Augmented Reality," *3rd IEEE and ACM international symposium on mixed and augmented reality (ISMAR'04), Arlington*, 2004.

[Reitmayr 2006] Reitmayr, G. and Drummond, T., "Going out: Robust model-based tracking for outdoor augmented reality," Proc. IEEE/ACM International Symposium on Mixed and Augmented Reality, October 22-25, 2006, pp. 109-118.

[Rekimoto 1995] Rekimoto, J., "The magnifying glass approach to augmented reality systems," Proc. International Conference on Artificial Reality and Tele-Existence '95 / Conference on Virtual Reality Software and Technology '95 (ICAT/VRST'95), November 1995, pp. 123-132.

[Ryan 1999] Ryan, N. and Pascoe, J., "FieldNote: a Handheld information System for the Field," Proc. Workshop on TeloGeoProcessing, Lyon.

[Samet 1988] Samet, H., "An overview of quadtrees, octrees, and related hierarchical data structures," *NATO ASI Series*, 40, 1988, pp. 51-68.

[Savolainen 2005] Savolainen, V., Cowan, R.S., Vogler, A.P., Roderick, G.K., and Lane, R., "Towards writing the encyclopaedia of life: an introduction to DNA barcoding," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1462), 2005, pp. 1805-1811.

[Schall 2009] Schall, G., Mendez, E., Kruijff, E., Veas, E., Junghanns, S., Reitinger, B., and Schmalstieg, D., "Handheld Augmented Reality for Underground Infrastructure Visualization," *Journal of Personal and Ubiquitous Computing*, 13(4), 2009, pp. 281-291.

[Schilit 1994] Schilit, B., Adams, N., and Want, R., "Context-aware computing applications," Proc. 1st International Workshop on Mobile Computing Systems and Applications, pp. 85-90.

[Schmandt 1998] Schmandt, C., "Audio hallway: a virtual acoustic environment for browsing," Proc. Proceedings of the 11th annual ACM symposium on User interface software and technology, San Francisco, CA, pp. 163-170.

[Schmidt 2000] Schmidt, A., Van de Velde, W., and Kortuem, G., "Situated interaction in ubiquitous computing," Proc. Workshop Proposal for Conference on Human Factors in Computer Systems, pp. 374-374.

[Shneiderman 1996] Shneiderman, B., "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations," Proc. IEEE Symposium on Visual Languages, Boulder, CO, Sept. 3-6,1996, pp. 336-343.

[Silberschatz 2004] Silberschatz, A., Galvin, P.B., and Gagne, G., *Operating System Concepts*. John Wiley & Sons, Inc., 2004.

[Sinclair 2002] Sinclair, P., Martinez, K., Millard, D.E., and Weal, M.J., "Links in the palm of your hand: tangible hypermedia using augmented reality," Proc. HyperText and HyperMedia, pp. 127-136.

[Slay 2001] Slay, H., Phillips, M., Vernik, R., and Thomas, B., "Interaction modes for augmented reality visualization," Proc. Australian symposium on Information visualisation, Sydney, Australia pp. 71-75.

[Slay 2002] Slay, H., Thomas, B., and Vernik, R., "Tangible User Interaction Using Augmented Reality," Proc. Australasian Conference on User interfaces, Melbourne, Victoria, Australia, pp. 13-20.

[Snyder 1996] Snyder, G., "Walking the New York Bedrock Alive in the Sea of Information," in *Mountains and Rivers Without End*. Washington, D.C.: Counterpoint, 1996, pp. 99-101.

[Sutherland 1963] Sutherland, I.E., "Sketchpad-A Man-Machine Graphical Communication System," Proc. Spring Joint Computer Conference, Washington, D.C., May 21-23, 1963, pp. 329-346.

[Sutherland 1968] Sutherland, I.E., "A head-mounted three dimensional display," Proc. AFIPS Conference, San Francisco, California, December 9-11, 1968, pp. 757-764.

[Szalavari 1997] Szalavari, Z. and Gervautz, M., "The Personal Interaction Panel-a Two-Handed Interface for Augmented Reality," *Computer Graphics Forum*, 16, 1997, pp. 335-346.

[Tanaka 2008] Tanaka, K., Kishino, Y., Miyamae, M., Terada, T., and Nishio, S., "An information layout method for an optical see-through head mounted display focusing on the viewability," Proc. ISMAR, September 15-18, 2008, pp. 139-142.

[Tovee 1996] Tovee, M.J., *Introduction to the Visual System*. Cambridge, UK: Cambridge University Press, 1996.

[Tukey 1977] Tukey, J., *Exploratory Data Analysis*. Addison-Wesley, 1977.

[Tversky 2002] Tversky, B., Morrison, J.B., and Betrancourt, M., "Animation: can it facilitate?," in *Int. J. Human-Computer Studies*, vol. 57, 2002, pp. 247-262.

[Ullmer 2000] Ullmer, B. and Ishii, H., "Emerging Frameworks for Tangible User Interfaces," *IBM Systems Journal*, 39(3), 2000, pp. 915-931.

[Underkoffler 1999] Underkoffler, J., Ullmer, B., and Ishii, H., "Emancipated pixels: real-world graphics in the luminous room," *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, 1999, pp. 385-392.

[Uratani 2005] Uratani, K., Machida, T., Kiyokawa, K., and Takemura, H., "A study of depth visualization techniques for virtual annotations in augmented reality," Proc. Virtual Reality, 2005. , March 12-16, 2005 pp. 295-296.

[Vickers 1972] Vickers, D.L., "The Sorcerer's Apprentice: Head-Mounted Display and Wand," Ph.D. Dissertation, University of Utah, 1972.

[Viega 1996] Viega, J., Conway, M.J., Williams, G., and Pausch, R., "3D Magic Lenses," Proc. UIST '96, Seattle WA, pp. pp. 51−58.

[Ware 2004] Ware, C., *Information Visualization: Perception for Design*. Morgan Kaufmann, 2004.

[Westeyn 2003] Westeyn, T., Brashear, H., Atrash, A., and Starner, T., "Georgia tech gesture toolkit: supporting experiments in gesture recognition," Proc. Proceedings of the 5th international conference on Multimodal interfaces, Vancouver, British Columbia, Canada, November 5-7, 2003, pp. 85-92.

[White 2006a] White, S. and Feiner, S., "Demo: Augmented Reality User Interfaces to an Electronic Field Guide," Proc. ISMAR 2006, Santa Barbara, CA.

[White 2006b] White, S., Feiner, S., and Kopylec, J., "Virtual Vouchers: Prototyping a Mobile Augmented Reality User Interface for Botanical Species Identification," Proc. IEEE 3DUI 2006, Alexandria, VA, March 25-26, 2006, pp. 119-126.

[White 2006c] White, S., Marino, D., and Feiner, S., "LeafView: A User Interface for Automated Botanical Species Identification and Data Collection," Proc. UIST 2006 Adjunct Proceedings, Montreaux, Switzerland, pp. 101-102.

[White 2007a] White, S., Lister, L., and Feiner, S., "Visual Hints for Tangible Gestures in Augmented Reality," Proc. IEEE ISMAR 2007, Nara, Japan, October 12-14, 2007, pp. 47-50.

[White 2007b] White, S., Marino, D., and Feiner, S., "Designing a Mobile User Interface for Automated Species Identification," Proc. ACM CHI 2007, San Jose, CA, April 28-May 3, 2007, pp. 291-295.

[White 2007c] White, S., Morozov, P., and Feiner, S., "Imaging for Insight: Site Visit by Situated Visualization," Proc. ACM CHI 2007 Workshop on Imaging the City, San Jose, CA, April 28-May 3, 2007.

[White 2007d] White, S., Morozov, P., and Feiner, S., Site Visit by Situated Visualization in Calabrese, F. et al., "Urban Computing and Mobile Devices," IEEE Pervasive Computing, vol.6, no.3, July–September 2007, pp. 52-57.

[White 2008] White, S., Morozov, P., and Feiner, S., "Progress Towards Site Visit by Situated Visualization," Proc. CHI 2008 Workshop: Urban Mixed Reality, Florence, Italy, April 5-10, 2008.

[White 2009a] White, S. and Feiner, S., "SiteLens: situated visualization techniques for urban site visits," Proc. 27th international conference on Human factors in computing systems (CHI), Boston, MA, USA, April 4-9, 2009, pp. 1117-1120.

[White 2009b] White, S., Feng, D., and Feiner, S., "Poster: Shake Menus: Activation and Placement Techniques for Prop-Based 3D Graphical Menus," Proc. IEEE 3DUI, Lafayette, LA, March 14-15, 2009, pp. to appear in.

[Yeh 2006] Yeh, R., Liao, C., Klemmer, S., Guimbretière, F., Lee, B., Kakaradov, B., Stamberger, J., and Paepcke, A., "ButterflyNet: a mobile capture and access system for field biology research," Proc. SIGCHI, Montréal, Québec, Canada, pp. 571-580.

[You 2001] You, S. and Neumann, U., "Fusion of vision and gyro tracking for robust augmented reality registration," Proc. IEEE Virtual Reality, 2001, March 15-17, 2001, pp. 71-78.