PhD Thesis

# Augmented Reality Interfaces for Mobile Environmental Monitoring

Eduardo E. Veas

Graz University of Technology
Institute for Computer Graphics and Vision

*Thesis supervisors*
Prof. Dr. Dieter Schmalstieg
Dr. Simon Julier

Graz, April 2012

To my families

---

*to Analía, my house*

## Statutory Declaration

*I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.*

| | | |
|---|---|---|
| Place | Date | Signature |

## Eidesstattliche Erklärung

*Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe.*

| | | |
|---|---|---|
| Ort | Datum | Unterschrift |

# Acknowledgement

This thesis would not have been completed, were it not for the support, collaboration and insightful discussion with many great people.

My gratitude goes to Dieter Schmalstieg for shedding light in every twisted idea I could come up with. But mostly for his enabling ability, for creating and letting me be part of this think-tank that is the graphics team in the Computer Graphics and Vision group, for facilitating the means to let us create, play, and discover, to probe and learn, to do research. I would also like to thank Simon Julier for accepting the role of advisor, and for all the discussions that raised my *awareness* about many topics I had overlooked.

The wonderful group of people at the Institute for Computer Graphics and Vision have made these 5 years an unforgettable, rich experience. I will always be grateful for their friendly fellowship, and for giving me the opportunity to learn from/with them. In particular, I would like to thank my friend Ernst Kruijff for countless hours of sheer creativity. Besides introducing me to user interface design, Ernst shared with me most enjoyable activities: sketching, skiing, swimming, designing and good music. I would also like to thank my friend Erick Mendez for hours of sitting in front of a monitor, developing and debugging openGL code. Among the many things I learned from Erick, he showed me how to think about visualization, and how to focus on the thing I need to develop and not on all its possible uses and nuisances. A badly learned lesson from my part. I would like to thank Alessandro Mulloni for enlightening discussions where I learned how to prepare and plan experiments to the detail. Thanks to Raphael Grasset for his French-ness. Raphael is a person with a gifted humor, who after coding and debugging continuously for over 24 hs, can smile and make you laugh. Raphael, the master, who took my framework and pulled it inside out in a couple of months, also taught me, and keeps trying to teach me how to think about experiments. I also want to thank Joe Newman for his friendship, even when our lives seem to take us in opposite directions. Thanks to Tobias and Steffi for reminding me of past lives, for tasty okonomiyaki and never ending support. I would further like to thank Gerhard Reitmayr and Denis Kalkofen for lending me time to write up this thesis. My thanks go to Markus Tatzgern and Lukas Gruber, without whom this dissertation would not have a proper german abstract. And I would not want to forget the staff at ICG that made my life easier all the way, thank you Christina, Renate, Manuela and Karin and our system staff Andy, Mark and Albert. Albert seems to always have another hard drive in his pocket waiting when I break mine. I want to extend my thanks to all the ICG group, who shaped me as a professional.

Further, I would like to thank Thomas Mauthner and Amir Saffari for starting the football sessions, to Alessandro for understanding funny playing a la italiana. I also would like to thank Alexander Lex for making me run again, and Manuela Waldner for her skiing tips.

I dedicate this thesis to my families. My father, Enrique, was patient enough and allowed me to learn and explore for myself, but was always there when I needed him. He taught me math and science. My mother Mary was tenacious and taught me to write properly. Neither of them taught me how to give up. My brother, Claudio, my best friend always, put up with me through the worse of times, and taught me how to play together.

Lastly, and most important of all, I would like to thank my wife, Analía for choosing to take the strangest adventures with me, for cheering and taking over when I have a deadline and for caring if I didn't sleep. Most of all, thank you for making your life where I am, for not letting me loose focus, and for believing in me. Thank you also to the best researchers I have known, Sarah and Agustín, my children. They have shown me what it is to systematically investigate an unpredictable environment, to probe, to fail, and to keep trying, they have shown me tenacity and audacity in this short three years of their life. I hope I can keep learning from you for a long time.

# Preface and Motivation

Every decade has its technological revolution that seems to change the way we experience the world. In the past, the introduction of the internet put a virtually limitless amount of information at the disposal of any individual sitting at a terminal. Recently, we started to let digital information follow us everywhere. Remarkable developments in mobile technology have vanished the boundary that the terminal represented. We can reach (and be reached by) information anywhere, but we are still lost in the tools of the past. We are yet to learn ways to use our locality to dive into the information connected to our surroundings. In scientific visualization, it is a recognized challenge that we need novel ways to examine and make sense of massive amounts of information, in any place, using every kind of display available. This dissertation takes a step to face this challenge.

This thesis concentrates on mobile scientific data visualization in the frame of environmental monitoring. It uncovers how data from a myriad of sensors can be presented to the observer in real-time, in direct relation with the surrounding environment. Augmented reality (AR), a paradigm that merges digital information with the real world, presents a unique opportunity to experience the real world and its associated data as a seamless information space. To make the observer aware of environmental data in the world, this dissertation looks at how to represent multi-dimensional data in the real world captured by a camera. It addresses issues ranging from reducing interference, to coherently conveying spatial properties such as shape and topology, and filtering, to avoid clutter and information overload. It also introduces novel view management metaphors that extend overview of traditional AR, and provide vantage zooming points in the real world for the visualization workflow. Finally, to fuel the mobile experience, this thesis analyses pose and interaction patterns of handheld AR in the frame of ergonomic devices.

In summary, this dissertation presents a compendium of studies and techniques for mobile environmental data visualization. Still, mobile AR visualization and the tools presented here are not restricted to environmental data. They will hopefully find their way into the mainstream mobile toolset.

# Kurzfassung

Diese Dissertation konzentriert sich auf mobile Visualisierung von wissenschaftlichen Daten im Rahmen der Umweltüberwachung. Sie zeigt wie Daten aus einer Vielzahl von Sensoren in Echtzeit in einem direkten Zusammenhang mit der Umgebung präsentiert werden können. Augmented Reality (AR) verschmilzt digitale Information mit der realen Welt und stellt damit eine einzigartige Gelegenheit dar, die reale Welt und die damit verbundenen Daten als nahtlosen Informationsraum zu erleben.

Diese Dissertation analysiert, wie man einem Benutzer multi-dimensionale Umweltdaten in einem Videobild, das die echte Welt zeigt, näher bringen kann. Sie behandelt Fragen im Bereich der Reduzierung visueller Interferenzen, der koherenten Vermittlung von räumlichen Eigenschaften wie Form und Topologie, und der Datenfilterung, um visuelle und mentale Übersättigung zu vermeiden. Ausserdem führt sie neue View Management Metaphern ein, die es erlauben Überblicksvisualisierungen fuer traditionellen AR-Anwendungen zu erzeugen, und damit auch andere Blickpunkte zu erreichen. Die Arbeit analysiert Körperhaltung und Interaktionssmuster für räumliche Interaktion in Handheld-AR im Zusammenhang mit ergonomischen Geraeten.

Diese Dissertation stellt ein Kompendium von Studien und Techniken für mobile Umweltdatenvisualisierung dar. Dennoch sind mobile AR-Visualisierung und die hier vorgestellten Tools nicht auf Umweltüberwachungsdaten beschränkt. Sie werden hoffentlich ihren Weg in den Mainstream mobiler Toolset finden.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Since the beginning of time, the environment has been in constant change. However, only recent technological advances enable us to analyze and study environmental change as it happens, and to act upon it to prevent damage and loss to people. Growing consideration about global warming and the natural environment has recently accelerated the need for more accurate, responsive and pervasive environmental monitoring [Hilbich et al., 2008a]. Supported by progress in wireless networking and sensor miniaturization, pervasive sensor networks for environmental monitoring nowadays play a major role in gathering data from the environment and ecosystems [Bogue, 2008] [Hart and Martinez, 2006]. Still, most monitoring activities demand physical presence at the site. In particular, personal observation is fundamental in judging and understanding the current situation, as well as in communicating with stakeholders.

In spite of the evident need for ubiquitous up-to-date information, on-site visualization and interactive tools for the exploration of environmental data are still really inadequate for domain experts. Current solutions are generally limited to tabular data, basic 2D plot visualization, or restraint to standard 2D GIS tools designed for the desktop and not adapted to mobile use. These tools often have restricted connectivity and limited interoperability. In that context, the next challenges in environmental monitoring comprise the ability to capture spatial and temporal variability of environmental parameters, to develop real-time image and data transmission [Shin et al., 2007], and data visualization in order to enhance environmental models, prediction tools, and decision making [Bogue, 2008].

This thesis addresses the topic of mobile visualization of environmental data. It uncovers how data from a myriad sensors can be presented to the observer in an intuitive manner, in the context of the observed environment. To that end, a novel augmented reality platform is introduced that, based on thorough perceptual and ergonomic studies, leverages the visualization of geo-referenced sensor measurement and simulation data in a seamless integrated view of the environment.

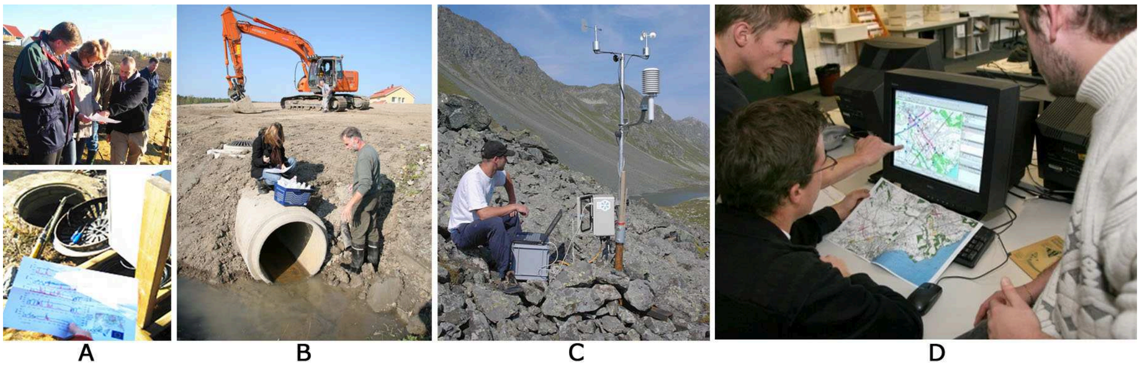


**Figure 1.1:** Environmental monitoring activities. A: analyzing and comparing observations with measured data (paper plots). B: manually measuring water quality and noting down observations. C: downloading data and performing maintenance on sensor station. D: analyzing correspondences between observed and measured data.

## 1.1 Environmental Monitoring

Environmental monitoring is the method used by geoscientists to study the development of environmental processes. It involves continuous observation and regular measurement of environmental parameters of a specific area, in order to identify and understand environmental changes, and possibly aid the decision making process related to the site.

In this context, the workflow of a geoscientist alternates visits to the field with work at the office. Site visits serve the purpose to document the visual appearance of the environment, to gather samples and personal observations, or to deploy/maintain infrastructure (see Figure 1.1, A – C). At the office, the geoscientist applies scientific visualization to compare results of physical models with reality (see Figure 1.1 D). This process highly depends on the availability of resources and tools: while on-site, the researcher rarely has access to all sensor readings or numerical analysis tools (closed network, complexity of simulation tools), let alone visualization possibilities. These data and tools are therefore left to traditional domain of office applications.

One drawback of this approach is the disconnection during analysis of the data from the observed environmental context (the site). Similarly, the digital representation of the site under consideration is most often not timely and/or spatially as accurate as the physical observation. Thereby, we can identify a real gap between the environment as observed on-site and its digital representation; a dissociation that the scientist likely needs to solve to comprehend the situation.

This thesis opens environmental monitoring to the mobile domain. Mobile technology and wireless networks enable timely access to a wealth of data from sensors, advanced numerical processes (e.g., simulations), and even previous studies. However, the availability of data is only one aspect of the problem. The challenge is that of combining heterogeneous data sets, with their own levels of scale, stored in various formats, while standing at the site of study.

## 1.2 Mobile and Outdoor Augmented Reality

Mobile AR is a technology by which computer generated information is registered onto the surroundings of a freely roaming person [Höllerer and Feiner, 2004]. Artificial information is thereby perceived as part of the real world. Mobile AR defines a human computer interface (HCI) paradigm, whereby moving and interacting with the real world serves as interface to the digital information [Schmalstieg and Reitmayr, 2005]. These concepts build upon the notion of an AR system, characterized by aligning virtual objects with physical ones in a real environment, interactively and in real-time [Azuma et al., 2001]. The mobile AR experience then depends on a device that represents the point of view and interaction capabilities of the user.

Early outdoor AR prototypes relied on head mounted displays (HMD) driven by portable computers in backpack setups. The Touring Machine [Feiner et al., 1997], the first outdoor AR prototype, experimented with wearable interfaces for navigation and information search in the frame of a 3D graphical tour guide. Thereafter, several HMD-based prototypes investigated the wearable, world-as-a-user-interface paradigm in the frame of outdoor gaming [Thomas et al., 2000], navigation [Thomas et al., 1998] and authoring [Piekarski and Thomas, 2001] [Baillot et al., 2001], among others. Beyond the applications, these developments contributed knowledge about user interfaces suitable for outdoor AR [Höllerer et al., 1999] and collaboration [Höllerer et al., 2001] [Reitmayr and Schmalstieg, 2001]. Notably, several works studied perceptual and ergonomic factors of these systems [Swan et al., 2007], and developed methods to study ergonomics in general [Goldiez et al., 2004]. However, even with sophisticated improvements, backpack-HMD systems are still rather obtrusive, cumbersome and/or unaffordable.

In the meantime, the pace in miniaturization geared further improvements in portable, mobile computers, seeing the development of ultra-mobile PCs (UMPC) into today's tablets, and the personal digital assistant (PDA) into smart-phones. These developments gave way to handheld AR, a less intrusive, more socially acceptable form of mobile AR.

This thesis concentrates on handheld AR as the display technology for outdoor visualization. In handheld AR, the user actively holds the display device and experiences AR as a magic lens, revealing information in the pointed direction [Wagner, 2007]. Ideally, an autonomous handheld device for AR needs to include sensors to accurately derive/represent the point of view of the user (e.g., location and orientation tracking, camera, etc.) and accommodate actuators for spatial interaction. The physical world is captured in real-time

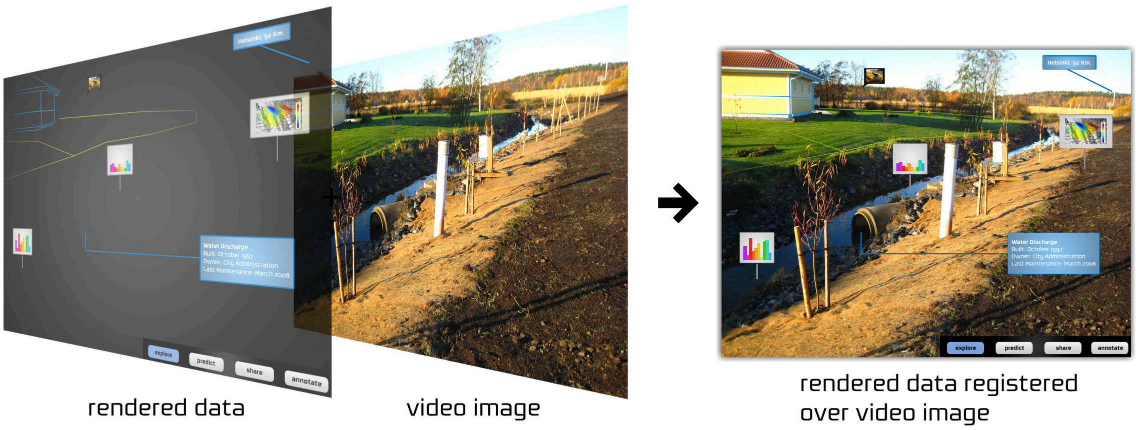


**Figure 1.2:** Augmented Reality overlaying data onto the real world. Left. Data is rendered in 3D, in geometric correspondence with the viewer position (e.g.,note the blue wireframe corresponding with the house location). Middle. The perspective of the viewer captured by a video camera. Right. In AR, data seems coexist with the assets in the video.

by a video camera upon which augmentations are overlaid.

While mobile technology offers a way to directly access heterogeneous datasets at different spatio-temporal scales from any location. Its combination with AR provides a natural way to integrate abstract content with the physical world, as shown in Figure 1.2. Outdoor AR can potentially complement the environmental monitoring workflow across numerous tasks, such as displaying sensor positions for deployment or maintenance, integrating representations for multivariate data and complex simulations by means of graphic overlays (e.g., simulation overlaid onto a mountain, sensor data represented visually and directly where they are measured, etc.).The ambitious goal is to exploit the fact that the actual site is available to integrate the information seamlessly in its real-world context.

## 1.3 Problem Statement

The development of mobile visualization techniques or platforms for environmental monitoring must overcome numerous challenges endogenous to the task (i.e., data visualization), situation (i.e., outdoors), and technology (i.e., mobile devices, AR). To form a clear picture about the situation, the scientist needs to assimilate data that differ in dimensionality, update rate, and format; often switching between representations (plots, visualizations, numbers, photographs, maps). In the case of environmental monitoring, all sources of data represent one aspect or another of the physical world. The potential of outdoor AR for visualization lies in that it establishes a spatial relationship across data sets by situating data in its spatio-temporal context. Issues of orientation and interaction with vast, heterogeneous data sets are reduced to spatial orientation and interaction, allowing people to take advantage of common navigation knowledge accumulated since childhood. However, the combination of data in such an integrated space is no trivial task. In particular, several ergonomic and perceptual factors of handheld AR have to be considered before a suitable visualization can be deployed outdoors.

Outdoor applications have to run on devices severely limited in terms of processing power, screen space, and network bandwidth, while at the same time deal with factors inherent to outdoor situations, such as bad weather conditions and rough terrain. From the ergonomics stance, a device for outdoor AR must enable the usage of such application, limiting fatigue and complying with environmental conditions.

In addition to the issues of data variability and heterogeneity, an integrated view of the situation requires consideration of well-known perceptual factors of AR, such as occlusions and shape perception [Kruijff et al., 2010a]. These add to the fact that environmental data tends to cover wide areas, where overlays can potentially occlude a large portion of the field of view.

Furthermore, due to its situated nature, outdoor handheld AR poses further perceptual constraints on the task. Visualization is an interactive process, characterized by tasks such as overview, zooming, filtering, and details on demand, that the user undertakes iteratively to analyze and understand data; trying to detect patterns, differences, connections or similarities in it [Shneiderman, 1996].

The situation that interests the geoscientist often spans a wide area. In contrast, the first person view of AR situates the user within the dataset and constraints visualization to a single perspective: the portion of the world captured by the camera. This narrow overview severely limits interaction and restricts awareness of the situation. In fact, it is essential to complement AR methods with suitable interactions to enable zoom-in and zoom-out tasks, to extend overview, and to provide different perspectives on the dataset. Ultimately, the toolset must enable the visualization workflow while maintaining the relationship with the real world context.

## 1.4 Contribution

This thesis presents a compendium of studies and techniques for mobile environmental data visualization. The aim of this work is to create a seamlessly integrated information space whereby scientific data are visualized in the context of their occurrence: the physical world. Here, a main, general contribution lies in the notion of **mobile environmental monitoring**, conceptualized in Figure 1.3, that presents both a promising field of application for mobile AR technology, as well as a novel paradigm for the monitoring task. The workflow supporting mobile environmental monitoring presented in chapter

**Figure 1.3:** Mobile environmental monitoring. Application concept designed with stakeholders and specialists in environmental monitoring. The potential for an outdoor AR solution is in the integration of the data space with the real world.

3 was sketched together with specialists in the field and gave way to an innovative infrastructure that enables access to a wide range of wireless sensors networks, deploys multivariate visualization, allows close-to real time data access, directly integrates simulation results, and integrates tools developed specifically to create a shared awareness in interdisciplinary teams.

Mobile environmental monitoring provides a suitable frame to study the characteristics and mechanics of AR in wide outdoor spaces. The second major contribution lies therefore in the systematic study of **human factors of outdoor handheld AR**. The studies and analyses hereby presented follow a systematic approach, where top-down analysis helps pinpoint issues, which are later subject to bottom-up studies in the frame of the proposed solution. The net result is the identification of human factors in general, along with best-practices and lessons learned in the frame of environmental monitoring. This approach and the accompanying evaluations strengthen the validity and applicability of the methods presented here to a large range of application domains.

To comply with severe outdoor requirements, a device for handheld AR needs to be robust, ergonomic and ultimately allow usage under extreme conditions. Chapter 4 describes the analysis, design and evaluation of a **handheld construction for spatial interaction in outdoor AR**. The process describes ergonomic requirements, prioritizing human factors of spatial interaction with handheld devices in relation to tasks in AR. It shows how to improve interaction and ergonomics of AR setups that require external extensions. These analyses and evaluations are part of an iterative process that follows the evolution of the platform towards the extreme outdoors device, as shown in Figure 1.4.

Equipped with a mobile platform and relying on a ubiquitous infrastructure to timely deliver data, the focus is drawn to the topic of **outdoor visualization of environmental**

Figure 1.4: Four generations of handheld AR platform. (Left) Vesp'r design to explore spatial interaction and usage patterns. (Middle-Left) Bulk'r focused on robustness and weather-resistant construction. (Middle-Right) Ice'r aimed at compactness. (Right) Cool'r improved modularity and extreme weather resistance.

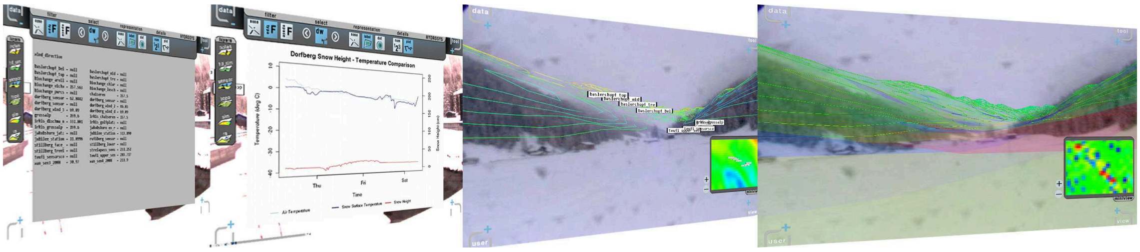


Figure 1.5: Visualization of Heterogeneous data. (Left) Tabular form summarizes details for a collection of sensors. (Middle-Left) Combined plots show measurements over time. (Middle-Right) Labels identify sensor locations in AR mode. (Right) Simulation results color-code the landscape in AR mode.

**data using AR**. A first stage considers what are appropriate AR representations of different data formats. These representations lead to perceptually-based techniques that strive to fuse captured aspects of the world in comprehensive outdoor AR visualizations (see Figure 1.5). Notably, due to physical constraints of a mobile device (i.e., reduced screen space), it soon becomes clear that only a handful of visualizations can be intelligibly active at the same time. Based on the workflow of environmental data visualization, this thesis develops interfaces to organize visualizations intuitively. Furthermore, with focus in the interactive aspects of visualization, two novel techniques are proposed to provide different perspectives and extend overview of the dataset, shown in Figure 1.6.

**View sharing** is the basis of a multi-view system, leveraging multiple cameras available in the deployment site to extend overview, provide vantage zoom-in points, and increase situation awareness.

**Variable perspective view** combines views from different perspectives in a single image; with the advantages of providing focus and context, and a better use of the screen space.

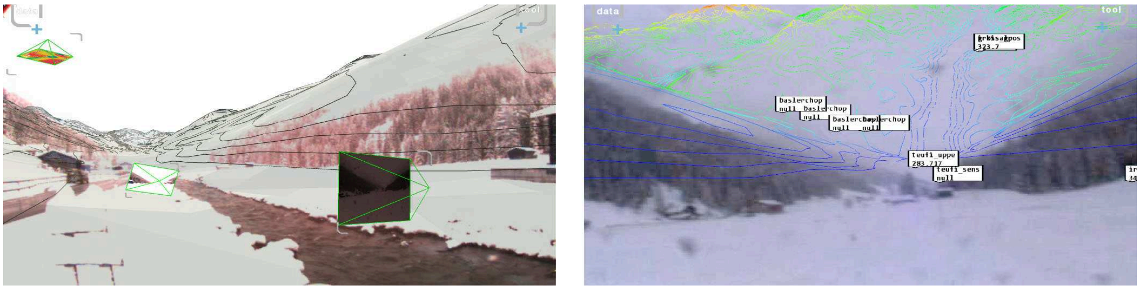


**Figure 1.6:** Awareness interfaces. (Left) In multi-view AR, tapping on the green camera icons switches to their associated view, providing a zoom-in to remote areas of a site. (Right) The variable perspective view deforms the datasets to overview the site.

## 1.5 Collaboration statement

The body of work described in this thesis would not have been achieved without the contribution of multiple researchers, summarized by the co-authors in related publications. The following list of people that closely collaborated and, in a sense, guided the direction of this work, deserve special mention:

**Erick Mendez** of Graz University of Technology co-developed the visualizations presented in chapter 5. He developed the saliency modulation technique used for the experiments of section 6.5 and collaborated in the design and performance of the memory study.

**Ernst Kruijff** of Graz University of Technology co-designed the hardware setups presented in chapter 4, and collaborated in the design of multiview evaluations presented in sections 7.1.4, as well as the final evaluations presented in section 3.6.5.

**Raphaël Grasset** of Graz University of Technology implemented the final user interface described in 3.5.3 and collaborated in the design and instrumentation of final evaluations presented in section 3.6.5. He also contributed to the design of the study on variable perspective presented in section 7.2.3.

**Gerhard Schall** helped carrying out initial evaluations of the Vesp'R and Bulk'R hardware setups (sections 4.3 and 4.4).

**Alessandro Mulloni** of Graz University of Technology implemented the transitional technique 7.1.3.3, and collaborated in the evaluation of multi-view transition techniques presented in section 7.1.4

**Holger Regenbrecht** of University of Otago provided insightful guidance on user evaluations in general, and on the evaluation of multi-view transitions of section 7.1.4 in particular.

**WSL Institute for Snow and Avalanche Research** was the main contributor of domain specific knowledge. In particular, Prof. Dr. Michael Lehning, Dr. Mathias Bavay, Nicholas Dawes and Thomas Grünewald actively participated in the conception of the alpine scenarios (see section 3.6.2.1), they provided infrastructure blocks referred to in section 3.3.1 and 3.3.3 upon which this thesis builds.

**Laboratory of Environmental Fluid Mechanics and Hydrology** at the Ecole Polytechnique Fèdèrale de Lausanne played an important role, providing domain specific knowledge and helping to define usage scenarios. The input provided by Dr. Vincent Luyet and Dr. Silvia Simoni was an invaluable contribution during user centered design phases (see section 3.1).

**Geoinformatics Department** at the Aalto University, Finland, contributed with domain specific knowledge focused towards urban scenarios. Prof. Ari Jolma and Ioan Ferenick contributed the conception and instrumentation of the nordic scenario described in section 3.6.2.2.

**LSIR Distributed Information Systems Laboratory** at the Ecole Polytechnique Fèdèrale de Lausanne provided infrastructure and support needed to access environmental sensors on-site (see section 3.3.2).

The outcomes of this thesis are largely based upon peer-reviewed publications. The following list is a road map from the contents of this thesis to the co-authored publications.

- Veas, E. and Kruijff, E. (2008). Vesp'R: design and evaluation of a handheld AR device. In *ISMAR '08: Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 43–52, Washington, DC, USA. IEEE Computer Society

  The publication contributes studies on ergonomics of handheld augmented reality (i.e., pose, patterns of usage, etc.) and provides the foundation for the work presented in Chapter 4. The author together with *Ernst Kruijff* performed a series of functional evaluations of handheld AR and ergonomics analyses using mock-ups, that delivered a set of guidelines on pose and interactions. Thereby, the author designed and developed the Vesp'r prototype, and the applications used for evaluation. The evaluation methodology and execution were shared among the authors.

- Schall, G., Mendez, E., Kruijff, E., Veas, E., Junghanns, S., Reitinger, B., and Schmalstieg, D. (2008b). Handheld Augmented Reality for underground infrastructure visualization. *Personal and Ubiquitous Computing*, 13(4):281–291

  This publication contributes initial evaluations of a handheld AR application deployed outdoors to Chapter 4. The author together with *Ernst Kruijff* contributed the mobile platform, its evaluation methodology, and analysis of results . All authors contributed to the experimental design for the evaluation of usability. The Vidente application was developed by *Erick Mendez and Gerhard Schall.*

- Veas, E., Kruijff, E., and Mendez, E. (2009). HYDROSYS - first approaches towards on-site monitoring and management with handhelds. In J. Hřebíček, J. Hradec, E. Pelikán, O. Mírovský, W. Pillmann, I.Holoubek, T. B., editor, *Towards e-environment, EENVI2009*, Prague, Czech Republic

  This position paper outlined conceptual ideas behind mobile environmental monitoring, discussed in Chapter 3. All authors shared the conceptual design of a system to supply a mobile AR application with real-time environmental data. The concepts of mobile environmental monitoring were sketched by the authors following discussions with the HYDROSYS consortium. The application scenarios were defined by partners of the HYDROSYS consortium.

- Kruijff, E., Veas, E., Gruenewald, T., Silvia, S., Luyet, V., Salminen, O., Nurminen, A., and Lehtinen, V. (2010b). HYDROSYS: on-site monitoring and management of environmental processes using handheld devices. In Anand, S., Ware, M., and Jackson, M., editors, *GeoHydroinformatics: Integrating GIS and Water Engineering.* Taylor & Francis Inc

  This book chapter extended the concept of mobile environmental monitoring with scenarios and initial visualizations from the first prototypes. All authors contributed further concepts that outline mobile environmental monitoring.

- Veas, E. E. and Kruijff, E. (2010). Handheld Devices for Mobile Augmented Reality. In *MUM '10 Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia*, Limassol, Cyprus

  This paper contributed complementary work in the development of hardware interfaces for handheld AR in Chapter 3. The authors shared the design, development and construction of the four platforms.

- Veas, E. E., Mulloni, A., Kruijff, E., Schmalstieg, D., and Regenbrecht, H. (2010). Techniques for View Transition in Multi-Camera Outdoor Environments. In *Graphics In-*

*terface 2010 (GI '10)*, pages 193–200, Toronto, Ont., Canada. Canadian Information Processing Society

This paper evaluates techniques for navigation in multi–camera systems under the premise of maintaining spatial awareness at low cognitive cost, as described in Chapter 7. The conceptual foundations of multi-camera AR were developed by the author together with *Ernst Kruijff*T. The author developed the Tunnel and Mosaic techniques for camera transition. The Transitional technique was developed by *Alessandro Mulloni*. All authors contributed to the study design and execution.

- Veas, E., Mendez, E., Feiner, S., and Schmalstieg, D. (2011). Directing Attention and Influencing Memory with Visual Saliency Modulation. In *to appear in Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI2011)*, Vancouver, British Columbia, Canada

  Provides a theoretical background and prospective outlook in future directions of visualization in mobile AR to close Chapter 6. The Saliency Modulation Technique was developed by *Erick Mendez* and is reported in his thesis [Mendez, 2011]. *Steven Feiner and Dieter Schmalstieg* guided *Erick Mendez* and the author in the design of the different evaluation stages. The author worked jointly with *Erick Mendez* to identify levels of modulation that can go unnoticed by the viewer, while still influencing recall of selected areas in the visual input.

- Veas, E., Grasset, R., Kruijff, E., and Schmalstieg, D. (2012b). Extended overview techniques for outdoor augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 18:565–572

  This article details techniques that improve site understanding for outdoor Augmented Reality (AR). In particular, it contributed two advanced view management techniques to support the visualization workflow, described in Chapter 7. The author designed and developed the view-sharing system and the variable perspective view techniques. All authors contributed to the study of the techniques.

- Veas, E., Grasset, R., Ferencik, I., Grünewald, T., and Schmalstieg, D. (2012a). Mobile Augmented Reality for Environmental Monitoring. *Submitted for review to Personal and Ubiquitous Computing*, 0(0):0–16

  In this article, functional and workflow analysis and initial human factors studies guided the implementation of a fully functional prototype and its acceptance evaluation, the core of Chapter 3. Whereas previous publications on mobile environmental monitoring were mostly conceptual, this article contributes a formal definition based on three years of research in the topic. The topic of mobile environmental monitoring is the result of a joint study by the HYDROSYS consortium. The author, together with *Raphaël Grasset*, and *Ernst Kruijff* guided this study and outlined the resulting concepts. *Raphaël Grasset* and the author developed the last prototype of the system. *Raphaël Grasset* developed the final user interface and the filtering methods. The author developed the final visualization interface and view management techniques. *Ioan Ferencik and Thomas Grünewald* instrumented the field scenarios and assisted in several stages of concept. The evaluation was a joint effort by all the authors and was assisted by people in the HYDROSYS consortium.

## 1.6 Outline

The rest of the dissertation is organized as follows:

**Chapter 2: Related Work** guides the exposition of previous research in various fields that this dissertation touches. It visits environmental monitoring to understand data cycle and tools for visualization and interpretation.

**Chapter 3: Mobile Environmental Monitoring** takes a top-down, user centered design (UCD) approach to analyze environmental monitoring from the perspective of the professional in the field. UCD lends functional and user analysis guiding the definition of a set of challenges that ultimately lead to the solution: a mobile AR approach to scientific visualization. As the concept of mobile environmental monitoring is drawn, the chapter presents design and implementation of a fully functional prototype and its evaluation in two real deployment scenarios.

**Chapter 4: Handheld Devices for Mobile Environmental Monitoring** builds upon the needs of high-quality AR and the motivation to experiment with interfaces for spatial interaction. This chapter analyzes the design and construction of handheld platforms and their evolution towards the extreme outdoors device. The first generation, reported in [Veas and Kruijff, 2008], was a result of various studies on ergonomics and human factors. Thereafter, each following iteration in the design-production process was guided by experiences and evaluations that resulted in new guidelines for future versions.

**Chapter 5: Environmental Data Visualization and Handheld AR** describes novel AR visualization techniques for a seamless information space, whereby heterogeneous datasets share a common reference frame. Initial, basic techniques for sensor and overlay visualization concern the integration of different datasets in a single AR view. Thereafter, advanced, perceptually-based techniques lend subtle extensions to overcome AR issues.

**Chapter 6: Subtle Visualizations Based on Saliency Modulation** evaluates a novel technique to emphasize particular data characteristics or enhance visualization goals.

Based on the executive role of working memory in the retrieval, processing and integration of data during visualization, the chapter analyses a method for rapid, accurate, and effortless visual exploration. It concentrates on a single method; modulation of visual saliency to *suggest* preattentive targets, and contends that subtle modification of visual saliency, even when seemingly imperceptible, can influence working memory. The chapter presents empirical studies, as well as formal evaluations of techniques that push the boundaries of AR visualization.

**Chapter 7: View Management in the Workflow of Interactive AR Visualization** introduces an unobtrusive user interface to control extensive features of a tool for environmental specialists. The chapter introduces view management tools to support the visualization workflow, enhancing overview and zoom capabilities based on virtual views and view-sharing. The view-sharing infrastructure gives way to the discussion on collaborative interfaces for outdoor AR.

**Chapter 8: Sharing Observations in Collaborative Mobile AR** extends the analysis to a multi-user environment. This chapter studies the mechanics of collaboration for a group of people engaged in mobile AR exploration of a site. Based on view-sharing, the proposed model of mobile collaboration describes mechanisms that can be readily adopted as general solutions for collaboration in outdoor, mobile AR.

**Chapter 9: Conclusion** concludes this dissertation with a discussion of the implications of results for environmental monitoring and potential applications in other fields, as well as presenting promising directions for future research.

# Chapter 2

# Related Work

AR for scientific visualization is a relatively new area of research, largely open for exploration. This chapter provides a foundation on previous research upon which the contributions of this thesis build. It sets the scope of this thesis in the context of each of the fields it touches: mobile AR, environmental monitoring and scientific visualization.

Mobile AR proposes the integration of data visualization in the surroundings of the user, subordinating interaction to the user's mobility capacities. That is, by registering the information space to the user's viewpoint; AR relies on natural motion and the user's understanding of space for interaction. In the case of environmental monitoring, all data refer, in one way or another, to the physical world. Thus, we automatically obtain a spatial mapping between data and physical world that can be exploited for visualization. Scientific visualization describes how scientists experience spatialized data. That is, it describes the possibilities to represent numerical values in the data, and the patterns of interaction usually followed for their analysis.

The chapter starts by briefly revisiting the history of AR, with enough information to understand how mobile AR came to be and the technology involved in the outdoor AR experience. Thereafter, the current trends in environmental monitoring, data gathering and visualization are introduced, followed by a description of the topic of interactive visualizations. With this frame, we proceed to analyze previous works in AR for visualization that motivate this thesis.

## 2.1 Augmented Reality

AR owes its origin to the seminal work of Ivan Sutherland and colleagues. In their quest to create the ultimate display (Figure 2.1), they built a mechanically tracked 3D see-through head-worn display capable of surrounding the user with computer generated 3D

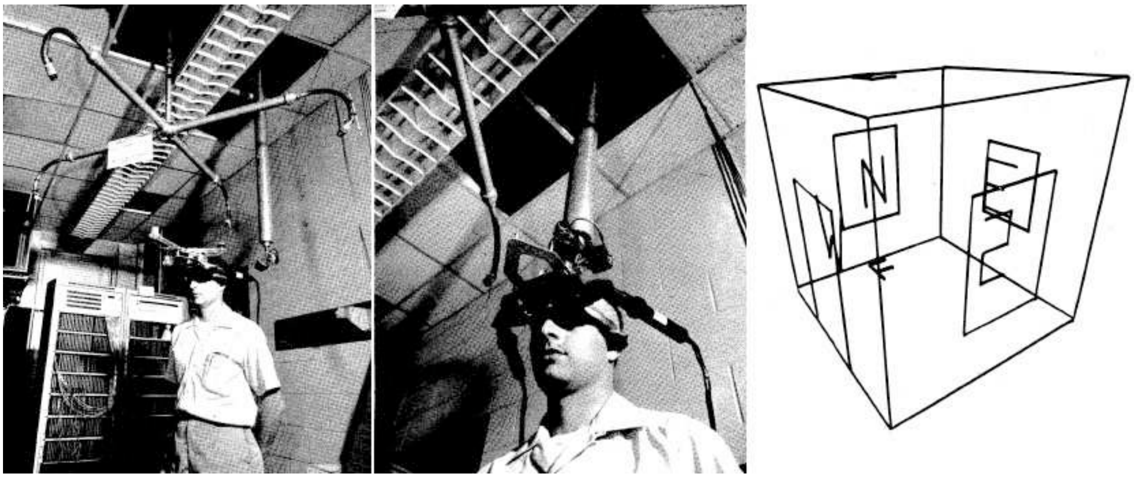


**Figure 2.1:** Sutherland's head mounted 3D display used miniature CRTs to render wire-frame models (Right). It could use a mechanic head position sensor (Left) or an ultrasonic one (Middle). Images from [Sutherland, 1968]

information [Sutherland, 1968]. Though research in computer graphics and later virtual reality advanced at renowned institutions, several years passed till the next technological milestone for AR appeared. Tom Furness developed a high-resolution heads-up overlay display for the US Air Force Super Cockpit project [Furness, 1986]. The term *augmented reality* is attributed to Caudell and Mizell and their work on conveying wiring instructions for aircraft assembly performed at Boeing [Caudell and Mizell, 1992]. Although early works positioned AR as a mobile interaction paradigm, whereby we rely on our sense of involvement with the physical world to explore concepts not realizable in it [Sutherland, 1965], tracking technology constrained it to limited, controlled spaces. Clearly mobile AR depends on a device that represents the point of view and interaction capabilities of the user.

### 2.1.1 Mobile AR

Early outdoor AR prototypes relied on head mounted display (HMD) hardware driven by portable computers in backpack setups, like the one shown in Figure 2.2. The Touring Machine [Feiner et al., 1997], the first dated outdoor AR prototype, experimented with wearable interfaces for navigation and information seeking in the frame of a 3D graphical tour guide. It was later extended for the placement of multimedia presentations in the physical world (termed multimedia documentaries) [Höllerer et al., 1999]. Thereafter, several HMD-based prototypes investigated the wearable, world-as-a-user-interface paradigm in the frame of outdoor gaming [Thomas et al., 2000], navigation [Thomas et al., 1998] and authoring [Piekarski and Thomas, 2001] [Baillot et al., 2001], among

**Figure 2.2:** Tobias Höllerer wearing the MARS system (Left,Middle). Augmented scene from situated documentaries (Right) [Höllerer, 2004]

others. Beyond the applications, these developments contributed knowledge about user interfaces suitable for outdoor AR [Höllerer et al., 1999] and collaboration [Höllerer et al., 2001] [Reitmayr and Schmalstieg, 2001]. Notably, several contributions studied perceptual and ergonomic factors of these systems [Swan et al., 2007], and developed methods to study ergonomics in general [Goldiez et al., 2004]. However, even with sophisticated improvements, backpack-HMD systems are still rather obtrusive, cumbersome and/or unaffordable [Wagner and Schmalstieg, 2003].

In the meantime, handheld AR, a novel configuration for AR display, was increasingly geared by the pace in miniaturization. Improvements in portable, mobile computers saw the development of ultra-mobile PCs (UMPC) into nowadays tablets, and the personal digital assistant (PDA) into what we know as smart-phones. Handheld AR can be traced back to the work of Rekimoto on augmented interaction in the NaviCam [Rekimoto and Nagao, 1995]. With the goal to use real-world information as implicit interaction method, the NaviCam relied on a video camera and marker-based tracking to detect the real-world situation and enhance it with context sensitive information. Rekimoto described the *magnifying glass approach* [Rekimoto, 1995], whereby the real-world is enlarged in terms of information. Handheld-AR is defined by such metaphor, where the user actively holds the display device and experiences AR as through a magic lens revealing information in the direction pointed. Handheld-AR received a major boost with its deployment on consumer devices. The AR-PDA project used a thin-client approach to show AR contented generated and streamed wirelessly from a remote server [Kleinnjohann, 2001]. Wagner and Schmalstieg developed the first self-contained PDA AR application [Wagner and Schmalstieg, 2003]. Wagner and colleagues created a software infrastructure supporting an ever increasing number of mobile devices and phones before they became smart, and went on to deploy applications for collaboration, museum guides and navigation applications among others [Wagner, 2007]. In these applications, registration and tracking are based

**Figure 2.3:** Rekimoto's Navicam (Left) used fiducial markers to register augmentations in the surroundings of the user (Right). [Rekimoto and Nagao, 1995]

**Figure 2.4:** Handheld AR collaboration. Wagner's invisible train was an AR game where users collaborated to direct their virtual trains through a physical railroad track. [Wagner, 2007]

upon recognition of known targets (e.g., markers, planar targets such as photographs, or objects with a known, distinguishable shape), referred to as closed-loop tracking by Höllerer and Feiner [Höllerer and Feiner, 2004].

Other works have considered the general mobile AR case, independent of any preparation of the environment. Hereby, registration depends on sensing, usually based on GPS and orientation sensors, the 6DOF of the person, as was the case in early HMD-based systems [Höllerer and Feiner, 2004]. In general, this modality depends on high-end sensors to deliver accurate pose, and advanced sensor fusion mechanisms to combine their output. A compilation of techniques for sensor fusion in tracking can be found in [Julier and Bishop, 2002].

A handful of testbed systems demonstrate outdoor vision-based tracking techniques [Reitmayr and Drummond, 2006], or the combination of sensors and vision based tracking in a multi-sensor fusion algorithms [Schall et al., 2009].

### 2.1.2 AR Definition and Modalities

The notion of an AR system is characterized by aligning virtual objects with physical ones in a real environment, interactively and in real-time [Azuma et al., 2001]. This formal definition of AR is somewhat present since its inception in Sutherland's seminal work: "the image presented by the three-dimensional display must change in exactly the way that the image of a real object would change for similar motions of the user's head" [Sutherland, 1965].

Taking a comparative perspective, Milgram and Kishino define an application depending on its proportion of virtual vs real content [Milgram and Kishino, 1994]. This results in a full continuum ranging from VR, with purely virtual content, through augmented virtuality, where virtual content is dominant and extended with bits of reality, and augmented reality, where real content dominates, (see Figure 2.5). In their work about transitioning between VR and AR, Grasset et al [Grasset et al., 2008] define these stations (AR and VR) as contexts to present information. In chapter 7, we extend the notion to all modalities in the Milgram-Kishino continuum. To allow seamless interaction with the data from multiple perspectives, we define transition techniques between different AR contexts, but also enable smooth travel through the continuum. As Figure 2.5 shows, to gradually change the level of awareness, we designed a set of interaction metaphors that take the visual display to different stations within the continuum. Our view management techniques combined with filtering approaches control the amount of virtual and real content displayed, and allow the user to selectively attend to portions of the data.

### 2.1.3 Mobile Devices for Outdoor AR

Wearable AR systems can be classified as a subfield of wearable computing [Mann, 1996]. A wide range of mobile, wearable AR systems exists, and make use of mobile computers in different sizes. Two major directions can be identified: The backpack-mounted systems [Feiner et al., 1997] [Thomas et al., 1998] and the handheld AR system [Rekimoto, 2001]. Nowadays, a notable number of AR applications are available on smart phones. Smart phones, though, have inferior positioning sensors compared to custom-made platforms. More recently, platforms are appearing that use handheld projectors for augmented reality [Karitsuka and Sato, 2003].

A few AR platforms, including previous work in our lab [Schall et al., 2007] [Schmalstieg and Reitmayr, 2005], extend tablet form factor PCs to create a computationally and sensory more powerful platform. Most of these platforms were not build with ergonomics in mind, but rather as a straightforward way of attaching sensors to a mobile computer, as shown in Figure 2.6. The general approach of attaching sensors and actuators to a tablet PC is a combination of a Perspex case and duct tape, which normally does not lead to a

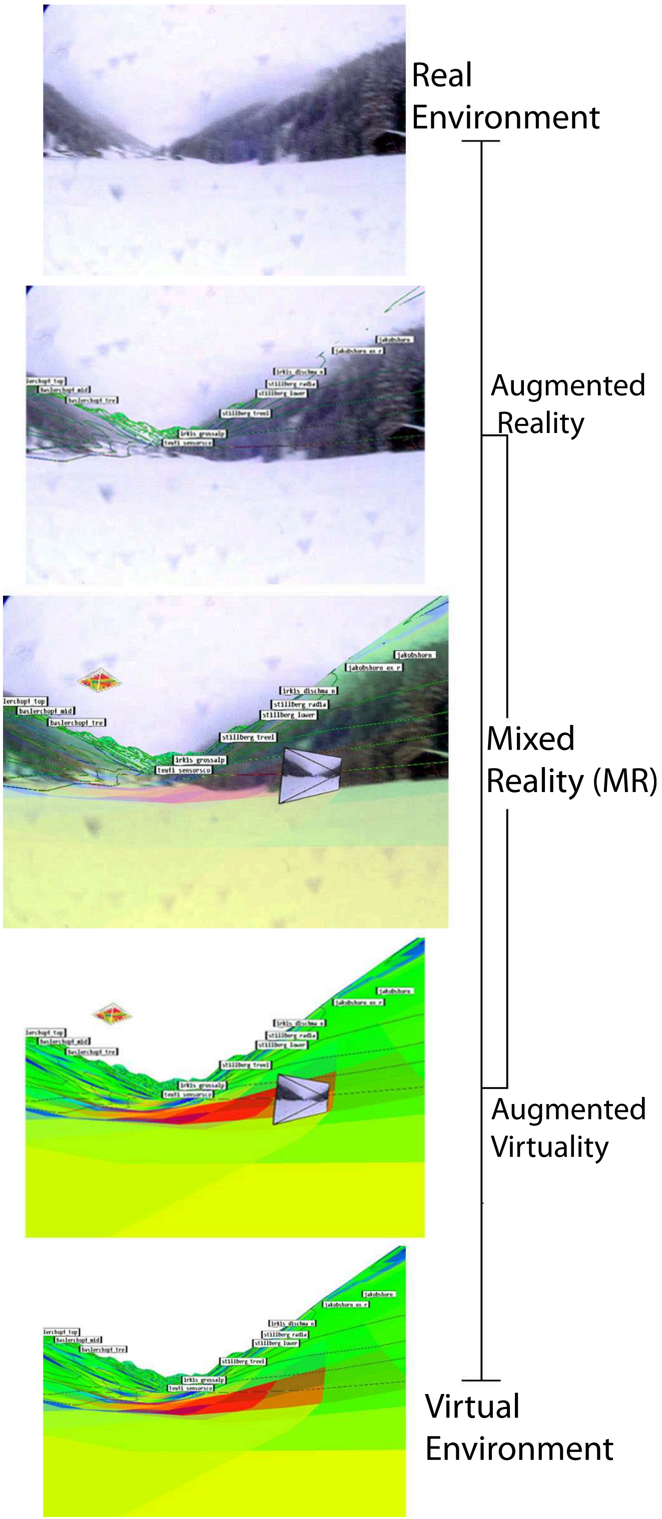


**Figure 2.5:** Navigating the Reality-Virtuality (RV) Continuum. Our view management metaphors visit different stations of the Milgram continuum [Milgram and Kishino, 1994]. At the virtual end, a 3D model shows snow surface skin temperature (TSS) and sensor data in labels. Moving up, an inset video from another user adds minimal real content. The middle of the scale blends all content in MR. Moving further towards the real environment minimizes virtual content,e.g., by tracing isolines for the TSS overlay.

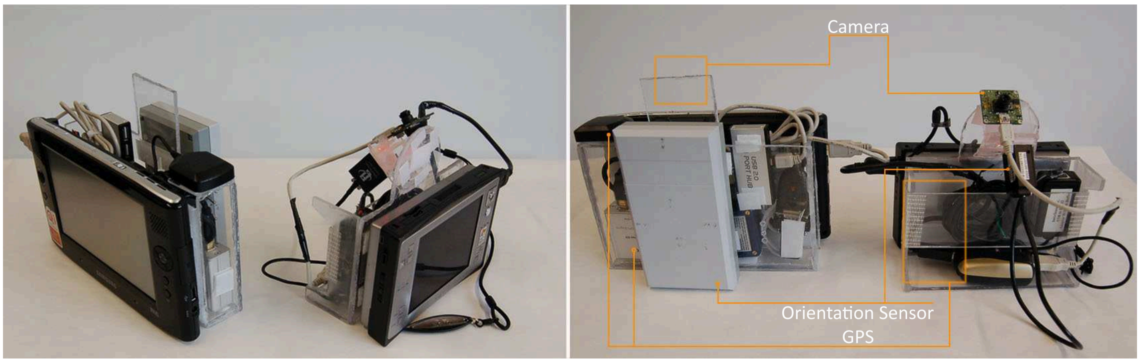


**Figure 2.6:** Perspex encasings for two devices, the Samsung Q1 and the Sony U-70 (left and right in each image respectively). The image on the right shows the sensors where available or a frame indicating the space they normally occupy. [Veas and Kruijff, 2008]

highly ergonomic construction. Some exceptions are the AR mask by Grasset et al [Grasset et al., 2005] and MARTI demonstrated by Stutzman and colleagues [Stutzman et al., 2009]. Other devices like the Xybernaut setups separate the processing unit from the display, lowering the weight held in the hands, though they are hardly used anymore.

Chapter 4 describes the design and development of a mobile device for outdoor handheld AR. In particular, it informs on rapid prototyping, with foundations on the design of 3D devices [Bowman et al., 2004], and encompasses the reasoning on the usage of different materials, similar to the design of some haptic devices [Ju et al., 2003]. The chapter focuses on multiple iterations of design-evaluation, and the lessons learned in deploying a device for extreme outdoor usage.

## 2.2 Environmental Monitoring

Environmental monitoring is a rather specialized activity and several factors influence the modality applied (i.e., how it is done). The process and location of interest often dictate what is to be measured, then technology and budget define how to measure it. The following discussion focuses on the areas of snow sciences and hydrology sciences.

In the case of water quality, the state-of-the-art in monitoring and management is based on manual sampling schemes, where water samples are taken and analysed biweekly or even more sparsely. Real-time monitoring is rarely applied, but can be done in specific cases and always requires particular arrangements. The information transfer and utilization comprise mostly custom-built, non-generic solutions. Systems that record water quantity and quality data in the field and transmit it to the Internet using GPRS and other wireless communication protocols exist, but to deliver such data to end-users on-site in near-real-time is a novel feature. For example, pollution caused by stormwater in urbanized areas is a poorly understood process, for which better simulation methods based on real-time data gathering are needed. The ecological treatment of stormwater has not received much attention [Hood and Reihan, 2007].

The state-of-the art in alpine watershed monitoring and management is based on field campaigns that generally focus on deploying relatively few “expensive” base stations having traditional sensors, limiting the spatial coverage, the online data access as well as the possibility to have active control to respond to changing conditions. Moreover, experimental field work in hydrologic sciences is still very much dominated by single-use, often small-scale efforts of one or a few research groups. A few exceptions exist, such as the national programmes LOCAR [1] and CHASM [2].

Thus, environmental monitoring mostly involves on-site observation, few expensive sensing stations with data logger (e. g. IFKIS network in Switzerland) and Geographical Information System (GIS). With this current approach, many processes cannot be monitored. Sensor data is still (timely) limited and scattered in the environment: even when a problem is detected by a sensor, it can often not be traced back directly, since detailed information on the area is missing. As a result, many processes are poorly understood and both the representation (i.e., physical model) and visualization thereof are incomplete [Barrenetxea et al., 2008a].

For example, warming, melting and disappearance of permafrost have accelerated in recent times, increasing damage to structure and raising public concern. By the middle of the 21st century, climate change may cause 2 to 3 °C warming of the frozen ground and a $10\%$ to $16\%$ reduction of the total permafrost [Anisimov, 2001]. Changes such as these modify the water balance, raising social problems, and making environmental monitoring with a high resolution in space and time more and more important [Hilbich et al., 2008b].

In recent years, trends are moving toward deploying a large number of wireless sensing stations in order to provide high spatial and temporal density measurement. Wireless Sensor Networks (WSN) [Chong and Kumar, 2003], [Culler et al., 2005], [Aberer et al., 2007] have become a widespread tool for monitoring a wide range of environmental phenomena. Many research projects are investigating possible applications of sensor networks, ranging from habitat monitoring [Szewczyk et al., 2004] to agriculture [Langendoen et al., 2006] and to environmental monitoring [Selavo et al., 2007], [Martinez et al., 2004], [Werner-Allen et al., 2005]. However, deploying a WSN in the field has always been reported as a difficult task and remains challenging [Barrenetxea et al., 2008b]. In many cases, sensing stations have severe limitations, such as no real time data and short spatial coverage, elevated cost, etc.

Enhancing environmental modeling, prediction tools, and the decision making process

---

[1] www.nerc.ac.uk/research/programmes/locar/background.asp

[2] www.ncl.ac.uk/chasm/ChasmOverview/index.htm

in general with real-time image and data transmission [Shin et al., 2007], and visualization of spatial and temporal variability of environmental parameters have been identified as the next challenges in environmental monitoring [Bogue, 2008].

Visualization is a frequently used technique for analyzing scientific data, and well-accepted within geoscience disciplines [Nocke et al., 2008]. The heterogeneity of data has brought forward a vast number of techniques to deal with various issues, including spatial and temporal aspects, different view types (region-based or station-based), or various ways of relating data (2D maps or 3D globes). As comes forth from a study by Nocke et al [Nocke et al., 2008], spatial reference of data is of high importance. As users with different levels of experience and interest access geo-scientific data, there is no uniform type of visualization. Most users apply standard 2D presentation techniques, predominantly for scientific purposes. Nonetheless, both scalar and vector based 3D data representations are an active field of research and increasingly gaining acceptance.

A Geographic Information System (GIS) is a system of hardware and software used for storage, retrieval, mapping, and analysis of geographic data. GIS is often used for the management and visualization of geo-spatial data. GIS, combined with remote sensing and GPS, is currently a common tool for environmental monitoring [Gao and Liu, 2001].

Mobile GIS combines GIS software with handhelds equipped with GPS and network access. They predominantly use map-based representations, similar to their desktop counterparts. A large part of the current solutions (like the Arcpad system [ESRI, 2004]) are unfortunately single-user, limited to 2D representation, lack support for real-time (up-to-date) data or are limited in term of data interoperability or their visual combination.

## 2.3 Scientific Visualization

The task of visualization is one of communication; it intends to effectively communicate the information contained in the numbers from a dataset using graphical means. For scientific visualization, the goal is to facilitate accurate, quick, and unbiased interpretation of physical phenomena from its measured physical quantities [Laidlaw et al., 2005].

Understanding and insight are the main goals of scientific data visualization [McCormick, 1988], which relies on the innate perceptual abilities of people to detect patterns, differences, connections or similarities in graphical representations [Shneiderman, 1996]. With the vast amounts of data available, a suite of visual metaphors and associated visual approaches becomes necessary to provide users with multiple complementary views of their information [Thomas and Cook, 2005]. Our research concentrates on developing visualization metaphors seamlessly integrated within their physical world context, with the goal of emphasizing the sensation of correspondence of the data with the surroundings in a way that allows the exploration of the relations between them. The experience elicited with mobile AR of the information being part of the surroundings enhances situation awareness, supports spatial judgements, and provides a basis of natural navigation for interaction with both data and physical world. The physical world effectively becomes part of the dataset. Thereby, relying in the notion pushed forward by Van Dam et al., that exploring and understanding the complex interrelationships between data and the physical world can readily be facilitated through kinesthetic feedback gained by peering around at them from within, taking alternative perspectives or interacting with them [Dam et al., 2000].

### 2.3.1 Visual Representation

The choice of visual metaphor is crucial for comprehension analysis and understanding of the phenomenon underlying the data [McCleary, 1983]. The first step is figuring out what visual entities can be used to represent numbers in our dataset, so it is easy to explore the relationships among the variables in it. The components used to create a representation are our visual dimensions. Color, shape, size and movement are examples of some of those dimensions [Acevedo, 2007]. Our focus lies on the combination of visual data representations with live video streams of the real world. We rely primarily on existing visualizations used to represent environmental data. In the case of 2D overlays, the representation mainly takes color dimensions.

We intend to produce perceptually high quality imagery. One important factor is including informative labels and rendering them using clearly readable fonts [Hibbard, 2004]. Another factor is related to visual coherence of the scene presented. In the case of scientific visualization, the interest is not so much on mimicking the visual properties of the physical world (e.g., global illumination), but on maintaining the color scheme used for visualization while conveying correct spatial relations. Shape perception of 3D scenes can be supported by adding supplemental line renderings [Girshick et al., 2000]. Girshick et al. suggest using principal directions to convey shape in 3D line drawings. We rely on silhouetting with contour and isolines to enhance the perception of shape and depth.

A visualization metaphor supplies a method for rapid, accurate, and effortless visual exploration by directing preattentive features [Healey, 2001]. Various techniques can be used to emphasize particular data characteristics or visualization goals. In most cases, some methods may present salient information more quickly and accurately. However, users are engaged in a thought process [Hibbard, 2004], and we want to avoid disruptive visualizations. The understanding of visualizations is performed by our cognition; the process of knowledge acquisition and reasoning [Anderson et al., 2011]. Working memory is a central construct of the cognitive process, responsible for retrieval, processing and integration of data during executive decision making [Baddeley, 1992].

We concentrate on a single method: modulation of visual saliency to *suggest* preattentive targets. We contend that subtle modification of visual saliency, even when seemingly imperceptible, can influence working memory.

#### 2.3.1.1 Visual Saliency

Visual salience (or visual saliency) is the distinct subjective perceptual quality that makes some items in the world stand out from their neighbors and immediately grab our attention [Itti, 2007]. It refers to the evolved process in primates and other animals that restricts complex object recognition to small areas or objects at any one time that are in analyzed serially. Saliency is commonly agreed to have bottom-up and top-down factors. Bottom-up (memory-free, stimulus-based) factors refer to pure sensory information, such as a bright light suddenly appearing in front of us. Top-down (memory-bound, goal based) factors involve a conscious effort, such as ignoring more salient stimuli while scanning a book index. Bottom-up factors announce to the organism whether a location is different enough from its surroundings to warrant attention. Measurements of the attention process of an organism are typically focused on stimulus-only factors. The most influential works on understanding this were carried out by Treismann and Gelade [Treisman and Gelade, 1980], and by Koch and Ullman [Koch and Ullman, 1985]. Koch and Ullman, in particular, proposed the idea of a single map that is a combination of individual salient contributions; the normalized result is referred to as the saliency map. They state that the saliency at a given location is determined primarily by how different this location is from its surround in color, orientation, motion, depth, etc.

#### 2.3.1.2 Saliency and Visual Attention

There is much evidence indicating a correlation between visual attention and the saliency map. Ouerhani et al. [Ouerhani et al., 2004] and Santella et al. [Santella and DeCarlo, 2004] used an eye tracker to confirm the relationship between the saliency map and human visual attention. Lee et al. [Lee et al., 2007] went one step further by using the saliency map to track objects being attended to by the user.

Practically any change done to an image will modify its saliency map. Blurring, (de)saturating, harmonizing and distorting are operations that implicitly change the saliency of an image. Recent research has focused on directing attention through saliency manipulation for volume rendering [Kim and Varshney, 2006], non-photorealistic stylization [Santella and DeCarlo, 2004] and geometry [Kim and Varshney, 2008]. These works concentrate on creating salient features; in contrast, our work receives an existing image as input and pursues the manipulation of its existing salient regions. Our technique works with dynamic live video and can thus support augmented reality applications with arbitrary scenes and without requiring an eye tracker.

#### 2.3.1.3 Saliency and Memory

There is a two-way relation between attention and memory that has been widely studied in the past [Awh et al., 2006] [Itti, 2005] [Chun and Turk-Browne, 2007]. Awh et al. [Awh et al., 2006] identified experiments leading to the conclusion that attention influences processing during both early sensory and post-perceptual stages. They also collected evidence supporting that the same attentional processes that facilitate early sensory identification of new information are recruited for active maintenance of information in memory. Two recent studies have proven the influence of saliency in memory, albeit with different results regarding the reasons. Berg and Itti [Berg and Itti, 2008] came to the conclusion that saliency contributes to memory through influencing overt attention. They had participants examine a shopping-related scene for 2s and then asked if a target item was contained in the scene. They found that fixation times, but not saliency, influenced performance. Fine and Minnery [Fine and Minnery, 2009] found that the influence of saliency extends beyond occulomotor behavior to higher order centers involved in spatial working memory. They presented participants with maps that included a number of icons to memorize. After a pause, participants had to drag each icon to its original position. They found that participants attended to icons equally regardless of their saliency (quantified using the model from Itti et al. [Itti et al., 1998]), but errors in placement were significantly reduced for salient icons. Thus, results could not be explained by a biasing of overt attention. Both cases support the fact that saliency influences memory. We assume that by actively modifying an object's saliency, we can influence memory.

### 2.3.2 Interaction Metaphors

To effectively examine data, users need methods to organize their information, gain overview of it, explore and examine it at different levels [Thomas and Cook, 2005]. Visualization can be oriented towards data exploration, in which case raw data are presented graphically to promote visual thinking. Alternatively, once the scientist recognizes certain patterns and starts to hypothesize about the situation, visualization takes an explanatory role, where users want to highlight some characteristics of the data [Acevedo, 2007].

These two modalities directly correlate to a well-known visualization workflow; alternating overview with zooming and filtering or getting details on particular datasources [Shneiderman, 1996] [Thomas and Cook, 2005]. Exploratory visualization provides a broad understanding of the data, and serves as initial, qualitative assessment of data gathering, modeling and experimental methods. It concentrates on providing insight into the spatial organization of data, in particular the spatial relationships among data variables of multi-varied datasets [Acevedo, 2007]. Explanatory visualizations support the user in isolating patterns and asking questions about them.

We do not aim to replace a full data analysis with our outdoor AR application, but the paramount challenge we are after is a workflow that supports mobile visualization of real-time data. Primarily, our application focuses on instrumenting exploratory visualization, setting an integrated information space based on the (geo-)spatial quality of data. AR constitutes an interaction paradigm in itself, that leverages the human psychophysical knowledge about the world to enable intuitive locomotion and navigation through the dataset [Billinghurst and Thomas, 2011].

Nevertheless, the visualization workflow poses a challenge for the AR paradigm. AR needs a device to capture the real world, and is constrained to the physical properties of such device. Some approaches have laid out a theory to deal with information overload by filtering [Julier et al., 2002]. Sill several questions remain open. How do we gain overview on the dataset? What does zooming entail in AR? The limitations of AR for visualization are further analyzed in section 2.4.1, after reviewing other approaches for AR visualization.

## 2.4 Mobile visualization and AR technology for interactive data representation

Interaction with sensors is an active area of research in the ubiquitous computing community. In [Paxton and Benford, 2009], Paxton and Benford evaluate children experiences in participatory sensing. Kim and Paulos [Kim and Paulos, 2009] analyzed behavior changes resulting from the availability of indoor air quality measurements.

Whereas several well-known examples can be found in virtual reality (VR) e.g [Lin and Loftin, 1998], AR for scientific visualization is a largely unexplored area. Lifton et al. experimented with combining data collected from electronic sensors in their building floor into the Second Life virtual environment. In what they termed “cross realities”, sensed phenomena from the real world are extrapolated, presented, and interacted with in immersive virtual environments, before being projected back to reality through incarnations [Lifton et al., 2009]. They deployed two mobile, handheld platforms to browse and interact with sensors. These applications used a simple map to represent topology of deployed sensors and associated data.

Using marker based AR, Meiguins et al. deployed scatterplots of multidimensional data, ranging from 1D to 3D [Meiguins et al., 2006]. Motivated to explore physical interactions, a menu maps markers to well-known visualization interactions, such as general view, zooming, filtering, and details. Interaction took place by occluding markers associated with menu items. The work from Meiguins uses AR as a user interface and the plots displayed are completely disconnected from reality. Conversely, the work from Nikishkov and Tsuchimoto, based on the same tracking technology, situates visualizations in context displaying tactile data (i.e., measured stiffness) projected in a human organ (i.e., a breast dummy) for breast cancer examination [Nikishkov and Tsuchimoto, 2007]. To monitor real-time information from patients, Claros et al. attached plots from biometric sensors to fiducial markers [Claros et al., 2007].

The same technology was used to inspect ambient information in buildings. Both Rauhala et al and Goldsmith et al introduced fiducial markers at specified locations, to evaluate a concept for location based browsing of information from sensor networks using mobile devices [Rauhala et al., 2006] [Goldsmith et al., 2008].

Noteworthy contributions to outdoor AR for visualization include ARVino from King et al. [King, G.R.; Piekarski, W.; Thomas et al., 2005] and SiteLens from White and Feiner [White, 2009]. In ARVino, King et al. used a tripod mounted AR system to visualize GIS data for viticulture. They noted what they termed the “long flat view” problem, arising when viewing flat virtual objects from a first person perspective. The effect is that due to depth perspective distant objects are relatively small and difficult to see. Although it was not further analyzed, this problem is just an instance of the narrow overview issue described in chapter 7. In SiteLens, White and Feiner experimented with a mobile system to present novel visualizations of CO2 sensor data. Their system limited to static data-sets of a single (CO2) 1D datatype.

A related project for outdoor AR targeted visualization of subsurface infrastructure for utility companies [Schall et al., 2010] (see Figure 2.7). The project dedicated considerable effort to the offline conversion of large spatial databases to generate appropriate AR models [Schall and Schmalstieg, 2008], and to the study of outdoor tracking techniques [Schall et al., 2008a] [Schall et al., 2009]. It provided a frame for the development of the handheld AR platform prototype, described in chapter 4, as well as for initial evaluations of spatial interaction in outdoor AR [Schall et al., 2008b].

In contrast to these projects, chapter 3 defines a fully general infrastructure that addresses the complete cycle of mobile monitoring, and leverages visualization of heterogeneous, static or dynamic multidimensional data updated in real-time.

Chapter 5 addresses the development of visual presentation modes for environmental data in AR. The tools to support the interactive workflow of visualization in outdoor AR are laid out in chapter 7.

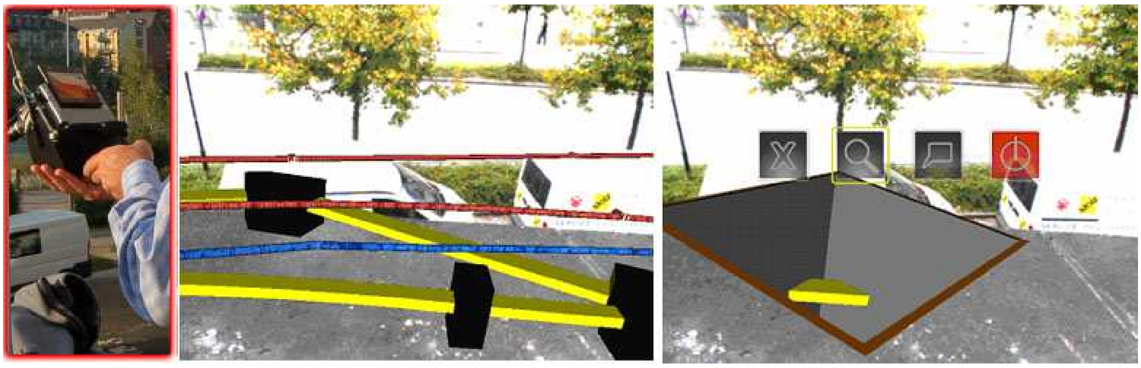


**Figure 2.7:** Vidente Handheld AR for Underground Infrastructure Visualization. Our first handheld AR device had a single handed mode (Left). Simply rendering underground infrastructure over the video stream does not provide depth cues: the pipes appear to float on the street. A virtual excavation tool, one of the visualization tools deployed by Vidente to support depth perception. [Schall et al., 2008b]

### 2.4.1 Limitations of Outdoor AR for Visualization

With traditional AR alone, the workflow of visualization described in section 2.3.2 falls short, as it is restricted in ways of gaining overview or zooming into portions of the dataset. The first person view of AR situates the user within the dataset, but having access to other viewpoints such as the physically unreachable ones (e. g., bird view, peak of a mountain, in the middle of a forest) or remote cameras (static or mounted on drones) would be of great help for the scientist. The inside-out perspective of mobile AR generally restricts visualization in three ways:

- It narrows the overview to the portion of the world captured by the camera.
- Variable elevation in the terrain causes multiple occlusions and the spatial relationships between objects and the environment become unclear.
- There is no way to zoom-in on a portion of the dataset without losing reference to the physical world.

Elmqvist and Tsigas [Elmqvist and Tsigas, 2007] considered general occlusion problems in virtual environments, and presented a taxonomy of techniques for the acquisition of spatial information. Among other interesting findings in their paper, Elmqvist and Tsigas identified four object interactions that define the types of occlusion problems: proximity, intersection, enclosement and containment. We address the first three cases with the techniques presented in this paper, these cases being the most common ones in outdoor environmental data visualization.

Often, mobile AR techniques are developed as technological showcases, without a human factors foundation [Livingston, 2005]. In contrast, we grounded the design and development on user centered design, iteratively performing a range of human factors studies and usability experiments in collaboration with specialists in the field.

Understanding the contents of complex visualizations requires consideration of mental workload and situation awareness. The term situation awareness was coined in aviation, and has found its application in user interfaces [Endsley, 1988] [Endsley, 1995]. The process of acquiring and maintaining situational awareness and its relation with mental workload are illustrated by [Tsang and Vidulich, 2006].

### 2.4.2 Interactive Perspectives in Outdoor AR

In AR, occlusion is a recurrent problem and numerous techniques have been introduced to deal with it in different scenarios. Bane and Höllerer [Bane and Höllerer, 2004] experimented with tools for x-ray vision in mobile AR. They used a static, tripod mounted system to experiment with tools to interactively select depth levels for x-ray vision. Mendez and Schmalstieg [Mendez and Schmalstieg, 2009] and Sandor et al. [Sandor et al., 2010] experiment with techniques to properly convey depth differences between occluders and occluded elements of a scene using x-ray vision. These techniques rely on accurate 3D models of the objects visualized and equally accurate tracking. To convey up-to-date features of a dynamic 3D environment, Kameda et al. [Kameda et al., 2004] and Avery et al. [Avery et al., 2007] used remote cameras to capture the occluded objects. Kameda et al. [Kameda et al., 2004] relies on a static infrastructure of cameras to capture the environment, whose imagery they texture map on an accurate 3D model of the environment using advanced tracking techniques to find texture coordinates. Avery et al. [Avery et al., 2007] used a drone to explore unknown territory. They used a picture-in-picture technique to render imagery of the occluded scene. A difficulty with the picture-in-picture technique when viewing occluded scenes is that it does not convey spatial orientation of the remote camera. This issue relates to human factors derived from situation awareness, as are analyzed in multi-camera systems for surveillance.

Several approaches have proposed integrating multi-camera systems with virtual environments to enhance perception of spatial information. Examples of these works are the contextualized videos of Wang et al. [Wang et al., 2007] and Video Flashlights of Sawhney et al [Sawhney et al., 2002].

We deploy a multi-view system with very little infrastructure, which builds in a rather ad-hoc manner on mobile users and drones. This system allows the user to take different perspectives, showing vantage zoom points in the physical world without having to move to them. Sukan and Feiner [Sukan and Feiner, 2010] allow a user to take snapshots of an object from different perspectives and then use them for overview purposes. They experiment in an indoor marker-based tabletop setup, where an object of interest is completely in view, as opposed to our outdoor mobile AR scenario, where the datasets spread over large areas, and cannot be fully captured by the camera.

### 2.4.3 Combining Perspectives in a Single View

The problem with multi-cameras is that once the user is navigating the 3D representation, the connection with the physical world is lost. We wanted a metaphor that enforces and maintains the connection with the physical world, while enhancing overview and providing correct spatial relationships.

A technique that combines properties of enhancing overview of a scene while dealing with occlusion is called multiple viewports [Elmqvist and Tsigas, 2007], and is commonly found in CAD software. The issue of combining multiple perspectives in a single image has been used as an artistic form to draw panorama maps. Jobst and Döllner [Jobst and Döllner, 2008] studied how this technique enhances perception of 3D spatial relations, increasing overview and information density. In subsequent research, Lorenz et al. [Lorenz et al., 2008] studied how to navigate a virtual environment for tourism when using multi-perspective views. Pasewaldt et al. [Pasewaldt et al., 2011] reported on a authoring tool for the multi-perspective deformation. The interaction with the multi-perspective view itself requires manipulation of visualization parameters and has not been considered up to now.

Kim and Dey used a simple technique to display a road map merging with the real world view in the free area of a wind-shield [Kim and Dey, 2009]. They proved that this technique can minimize issues of divided attention and cognitive load during navigation, while attending to the real driving space and a GPS-based map visualization. Sandor et al. [Sandor et al., 2009] described a space distorting technique to visualize points of interest that are occluded or outside the field of view in mobile AR. Our techniques are conceptually similar, but use a different approach and have a different application setting.

In chapter 7 we introduce a novel technique to combine multiple perspectives. The technique works from an ego-perspective, preserving the real world context, and extending it with overview virtual perspectives. Its effects can be seen in Figure 2.8

## 2.5 Summary

This chapter provides a frame for the topics addressed in this thesis, derived from AR, environmental monitoring and visualization. In particular, it identifies perceptual issues across these disciplines that challenge outdoor AR data visualization, and guide the

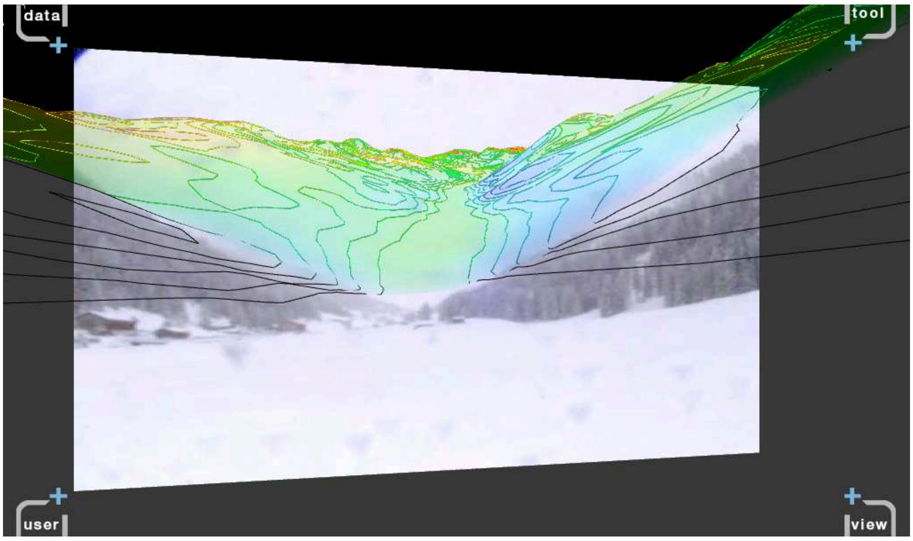


**Figure 2.8:** Combining perspectives in a single view. The variable perspective technique described in chapter 7 combines multiple perspectives in a single image, preserving the connection with the real world.

ideation of novel metaphors described in the following chapters. The basic components of our research can be characterized, based on concepts visited in this chapter.

The next chapter will discuss the topic of mobile environmental monitoring, analyzed together with experts in the field. It will detail its workflow and propose a fully functional infrastructure for deployment along with validation of the overall system.

# Chapter 3

# Mobile Environmental Monitoring

This chapter takes a closer look at environmental monitoring, analyzing characteristics inherent to the task from the perspective of the professional in the field. In particular, the focus is drawn upon the role of mobile tools within environmental monitoring.

This chapter contributes 1) the concept of mobile environmental monitoring, its requirements and workflow along concrete usability analysis, 2) an innovative infrastructure that enables access to a wide range of different wireless sensors networks, deploys multivariate visualization, allows close-to real time data access, directly integrates simulation results, and includes a range of tools developed specifically to create a shared awareness in interdisciplinary teams, 3) usability and acceptance evaluations of the prototype with real scenarios.

To unveil the process of mobile environmental monitoring, section 3.1.1 starts by identifying groups of stakeholders and, based on their goals and activities, the context of use. Thereafter, section 3.1.2 introduces concept and workflow, followed by challenges in section 3.1.3.

## 3.1 What is environmental monitoring

The process of environmental monitoring using mobiles was studied in the frame of a three-year project following a user-centered design (UCD) approach. Thereby, a large group of end-users took part in incremental, iterative phases (participatory design [Schuler and Namioka, 1993]) with around 65 users with various backgrounds, and focused on creating a thorough understanding of the context of use (contextual design [Beyer and Holtzblatt, 1998]).

**Table 3.1:** Classification of stakeholders based on their activities.

| User Group Description | | |
|---|---|---|
| User Group | Profession and Practise | Main Activities |
| Regional authority, water project manager | Engineers and specialists* | Designing and managing monitoring network.<br>Decision making and management of natural disaster.<br>Design mitigation/restoration plans. |
| Environmental company / watershed contractor | Environmental engineers* | Administration of resources.<br>Supervision of construction. |
| Researcher | Environmental scientist* | Deploying long-term monitoring system.<br>Multidisciplinary field work, counseling.<br>Model development and validation. |
| Private companies with infrastructure in the Alps | Entrepreneur, various specialists (civil, electricity,. . . ). | Designing, building and maintaining infrastructure.<br>Responsible of security related to their infrastructures. |
| * hydrologist, biologist, geologist, geographer, forestry | | |

### 3.1.1 User Groups and Goals

Built on a hydrological context of the water cycle model, two user scenarios were selected, targeting wet snow avalanches and watershed modeling. Stakeholders involved in monitoring these events and their activities can be classified as shown in Table 3.1. Each user group has particular reasons to monitor the environment, mostly dictated by their activities. For example, environmental scientists aim at creating mathematical representations of a particular phenomenon (physical models), defining the inputs that trigger certain environmental conditions. Conversely, a company or government institution tries to identify environmental changes and obtain aids for decision making.

Nonetheless, these goals often entail relying on the findings and understanding of others. An implicit goal is to create a *shared* understanding of the environment and discuss potential solutions to problems found. On-site actions should aid in more closely connecting captured process data with its actual context, the environment itself. It is expected that improving the understanding of context improves the general understanding of processes and eases solution finding.

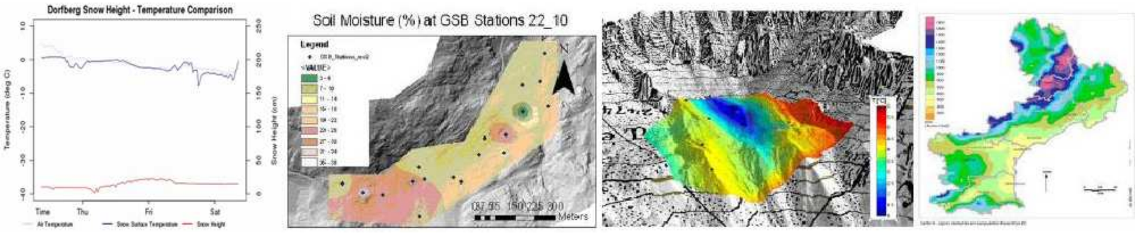


**Figure 3.1:** Traditional Visualizations. Left: combined plot comparing measurements of snow height and temperature over time. Middle Left: color coding of soil moisture values. Middle Right: interpolations for surface skin temperature. Right: color coding of land use.

### 3.1.2 Mobile Environmental Monitoring: Definition and Workflow

Environmental monitoring can be defined as the process of continuously observing and regularly measuring environmental parameters of a specific area in order to understand a phenomenon. Mobile (on-site) monitoring comprises all activities conducted in the field. It does not replace, but complements monitoring and data analysis in the office.

Monitoring comprises several tasks closely interconnected in work cycles, alternating visits to the field with work at the office. Upon identification of a suitable site for monitoring, the data pipeline is setup (i.e., acquiring legacy data, creating, placing and maintaining sensors, preparing storage and network). Site visits serve the purpose to maintain infrastructure, to gather samples and personal observations later examined at the office, or to compare simulation results with reality in a decision process. The latter reflects management of environmental processes, where plans are thought out and potentially enforced.

### 3.1.3 Challenges

Environmental monitoring entails several challenges, some of them with direct impact on mobile activities.

**Site: real-world conditions.** Weather conditions in remote areas pose a serious challenge to devices, as simple consumer electronics are prone to failure in extreme conditions (e.g., low temperatures). This affects sensors, but also devices brought by the scientist. In some cases sensors can be conditioned to work in extreme conditions, whereas in others they have to be removed and redeployed following seasonal patterns. Restricted accessibility locations incur high mobility costs. The challenge is to make the most out of each visit, reducing the need to repeat visits due to unexpected events.

**Data(-source) heterogeneity.** Environmental data sources vary in dimensions, periodicity and other characteristics. Sensors usually produce unidimensional data, with exceptions (e.g., wind direction). However, the periodicity of measurements varies widely from fractions of a second to hours or days.

Numerical post-processing can output estimates of measurements covering an area (i.e., interpolations), or predictions of future states of the environment using complex physical models (i.e., simulations). The output format varies (e.g., single value, ranges, 2D or 3D), and the frequency is subject to the input (e.g., sensor readings) and its complexity. Simple interpolations run in fractions of minutes, while simulations can take days.

Legacy data such as plans, maps and 3D models (DEM, DTM) are not updated frequently, and their representation is only partially faithful (e.g., DEM does not represent man-made structures or temporal characteristics like snow cover or vegetation). The scientist often complements these data with photographs that capture up-to-date information. A noteworthy, problematic aspect of spatial data is that of geographic scale. Finding the right pixel size is often a complicated task [Hengl, 2006]. An AR overlay of spatially distributed environmental variables supplies valuable information to assess pixel size, or even to suggest a scale for a phenomenon.

**Data acquisition.** The acquisition involves all stages from when data is gathered until it is delivered to the user. These stages vary for different data types. Automatic sensors are deployed in the required density to gather frequent measurements. Conversely, data from manual sensors is collected at locations of interest, pre-planned or identified ad-hoc by the scientist, and uploaded later.

All data must undergo sanity and quality checks to ensure validity, and security mechanisms enforcing privacy restrictions. Data may then be stored and indexed for future access. Legacy sources are not often indexed, and must be obtained from a central repository. Numerical processes can be automatically triggered upon availability of input, or manually configured depending on the model. For automatic processes, it makes sense to store and index the output. Complex simulations require powerful dedicated servers, and each has different ways to specify input and access to results.

In remote sites, the infrastructure needs to account for network availability and latency. At the same time, the data infrastructure must allow seamless storage and retrieval of data with different characteristics.

**Interpretation.** Numerical data are converted to visualizations that, based on perceptual abilities of people, provide an advantage to detect patterns, differences, connections or similarities in numerical data [Shneiderman, 1996]. Typical visualizations for environmental data include simple and combined plots for point data, interpolations of measurements from several sensors, color coded 2D diagrams or maps (Figure 3.1). To form a clear picture about the situation, the scientist needs to assimilate all these data that differ in dimensionality, update rate, and representation. Often, this process entails switching between representations (plots, visualizations, numbers), and correlating with the real world (maps, models, photographs) to understand where effects originate.

Monitoring and management actions require communication support, whereby users with potentially various backgrounds discuss findings and potential solutions.

## 3.2 Approach: In-context visualization of environmental data

The ultimate goal of mobile environmental monitoring is to visualize abstract data, such as sensor measurements and simulation results in the context of their occurrence. We describe below two general concepts to support this approach.

### 3.2.1 The notions of context and situation

While monitoring events on-site, users operate within a certain context, entangling the various participants, the actual environment with its artifacts, and high-level conditions, such as weather or noise. Context and situation are overlapping concepts that refer in general to the action-space within which a user operates. We are interested in situation awareness, which encompasses user understanding of the current action-space. Context is important to consider the characteristics of the action-space: physical processes are being observed, that originate in an environment, which may be heavily under change. This notion is also the crux behind the infrastructure presented in this chapter: visual representations of environmental data are placed in the context of the actual environment.

The spatiotemporal characteristics of the environment are of utmost importance for creating a correct understanding: bringing representations of physical processes in relation to spatio-temporally outdated or only partial representations of the physical environment may lead to the wrong interpretation of the situation. Within the boundaries of update rates, data always refers to the latest spatio-temporal stage of development of this environment. Assuming that the office is not in the direct vicinity of the environment being observed, this separation may well lead to misinterpretations when environments are under heavy change, or only limited representations exist. Ideally, we want users in the field to obtain the latest sensor data without manual intervention. All this must happen in a collaborative framework whereby users can interact, communicate and discuss findings and potential solutions.

Hence, our aim is to create (a) a correct per-user understanding of the data representation in relation to the actual environment in its latest spatio-temporal state, as well as (b) a shared understanding of the knowledge gained by users with potentially different backgrounds and different perspectives on the site.

### 3.2.2 Visualization using mobile AR.

The method of choice, augmented reality, strives to render computer generated artifacts correctly registered with the real world in real time. The term “correctly registered” means that these artifacts appear in the correct position relative to the point of view of the user. We propose a mobile AR application for visualization of environmental data in the context of the site of study. This in-context interactive visualization supports the scientist in monitoring tasks, from displaying sensor positions for deployment or current readings for maintenance, down to integrating representations of multivariate data and complex simulations. With AR, complex data sets are contextualized within the environment they originate. Similar to a multi-layer approach, multivariate data can be compared and analyzed through a combination of representations (mixed dimensionality).

To support situation awareness, the in-context capabilities of the system are extended with multiple perspectives on the site, deploying and sharing footage from multiple imaging devices. This multiview infrastructure improves understanding of spatial relationships and provides a solid basis for cooperative work.

### 3.2.3 Requirements

The aforementioned challenges outline requirements for a mobile AR system that necessarily differs from general mobile applications and from in-office GIS systems.The main requirements include:

**Geo-referenced data:** all data gathered must be geo-referenced to generate in-context visualizations.

**3D digital model:** terrain models are needed for visualization and simulations.

**Network access:** although it is possible to use prerecorded data, the full potential of the application comes with real-time sensor/simulation updates and communication with other users.

**Accurate tracking:** high accuracy estimation of the user’s view (pose + video) require high-quality sensors such as camera, orientation sensor and GPS.

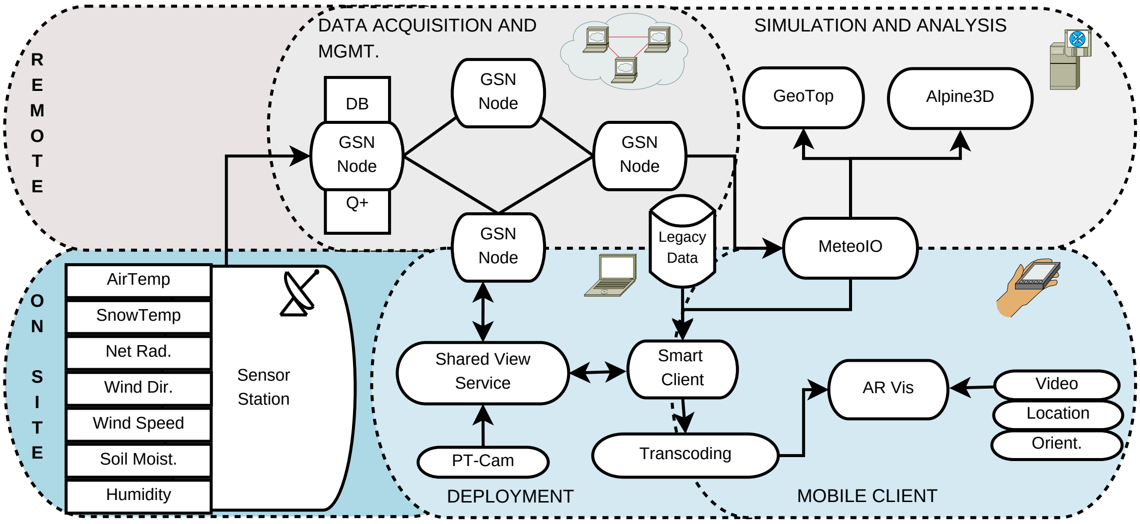


**Figure 3.2:** Overall system diagram. Component-blocks are organized according to their deployment (i.e., on-site, remote). Icons at the top-right of each block indicate the (usual) platform running the components (i.e., dedicated sensor station, distributed systems, mainframe, laptop, handheld device).

**Interaction:** visualization and user interface should be clearly visible, preferably show undistorted colors, support correct depth perception.

**Robust mobile system:** the mobile platform needs to be robust, ergonomic and have an outdoor readable display. The interface must allow usage under extreme conditions (e.g., wearing gloves).

## 3.3 System infrastructure and workflow

Our infrastructure is divided in five main components, each presenting clear interfaces to distinct services (see Figure3.2). The design relies on a data flow model for communication and interoperability. Our system leverages on-site/off-site communication and data exchange, supported by a network layer underlying the system components. *Sensor and acquisition* components manage sensor measurement, filtering and aggregation. Furthermore, *simulation and analysis* components furnish additional scientific data-sets that can be used on-site. The mobile support is divided in *deployment*, providing infrastructure to access data (environmental and other) and prepare it for rendering, and the *run-time client*, providing visualization and interaction capabilities.

### 3.3.1 Network Layer

Our system follows an heterogeneous configuration of wired and wireless network. Internet and LANs provide off-site access to sensor information such as databases or web portals, while wireless links support sensors and the mobile interface.

Interactive mobile monitoring requires a network with high availability and real-time refresh rates. We rely on the mobile phone network (GSM) where available. A special, low powered WiFi bridge was developed for remote areas, where connections are degraded or entirely absent. It operates either in a multi-hop setup, using more than one link to extend WiFi connection to the sensing location, or connected to the mobile data network in an area of good reception. Links are mobile and can be set up in a relatively short amount of time.

### 3.3.2 Data Acquisition and Management

This component encompasses sensor measurement, transmission, and low level data management (physical and logical data services). Our system unifies the hardware and software aspects of sensor acquisition and management, offering the possibility to plug in various sensor networks (e.g. Pachube, OGC SWE[1] or ad-hoc solution [Yang et al., 2009]).

The current implementation relies on the Global Sensor Networks (GSN), a distributed sensor network middleware, developed to provide uniform, ubiquitous interface to a large deployment of varied sensor types (i.e., mechanical, electrical, thermal, digital/analog, simple or multi-variate) [Aberer et al., 2007]. Similarly to OGC SWE standards, GSN supplies a multi-level architecture (implemented in Java), offering the possibility to describe, aggregate or filter sensors. GSN stores measurements from automatic and manual sensors, it employs parametric and probabilistic methods for quality control, and supplies federated access to data [Jeung et al., 2010]. It offers temporal and spatial queries, as well as dynamic registration for push/pull data retrieval. It also processes plot queries, resulting in 2D graphical content, avoiding transfers of a high volume of data over potentially expensive network links. GSN uses eXtensible Markup Language (XML) for data exchange and interoperability. We use this format to communicate and interact with sensor measurements and between the components of our infrastructure.

### 3.3.3 Analysis and Simulation

Collected and filtered sensor measurements serve as input to data analysis and simulations. These components and subsystems generate advanced data-sets that can be used and visualized on-site.

---

[1] http://www.opengeospatial.org/standards/swes

A generic library aids in the conversion of sensor data to scientific analysis tools such as Alpine 3D (A3D) [Lehning et al., 2006], Matlab, etc. The library MeteoIO[1] presents a uniform, format independent interface to integrate meteorological data in an application. It is plug-in based, and enables easy access to different data sources (e.g., GSN, A3D, GeoTop) and to hydrological simulations. MeteoIO performs interpolations of meteorological parameters, and includes plug-ins to access simulations in real time (e.g., GeoTop [Simoni et al., 2008], Snowpack [Liston and Sturm, 1998]). To run a simulation, the library obtains all the parameters and the specification of data (GSN nodes, data types, etc.) from a file. MeteoIO then retrieves data from GSN and feeds it to the appropriate model via a plug-in. Upon reception, results can be converted to other formats (transcoded) using filter plug-ins. Thus, 2D data-sets can be transcoded to image formats, later used by the mobile client.

### 3.3.4 Deployment

To implement our awareness concept, we introduced the notion of on-site campaign services deployed to support on-site activities, in particular collaboration. During campaigns, different services run on a portable computer at the site of study and provide access to shared information (e.g., remote views, annotations).

A shared view service (SVS) supports the notion of multiple perspectives on the environment. It allows users to look through the "eyes" of a collaborator. The SVS also shares views from pan-tilt and infrastructure cameras, enabling remote observation of the location. To this end, SVS can connect to GSN to retrieve images from cameras deployed as part of the fixed sensor network. Moreover, it can relay images to GSN, which can in turn be used remotely. Finally, the SVS can be extended to record geo-referenced annotations linked to snapshots from mobiles.

The SVS is based on a publish/subscribe model. A hand-shake protocol acquires identity of clients and information about the camera-lens system (i.e., intrinsic parameters). Thereafter, clients regularly supply tracked frames (i.e., jpeg image + 6D position). Clients can subscribe to frames from any camera. As the SVS is deployed on a dedicated computer, we avoid network overload and latency compared to running directly in GSN. The SVS uses its own on-site ad-hoc network (private WiFi) guaranteeing close to real-time performance (dependent on frame update rate and resolution).

As our on-site computer is connected to GSN, a data service can cache incoming sensor/simulation data for all users, reducing outbound traffic. This component leverages data conversion to assure interoperability with the client. Incoming geo-referenced data undergoes geometric conversions, transforming its reference frame to that of the ap-

---

[1] http://slfsmm.indefero.net/p/meteoio/

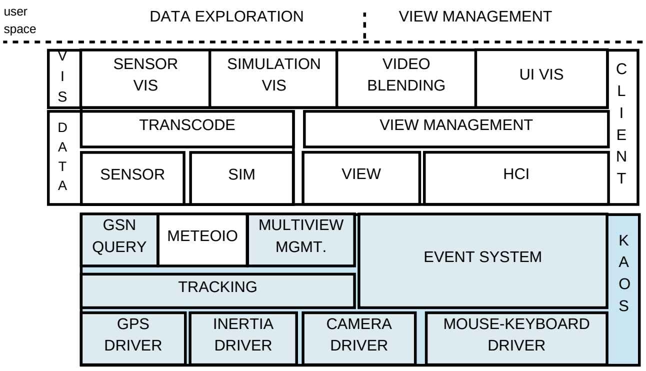


**Figure 3.3:** Block diagram of the mobile client.

plication (i.e., projection WGS84–UTM, and reference adjustment). Furthermore, 3D simulation results are transcoded, obtaining geometry as a regular polygonal mesh and 2D geo-registered overlays (e.g., textures) with corresponding coordinates. These conversions occur in accordance with the data format supported by the client and considering the mobile platform capabilities (downscaling large polygonal models, downscaling image content, etc).

### 3.3.5 Mobile Client

The mobile client defines the interface for end-users. It delivers access to the data on-site, as well as visualization, interaction, analysis or reporting.

**Hardware Platform.** The target platform is a Panasonic CFU1 tablet PC (Intel Atom CPU Z520 1.3 GHz , 1 GB, Intel GMA500). Although its processing power is relatively limited, the platform is ideal for outdoor, rough environments: it is ruggedized, waterproof, has a long lasting battery life complemented with two hot-swappable batteries, and a 5.7” touch-screen. The system is complemented with a UBlox differential GPS, an InertiaCube3 or XSense tracker, and a UEye camera with a Pentax 4.2mm wide angle lens.

**Software Platform.** Our client is built on a specifically designed mobile infrastructure, KAOS, which supports multi-threaded plugins and an event based communication mechanism. The KAOS infrastructure has been optimized and developed to handle low level mobile and portable devices.

Above the KAOS framework we implemented our specific mobile AR/VR client (see Figure 3.3), handling graphics rendering (OpenGL), graphical user interface, AR registration and audio communication (Skype API). We made prolific usage of the plugin

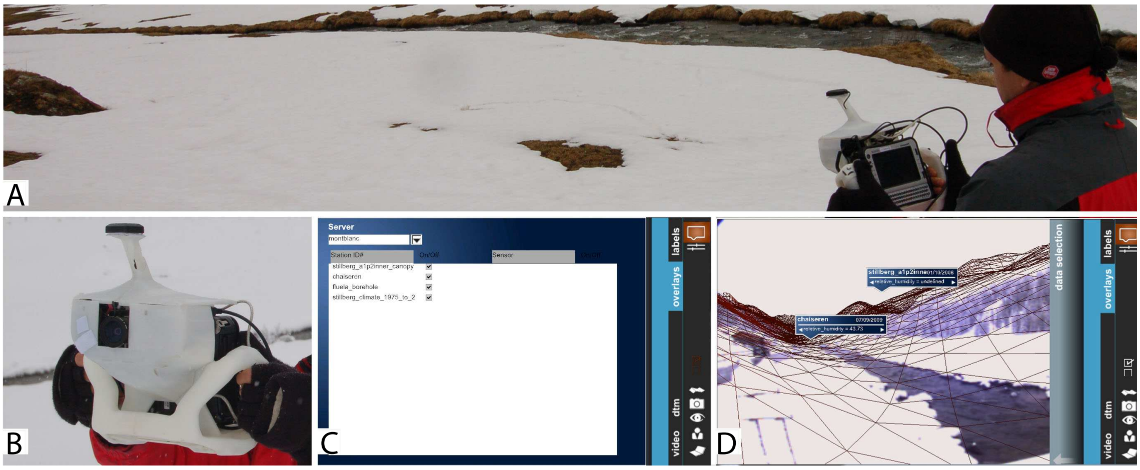


**Figure 3.4:** First Prototype: the first prototype evaluated during expert workshops displayed sensor label and a wireframe DEM.

mechanism to support a variety of tracking technology, but also for handling and importing different types of data-sets.

For tracking, we developed a standard inertial+GPS plugin as well as an advanced hybrid tracker. We also developed a simulator/tracking recorder leveraging the possibility to debug and test the interactive platform off-line.

## 3.4 Interactive Tools for Mobile Visualization

Visualization, more than pure viewing, is an interactive task, defined by a workflow of well-known interactions such as overviewing, zooming, filtering, and obtaining details [Shneiderman, 1996]. The evolution of the tools envisioned for visual presentation of heterogeneous data in AR is the focus of chapter 5. The iterative development of novel tools to enable the visualization workflow in outdoor AR will be discussed in chapter 7. Each of these tools and techniques have been evaluated independently and the results are presented in the corresponding chapters. Instead, this section gives a preview of the features included in each of the three functional prototypes that were showcased and evaluated by experts users in incremental stages. Although each prototype was fully functional, the presentation, interaction and functionality matured incrementally with each stage.

The first prototype was based on a robust platform created to withstand abuse in extensive demo sessions, see Figure 3.4 B and section 4.4 for a detailed description. It boasted an animated user interface based on a flash plugin that enabled OpenGL rendering. While compressed, the UI took the right hand side of the screen unused by the video background (i.e., $4/3$ camera vs $16/9$ screen resolution). In this prototype, each label represented a

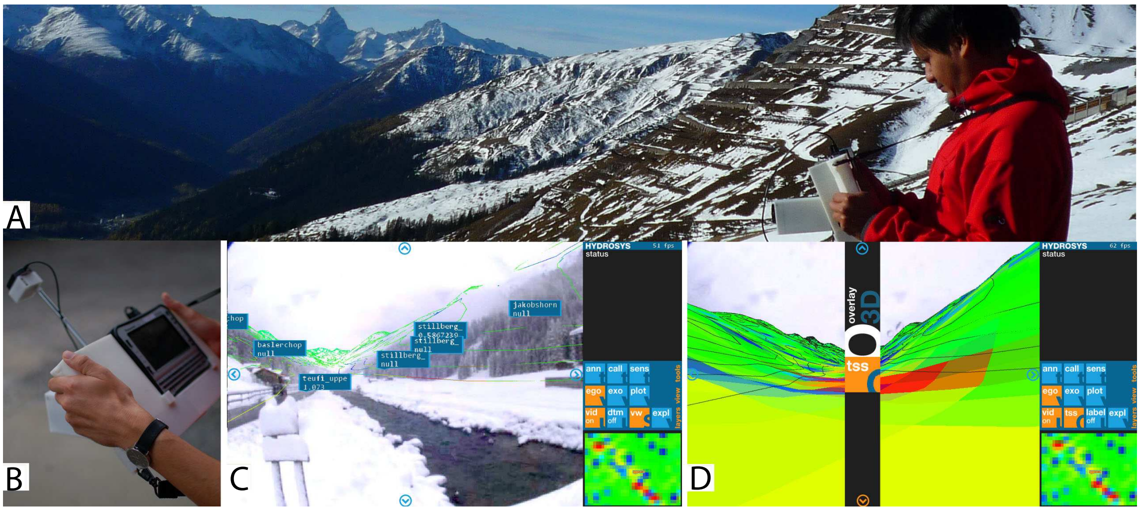


**Figure 3.5:** Second Prototype: the second prototype, showcased in expert interviews, enabled selection from a variety of sensors/simulations, showing line or fill overlays.

station that could be controlled independently to display the selected data type. Thereby, each label could display data for a different sensor in the station, showing for example temperature in one sensor station, and relative humidity in another. This feature proved completely counterintuitive and difficult to control. The prototype was capable of displaying 3D data in wireframe or fill mode. However, it was not possible to combine these with 2D overlays of data. Technical limitations plagued this first implementation, due to poor graphics support, texture operations were prohibitive and its performance could scarcely be called interactive, running at 8fps at its best. Still initial evaluators were tolerant and envisioned high expectations both for the technology and approach. The prototype was evaluated in expert workshops described in section 3.6.1.

The second platform retained robustness while having a compact design with little movable parts, as shown in Figure 3.5 B and detailed later in section 4.5. Functionally, the prototype displayed labels, this time for a single, selectable type of measurement in all the stations, e.g., all the labels displayed temperature (if available). A table with details per station was available upon selection. Overlays were finally available, mapped onto 3D representations. In addition, this prototype introduced a new form of 3D data, that represented a terrain using contours and iso-lines. Data selection was based on a crossbar user interface, that allowed to change data type for labels and overlays independently, and several switches. Technically, the poor performance of the previous implementation prompted a full redesign of the client, which adopted the structure already described in section 3.3.5. All the graphical items, including the user interface, were redeveloped using advanced OpenGL features, such as vertex buffer objects and shader programs. Stress tests showed that the prototype could handle large amounts of data, over 100 labels and

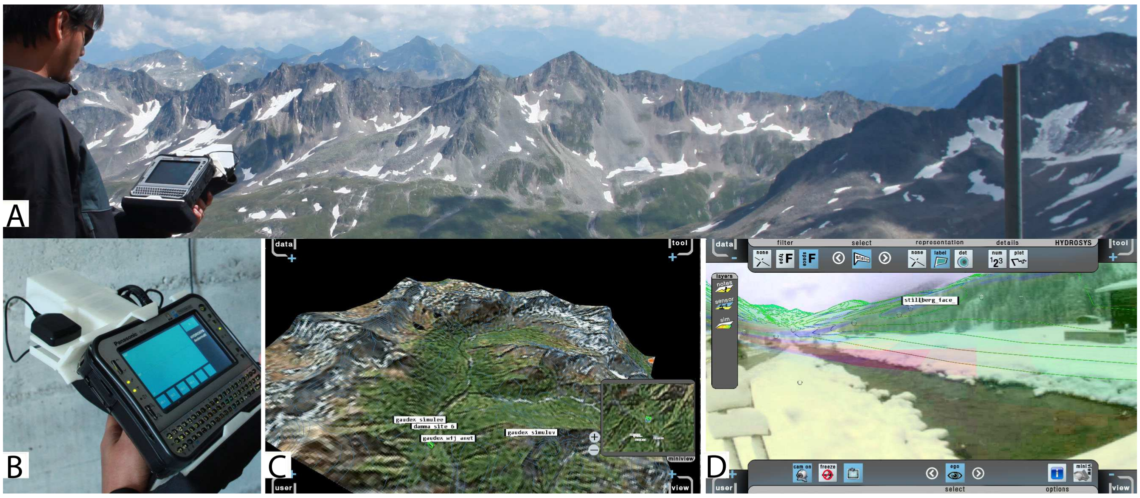


**Figure 3.6:** Final Prototype: the third prototype, used in the formal evaluation, applied a workflow oriented UI to organize visualization activities, and introduced transparent overlays, multiview and collaboration tools.

large datasets at interactive framerates (24–30 fps). The prototype was evaluated subjectively in showcase interviews described in section 3.6.1. The platform for the last prototype was lightweight and modular, and maintained robustness by combining different materials in the construction, see Figure 3.6 B and section 4.6 for more details. The prototype enabled novel features that allowed users to take different perspectives on the site without moving physically, thus enabling zooming and overviewing the datasets. It also enabled transparency control for overlays, and a handful of collaboration tools. A novel UI structured around the monitoring workflow extended the functionality, described in section 3.5.3. A final, formal evaluation validated the prototype, its usability and functionality, as shown in section 3.13.

## 3.5 Towards a Workflow Oriented User Interface

This section takes a look at the user interfaces evaluated at different stages of validation (see section 3.6). In the particular case of outdoor AR, as the actual screen space and resolution restrict the amount of data that can be visualized, the role of interactions gains added value. The interface has to be lean and intuitive, to disappear behind the user's workflow, and let her experience the data. The challenge is to find an underlying methodology, a logic to organize activities in metaphors and techniques that reflect the "way of thinking" of the user for handling data in the field. In our outdoor application, workflow foundational tasks of Shneiderman's information seeking mantra [Shneiderman, 1996] are orthogonal to the functional activity groups afforded by the application:

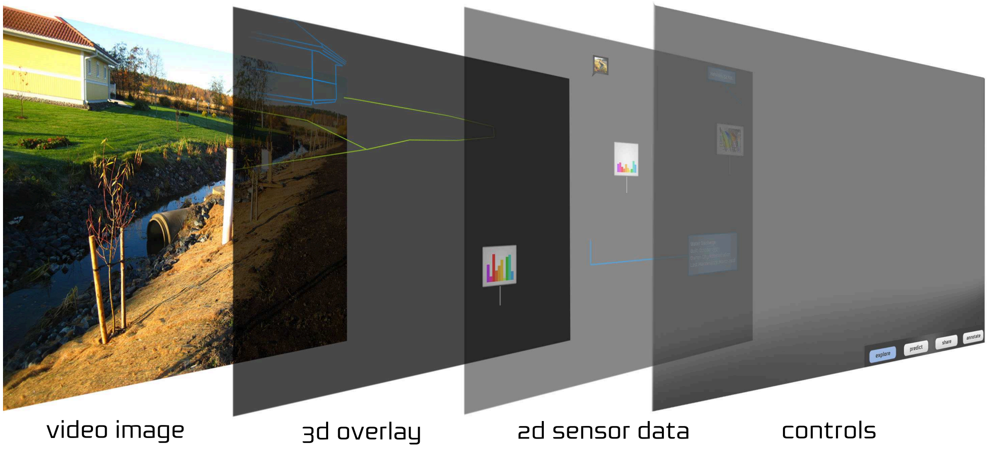


**Figure 3.7:** Concept of Layers as Organizational Units. Initial concept of layer for organization of activities with the video as the first layer, followed by data layers –overlay and sensors– and even a layer for controls.

data exploration, view management and collaboration, presented in chapters 5, 7, and 8 respectively.

Initially, the logic in the workflow of the user was captured in a range of diagrams (flow charts, UML use-cases, state, and transition diagrams). The process served to uncover and document the internal structure of user activities. Hence, an interplay was identified between possible metaphors and the higher level logic.

At first, the organization of the user interface followed this workflow only loosely, but it was more guided by functionality. As functionality was made available, the user interface design had to be revisited time and again. This helped maintain a lean user interface at all stages, without having to go through loads of empty options or menus in early stages, when functionality was scarce. Towards the final version, the application was more functionally complete. The final version closely reflects the workflow captured in early diagrams.

### 3.5.1 Early Prototypes and Concepts

The user interface in our mobile AR visualization system has to organize available graphical representations in an effective way, supporting unobtrusive interaction using minimal UI elements where possible. In early definition phases, a layer metaphor was the guiding organizational technique (see Figure 3.7). Although it was only functionally realized later with the multi-pass renderer, from the beginning, layers logically organized data exploration activities by grouping tools and activities corresponding to data. We hypothesized that we would have a number of data layers associated with 3D or 2D representations

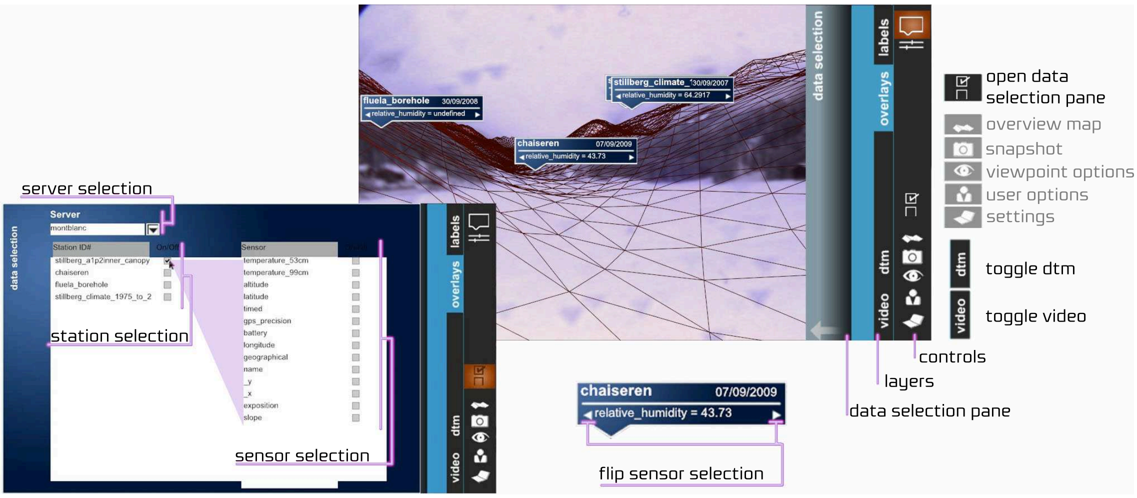


**Figure 3.8:** Flash Animated Interface. This prototype had an animated interface. A visual feedback shadow (orange block in the figure) followed the cursor in the menu and the data selection panel slid in-out upon invocation. The labels were also rendered by the UI. Arrow buttons in each label let the user change the datatype per label.

for which the main activities would be: data association, appearance tuning (e.g., opacity change for overlays), filtering, reordering (moving a layer up or down a stack), image based optimizations (e.g., edge extraction). At the extremes, there would be the video and control layers that could not be reordered [Kruijff et al., 2009].

The first prototype integrated the most basic components needed for on-site monitoring, covering functionality from very basic data retrieval modules bound to early visualization efforts. Thereby, one could visualize raw sensor data and a wireframe 3D model of the environment. The AR mode occupied $800 \times 600$ of the screen. The remaining $224 \times 600$ were available for the user interface, which divided in controls, layers and data selection, refer to Figure 3.8. Besides toggling overlay and video on and off, the main functionality in this prototype was on data selection.

Data selection was based on cascading checkboxes, and a per-station data selection. For each station, one label was shown, and the user could change the source of data individually. During the evaluation in expert workshops, data selection turned out to be tedious and confusing. Users wanted a quick start with default settings that they could tune later. Hence, in subsequent prototypes, we moved technical configuration of device and data servers, as well as the site and static data sources to an XML profile, created once and reused for regular site visits.

From a technical standpoint, a Flash extension to Studierstube [Schmalstieg et al., 2002] drove the prototyping for this initial interface. Ironically, fast prototyping turned to slow pre-development and slower performance, as we had to develop a set of Flash-based

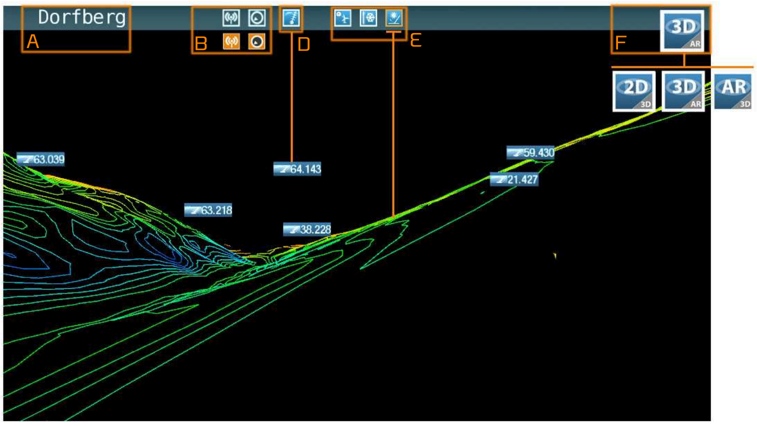


**Figure 3.9:** Icon Only Interface. This test-prototype used only icons. Except for the name of the site (A), there were state icons (B) showing network and compass availability (plain) or lack thereof (orange), a sensor reading icon (C) showing temperature labels, overlay selection icons (D) for wind speed, snow cover, and solar wave radiation, and view mode icons for 2D map, 3D view and AR view.

components from scratch to provide basic UI widgets, for the plugin supports only an older subset of Flash, and its runtime considerably lowered the frame rate.

A new engine, mentioned in chapter 5, opened up numerous possibilities to try new interfaces. We tried a purely iconic interface. It used the top of the screen to organize status and selection icons for the type of information available, e.g., sensors or simulations, as well as for view modes. This prototype was tested internally for performance and only reached a small number of end-users. Experimentation showed that the icons were too small to be tapped with bare fingers, not to mention with gloves. Besides, we soon realized that the interface would not scale with the number of data types, and soon it could not accommodate the icons. But its main drawback were the icons. First, as there is no convention for the graphical representation, we needed to come up with new icons for every sensor reading or overlay. Second, poor legibility made the small icons confusing.

### 3.5.2 Data-Driven User Interface

The second generation UI used a “cross menu bar”, with horizontal and vertical entries. In a sense, it resembles common approaches in consoles, such as the Sony XMB menu[1], but in a slightly simplified manner. The cross-bars organized functionality around datatypes in two independent dimensions: 1D/2D data (horizontal), display labels and plots, and 3D data (vertical), display texture overlays over the terrain. The cross menu bar scales well

---

[1] `http://en.wikipedia.org/wiki/XrossMediaBar`

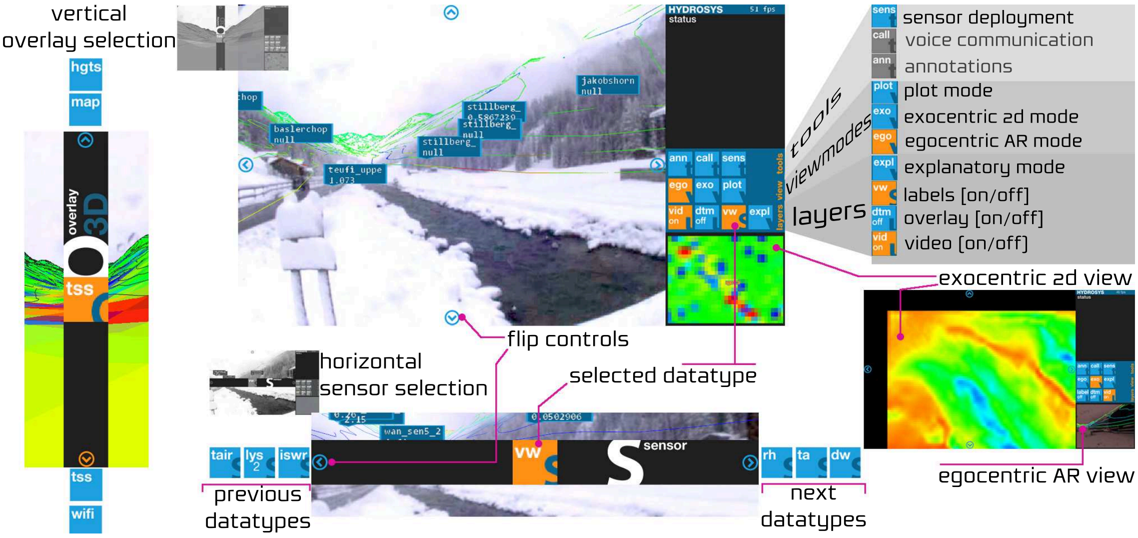


**Figure 3.10:** Data-Driven Interface. Data selection using the cross-menu drove the modes in this interface. Organizational layers were packed in a visual feedback panel together with tools and view modes. The switches in the panel served to activate functionality.

with increasing number of entries, as it only needs to display the current selection and directional arrows. The menus are easily accessible with limited controls, and compatible with the control structure of the UMPC. Figure 3.10, shows the flick controls that serve the interface.

The interaction with this interface is mainly data driven, in the sense that changes in data selection modify the appearance of the interface. The main control, the cross menu bar, changes selection for the principal data sources. In this version, the overlay representation was always on, either as lines or as in fill-mode.

An info panel provided visual feedback and organized the workflow. Functionally, the interface already started to organize the workflow into layers of data, view modes, and tools. Layers include the video (on/off), overlays (on/off in panel, type in vertical cross-bar), sensors (on/off in panel, type in horizontal cross-bar), full data collection (on/off in panel). The latter collects all readings for the selected sensor type in a full-screen table. In addition, two view modes were supported: the AR mode, and the exocentric mode, a top-down view on a simplified dataset (no elevation). Sensors could query plot modes for time series or comparison. Tools included sensor deployment, that displayed the expected location of the next sensor in a sensor deployment plan, calls, and annotations (non-functional at this stage).

The interface was evaluated during showcase interviews and initial deployments in case studies, as described in section 3.6. Expert users gave positive feedback on its functionality. But there was plenty of room for optimization, particularly regarding interface

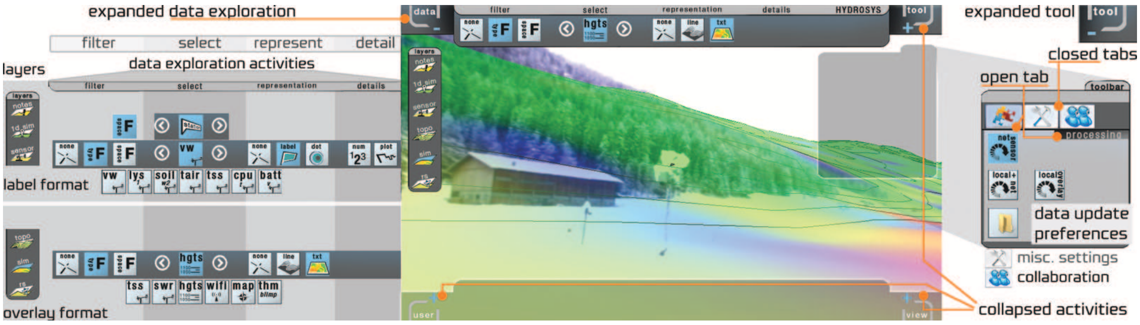


**Figure 3.11:** Guided Exploratory Visualization Interface. The most complete interface in terms of functionality. Functional categories, activated by corner buttons, group workflow activities. The Data category activates the guided visualization toolset, organized following the visual information search mantra.

dynamics: it seemed during evaluation, that users were unable to find various modes of operation even when all the mechanisms are represented in the side panel.

For the final interface, having increased functionality in all aspects, we would have to come up with a structure to guide the user through the possible options while remaining unobtrusive.

### 3.5.3 Guided Exploratory Visualization

In previous examples, the availability of functionality triggered fundamental changes in the user interface, the final stage saw the introduction of most new functionality, pushing the requirement for a more structured approach. The conventional visual types added novel options for filtering, visual, categorical and spatial filters for sensors described in section 5.1, as well as lines, fill and transparent modes for overlays. New types of visual representations were added to reflect features of the data pipeline, e.g., annotations, 1D simulations. Collaboration tools such as annotations and voice communication were finally deployed. On top of this, view management was brought to a whole new level, with a multiview system to improve situation understanding, described in section 7.1.

The three grouping concepts – layers, view, tools – in the panel of the second prototype (Figure 3.10) were promoted to full blown functional categories, each subsumed by a collapsable menu for data exploration, tools and view respectively. View management is described in chapter 7 and collaboration in chapter 8. A fourth user management category was also added. In this interface, layers turned into the structuring tool for exploratory visualization. Note that the main data region spans the full screen resolution 1024×600px, and the UI comes on top when visible.

Data exploration occupies the left and top icon bars (Figure 3.11). Layers, the organizing scheme for data exploration, group data categories in: RS (remote sensing, such as aerial or ground photography), sim (simulation data), topo (topographic data), sensor (sensor data), 1d sim (one dimensional simulation data), and notes (annotations).

Each layer has properties and options that can be categorized logically to reflect a workflow reminiscent of the visual information seeking mantra. These options are categorized in the top menu for the selected layer. Users can apply specific data filters, or select the data type they want, the way they want to see it displayed (for example, a station can be shown as label or as dot), and request details where available. The methods for different filters and visualization modes are described in sections 5.1 and 5.2.

Such careful consideration of the visualization workflow in categorical functional units induces an intuitive structure that unobtrusively guides the user through the dataset: users do not need to search for functions which are logically represented and structured. The collapsible bars associated with each category keep the screen clean, and their position at each corner of the screen clearly identifies them as disjoint functionality groups.

This design was validated in the formal evaluation described in section 3.6.5. In the evaluation, users without any experience with previous prototypes could easily perform tasks on data selection and interpretation. The user interface went unnoticed and did not receive critique.

## 3.6 Validation and discussion

This section presents the methodology, different application scenarios we explored along the duration of the project and the lessons learnt. Two different scientific areas were targeted for our case studies: snow sciences and hydrology sciences.

### 3.6.1 Methodology

We used different evaluation methods and UCD tools to validate our system: active usage by partner experts, showcases at expert workshops and conferences, formal human factors experiments for individual components. The process followed a systematic approach, where top-down analyses helped pinpoint issues, subjected later to bottom-up studies in the frame of the proposed solution. Experts participated in this user-centered development approach with different levels of involvement. A group of ’partner’ experts (15 people, representing groups in hydrology, sensor technology and sensor networks) participated closely in every stage of development. They helped to select a second group (16 people, stakeholders from government entities and companies), appointed as ’advisory’ experts, who participated sporadically at specific development milestones.

**Prototyping cycle.** A definition phase in the first six months of the project involved numerous surveys, interviews and discussion with a large group of experts (52 people) outside our partner group. These sessions helped establish the role of mobile AR in monitoring activities, and define conceptual tools. During a development phase, we worked closely with partner experts to define and implement tools and overall infrastructure. Conceptual tools were studied and validated from a human factors and ergonomics perspective, but also to analyze phenomena from deployment sites, and conceptually re-defined where appropriate.

**Expert workshops.** Two expert workshops were scheduled to showcase the first functional mobile AR prototype (Figure 3.4-A). Firstly, in Davos, Switzerland, 12 end-users from research and environmental organizations were introduced to mobile monitoring concepts, scenarios, and the prototype. A two-hour (plus) intense discussion followed the presentation. Secondly, as part of a large hydrology workshop in Lahti, Finland, 40 end-users from research, government/municipalities, and companies went through workshops in groups of 3-5 people. After a brief introduction, each group took 10-15 minutes to collect thoughts and notes for a longer discussion. Feedback from the workshops helped refine, and in cases re-define, tools and components of the system.

**Showcase interviews.** The prototype in Figure 3.5-B was subject to showcase interviews. During the interviews, participants were demonstrated the main features of the system, asked to explore them, then we collected their feedback through an unstructured interview as well as a questionnaire. Additionally, 7 experts from snow science with varying backgrounds (some regular site visitors, others mainly office-based) were also interviewed.

### 3.6.2 Case Studies

Two long running scenarios were chosen as a testbed for our system. The snow science scenario was tested in Davos (Switzerland) in the context of snow avalanches (in collaboration with SLF and EPFL). The hydrology scenario was explored with Aalto University in the context of watershed modeling in Kylmäoja (Finland).

#### 3.6.2.1 Snow Avalanches in Dorfberg, Davos

**Context** Wet-snow avalanches are significant, frequent hazards in mountainous regions which have a large degree of potential damage to infrastructure and residents ( [Mitterer et al., 2009] [Mitterer et al., 2011] [Techel et al., 2011]). Nevertheless, the formation and triggering of these avalanches are very complex and poorly understood ( [Schweizer et al., 2003] [Baggi and Schweizer, 2009]). The Dorfberg is a steep south-east facing slope above the city of Davos (Grisons, Switzerland) which is an ideal study site for wet-snow avalanches. The scenario aims at improving the understanding of wet-snow avalanches and their related processes. The system described in this paper was applied in order to test its applicability in real case scenarios in remote terrain and rough weather conditions, like those found in Dorfberg.

**Implementation** The site has been equipped with numerous automatic meteorological sensors including a full weather station (e.g. temperature, wind speed, solar radiation, snow depth), and several sensors to measure snow-cover and soil characteristics like snow temperature or soil moisture content. The site is observed from different perspectives with time-lapse photography and researchers visited the sited regularly to perform manual field measurements and to maintain the sensors. To provide reasonable network coverage, a multi-hop WiFi link was deployed, and all sensors have been integrated into GSN. Using the sensor data as input, Alpine 3D, an advanced spatial numerical model ( [Lehning et al., 2006]) was run in order to produce spatial information of the interaction between meteorology, terrain and snow cover. Moreover, simple spatial interpolation methods were applied for the data. Sensor data and simulation output could be queried by the platform in near real-time, and the variable could be displayed while being on site.

#### 3.6.2.2 Watershed Modeling in Kylmäoja

**Context** The catchment of Kylmäoja is located in the south of Finland. The area is drained by Kylmäoja, a stream formed by the merge of three branches close to its geometric center. The catchment features clayey soils that closely follow the valleys, with a topography dominated by moraines structured as shallow layers on short rocky hills. The site scenario is a pond and its neighborhood, spanning 391 $\times$ 390 meters, and yielding a surface area of 15.24 hectares. The land use in the site is highly heterogeneous ranging from grassland to forest and containing natural and man made features.

**Implementation** Sensors were deployed in key points along the Kylmäoja stream, spatially arranged to optimize their representativeness (drained area). On-site campaigns aimed at the collection of data representing specific qualitative (turbidity, conductivity) as well as quantitative (water level) hydrological parameters, and familiarization of end-users with the site towards the evaluation of the system. The parameters selected for sampling can be used for several environmental hydrological models, and are simple enough to be understood by non-expert users when evaluating and using the system. The sampling was performed using commercial Luode water sensor stations equipped with a

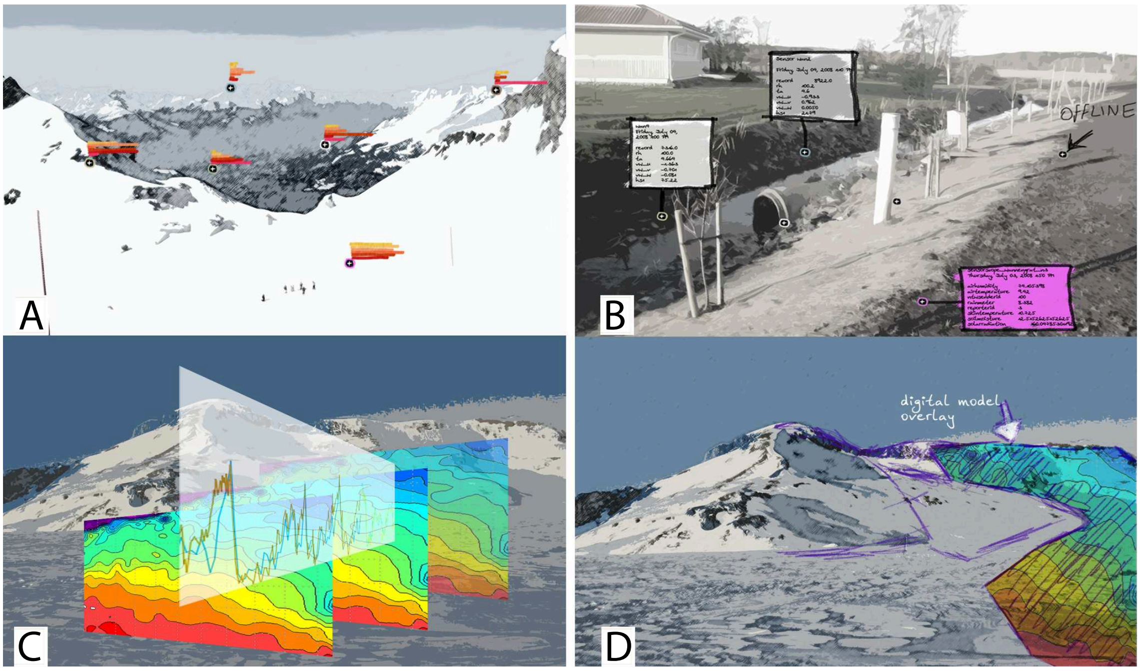


**Figure 3.12:** Conceptual Visualization Designs. Top: Sensor representation. A: miniplots show data distribution of sensor over time. B: labels present observations from all the stations. Bottom: Overlay representation. C: A series of 3D overlays to show temporal measurements. D: 3D spatial overlays mapped on the terrain.

submersible optical YSI-600 water quality sensor and a pressure gauge.

The 2d hydrological model used for the campaign was r.sim.water, a spatially distributed, dynamic 2D hydrological simulation model based on the continuity equations solved by Green's function Monte Carlo method [Mitas and Mitasova, 1998]. The simulation is based on a 70mm/hr rain event and a value of 0.13 for Manning's coefficient to account for the effect of land use. The model was run with a ten minutes time step for a period of an hour. To accurately represent the site, an average value obtained through empirical evaluation of 30 cm was added to the final water level to account for the depth of the pond in raster cells located inside the pond. The simulation results consists of six GIS rasters representing predicted water level at the requested time (see Figure 3.13).

### 3.6.3 Prototyping Cycle

The foundation of our work, summarized in section 3.1, was outlined in a definition phase. During this phase, we created mock-ups that experts refined into the initial concepts for mobile monitoring.

Figure 3.12 shows examples of visualization concepts. The tool in Figure 3.12-A draws mini-plots of accumulated sensor values in AR. The concept was abandoned due to lack of support in the early pipeline, and after initial testing, the screen form factor (outdoor lighting, resolution) proved that it would be difficult to interpret the values. Similarly, the tool to visualize sensor measurements in Figure 3.12-B was refined to reduce screen clutter by showing at most one measurement per station at a time. An initial idea to display several overlays for comparison purposes (e.g., progression of a simulation) was also sketched ( Figure 3.12-C). Complexity to implement this concept on our mobile platform made us discontinue its development. We re-introduced it later through a temporal approach: the model for the Kylmäoja catchment produces a series of overlays that are animated in place, but they cannot be shown concurrently. Figure 3.12-D shows conceptual overlays as finally implemented in our system (without support for subregion overlays, again for performance reasons).

Besides developing conceptual tools, experts also helped us specify tasks where mobile technology is foreseen to have important impact, such as: sensor deployment and maintenance, monitoring and understanding environmental processes, communication, collaboration and management of environmental processes. These outcomes set the cornerstones to define test-scenarios where the technology could be gradually deployed and tested.

### 3.6.4 Expert Workshops / Showcase Interviews

Overall, experts confirmed that our mobile monitoring approach complements effectively current practice. Our system workflow was noted useful for *sensor setup and maintenance*, ”real-time feedback for sensor setup or manual measurements insures that the data make sense, saving time and avoiding returning to site for new measurements”. It was also found useful as information exchange for *decision making support*, when multiple users with different perspectives need to discuss a situation.

*Usage Scenarios and Applications:* The discussions brought up an extensive list of potential scenarios for on-site monitoring, e.g., hydrological processes, spreading of algae, or events that are difficult to predict. Experts regarded the case studies as good, representative examples. Lessons learnt from them can be transferred to scenarios with akin situations elsewhere, e.g. transport of pollutants in streams.

Of particular interest are situations that can be re-produced in the field by visualizing results of the model, showing how the modeled situation affects the environment (e.g., effects of rain, such as flooding, can be observed even if it is not raining at the moment). Another noteworthy use case in management is on-site visualization in conjunction with warning systems, e.g., flood, avalanche or landslide warnings.

The biggest challenge for the system is nature itself, e.g., avalanche or flooding destroying sensor stations. In general, on-site monitoring increases the amount of work (e.g., testing, calibration, and maintenance stages), but the benefits gained with it may well compensate (e.g., immediate notification of failures in measurements).

On a functional level, an annotation feature - implemented in a later version of our platform - was regarded highly beneficial for professionals (employees of forestry and environmental inspectors who spend most of their time in the field). Participants also raised the need to integrate sensors for mobile measurement. For example, the SLF SnowMicropen, a manual instrument to assess the hardness of the snow cover, would certainly benefit from direct feedback while taking samples in the field.

*Visualization:* In general, users found 3D visualization useful for those situations in which spatial data sets are the basis for analysis. However, for some tasks, it is overrated. Single-point data measurements that are not further processed may fall under this category. Nevertheless, 3D visualization is beneficial for users that have difficulties reading maps. This was confirmed by all experts in Davos and Lahti: ”Experts might prefer using 2D, but a non expert understands 3D visualizations more easily. When working in unfamiliar environments 3D visualization is more useful than 2D visualization”.

Experts found the integrated visualization and comparison of different sources of information was a major advantage for decision-making (e.g., overall situation/background information). AR was rated useful in the field, though not for all situations. When setting up sensors, a camera is not needed, but the footage can help in documentation. One situation where AR finds a major value is for showing sub-surface structures or different layers of ”material” (soil, grass, snow). Further, experts expressed that AR also merits use in the office, in a form of ”tele AR”. As such, it may certainly help multi-disciplinary teams of users cooperate.

*Mobile Device and User Interface:* Functionally, users do not always need all the features and may want to customize the system. Users stressed having a steep learning curve with new user interfaces (UI), and expressed a preference for ”single button” interfaces and metaphors known from smartphones. The UI should be controllable with gloves; for this purpose, the keypad of the device proved to be too small. Thus, buttons on display have to be big enough. Besides, users judged the device setup too large and cumbersome, particularly for transport. They would rather have a modular device that could be stripped down if needed.

*Collaboration:* Experts in showcase interviews were positive about the notion and system for a shared understanding (6 experts rated very positively, one very negatively: M=4.71, SD=2.06). Annotations and marking tools to make ”notes” on screen-shots were well received and, although not devised for complex documents, they seemed good enough for reporting plans (M=4.28, SD=1.38). Experts noted the tools improve office to site communication (M=4.43, SD=1.61).

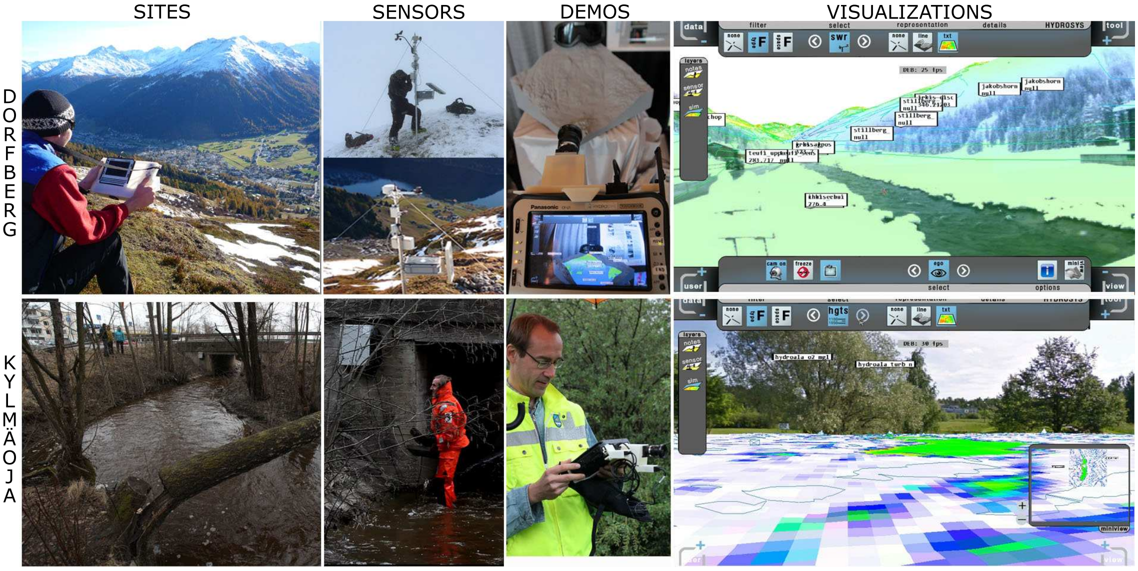


**Figure 3.13:** Scenario matrix. Sites, sensors, demos, and visualizations for two scenarios. Top row: Dorfberg, Alps, Switzerland. Bottom row: Kylmäoja, near Helsinki, Finland.

### 3.6.5 Scenario Evaluation

In the context of two deployment scenarios, this section reports results of two formal evaluations, and feedback from end-users regarding experiences and lessons learned through the usage of our system in the field.

#### 3.6.5.1 Dorfberg, Davos

**Formal Evaluation** We recruited 20 specialists (age 18-75) including eight geoscientists for the evaluation. The evaluation focused on outdoor visualization issues (such as visibility and readability), underlying the use of AR and the general workflow. After a short introduction, we conducted a walkthrough of each feature, allowing users to tests specific functions. Subsequently, they completed a questionnaire (7 point Likert-scale questions with 7 for high, 1 for low). An extended questionnaire was given to geoscientists. Our system was generally rated good to very good. Without exceptions, all participants were enthusiastic about accessing and comparing different data representations (M=5.88, SD=0.99). With respect to in-context visualization, participants successfully located sensors in the real world from the information presented on the screen (M=4.93, SD=1.16), and interpreted the visual AR information with the physical environment (M=5.47, SD=1.07, geoscientists score was M=6, SD=1.25). One main issue raised by participants was due to the screen visibility under really bright conditions during the evaluation (we addressed it later with a visor that can be attached to the device).

*Visualization Pipeline*: Throughout the on-site monitoring deployment, the data pipeline was well received by experts at SLF. The different types of data could be queried and directly displayed in the platform. The user was able to choose a format appropriate for display (1D / 2D / 3D) and to easily control overlays of the different sensors and layers. Such a direct feedback on sensor data can be beneficial when maintaining or checking sensors. The different display formats can help to get a feeling for the site even though this was not applicable to the current scenario, as the users obviously knew the site very well. Switching between different view angles was not applied directly for the scenario, as the site is very limited in size. Nevertheless, the different view angles (e.g. webcam) proved valuable for data analysis in the office.

*Workflow Integration* : On-site data access and visualization options were noted to improve the on-site workflow and provide information about the actual situation. This can affect decisions about additional action or measurements while being on site. Nevertheless, on-site data analysis is still a rough simplification of the workflow and data analysis usually performed in office after field campaigns. Along these lines, on-site analysis only provides an overview of the situation, but detailed scientific in-office analysis and interpretations cannot be replaced. It must be noted that in a highly complex research project such as represented by this scenario, the analysis is usually not straightforward, but complex and time consuming. Thus, most data analysis has to be performed in the office. Additional needs defined in the course of the development of this system were tools, which can be used for georeferenced mapping while being on site (e.g., directly map an observation, like an avalanche release zone or a specific layer in a snow profile on a screenshot).

*Usability*: The system presented limitations in terms of ergonomics: The setup was relative bulky and large and therefore not very convenient for field work. Moreover, handling the system with gloves was not simple and the quality of the display was limited - especially in bright sunlight. The quality of the displayed data (e.g., overlays on DTM) frequently showed significant deviation from reality. This can probably be attributed to the low spatial resolution of the elevation models.

In the context of snow sciences and based on comments of experts in this field, it can be concluded that the concept could be valuable in supporting the on-site workflow within the limits of a meaningful applicability.

#### 3.6.5.2 Kylmäoja, Finland

**Formal Evaluation** The system was demonstrated to the public and to specialist end-users. A formal evaluation was also conducted with experts following the aforementioned methodology. The evaluation of the system was performed in the last on-site campaign and two main features were tested, namely displaying sensor measurements and visualizing outputs of a 2D hydrological model. Nine participants took part in the evaluation, all experts in the field of geospatial analysis. Participants successfully matched scientific visual information with information from the real world (M=5.83, SD=0.75) and could easily locate the sensors (M=5.67, SD=1.00). With regards to accessing heterogeneous data, responses were mixed. We hypothesize that this is due to a difference in area of expertise, as users were not purely hydrology oriented. Registration of the data with the video scored acceptable (M=4.33, SD=0.52), also probably due to the planarity of the terrain, errors were more noticeable.

*Platform Components*: Professionals in the field noted the important aspect of independent and interdependent tasks for mobile AR environmental monitoring and how this should be integrated in such a system (e.g. data availability for running a model). Some of these tasks were technically too difficult or unfeasible regarding the choice of our mobile platform. As an example, the idea of building a simulation machinery to run models in real time atop GSN was abandoned in favor of a separate Web Processing Service for interoperability reasons. Another meaningful example in the above sense is the replacement of the hydraulic 1D model developed in-house with a 2D hydrological model because of difficulties in conducting hydraulic modeling with LiDAR based stream geometry on very small (less then 1m depth) stream. More flexibility and adaptability in term of adding and supporting more scientific processes (modeling, analyzing, simulating) are both important components that should be integrated in future version of AR environmental monitoring.

*Standards*: Implementing specific OGC standards (WPS, WFS, WCS) would greatly enhance the usability of the mobile AR client. This would allow instant in context retrieval, query and even editing of environmental spatial data. The client could provide complex functionality by chaining OGC services hosted on remote sites. With the envisioned developments in the network transmission capabilities (4G), the manipulation of the payload would not be problematic anymore.

*Modeling*: The main aim of the project/scenario was to trigger the execution of environmental models in the field, in near real-time while feeding them sensor data. This goes beyond the concept of on-site environmental modeling, as it encompasses simultaneous modeling and monitoring activities. Models that employ adimensional abstract parameters can benefit from on-site modeling, because it is possible to adjust the value of the input (parameter or variable) and directly relate the change in result to the AR context and reality. This facilitates association of abstract input with observable results and thus the role of the abstract data are better understood. Looking further, spatial autocorrelation is a property of spatially distributed environmental variables and failure to account for it can yield false results [Falk et al., 2009]. Because the spatial autocorrelation is an abstract property, it cannot be observed in nature but an in-context AR visualization would allow a better estimation and understanding of this property specially for non GIS experts.

### 3.6.6 Discussion

*Context*: Environmental monitoring and on-site visualization can provide clues into and facilitate the understanding of various abstract environmental parameters in respect to specific natural processes. In general, end-users confirmed both the technical quality of the infrastructure and the usefulness of our approach. Our mobile AR system accomplishes the goal of presenting information from an environmental sensor network in-context with its location of occurrence. With few additions (modeling), it could support model validation and calibration for certain classes of environmental models.

*3D Visualization*: Not all environmental scientists had prior experience with 3D visualization, though those who did rated the tool very positively. The progress of the prototype also influenced end-user attitude. The nordic expert workshop confirmed that 3D is generally believed to be the ”future” of hydrology/environmental visualization. During the final demonstration with a more mature prototype, a higher preference for 3D visualization was noted in general. A number of users stated that they would use the system in its current form, and others would adopt its component-based architecture. We must concede that the combination of AR and mobile monitoring entails a new paradigm compared to just traditional tools. In this sense, it is reasonable that experts will need time to integrate this solution in their scientific workflow (which will imply an integration of a broader range of data format interoperability in our system).

*Performance*: A major limitation was the computational capability of ruggedized handhelds for environmental science application: as we modified our hardware prototype throughout the project (i.e., specific casing, tracking devices, ergonomics), adapting it to all weather conditions was really challenging, notably as ensuring reasonable performance computation with the device (i.e. real-time graphics, computational steering simulation trials, etc). We hope to address these issues in the future by redeveloping our system on more powerful handheld that can benefit of the recent trend of advanced smartphones platform (e.g., iOS, Android), but also a lightweight version of our platform for the general public.

*Domain*: Along the project, we also started to develop more domain-dependent tools (like marking tools for snow profile), which will be integrated to a specific domain-based layer in our infrastructure. Similarly, this work was not focused on issues related to network reliability, data quality, data uncertainty or simulation accuracy. More work on these topics, especially regarding novel ways to combine different visualization techniques in AR, need to be investigated. Although some of the proposed features are quite application-specific, most of our concepts and results can readily be deployed in other scientific or engineering fields, where pervasive sensors are deployed.

## 3.7 Summary

Over the last decade, we have seen significant progress in research about capturing and visualizing environmental processes to enhance modeling, prediction or any decision making process in general. The contribution of this chapter is a 3D mobile AR platform, allowing a researcher to visualize and interact with data in-context, integrated in an infrastructure covering wireless sensor acquisition/management to mobile visualization. This solution targets real-time access to sensor data, simulation results and the physical world, while providing dedicated tools for analysis and comparison.

Overall, on-site AR environmental monitoring can be regarded as a promising technique, and we expect that it will mature in upcoming years and position itself amongst the fundamental techniques of environmentally aimed scientific inquiries.

# Chapter 4

# Handheld Devices for Mobile Environmental Monitoring

How to design a physical interface for mobile environmental monitoring? In this chapter we analyze the design and construction of four generations of handheld platforms and the evolution towards an extreme outdoors device. The first generation, reported in [Veas and Kruijff, 2008]), was a result of various studies on ergonomics and human factors. Thereafter, each following iteration in the design-production process was guided by experiences and evaluations that resulted in new guidelines for future versions.

## 4.1 Introduction

Fueled by the increase of available mobile phone hardware and software, handheld Augmented Reality (AR) has become a paradigm for mobile AR applications. Thanks to their form factor, mobile phones come out as the preferred platform to bring mobile AR to the wide public. Nevertheless, the technically limited mobile phone hardware only provides a constrained platform for AR that does not necessarily match the requirements of all applications.

In recent years, mobile AR has also become a paradigm for industrial applications. These applications pose strict requirements on accuracy of the components that are used to generate the AR experience, namely tracking, video and graphics. Our work focused on designing and producing a platform primarily for this breed of high quality AR applications. The hardware consists of a display device with externalized sensors and controllers. The development has been driven by qualitative evaluations consisting of experiences followed by interviews and surveys with experts in the public sector and industry. These interviews and observations provided requirements and guidelines for each generation of our hardware platform.

The analysis, design and evaluation of our platform in all its flavors has been part of three projects: Vidente and SMARTVidente focus on on-site modification and surveying of geometric and semantic attributes of geo-spatial 3D models, whereas HYDROSYS aims at monitoring and management of environmental processes through interactive visualization of sensor network data. Applications for these projects require accurate tracking, high quality graphics and video and are operating in potentially harsh environments: A robust platform is required that can hold all necessary sensors and controllers and can cope with the external conditions.

The novelty of this article can be found in both the generation of robust and ergonomic devices for handheld outdoor AR, and the experiences gained and reported. Ergonomics and human factors are the foundation upon which all our research is constructed. We present a compendium of guidelines and best-practices, going from very experimental design for research purposes, to applications requiring limited weight while still providing the best AR experience in extreme outdoor environments. We show how the functional requirements evolved from research on interaction with handheld AR standpoint towards more elaborate applications. In our application domain, usage duration is longer than the generally “peek through the hole and decide” actions supported by most current AR applications. Furthermore, to gain acceptance by end-users, devices have to appear less ”experimental” and be more robust to endure actual usage.

The different stages of analysis and design of platforms reacted to four main categories of requirements that have a strong interplay:

**Ergonomics:** user should be able to hold the device ergonomically in longer interactive sessions

**Robustness:** the construction should protect in particular the sensors from possibly harsh weather and usage conditions

**Compactness:** the device-sensor combination has to be small, held comfortably and easily transportable

**Modularity:** the device should be reconfigurable based on user needs: sensors should be mounted or removed at will

Whereas the first generation was mostly driven by ergonomics, the following stages proceeded towards robustness and compactness, and the final version aimed at obtaining modular construction. At every stage, we reflected the knowledge gained from the previous generations: hence, the fourth generation considers all categories of requirements and feedback we received over time.

We will illuminate the foundations that provide requirements for the design of handheld AR platforms. Thereafter, we proceed through all four generations of devices, showing how and what we learned throughout the iterations of requirement analysis, design and evaluation. This chapter concludes with a reflection on the experiences we gained, describing guidelines for other researchers and practitioners to design apt platforms.

## 4.2 Foundations of Handheld AR

The analysis, design and evaluation of platforms for handheld AR is driven by three aspects: task domain, sensors and controllers needed, and ergonomics associated with operating the device. This section provides the background on all three aspects and thus the crux to all the design stages of the four generations of devices. Throughout the chapter, we will predominantly focus on the interplay between sensors and controllers, and ergonomics. Part of the foundations described in this section are derived from a larger analysis that formed the basis for designing the first generation of platforms, the Vesp'R [Veas and Kruijff, 2008], but in principle affect all generations of device platforms.

### 4.2.1 Task

Before designing the first handheld platform, we performed a detailed functional analysis on handheld AR applications [Veas and Kruijff, 2008], to better understand the interaction space. The result anticipated several forms of interaction that are desirable under this paradigm. Handheld AR applications stem from many domains, varying from entertainment, to city navigation, and engineering. Many of these applications share features of mobile applications, in particular of location based services (LBS) [Jones and Marsden, 2006]. Generally speaking, users perform navigation actions (viewpoint manipulation and maneuvering, including map interaction), simple system control actions such as visualization mode changes, straightforward object manipulation actions (often using a lens-metaphor), and only limited numerical input. It was found that in most applications, interaction is clearly dominated by the viewing of data. When analyzing the functionality of these tasks, we noticed a high variety between accuracy, speed, frequency and duration of actions and no clear association between a task and a controller.The duration of interaction also varies widely, between less than a minute for simple tasks, up to about 30 minutes for complex, and possibly collaborative tasks. For a more detailed taxonomy of tasks, refer to [Veas and Kruijff, 2008].

### 4.2.2 Sensors and Controllers

One of the main requirements for generating an augmented reality in a mobile domain is a self contained system holding all devices that are required for the application. The platform has to serve all requirements of these devices, such as power and connections.

AR requires several **sensors** to accurately register augmentations in 3D. For graphical augmentations in (handheld) AR, this implies that the pose of the camera capturing the real world must be known to calculate relative poses to render augmentations. Registration accuracy is determined by the quality of sensors and the implemented technique. Localization is generally achieved using GPS or a vision-based approach. For higher-accuracy localization, it may be needed to fuse the input from different sensors, or to rely on methods such as differential GPS together with correction signals [Höllerer et al., 1999]. An orientation sensor is often included to complete the pose estimation, in particular when relying on non-vision based tracking. Finally, live video is needed to generate AR for which, theoretically, any camera can be used. However, several characteristics of the camera affect perception of the AR experience. The field of view and focal length define the active viewing area. A reasonably high resolution that matches the screen aspect ratio is desirable, whereas a high framerate helps to generate a smooth experience.

Allocation plays an important role both for sensors and controllers. With regards to sensors, all distances between the camera and sensors used for localization must be measured and calibrated for accurate localization. To simplify the calculation and the calibration procedure, the orientation sensor is mounted preferably below (or above) the camera. The GPS antenna should be mounted at a fixed location, where signals are not blocked. Also, offsets to accurately specify location should be known, although for normal GPS with error larger than 1m, this is not necessary.

Most handheld computers are equipped with a number of **controllers**. At a UMPC, control can range from 2DOF to 6DOF, and make use of both isometric and isotonic control: as we noted in [Veas and Kruijff, 2008], handheld AR interaction is often bound to the platform at hand. Handheld computers generally include a micro-joystick, a couple of buttons that are possibly associated with a keyboard, or click switches. These controllers often do not afford fine-grained action and may be difficult to reach. Besides buttons and micro-joysticks, most platforms also include a touch screen that nowadays can be operated by finger (direct input or gestures) or pen. Apart from for viewpoint interaction, camera and orientation sensor can also be used for gestures, though this is not very common due to the decoupling of visuals and movement. Finally, in cold weather, the device needs to be operated while the user is wearing gloves, which poses severe restrictions on the choice of controllers.

### 4.2.3 Ergonomics

User acceptance of a handheld device can be ascribed to its ergonomics. Ergonomics is determined by several interrelated issues: pose, grip, controller allocation, weight and size.

The **pose** is defined by the bio-mechanic system of the wrist, arm and shoulders and mainly influenced by the angle at which the device is held. Regulated by the way the user has to look at the screen to view the world, the pose may be at eye-height or lower, with varying distances from the body.

The **grip**, the way the user holds the device with the hands, affects the comfort of holding the device, may limit fatigue, and can support simultaneously holding and interacting with the device. Users normally require a power grip, which avoids the device from slipping from the hand, which may occur especially in heavier setups. The ideal diameter of a grip is around 76mm (from fingertips to palm), at which it increases the strength of the wrist. Unfortunately, with smaller devices the size is hardly achievable [Marras, 1997]. Depending on the device weight and size, the user makes use of a single or two-handed grip. A two-handed grip os often required to avoid the tilting of a heavier device, or when the user needs to interact with the content on the screen (without using a pen). Balancing is frequently a result of fatigue caused by holding the device up: users may need to hold a device at eye height for several minutes, which may cause fatigue such as muscle tremble. Lowering the device by one foot can easily double the duration of holding up a device without experiencing fatigue [14]. Weight-balance strongly affects the ergonomics of a pose – when a setup is off-balance, in certain poses the tilting of a device will quickly result in fatigue, in particular when a non-ideal grip on the device needs to be maintained.

The **control allocation** is in direct interplay with both pose and grip. The locations of the controllers in the device, and the relationship between location of controller and grip defines if a user can directly control an application, or needs to change the grip on the device. Whereas micro-joysticks are mostly placed at reachable locations, touch screens require the user to grasp the device differently.

Obviously, ergonomics is highly affected by the **weight** and **size** of the different kinds of computers and sensors that are coupled in the handheld construction. The higher the weight, the more restricted the user may get. In Table 4.1, we provide an overview of the weight and approximate size of the various computers, sensors and associated cabling that are carried around.

### 4.2.4 Requirement summary

To summarize, we can make several statements that apply to the design of handheld platforms for high-quality AR. These requirements are tackled throughout the four genera-

| Device | | | Cabling (bundle) | | Version* |
|---|---|---|---|---|---|
| Name | Size (WxLxH) | Weight | Length/Size (WxLxH) | Weight | |
| **Computer** | | | | | |
| Sony Vaio UX | 150 x 100 x 38 mm | 517 gr. | N/A | N/A | 1,2 |
| Panasonic CF-U1 | 184 x 151 x 57 mm | 1060 gr. | N/A | N/A | 2,3,4 |
| **Sensors** | | | | | |
| Intersense IC3 | | | | | 1,2,3,4 |
| – Sensor | 26 x 39 x 15 mm | 17 gr. | 4.57 m / 120 x 45 x 22 mm | 120 gr. | |
| – Serial/USB adapter | 60 x 35 x 20 mm | 85 gr. | 2 m / 120 x 42 x 20 mm | 85 gr. | |
| Ublox GPS | | | | | 1,2,3,4 |
| – Antenna | 39 x 46 x 12 mm | 42 gr. | 5 m /110 x 40 x 22 mm | 65 gr. | |
| – Electronics | 55 x 54 x 24 mm | 70 gr. | Short USB cable** | 8 gr. | |
| Ueye Camera with lens | 89 x 42 x 42 mm | 190 gr. | Short USB cable** | 8 gr. | 1,2,3,4 |
| Standard Webcam | 62 x 30 x 21 | 30 gr. | 0.6 m / 50 x 10 x 4mm | 15 gr. | 4 |
| Standard USB Hub | 58x 35 x 15 mm | 21 gr. | Short USB cable** | 8 gr. | 2,3,4 |
| Battery pack (4 x AA) | 54 x 54 x 15 mm | 110 gr. | N/A | N/A | 1,2 |
| **Controllers** | | | | | |
| ICube-X Midi wireless hub | 61 x 22 x 14 mm | 20 gr. | N/A | N/A | 1 |
| Various midi actuators | N/A | N/A | N/A | N/A | 1 |
| Bodnar USB board | 55 x 33 x 9 mm | 10 gr. | N/A | N/A | 1,2 |
| Genius Maxfire Pandora Pro | 41 x 81 x 12 mm | 27 gr. | Short USB Cable** | 8 gr. | 3 |

* 1=Vesp'r, 2=Bulk'r, 3=Ice'r, 4=Cool'r

** Short USB cable= cable < 15cm

**Table 4.1:** Dimensions of computers, sensors, controllers and associated cabling.

tions of devices we present in Section 4.2.5. As we will show throughout the discussions, these requirements relate directly to ergonomics, robustness, compactness and modularity.

- Match sensors and controllers to the needs of the application
- Beware of the dependencies between sensors and place them correctly in relation to the display device
- Provide easily accessible and well performing controllers
- Integrate all devices without destroying ergonomics
- Support ergonomic poses and afford a good grip (or grip variety) on the handheld construction

- Keep weight and size as limited as possible and balance the construction

The requirements are in direct relation to the display device (hence, screen and computing platform) being used: the device forms the actual starting point for defining the needs on the sensors and controllers. It may come with sensors that could potentially be used such as a build-in camera or tilt sensor. However, in most cases, these sensors have similar qualities as mobile phone sensors and need to be replaced by high quality devices. It should have a bright and high-contrast screen that can be used outdoors, and preferably includes an anti-reflective surface. Furthermore, it should have a low-power consuming CPU to ensure long operation. Both the data received from the sensors and the actual graphics overlaid on top of the video may pose higher demands on the processing capacities of the platform. As a result, most AR applications will benefit from a computer that has a graphics processor too. In the past, the Sony Vaio UX platform was the dominant choice, but is not produced anymore. Meanwhile, researchers try out different platforms, including small tablets or robust UMPCs such as the Panasonic CF-U1. At present, the market for tablets has extensively diversified. Manufacturers have turned to ARM processors that deliver stable platforms with sustainable battery life, and superior display resolution. In the near future, the inclusion of graphics cores will be the deciding factor for high-quality mobile AR.

### 4.2.5 Handheld Devices for Mobile Augmented Reality

The motivation behind the iterative design cycles was the creation of a platform that affords mobile, high quality AR. Table 4.1 lists each piece of hardware we have used, its weight and dimensions, and dimensions for the cabling it requires. The table includes a column indicating what version(s) of the platform use each piece of hardware. The cabling is relevant for all but the first generation of our device, because this version relied on specially tailored electronics (USB hubs, and joysticks) and cabling to reduce space requirements. Sections 4.3 to 4.6 describe each generation of our platform, emphasizing its requirements, design and evaluations. In particular, we stress how the results of each evaluation influenced requirements for the next generation of the platform.

## 4.3 Vesp'R: ergonomics and experimentation

The particular motivation for the first generation was to open new possibilities for spatial interaction using handheld AR, as well as to experiment with ergonomics of one and two-handed interaction. We explored both conventional controls (joysticks and extra buttons) and non-conventional controls ( bend sensor, grip camera) and set out to create a new experience for handheld AR.

**Figure 4.1:** Vesp'R: the single-handed grip, the handles with different controllers, the inside of the new single-handed grip/handle and the backpack.

This generation has a strong focus on ergonomics and compactness, while modularity is only considered for controllers and interaction modality.

### 4.3.1 Requirements

The requirements for the first generation of the platform derived from evaluations on UMPCs, on ad-hoc constructions [Veas and Kruijff, 2008], and from analysis on handheld AR. Adding to the requirements introduced in the previous section, this platform was aimed at studying pose and interaction possibilities: the platform had to allow both single-handed and two-handed interaction. It also needed to support multiple interaction methods, to enable experimentation with non-conventional controls and interaction modalities. The controls had to be allocated such that when interacting with the application in one pose, the user does not need to change pose to reach a certain control. After studying ergonomics on grip, it was decided that a power grip is needed to balance weight. Allocation of controls and sensors needs to be carefully planned with respect to the expected poses for interaction, so that weight balance is maintained without causing unnecessary strain in the lever-system biomechanics of the arm.

### 4.3.2 Design

We initiated the design process by going through a number of iterations of designing and evaluating mock-ups with a small user group. The mock-ups were used to evaluate different kinds of poses, grips and control allocations that lead to the general form of the platform. To comply with multi-modality of pose and interaction requirements, the platform was designed as central (base or ”backpack”) unit integrating the UMPC and all sensors. Controllers are placed on handles that connect to the base unit at specified locations. For two-handed usage of the construction, we adopted a design that placed two handles at the side of the base unit. The base unit was designed to hold most of the sensor packages in a box. A mount for the camera with room for an orientation sensor below was allocated externally, as well as the GPS antenna (mounted in a short pole at the side). All cables were measured to the smallest distance, and shortened such that each sensor would be connected without wasting space in spare cabling. A USB hub was modified to use on-board connectors to save space (from 4cm for normal USB to 5mm for on-board 5-pin connector).

We experimented with general forms of **grips**, in which different kinds of controls could be embedded using plastic boxes, foam, clay and other materials for mock-ups. From these studies, we found out that the grip itself resembles that of similar devices like a drill or joystick: other than changing scale to hold electronics in a balanced way, the form is hard to improve. In the first handle, based on the **controller allocation** plan, we mounted two joysticks, usable by index finger and thumb respectively. The idea was to map constrained interaction techniques on both micro-joysticks to control specific axes in translational task with a dedicated controller (see Figure 4.1). We also included 3 thumb-operated buttons, and one that could be reached by either the index finger or the middle finger. All controls are mapped to a USB board from an off-the-shelf joystick. The second handle is a test bed for alternative, unconventional MIDI controllers, and including quasi haptic input methods relying on touch sensitive Piezo sensors. Initial experimentation shows that the usage of Piezo based elements has limitations: in single handed configuration, the force needed by the fingers to balance the construction prevent fine-grained control. As a result, only the secondary control unit includes a Piezo sensor accessible by the thumb, and a bend sensor that can be used by the index finger. The latter can be used well to control ranges of values. In addition, a wireless camera and laser pointer are mounted in the joystick for additional tracking and interaction purposes. The single-handed version requires to detach both handles from the sides, one handle can be mounted below; the second handle could be put away, or used for freehand (spatial) interaction. The handles were attached to the base unit with normal screws. Later we designed a handle specially for one handed interaction. Calculating the approximate weight of the case and the peripherals against the weight of the UMPC, we placed the handles directly behind the back of the UMPC, close to the weight equilibrium. This supports the device tho be handled in a balanced way. In this version the weight and **size** were quite optimized due to the alternative cabling of devices which, however, is difficult to achieve.

This generation is produced using nylon-based stereolithography (STL), having a wall-thickness of around 3mm. The model is covered by a thin velvety-like rubber and is partly glued, partly screwed.

### 4.3.3 Evaluations

We performed two formal evaluations to test acceptance of the platform and the interaction modes it affords (for complete results see [Veas and Kruijff, 2008]). The first test was set up as a user attitude evaluation, exploring a more application specific setup. The second test, intended as comparative study, evaluated different device setups in a more abstract task setting. Both tests looked specifically into pose, grip and controller issues.

#### 4.3.3.1 First Evaluation

In the first evaluation, 17 users (16 male and one female, all with a computer science background, but no AR/MR specialists) made use of the Vesp'R in single handed mode. The large handle with the micro-joysticks was mounted below the BatPack (Figure 4.1 Left); the other hand could be used freely. The users could explore an excavation site, using a magic lens tools, but did not have a specific task sequence they needed to follow. Average usage time was about 10 minutes, during which the evaluators observed users and were open for discussion. After the practical part, users were asked to fill out a questionnaire with 13 questions (7 point Likert scale, higher scores are better).

We predominantly focused on the ergonomics of the device construction and grips, next to the usage of the controls. In addition, we stated questions regarding optical issues and the user interface.

**Results** The overall weight of the device was not rated negatively, but obviously users preferred using a lighter device construction. Both the weight balance and the grip on the construction were rated mid range. Nonetheless, most users did not report noticeable fatigue after using the construction for about 10 minutes. Consequently, the comparative study (section 4.3.3.2) needed to provide better insight on the weight, balance and grip issues. The placement of the controllers was rated positively, similar to the control effectivity. The interaction techniques (not discussed in detail here) also performed well, in the range of avg. 5.00 and 5.50.

The test showed weight balance could still be improved. The device slightly tilts to the back resulting in some fatigue after extended usage. Better weight balance could be achieved by placing the handle further to the center of the weight equilibrium, what we did in a second smaller handle with the changed control structure.

We also observed that users held the device in an unintended but obviously convenient manner: by holding the BatPack itself, instead of the handle. Five users held the handle and used the second hand to grab the BatPack below or behind (see Figure 4.1 Left). Another user grabbed the BatPack with both hands reaching the handle only to control the application. In the second evaluation, we intended to observe this phenomenon in

| Question | Avg | Stdev | Correlation with Second Study |
|---|---|---|---|
| Overall weight | 3,93 | 1,58 | +/- |
| Weight balance | 4,07 | 1,10 | +/- |
| Ergonomics of grip | 4,27 | 1,16 | - |
| Grip material | 4,80 | 0,94 | - |
| Fatigue | 5,21 | 1,58 | + |
| Placement of controllers | 5,07 | 1,39 | +/- |
| Switch between focal planes | 5,20 | 1,37 | |
| Switching nuisance | 5,38 | 1,56 | |
| Control effectivity | 4,80 | 1,42 | +/- |

**Table 4.2:** Results of first evaluation. Correlation "+" means the rating was higher/better than the second evaluation.

more detail, by using the support hand as one condition. The first impression was that the non-dominant hand was used to relieve the force on the dominant hand, providing a steadier grip and avoiding tilting during interaction.

Overall, the results of the test were quite positive, even though many ratings were mid-range. The explanation, supported by the second evaluation, is quite simple: single-handed interaction is tedious, and clearly affects the rating. The weight and associated fatigue (even when not rated dramatically in test 1) lowers the overall attitude towards the device.

#### 4.3.3.2 Second Evaluation

This evaluation compared traditional UMPC setups with the different configurations that Vesp'R affords. 15 users (12 male, 3 female, 14 right handed, one left handed) participated in the test. All users had a computer science background, 4 people were non-specialists in the field of AR/MR.

We considered five conditions: UMPC only (535 grams), the UMPC construction with a perspex encasing (739 grams, referred to as "UMPC with plastic case"), Vesp'R with two large handles mounted on the sides (totaling 1249 grams, referred to as "handles at side"), Vesp'R with one large handle mounted below (the one with micro-joysticks, 1091 grams, referred to as "big handle below"), Vesp'R with the newer, smaller handle mounted below (1105 grams, referred to as "small handle below"). The different weights provided us with insights in the weight balance and ergonomic factors influencing fatigue. Each condition took about five minutes, totaling test time of about 40 minutes (including answering the questionnaire).

The test was laid out in two spatial areas: a selection and a placement area separated by about 5 meters. The selection area consisted of a poster placed at eye-height with markers over which different buildings were overlaid. The placement area consisted of two posters with a city map. One poster was placed on the wall at eye sight ("wall-mode"),

| Device | Mode | Overall weight | | Weight Balance | | Operate 1 Hand | | 1-handed view | |
|---|---|---|---|---|---|---|---|---|---|
| | | avg | stdev | avg | stdev | avg | stdev | avg | stdev |
| UMPC only | W | 6.53 | 0.83 | 6.67 | 0.82 | 2.10 | 1.73 | 5.70 | 1.95 |
| | T | | | 6.73 | 0.59 | 2.10 | 1.64 | 5.60 | 2.12 |
| UMPC with plastic | W | 6.07 | 1.03 | 6.00 | 1.20 | 1.80 | 1.65 | 4.70 | 2.21 |
| | T | | | 6.07 | 1.10 | 1.90 | 1.90 | 4.90 | 2.42 |
| Handles at side | W | 5.00 | 1.36 | 6.13 | 1.30 | | | | |
| | T | | | 6.20 | 1.01 | | | | |
| Big handle below | W | 3.73 | 1.91 | 4.07 | 1.75 | 4.60 | 1.64 | 6.07 | 1.58 |
| | T | | | 4.27 | 1.75 | 4.07 | 1.67 | 5.60 | 1.76 |
| Small handle below | W | 3.67 | 2.02 | 4.00 | 1.81 | 4.53 | 1.46 | 6.33 | 1.35 |
| | T | | | 4.07 | 1.83 | 3.93 | 1.39 | 6.00 | 1.41 |

**Table 4.3:** Vesp'R weight balance ratings. Weight, balance, operation ratings. (W = wall, T = table).

one on a desk ("tablemode").

The purpose of these modes was to analyze how pose affects usage for each construction, since both poses differ in weight balance and forces on the hand, wrist and fingers. Participants had to pick and then place buildings on the correctly, placing two objects (buildings) at each wall and table placement area. Upon placement, objects snapped to a specific position close to the final one. The snapping forced users to perform some translational actions. During the test, we noted down the specific pose and grip observations, and watched muscular activity to detect signs of fatigue. Afterwards, users were asked to answer 13 questions (total 95 answers), using a 7 point Likert scale questionnaire (higher is better).

**Weight, balance and operation** The overall weight was perceived as very good for both the UMPC and the UMPC with plastic case. The rating was followed by a still very good note for the Vesp'R with side handles, considering that it weights about 500 grams more than the other constructions. The single-handed versions both scored the same: acceptable, but rather on the heavy side. Actually, both are about 100 gram lighter than the two-handed version, but the weight mostly leans on the dominant hand.

As expected, independent of the mode (table/wall) weight balance was very good for both the UMPC only and the UMPC-plastic case combo. The two-handed Vesp'R received high ratings, confirming our work on weight balance. Again, considering that the two-handed version is considerably heavier than the traditional light weight UMPC construction, we believe this result is very good. In line with the first evaluation, weight-balance for the singlehanded configurations rated worse. Users did not notice a major difference between the two handle placements, event though the smaller handle is placed far closer to the weight equilibrium. Possibly the weight was too high for one hand that a difference could not be clearly noticed, even when the second hand was used to support

| Device | Mode | Fatigue | | Regrasp weight | | Regrasp control | | Comfort | |
|---|---|---|---|---|---|---|---|---|---|
| | | avg | stdev | avg | stdev | avg | stdev | avg | stdev |
| UMPC only | W | 5.87 | 1.19 | 6.73 | 0.46 | 6.13 | 1.46 | 5.33 | 1.54 |
| | T | 6.00 | 1.07 | 6.73 | 0.46 | 6.20 | 1.42 | 5.33 | 1.54 |
| UMPC with plastic | W | 5.20 | 1.70 | 5.67 | 1.99 | 5.40 | 2.23 | 4.00 | 1.89 |
| | T | 5.47 | 1.41 | 5.80 | 1.86 | 5.47 | 2.10 | 4.07 | 1.79 |
| Handles at side | W | 5.27 | 1.49 | 6.33 | 1.11 | 6.47 | 0.74 | 6.07 | 1.39 |
| | T | 5.60 | 1.35 | 6.40 | 1.05 | 6.47 | 0.74 | 6.00 | 1.36 |
| Big handle below | W | 4.07 | 1.79 | 4.47 | 1.84 | 5.07 | 1.71 | 4.27 | 1.94 |
| | T | 4.00 | 1.81 | 4.53 | 1.77 | 5.00 | 1.73 | 4.13 | 1.84 |
| Small handle below | W | 4.00 | 1.81 | 4.27 | 1.91 | 4.53 | 2.07 | 4.60 | 1.80 |
| | T | 3.93 | 1.62 | 4.40 | 1.80 | 4.40 | 2.03 | 4.27 | 1.53 |

**Table 4.4:** Vesp'R fatigue ratings. Fatigue factors ratings. (W = wall, T = table).

the BatPack.

Single-handed operation ratings changed the appreciation of quality for some devices considerably. Users rated the usage of both UMPC and UMPC with plastic case as extremely bad for one handed interaction. Both single-handed Vesp'R scored much better with acceptable, but not impressive scores. This indicates that we improved the construction to such extent that singlehanded interaction (control) of applications is better than with the traditional configurations. Nonetheless, participants believed that all configurations are appropriate for one handed viewing (navigation only tasks). The UMPC-only condition scored well in both modes, with the plastic case only slightly lower. Still, the single-handed Vesp'R outperformed them both, with the smaller handle (wheel stick) performing best in all modes.

**Fatigue** The results on fatigue were in line with the weight balance ratings: the UMPC only, the UMPC with plastic, and the Vesp'R in two-handed condition performed good to very good, whereas the single handed version performed worse. Due to the mixed duration of the tests (5 minutes per configuration and 25 minutes maximum usage), it is difficult to truly grade the fatigue ratings. The 25 minutes were not completely continuous: the arms of the user could relax in between configuration changes, even though these mostly took just 20 seconds. Nonetheless, we believe that for normal usage of the device, the ratings are quite representative, even though making a true duration test would be an interesting and needed alternative.

The UMPC only and the UMPC with plastic received high scores on regrasping. This came as a surprise, since the grip did not seem very comfortable when observing participants perform.

The two-handed Vesp'R performed extremely well in both wall and table mode. Even with the increased weight, no regrasping of the construction / grips was necessary.

In line with results on weight, both single handed grips performed worse. The increased force on the single hand clearly resulted in regrasping due to fatigue on the hand holding the joystick, even when the construction was balanced by the second hand.

With regards to user comfort, results were clearly in favor of the two-handed Vesp'R. The UMPC only still scored reasonably well in both wall and table mode, the UMPC with plastic case only proved mediocre. The two-handed Vesp'R received a high score in both modes, supporting our ergonomic studies and design of the device.

**Grips and interaction** Investigating the ergonomics of the grips and the interaction with the controllers, showed surprising results. None of the device grips rated extremely well. Surprisingly, people did not mind holding the UMPC, which was rated about the same as the big handles we used beside and below the BatPack. The small handle received a lower score, whereas the UMPC with the plastic case performed worst.

We have the strong impression users only rated ergonomics of the construction as one instead of the joystick: hardly without exception, the large joystick performed extremely well in the two-handed configuration, while the same joystick received lower ergonomics values in the single handed configuration.

The material of the joysticks, the velvety rubber, was highly appreciated by the users (avg. 6.00 / stdev 1.20). The low rating of the UMPC with plastic could be higher if it had a better mounting for connecting the UMPC: multiple users reported on being afraid to drop the UMPC from the casing.

Users found the UMPCs buttons easy to reach. Both the big and the small handle scored about equally well. Surprisingly, when asking the users about the placement of the controllers on the big handles in two-handed configuration, the score was much higher. Obviously, the weight balance, pose and force on the hand / wrist have a large effect on how the user reaches and uses the controllers.

We received quite diverse ratings on effectivity. All UMPC conditions had similar interaction mechanisms: select and drop objects with a button and move them with the finger mouse. The two-handed configuration made use of the same mechanisms as the single handed big handle configuration: objects were selected by button click and moved in a constrained way using the micro-joystick. For the wheel mouse, we also constrained the interaction. Users could select a specific axis using one button and scroll the wheel to translate the object over this axis.

We observed that most users had problems with the UMPC only condition due to limitations while translating objects. Since the micro-joystick is mapped to the mouse pointer, once it hits a border it does not translate in this direction any further. Thus, users needed to clutch by tapping on the screen, to move the mouse pointer back. This obviously limited interaction for some users. On the other hand, the wheel mouse has a mechanical disadvantage, since users need to put too much force on the wheel to click in

| Device | Mode | Grip ergonomics | | Control placement | | Effectivity | |
|---|---|---|---|---|---|---|---|
| | | avg | stdev | avg | stdev | avg | stdev |
| UMPC only | W | 5,30 | 1,77 | 6.11 | 1.27 | 3.60 | 1.18 |
| | T | | | | | 3.73 | 1.10 |
| UMPC with plastic | W | 4,00 | 2,16 | | | 3.53 | 1.30 |
| | T | | | | | 3.67 | 1.29 |
| Handles at side | W | 5,20 | 1,82 | 6.00 | 0.88 | 6.13 | 0.64 |
| | T | | | | | 6.20 | 0.56 |
| Big handle below | W | 5,20 | 1,82 | 5.47 | 1.68 | 5.00 | 1.77 |
| | T | | | | | 4.93 | 1.79 |
| Small handle below | W | 4,80 | 1,86 | 5.33 | 1.72 | 4.60 | 1.64 |
| | T | | | | | 4.33 | 1.80 |

**Table 4.5:** Vesp’R control ratings. Grip and controllers ratings. (W = wall, T = table).

either direction. The low rating of the wheel joystick was not only due to the mechanical construction, but also in relation to the force on the hand in single handed usage. Without support, the index finger can hardly control the wheel, but is rather used to balance the construction. At the end, and most importantly, the two-handed Vesp’R configuration outperformed all others.

#### 4.3.3.3 Discussion of results

Interaction with the single-handed version is possible, but not ideal. Both evaluations showed that holding the device single-handedly is not very ergonomic for longer sessions. Leaving the second hand free for pen-input or real-life communication would require another approach, particularly for longer sessions. A positive whilst unintended effect of the base unit is the rather ergonomic pose it affords: when holding the device from the bottom, single-handedly. This pose is particularly comfortable when the elbow if placed against the waist, resulting in what we call the waist-pose. The two handed version advances ergonomics of handhelds significantly, in such a way that people perform comfortably even with double the weight of other devices. Proper controller allocation affords fine-grained actions even with the non-dominant hand. This version has been continuously in use since its construction by researchers and end-users.

The main drawback of this first version is its structural instability. The mechanism to attach handles weakens the construction, and it often had to be repaired and reinforced. Specially tailored electronics made it difficult for people to fix problems or extend the platform by themselves. Extending the platform with new sensors is possible insofar as these fit in the base unit, while new controllers require designing a new handle.

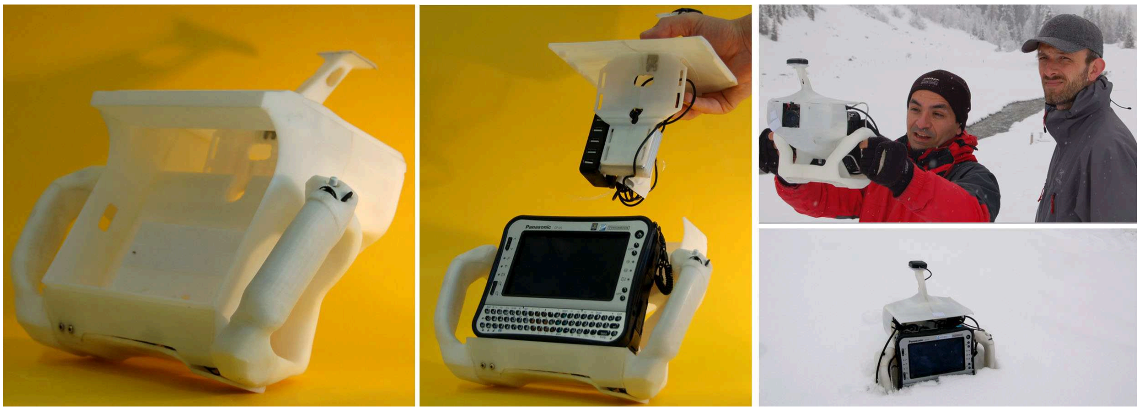


**Figure 4.2:** Bulk’R: the grip with the backpack, the fin construction for attaching devices and cables, and outdoor usage in harsh conditions.

## 4.4 Bulk’R: robustness and weatherization

This generation came as a first attempt to cope with requirements for a **robust**, all-weather device for outdoor use. The previous generation, developed for research, was too fragile. Its design, aimed at minimizing weight and dimensions of all parts, led to specially tailored electronics that were difficult to replace. From the usability and human factors stance it was well received, and it drew enough attention that it was constantly in demos and presentations, going back to the lab only for repairs. However, the multi-configurability facet (two-handed, one-handed) and its specially tailored electronics made it too much of a prototype. Robustness was not among its requirements, and certainly not among its features. When the HYDROSYS consortium issued its request for an all-weather platform, it was viewed as an opportunity to build on lessons learned and create a robust platform for outdoor use. This platform was designed with two very different computers in mind, and initially covered the same external sensors as the previous one; although it has been extended to use more, even experimental, sensors (see Table 4.1).

### 4.4.1 Requirements

To meet the weather resistance requirements of outdoor use, the unit had to enclose and protect all sensors. Furthermore, the strict robustness requirements called for a specially tough body, and standard electronics that could be replaced by off-the-shelf components. A new requirement brought about after testing the old platform was the need for batteries to supply the external USB hubs and sensors. The Sony UMPC was tested with all devices and could only run up to 30 minutes when everything was connected and generating data. This limited its usability mainly to demo sessions. To extend the runtime expectation, external batteries had to be used to supply the USB hub(s) and all the external hardware. The Panasonic CF-U1 tested under the same conditions, has a battery life close to 4 hours. The requirements for interaction from the applications that used the Vesp'R proved more limited than initially evaluated, thus our functional evaluation had to be revisited to limit spatial interaction. Most of the non-conventional controllers were removed, adding 1D controllers and some buttons.

### 4.4.2 Design

The design was divided in the grip, the housing to hold the sensors and the housing for the computer. A housing in the size of the Panasonic computer was designed to enclose the Sony computer to protect it from weather effects and sunlight. When using the Panasonic this housing is removed. The three parts are joined semi-permanently using a screw mechanism, only to be separated for exchanging the computer. The part designed to be removed is the top cover. This part has a fin that reaches to the bottom of the backpack. All the hardware is attached to the fin, and can be easily removed by removing the top cover. We exploited the findings on the **pose**, **grip**, and **weight balance** from the previous version; creating a new version of the grip that would improve robustness. From the previous version we knew that single-handed interaction can be ergonomically supported. This time, however, we needed to create a stable, robust single-handed grip. The grip part integrates single and two handed grips in a stable but massive construction. This grip allows the user to hold the device with one hand from below and operate the touch display or pen input single-handedly for short periods, or to operate it in two-handed mode using the build in controllers (a micro-joystick and two 1D-controllers). To change between single and two-handed operation modes, the user has to regrasp the unit.

This generation is produced using nylon-based STL in varying wall-thicknesses of between 3 and 4mm, with several solid parts (the grip). The model only holds few connections that are glued, and several nylon screws to connect the grip with the container, holding the cables and devices. The top of this container is screwed on with nylon screws.

### 4.4.3 Evaluations

This version was evaluated mainly through informal observations and interviews with at least 50 users. The unit was brought to numerous demonstrations of outdoor projects where specialists from industry and the public sector could test it. During these sessions, we noted observations and critiques from people working in the fields we are concerned with. In general, the unit robustness was appreciated. Throughout all evaluations people were confident that the unit was stable and would not break or fall apart. However, it was often criticized by its sheer weight and bulkiness. With respect to controllers and their allocation, the interactive aspect of the unit was well received except by researchers, since the controllers are just the necessary ones and do not allow further extension.

The controllers were found appropriate for prototype tasks that involved mainly browsing data, without having to control more than a couple of menus. From one of the associated projects, we received a critique as to why we were including joysticks and advanced controllers, if only a bunch of buttons would do (to just switch on and off some view modes). However, the main critique we got from experts was that they cannot fit it in a backpack or bag and, of course, its sheer weight make it unusable after 10 to 15 minutes.

**Discussion.** The first thing people note about this generation is how bulky it is. Even after adding a belt, it was still considered heavy and not ideal to operate. The fact that the pack needs to enclose all sensors implies considerable space requirements. A subtle matter that might have gone unnoticed is that we tried to keep the customized electronics to a minimum in this version. Consequently, the new platform had to allocate not only bulky connectors (USB vs on-board 5 pin connectors), but also lengths of cabling. The IC3 from Intersense alone comes with a 5 meter long cable, plus a serial-to-USB converter that is bigger than the sensor itself. The cabling and electronics increased the space requirements. Weight balance was still appropriate, but the extra material needed for robustness added to the cables and electronics. The differences with the old version were quite evident. By the time the Panasonic computer was tested on this platform, it was already considered heavy and bulky. This UMPC only added more weight to the setup. Even if the computer could run for four hours on batteries with all the sensors connected and generating AR content, the added weight and bulkiness reduced the user experience to below the older unit (less than 15 minutes). Production-wise, building the unit once all the parts are available is easier than for the previous version, but not at all a simple task. Finally, with respect to ergonomics, the waist-pose was increasingly used, which was no surprise considering the weight of the robust yet heavy setup.

## 4.5 Ice'R: compactness, portability and assembly

After the second generation, we had to revisit several issues. Strict requirements from end-users impose that the unit be portable in an **ergonomic** manner, in the sense that the user must be able to move relatively larger distances carrying the unit, albeit not while using it. Further restrictions on size of devices required the new generation to be more **compact**, to be carried more comfortably and stashed away easier: simply said, the previous version was too bulky. Furthermore, the device should be easy to assemble by an end-user, and could benefit from better production methods, hence reducing production time.

**Figure 4.3:** Ice'R: the different poses afforded by Ice'R, being two-handed at eye-height, single or two-handed with belt, and mounted on a tripod.

### 4.5.1 Requirements

This generation of the handheld platform had to react to the portability and production requirements being made. Exchanging a sensor or accessing cables should be a simple task and the device should be easy to produce to cut production time. Simultaneously, the device should become smaller and more portable. The Panasonic CF-U1 replaced the Sony as the preferred platform in light of its robustness, weather resistance and long runtime expectation; all three most wanted requirements for outdoor applications. With the increasing importance of the waist-pose, we also had to reconsider ergonomic issues, in particular pose and grip. As noticed while evaluating the first and second generations, in longer duration sessions users are often forced to take the waist-pose. The waist-pose implies a potentially dynamic angular offset between body and screen: the screen of a heavier device is held under different angles, to balance the weight and limit fatigue. This dynamic aspect affects the angle in which the camera has to point forward, which is generally fixed or difficult to change. At this stage of the design phase, finding a mechanism to dynamically angle the camera-orientation sensor pack was thought to be beneficial. With respect to compactness, the added weight of the computer and cables brought us back to a struggle with weight and space. Knowledge on grip and weight balance gained from the first platform was crucial to solve this problem: in particular weight-balance and tilt aspects should be taken into regard when dealing with dynamic angular offsets.

### 4.5.2 Design

To match portability and assembly aspects, we had to refine in particular the outer form and all movable parts. As a starting point, we reconsidered the **grips** of the Vesp'R:

however, this time we fixed the joystick-like grips to the body of the construction to avoid unwanted rotation of the handles. Furthermore, driven by the need to make the device more compact and easy to stash away, we avoided the usage of round and open forms: such forms take up too much space when stashing the device away. We ended up with a box-like form which is easily packable, yet still quite large. The platform was designed to be used in three alternative **poses**: strapped to the body with touch/pen input, held at arms length and closer to the eyes for shorter interaction sequences, or placed on top of a tripod. The first pose makes use of a strap around the back that keeps the unit in-place against the waist of the user, while leaving the hands free for operation or balancing the handheld construction. The second pose relies on the grips and a joystick input. To support the tripod mount, we embedded a tripod connector below the handheld platform. We envisioned that users would rely mostly on the waist-pose while surveying and only use eye-level pose for short periods at a time. This pattern of usage allows for longer sessions, because the arms can rest while in waist-pose. Furthermore, the tripod mount affords continuous usage and collaboration with others, since multiple users can look at the (small) screen, and no ergonomic conflicts are caused by weight.

This generation applies a slider mechanism to put the parts together, simplifying production and reassembly, while still being weather resistant. A single panel at the back covers all the sensors and cabling that do not need to be accessed for normal operation. For the first time, we also focused on a robust yet flexible mechanism for the camera-orientation sensor pack. Both are mounted at the side of the computer with a lever mechanism to allow changing inclination of the camera-orientation sensor pack while using the device, or to stash it away for transport. Mounting it at the side of the computer greatly reduces space at the back, making the unit flat and rectangular. We also included a detachable antenna for the GPS unit: during transportation, the antenna takes too much space and thus increases the overall **size** and is a potential source for damage. Finally, to reduce problems with custom **controllers** (loose cables, size) we dismantled a small controller (a Genius Maxfire Pandora Pro) and placed it in a small box that could be slid into the left-grip of the device. The controller box holds several buttons and a mini joystick that can be reached well.

This generation is produced using nylon-based STL with a wall-thickness of around 3mm. The model only needs to be glued on few places - most other parts are slid in. Miniature magnets were used for opening and closing the lid on top of the camera-orientation sensor pack: test showed the magnets do not disturb the orientation sensor when placed away far enough.

### 4.5.3 Evaluations

We performed a range of structured interviews with around 40 expert end-users. During the interviews we discovered that particularly those users that need to travel with light luggage found the device still to be too bulky. They suggested a modular platform: users wanted to put together their own display-sensor combination to save space by disconnecting unused devices, or exchange devices for smaller versions in less demanding applications. Experts were happy with accessing every part of the platform easily, but felt uncomfortable operating the unit. In particular the mobile part for the camera tilting resulted in undesired constant corrections.

**Discussion.** The flat form factor of this generation makes it easy to pack in a backpack, however, it is still very bulky. Although its form factor is not suitable for longer operation in grip mode, in the waist-pose and when alternating poses the operation can be extended to longer periods. Most handheld AR applications require the user to constantly make correlations between the real world and the view on the screen. This effect is accentuated when operating in waist-pose, because the user is facing down directly, while the camera is pointing forward: the user needs to look up and down frequently. Production-wise, this generation proved to be very successful - the limited glue connections and the sliders reduced production time to a few hours. Finally, we were not satisfied again with the controller box, which still causes too much trouble: the small size of the buttons often causes problems while pressing.

## 4.6 Cool'R: modularity

**Figure 4.4:** Cool'R: device with connected sensors, device taken apart showing different modular boxes, lens protection and open neoprene hull, and the single-handed pose with pen operation.

The goal for the last version was to devise the simplest construction possible, allowing a user to attach and remove sensors from the unit with ease: the construction should be as compact as possible. As such, a truly modular construction had to be found. Furthermore, weight restrictions made us consider different materials to reduce weight where possible.

### 4.6.1 Requirements

The main goal of the last handheld generation was to contrive a **compact** yet **robust** platform that could be dynamically modified based on user needs. Still, the setup had to be resistant to external influences such as weather, dirt, and object protrusion. Based on the comments on the third generation, we had to compress the size of the platform as much as possible. Users should be able to attach/detach various kinds of sensors: different tasks may require different sensors, including unforeseen ones such as a temperature or moisture measurement devices. Some users require the construction to be stripped down to the minimum: field workers often are constrained in what they can carry, in particular when exploring remote sites. Hence, the construction had to be **modular** in such a way that leaving out a sensor would actually reduce the overall size: here, “modular” should be understood as a product of extensibility and compactness.

Coming up with device that is both ergonomic and robust, yet also compact and modular is far more challenging than it may seem at first sight. As our experiences show, particularly compactness is hard to achieve, though very important to keep the device transportable and improve user acceptance. Furthermore, compactness should not minimize **ergonomics**.

### 4.6.2 Design

After analyzing different possibilities to mount external sensors and possible controllers in the previous devices, we decided to strip as much material as possible. We wanted to remove all parts that have a function that is performed by other or smaller parts to save **weight**. We decided to use the **grip** of the computer itself: the Panasonic affords a good, close to power grip due to its thickness, which is consistent with our findings for the Vesp’R (diameter of around 76mm). Thus, we avoided the usage of external handles. Nevertheless, we also had to devise a construction to attach external sensors without fully blocking the back of the UMPC. The result is an X-shaped exoskeleton mounted behind the UMPC, which leaves the hands free to grab the device from both sides. The exoskeleton also allows easier access to the hot swappable batteries of the UMPC, which was impracticable with the older versions that completely covered the back.

Unfortunately, the exoskeleton has a reduced amount of material, rendering it vulnerable to structural damage from the weight of modules attached to it. To withstand forces caused by weight, we strengthen it with an aluminum-enforced backbone. It gives the central part of the exoskeleton its stiffness against bending, and strength without uselessly increasing the weight.

Bound by the **pose** users take when interacting with the device, the next stage was to define attachment points for the sensors. In particular the camera and the GPS antenna need to be mounted in a particular way: the camera needs to point forward, and blocking the tracking signals should be avoided. Users tend to hold the device in a 60-degree angle during operation: as such, reflections by ambient light can be avoided, and users can interact with the screen using an ergonomic wrist angle. To relieve the weight on the wrist we rely a belt that connects diagonally to the UMPC, allowing direct access to the screen by the dominant hand. The belt also allows the user to rest the arms in the hip, in "waist-pose", which has shown to be very beneficial in previous generations.

We added a slider mechanism on the topside of the exoskeleton: the user can easily slide devices in and out that are needed for the task at hand. The variety of boxes we generated to protect the different camera-orientation sensor pack, and the GPS sensor all have a 30 degree angle; pointing forward when the user holds the device. This time, we also made a considerably smaller box for a webcam-orientation sensor pack that can be used when slightly lower camera footage quality is still acceptable.

Unless a vision-only tracking solution is used to estimate pose, additional devices for tracking generate a considerable bulk of cable that need to be stashed. As noted before, the volume of cables varies widely (see Table 4.1), leading to large casing in previous experiences. To keep the device as compact as possible, we decided to use a neoprene hull. A user can mount a simple and small bracket below the UMPC construction to which a deformable piece of neoprene is attached. Cables can easily be stashed in the neoprene form and folded behind the UMPC, taking up relatively little space. The hull affords grabbing the device construction from the back single-handed, leaving the other hand available for other tasks. This pose was found ergonomic and useful in the first and second generations. Neoprene is extremely robust and weatherproof: it can hardly be ripped or punctuated, is waterproof, and absorbs shocks hence protects the devices inside. Furthermore, due to its flexibility and stretchiness, we can actually compact the cables and additional devices by tightly folding the hull: the neoprene will compress everything to a limited amount of space without damaging the contents, which is impossible with any solid material we could have used. Hence, the neoprene "scales" well with additional devices by always compressing to the minimum space and thus **size**. Finally, we did not add any **controllers** to the unit this time: The UMPC has several buttons that can be reached well, and pen-input is possible using the grip and weight relieve by the belt. The

**Table 4.6:** Dimensions and weight of each platform without electronics.

| Platform | Dimensions ( w x l x h) | Weight (cm3 material) |
|---|---|---|
| Vesp'R | 17.15cm x31cm x19.4cm | 410 gr. (447.86 cm3) |
| Bulk'R | 22cm x31 x25cm | 1130 gr. (1235.10 cm3) |
| Ice'R | 8.3cm x 36cm x18.6 | 800 gr. (870.65 cm3) |
| Cool'R | 12 cm x 19.5cm x 24cm | 252 gr. (273.35 cm3) |

main construction, the exoskeleton and the boxes, were printed in nylon-based STL. Wall thickness is around 4mm for all structural parts, and 3mm for the boxes. The exoskeleton has a 2mm thick aluminum profile glued in for stability. The hull is made from 4.5mm thick neoprene.

**Discussion**. This generation has not been evaluated yet, though we can report on several initial observations. The combination of different materials in the device construction guarantees that the device is robust, compact, and easy to produce. Due to the mix of materials, we were also able to considerably reduce the **weight** of the construction in comparison to the last two models that were designed for robustness (see Table 4.6). Enclosing each module in its own hull helps isolate and protect delicate hardware. However, we still need to evaluate the potential blocking of signals with the new GPS position, and the effects of the singular pose the device affords: users cannot look closer at the content on the screen while holding the device at eye height, since the camera would be facing up instead of forward.

## 4.7 Lessons learned

Throughout multiple phases of analysis, design and evaluation, we gathered experience on different aspects of improving handheld AR setups. In this section, we summarize the lessons learned, hoping that other researchers can benefit from the gained knowledge.

The way the users hold the device, the **grip**, seems mostly associated with the primarily navigation-oriented task space. Even when performance of fine-grained actions is not needed, most users tend to hold the device in a power grip - there is not always a need for a precision grip to perform actions. The particularly thicker Panasonic UMPC was beneficial to apply a power grip on the device itself: the thickness and robustness of the device nicely affords a power grip, whereas the smaller and less robust Sony UMPC proved too small and unstable. The robustness of the Panasonic thereby reduced the necessity for an external grip (handle), and provided a solid backbone for mounting the additional construction that holds the sensors. Thus, the bigger and heavier Panasonics saved us some weight and size at the end, while still affording a good grip.

**Lessons learned:** provide a power grip to hold the construction. Reuse, if possible, the device grip to save weight and size of the total construction.

A secondary grip to hold the device single-handedly appeared in the first two generations. Though some users make use of the single-handed grip to hold the construction continuously, it is used mainly to stabilize: the non-dominant hand stabilizes the device for the dominant hand to perform actions. This kind of behavior was originally studied by Guiard [Guiard, 1987], and applies to two-handed interaction in general. The secondary grip relates directly to the dominant **pose** the users seem to take: to avoid hand trembling and fatigue caused by the weight of the platform, users tend to hold the device at an angle at waist-height. Hereby, users tend to rest their elbows in the sides against their hips, resulting in what we call the ”waist-pose”. To further improve this pose, from the second generation on we introduced a diagonal belt that distributes the weight to the shoulders and back: in particular with pen-input, the stabilizing pose combined with the belt proved useful. Moving the device closer to the eyes and at eye height was infrequent, and may be associated with the small size of the screen: during navigation, exploratory actions do not require much detail on the screen. Once a particular asset is noticed, often the device needs to be moved closer to the eyes to observe it in more detail, hence improving the visibility. Similarly, the pose may change under effect of reflections, which are still a significant hindering factor in outdoor scenarios. The pose will require the allocation of the camera at an angle that allows it to observe the real world. Though we experimented with movable parts, finally we ended up with a fixed angle: the camera is often equipped with a wide-angle lens, affording a wider angular range of operation before objects are outside view.

**Lessons learned:** support the use of the waist-pose in combination with a belt, and rotate the sensors adequately.

Not surprisingly, the pose is often also dictated by the allocation and usage of **controls**. In particular with pen-input, the grip and pose on the handheld construction needs to be changed considerably. Similarly, with badly located controllers, users will need to angle their wrists in unnatural angles [Veas and Kruijff, 2008]. The number and allocation of controls is highly dependent on the kind and frequency of tasks. For exploratory tasks (navigation), a simple control structure is often adequate. Similarly, for short task sessions, users can cope often with badly located or general-purpose controllers like micro-joysticks or buttons. Once applications become more complex, controls or interaction techniques should match the control structure: when touch screens are available, they can often match the complexity by providing adequate menus. In relation to the power grip, we can often refer to game joysticks, which generally have a good grip and ergonomic allocation of controls. However, care should be taken when building in controls: they are not only difficult to build-in in a compact manner, but also the electronics and cabling have to be planned thoroughly to reduce changes of malfunctioning. Specially tailored electronics require assembly time and can not be done by non-experts.

**Lessons learned:** carefully plan controller allocation to reduce fatigue and pose changes. Use off-the-shelf controllers when available to simplify assembly.

Size is in direct relation to weight: the smaller setups we created were obviously lighter than the bigger ones. Nowadays, an influential factor on size acceptance is the mindset of end-users that is often influenced by mobile phones. Not necessarily understanding the technical foundations and differences between mobile phones and UMPCs, it is often difficult to explain without showing end-users the effects of a larger and better screen, more processing power and better sensors: the required quality of sensors and screen size often becomes evident when end-users use the platforms for a longer time. Notwithstanding, cell phones and UMPCs are two different platforms.

An alternative that proved most beneficial was to strive for full modularity. Attaching and removing sensors with their encapsulation saves space and weight once it can be achieved in a robust enough way: till now, our latest platform generation has proven well. The reader might wonder why modularity came in last in our list of requirements, only after a few iterations. The fact is that for the first generations, the set of sensors used for AR was pretty much fixed by project requirements. These sensors had been tested and provided the best performance at that time, thus we designed the first prototypes mainly for them, leaving extra space for device controllers that were planned before hand. Modularity often comes at the cost of robustness. Even though from the first version we used mechanisms to attach parts (e.g. the handles on Vesp'R), the first iterations of this mechanism were unsuccessful (either too unstable or too cumbersome to operate). Modularity at the level of attaching a module for each sensor was impossible under such circumstances, and only after experimenting with a sliding mechanism in Ice'R we could start thinking about modularity as a possibility. Once a robust mechanism is available, the sensors can be grouped in functional units to create modules, or they can also be considered each a different module. The last generation of our platform uses a sliding attach/detach mechanism and STL boxes for modules. By using removable boxes for sensors, potential transportation problems caused by parts that stick out can be avoided, increasing robustness when not in use.

**Lessons learned:** when striving for modularity, first experiment with the mechanism to attach/detach modules until it proves to be robust. Subsequently, create modules out of functional units and plan their allocation with respect to weight balance.

With the aim to protect in particular the sensors, robustness often correlates with **weight**: the more robust the construction needs to be, the higher the weight, especially when complex production and assembly methods such as molds cannot be used. Stereo lithography (STL) is a useful method to rapidly create constructions that are not necessarily too heavy. Nylon-based STL weights about 0.9 gr per cm3, hence, larger constructions can be made within weight boundaries: we used between 273.35 and 1235.1 cm3 for our constructions (see Table 4.6). Nonetheless, combining STL with other materials is advisable: aluminum is light and can make the construction considerably stronger when used as aluminum-nylon composite, avoiding potentially far thicker STL models. Neoprene is an excellent choice for robust and flexible ”containers”, compressing the contained parts. Not to be forgotten is the effect of the battery life of different platforms on weight: the Panasonic comes with two hot swappable batteries that have a long runtime. Hereby, we avoided the necessity for a battery-powered USB-hub, which requires changing the batteries every 90 minutes: battery and cabling consume space and increase weight. With respect to **size**, the main denominator is mostly not the sensor itself, but the cabling. Some devices come with meter-long cables (up to 5 meters / 15 feet), which can often not be shortened due to warranty issues. We often ended up with wasting about 2/3 of our total encapsulation for cabling. Thus, as an effect of the externalization of higher-quality sensors we often end up with major size restrictions. With the Vesp’R we successfully tried to shorten most cables, however, often had to deal with production and operation difficulties due to broken or loose cables. Nonetheless, we have greatly improved the weight and size of the devices, as well as the production of all parts: STL produces professionally looking prototypes, and is still affordable. Hereby, a good-looking prototype often helps: end-users certainly are more positive towards good-looking prototypes [Norman, 1990]. Furthermore, we have replaced most screwing connections with sliding mechanisms or glue: whereas it took around 2 to 3 days to build the first generations, we can now make a device in around 2 hours.

**Lessons learned:** Mixed materials are great for reducing weight and size of setups, yet still protecting the devices against external influences, and can often be produced easily. Cabling currently is the main factor in the external construction: devices with short cables can save much space.

## 4.8 Summary

In this chapter, we presented the results of designing four generations of hardware platforms for handheld AR. Based on the numerous iterations that were part of each design cycle, we extracted requirements and analyzed the interrelationship between key factors: ergonomics, robustness, compactness and modularity. We have visited evaluations, both formal and informal, that validated each iterations. Each platform with its motivation and requirements has elevated the trade-off between the key factors, influencing the resulting design. We have described valuable lessons learned throughout this process, but the main contribution of this work is in the study of patterns of interaction: the poses users take, the role of ergonomics and controller allocation.

Regarding the platform maturity, we believe we have reached a level that opens immense possibilities. Not only are we testing our platforms in research installations, but also bringing them to the field and pushing them to extreme situations. The new, modular platform opens up new possibilities to test diverse sensors and controls under extreme conditions.

# Chapter 5

# Environmental Data Visualization and Handheld AR

Visually conveying the information delivered by real-time sensors in a comprehensible manner is not an easy task. In fact, intrinsic characteristics of data, its sheer vastness, multi dimensionality and high spatial distribution make visualization one of the most demanding tasks, both for the user and for the platform. This chapter concentrates on the visual presentation of spatial data in outdoor AR. We prioritize visual results, while at the same time attending at the performance and hardware restrictions of the mobile platform. After all, AR is an interactive paradigm, and must be if the user is to visualize vast, extensive datasets on a small screen, as described in chapter 7.

Scientific visualization intends to translate the set of numbers in a spatial dataset, generated by a scientific process, into a graphical representation which allows interpretation of the data characteristics [Gershon and Eick, 1995]. Unlike numerical presentation, visualization allows the user to better understand fast changing information. Our research exploits the spatial reference of the dataset to place it in its real world context, effectively converting the real world in part of the data. In the first part of the chapter, we concentrate on how to successfully accomplish this combination for the case of sensor measurements, 2D overlays and 3D data. The topics are constructed from initial mock-ups, going through representations validated with experts, as explained in chapter 3.

Understanding the situation is not only encumbered by heterogeneity of data, but also by other extrinsic attributes that have several cognitive implications. The interplay between background imaging (video, 3d model or map), and foreground visuals such as sensor data visualization and user interfaces raises visual complexity. In outdoor situations, color and contrast play an important role in the “readability” of information. Regarding visual perception, the main interest lies in conveying/analyzing how the data interacts with the real world. Thus, after setting the basic techniques, we concentrate on well-known perceptual issues of AR, shape perception and occlusion, as applied to the case of scientific data visualization. To simplify its combination with the real world, the representation of continuous data is based on color dimension or color coding, the dimension most commonly used in the field. Based on this, we analyze techniques to reduce the interference and coherently convey spatial properties such as shape and topology.

When visualizing large amounts of data, it becomes crucial to discriminate and isolate interesting portions. Filtering methods in AR serve to avoid clutter, information overload, and to reduce massive datasets to subsets manageable within the performance constraints of mobile platforms. Data filtering methods behave differently depending on the type of data. In section 5.1.3, we analyze filtering methods for sensor data. Chapter 6 concentrates on visual discrimination of data guided by subtle modifications of the visual input.

## 5.1 Sensor Data Visualization

The goal behind visualizing sensor measurements in AR is to convey the semantics of discrete measurements in the context of the real world, e.g., given by their spatial relationships with real world entities. Sensor measurements correspond to measurements at specific locations, that is, they are discrete in nature (spatially and temporally). For example, in some cases a measurement can result from structure or objects in the spatial vicinity of the measured location, like increased pollutants due to waste disposal, or temperature changes after building certain structures. For AR to work in these situations, sensor measurements have to be spatially registered in 3D at the geo-located position.

### 5.1.1 Background: Visualizing Discrete Measurements

The representation of sensor measurements raises several interesting challenges. First and foremost, in terms of *legibility*, any representation, graphical or textual, must have discernible features to clearly convey its meaning. One can naïvely consider the meaning to be “at position XYZ, the measurement of sensor V was W units”. However, close consideration reveals that the goal in reality is more complex. For example, regarding location, we want to convey the position of the sensor relative to the user, to other real-world objects, and to abstract (virtual) objects that form its spatial context. Furthermore, the sensor has a data type and other measuring context.

Legibility is limited by physical properties of the display and issues that affect perception such as foreground-background interference and visual clutter. Foreground-background interference occurs when overlaying an image (foreground) onto another (background) with similar color properties. In a dynamic environment, such as outdoor AR, the color properties of the background constantly change, complicating legibility.

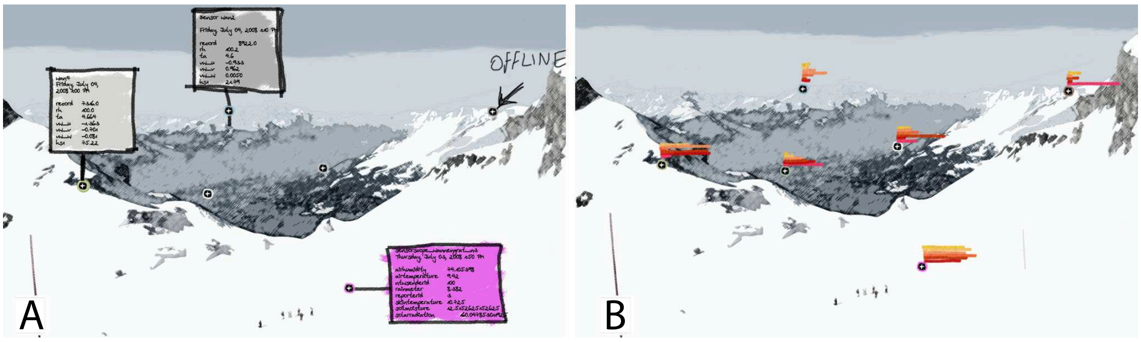


**Figure 5.1:** Sensor visualization concept. Labels grouping all data available for a station (A). Mini plots of time series for sensors in a station (B).

Visual clutter leads to information overload, where a person fails to assimilate information, undermining situation awareness. The natural response is to try and reduce the amount of information to be assimilated at a given point in time, to recover awareness. In the case of visual search, when more than one separable feature is required to find a specific item in a display, items must inspected serially [Treisman and Gelade, 1980]. Thus, with increasing number of items, search time increases. Furthermore, nearby similar stimuli are not processed independently by the human visual cortex [Kastner et al., 2001], and can cause sensory suppression. This implies that the absence of visual features uniquely identifying objects or subsets thereof gives rise to visual clutter. Furthermore, with increasing number of objects, information density increases and objects overlap to such an extent that identifying objects might become impossible [Peterson, 2009]. In summary, the representation for sensor measurements aims at conveying the value within its spatial context of virtual objects (e.g., other measurements) and physical entities in the real world, in a readable, organized manner. We are searching for the minimal representation that retains legibility, and avoids clutter.

### 5.1.2 Sensor Display Prototypes

In an attempt to follow well-known means accepted by our expert community, we chose to represent sensor measurements with 2D labels, featuring textual, graphical or combined content. Alternatively, one could devise and evaluate novel 3D representations for each type of data and their combinations, following the work of White and Feiner [White, 2009].

In most cases in our scenarios, sensors are physically connected to a sensor station, a construct that provides basic infrastructure such as energy, and network connectivity to a bunch of sensors. The measurements of sensors are located at the geographic position of the corresponding station. Therefore, out of discussions with experts during the def-

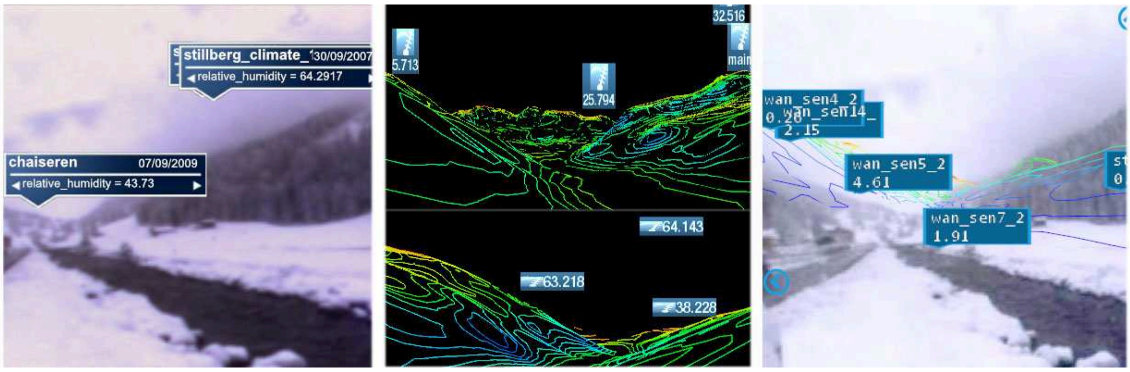


**Figure 5.2:** Label Prototypes: The first prototype included contextual information such as sensor name, data type and timestamp (Left). During tests of icon and measurement labels (Middle), we experimented with vertical and horizontal layout (M-Top and Bottom respectively). The second prototype conveyed sensor name and measurement (Right).

inition phase, the conceptual representation of sensor measurements grouped all sensors corresponding to a station in a single label, as the sketch in Figure 5.1 A shows. We also speculated that we could graphically represent time series of measurements in mini plot labels, as shown in Figure 5.1 B. The latter could display measurements at specific time intervals for a single sensor, or latest measurements from different sensors.

Legibility issues of individual items and the physical capacity of the screen hampered the conceptual representations. The full station label shown in Figure 5.1 A, when displayed with a legible font at the resolution of the device's screen, could occupy the half of the screen for a single station. Similarly, in the case of the mini plots, early tests showed that it would be difficult to interpret the values, and rendering them in an intelligible manner takes a large portion of the screen. The concept of graphically representing measurements in AR was discontinued after these initial results. In the following, we concentrate on the label representations, and return to graphs at the end of the section.

For the first prototype, we decided to include a single sensor reading per label. Each label included data complementary to the sensor reading, such as station name, data type, and timestamp, as shown in Figure 5.2 Left. The expert workshop evaluation of this prototype revealed that, in terms of legibility, the representation was well readable, but the screen was quite cluttered with a low number of labels (around eight).

In practice, studies with our partners revealed that we would need to represent a maximum of 15∼20 labels. To increase the number of labels without cluttering the screen, an icon based representation was devised, showing an icon for the data type and a number for the reading. The layout could be adapted to vertical or horizontal, allowing to reduce the screen coverage in one of these dimensions (see Figure 5.2). As expected, vertical layout reduces clutter in the horizontal dimension. Conversely, horizontal layout, reduces it in the vertical dimension. This representation required icons for each of the datatypes used. However, end-users had a different request, although they found these labels interesting, the minimal information had to include the station name.

Based on these experiences, we designed the label representation for showcase interviews, that experts validated with positive ratings. It included an abbreviated station name and a single reading (shown in Figure 5.2 Right). The data type was incorporated in an info panel for visual feedback. A spin crossbar was used to change the datatype as described in chapter 7. Selecting or querying a station (by tapping the label or icon) brings up a summary table collecting all data for the station, as was the original idea. The table included the timestamp among the other sensor readings.

It is interesting to consider these representations in light of the desired goal, to convey the meaning of measurements in connection with the real world. Clearly, showing all the information for each station would not have worked, since the user is most ofter interested in one or two datatypes at a time. Comparing them with different datatypes in another station also makes little sense. In contrast, having a single reading per station allows the user to see more stations, and compare readings of the same datatype. If they need details, they can always tap on a station, and get the full description. This format is inherently a filter by datatype mechanism: e.g., a user can display all stations that report temperature.

### 5.1.3 Alternative Representations for Visibility Management

The last iteration is already a step forward in determining the meaning in labels and how sensor measurements are communicated. It deploys a filter by datatype that allows visualization of related items, and a visibility strategy that gives users options to reduce clutter. The visibility management strategy becomes crucial to avoid visual clutter and information overload. One approach would be to deploy automatic filtering algorithms to only show information which is important to the user at a given time [Julier et al., 2002], or to reorganize objects such that information transfer is improved [Bell et al., 2001]. However, automatic methods require an expert definition of a set of criteria upon which objects can be removed [Julier et al., 2002], or accurate models of all the information displayed (virtual and real) to reorganize it appropriately [Bell et al., 2001]. For exploratory visualization, one of our main interests, this is unfeasible. Our experts emphasized that they want to be in control of what information is filtered out, and the decision will be reached only at run-time.

In general, clutter can be managed by controlling the visual dimensions used to represent labels, (e.g., size, transparency, color, intensity) or by filtering [Peterson, 2009]. For the final prototype, we developed alternative formats and filtering techniques that the user can control at will (see Figure 5.3). A minimal, iconic format indicates only the position

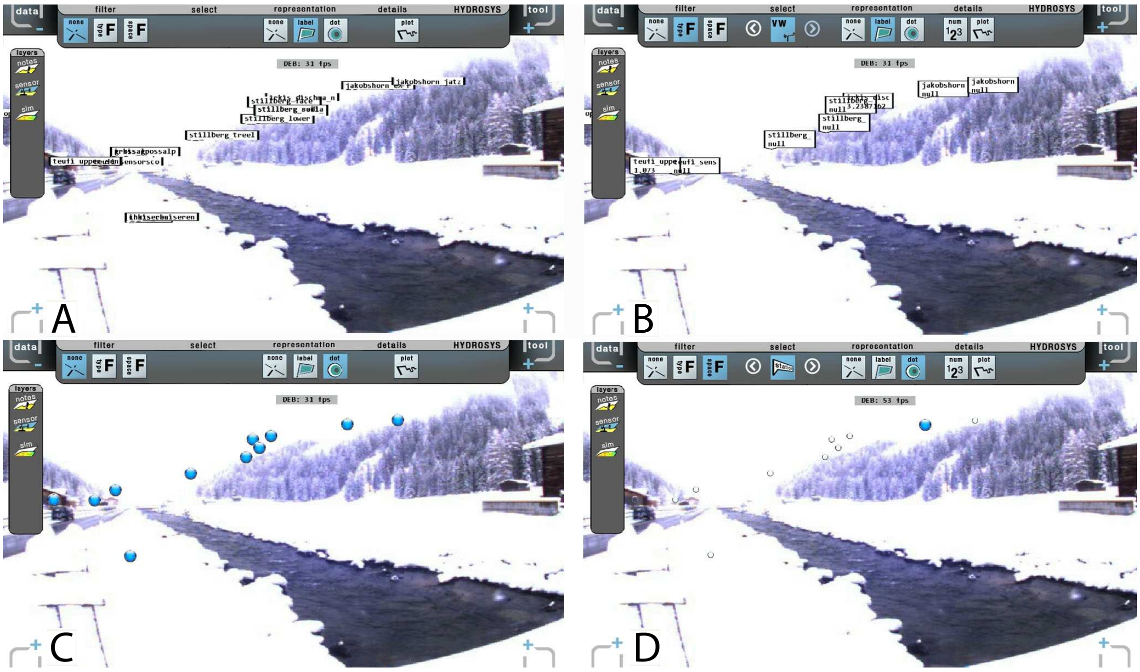


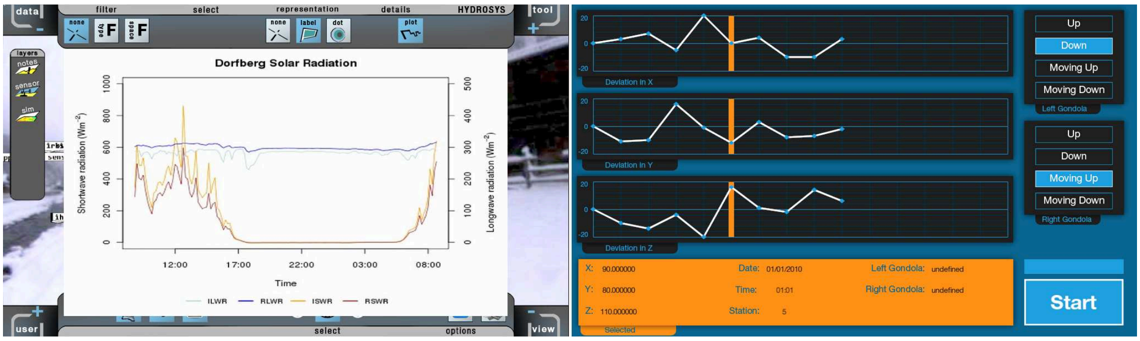
**Figure 5.3:** Sensor visualizations. A: station name. B: station name and single reading. C: station locations. D: station locations and selection.

of stations (Figure 5.3 A). An informative format shows station names (Figure 5.3 B), while an extended format shows each label with a single reading for a common data type (Figure 5.3 C). In addition, the user can filter stations spatially by browsing through a spin menu, or based on proximity by clicking on the screen. Thereby, all but the selected station are displayed as icons, while the selected station shows informative data in the AR view (Figure 5.3 D).

Thus, the options to leave some information out are left to the user, who can choose among different filters to explore the information space following the desired strategy. These visualizations were validated in a formal evaluation, where participants had to find data and relate it to the real world.

### 5.1.4 Accumulated Measurements Display

In the concept sketches, a tool to draw mini plots of accumulated sensor values was included, see Figure 5.1 C and D. Two issues conspired against its deployment. First, the initial data pipeline was not devised to obtain a time series of data in a sustainable way, due to high bandwidth costs. Second and most influential, we found that the plots would be difficult to interpret after initial tests of the screen form factor and physical properties of the deployment (i.e., screen resolution vs size, outdoor lighting). Instead, the plot



X: 90.000000
Y: 80.000000
Z: 110.000000
Date: 01/01/2010
Time: 01:01
Station: 5
Left Gondola: undefined
Right Gondola: undefined

**Figure 5.4:** Sensor Plots. Left: plots queried from GSN . Right: real-time plot for mobile sensor, in this case, plotting the inclination variability measured with the orientations sensor.

mode can be queried to display graphs comparing measurements of multiple sensors for a single station or across stations. These plots are computed by the sensor network, and provided in image format, which reduces the payload on sensitive parts of the network preventing the transfer of large amounts of data all the way to clients, see Figure 5.4 B. However, when a mobile sensor interfaces directly to the mobile AR platform, a line plot functionality stores and plots latest readings. Figure 5.4 C shows plotted inclination data measured with an InertiaCube orientation sensor.

## 5.2 Overlay Visualization

Overlays originate from numerical processes executed over a range of measurements, such as simple interpolations and simulations. They can also originate from authorities (e.g., land use) or other users. This section considers 2D and 3D output. Along the lines of the previous section for 1D measurements, representing 2D and 3D data in AR aims to convey the relationship of these data with the world, e.g., to compare or associate results with physical entities. In the case of 3D, one can observe the effect these results have on structure. For example, the Kylmäoja hydraulic model outputs changes in water level at different scales. Viewing these results overlaid gives a good idea of the distribution of water during a rain event with the specified parameters (e.g., 70mm/hr). In the case of 2D, consider, for example, viewing a land use color map on-site describes the surrounding topology, roads, public infrastructure, and community facilities, among others.

To create a visualization, data are mapped to visually perceivable elements in a graphic representation, often a surface containing the data. The individual elements that can be used for mapping are called the visual *dimensions*, as defined by Acevedo [Acevedo, 2007]. Much of the research in visualization goes into studying what visual dimen-

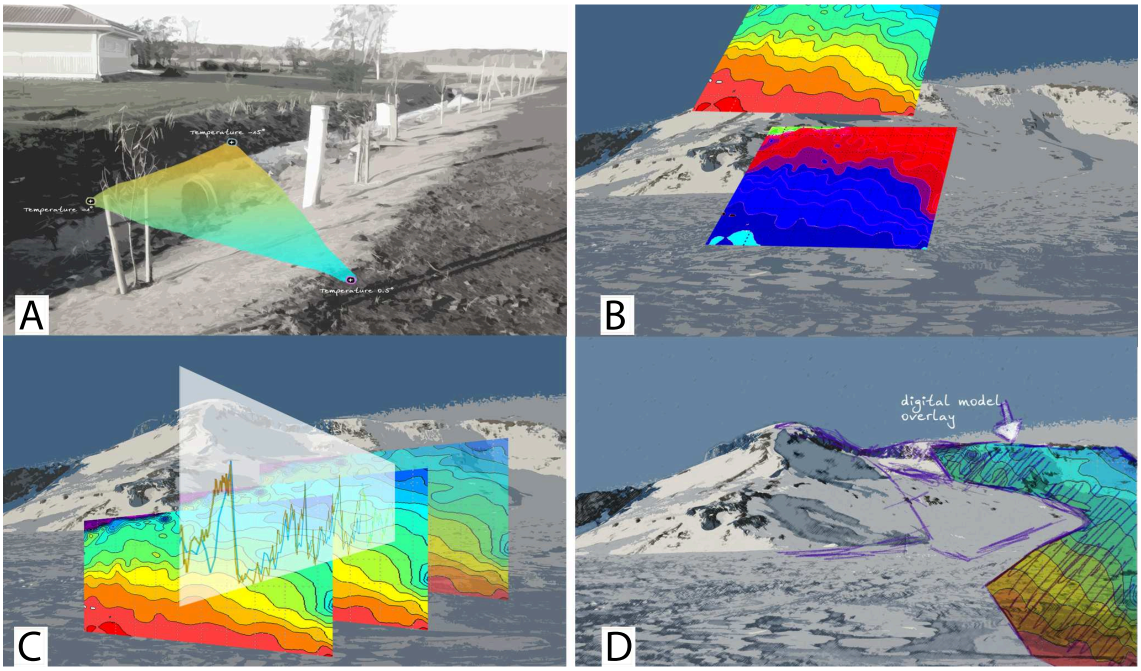


**Figure 5.5:** Overlay visualization concept. A tapestry interpolation covering a small portion of the video (A). A stack of 2D data representations shown in-place without registration (B). A time series correlated with a plot (C). A data interpolation registered with the real world (D).

sions can be used to convey information, how can they be combined to represent multiple datasets, and which are the most efficient ones. The predominant way to represent a scalar field on a 2D surface is color. It is the main dimension used by the overlays presented in this thesis. Alternatively, other dimensions studied in visualization include texture [Interrante, 2000] and glyphs varying in size, distribution and orientation [Acevedo, 2007]. Their integration and interplay with video background in AR, as well as their interpretation for the users, remain an open research direction.

To convey the meaning of 2D and 3D data in relation to the real world, it is necessary to register 2D data to some 3D representation of the terrain, and both of them in relation with the perspective of the user. Several conceptual sketches were produced for overlays. We assumed that simple interpolations could be visualized as a tapestry, with minimal registration on the terrain (see Figure 5.5 A). However, it turns out that a large number of values are actually needed for interpolations to make sense, such that the sensors cover a wide area. Then, their visualization is more clear if correctly registered. We also considered different ways to stack, or combine results from different processes, as shown in Figures 5.5 B and C. The main drawback here lies in the amount of screen space each requires. In the case of Figure 5.5 C, displaying the progression of a simulation in time was

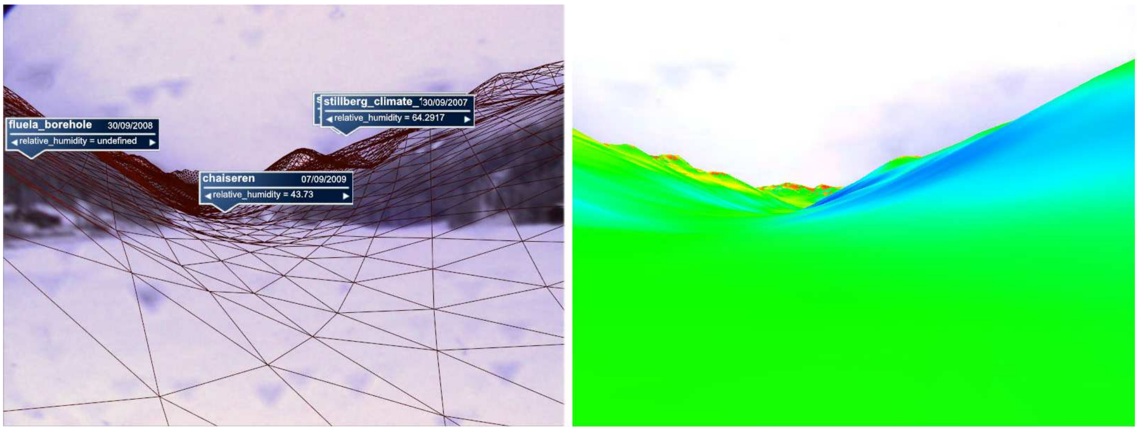


**Figure 5.6:** Visualization Prototypes. The first prototype dealt with registration, rendering a wireframe 3D model over the video frame (Left). Registering results from interpolations and simulations with a 3D model gives a colormap of the terrain in egocentric mode with poor shape cues (Right).

later implemented as in-place animation, still these cannot be shown concurrently. The sketch shown in Figure 5.5 D relays the intention to map the 2D representation onto the real world, which is what our AR representation pursues. For experts, this was the most puzzling and interesting representation, for they can well understand distributions of colors, but the AR experience brought a new combination of several aspects that they would need to experience first in order to comprehend (these were just sketches). The sketch in Figure 5.5 D also shows the concept of applying the overlay to a selected, segmented section of the real-world, this concept is not further developed in the thesis.

The first visualization integrating 3D data with the real world was based on a wireframe rendering of the 3D model (see Figure 5.6 A), without color space modulation (i.e., no other data mapped on it). The wireframe of such a dense model already conveys some structure information, that serves to make sense of the spatial placement of labels, for example. But, it is so dense, that it effectively ends confusing users, as we learned during expert workshops. In fact, this was never meant as an end result, but as a way to actually showcase what AR really is, since the static nature of sketches gave a wrong impression of a manually post computed, non real-time method.

For the next prototype, our new engine supported the visualization of a textured DTM of $36km^2$ ($6km \times 6km$)at a resolution of a sample per 10m at 30fps. The transcoding of data took approximately three minutes and was done off line in a high end desktop computer. The textures for the DTM were obtained from common interpolations used in the field, such as snow surface skin temperature, solar wave radiation, or from simulation outputs such as the hydraulic model described in chapter 3 for the Kylmäoja scenario.

The domain expert can filter and modify the type of simulation by browsing through a spin menu. Currently, only one simulation can be displayed at a time. Multi-modal visualization is supported by combining overlays with sensor visualization. For example, by displaying a surface skin temperature interpolation overlay with solar radiation measurements, one can analyze the effect of solar radiation.

Often in geosciences, the 3D models from simulations do not reflect completely the real world, for example DTMs and DEMs lack structures such as buildings. By overlaying a 3D model and 2D texture data covering a large area, we have effectively created a self orienting, egocentric 3D representation of the data, mostly occluding the video background (see Figure 5.6). Without landmarks in the AR view, it is impossible to establish its relationships with what one sees with the naked eye. The next section relates the techniques deployed to compensate for these effects.

## 5.2.1 Perceptual Optimizations

The choice of AR brings about several perceptual issues that plague AR visualizations. In particular, mobile AR is affected by depth distortions related to incorrect shape interpretation of virtual objects, and visibility issues related to screen problems and physical conditions of outdoor environments. Perceptual optimization aims at improving the visibility of content rendered on the screen, that is, the visualizations of environmental data. This section deals with issues of color optimization, using isophotes and transparencies to improve visual coherence.

### 5.2.1.1 Managing Occlusion and Shape Perception of Large Overlays

In our case, when activating an overlay, the 3D terrain is rendered from the perspective of the video camera. This is the normal AR approach, since 3D models often span a fraction of the visible video. But the terrain model spans a wide area, most often occluding a large portion of the video. The literature on mobile AR in these cases uses simplified wire-frame models [Höllerer and Feiner, 2004], suitable to convey structure of buildings and urban models. In the case of environmental monitoring, 3D models often represent only elevation (lack structures). In addition, to make sense of the situation, the user needs to see the color mapping of the result.

We overcome the issue with a dual solution, using the multi-pass renderer described in section 5.3. The user is given control of the opacity with which the terrain model is rendered. The terrain can then be viewed in full color, or semitransparent allowing to view the real world underneath. Thereby, a user can see trees on a slope or a barn next to the river, where the digital model just displays a flatland, as shown in Figure 5.8.

Nonetheless, a 3D terrain flat-shaded with artificial colors offers particularly poor shape cues, even more so when rendered semitransparent. Shape perception of 3D scenes

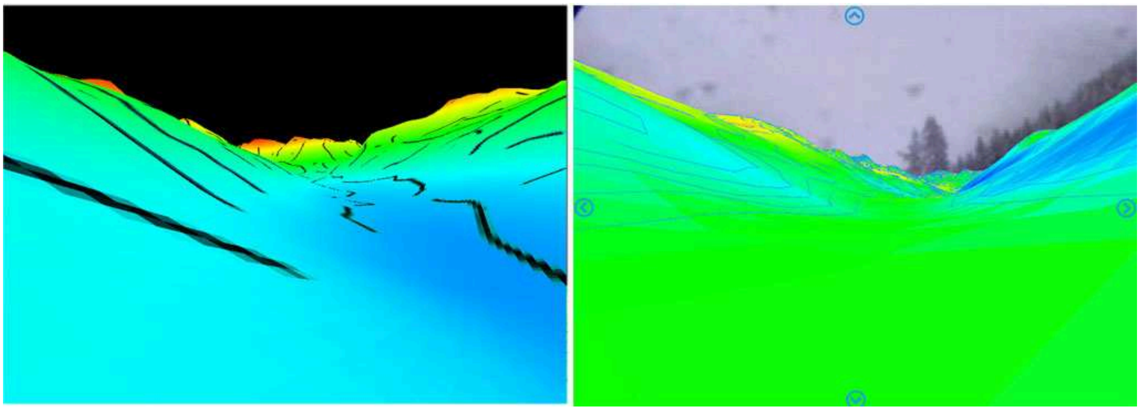


**Figure 5.7:** Lines representing structure of elevation models. Left: ridges and valleys rendered first to texture used by the shape rendering pass. Right: Isophotes rendered directly as an additional pass after the 3D elevation model. Note the artifacts in nearby lines in the left figure, due to texture interpolation.

can be supported by adding supplemental line renderings [Girshick et al., 2000]. Girshick et al. suggest using principal directions to convey shape in 3D line drawings, we investigated several methods to abstract the DTM in a way that is representative of the topology, but less obstructing. Two candidates were tested. The first technique computed the locations of high curvature changes, based on positive (valleys) and negative (ridges) derivatives. This technique of ridges and valleys represents a number of descriptive features of the geometry that are unobtrusive to the viewer (see Figure 5.7 Left). The second technique is called isophotes and it reflects the locations on the image that have similar reflectance values, calculated from the top of the DTM (simulating the position of the sun). Isophotes are more numerous than ridges and valleys (see Figure 5.7 Right).

In an attempt to try and reduce load in the rendering, we created textures out of the line models during transcoding. By plotting the lines in the texture, an expensive line rendering pass was skipped, which we hypothesized would increase performance. The end result did not provide any speed-up on the target platform, but it did introduce undesirable visual artifacts around the lines in comparison to the direct line rendering (see Figure 5.7).

To extract isophotes from the DTM, during transcoding, we apply the algorithm proposed by [Burns et al., 2005].

> Our input data $\phi(i,j,k)$ is a 3D matrix of real-numbered data values. We begin by rendering contours on isosurfaces of $\phi$. An isosurface F can be defined as the zero-set of the function
>
> $$f(i,j,k) = \phi(i,j,k) - \tau, \tag{5.1}$$
>
> where $\tau$ is a threshold within the range of the data values. This equation

> $f = 0$ implicitly defines a 2D surface in 3D space. Contours on a continuous surface are those locations where the surface normal is perpendicular to the view. They are locations at which $n \cdot v = 0$, where $n$ is the surface normal and $v$ is the view vector. In our application, we could use this definition directly by first extracting an isosurface at some threshold $\tau$ , computing the surface normal as $n = -\nabla\phi$, and extracting the contour. Alternatively, we can consider the set of contours on all possible isosurfaces of $\phi$, which is itself a 2D surface defined as the zero set $C$ of the function
>
> $$c(i, j, k) = -\nabla\phi(i, j, k) \cdot v(i, j, k). \tag{5.2}$$
>
> Finding contours at a specific threshold $t$ then reduces to finding the intersection of the 2D surfaces F and C. Generically, this intersection takes the form of a set of 1D loops in 3D space. To extract a contour, we first locate cubes containing zeros of both the $f$ and $c$ implicit functions. The algorithm finds intersections of the two implicit functions on the faces of the cube, using linear interpolation based on the values at the eight corners. Then it finds any intersections between these two sets of lines on each face. The result is a set of points on the faces of the cube, which when connected yield segments of the contour. We stylize the contours based on properties such as their length and visibility. We detect these “interior” contours by defining
>
> $$s(i, j, k) = \nabla\left(\hat{n}(i, j, k) \cdot v(i, j, k)\right) \cdot v(i, j, k), \tag{5.3}$$
>
> where $\hat{n}$ is the unit-length normal. As with the normal, the gradient is computed numerically, on demand. Since $s$ has the same sign as radial curvature, interior contours are found by testing $s < 0$ once a segment has been extracted.

The terrain can be rendered in line mode, with the overlay textured onto the contour lines; or in fill mode, where both lines and polygons are rendered. The line mode displays colors overlaid on the lines, while the video background is clearly visible. Isolines provide additional depth cues, as they become closer together with increasing distance (see Figure 5.8). Depth ordering is controlled by the polygonal 3D model, it must be rendered even when invisible to compute occlusions in the graphics pipeline.

## 5.3 Implementation

The first prototype of our AR system was integrated in a fully featured scene graph engine driven by Studierstube [Schmalstieg et al., 2002]. It included nodes for the representation

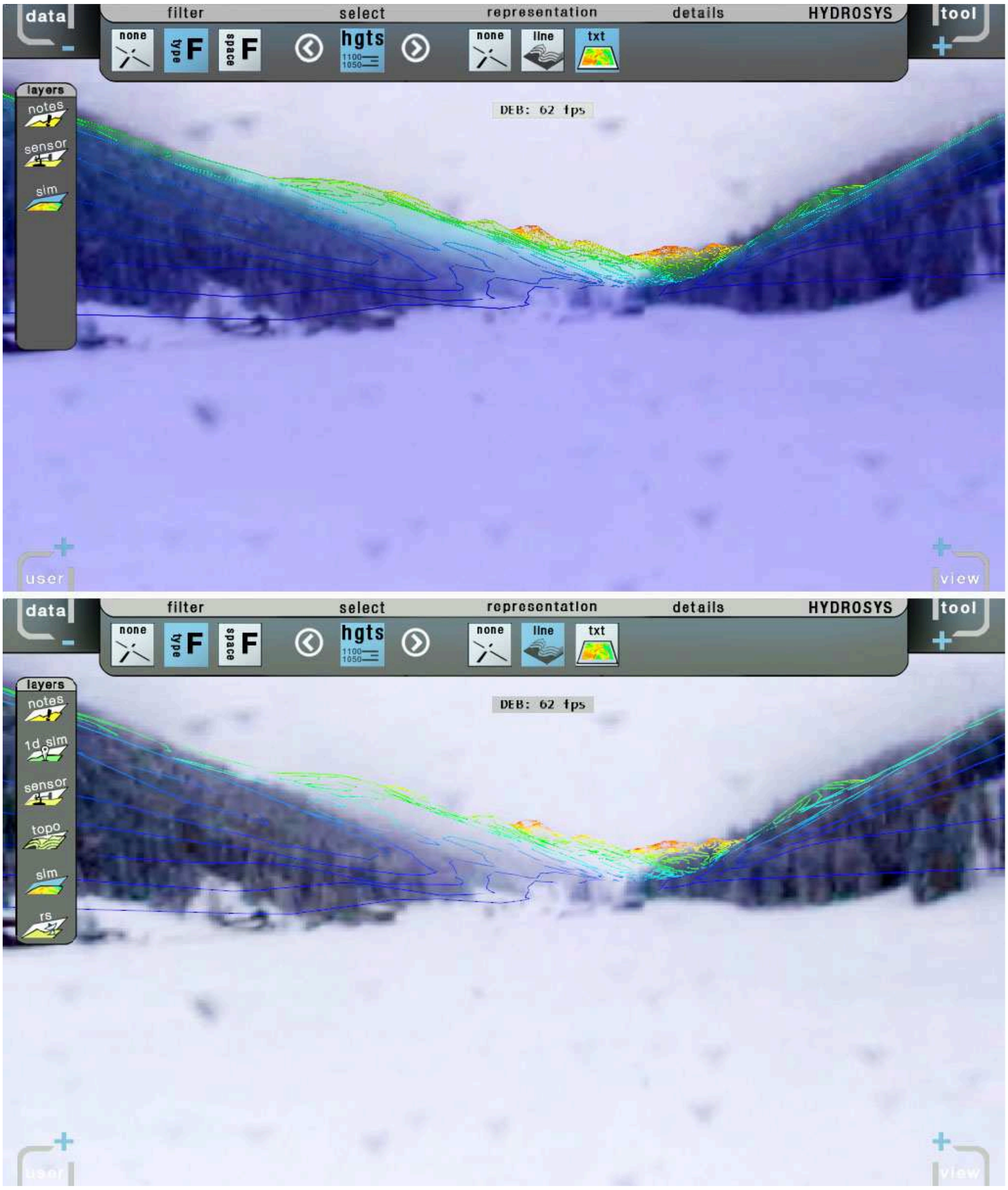


**Figure 5.8:** Terrain overlay visualization. Top: terrain in fill mode with transparency. Note how the transparency shows the real world underneath while preserving the color coding of the overlay, and isolines provide a sense of depth. Bottom: overlay mapped on isolines. The lines give a general idea of the color distribution while clearly showing the video.

of pose sensors, and their connection with graphical data. A Flash extension to Studierstube drove both the user interface and the rendering of labels. In spite of the flexibility of such a system in terms of available components for visualization, it had a severe performance impact in our platform. With ten labels the framerate dropped to a crawl, below interactive rates, mainly due the poor graphics performance of the device and the high demands of the graphics engine. Similarly, even after removing detail from the DTM, rendering it without textures almost exhausted the possibilities of our poor device.

This forced a redevelopment of a basic rendering engine, that relied mostly on the vertex buffer objects extension (VBOs). We created implementations of specific renderers to handle geometry (e.g., elevation models), and the user interface (e.g., tiles, text). The vertex buffer objects (VBO) organizes and reduces the number of communications between CPU and GPU. The geometry (vertices, indices) and texture coordinates for the 3D model are transferred once to the GPU, and thereafter used for rendering continuously. As the rendering is based on shader programs, each renderer implements a collection of vertex and fragment shaders to extend its graphical possibilities. Stress tests on the new implementation, with up to 500 randomly placed labels still maintained interactive rates (20 fps), although they were unreadable due to clutter. Similarly, the new implementation allowed to render the DTM textured multiple times, to achieve the effects described in section 5.2.1.

To extend the control over rendering, in particular the blending between all these representations, we built a multi-pass renderer. Multi-pass rendering allows the rendering of disparate parts to textures or render buffers which are blended together in a final compositing step. In the case of our mobile platform, we had to rely on expensive render to texture modes, since framebuffers were not available. The parts of the application are separated in layers. In the minimal case, the 3D model, the labels and the user interface represent three different layers. Figure 5.9 summarizes components that contribute to each layer. Each of these layers is rendered first to a texture. Finally the textures are rendered in order to screen. This separation gives more control over the blending between layers, and enables the application of image processing techniques to the layers in the final rendering pass. For example, one can apply a “sharpen” filter to the video background layer, to enhance edges of a blurry video. The multi-pass rendering proved to be the extreme case for our engine in the platform of use, as it really taxes the device, bringing the framerate to 10 fps (still faster than the first prototype). Thus, its use is optional, the user can decide to use normal rendering (no layers) at full speed, and deploy perceptual enhancements when required.

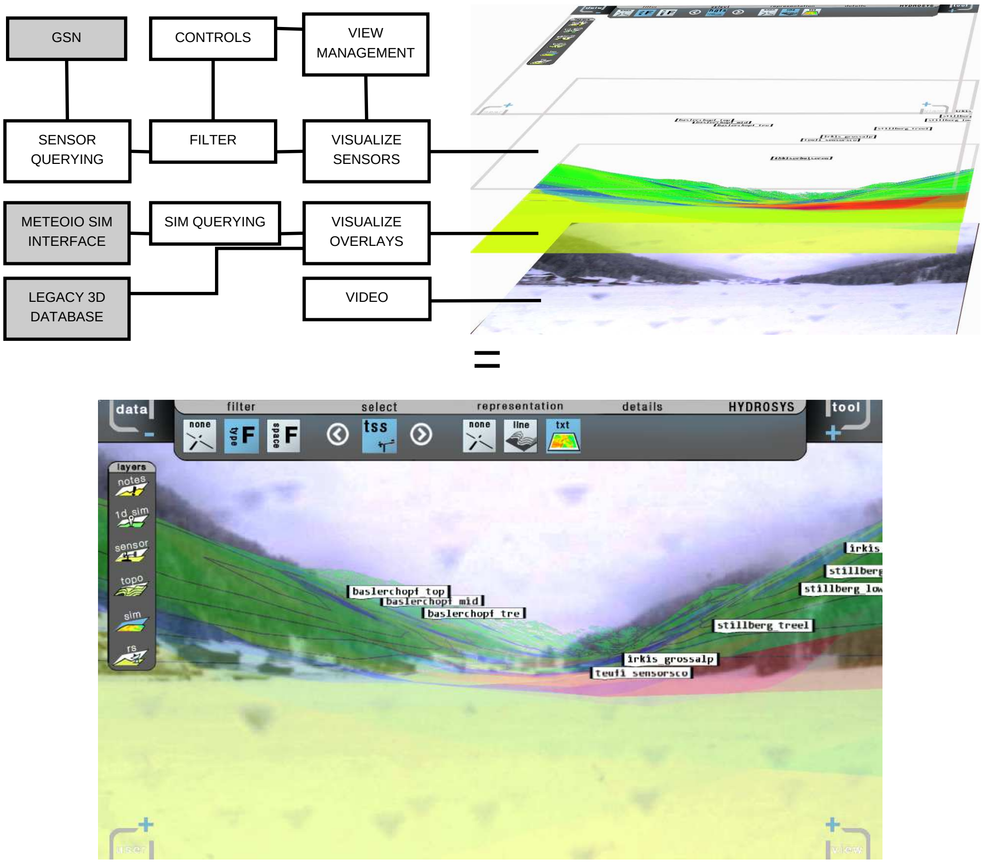


**Figure 5.9:** System components contributing to different layers. After rendering, layers are blended together in a compositing step.

## 5.4 Summary

This chapter has taken a broad look at scientific data visualization using outdoor AR. In the first part, we focused on representing measurements from multi-sensor arrays. These representation aimed at conveying the value within its spatial context of virtual objects (e.g., other measurements) and physical entities in the real world, in a readable, organized manner. Thus, the minimal representation that retains legibility, and avoids clutter. Thereafter, we develped a seamless integration of environmental data in outdoor AR visualizations. The main accomplishment was in achieving coherent visualizations for data in AR, concentrating on allowing the user to understand the effects of the process in light of its real-world context. The next chapter looks at the topic of visually segregating data in a subtle maner. It introduces a novel approach to represent data and to guide interpretation when certain knowledge about the data can be assumed.

# Chapter 6

# Subtle Visualizations Based on Saliency Modulation

In this chapter we deviate from the environmental visualization scenario to take an abstract view on representation and the question of how to subtly guide the user to selected portions of the visual input. Augmented reality (AR) applications intended to call attention to real objects often do so by overlaying on the real world highlighting effects or virtual objects such as arrows. At times, it would be desirable that these effects were more subtle, in part to avoid exacerbating perceptual issues, but mostly in cases where the objects highlighted are secondary to the user's task. In many cases, the application needs to appeal to post-perceptual processes, to tell the user that a particular object is somehow related to their current task, but without alerting or interrupting the user's workflow. For example, when visualizing simulation results overlaid on a mountain landscape, the application wants to highlight the sensors that contributed data to the simulation.

The literature on psychology and vision identifies saliency as a model of attention [Itti, 2007]. Moreover, attention influences memory at different stages of processing [Awh et al., 2006]. Thus, we assume that by manipulating the saliency of a region in the visual input, we can potentially influence attention and memory. We apply a saliency modulation technique (SMT) to modify videos so that a region of our selection contains the highest saliency. The technique was first described by Mendez et al. [Mendez et al., 2010], and later evaluated by Veas et al. [Veas et al., 2011] in terms of its effectiveness to 1) direct attention to a selected region of the visual input and 2) influence the recall rate of certain objects, 3) without the user becoming aware of any modifications. The SMT enables an AR approach known as ”mediated reality,” in which existing features of an image are modified, instead of adding discrete new objects. This section describes three studies, measuring modulation awareness, attention, and memory. The modulation awareness study finds the largest amount of modulation we can apply that is imperceptible to the viewer. The attention study evaluates whether this modulation threshold shifts attention towards selected regions of videos. The memory study evaluates whether it increases recall for selected objects. Our results indicate that regions modulated with the SMT will draw a first fixation faster than without modulation. Moreover, modulation can increase recall for selected objects. In summary, the SMT can significantly shift attention to selected areas and influence memory of selected objects from a video in a way that is imperceptible to the viewer.

## 6.1 Saliency Modulation Technique

The SMT developed by Mendez [Mendez, 2011] is capable of manipulating the saliency of a video at interactive framerates [Mendez et al., 2010]. For each frame, it computes a saliency measure on every fragment according to a hierarchical multi-channel contrast measure [Itti et al., 1998]. It then modifies the image, changing contrast in lightness and color to have the highest attention salience inside a designated focus region. Changes are applied so that spatial and temporal coherence are maintained. In detail, the SMT works by analyzing and modulating conspicuities in three dimensions: lightness ($L$), red-green color opponents ($O_r$), and blue-yellow color opponents ($O_b$). Each frame is first converted to $CIEL * a * b$ space, thereby obtaining the values for each dimension $k \in \{L, O_r, O_b\}$. A pyramid of images is created with $p$ levels. Modulation progresses from coarse levels to fine levels of the image pyramid. This allows changes affecting a large region to occur early in the process, while later steps progressively refine the result, introducing less noticeable artifacts. For each level, *analysis* and *modulation* steps are carried out iteratively in each dimension $k$.

**Saliency analysis.** During this step, the conspicuities of the image are computed to measure the naturally salient objects in the scene. A conspicuity is given as a signed sum of the center-surround differences at multiple scales of an image pyramid. The conspicuity $c_k$ is the defined as:

$$c_k = \frac{\sum_{n=0}^{n=2} \sum_{m=n+3}^{m=n+4} k_n - k_{n+m)}}{p},$$

where $p = 6$, and $k_i$ is the conspicuity $k \in \{L, O_r, O_b\}$ at mipmap level $i$. The conspicuity $c_k$ is normalized using the global conspicuity maxima [Lee et al., 2007]. The normalized conspicuity $\hat{c}_k$ is:

$$\hat{c}_k = \frac{c_k}{max(c_k)},$$

where $k \in \{L, O_r, O_b\}$.

**Saliency modulation.** Given a dimension $k \in \{L, O_r, O_b\}$, let $\hat{c}_k$ be the normalized conspicuity of a location and $t_k$ be the threshold of the conspicuity, a floating point number that governs the amount of modulation applied to the location. A modulation adjustment $m_k$ is calculated for this location as,

$$m_k = \begin{cases} 0 & \hat{c}_k < t_k \\ \hat{c}_k - t_k & \text{otherwise} \end{cases}.$$

For a feature value $f_k$ of a location, the modulated value $f'_k$ is calculated by applying the modulation $m_k$ in order to increase the conspicuity of the focus, and correspondingly decrease that of the context. Thus,

$$f'_k = \begin{cases} f_k + m_k & \text{if the location is marked as focus} \\ f_k - m_k & \text{otherwise} \end{cases}$$

Modulation is performed in the order of sensitivity of the human visual system [20]: first, lightness is modulated, then red-green opponents, then blue-yellow opponents. Note that other contributors to saliency remain unaffected (e.g., motion, size, and orientation). Finally, the image is converted from $CIEL*a*b$ to RGB. For a detailed implementation refer to [Mendez, 2011]. In this section we concentrate on finding thresholds that make modulation imperceptible, and validating that the modulation of a video successfully influences memory.

## 6.2 Methodology

To prepare the stimuli for the awareness and attention studies, we recorded $\sim$10h of video under various situations (indoors, outdoors, night, day, with moving objects, free moving camera, panning camera). The idea was to have a manageable variety of videos that represented day-to-day situations. From these, we extracted clips, each lasting 10s, with the restriction that no human body parts appear in the clips because they represent a high attention sink. Videos were recorded at a resolution of $1280 \times 720$ at 29.97fps and presented without resizing and uncompressed to avoid interpolation artifacts by the graphics card. For each experiment, we recruited a balanced number of participants from the university population and the general public. All participants had normal or corrected to normal vision, and were screened for color-sensitivity deficiencies by an on-screen Ishihara test. We used an SMI desktop-mounted eye tracker, operating at 60 Hz. Stimuli were presented on a $19''$ monitor at 70cm from the participant. A chin rest was used to limit head movements. All studies were performed in an empty office with lights off, and windows and doors closed, to minimize attention distracters.

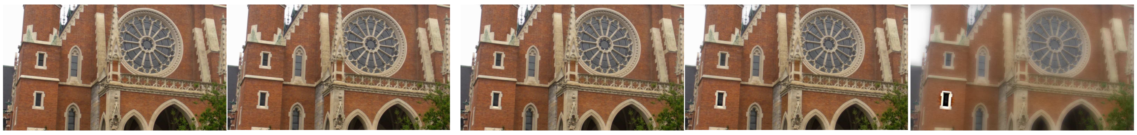


**Figure 6.1:** Modulation thresholds. Frame from a video modulated to emphasize a window using different modulation thresholds. From left to right, modulation thresholds zero (no modulation), three, four, five, eight (full modulation).

Focus regions (FR) for the awareness and attention studies were chosen by analyzing the videos and selecting low salience regions. The selection methodology is presented in the next section. Each clip contained one or more FR. Each FR was visible for at least 2s.

## 6.3 Exploratory Study: Modulation Awareness

In the SMT presented above, the amount of modulation is governed by a threshold ($t_k$) for each modulation dimension. Thus, the SMT can be configured to produce different modulation thresholds (see Figure 6.1). Our initial concern was how to apply the SMT so that the viewer is unaware of the manipulation. In other words, we were seeking the maximum modulation that is imperceptible to the viewer. To investigate viewer attitude towards modulation, we conducted a series of studies on modulation awareness.

A threshold is a floating point value in the $[0 \ldots 1]$ range. To reduce the search space, we discretized this range into a set of seven samples. Additionally, we used the extreme values 0 (no modulation) and 1 (full modulation), for a total of nine thresholds. We performed three studies to investigate the appropriate modulation threshold. A challenge in these studies is that participants need to evaluate different modulation thresholds for videos by actively checking for visual manipulations in them, a goal-based task. This type of task is known to modify the gaze path of participants and suppress stimulus-based attention. Thus, analysis of attention cannot be performed at the same time as the study on awareness of modulation.

### 6.3.1 First Pilot Study

Our intention for the first pilot was to identify and discard thresholds for which modulation is clearly perceivable, thus reducing the search space for a subsequent formal study. The stimulus was a series of 18 clips, two for each of the nine modulation thresholds, lasting $\sim$10s each.

Three people (all male, ages 28, 33, and 35) participated in this study. Participants were requested to look at the videos and verbally rate each of them on a 7-point Likert scale for naturalness (where 1=natural corresponded to the video is as it came from the camera, and 7=unnatural meant the video has been manipulated to such an extent that it feels unnatural). The videos were shown in randomized order, in two sets of nine with a short break in-between. It is important to note that participants had to judge each video in isolation and the videos for each modulation threshold were different. Therefore, participants were not given the chance to compare a modulated video with the original version.

**Analysis and Results** We did not perform statistical analysis in this set due to its small sample size. We confirmed, however, that the higher the modulation threshold; the higher the score given by the participants (see Figure 6.2 Left). Thresholds 0-5 scored below the middle of the scale (*somewhat unnatural*). In fact, thresholds 0 and 4 had an average score of 3.

### 6.3.2 Second Pilot Study

In the first pilot study, participants judged each modulation threshold in isolation. This raised the doubt of whether they would detect a difference if they were given the chance to see both modulated and unmodulated versions of the same stimulus. The goal of this pilot was to verify whether participants could notice a difference between modulated and unmodulated images. We randomly selected screenshots from the stimulus videos. These were presented in pairs with a change-blindness break in between, following the setup suggested by Rensink et al. [Rensink et al., 2000]. For each pair, the images were presented in the order FBFBSBSB, where F corresponds to first image shown for 240ms, B to blank image shown for 320ms, and S to second image shown for 240ms. There were nine change-blindness sets, one for each threshold considered. We modulated two images for each of the nine modulation thresholds, totaling 18 image pairs.

Three participants took part in this study (2 male, 1 female, ages 28, 29, and 24). They were instructed to observe the images and state whether or not the images were different. Each participant saw each of the 18 image pairs once. The presentation of the image pairs was randomized. As suggested by Rensink et al., each change-blindness pair was presented for 60s. Participants, however, had the possibility to interrupt the sequence by stating a judgment.

**Analysis and Results** We did not perform statistical analysis for this set due to its small sample size. Figure 6.2 (Right) shows responses for this study as stacked bars. We interpreted each affirmative response as a value of 1, and each negative response as a value of 0. As shown in the figure, the pair zero-zero was always correctly judged as being

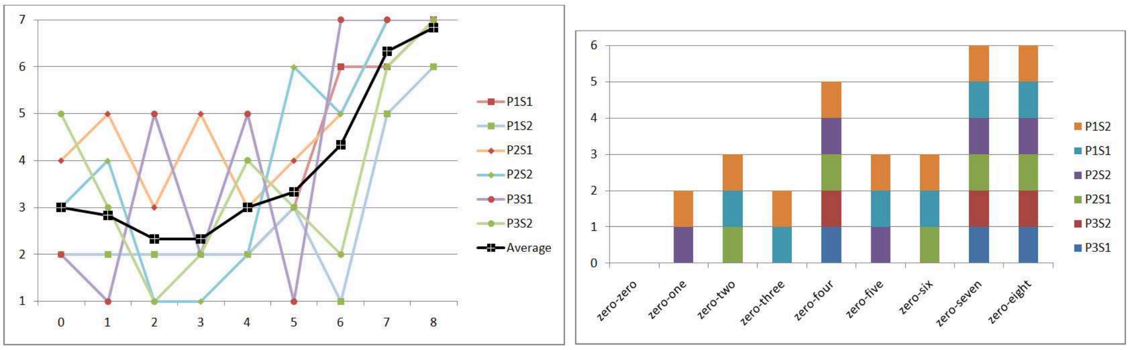


**Figure 6.2:** Responses of pilot studies on perceived modulation. (Left) Responses corresponding to the first pilot study. (Right) Responses corresponding to the second pilot study. Notation PmSn means Participant m, set n. Notation zero-number means a pair with the same image unmodulated and modulated at threshold number.

unmodulated (it never received an affirmative response). Pairs zero-seven and zero-eight were also always correctly judged as being modulated (6 affirmative responses each). Intriguingly, pair zero-four was graded higher than zero-five and zero-six.

### 6.3.3 Formal Awareness Study

Building on the pilot studies, we conducted a formal study to further evaluate the reaction of people to modulation thresholds 3-5. Our aim was to verify that the threshold used in subsequent experiments was imperceptible.

**Method** The stimuli for the awareness study were obtained using the same 20 clips used in the attention experiment presented below. Therefore, three candidate thresholds (3, 4, and 5) plus the control (no modulation) times the 20 stimulus videos resulted in 80 video-threshold pairs. We arranged the videos so that each video-threshold pair was seen by four participants. Each participant watched each video with a randomized modulation threshold. No participant watched the same video twice with different modulation thresholds.

We recruited 16 participants for this study (12 male, 4 female, 18-35 years old, $\overline{x} = 27.8$), none of whom participated in subsequent experiments. The procedure and instructions were the same as those described for the first pilot.

**Analysis and Results** To analyze results, we considered the four modulation thresholds (0, 3, 4, and 5) as related samples. We then conducted three Wilcoxon signed tests for two related samples, to determine whether participants noticed significant damage to the

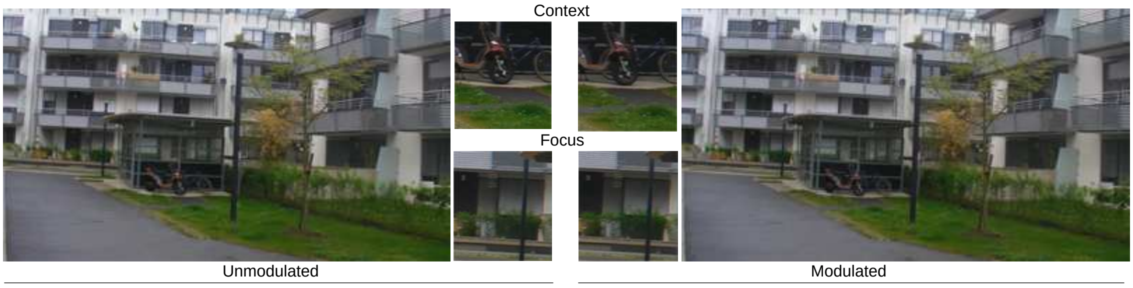


**Figure 6.3:** Comparison between modulated and unmodulated video. (Left) Frame and details from an unmodulated video. (Right) Same frame and details after modulation. Differences may be seen when compared side by side, but evidence of modification is difficult to see when viewing the modulated version in isolation.

videos compared to the ground truth. Our pair samples were zero-three, zero-four, and zero-five. We applied a Bonferroni correction to account for the number of pair samples and keep the $\alpha$ level below .05. The analysis showed no significant difference for any of the pairs. Thus, there was no evidence that the general population would be able to distinguish which videos had been modulated and which had not. However, we decided to take a somewhat conservative approach and use threshold four for our modulation procedure. Figure 6.3 illustrates the results of modulation. The left image was obtained from the unmodulated video (first condition). The right image was obtained from the video modulated with threshold 4 (second condition). When comparing both images side-by-side, changes are barely perceptible. If, however, one is allowed to see only the modulated video in isolation, the changes become imperceptible. The insets in Figure 6.3 show a detailed comparison of the changes. Observe that the focus after modulation has slightly more vivid colors and more contrast, while the context has slightly duller colors and less contrast.

## 6.4 Attention Experiment

The goal of this study was to verify, through use of an eye tracker, whether the SMT can direct the visual attention of participants to selected regions. Here, we assume that visual attention can be characterized by eye gaze. As stimuli for this experiment, we selected 20 clips lasting roughly 10s each.

### Hypotheses

**H1:** The time before the first fixation on the FRs will be smaller for the videos modulated with our procedure than for the original unmodulated videos.

**H2:** The fixation time in the FRs (i.e., sum of durations of all fixations on the focus) will be higher for the videos modulated with our procedure than for the original unmodulated videos.

**H3:** The percentage of participants that have at least one fixation on the FR will be higher for the videos modulated with our procedure than for the unmodulated videos.

### 6.4.1 Method

Since we wanted to compare eye-gaze for regions between unmodulated and modulated versions of a video, we used a between-subjects, repeated measures design with independent variable modulation (unmodulated, modulated), and dependent variables time before first fixation, fixation time (i.e., sum of durations for all fixations on the focus), and percentage of participants with at least one fixation.

We recruited 40 participants to take part in this experiment. They were divided into two conditions for the between-subjects setup (20 in the unmodulated condition, 20 in the modulated condition): (20 regions × 20 participants) = 400 trials per condition = 800 trials total.

Twenty participants (14 male, 6 female, 24-52 years old, $\overline{x} = 31.4$) took part in the first (unmodulated) condition. Each participant was provided with the following instructions:

> You will sit in front of a computer screen. We will display a series of short video clips. All you have to do is look at the clips. That's it! This test is divided into two parts so you can have a break in between. Your eye gaze will be tracked with a non-wearable system. It will be using an infrared camera and light placed in front of you. Infrared light is invisible to the eye and poses no harm to you.

Care was taken not to mention the number of video clips in order to avoid counting (which would trigger a top-down task). It was emphasized that there was no task and that all that was required was to watch the clips. The eye tracker was calibrated for each participant before the stimuli were presented. Each participant watched each of the 20 unmodulated videos once, in random order. Between videos, a blank slide was shown for 2000ms.

By analyzing eye-gaze data from the first condition, we determined a visually unattended region for each clip in the unmodulated stimuli. We define unattended regions as those that have fewer than five fixations by less than twenty percent of the participants. These unattended regions were then designated as the FRs of the study. To increase the saliency of FRs for the second condition, the clips were modulated with the SMT at threshold 4, as suggested by the awareness study. The clips derived through this process were used as stimuli for the second condition.

Twenty participants (14 male, 6 female, 25-42 years old, $\overline{x} = 32.1$) took part in the second (modulated) condition. They went through the same procedure as in the first condition, the only difference being that the stimuli were modulated.

### 6.4.2 Analysis

Analysis was performed with independent samples *t*-tests whenever our data satisfied the condition of normality and with Mann-Whitney $U$ tests otherwise. In both cases, tests were one-tailed. Two-tailed tests would be able to indicate whether there is a significant difference between both conditions, but not whether this difference is in the intended direction (increasing attention). A Shapiro-Wilk test indicated that the data for H1 and H3 satisfied normality. However, the data for H2 did not. We adjusted $\alpha$ levels using a Bonferroni correction to ensure a level of .05.

**Results for H1.** The results of the one-tailed *t*-tests indicate that the mean values of the second condition (modulated) were significantly smaller than the mean values of the first condition (unmodulated), t (35) = 2.916, $p < .01$. Therefore, the mean duration before the first fixation on the FRs for participants in the second condition (M = 9231.58, SD = 468.16) was significantly smaller than that of the participants in the first condition (M = 9638.86, SD = 363.10).

**Results for H2.** There is no significant difference in total fixation time between the unmodulated (M=26.79) and the modulated (M=43.31) conditions (Mann-Whitney $U$=99.0, n1=17, n2=20, p=.3, one-tailed). Despite the lack of normality of the data for this hypothesis, we performed a *t*-test confirming the results t (35) = -2.117, p = .02. Therefore, the mean total fixation time for participants in the second condition (M = 43.31, SD = 24.32) was not significantly different from that of the participants in the first condition (M = 26.79, SD = 22.82).

**Results for H3.** The results of the one-tailed *t*-tests indicate that there is no significant difference in the number of participants that had at least one fixation between the unmodulated and the modulated conditions, t (35) = -2.028, p = .05. Therefore, the mean number of participants with at least one fixation in the second condition (M = .13, SD = .08) was not significantly higher than that of the participants in the first condition (M = .08, SD = .06).

As can be seen, H1 proved statistically significant; however, we were unable to find significant differences for H2 and H3. We further examined the gaze data of our participants to try to find consistent failures in our modulation procedure. By visually analyzing heat maps of our videos in the second condition, we found what seemed to be a consistent pattern where our modulation procedure failed: whenever the camera panned directly away from the FR, the technique seemed to be unable to attract fixations. This did not happen on videos where the camera was static, or whenever the panning was not directly away from the FR. Subsequently, we filtered out the information from FRs that fit this criterion (5 regions out of 29 were excluded). Then we repeated the analysis. Once again, we performed a Shapiro-Wilk test to verify the normality of our filtered data. The results indicated that the filtered data for H1 and H3 satisfied normality, but the filtered data for H2 did not.

On the filtered data for H1, the results of the one-tailed *t*-tests indicate that the mean values of the second condition (modulated) were significantly smaller than the mean values of the first condition (unmodulated), t (35) = 3.386, $p < .01$. Hence, the mean duration before the first fixation on the FRs for participants in the second condition (M = 9126.98, SD = 499.44) was significantly smaller than that of the participants in the first condition (M = 9616.66, SD = 352.59). On the filtered data of H2, results showed a significant difference in the total fixation time between unmodulated (Mean=14.03) and modulated (Mean=23.23) conditions (Mann-Whitney $U$=85.5, n1=17, n2=20, $p < .01$, one-tailed). Despite the lack of normality of the data for this hypothesis, we performed a *t*-test, which confirmed the significant difference in total fixation time between conditions t (35) = -2.659, $p < .01$. Consequently, the mean total fixation time for participants in the second condition (M = 49.52, SD = 26.04) was significantly higher than that of the participants in the first condition (M = 27.65, SD = 23.55). On filtered data for H3, the results of the one-tailed *t*-tests indicate that the mean values of the second condition were significantly higher than the mean values of the first condition, t (35) = -2.478, $p < .01$. Consequently, the mean number of participants with at least one fixation in the second condition (M = .15, SD = .09) was significantly higher than that in the first condition (M = .08, SD = .06).

### 6.4.3 Discussion of Attention Experiment

As can be seen from the analysis, we could always draw the eye gaze of participants significantly sooner with our modulation technique. However, once we filtered out situations in which the camera panned directly away from the FR, analysis revealed additional effects of the SMT in the modulated condition. On filtered data, the average duration before the first fixation on the FRs was significantly shorter. We could also retain the visual attention of participants for a significantly longer time. And finally, the number of participants

**Figure 6.4:** Heatmaps of the user studies. (Left) Unmodulated condition. (Right) Modulated condition. This is a handpicked example chosen to illustrate the effect of the SMT.

with at least one fixation on the FRs of the modulated videos was significantly larger than for the unmodulated videos. It is difficult to illustrate accumulated fixations on a region in a video, since they spread out throughout the duration of the clip. Nevertheless, Figure 6.4 illustrates one frame of one video in both conditions, showing eye fixations accumulated over multiple frames, in which the effects of the SMT are clear. The image on the left comes from the unmodulated video and the image on the right from the modulated video. A white outline denotes the position of the FR. In the general case, however, the effect is not so apparent throughout the entire duration of the video.

The technique was not always effective for each of the video clips, nor for each of the participants in the tests. In the general case, the SMT will draw a first fixation faster than without modulation. In cases where the camera is not moving away from the focus regions, the number of participants that had at least one fixation in the focus region also increased, and the fixation time was significantly higher. Thus, we can state that attention direction with SMT was successful.

## 6.5 Memory Experiment

The goal of this experiment was to assess whether the SMT increases recall of selected objects in the video without suppressing recall for others. With the aim of comparing recall for regions between unmodulated and modulated videos, we used a between-subjects, repeated measures design with independent variable modulation (unmodulated, modulated), and dependent variable recall hits.

In order to prepare the stimuli, we recorded $\sim$2h video in a furniture store and extracted two clips (identified as video A and video B) lasting 1m each. These clips include people walking by, but no faces. The choice of location ensured the appearance of many different objects in the videos.

**Hypotheses**

H4: There is no significant difference in recall hits between the first condition (unmodulated) and the second condition (modulated) for recalled objects.

H5: There is a significant difference in recall hits between the first condition (unmodulated) and the second condition (modulated) for non-recalled objects.

Hypothesis H4 concerns losses caused by the technique in the normal condition in terms of suppressing recall of normally recalled regions. H5 concerns gains due to the technique in terms of increasing recall of selected regions.

### 6.5.1 Prerequisites: Recall Study

The memory experiment requires a set of regions that appear in each video from which participants would select those they remember. These regions are associated with objects and are regarded as objects for the rest of the discussion. We expected to be able to determine the set of objects by examining the videos using Itti's model for saliency, but the results were mostly coarse, and would not help identify individual objects. We then decided to use a mixed approach in which we preselected some regions based on visual inspection and validated them by means of a pilot study. Thus, we visually examined the videos and selected scenes containing both low and high salience objects. Factors for scene selection included being clearly visible for an acceptable amount of time (about 2s), and that the objects in it be clearly distinguishable. We selected 18 scenes in total and, for each we extracted one object with high saliency and one with low saliency. Pictures of these objects were printed on 36 cards, each 11cm × 10cm.

To refine the set of objects, we carried out an exploratory study with six participants (5 male, 1 female, ages 25-35), who did not participate in any subsequent test. The procedure and apparatus were the same as those for the formal memory experiment. Based on eye-gaze analysis and on recall hits, seven scenes were removed, and three objects were changed in the remaining scenes resulting in deck A with 10 cards from video A, and deck B with 12 cards from video B. We classified five objects from deck A and seven from deck B as highly salient. The remaining objects were classified as having low saliency. This classification served as a control, as the experiment assumes that objects with high saliency will have high recall hits. We added five and six distracter objects to decks A and B, respectively, to assess whether a participant was picking cards randomly.

### 6.5.2 Method

The same 40 individuals that participated in the attention experiment took part in the memory experiment. Participants were divided into two conditions for the between-subjects setup (20 unmodulated, 20 modulated): (22 objects $\times$ 20 participants) = 440 trials per cond. = 880 trials total. For each condition, videos A and B were shown in interleaved order; so that 10 participants experienced video A first and 10 participants experienced video B first. Before starting the experiment, participants were instructed to:

> Observe the video and try to memorize the objects that you see. At the end you'll be presented with a deck of cards picturing objects printed from the video and you'll be asked to select those that you remember. Be careful, the deck of cards also contains objects that did not appear in the video.

Participants experienced the first video, and were subsequently presented with the corresponding deck of cards from which they could pick those objects that they remembered. A recall hit was recorded for an object if it was selected by a participant. After a short break, the same procedure was applied for the second video. For each video, participants answered a questionnaire in a 7-point Likert-scale format to assess the difficulty of the task. After finishing the procedure for the two videos, they answered general questions about the naturalness of the videos.

Based on analysis of the first condition, we classified objects as high recall ($HR$, recall higher than $60\%$) or low recall ($LR$, recall lower than $40\%$). The $40\%$ and $60\%$ thresholds were arbitrarily selected based on results of the first condition. Visual inspection of recall hits for this condition showed a gap in results: no object scored between $40\%$ and $60\%$. Decks A and B had four $HR$ objects each, totaling eight $HR$ objects, all of which had been classified as highly salient. Three objects that had been classified as highly salient had low recall in the first condition, whereas objects classified as having low saliency all had low recall hits. In preparation for the second condition, videos A and B were modulated using the SMT to increase the salience of objects in $LR$. For the second condition, the only difference was the modulated stimulus; the procedure was the same.

### 6.5.3 Analysis

A Shapiro-Wilk test proved that the data for recall did not satisfy the condition of normality; the data are binary and not interval-scaled. Analysis was performed with Mann-Whitney $U$ tests, due to their robustness under these conditions. Since our hypotheses focus on one side of the distribution, all the tests are one-tailed. We adjusted $\alpha$ levels with a Bonferroni correction to ensure a level of .05.

**Results for H4.** For objects $o \in HR$, mean recall hits in unmodulated (M=.69) and modulated (M=.68) conditions show no statistical difference (Mann-Whitney $U$=$1.1272e^4$, n1=n2=160, p=.46, one-tailed). The result supports H4.

**Table 6.1:** Recall differences for objects in $LR'$.

| Object | o1 | o2 | o3 | o4 | o5 | o6 | o7 |
|---|---|---|---|---|---|---|---|
| Unmodulated | 2 | 2 | 3 | 2 | 8 | 1 | 8 |
| Modulated | 6 | 6 | 5 | 3 | 13 | 2 | 11 |

**Results for H5.** For objects $o \in LR$, the mean recall hits for unmodulated (M=.19) and modulated (M=.22) conditions show no statistical difference (Mann-Whitney *U*=$3.78e^4$, n1=n2=280, p=.15, one-tailed). There is not enough evidence to support H5.

To further analyze these results, we classified $LR$ objects into those that increased in recall hits in the second condition, and those that did not show any change or showed a decrease in recall. We first confirmed the relationship between recall and attention, correlating recall with fixation count r(878) =.35, $p < .001$; and with fixation time r(878) =.666, $p < .001$. Then, we analyzed features of these objects that contribute to saliency and how they affect recall. We found moderate correlations between recall and size (in pixels) r(878) =.449, p=.032, and the average size in time of the region ($\%$coverage × $\%$visible time) r(878) =.428, p=0.042. We observed that objects $o \in LR$ that decreased in recall in the second condition were either $< 2e^4px$ or appeared for less than 2s. Since the SMT cannot control contributions to saliency due to size or spatial frequency, we filtered data in LR based on these criteria, yielding two datasets: $LR' = o \in LR, size(o) > 2e^4px$ and $MR = LR - LR'$. Subsequently, we analyzed the filtered data.

On filtered data for H5, objects $o \in LR'$, the mean recall for unmodulated (M=0.19) and modulated (M=0.31) conditions differed significantly (Mann-Whitney *U*=$1.128e^4$, n1=n2=160, p=.008, one-tailed). The result supports H5, meaning that objects that had low recall hits in the first condition significantly increased in score when modulated with the SMT (see Table 6.1). Furthermore, objects $o \in MR$ did not suffer a significant reduction in recall from unmodulated (M=.18) to modulated (M=.11) (Mann-Whitney *U*=$6.660e^3$, n1=n2=80, p=.05, one-tailed). The results support H4 in the general case. This means that the SMT does not suppress recall for objects with otherwise high recall. The results did not provide enough evidence to support H5 in the general case. Nevertheless, for objects that cover more than $2e^4px$ and come into view for over 2s, the results showed a significant increase in recall. This suggests that the SMT increases recall of regions $> 2e^4px$ with durations $> 2s$, without a significant loss to other regions.

### 6.5.4 Discussion of Memory Experiment

Exit interviews showed no difference in mean difficulty of the task between the first (M=5.28, STD=.987) and second conditions (M=4.98, STD=1.250), t(40)=1.190 p=.237 (2-tailed *t*-test). Furthermore, there was no difference in mean difficulty between video A

**Figure 6.5:** Scene extracted from the modulated condition. The high salience object (top left) had equally high recall hits (12) in both conditions. The low salience object (bottom left) had a score of 8 (LR) in the unmodulated condition and achieved a score of 13 ($HR$) in the modulated condition.

(M=5.25, STD=1.171) and B (M=5.00, STD=1.086), t(40)=.990 p=.325 (2-tailed *t*-test). The main contribution of this work is to show that the SMT introduces imperceptible changes to a video that increase recall of selected objects, without significantly reducing recall of others. The resource addressed, namely memory, is limited. In this study, participants in both conditions tended to remember the same number of objects, $\chi^2$ (1, N = 880) = .39, p = .53. There is a tradeoff where the recall of some objects is reduced, while that of others is increased. In practice, our observations showed that a participant would recall on average five objects (at most eight) with certainty.

Some objects in $HR$ decreased in recall, but not significantly (H4). Conversely, objects that were filtered out (objects $o \in MR$) also decreased in recall, albeit not significantly. In comparison, recall of objects $o \in LR'$ significantly increased. This comparison is between scores for the same object; it does not mean that we can increase recall hits of an object over those of another object. In particular, it does not mean we can increase recall hits of an inconspicuous object over those of a conspicuous object. The results merely show that the SMT increases the chances of an object being remembered. Having clarified this, there were two cases where the scores of an object in $LR'$ increased to equal those of its scene counterpart in HR (Figure 6.5 shows one example). In both cases, recall hits in the first condition for the LR objects were at the $40\%$ limit.

## 6.6 Discussion of Saliency Modulation Technique

Our results indicate that the SMT can significantly shift attention to selected areas of a video, and it can increase recall of selected objects, without the viewer becoming aware of any manipulation. This provides strong evidence that the technique can influence the viewer's experience of a video at different levels of processing: it has applications in stimulus based conditions (bottom-up) and task based conditions (top-down). Furthermore, there was no noticeable increase in difficulty in assessing the modulated videos in comparison with the unmodulated ones.

### 6.6.1 Limitations

The restriction that no faces/hands appear in the clips seems to impact generalizability. However, in our experience, AR applications in non urban areas easily meet this requirement, and for the case of environmental monitoring the restriction does not pose a limitation. The results of our experiments do not guarantee that every viewer will attend to and/or remember the selected objects, but that they are more likely to, as compared to the original unmodified condition. However, thresholds can be adjusted interactively by passing a parameter to the SMT implementation (e.g., in response to eye tracker feedback). So, if an application needs to make the effects of the SMT perceptible, it only needs to increase the modulation threshold. Several factors have been identified that contribute to saliency (e.g., see Wolfe and Horowitz [Wolfe et al., 2007]). Of these, the SMT controls contributions in lightness, red-green color opponents, and blue-yellow color opponents, while other factors remain unaffected. In our studies, factors such as motion and size negatively affected results. Future research will need to address how contributors to saliency not controlled by the SMT affect its application. Meanwhile, the effectiveness of the SMT depends on the balance of these factors throughout the input. Avoiding extremes (e.g., small objects) can help in using the SMT successfully.

A major limitation for the SMT is registration: how do we decide that a certain portion of the video frame corresponds to a real world object that we want to emphasize. Vision based object recognition can provide an answer to this question, albeit with limitations of its own.

### 6.6.2 Application in Mobile Environmental Data Visualization

Our main motivation for developing and experimenting with the SMT is AR, in particular, information rich visualizations. Several applications can be imagined in the environmental monitoring domain. In this section, we will sketch two possible use-cases: first, to

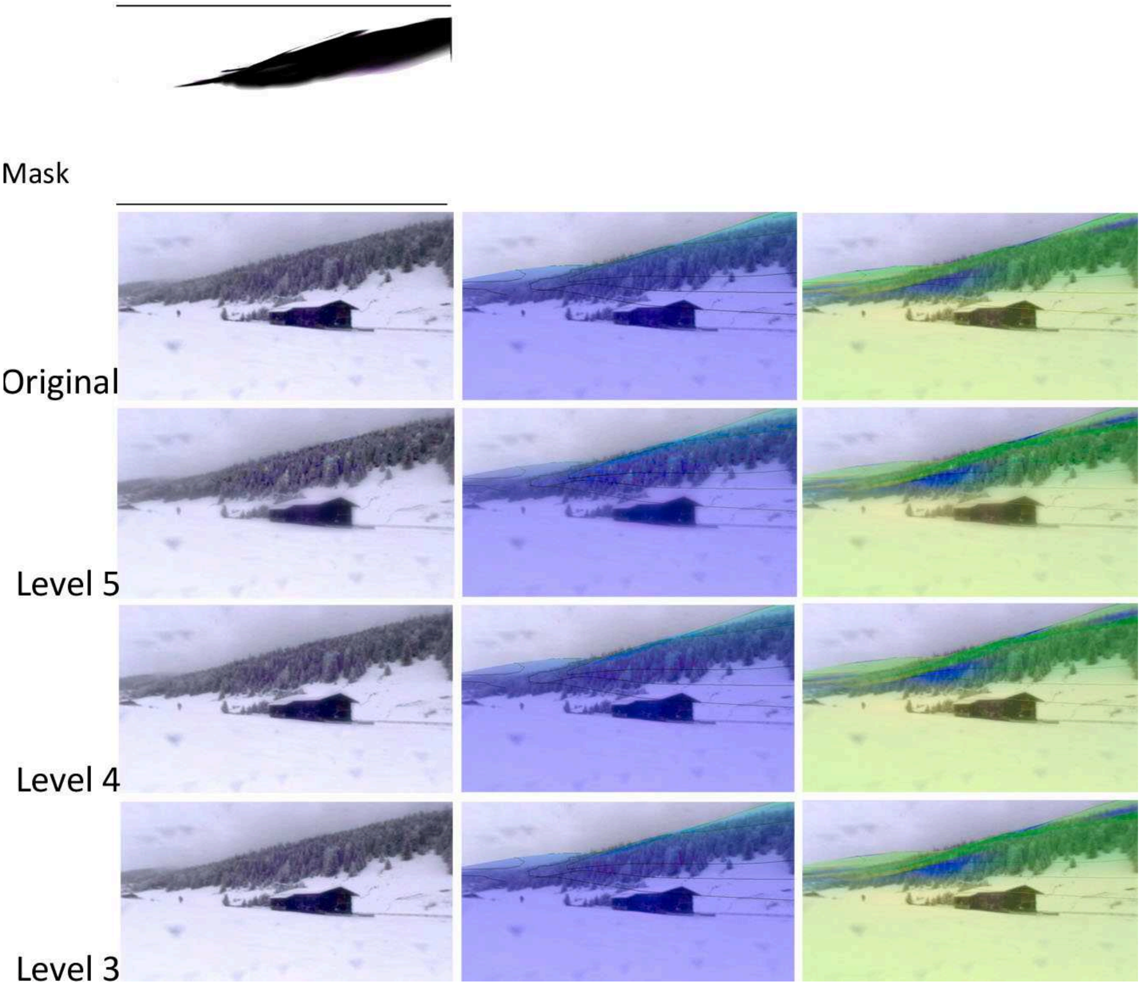


**Figure 6.6:** Saliency Modulation of Importance Segments. The mask is obtained by projecting a texture with the importance segment marked black onto the DTM. It determines where the SMT is applied. Three original images (Top row), are modulated at levels 5, 4 and 3.

emphasize spatial partitions on the information space, second, to modulate the visual input with discrete area data, such as land use.

Throughout this chapter and the previous chapter, a topic that was not address was semantic segmentation of the environment. Associating semantics to topological features of the environment enables powerful metaphors for filtering and selection based on semantics, or for focus and context visualizations. Based on this approach, a viewer could focus on a specific segment or zone (e.g., river bank), and observe results directly related to that zone, *blending out* data corresponding to other zones. Hereby, perceptual optimizations in term of screen space and clutter reduction are automatically elicited.

Nevertheless, the approach requires an appropriate segmentation of the area of interest. This is normally done or refined manually, in a tedious 3D editing process. Instead, we deploy a simple, straightforward method to define our segments and use them in combination with the SMT. Based on a terrain texture, we create texture partitions for each topologic feature of interest where to apply the SMT. We modify our engine, so that during rendering it outputs, beside the normal AR view, a rendering of the selected partition mask mapped onto the 3D terrain from the same perspective, both to textures. The former will be used as normal input to the SMT, and the latter will serve as mask, to define the region of interest. Figure 6.6 shows a segment created for the slopes, and the results after applying the SMT at thresholds 3, 4 and 5. Similarly, Figure 6.7 shows the results using a partition for land use. The portion of the mask shown in black is where the SMT enhances the saliency, whereas the rest is reduced to match the threshold. As can be seen in the images, the effects are hardly noticeable. Furthermore, the color mapping is preserved, maintaining coherence with the unmodified version.

These examples show how to avoid the issues and limitations of the SMT in practical applications, with the exception of its performance requirements. Instead of tracking an object of interest, we define it globally, by marking it on an otherwise empty texture map of the DTM. The projected texture serves to track the desired region. However, inaccuracies in tracking will directly affect the deployment of the SMT with this simple approach. Its effects in the described scenario are purely pedagogical, and need further study in order to make them of practical use.

## 6.7 Summary

This chapter took a novel approach on how to represent data and how to convey its properties in AR. The field of visualization in outdoor AR is open. Several future directions of research fan out, from novel representations to combine different data types, through algorithms for on-the-fly segmentation of 3D environments, to vision assisted methods for subtle guidance and visibility management. The next chapter complements the discussion on visual representation with the interactive workflow needed for visualization.

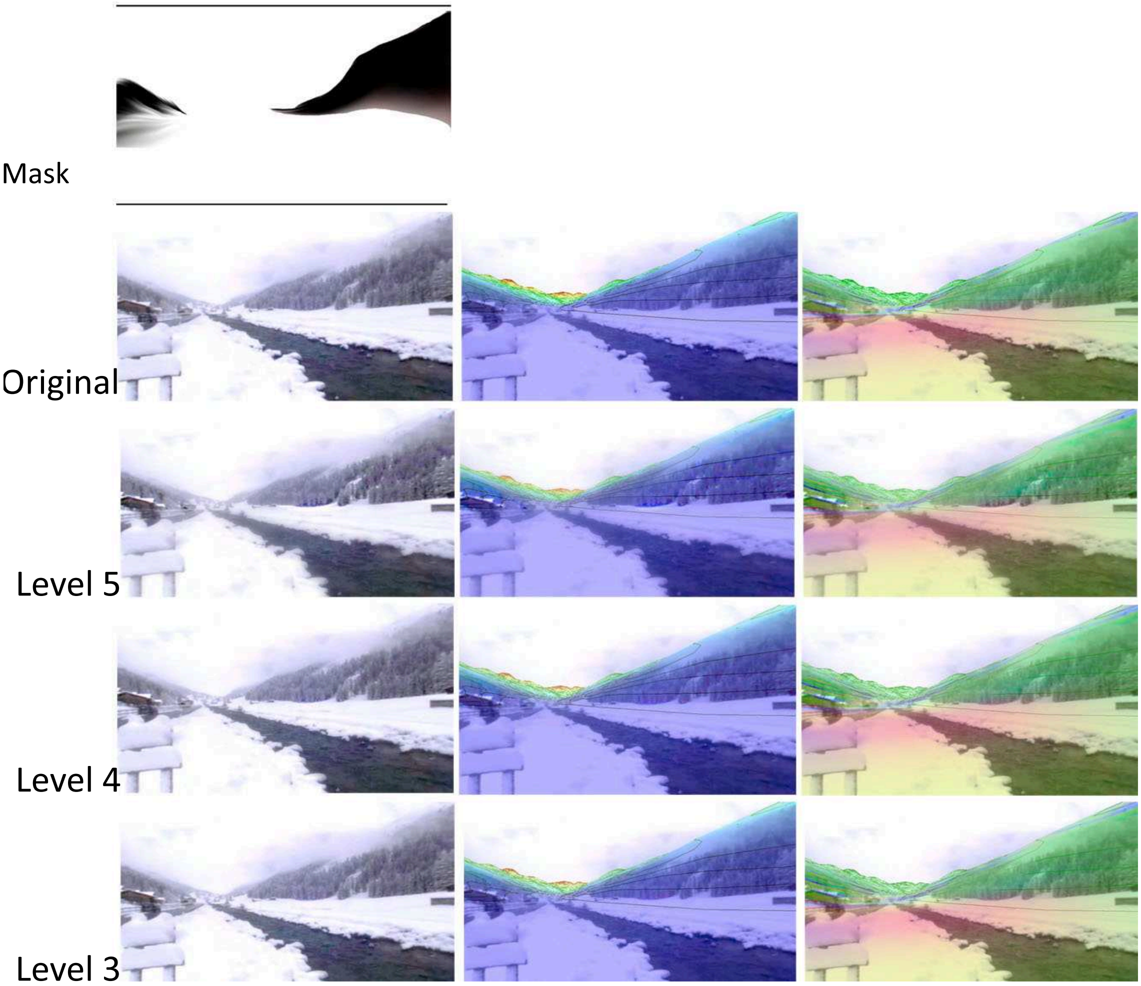


**Figure 6.7:** Saliency Modulation of Importance Segments. The mask is obtained by projecting a texture with the importance segment marked black onto the DTM. It determines where the SMT is applied. Three original images (Top row), are modulated at levels 5, 4 and 3.

# Chapter 7

# View Management in the Workflow of Interactive AR Visualization

Interaction plays a crucial role in the workflow of visualization. As we search through information, we take different perspectives to build a mental model, at times getting an overview, at times zooming or changing to close-up on a portion of the data. To better understand the situation, we filter out data, switch representation modes, or query further details on the data, a workflow defined by the well-known “information seeking mantra” of Shneiderman [Shneiderman, 1996].

It is important to note that we are not seeking to perform full data analysis in the field. This is a task reserved for comfortable space, with large displays and fast access to all sorts of data. Instead, our on-site methods are complementary, and provide a unique way to experience the scientific information in the context of the real world, in real-time. In this sense, the world captured by camera(s) is a distinctive part of our dataset, and we attempt to harness its expressiveness in the toolset presented in this chapter. Making sense of data, in this frame, implies an understanding of the relations between the real world and the rest of the data, creating a picture of the situation and maintaining it as new facts from real-time data become available.

AR presents an interaction metaphor with interesting advantages, but also shortcomings to this end. On the one hand, registering the information space to the user’s viewpoint implies intuitive navigation; AR relies on natural motion and the user’s understanding of space for navigation. On the other hand, it restricts the workflow of visualization in ways of gaining overview or zooming into portions of the dataset. To change the viewpoint in a traditional AR setting, the user has to move to the desired position. Visual exploration can greatly benefit from quickly accessing different perspectives on the information space with reduced physical movement.

The discussion on view management unveils novel metaphors and techniques to “view

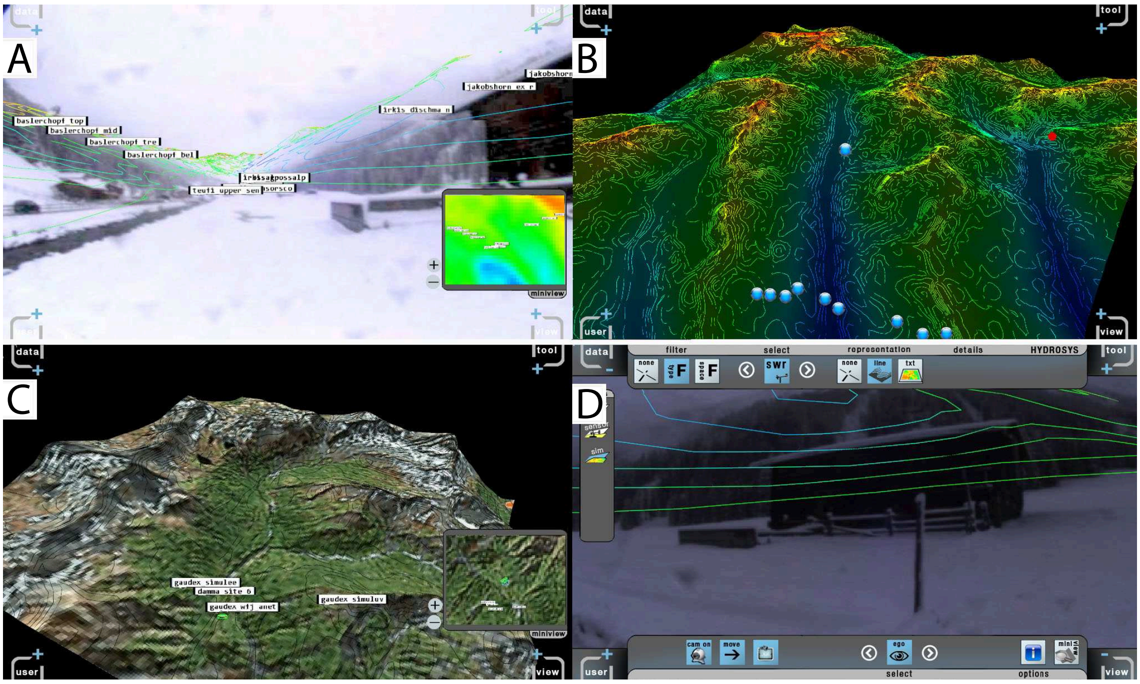


**Figure 7.1:** Multi-view AR used in a snow science scenario including an AR view (A), a virtual top down view (B), a 3rd person virtual view (C) and a second user's view (D).

the situation as a whole", responding to questions such as: How can situated users browse unhindered through large amounts of data, how can they zoom out, zoom in, or gain overview on the dataset while maintaining the context of the real world? How do they make sense of data from different perspectives? The answers to these questions are grounded in methods that combine viewpoints from virtual representations and real cameras. Leveraging a multi-camera infrastructure, a user can quickly access several points of interest in a large information space, calling augmentations and making judgements that could potentially reduce unnecessary physical movement. Furthermore, when working in teams, users can benefit from seeing what their peers are viewing and how it relates to what they see, a topic that will be discussed in the next chapter.

## 7.1 Extending the workflow with multiple perspectives

Traditional AR is restricted in its ways of gaining overview or zooming. As its ego-centric perspective situates the user within the dataset, a change of viewpoint in AR requires the user to move to the desired position. During visual exploration, users try to build a mental model of the situation (i.e., acquire situation awareness), and then visit interesting parts. This visualization workflow is compromised by traditional AR in three ways:

- AR narrows the overview to the portion of the world captured by the camera.
- Variable elevation in the terrain causes multiple occlusions and the spatial relationships between objects, and the environment become unclear.
- There is no way to zoom-in on a portion of the dataset without losing reference to the physical world.

We develop multi-view techniques to support the observation of a site without physically moving around. Our first solution extends the ego-centric AR view with additional real or virtual views, offering various fixed or dynamic perspectives. Consider for example, a surface skin temperature (TSS) visualization. By controlling or accessing a virtual camera, the user can leave the physical view and navigate the data representation, observe sensor locations, identify hot spots, cold spots, etc.; (see Figure 7.1).

However, virtual views breach the connection with the real world context. Hence, the multiview technique complements virtual views, with perspectives from other imaging devices in a single, generic framework for view management, including views from remote static or pan-tilt-zoom (PTZ) cameras, views generated by other users' devices, a view of 2D optical sensors (e. g., infrared cameras), cameras on unmanned aerial vehicles (e.g., blimp, drones, see Figure 7.2). These provide an interesting interface to browse not only the 3D representation, but also the real world associated with it.

Thereby, complementary and diverse perspectives on the area of interest extend understanding of the data by presenting overview, vantage zoom-in points, and generally increase situation awareness of the site. The major challenges of this approach are to deliver a simple way to represent and access these other views, and to define a clear understanding of the spatial relationship among their corresponding perspectives.

### 7.1.1 Understanding Multiple Distant Perspectives

In the process of dealing with the video from remote cameras, an observer must deal with a view discrepancy: The presentation of a remote camera view on a mobile device is separated from what the observer sees with her own eyes and perceives with the rest of her sensory system. The sensory qualities of remote camera feeds are necessarily reduced compared to direct perception, and users may lack information on how to get to the remote location or what lies in between them and the camera. The available information is mostly egocentric, though the user's persistent representation – the mental map – is exocentric. The process of building mental maps is affected by knowledge of the site, the spatial ability of the user and the features that can be matched from the environment to the remote camera perspective (e.g., shared landmarks).

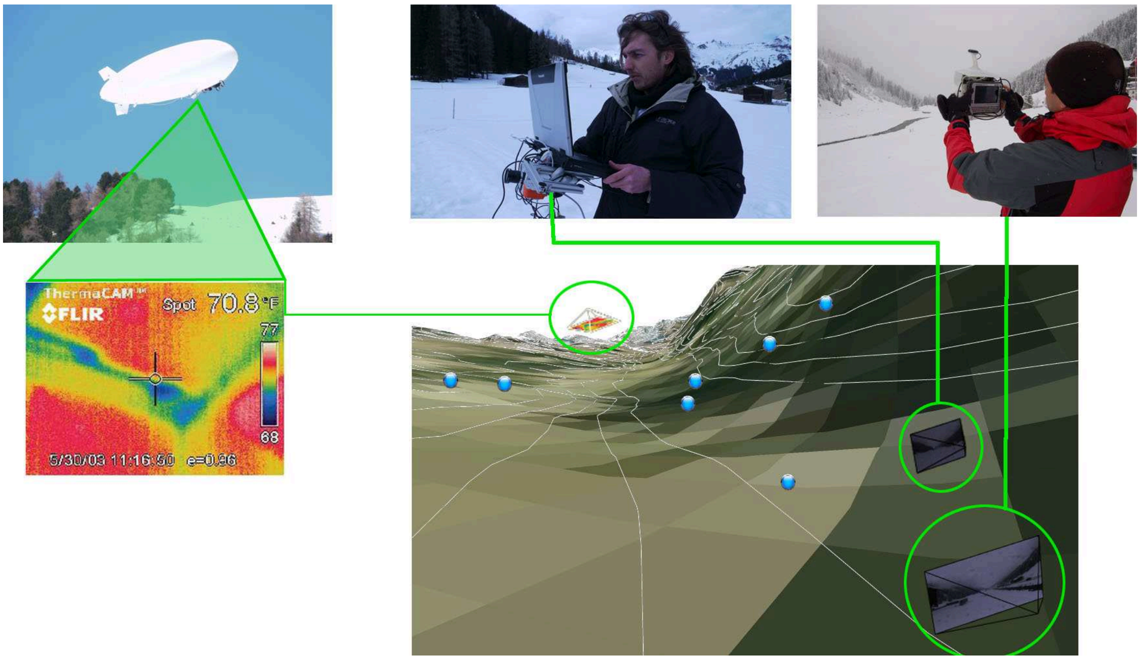


**Figure 7.2:** Combining Real and Virtual Views. A flying drone captures heat images, and two other users observe the environment with handheld devices. The three views are merged in the virtual representation of the environment.

Navigation techniques are designed to assist the process of building mental maps, by substituting missing sensory input in different forms. Previous work in this area concentrates on indoor surveillance and does not consider the characteristics of agile outdoor users with small mobile computers.

The two main issues for demanding tasks, such as multi-camera observation, are situation awareness and mental workload. Following Tsang and Vidulich [Tsang and Vidulich, 2006], we regard them as two separate constructs: situation awareness as a cognitive construct, and mental workload as its ”energetic” counterpart, mostly referring to the effort a user needs to invest. This section aims to explain what affects the design of a user interface that imposes low workload, but allows for high situation awareness.

#### 7.1.1.1 Situation awareness

The reason for deploying multiple cameras is because we assume they will likely provide a better overview to assess specific situations. This introduces the concept of situation awareness, a dynamic construct that results from a cognitive process entailing perception of cues in the environment, comprehension of the current situation and projection of future status [Endsley, 1995]. Situation awareness encompasses a person’s tasks and forms a basis for decision-making. Spatial awareness is a part of situation awareness that deals with the understanding of space. Spatial awareness includes a person's knowledge of self-location within the environment, of surrounding objects, of spatial relationships among objects and between objects and self, as well as the anticipation of the future spatial status of the environment.

Navigation involves gathering and applying spatial knowledge. While navigating, multi-sensory input is processed and stored in a mental map that represents spatial knowledge [Bowman et al., 2004]. A person experiences the world from an egocentric perspective, where all perceptual input is relative to the personal frame of reference. However, spatial knowledge is presumably stored in exocentric form, representing elements relative to each other and to a spatial reference frame [Easton and Sholl, 1995]. Mentally computing relations between different reference frames introduces distortions to spatial information [Tversky, 2000].

#### 7.1.1.2 Mental workload

Reasoning about inter-object relationships causes mental workload, an energetic construct that refers to the supply and demand of attentional and processing resources [Tsang and Vidulich, 2006]. Mental workload is affected by both the structure and the situation of a task (exogenous factors), and the abilities of a person (endogenous factors) such as spatial ability and experience with a system.

A task can be characterized by its difficulty, priority and related situational contingencies. For example, users may or may not have prior knowledge of the environment and the cameras located therein. Apart from the task's characteristics, both attention and processing of spatial information play an important role. Attention is a critical factor in the usage of handheld devices: users get easily distracted by occurrences in the environment and may be further challenged by adverse conditions such as reflections on the screen or strong sunlight. Attention is a scarce resource in the acquisition of situation awareness. The concentration required for switching between screen and environment likely affects the success of a navigation technique. When viewing a remote camera feed, the user has to make a mental transformation of her own location to the remote camera to establish a mental relationship. Such transformations are prone to errors, and may be further challenged by the size and possible bad legibility of the screen.

Research has indicated that the nature of the transformation affects the underlying neural implementation [Zacks et al., 2001] , workload and accuracy [Shepard and Metzler, 1971]. Minimizing complex mental transformations is a key requirement for achieving low mental workload and sustaining awareness. This can involve conveying the (unknown) location of the camera and indicating how to arrive from the current point to the remote point. Moreover, conveying camera orientations helps in disambiguating common objects.

#### 7.1.1.3 View discrepancy

Regular navigation techniques are based on the premise that they should provide enough sensory input (even though artificial) to give a sense of (or replace) the sensation of moving from one place to the other. This premise does not hold for handheld users and introduces a *view discrepancy*.

Instead of traversing between positions, the user remains at one position, but looks at the video feed from another location. While observing a remote video feed in the field, the view shown on the display device is inconsistent with what the observer perceives directly. Perception of the immediate physical environment is not completely discontinued while concentrating on the remote view. Hence, a perceptual conflict is introduced: users can observe multiple information sources that do not necessarily match in content or fidelity.

The extent of the discrepancy is affected by the relationship between user and remote camera Figure 7.3. We classify the relationship between the camera held by the user (L) and a single remote camera (R) by considering the remote camera as seen from the user's point of view. The classification takes into account whether the viewpoint of the remote camera is visible or not, and whether the camera is observing the same or a different scene (see Figure 7.3):

**camera in view – same scene (CS):** the viewpoint of the remote camera is visible and is (partially) observing the same scene as the user. Mental transformations required to connect viewpoints can be derived by inspection.

**camera in view – different scene (CnS):** the viewpoint of the remote camera is visible, but due to occlusions in the environment, it is observing a scene that the user cannot see.

**camera not in view – same scene (nCS):** the viewpoint of the remote camera is not in view, but it is (partially) observing the same scene as the user. This is the case when the camera is occluded: If the user requires only rotating in order to see the remote camera, the camera is considered in view. To derive the relation between viewpoints, the observer must match common objects in both views and perform mental transformations about those objects.

In a fourth category, camera not in view – different scene, the viewpoint of the remote camera is not in view and it is not observing the same scene as the user. We do not consider this case, because the relation between viewpoints cannot be derived without further knowledge. Although previous knowledge of the observer can aid in the process, the user may know the site and could thus derive the location of the camera.

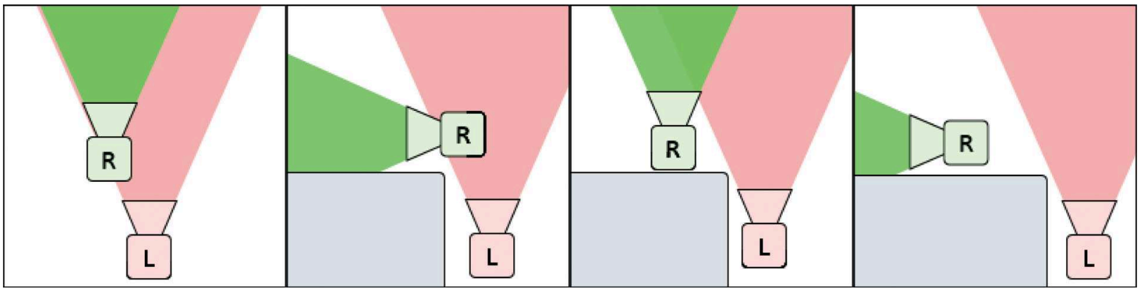


**Figure 7.3:** Camera Configuration Types. For a local camera (L) and a remote camera (R), we classify camera configurations based on the location of the remote camera from the viewer's point of view, and on the scene the remote camera is looking at.

### 7.1.2 Multiview User Interface

Our generic multi-view framework is associated with a shared-view software infrastructure (back-end system) that provides run-time access to parameters and content for each view, introduced in section 3.3.4. It stores intrinsic parameters for each camera and regular updates of tracked frames. From there, the user can access any of the views through the *view* interface, either by selecting it from a scrolling list (see Figure 7.4-A) or by tapping its iconographic representation (see Figure 7.4-B). These models are complementary, as a specific viewpoint may not be visible from the current user location and orientation, or it may be out of the current camera range (e. g., behind the user).

By default, the user is shown the first person perspective generated by her own handheld camera (standard AR view). The user can transition from this current view to a default virtual view, which allows navigating to any location in the virtual environment. Virtual views can be saved for later reference. Additionally, the application includes two predefined and frequently used virtual views (see Figure 7.1): a top-down overview of the environment (navigation restrained to pan and zoom), and a 3rd person virtual 3D at 45° behind and above the user (navigation through key controls or through graphical buttons mapped to camera controls). Some of the additional views can be associated with navigation controls (e. g., 3D virtual view, remote PTZ camera), some others are only defined as an end-point view (e. g., view of a remote static camera). For this purpose, each of the views is also associated with navigation control parameters. When a view is switched, graphical controls are automatically made available to the users (see Figure 7.4-D).

Two alternatives for contextualized videos were considered: a projected texture in the fashion of AVE [Neumann et al., 2003] and a billboard always facing front. The latter maximizes the viewable region of the video; but it does not convey orientation or view direction. The former was dismissed because it produces rendering artifacts. As described in [Neumann et al., 2003], when the 3D model is inaccurate, projected textures deform

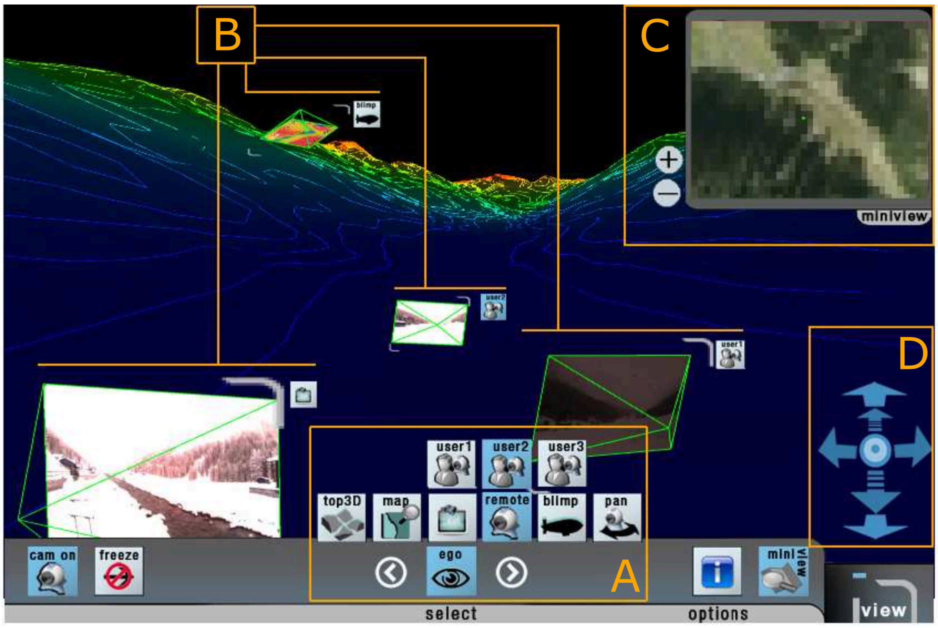


**Figure 7.4:** Multi-view AR interface. (A) scrolling list of views. (B) iconographic representation of views. (C) minimap from a top-down virtual camera. (D) navigation controls for additional views, here for a virtual 3D camera view.

features of objects that do not exist in the model (a barn would project all the way to the nearest mountain). Then, for the iconographic representation, we chose a wireframe frustum model that we spatially register with the pose of the additional view. The frustum is created from the intrinsic parameters of the camera, and its position and orientation are dynamically updated from the shared-view system. We also added a thumbnail representation for each view, providing a preview of what content is visible from it.

All these spatial awareness aids ensure that the user can understand where the remote view is located (spatial awareness), what portion of the world it observes (referential awareness), and what it sees in that portion of the world (view awareness). Finally, the application can present the video-feed of a secondary view in a mini-frame. In the current implementation, the secondary view is restricted to the top-down view, showing a mini-map centered on the user (see Figure 7.4-C) providing and supporting better spatial awareness of the user position, view direction, scene content and the surroundings, thus delivering an overhead view of the site for contextual information.

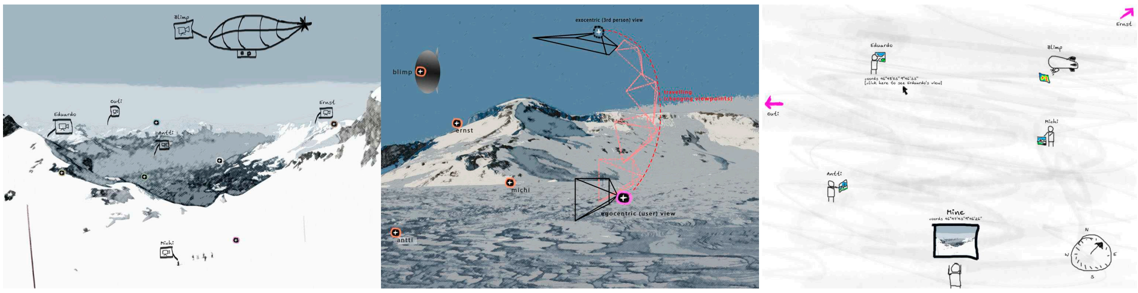


**Figure 7.5:** Viewpoint Transition Concept. A decides to change from AR mode (Left), to top-down virtual view (Right). The transition path brings him smoothly to the desired viewpoint (Middle). [Kruijff et al., 2009]

### 7.1.3 Viewpoint Transition in Multi-Camera Environments

Following the discussion about understanding different dynamic perspectives, we considered options to navigate the cameras. Figure 7.5 shows the concept of transitioning between views. A user changes from egocentric viewpoint to exocentric viewpoint, and the system makes a smooth transition following the path plotted in pink in the middle image. As navigational support, we developed and evaluated three techniques for viewpoint transition (Figure 7.6). Different considerations can be taken to transition between a local and a remote view. Techniques can be either egocentric or exocentric, and may take an uninformed or an informed approach. The latter property refers to data about the environment, such as maps or 3D models. Informed techniques require the user to interpret extra data to make sense of the environment. Our techniques are uninformed in the sense that they only convey information about the spatial configuration of the camera setup, but not on the full environment. We expect that additional spatial knowledge is retrieved from observation of both the camera feeds and the environment itself.

The three proposed techniques (see Figure 7.6) represent main research directions on navigation and multi-camera systems, adapted to mobile AR. They convey similar information in different ways, which allows for a proper comparison of workload and user acceptance. The Mosaic technique represents a typical surveillance-system solution [Girgensohn et al., 2007], displaying camera feeds in tiles organized topologically with respect to the user pose. The Tunnel technique is adapted from a recent technique used in AR to guide the user to an object [Biocca et al., 2006]. The transitional technique is based on an interface that allows users to transition between contexts [Grasset et al., 2006].

#### 7.1.3.1 Mosaic technique

The Mosaic technique allows users to transition between local/remote views using a mosaic of video thumbnails. The thumbnails resemble the visualization used in the control

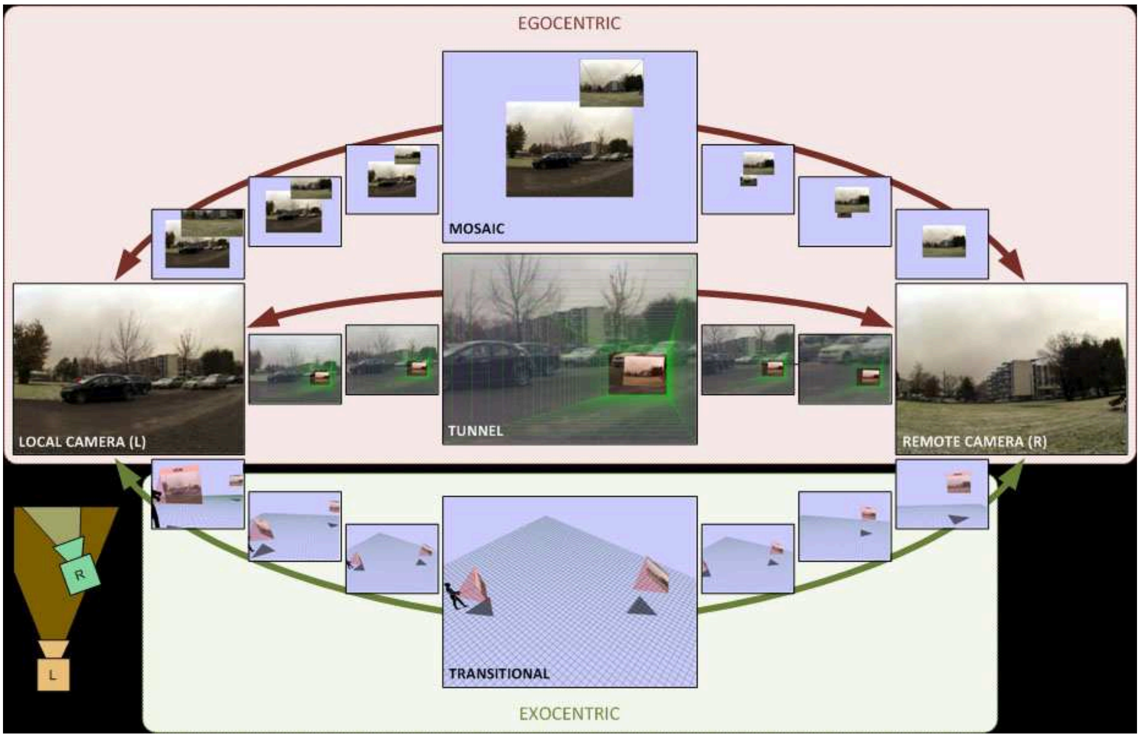


**Figure 7.6:** Transition Techniques. Proposed techniques applied to a CS condition. Using the techniques, users can browse the video stream from either the local or the remote camera, or they can smoothly move to a view where both videos are visible.

room of surveillance systems. We organize the thumbnails based on topology, similar to DOTS [Girgensohn et al., 2007]. The technique uses the angle between the viewing direction of the local camera and the position of the remote camera to position, on the screen, minimized versions of both videos. This conveys to mobile users "how they should turn" in order to "see" the camera, like a compass. As a user moves, the visualization updates accordingly.

The mosaic technique does not show the 3D spatial relation and the distance between cameras. It is primarily a 2D technique, providing a directional cue towards the remote camera. The thumbnails show both videos, allowing users to get a minimized view of both cameras at the same time. Since the organization of the thumbnails does not depend on distance, this technique allows the visualization of several cameras simultaneously, as long as the cameras are not in the same direction, see Figure 7.6.

#### 7.1.3.2 Tunnel technique

The Tunnel technique is a variation of the attention funnel, first introduced by Biocca et al. [Biocca et al., 2006]. This is an egocentric technique to guide a user to an object of interest. The technique displays a tunnel oriented to the remote camera. Users can travel down the tunnel to the other camera. We blend the tunnel over the video background so that tunnel and video are both visible. When the remote camera is in view, the user can see its video feed at the end of the tunnel. If the remote camera is not in view, the tunnel indicates the turning direction.

Traveling down the tunnel translates the user to the other camera. The technique is expected to work best when a user first rotates until the remote camera is in view and then travels to it. In this case, the technique conveys a complete spatial relation, including rotation, translation and view direction. The distance to the camera is correctly shown in 3D from the perspective of the user. The Tunnel shows a wireframe with one meter segments, see Figure 7.6.

#### 7.1.3.3 Transitional technique

The Transitional technique implements the concept of transitional interface in the sense of [Grasset et al., 2006]. In a transitional interface, users can transition between contexts, each possibly having a different space, scale and representation. In our case, users can move between an egocentric AR context, where a full-screen augmented video is visible, and an exocentric VR context, where users get a bird's eye overview on both cameras and their respective spatial position and orientation. In the exocentric view, an avatar is used to disambiguate the user's camera from the remote camera. We employ smooth animations to support coherent transitions, shown in Figure 7.6 Bottom.

### 7.1.4 Evaluation of Transition Techniques

We conducted a user study to compare the techniques asking users to infer relationships between a local camera and a remote camera. For each type of camera configuration (see section 7.1.1.3), we focused on the impact of the techniques on the users' spatial awareness and mental workload. The test setup used a handheld device consisting of an ultra-mobile computer (Panasonic CF-U1 with a 5.6" screen), a Ublox GPS sensor and a uEye UI-2210 color camera (640x480 resolution) mounted with a 4.2MM Pentax wide-angle lens. The uEye camera was physically bound to the whole setup, and acted as the local camera in the user study. All three techniques were used in combination with pre-recorded video feeds from static cameras. During the experiment, a tripod was positioned in the field to represent the remote camera.

The aim of the comparative test was to answer the following questions, for each type of camera configuration:

**Q1.** Were there differences in spatial knowledge obtained from the different techniques?

**Q2.** Which technique has less impact on the user's workload?

**Q3.** What is the user's preference: Do users pend towards one technique when they are asked for subjective impressions on various parameters?

#### 7.1.4.1 Methodology

The study was conducted as a single experiment and employed a 3x3 factorial design. We treated the 3 camera configurations (CS, CnS, nCS) as a between-subject independent variable: Users were divided in three groups, and each group experienced a different type of camera configuration. The type of transition technique (mosaic, tunnel, transitional) was treated as a within-subject independent variable. Hence, every participant was assigned the same camera type over three locations and varied the technique per location (all camera configurations were available at every location).

We used a Latin square distribution to balance the order in which the techniques and the locations were assigned to each user. Before the experiment, we collected demographic data, some information on the amount of time users spend with both paper and digital maps, and information on their spatial abilities, for which we used the SBSOD questionnaire [Hegarty et al., 2002].

The experiment started with an outdoor introductory session, where users could get familiar with the handheld device and the techniques. A dummy remote camera was provided for practicing. After the introduction, we blindfolded and walked the users to three different locations at our University campus. The three locations had varying levels of features, including different types of buildings and varying density of trees.

At each location we asked the user to identify the position of one remote camera, using a single technique. To prevent user biasing, the techniques were named with neutral names (A, B and C). Users were asked to draw a map representing the main objects in the scene, the camera locations and their orientation. Thereafter, they filled in a short questionnaire to assess spatial awareness and workload, using parts of the NASA TLX questionnaire [Hart and Stavenland, 1988] and an RSME (Rating Scale for Mental Effort, scale 0-150, 150 for maximum effort [Zijstra, 1993]). Finally, we collected the level of user confidence at each location. After the three techniques were used, we asked participants to state their preference for a variety of factors related to spatial awareness and workload. In total, participants had to fill out 12 pages of questions (31 spatial ability questions and 41 ratings based on Likert scale) and draw three maps.

27 users (16 male, 11 female, aged between 22 and 48) participated in the study. All (but one) participants had normal or corrected to normal vision. To partially compensate for the effects of prior knowledge of the environment, we invited 17 users that regularly visited the campus and 10 users that had hardly visited the campus before. Of the 17

regularly visiting users, 9 users did hardly know at least one of the locations. Results were collected from the 27 users x 3 locations, totaling 81 trials. The duration of the experiment was about 1.15h per participant.

#### 7.1.4.2 Results on situation awareness

Based on the median score of the SBSOD ($65.56\%$) and on differentiation of the results, we separated users in three groups:

**G1:** below average ($< 55\%$), 2 male / 7 female

**G2:** average ($55 - 75\%$), 9 male / 3 female

**G3:** above average ($> 75\%$), 5 male / 1 female

Female users were more prevalent in the G1 and G2 spatial ability groups, which is in line with other studies [Lawton, 1994] (Pearson correlation, $p < .01$). It is important to note that, though we applied a Latin square distribution of users and we enforced balance of gender and familiarity with the environment, we cannot make exact statements on user group effects per camera type: no high spatial ability users fell within the nCS condition.

We started with interpreting the maps drawn by participants. Due to the diversity and quality of drawing, we only made rough estimations on errors. We resorted to a voting approach, whereby researchers check for errors in the overall spatial configuration (VS), and separately in the position (VP) and orientation (VO) of the remote camera. The results of this analysis provide very interesting indications.

| | Mosaic | | | Tunnel | | | Transitional | | |
|---|---|---|---|---|---|---|---|---|---|
| | VS | VP | VO | VS | VP | VO | VS | VP | VO |
| CS | 0 | 1 | 1 | 3 | 2 | 3 | 2 | 1 | 3 |
| CnS | 2 | 1 | 1 | 3 | 1 | 2 | 3 | 1 | 0 |
| nCS | 1 | 3 | 0 | 2 | 3 | 3 | 1 | 1 | 1 |
| G1 | 1 | 2 | 0 | 5 | 3 | 3 | 3 | 1 | 1 |
| G2 | 2 | 3 | 2 | 2 | 3 | 3 | 3 | 2 | 1 |
| G3 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 2 |
| TOTAL | 10 | | | 22 | | | 13 | | |

**Table 7.1:** Total number and types of errors in the map drawings for each technique. Results are shown for each camera configuration, each spatial-ability group, and overall.

Mosaic caused the least errors in the drawings (Table 7.1). Of particular interest is that users made surprisingly few errors when drawing the remote camera position, even if the

| | G1 | G2 | G3 | Male | Female | AVG |
|---|---|---|---|---|---|---|
| Mosaic | 3.78 | 2.67 | 2.17 | 2.50 | 3.55 | 2.93 |
| | *1.09* | *1.07* | *0.75* | *1.10* | *1.04* | *1.17* |
| Tunnel | 3.11 | 2.58 | 2.67 | 2.31 | 3.36 | 2.74 |
| | *1.27* | *1.51* | *1.05* | *1.20* | *1.29* | *1.32* |
| Trans | 3.44 | 2.33 | 1.67 | 1.88 | 3.55 | 2.56 |
| | *1.24* | *1.37* | *0.52* | *0.96* | *1.21* | *1.34* |

**Table 7.2:** Averages and standard deviation (in italic) for self-assessed success ratings (for each spatial ability group and for gender). 7 point Likert scale, lower scores are better.

technique itself does not provide any distance information. Transitional performed very well in the nCS condition, especially when one considers that no high spatial ability users fell within this condition. The technique seems to provide quite accurate information on the placement and orientation of the remote camera, when it is not visible by the user. High-ability participants made significantly fewer errors (Pearson correlation: $p < .01$). Previous knowledge of the environment only had a significant main effect on errors for the Tunnel, but not on the other techniques (one-way ANOVA, $F1, 25 = 9.04$, $p < .01$): Users with previous knowledge performed better with Tunnel than users with no previous knowledge.

For each technique, users assessed their success after drawing the map (Table 7.2). In general, participants with higher spatial ability felt more confident. The previously stated correlation between spatial ability and errors in drawing the maps supports the personal assessment. Regarding the techniques, groups G2 and G3 felt most confident with the transitional technique, whereas group G1 preferred the tunnel. Hence, it is important to note that, though Mosaic caused the least errors, users did not report the highest confidence in this technique for drawing the map.

We noticed an interaction effect between spatial ability groups and the success rating ($F_{2,24} = 5.17, p = .01$). A post-hoc test showed a main effect of spatial ability on success ratings for Mosaic ($p = .02$) and Transitional ($p = .03$): Higher-ability users felt more confident using these two techniques than users with lower ability. Spatial ability didn't show a significant impact on the self-assessed success for Tunnel. One-way ANOVA also showed that camera types did not have a significant effect on the subjective success rating. The only noticeable result is that nCS is rated slightly higher, whereby participants tended to be less confident in that condition. There was no significant effect between the techniques.

In general, users estimated that with either technique they needed to retrieve as much information from the screen as from the environment itself. This partially confirms our expectations: the techniques provide only limited information, users would have to observe the environment to fulfill their task. We expected that Mosaic would force participants to observe the environment more, whereas the Tunnel would require more attention to the screen. However, a repeated measures ANOVA provided no significant difference among techniques, suggesting that for all of them participants had to pay as much attention to the screen as to the environment. Neither did we find significant effect of camera type on the focusing of attention on either screen or environment.

**Q1.** Were there differences in spatial knowledge obtained from the different techniques?

Mosaic tends to perform better, producing the least errors but not causing the highest confidence among users. Transitional tends to give higher confidence (for drawing maps), and seems to perform better when the remote camera is hidden from the user. Users with higher spatial ability apparently prefer either Mosaic or Transitional to Tunnel. Users in the lower spatial ability group had a slight preference for Tunnel, although it seemed to produce more errors.

#### 7.1.4.3 Results on mental workload

To analyze workload, we considered both the workload-related questions (derived from TLX) and the RSME scale. We found a direct correlation between mental workload and RSME per-technique ratings (Pearson correlation, $p < .01$ for all techniques), which gives a reliable base to judge the workload.

There is a tendency of Mosaic to require less workload and of Tunnel to require more, but a one-way ANOVA did not show any significant effect between technique and mental workload, nor between spatial ability and technique. Additionally, no significant effect could be found between group, technique and camera conditions after multivariate analysis. Finally, a one-way ANOVA did not show any effects on the order and progress of the test on the mental workload.

**Q2.** Which technique has less impact on the user's workload?

There was a tendency of the Mosaic technique to require less workload. Although not significant, Transitional received a better rating than Tunnel. No significant effects of the camera types on the workload could be found.

| | Ease of navigation | Ease of use | Drawing map | Effort | Helpfulness | Attention | Confidence | Like |
|---|---|---|---|---|---|---|---|---|
| Mosaic | 2.67 | 2.70 | 2.70 | 2.33 | 2.89 | 2.37 | 2.74 | 2.70 |
| | *0.62* | *0.67* | *0.54* | *0.83* | *0.32* | *0.84* | *0.53* | *0.67* |
| Tunnel | 2.37 | 2.41 | 2.63 | 2.37 | 2.59 | 2.19 | 2.26 | 2.11 |
| | *0.79* | *0.64* | *0.63* | *0.69* | *0.57* | *0.79* | *0.76* | *0.75* |
| Trans | 2.74 | 2.67 | 2.70 | 2.44 | 2.85 | 2.48 | 2.63 | 2.56 |
| | *0.53* | *0.55* | *0.47* | *0.70* | *0.46* | *0.70* | *0.56* | *0.64* |

**Table 7.3:** Average preference ratings (3 point Likert scale, higher is better) and standard deviations (in italic).

#### 7.1.4.4 Technique preference

Analyzing the general preferences of the techniques (3-point Likert scale), most users liked Mosaic best, followed by Transitional. A repeated measures ANOVA showed an effect on user preference: A post-hoc t-test showed that Tunnel was significantly less preferred than Mosaic ($p < .01$) and Transitional ($p = .03$). We noticed no significant effect of spatial ability on the technique preference. However, if we look into the details of the ratings, several differences can be noticed (Table 7.3).

Users in the G2, G3 groups liked the Tunnel technique less, consistently for all camera conditions. Mosaic was liked most in all camera conditions, whereas Transitional is not considerably disliked, especially in the CnS and nCS conditions.

There were no significant differences between the ratings of the techniques for attention, effort, general navigation preferences, as well as the usage of the techniques for drawing a map. We noticed that Mosaic required more attention than the other techniques in the CS condition, performing worse than Transitional. Tunnel performed worst in all conditions. Subjective effort ratings support the findings from the workload section.

Users gave high ratings to the usefulness of all techniques in helping them to draw a map. A repeated measures ANOVA showed a significance in the confidence rating (p = .02): a post-hoc t-test showed that users were significantly more confident in Mosaic than Tunnel (p = .03), but not significantly more confident in Mosaic than in Transitional. Spatial ability did not have a significant effect on the preference ratings. In general, lower spatial ability resulted in slightly lower preference rates, and worse attention rates. Also, effort was slightly increasing with decreasing spatial ability. It is interesting that increased effort did not significantly reduce the preference for a technique.

**Q3.** What is the user preference: Do users tend towards one technique when they are asked for subjective impressions on various parameters?

There was an overall preference for Mosaic. Though it does not always perform significantly better than Transitional, for most camera types and spatial ability groups it seemed to receive higher preference ratings.

#### 7.1.4.5 Discussion of View Transition Results

The way in which the relationship between user viewpoint and remote camera is communicated varies widely between techniques. Tunnel and Mosaic work in an egocentric mode – Tunnel uses AR methods to overlay all information on the video stream, whereas Mosaic shows only the relative location of the remote camera, hence just giving a directional cue. Transitional, on the other hand, works in exocentric mode, and reveals the full spatial relationship, including distance. We found that with all techniques, users had to gather information equally from both surroundings and video feeds, to infer the spatial relationship between cameras and the environment. To our surprise, the fact that some techniques provide more information does not seem to have a strong impact on the behavior to observe the environment.

The preference evaluation shows that all users found the techniques equally helpful in drawing the maps. It is interesting to contrast this result with the self-assessment on spatial awareness, where users of group G3 thought they were most successful with any technique but the tunnel, while users in G1 believed they were mostly successful with the tunnel. In the preference evaluation, most users considered Mosaic and Transitional most helpful, while Tunnel rated lower. Similarly, users deemed Mosaic and Transitional were easier to use and to navigate with compared to Tunnel. Participants felt significantly less confident with Tunnel. This actually contrasts with the confidence users reported after drawing the map, where Mosaic rated lowest. This is quite surprising and might infer either irregularity during the ratings of the users, or a difference between how users interpret confidence and success. It should be noted at this point, that users found it difficult to assess the success rating during the experiment.

An unexpected result from the preference survey is that users found that they needed as much effort for any of the techniques, while attention scores where similar, with tunnel scoring a bit lower. However, when comparing these results with the ones obtained for workload, the results show a difference especially in favor of Mosaic, and partially of Transitional. This may indicate that users prefer a technique that requires less workload and provides more confidence.

| | CS | CnS | nCS | G1 | G2 | G3 | AVG |
|---|---|---|---|---|---|---|---|
| SA errors | M | M<br>Tr | Tr | M | Tr | M | M |
| SA success | Tu<br>Tr | Tr | Tu<br>Tr | Tu | Tr | Tr | Tr |
| Mental load | M | M | M | M | Tu | M | M |
| Like | M | M | M | M | M | M<br>Tr | M |
| Attention | Tr | M<br>Tr | M | Tr | M | Tr | Tr |
| Confidence | M | Tr | M | M | M<br>Tr | M | M |

**Table 7.4:** Overview of best-performing techniques (highest averages) on main factors in situation awareness, workload and user preference. Mosaic (M), Tunnel (Tu) and Transitional (Tr).

Regarding camera types, we can conclude that all techniques help users infer where the cameras are, and produce similar results even for the nCS condition. Camera types mostly did not have a significant effect on the results. Still in the nCS condition, participants drawing maps from transitional tended to commit fewer errors (Table 7.4). Hence, for sites with "hidden" cameras, Transitional might be preferable over Mosaic. It is instructive to compare the results on workload with those on spatial awareness. When observing the self-assessment on spatial awareness, we noted that users in groups G2 and G3 felt they were most successful with Transitional, followed by Mosaic and then Tunnel, albeit with no significant difference. In comparison, the number of errors was lowest using Mosaic followed by Transitional and Tunnel. This implies that users are not fully aware of their overall performance. Compared to workload, we can infer that users prefer a technique that lets them perform reasonably well while imposing lower workload.

While observing participants, we noticed minor technical difficulties that may have an effect on the practical usage in real-world scenarios. The techniques vary in their resilience to accuracy and registration error (error in the alignment between virtual and real objects): Mosaic can cover better for errors than the other two techniques. We noticed that some participants were disappointed when registration errors occurred (the remote camera appeared to be off in the transition view) with either technique, and were forced to observe the environment more closely, even though this is not directly confirmed in the ratings. Time allocated to screen and environment was about equal throughout all techniques. In the comments and discussion after the experiments, participants did not raise the issue.

Some participants complained about the view distortion introduced by the use of wide-angle lenses, and the sometimes limited legibility of the screen. Illumination and weather conditions were roughly the same for all users, and should therefore not have had a significant influence on our findings.

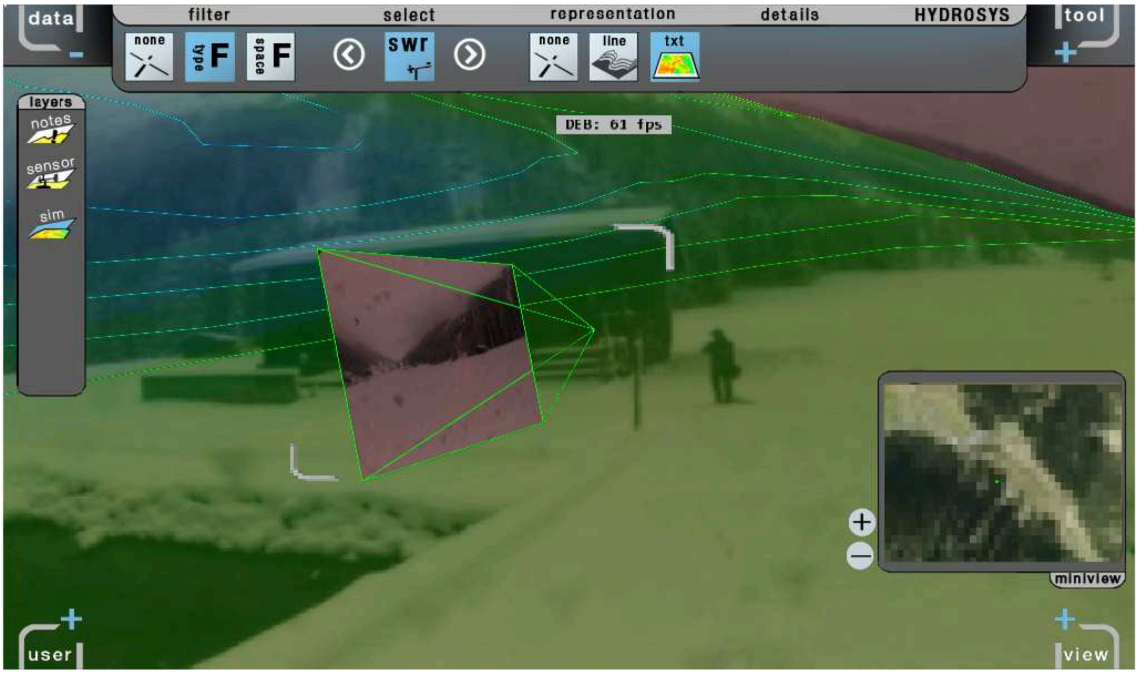


**Figure 7.7:** Unresolved occlusion. The remote view is behind the barn, which does not exist in the 3D model. The system is unable to compute the occlusion.

### 7.1.5 Multi-View Evaluation

Our studies of multiview techniques focused on awareness acquisition and cognitive load, both factors that guided our designs. The tasks of mentally deriving spatial relationships, maintaining referential awareness and other factors inherent to view sharing require a certain level of concentration from the user. Nevertheless, we still believe that the advantages of such a system are well worth the effort.

Moreover, the multi-view system requires a network infrastructure to communicate views amongst peers, which increases the costs of deploying such a system. It enables interesting features such as temporal queries, which have not been discussed in this article, but provide added value to the system.

In addition to the study on view transitions presented above, as part of a public scientific demonstration for geoscientists, we selected a number of participants to assess the general usability/acceptance of the multiview system. The day was a cloudless, very bright / sunny: viewing conditions on the used handheld platform (UMPC) were very limited. From the randomly selected 22 participants, 8 had a geosciences background. Without exception, the access to different perspectives on the field was found useful. 3 users would directly like to use the setup/system in its current form, 2 more users would use it after changes (including small software changes).

### 7.1.6 Multi-View Benefits

By using the multi-view system, users can take different views, i.e., foci on the complete data set. The system empowers users to exploit the whole range of visual information seeking activities, while maintaining a close link to the real world. They can overview the dataset in virtual view, with embedded videos of the real world situation where available.

If, at any time, users require a broader overview of the site, they can take control of the camera and navigate the virtual representation of the world using a virtual view. They can zoom-in to remote views, getting a closer eye on the real-world. Hereby, users can share their viewpoints and access remote cameras that are not directly connected to them. Remote views provide an interesting interface to interact not only with the 3D representation, but also the real world. Bringing up data overlays on real world views in an integrated manner, applying visualization parameters and tools of the datasets to get further details from the selected perspective.

When a point of interest is hidden from all the cameras, it is impossible to observe it except in the virtual view. In this case, the nearest user can be instructed to point a device in the desired direction, or to move to get a better view. By communicating and collaboration, the multi-view system improves situation awareness at a low mental effort, a topic further explored in section 8.

In spite of all the advantages of the multi-view system, such as enabling visual information search as required for visualization of large dataset, occlusions continue to pose a challenge. In particular, when observing virtual objects and embedded cameras in AR, it is impossible to ascertain their topology. Consider the case shown in Figure 7.7. The remote camera is behind a barn. However, because the barn is not in the 3D model, it is impossible to compute such information. Furthermore, if the camera was behind the mountain and it was displayed as occluded, it would be impossible to know which object is the occluder (the barn or the mountain or a mountain further away?). As a solution we could send a drone (blimp) or a user to inspect every object in the line of sight until they found the occluder, but a more efficient method exists in the form of VPV: Variable perspective view improves overviews and topology relations.

## 7.2 Combining Perspectives in a Single View

We drafted some prototypes combining perspectives in a single view, as shown in 7.8. These concepts were based on rendering a panorama to extend the video background.

Variable perspective view is an AR visualization technique developed to combine views from different perspectives in a single image. The goal of this combination is to provide a wider overview of the dataset and to allow the discovery of occluded objects

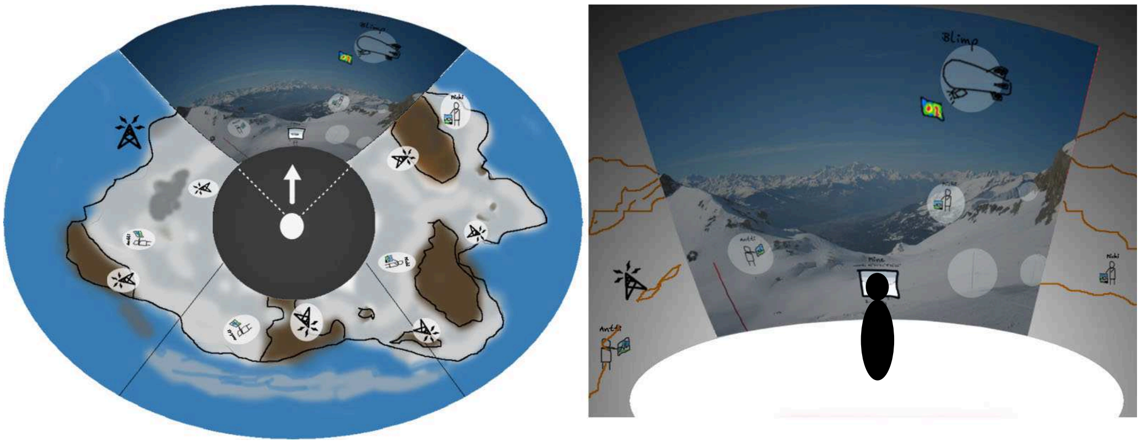


**Figure 7.8:** Extended Overview Concepts. Both concepts relied on rendering the 3D model from multiple viewpoints combining them with the AR view. (Left) renders a panorama centered on the user, and warps it around a cylinder together with the AR view. (Right) renders the AR view from a close exo-perspective, possibly showing a ghost of the user and bending the virtual panorama about the viewpoint. [Kruijff et al., 2009]

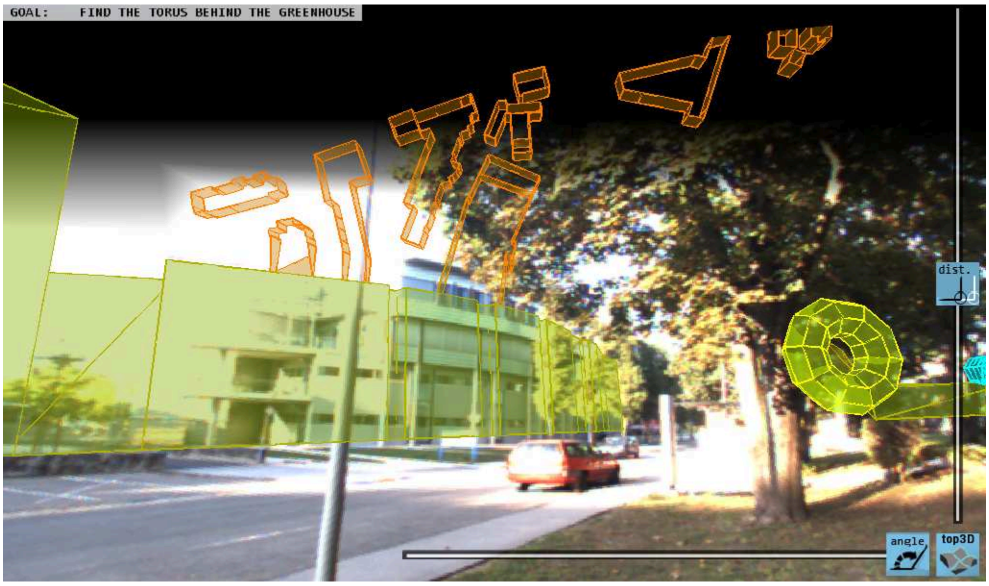
in a simple way, without leaving the egocentric AR view.

Our approach is inspired by multi-perspective techniques for VR [Jobst and Döllner, 2008]. Multi-perspective views rely on non-linear 3D projections, and include several deformation operations to combine multiple viewpoints in a single image [Jobst and Döllner, 2008]. One advantage of multi-perspective views is the increased usage of the screen real-estate to convey spatial context information.

In our case, we aimed to provide a solution tackling the AR aspect (real and virtual content integration) whilst being interactive (changing at runtime deformation parameters) on mobile devices. Additionally, we wish to avoid operations that distort the spatial relationships between objects (e. g., scaling). Finally, the implementation must account for the deformation of different information sources (e. g., 3D model and sensor data).

### 7.2.1 Technique

We developed a variation of the multi-perspective view that we call variable perspective view (VPV) technique. It combines the registered AR content with extra contextual information to extend overview in AR (see Figure 7.9). The VPV combines two virtual cameras: the main camera ($mc$), corresponding to the AR egocentric view, and the secondary camera ($sc$) or far camera. The method applies a skinning algorithm for skeleton animation as shown in Figure 7.10-A. We use a single joint with two bones with the following parameters: $d$ is the distance to the rotation axis (distance from $mc$ to the joint), $\alpha$ is the angle of rotation and $\phi$ is the transition zone (i.e., a volume of interpolation between



GOAL: FIND THE TORUS BEHIND THE GREENHOUSE

**Figure 7.9:** Variable perspective view. The video background is registered with its 3D representation from two perspectives.

$d + \phi/2$, and $d - \phi/2$). Note that the units are those of the rendered model (e.g.,

All vertices in the virtual scene are weighted according to their distance from the main camera to the rotation axis. The weight of vertices defines whether they fall in the view of the main camera, secondary camera or in the transition zone $\phi$, where they are interpolated. To further extend overview, the secondary camera is placed at a distance from the AR view, allowing to capture more information from the digital data, while correctly registering the video for real world context, as shown in Figure 7.9.

All the parameters mentioned above can be controlled at runtime through keyboard or using the graphical user interface (Figure 7.9). Controlling the angle of rotation, the user can alter the perspective of the secondary camera involved in the deformation. This becomes particularly useful to change the amount of overview: Smaller angles allow viewing further, larger angles allow to see behind objects. as shown in Figure 7.10-B.

Varying the distance to the joint provides direct control over what area is visible in the rotated view (see Figure 7.10-C). Control over the transition zone lets users vary between smooth and rough interpolations (see Figure 7.10-D). Smooth transitions provide more visually pleasing results. Rough transitions are useful to overview objects that are packed closely together.

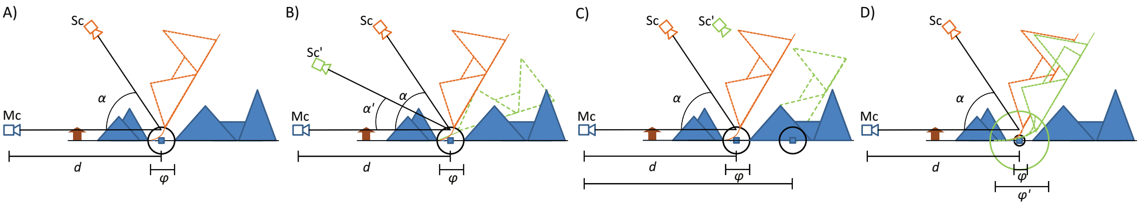


**Figure 7.10:** Interactions with variable perspective view. (A) The secondary camera ($sc$) is rotated by $\phi$ about an axis at distance $d$ from the main camera ($mc$). Vertices that fall within the transition area $\phi$ are interpolated between $mc$ and $sc$ depending on their distance from $mc$. (B) Effect of changing the angle of rotation $\alpha$ to $\alpha'$. The new camera $sc'$ has a better overview. (C) Effect of changing the distance $d$ to the rotation. The new camera $sc'$ (at distance $d'$) observes a portion of the environment further away from $mc$. (D) Effect of changing the size of the transition area $\phi$.

### 7.2.2 Implementation

The VPV was developed with a dedicated OpenGL-based framework. Its computation relies mainly on GLSL Shaders. Firstly, the CPU, in charge of the interactive aspect, computes the model-view transformation for the main and secondary cameras ($M_{mc}$ and $M_{sc}$), and a weight variability $w_\phi$ in the transition zone ,

$$w_\phi = \frac{1}{\phi}.$$

These parameters remain constant and are computed once for all vertices of objects in the scene. For each vertex $v$, a weight $w_v$ is computed – in a vertex shader – from its depth in world coordinates:

$$v_{mc} = M_{mc} \times v, w_v = (v_{mc}.z - d + \phi/2) \times w_\phi,$$

where $w_v$is clamped to a $[0 \dots 1]$ range. Based on $w_v$, the final, world coordinates of the vertex are obtained as follows:

$$v_{out} = \begin{cases} v_{mc} & w_v = 0 \\ (1.0 - w_v) \times v_{mc} + w_v \times v_{sc} & \text{otherwise} \end{cases}.$$

where $v_{sc} = M_{sc} \times v$. Finally, $v_{out}$ is projected; we assume that both cameras have the same intrinsic parameters (i.e., use a single projection matrix). This assumption helps in maintaining a sense of scale across VPV views. Although we have not experimented with this yet, we assume that assigning different projections to each camera will degrade perception of distances and depth.

An initial, exploratory evaluation, helped us identify perceptual issues in the first implementation, in particular regarding how we visually convey the separation between areas (main camera, interpolated, secondary camera) in combination with the view of the real world (background). To address the former, we pass the vertex weight $w_v$ to the corresponding fragment program, whereby the area is identified. We can then apply subtle colour changes to differentiate each area (e.g., color coding, blending), ensuring correct perception of the original colour. This prevents mistaking the effect of the VPV for a real change in the terrain.

To prevent interference between the deformed virtual content and real content, we implemented a masking operation. We calculate a horizon line of the virtual content associated with the near camera (mc), and fade out the video background above this horizon line (see Figure 7.9). This measure is highly dependent on the virtual content, but it insures that parts of the real world that will be ignored (e.g., sky) do not interfere with the VPV.

### 7.2.3 Evaluation

#### 7.2.3.1 Explorative Study

We applied rapid usability testing [Pawson and Greenberg, 2009] to get an initial impression on the technique's usability. During a demo of our overall system carried out at CHI 2011, five experts in the field of HCI and computer graphics were exposed to the technique. Instead of using questionnaires, we relied on informal conversation, as suggested in [Pawson and Greenberg, 2009], guiding the discussion around the topic of interacting to change overview or discover spatial relationships, and collecting suggestions.

As expected, participants showed enthusiasm at being able to control the effects of the technique, and were pleased with its fast response. One of them manifested satisfaction at being able to see the horizon while increasing overview. The user was satisfied to be able to see the whole extent of the area by simply rotating in place and manifested that ”seeing the horizon improved the navigation experience, while observing a larger area”. Another one was enthusiastic at the interaction that allowed him to discover occluded objects. ”It's like a tsunami effect, a wave that carries the objects to the top of the screen”. Notwithstanding these positive comments, some issues caused concern about the applicability of the technique. In particular, users declared that in some cases it is difficult to note the effect of the VPV. It is clear that the terrain is deformed, but it is unclear where the deformation takes place. In addition, the fact that the screen is full with information now poses a challenge in terms of cognitive load required to understand all information. To reduce these effects, we applied different representations for the different areas of the technique on fragment shader program.

#### 7.2.3.2 Comparative Study

We performed a formal study to analyze the effects of the VPV in a search and exploration task. We aimed at assessing usability and performance of the VPV by comparing it with a conventional overview interface: a self-orienting, forward-up map. The experiment followed a within-subject, repeated-measures design with the technique as independent variable (map, VPV), and dependent variables completion time, errors and subjective measure of cognitive load (measured using RTLX questionnaire).

**Tasks.** The study task was to interact with the technique to discover the location of virtual objects. The focus was not only in finding virtual objects, but in assessing whether participants (mentally) establish the spatial relationship between these objects and the real-world. We used colored 3D objects as targets for the search task, as these correspond to search by color pattern. To represent a search and exploration activity, we defined two categories of tasks:

Category 1: Finding a real object, then a virtual object related to it, another virtual object, and finally real one (RVVR). This included subtasks for the user such as:

- Locate the tallest building and find a cyan sphere near it.
- Find a yellow cone near the sphere.
- What building/location is the cone pointing at?

Category 2: Finding a virtual object, then another virtual object related to the first, and finally a real one (VVR). This included subtasks for the user such as:

- Find a purple torus.
- Find a yellow sphere to the left/west of it.
- What building/location is the near the sphere?

**Methodology.** The study took place in an area of approximately 2km$^2$ around our university campus. The 3D scene was composed of virtual representations of buildings and extended with ten to fifteen 3D objects (tori, cones, spheres) with varying colours (cyan, yellow, magenta). The size and colour scheme, as well as the representation of the scene were chosen after a pilot study carried out beforehand during daytime. Thereby, we estimated what general schemes were visible under the lighting conditions. The final representation is a combination of wireframe and fill rendering modes, as shown in Figure 7.9.

We identified two locations around our campus for our two conditions. We defined six trials per location (six scenes with objects at different locations and six different sets of instructions). We randomized the initial location across participants, the conditions across locations, and the order of the subtasks per location.

The test application had two operating modes: normal AR, and overview. The participant could only access one mode at a given time, and was forced to press a button to switch to the other mode. The overview mode was associated to our 2 conditions and consisted of the map during the map condition; or the VPV during the condition of the same name.

We logged the time in each mode, the overall duration of the task, all object selections, and errors when relevant. Before the measured trials, the participant was led to a third location for training. The procedure for the experiment was explained, and thereafter the participant tried the two techniques to get used to the controls and the instructions. Afterwards, the participant was brought to the location for the measured trials of each condition. After each trial, the participant filled the TLX questionnaire, and subsequently a subjective questionnaire on usability. Upon finishing six trials, the participant filled an exit questionnaire on general experience with the technique. Thereafter, she/he was taken to the second location to perform the second condition. Comments from participants were noted down throughout the experiment for post- analysis.

**Apparatus.** The platform for the experiment was Panasonic CF-U1 (screen resolution 1024x600) equipped with an external uEye Camera (800x600, 4.2mm wide angle lens). Location tracking was accomplished using a differential GPS (Ublox AEK-4H), and orientation with an inertial tracker (Intersense InertiaCube 3).

**Participants.** Ten participants were recruited from the university, (9 male, 1 female, average age 27.6 years old). All participants had normal or corrected to normal vision.

**Results.** Six trials out of the 60 were not completed due to unrecoverable tracking errors and removed from analysis. A paired-samples t-test revealed a significant difference in the task duration for map and variable perspective technique, $t(54) = -4.65$, $p < .01$. Participants took significantly longer to complete the task with the variable perspective technique (M = 119.84s) than with a map (M = 73.38). The effect was still present when analyzing results per each task type.

For RVVR, t (27) = -3.6, $p < .01$, completion time in VPV (M = 123.7s) was significantly longer than in map (M = 68.30s). For VVR, t (26) = -2.8, $p < .01$, completion time in vp (M = 115.8s) was significantly larger than in map (M = 78.60s). Besides, we noticed that participants spent most of the time in the overview mode, only switching back to AR for brief periods of time. For the VPV condition, the mean percentage of time in overview (M = 105s) was larger than in AR (M = 15.4s). For the map condition, the mean time in

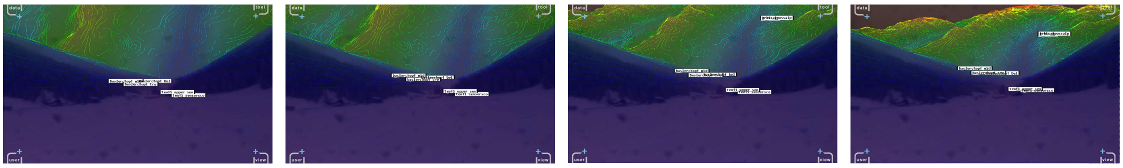


**Figure 7.11:** Increasing overview. Image sequence for increasing overview by changing the angle of rotation $\alpha$.

overview (M = 59.6s) was also larger than in AR (M = 13.7s). A paired-samples t-test revealed a significant difference in time in overview between VPV and map conditions, t (54) = -4.61, $p < .01$. It is worth noting that in VPV, the virtual content is still connected to the real-world, as opposed to the map.

A paired-samples t-test showed a significant difference in the separation between virtual and real for the VPV and the map, t (9) = -2.899, $p < .02$. Participants noted an acute separation between virtual and real in map (M = 2.8) than in VPV (M =4.1s) (in a 7-point Likert scale, lower means content separated). With regards to performance, there was no significant difference in error count across conditions (paired-samples t-test, t (54) = .087, $p > .9$). Differences in errors between the vp (M =1.22) and the map conditions (M =1.25) can be attributed to chance. Analysis of workload measures based on RTLX revealed no statistical difference across conditions (paired-samples t-test, t (54) =-1.04, $p > .9$). Thus, workload for the VPV condition (M =41.6) was not perceived differently than that of using a map (M =39.01). Finally, the VPV was well received. Participants were enthusiastic at trying the interface, and there was a trend towards preferring the VPV (M =1.9, STD=1) over the map (M =2.7, STD=1.2), albeit not significant in a 7-point Likert scale. Three participants commented that they preferred the VPV when they had to discover relations between virtual and real, while other three noted that with the VPV they got a better idea of the orientation of objects.

### 7.2.4 Variable Perspective Benefits

This initial comparative study provides preliminary results on usability and performance of our technique. People took more time using the VPV than with the map. Digital maps are well known and we did not expect to get better quantitative performance with our novel technique (without proceeding to a longitudinal study). We noted a learning effect during the study, and four participants reported it in their comments. Still, the ratings for workload did not differ, which suggests that performing with the VPV was not more demanding than with a map.

Participants spent longer time in map and VPV than in AR. This, we believe, led to the perception participants had of separate sources of information. With the map technique,

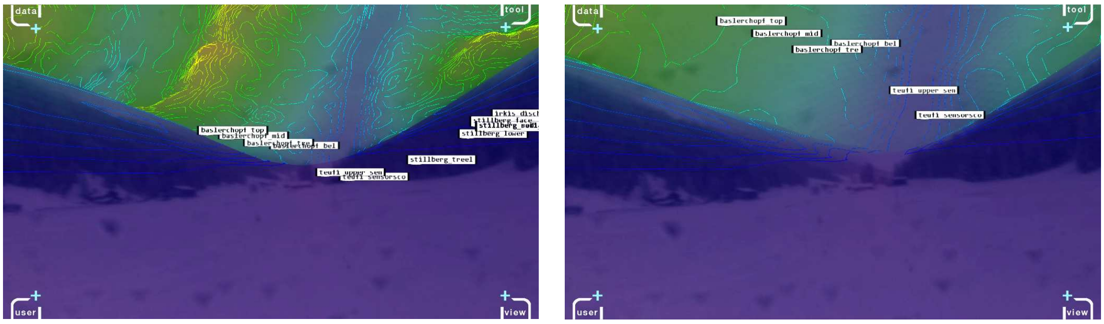


**Figure 7.12:** Unveiling occlusions. Spatial relationships between occluded objects can be discovered by changing the distance d to the rotation.

participants needed to shift to another context. Conversely, the VPV keeps them in a similar AR context even with a slightly higher execution time for the different tasks. Exit questionnaires showed that the VPV gave a significantly higher feeling of integration of virtual with real content, whereas, when performing with the map, participants mostly ignored AR. Kim and Dey [17] showed the advantage of this in tasks involving high levels of attention (e.g., driving).

The study helped identify further perceptual issues that, when addressed, will help the VPV reach its full potential. For example, close, large physical objects can block the VPV. To address the issue, the VPV needs to be extended with other, see-through occlusion management techniques (e.g., vanishing, ghosting).

Beside perceptual issues, the outcomes of the experiment open up several paths for future work. There is a first indication that the VPV can benefit from a more intuitive user interface.

The VPV extends overview for AR applications with the advantage of using the full screen to provide information. The combination with a registered video in AR allows direct access to the real world context. Thereby, users uncover spatial and topological information about the environment, while interacting with the technique to increase overview.

Increasing overview: Overview can be changed dynamically by manipulating the rotation angle and the distance to the rotation. These two parameters allow fine control over how much deformation is applied to the terrain (angle) and where is it applied (distance). Figure 7.11 illustrates the effect of changing the angle of rotation. To the left, the VPV occupies the whole screen, and there are little or no distant landmarks for the user to orient. To the right, the horizon is visible, showing distant landmarks for orientation and a larger portion of the terrain, although occlusions caused by changes in elevation become more prevalent.

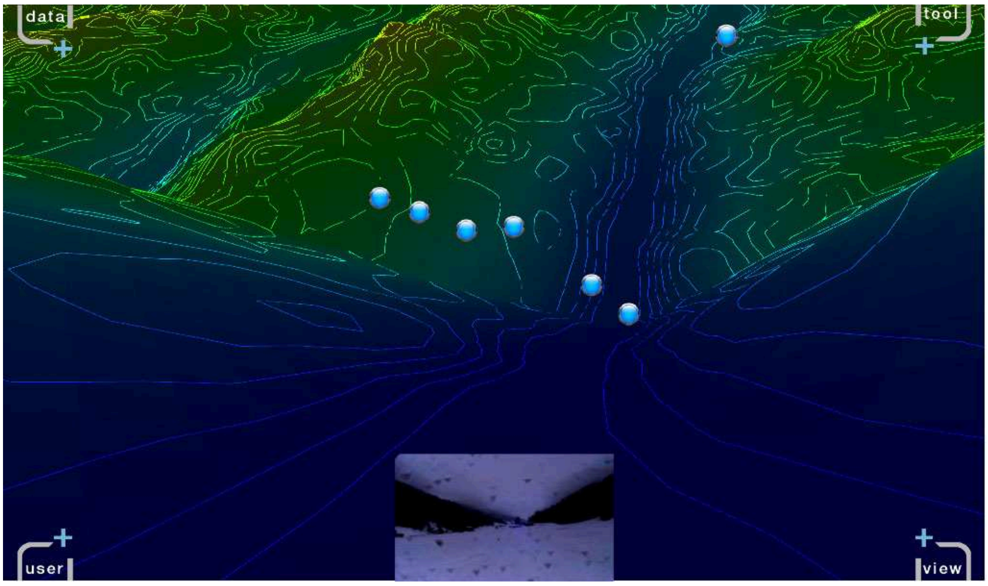


**Figure 7.13:** Detached variable perspective view. The view is detached from the AR context to increase overview over the 3D model.

Unveiling occluded objects: Spatial relationships between occluded objects are discovered interactively by controlling the distance to the rotation, and the size of the transition zone. Figure 7.12 illustrates the interaction to discover the positions of sensors. Note how these sensors appear packed together in the initial case, but become gradually separated as the interaction with the VPV changes the secondary viewpoint.

### 7.2.5 Combining Techniques

A combined version of VPV and multiview provides advantages worth mentioning. The views shared by other users or available from devices deployed in the field can be browsed and selected from the VPV. Thus the VPV provides a form of extended overview to the AR context, and the remote views complement this with zooming possibilities for remote points of interest. But it can also be detached from the AR context, while still keeping it in view. This comes in handy to experience the variable perspective from a vantage point, as depicted in Figure 7.13. The combination of virtual views and variable perspective AR allows the user to experience such view in an AR context.

## 7.3 Summary

In this chapter, we discussed interactive aspects of outdoor visualization. The initial motivation was that of finding metaphors that make the interaction intuitive, whereby users manipulate the dataset, the interface having become unobtrusive. The main challenge ahead is the limitation of AR when it comes to "viewing the situation as a whole" without moving. Our concern was to extend the overview for AR, and provide a situated user with tools to observe large parts of the dataset in relation with the real world context where the data is generated. We develop a range of functions to substitute operations common in a visualization for AR counterparts. The main contribution of this chapter is in describing a combination of user interfaces, methods and metaphors that enable data visualization in outdoor augmented reality.

# Chapter 8

# Sharing Observations in Collaborative Mobile AR

Throughout this thesis we have developed tools for mobile scientific data visualization, and insisted that these are complementary to visualization activities carried out at the office. Furthermore, chapter 3 stated an implicit goal to create a *shared* understanding of the environment for users to discuss potential solutions to problems found.

At this stage, with a better idea of what outdoor AR visualization can accomplish, we turn to the topic of collaboration. The focus, beyond office-site collaboration (or tele-collaboration), falls on letting users work together *regardless of location*, when they are free to move around. Many interesting situations arise with freedom of mobility. Compare the sketches in Figure 8.1 A, where users cooperate by sharing observations on a single device as compared to Figure 8.1 B, where users freely roam about performing different tasks in the same space.

We briefly analyze collaborative situations and their requirements, and discuss how our tools can be exploited to create shared awareness in interdisciplinary teams. Taking our previous discussions about sharing perspectives as starting point, we build a collaborative framework whereby users can interact, communicate and discuss findings and potential solutions.

### 8.0.1 Collaborative Situations of Mobile Environmental Monitoring

The deployment of mobile AR assumes a level of autonomy in the users. When each participant carries a device, it is expected that they can perform on their own. Taking this in consideration, it becomes evident that collaboration opportunities will arise spontaneously. Several mobile AR tasks can trigger spontaneous collaboration, even with minimal support.

**Figure 8.1:** Collaboration Situations. Users discuss observations sharing a display device (Left). Freely moving users inspect different portions of the environment, and share their experiences (Right).

**Acquiring awareness of the environment.** As already indicated, a single user has a limited perception of the deployment area. Still the whole mobile AR setup is about getting an enhanced perception of the physical environment with all its dynamic aspects. The lack of perception is a good enough reason to initiate collaboration with a nearby party. Asking "what do you see over there" to a peer nearby can help a user acquire extra information about the environment without moving. The collaboration can stop with the answer, or it might trigger a close session, where users discuss properties of that what they experience from different perspectives. Consider two users in a valley. One of them might have climbed a hill to get a vantage overview. The other one might ask her for overview information, while she might be interested in details that are accessible to the user down the valley.

**Planning for movement.** Any activity elicited through outdoor AR can surely benefit from different levels of coordination. If users can act autonomously and have the same or interchangeable objects of interest at different locations, the most efficient way to coordinate them is to assign each user a set of objects or locations that do not overlap. A campaign can be previously planned and then executed. However, dynamic conditions can cause one or more users to deviate, or to change their routes. This task offers numerous opportunities for collaboration. A coordinating entity will notice such changes, and initiate re-planning. Conversely, if users know the initial plan, they can communicate changes and attempt to reorganize. The success of the task depends on different pieces of information. Consider the case where a person at the office performs a routine sanity check on previously deployed sensors, while users in the field are performing their task. The office user might detect a malfunctioning unit in the surroundings of a new deployment. If the office user knows that there is a team at the site, she can request a member to deviate and check the malfunctioning unit, instead of planning a new campaign.

At this point, it is clear that the nature of mobile AR introduces numerous possibilities for collaboration, in particular spontaneous collaboration. A real example can better illustrate how real the need for collaboration in mobile monitoring can be.

> In summer 2008, cantonal administration received an extreme rainfall warning from MeteoSuisse for the Dranses catchment.[. . . ] During the event, cantonal experts [assembled at the headquarters in Sion] analyzed data coming from sensing stations in the Dranses catchment. They requested on-site observers through mobile or radio to describe the situation, informed local municipalities and tried to set up a coordinative effort. However, they were "blind" and could only consult the workplace data of the area. Cantonal experts asked the person on-site to give his/her opinion and to take some measures in order to quantify risks. [. . . ] Cantonal experts in conjunction with local experts and municipalities sent guidelines and directives for the on-site observer to coordinate onsite actions. [Veas et al., 2009]

The tasks above indicate a diversity of information that needs to be available to support collaboration. These requirements have different connotations depending on the proximity of the users. In some cases, the information is readily available (e.g., when in close proximity), while in others, attaining such information entails a possibly cognitive, but also physical effort. To successfully support collaboration, an application must be designed to provide such information on demand, that the effort needed to collaborate is reduced and it warrants a benefit for the parties involved.

### 8.0.2 Collaborative Interfaces

Collaborative tools in our outdoor AR system cover multiple one to one on-site collaboration, on-site to off-site collaboration (both being asynchronous or synchronous).

The shared view service introduced in section 3.3.4 and its associated interface discussed in section 7.1 offer a shared awareness model between participants of an interdisciplinary team. Users are aware of other users, their view-point and can even "peek" at the content of their view (i.e. device's screen).

These features are extended with functions for communication-centered tasks. Users can toggle small overlaid presentations of other users to locate the whereabouts of team members, and they can initiate communication sessions using pre-configured voice channels. The latter is particularly useful when users collaborate in distributed fashion with mobile or remote peers. The combination of view sharing with the voice communication channel helps to generate a shared awareness for this kind of task.

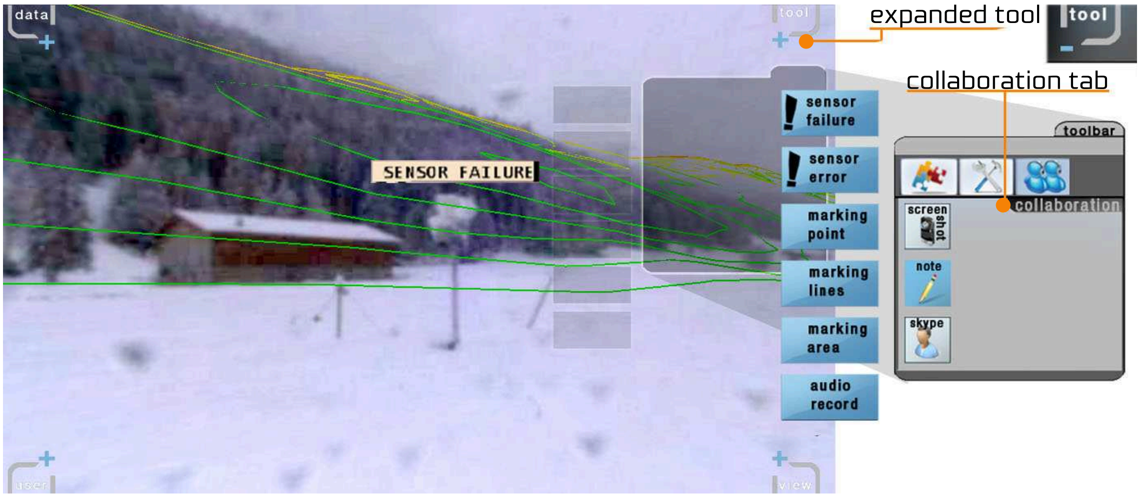


**Figure 8.2:** Collaboration Interface. The collaboration tab activates mechanisms to convey observations to peers, leaving annotations, or initiating real-time communication.

Asynchronous collaboration methods are built around the notion of geo-referenced annotations, possibly containing graphical content (e.g., a screenshot). They encompass several ways of noting down remarks, by using semantic text-based annotations, voice notes and marking tools. Semantic annotations are ordered logically according to the application at hand – examples include sensor failures, or specific observatory findings. A set of simple marking tools can be used to draw upon screenshots taken by the user. An example of a marking task is taking a picture of a snow profile, and marking the various layers that can be observed.

### 8.0.3 Discussion: Situation Awareness Revisited

The collaboration toolset proposes, from an AR perspective, novel ways to share discoveries and build the understanding of an environmental phenomenon or process. During evaluations (see section 3.6), users continuously expressed their enthusiasm and gave positive feedback towards these tools. Nevertheless, in light of the situations introduced in 8.0.1, these tools can only capture a partial model of the information needed. This section establishes a model to evaluate the functionality of the collaboration toolset, and identify areas for future extension

For successful collaboration to occur, users need to be aware of other users and the workspace [Gutwin and Greenberg, 2002]. In outdoor AR, the workspace includes the portion of the real world and the abstract (virtual) information associated with it. Spatial knowledge occupies a preference position among the requirements of awareness. Other aspects are subsumed to the knowledge of what place they occupy, how their dynamics affect the disposition of space, and how actions (moving) affect such relationships.

Gutwin and Greenberg [Greenberg et al., 1996] define knowledge that is relevant to maintain awareness in a shared workspace for the purposes of successful collaboration. This information identifies knowledge about others, their activities and where they are carried out.

#### 8.0.3.1 Viewpoint Awareness

For mobile AR, knowledge about others is relative to where they are in relation to one another, and possibly to objects of interest. Since AR is predominantly view oriented, part of understanding the other's location requires to know where and what they can see (e.g., in what part of the world they can act).

The shared view system addresses the requirements of viewpoint awareness, by sharing key aspects of the user's camera. The pose of the camera determines location awareness (i.e., where is a user working). Furthermore, the camera intrinsic parameters define the viewport – the portion of the world – that the camera can capture. These, together with the camera frames define gaze awareness (i.e., where they are looking) and view awareness (i.e., where they can see). Hereby viewpoint sharing entails sharing tracking information as well as intrinsic parameters of a camera, the id of the person using it, and the video frames. This strictly refers to what a user can see on the AR screen, and disregards what they can see with their own eyes.

#### 8.0.3.2 Feedthrough Awareness

In mobile AR, objects of interest (OOI) have particular properties. From an awareness stance, a user needs to know where is what and which qualities it has: Is it real? Virtual? Is it a real object with a virtual representation?

For collaboration, several relationships between users and OOI become important. The relation of proximity between the other party and an OOI answers to questions such as: Is anyone near an object of my interest? Can another user reach it? Manipulate it? Conveying what objects or parts of the workspace each user is interested in acts as a facilitator for informal collaboration. Furthermore, awareness of objects, where they are, and how others see them, also aids in referencing during collaboration.

In a broad sense, annotations can cover how users perceive certain objects or document experiences at given locations. Still, the current implementation does not identify OOI nor the possible actions at different locations.

#### 8.0.3.3 Activity Awareness

An important part of workspace awareness lies in knowing what users are working on. Viewing, the most common interaction in AR, has been partially addressed in viewpoint awareness. Authoring applications particularly benefit from representing users activities that modify virtual objects (e.g., adding annotations). However, it becomes rather difficult to represent tasks where the user has to interact with the physical world in a way that the AR application cannot track (e.g., fixing a sensor, gathering samples). In such cases, we need to rely on the users themselves to report status.

#### 8.0.3.4 Intentional Awareness

Beside knowing what is going on now in the shared workspace and what has happened before, it is interesting to consider what is going to happen, at least regarding user activities. The collaborating group can benefit from knowing what each participant intends to do [Pinelle and Gutwin, 2005]. This piece of information is very task specific, but we can still refer to tasks inherent to mobile AR (moving and viewing). A user that needs to visit areas of interest separated from one another will have to move sequentially to each of them. In turn, the user will consider her options, decide on the next waypoint, and move there, possibly choosing a route. Other users may benefit from knowing this, some of them may want to get a look from that perspective, and ask the user to share her viewpoint, others may have an OOI in the way and ask the user to prompt when approaching that object.

Sharing information about movement plans requires a minimum of one point per user: the next waypoint that the user intends to reach. This can be communicated with annotations. Of course, a richer representation, such as a sequential list of points, or a route, is more helpful. One important consideration is that this information is highly unreliable. A user might need to change plans, removing waypoints in her route, re-routing, etc. and other users will need to be notified and act accordingly. A notification system integrating information about points of interest and movement plans can keep the users aware of interesting activities and act as enabler for collaboration. A user might request a notification if another user is close to point A, and initiate collaboration when such a thing happens.

## 8.1 Summary

This chapter has taken a theoretic perspective, and analyzed the requirements for successful collaboration in outdoor AR. Taking the shared view service described in the previous chapter as a backbone, the toolset was extended to manage synchronous and asynchronous collaboration. Thereby, we considered the aspects of collective interaction and how they affect the workflow of a single user. First hand, empirical study of the toolset guided a theoretical analysis of awareness aspects in outdoor AR, in particular affected by mobility. We validated the compliance of our toolset with this awareness model.

# Chapter 9

# Conclusion

This dissertation investigated the concept of outdoor AR as a paradigm for mobile scientific visualization. The research was motivated by the disjoint picture presented by data from a myriad of sensors and numerical processes with various update rates, that force the scientist to switch between representations (plots, visualizations, numbers) and correlate with the real world (maps, models, photographs) to make sense of the situation. Instead, this dissertation proposes to create a seamlessly integrated information space, whereby scientific data are perceived in the context of their occurrence: the physical world.

The scientific area of study was that of environmental monitoring, where sources of data represent one aspect of the physical world. Thus, their geographic anchor simplifies the mapping to the real world. Nevertheless, the combination of data with the real world must confront several challenges endogenous to the task (i.e., data visualization), situation (i.e., outdoors), and technology (i.e., mobile devices, AR), that drive the contributions of this thesis.

## 9.1 Contributions

Several contributions can be elicited from the methods presented in this dissertation. Major contributions define research directions and results described in chapters 3 to 7, whereas, in our case, minor contributions are outcomes that can be taken individually as starting point or building blocks for other work. In this section we highlight contributions and describe their limitations with the goal of suggesting future improvements.

### 9.1.1 General Contributions

In chapter 3 we introduced **mobile environmental monitoring**, analyzed potential deployments, and proposed an infrastructure that reflects the lifecycle of data, from acquisition through various stages of processing to mobile visualization. Our infrastructure for mobile environmental monitoring encompasses real-time access to sensor data, results of numerical processes and the physical world, while providing dedicated tools for visualization and collaboration.

With data ubiquitously available in real-time, we turned our efforts towards **outdoor AR environmental data visualization**. Throughout this thesis, we demonstrated a suite of interaction techniques and metaphors that facilitate visualization activities from within the data. The user's awareness is enhanced with perception of correspondence between the data and the surroundings in a way that allows the exploration of the relations between them. Thus, we endow a situated user with the ability to experience the information in relation to the surroundings. Thereby, understanding and insight are elicited through kinesthetic feedback gained by peering around at data from within, taking alternative perspectives or interacting with it.

The proposed methods have their limitations and are not meant to replace traditional data analysis, rather they are complementary, in that they supply near real-time localized views on the situation. Thereby, our tools facilitate awareness of the physical world when observing data, and vice versa, thus enhancing the overall monitoring workflow with timely experience of the physical world.

For validation, we performed a series of qualitative evaluations with experts in different scientific areas and varied scenarios, that iteratively shaped the concepts and tools presented (summarized in chapter 3). The results helped to define situations that benefit from outdoor visualization, and also its limitations in certain deployments. For many of the interviewed experts, 3D visualization and AR constitute novel paradigms and it is reasonable to expect that they need time to incorporate the technology in their scientific workflow. As this happens, it will trigger a wealth of opportunities to explore new functionality and improve the existing toolset.

Major limitations for the current deployment of mobile AR visualizations can be attributed to physical limitations of the platform. The form factor and weight of the device were a constant disadvantage in some occasions. Furthermore, as high dynamic range conditions are frequent in outdoor environments, lighting conditions vary widely to extents that render the screen almost unreadable.

Taking a human factors perspective, some of the limitations in applicability can be derived from the technology used. Whereas handheld AR is ideal for its relatively unhindered deployment and social acceptance, the handheld AR paradigm, as opposed to immersive AR, has the disadvantage of requiring complete attention, it keeps the hands of the user on the device. In contrast, field work often entails manipulation of tools, manual sensors, samples, and so on. In these cases, the proposed solution falls short and cannot be employed to complement the user's activities. Therefore, one must bear in mind the expected duration of interaction on-site in the face of each situation when considering a possible deployment.

In spite of the shortcomings, experts were enthusiastic to explore the possibilities that this paradigm brings about, and suggested several possible activities and scenarios where the technology is well applicable. Of particular interest are situations where timely data could reduce work, for example in verifying or installing a sensor. Even more interesting are situations where experiencing real-time information on site potentially improves understanding of a process. Consider the effects of quickly changing structural characteristics, such as during a flooding. Their interpretation depends on seeing the representation of processes in direct relation to the actual state of the environment. Along the same line, the verification of numerical models can benefit considerably from real-time, real world experience. For example, when timely updates can be guaranteed, executing a model with real-time data while experiencing its results within the real world context gives a better insight on the variability of parameters.

Overall, environmental monitoring with outdoor AR is a promising area of research that opens many possibilities for experimentation. We expect that it will mature in upcoming years and position itself amongst the fundamental techniques of environmentally aimed scientific inquiries.

### 9.1.2 Individual Contributions

The methods and techniques that drive our outdoor AR visualization toolset are based on a number of individual contributions. Each of these contributions were framed in the mobile AR visualization context, but they were evaluated in isolation, prior to their introduction in the toolset. Here we summarize the contributions and limitations in each case.

#### 9.1.2.1 Handheld Platforms for Outdoor AR

In chapter 4 we described the evolution of a handheld platform to support outdoor AR. The motivation was the ergonomic extension of a mobile computer with high end sensors as a basis for the AR experience. Though we successfully integrated numerous sensors in a modular, ergonomic platform, it was still highlighted as one of the limitations of the overall approach. In extreme outdoor scenarios, where the user can carry a limited weight and has limited space, a device needs to be as light and slim as possible, and users were not forgiving in the light of slim tablet PCs appearing in the market.

In this light, our contribution is limited to the ergonomic aspects that make a device of such characteristics usable: modularity, weight balance, control placement. With progress in technology, it is expected that other factors would be solved in production efforts. Still, until manufacturers start to prioritize high quality position sensors (GPS, orientation) that simplify the deployment of AR applications, researchers have to find alternative solutions.

AR researchers have turned their efforts to investigate vision based methods to supply proper registration. However, for general outdoor AR, the results of pure vision methods are quite limited, and the safest approach is still the use of combined GPS and orientation sensors, much the same as in the past 15 years. From this perspective, our contributions are still valid in that they provide a set of guidelines on how to devise a platform that can be gripped comfortably, where to position controls so that they can be reached while interacting, and how to avoid common pitfalls while doing it.

#### 9.1.2.2 Saliency Modulation for Subtle Visualization

The major contribution in our work with this technique is in validating that it has significant influence beyond low level attentional responses and it can be used in videos. Previous works had addressed saliency and utilized different approaches (even an early version reported by Mendez [Mendez et al., 2010]) to modify images to the end of guiding stimulus-based attention or to add artificial objects that change the saliency map in predictable way). In our case, we evaluated the saliency modulation technique developed by Mendez [Mendez et al., 2010] to modify videos so as to influence recall of selected objects.

Our interest in influencing memory was motivated by the fact that working memory is a central construct of the cognitive process [Baddeley, 1992]. In this frame, our contribution lies in the suggestion of preattentive targets to the cognitive process, and the realization that the modification influences processes (memory) at the cognitive level. The results of our studies validate the SMT as an alternative means to convey information to the user, suggesting attention shifts and influencing recall of selected regions without perceptible changes to visual input. These results represent fundamental research and, by no means, cover all the requirements of our motivating scenarios. Still, our experiments address two processes common in HCI: stimulus-based attention guides the user in the exploration of visual input, playing an important role in tasks such as visual search. Memory is involved in user tasks at several stages (e.g., navigation and visual search).

While other approaches to draw attention or influence memory exist, most lack the subtlety of the SMT. The SMT presents an alternative means of attention direction by modifying existing features of the real world image, instead of adding traditional augmentations (such as pointing arrows or frames). The SMT enables mediated reality, since its premise is modifying the existing video input instead of adding virtual artifacts to it.

One advantage inherent to this approach is that it protects context. While the saliency of the context is diminished as that of the focus is increased, the context does not suffer any other degradation. Perceptual issues arising from visual clutter or differences in depth between virtual and real objects are also prevented. We believe that the results provide sufficient evidence to justify further experimentation in tasks that better match real-world conditions.

#### 9.1.2.3 Multi-view systems

We have proposed a set of techniques to support data visualization in outdoor AR applications. Our initial concern was to extend the overview for AR, and provide a situated user with tools to observe large parts of the dataset in relation with the real world context where the data is generated. Our solution was to complement AR with multiple perspectives on the site. We resorted to a mixed approach, whereby virtual cameras facilitate overview of the dataset, notwithstanding the loss of the real-world context; while real imaging devices deployed in the environment provide sufficient real-world context and supplement the visualization workflow with zooming possibilities.

Several contributions can be elicited from our work in multiview systems for visualization. First, the identification of a hybrid system of multiple views that builds in a rather ad-hoc manner on virtual views, mobile users and drones. The second contribution lies in the identification of human factors related to interpreting multiple dynamic perspectives for usage in AR. Finally, we contributed studies on transition techniques to navigate remote views that enhance awareness while maintaining low workload costs.

The multiview approach has the shortcoming that it requires a basic network infrastructure, albeit minimal, to discover and update views from remote cameras. This infrastructure needs to be designed with clear notions of the number of users and scalability, as each supplier will submit video streams or frames, and each consumer might be interested and therefore subscribe to multiple suppliers.

The multiview system was well rated in discussions with experts, and received positive criticism in the form of suggestions. Besides enhancing AR visualization workflow, this system forms the basis for several interesting contributions such as collaboration and tele AR (see the discussion below).

#### 9.1.2.4 Combining Multiple Perspectives in AR

Following the motivation on the multiview system, we focused on developing a technique to enhance overview of AR within AR, i.e, without resorting to leaving the AR context. The main contribution is an interactive technique that extends overview by composing two perspectives on the data into the AR view. The main perspective is related to the AR egocentric camera, and the second one to an overview camera partially controllable by the user.

By controlling the secondary camera, the user can change the amount of overview or discover the topology of multiple occluding virtual objects in the environment. The main experience is as if the environment bent upwards at a certain distance form the user. We specified a number of alternative representations to convey the relations between the different perspectives.

In our evaluation, we found that although users performed slower in comparison with a map for the proposed task, they experienced the virtual and real as a unified environment using the variable perspective technique. We contend that this technique requires a more intuitive interface to work efficiently in different situations. The fact that participants experienced an undivided information space with the real world complemented by virtual objects from multiple perspectives encourages further experimentation, particularly in cases where switching contexts leads to loss in performance, as those suggested by Kim and Dey [Kim and Dey, 2009].

## 9.2 Outlook

After revisiting the contributions of this dissertation, this section explores future research directions. It is our hope that the results here presented motivate the reader to engage in new research, participate and contribute new methods, and enrich the overall mobile experience.

### 9.2.1 Mobile Environmental Monitoring

Some future directions for mobile environmental monitoring have already been hinted at. It is possible to take the individual activities and scenarios proposed in section 9.1.1 and pursue specialized solutions. For example, one option is to develop a modeling plugin and interfaces to experiment with model parameters on-site, then validate the contribution of that solution.

However, one of the most interesting follow up ideas derived from discussion with expert users was the extension of mobile sensing capabilities. In a nutshell, the idea is to define an interface whereby one can connect arbitrary sensors to the mobile device, maybe following an approach such as [Aharony et al., 2011], and enable sensing at arbitrary locations. From there, the system can be handed to expert users to explore novel scenarios and overall usability. Or, it can be developed as a social surveying tool, whereby citizens contribute with sensing of their environment to a general awareness of its condition, extending the work done by Kim and Paulos [Kim and Paulos, 2009] for a single office in a building to a full social environmental monitoring experience.

### 9.2.2 Mobile Visualization and AR

In their discussion about challenges for visual analytics, Thomas and Cook argue that new methods are needed for an analyst to examine and make sense of a massive, multi-dimensional, multi-source, time-varying information stream, and that we need to use every kind of display available [Thomas and Cook, 2005]. This dissertation is a contribution in that direction.

The visualization efforts in this dissertation were directed towards integrating information with the real world represented by frames from a video. The only major attempt at creating multivariate representations was using the SMT to modulate overlays. Still, our visualizations are multivariate in that they represent the real world, 2D/3D data as overlays, and 1D data in labels. Our overlays use isolines to convey 3D shape and colour mapping for sensor information.

Building on our outdoor AR application and the infrastructure to deliver data in real-time, interesting follow up steps are to explore how, if at all, multivariate visualizations integrate with the real world. In the information visualization field, much work has been dedicated to exploring dimensions to represent multiple values in a single picture, such as glyph density, size and orientation [Acevedo, 2007] or texture noise density, orientation and colour [Khlebnikov et al., 2012]. A full study of visual semiotics focusing on AR and the integration of data would indeed result in interesting contributions to the community. Similarly, exploring how to efficiently represent multiple readings per label, as described in the sketches of section 5.1, would without doubt open up numerous improvements.

The fields of application for mobile visualization and AR are not limited to environmental monitoring or outdoor situations. In general, wherever spatial data sets are the basis for analysis, mobile visualization can be utilized. Consider the case of presenting ambient information in buildings as described by [Rauhala et al., 2006] [Goldsmith et al., 2008]. These relatively simple systems could be extended to provide information ubiquitously within the building infrastructure, much in the sense of [Hailemariam et al., 2010] but using AR. Furthermore, as consumer level devices are targeted, visualizations can be geared towards the general public. This will fuel novel approaches, as the information the general public is interested in spans a wider range, from power consumption, to social interactions.

### 9.2.3 Saliency Modulation Use Cases

The SMT opens up several directions of research. One idea for mobile devices is to use the SMT as an aid to navigation. We would like to suggest objects related to landmarks and explore whether a navigator recalls having seen them along a path. Our results foster experimentation in this direction. Similarly, several new AR applications make use of panoramas for tracking or even for interaction [Langlotz et al., 2011]. The SMT could be used to emphasize landmarks or objects of interest in a panorama itself, without adding virtual highlighting objects that would hinder visualization of the panorama features.

Furthermore, we are aware that the SMT has applications beyond AR; for example, in training, the SMT could be used to suggest that a trainee shift attention towards areas of interest in a scene. A surgeon during training surgery might be reminded of sensitive organs near the work area without visually overlaying any information on the video feed. By varying the modulation thresholds, one could even support using more subtle levels for advanced trainees. Alternatively, physicians following a procedure in realtime could each have the SMT applied to different aspects, depending on a user profile.

### 9.2.4 Tele AR

The multiview system bred interesting follow up directions. Experts expressed that AR also merits use in the office, in a form of “tele AR”, where the office user would access video from the site, possibly using an extended version of the multiview system, and deploy visualizations. The idea is that by recording the environment, either automatically or in previous on-site operations, full data analysis can still be done in the comfort of the office without loosing access to the real world context, albeit with limited access restricted to camera footage, e.g., in an approach similar to Paczkowski et. al [Paczkowski et al., 2011].

Exploring this research direction implies taking full consideration of the temporal relationships between data and the environment and representing them during analysis. Furthermore, it would be worth investigating what kind of data would be required and how to incorporate live video (e.g., panorama recordings), and if it would be possible to mimic the AR experience remotely, with the intricate details of locomotion, for example in an immersive setup such as a CAVE. Note that the notion of tele AR does not necessarily imply collaboration, a user could rely on a remote camera recording frames, or on data freely submitted by the general public, and use it to overlay information.

### 9.2.5 Collaboration with Mobile Users

The multiview system also served as initial building block for the collaboration tools described in section 8. Although these tools afford more than simple communications and cover certain important aspects, we have only started to scratch the surface of the possibilities behind mobile collaboration.

The discussion in section 8 is already a starting point to describe a model for the collaborative situations of mobility. We seek to unveil what information is needed, how we can facilitate it, but in particular how can we provide information that serves as vehicle for collaboration. Freedom of mobility and autonomy create intricate possibilities worth investigating. In particular, we are interested in modeling the information necessary for successful collaboration in the face of mobility and autonomy. And, given a model of the different situations, how can we create a system that gracefully degrades/upgrades to provide the information that is needed and unavailable in the environment, such as when user's move away from one another and they cannot visually perceive where each other is looking.

Consider an emergency situation involving teams and teams of teams, which requires a coordinated and informed response. With regards to coordination, it is crucial to gather/distribute information about actors and their involvement in the situation. Actors involved in an emergency situation can be categorized based on different criteria (e.g., location, affiliation, roles, etc.), which will probably change during the emergency. To enable interaction between all parties involved, it becomes necessary to know first who is involved, where they are operating, what actions they are undertaking and with what goals. This information is a part of our tentative model from section 8. The goals of collaboration-awareness include enabling teams to understand each other's actions, identifying opportunities for cooperation, and facilitating coordination.

The mechanics of collaboration in the face of mobility and autonomy present a fascinating topic of research. The challenges and opportunities can be adapted to different scenarios that require collaboration. Investigating their peculiarities promises a rewarding experience.

### 9.2.6 Extending Information Transfer by Means of Deformations

The variable perspective view provides overview in a composition of different perspectives that results in perceiving a bent environment, deformed at a distance from the user where it bends upwards by a certain angle, both parameters controllable by the user. We seek to explore a generalization of such approach, by studying how the 3D data can be deformed to explore hidden information with minimized movement. We can start by considering object centric operations versus environment centric operations. The former apply to a particular object and maintain the environment relatively static. The latter deform the environment and its effects extend to all objects in the area of effect. Still, the operations that are available in each case, and how they affect perception, i.e., when does coherence breaks, are questions all open for future study.

## 9.3 Summary

This chapter has summarized the contributions of this dissertation in the light of discoveries made throughout the chapters. The overview is intended to set a frame for the discussion of possible future directions of research that we have identified while working and developing the individual contributions. It is a humble thank you to the reader for bearing with the author till this stage, in the hope of inducing motivation and generating fruitful ideas for future research.

# Bibliography

[Aberer et al., 2007] Aberer, K., Alonso, G., Barrenetxea, G., Beutel, J., Bovay, J., Dubois-Ferrière, H., Kossmann, D., Parlange, M., Thiele, L., and Vetterli, M. (2007). Infrastructures for a Smart Earth - The Swiss NCCR-MICS initiative -. *Praxis der Informationsverarbeitung und Kommunikation*, 30(1):20–25.

[Acevedo, 2007] Acevedo, D. (2007). *A Framework for the Perceptual Optimization of Multivalued Multilayered 2D Scientific Visualization Methods*. PhD thesis, Computer Science Department, Brown University.

[Aharony et al., 2011] Aharony, N., Pan, W., Ip, C., Khayal, I., and Pentland, A. (2011). Social fmri: Investigating and shaping social mechanisms in the real world. *Pervasive and Mobile Computing*, 7(6):643–659.

[Anderson et al., 2011] Anderson, E., Potter, K., Matzen, L., Shepherd, J., Preston, G., and Silva, C. (2011). A user study of visualization effectiveness using eeg and cognitive load. In Hauser, H., Pfister, H., and van Wijk, J., editors, *Eurographics/IEEE Symposium on Visualization 2011*, volume 30. Awarded 2nd Best Paper!

[Anisimov, 2001] Anisimov, O. (2001). Climate change, degradation of permafrost, and hazards to infrastructure in the circumpolar arctic. In *American Geophysical Union, Fall Meeting 2001*.

[Avery et al., 2007] Avery, B., Piekarski, W., and Thomas, B. H. (2007). Visualizing Occluded Physical Objects in Unfamiliar Outdoor Augmented Reality Environments. *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 1–2.

[Awh et al., 2006] Awh, E., Vogel, E., and Oh, S. (2006). Interactions between attention and working memory. *Neuroscience*, 139(1):201–208.

[Azuma et al., 2001] Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., and MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, 21(6):34–47.

[Baddeley, 1992] Baddeley, A. (1992). Working memory: The interface between memory and cognition. *J. Cognitive Neuroscience*, 4(3):281–288.

[Baggi and Schweizer, 2009] Baggi, S. and Schweizer, J. (2009). First results on characteristics of wet snow avalanche activity in a high alpine valley. *Natural Hazards*, 50(1):97–108.

[Baillot et al., 2001] Baillot, Y., Brown, D., and Julier, S. (2001). Authoring of Physical Models Using Mobile Computers. In *Proceedings of the 5th IEEE International Symposium on Wearable Computers*, ISWC '01, pages 39—-, Washington, DC, USA. IEEE Computer Society.

[Bane and Höllerer, 2004] Bane, R. and Höllerer, T. (2004). Interactive Tools for Virtual X-Ray Vision in Mobile Augmented Reality. In *IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 231–239. IEEE Computer Society.

[Barrenetxea et al., 2008a] Barrenetxea, G., Ingelrest, F., Lu, Y. M., and Vetterli, M. (2008a). Assessing the challenges of environmental signal processing through the sensorscope project. In *ICASSP*, pages 5149–5152.

[Barrenetxea et al., 2008b] Barrenetxea, G., Ingelrest, F., Schaefer, G., Vetterli, M., Couach, O., and Parlange, M. (2008b). Sensorscope: Out-of-the-box environmental monitoring. In *Proceedings of the 7th International Conference on Information Processing in Sensor Networks, IPSN 2008, St. Louis, Missouri, USA, April 22-24, 2008*, pages 332–343. IEEE Computer Society.

[Bell et al., 2001] Bell, B., Feiner, S. K., and Hoellerer, T. (2001). View management for virtual and augmented reality. In *ACM Symposium on User Interface Software and Technology*, pages 101–110, Orlando, USA.

[Berg and Itti, 2008] Berg, D. and Itti, L. (2008). Memory, eye position and computed saliency. *Journal of Vision*, 8(6):1164.

[Beyer and Holtzblatt, 1998] Beyer, H. and Holtzblatt, K. (1998). *Contextual Design: Defining Customer-Centered Systems*. Morgan Kaufmann.

[Billinghurst and Thomas, 2011] Billinghurst, M. and Thomas, B. (2011). Mobile Collaborative Augmented Reality. In L. Alem, W. H., editor, *Recent Trends of Mobile Collaborative Augmented Reality Systems*, pages 1–19. Springer.

[Biocca et al., 2006] Biocca, F., Tang, A., Owen, C. B., and Xiao, F. (2006). Attention funnel: omnidirectional 3d cursor for mobile augmented reality platforms. In *CHI*, pages 1115–1122.

[Bogue, 2008] Bogue, R. (2008). Environmental sensing: strategies, technologies and applications. *Sensor Review*, 28(4):275–282.

[Bowman et al., 2004] Bowman, D. A., Kruijff, E., LaViola, J. J., and Poupyrev, I. (2004). *3D User Interfaces: Theory and Practice*. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA.

[Burns et al., 2005] Burns, M., Klawe, J., Rusinkiewicz, S., Finkelstein, A., and DeCarlo, D. (2005). Line drawings from volume data. *ACM Transactions on Graphics TOG*, 24(3):512.

[Caudell and Mizell, 1992] Caudell, T. P. and Mizell, D. W. (1992). Augmented reality: an application of heads-up display technology to manual manufacturing processes. In *Hawaii International Conference on System Sciences*, pages 659–669, Hawaii. IEEE Computer Society.

[Chong and Kumar, 2003] Chong, C.-Y. and Kumar, S. P. (2003). Sensor networks: evolution, opportunities, and challenges. In *Proceedings of the IEEE*, pages 1247–1256.

[Chun and Turk-Browne, 2007] Chun, M. M. and Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current opinion in neurobiology*, 17(2):177–184.

[Claros et al., 2007] Claros, D., De Haro, M., Domínguez, M., De Trazegnies, C., Urdiales, C., and Sandoval, F. (2007). Augmented reality visualization interface for biometric wireless sensor networks. In *Proceedings of the 9th international work conference on Artificial neural networks*, IWANN'07, pages 1074–1081, Berlin, Heidelberg. Springer-Verlag.

[Culler et al., 2005] Culler, D., Dutta, P., Ee, C. T., Fonseca, R., Hui, J., Levis, P., Polastre, J., Shenker, S., Stoica, I., Tolle, G., and Zhao, J. (2005). Towards a sensor network architecture: Lowering the waistline. In *Tenth Workshop on Hot Topics in Operating Systems (HotOS X)*.

[Dam et al., 2000] Dam, A. v., Forsberg, A. S., Laidlaw, D. H., LaViola, J. J., and Simpson, R. M. (2000). Immersive vr for scientific visualization: A progress report. *IEEE Comput. Graph. Appl.*, 20(6):26–52.

[Easton and Sholl, 1995] Easton, R. D. and Sholl, M. J. (1995). Object-array structure, frames of reference, and retrieval of spatial knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(2):483–500.

[Elmqvist and Tsigas, 2007] Elmqvist, N. and Tsigas, P. (2007). A Taxonomy of 3D Occlusion Management Techniques. *2007 IEEE Virtual Reality Conference*, pages 51–58.

[Endsley, 1988] Endsley, M. R. (1988). Situation awareness global assessment technique (SAGAT). In *National Aerospace and Electronic Conference (NAECON)*, Dayton, OH, USA.

[Endsley, 1995] Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors The Journal of the Human Factors and Ergonomics Society*, 37(1):32–64.

[ESRI, 2004] ESRI (2004). ArcPad ® : Mobile GIS. Technical report, ESRI.

[Falk et al., 2009] Falk, M. G., Denham, R. J., and Mengersen, K. L. (2009). Spatially stratified sampling using auxiliary information for geostatistical mapping. *Environmental and Ecological Statistics*, 18(1):93–108–108.

[Feiner et al., 1997] Feiner, S. K., MacIntyre, B., Höllerer, T., and Webster, A. (1997). A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. *Personal and Ubiquitous Computing*, 1(4):208–217.

[Fine and Minnery, 2009] Fine, M. S. and Minnery, B. S. (2009). Visual salience affects performance in a working memory task. *Journal of Neuroscience*, 29(25):8016–21.

[Furness, 1986] Furness, T. A. (1986). The super cockpit and its human factors challenges. In *Human Factors and Ergonomics Society Annual Meeting Proceedings*, volume 30, pages 48–52. Human Factors and Ergonomics Society.

[Gao and Liu, 2001] Gao, J. and Liu, Y. (2001). Applications of remote sensing, gis and gps in glaciology: a review. *Progress in Physical Geography*, 25(4):520–540.

[Gershon and Eick, 1995] Gershon, N. D. and Eick, S. G., editors (1995). *IEEE Symposium On Information Visualization 1995, InfoVis 1995, 30-31 October 1995, Atlanta, Georgia, USA*. IEEE Computer Society.

[Girgensohn et al., 2007] Girgensohn, A., Kimber, D., Vaughan, J., Yang, T., Shipman, F., Turner, T., Rieffel, E., Wilcox, L., Chen, F., and Dunnigan, T. (2007). DOTS: support for effective video surveillance. In *MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia*, pages 423–432, New York, NY, USA. ACM.

[Girshick et al., 2000] Girshick, A., Interrante, V., Haker, S., and Lemoine, T. (2000). Line direction matters: an argument for the use of principal directions in 3d line drawings. In *Proceedings of the 1st international symposium on Non-photorealistic animation and rendering*, NPAR '00, pages 43–52, New York, NY, USA. ACM.

[Goldiez et al., 2004] Goldiez, B., Livingston, M. A., Dawson, J., Brown, D., Hancock, P., Baillot, Y., and Julier, S. J. (2004). ADVANCING HUMAN CENTERED AUGMENTED REALITY RESEARCH. In *Proceedings of the 2004 Army Science Conference*, Orlando FL, USA.

[Goldsmith et al., 2008] Goldsmith, D., Liarokapis, F., Malone, G., and Kemp, J. (2008). Augmented Reality Environmental Monitoring Using Wireless Sensor Networks. In *2008 IEEE International Zurich Seminar on Communications*, pages 539–544. Ieee.

[Grasset et al., 2008] Grasset, R., Dünser, A., and Billinghurst, M. (2008). Moving Between Contexts-A User Evaluation of a Transitional Interface. In *International Conference on Artificial Reality and Telexistence (ICAT 2008).*

[Grasset et al., 2005] Grasset, R., Looser, J., and Billinghurst, M. (2005). A Step Towards a Multimodal AR Interface: A New Handheld Device for 3D Interaction. In *ISMAR '05: Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 206–207, Washington, DC, USA. IEEE Computer Society.

[Grasset et al., 2006] Grasset, R., Looser, J., and Billinghurst, M. (2006). Transitional Interface: Concept, Issues and Framework. In *5th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2006)*, pages 231–232. IEEE Computer Society.

[Greenberg et al., 1996] Greenberg, S., Gutwin, C., and Roseman, M. (1996). Awareness of Others in Groupware. In *Conference companion on Human factors in computing systems*, volume 5, page 205. ACM.

[Guiard, 1987] Guiard, Y. (1987). Asymmetric Division of Labor in Human Skilled Bimanual Action: The Kinematic Chain as a Model. *The Journal of Motor Behaviour*, 19(4):486–517.

[Gutwin and Greenberg, 2002] Gutwin, C. and Greenberg, S. (2002). A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *Computer*, 11(3):411–446.

[Hailemariam et al., 2010] Hailemariam, E., Glueck, M., Attar, R., Tessier, A., McCrae, J., and Khan, A. (2010). Toward a unified representation system of performance-related data. In *eSim 2010, IBPSA-Canada eSim Conference*, pages 117–124.

[Hart and Martinez, 2006] Hart, J. K. and Martinez, K. (2006). Environmental Sensor Networks: a revolution in the earth system science? *Earth-Science Reviews*, 78(3-4):177–191.

[Hart and Stavenland, 1988] Hart, S. G. and Stavenland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Hancock, P. A. and Meshkati, N., editors, *Human Mental Workload*, chapter 7, pages 139–183. Elsevier.

[Healey, 2001] Healey, C. G. (2001). Combining Perception and Impressionist Techniques for Nonphotorealistic Visualization of Multidimensional Data. In *SIGGRAPH 2001 Course 32: Nonphotorealistic Rendering in Scientific Visualization*, pages 20–52.

[Hegarty et al., 2002] Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., and Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence*, 30(5):425–447.

[Hengl, 2006] Hengl, T. (2006). Finding the right pixel size. *Computers & Geosciences*, 32(9):1283–1298.

[Hibbard, 2004] Hibbard, B. (2004). The top five problems that motivated my work. *IEEE Comput. Graph. Appl.*, 24(6):9–13.

[Hilbich et al., 2008a] Hilbich, C., Hauck, C., Hoelzle, M., Scherler, M., Schudel, L., Völksch, I., Mühll, D. V., and Mäusbacher, R. (2008a). Monitoring mountain permafrost evolution using electrical resistivity tomography: A 7-year study of seasonal, annual, and long-term variations at Schilthorn, Swiss Alps. *Journal of Geophysical Research*, 113(F1):1–12.

[Hilbich et al., 2008b] Hilbich, C., Hauck, C., Hoelzle, M., Scherler, M., Schudel, L., Völksch, I., Vonder Mühll, D., and Mäusbacher, R. (2008b). Monitoring mountain permafrost evolution using electrical resistivity tomography: A 7-year study of seasonal, annual, and long-term variations at schilthorn, swiss alps. *JOURNAL OF GEOPHYSICAL RESEARCH*, 113(13).

[Höllerer, 2004] Höllerer, T. (2004). *User Interfaces for Mobile Augmented Reality Systems*. Phd, Columbia University, New York, USA.

[Höllerer et al., 1999] Höllerer, T., Feiner, S., and Pavlik, J. (1999). Situated documentaries: Embedding multimedia presentations in the real world. *Wearable Computers, IEEE International Symposium*, 0:79.

[Höllerer et al., 1999] Höllerer, T., Feiner, S., Terauchi, T., Rashid, G., Hallaway, D., and Hoellerer, T. (1999). Exploring MARS: Developing Indoor and Outdoor User Interfaces to a Mobile Augmented Reality System. *Computers and Graphics*, 23(6):779–785.

[Höllerer and Feiner, 2004] Höllerer, T. and Feiner, S. K. (2004). Mobile Augmented Reality. *Telegeoinformatics LocationBased Computing and Services*, pages 1–39.

[Höllerer et al., 2001] Höllerer, T., Feiner, S. K., Hallaway, D., Bell, B., Lanzagorta, M., Brown, D., Julier, S. J., Baillot, Y., Rosenblum, L., and Hoellerer, T. (2001). User interface management techniques for collaborative mobile augmented reality. *Computers & Graphics*, 25(5):799–810.

[Hood and Reihan, 2007] Hood, M. and Reihan, A.and Loigu, E. (2007). Modeling urban stormwater runoff pollution in tallinn, estonia. In *International Symposium on "New Directions in Urban Water Management*.

[Interrante, 2000] Interrante, V. (2000). Harnessing natural textures for multivariate visualization. *IEEE Comput. Graph. Appl.*, 20(6):6–11.

[Itti, 2005] Itti, L. (2005). Models of bottom-up attention and saliency. *Neurobiology of attention*, 582(1980):576–582.

[Itti, 2007] Itti, L. (2007). Visual salience. *Scholarpedia*, 2(9):3327.

[Itti et al., 1998] Itti, L., Koch, C., and Niebur, E. (1998). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259.

[Jeung et al., 2010] Jeung, H., Sarni, S., Paparrizos, I., Sathe, S., Aberer, K., Dawes, N., Papaioannou, T. G., and Lehning, M. (2010). Effective Metadata Management in Federated Sensor Networks. *2010 IEEE International Conference on Sensor Networks Ubiquitous and Trustworthy Computing*, pages 107–114.

[Jobst and Döllner, 2008] Jobst, M. and Döllner, J. (2008). Better Perception of 3D-Spatial Relations by Viewport Variations. In *Visual Information Systems. Web-Based Visual Information Search and Management*, pages 7—-18. Springer.

[Jones and Marsden, 2006] Jones, M. and Marsden, G. (2006). *Mobile interaction design*. John Wiley & Sons Ltd.

[Ju et al., 2003] Ju, W., MADSEN, S., FIENE, J., BOLAS, M., MCDOWALL, I., and FASTE, R. (2003). Interaction devices for hands-on desktop design. *SPIE proceedings series*, pages 585–595.

[Julier et al., 2002] Julier, S., Baillot, Y., Brown, D., and Lanzagorta, M. (2002). Information filtering for mobile augmented reality. *IEEE Computer Graphics and Applications (CG&A)*, 22(5):12–15.

[Julier and Bishop, 2002] Julier, S. and Bishop, G. (2002). Special Issue on Tracking. *IEEE Computer Graphics and Applications (CG&A)*, 22(6):22–80.

[Kameda et al., 2004] Kameda, Y., Takemasa, T., and Ohta, Y. (2004). Outdoor See-Through Vision Utilizing Surveillance Cameras. *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, (Ismar):151–160.

[Karitsuka and Sato, 2003] Karitsuka, T. and Sato, K. (2003). A Wearable Mixed Reality with an On-Board Projector. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality.*

[Kastner et al., 2001] Kastner, S., De Weerd, P., Pinsk, M. A., Elizondo, M. I., Desimone, R., and Ungerleider, L. G. (2001). Modulation of sensory suppression: Implications for receptive field sizes in the human visual cortex. *Journal of Neurophysiology*, 86(3):1398–1411.

[Khlebnikov et al., 2012] Khlebnikov, R., Kainz, B., Steinberger, M., Streit, M., and Schmalstieg, D. (2012). Procedural texture synthesis for zoom-independent visualization of multivariate data. In *to appear in Proceedings of EuroVIS'12*, pages nnn–nnn.

[Kim and Dey, 2009] Kim, S. and Dey, A. K. (2009). Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. *Proceedings of the 27th international conference on Human factors in computing systems CHI 09*, page 133.

[Kim and Paulos, 2009] Kim, S. and Paulos, E. (2009). inAir: measuring and visualizing indoor air quality. In *Proceedings of the 28th international conference on Human factors in computing systems CHI 10*, pages 1861–1870. ACM Press.

[Kim and Varshney, 2006] Kim, Y. and Varshney, A. (2006). Saliency-guided Enhancement for Volume Visualization. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 12(5):925–932.

[Kim and Varshney, 2008] Kim, Y. and Varshney, A. (2008). Persuading Visual Attention through Geometry. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 14(4):772–782.

[King, G.R.; Piekarski, W.; Thomas et al., 2005] King, G.R.; Piekarski, W.; Thomas, B., King, G. R., Piekarski, W., and Thomas, B. H. (2005). ARVino - outdoor augmented reality visualisation of viticulture GIS data. In Werner, B., editor, *Proceedings of the 4th IEEEACM International Symposium on Mixed and Augmented Reality*, pages 52–55. IEEE; ACM; IEEE Comp Soc; XSENS; NOKIA; Trivisio; SIEMENS; Engn Syst Technol; INTERSENSE; Fdn Stift, IEEE Computer Society.

[Kleinnjohann, 2001] Kleinnjohann, B. (2001). Mobile ar4all. In *Proceedings of the IEEE and ACM International Symposium on Augmented Reality (ISAR'01)*, pages 181–, Washington, DC, USA. IEEE Computer Society.

[Koch and Ullman, 1985] Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human neurobiology*, 4(4):219–27.

[Kruijff et al., 2010a] Kruijff, E., Swan II, J. E., and Feiner, S. (2010a). Perceptual issues in augmented reality revisited. In *Mixed and Augmented Reality ISMAR 2010 9th IEEE International Symposium on*, number 13-16 Oct. 2010, pages 3–12. IEEE.

[Kruijff et al., 2010b] Kruijff, E., Veas, E., Gruenewald, T., Silvia, S., Luyet, V., Salminen, O., Nurminen, A., and Lehtinen, V. (2010b). HYDROSYS: on-site monitoring and management of environmental processes using handheld devices. In Anand, S., Ware, M., and Jackson, M., editors, *GeoHydroinformatics: Integrating GIS and Water Engineering*. Taylor & Francis Inc.

[Kruijff et al., 2009] Kruijff, E., Veas, E., Mendez, E., Nurminen, A., Lehtinen, V., Luyet, V., Lehning, M., Grünewald, T., Jolma, A., Simoni, S., Papaioannou, T., Salehi, A., Jemai, J., Rosten, E., and Williams, B. (2009). HYDROSYS System Specification. Technical report, Graz University of Technology.

[Laidlaw et al., 2005] Laidlaw, D. H., Kirby, R. M., Jackson, C. D., Davidson, J. S., Miller, T. S., Silva, M. D., Warren, W. H., and Tarr, M. J. (2005). Comparing 2d vector field visualization methods: A user study. *IEEE Transactions on Visualization and Computer Graphics*, 11:59–70.

[Langendoen et al., 2006] Langendoen, K., Baggio, A., and Visser, O. (2006). Murphy loves potatoes: Experiences from a pilot sensor network deployment in precision agriculture. In *International Workshop on Parallel and Distributed Real-Time Systems (WPDRTS*.

[Langlotz et al., 2011] Langlotz, T., Mooslechner, S., Zollmann, S., Degendorfer, C., Reitmayr, G., and Schmalstieg, D. (2011). Sketching up the world: in situ authoring for mobile Augmented Reality. *Personal and Ubiquitous Computing*, pages 1–8.

[Lawton, 1994] Lawton, C. A. (1994). Gender differences in way-finding strategies: Relationship to spatial ability and spatial anxiety. *Sex Roles*, 30(11):765–779.

[Lee et al., 2007] Lee, S., Kim, G. J., and Choi, S. (2007). Real-time tracking of visually attended objects in interactive virtual environments. In *ACM VRST*, pages 15–24.

[Lehning et al., 2006] Lehning, M., Völksch, I., Gustafsson, D., Nguyen, T. A., Stähli, M., and Zappa, M. (2006). ALPINE3D : a detailed model of mountain surface processes and its application to snow hydrology. *Hydrological Processes*, 2128(May 2005):2111–2128.

[Lifton et al., 2009] Lifton, J., Laibowitz, M., Harry, D., Gong, N.-W., Mittal, M., and Paradiso, J. A. (2009). Metaphor and manifestation - cross-reality with ubiquitous sensor/actuator networks. *IEEE Pervasive Computing*, 8(3):24–33.

[Lin and Loftin, 1998] Lin, C.-R. and Loftin, R. B. (1998). Application of virtual reality in the interpretation of geoscience data. In *Proceedings of the ACM symposium on Virtual reality software and technology 1998 VRST 98*, pages 187–194. ACM Press.

[Liston and Sturm, 1998] Liston, G. E. and Sturm, M. (1998). A snow-transport model for complex terrain. *Journal Of Glaciology*, 44(148):498–516.

[Livingston, 2005] Livingston, M. A. (2005). Evaluating human factors in augmented reality systems. *IEEE Computer Graphics and Applications*, 25(6):6–9.

[Lorenz et al., 2008] Lorenz, H., Trapp, M., Jobst, M., and Döllner, J. (2008). Interactive Multi-Perspective Views of Virtual 3D Landscape and City Models. In *11th AGILE International Conference on GI Science*, pages 301–321.

[Mann, 1996] Mann, S. (1996). Smart Clothing: The Shift to Wearable Computing. *Communications of the ACM*, (August 1996).

[Marras, 1997] Marras, W. (1997). Biomechanics of the Human Body. In Salvendy, G., editor, *Handbook of Human Factors and Ergonomics*. John Wiley & Sons, second edition.

[Martinez et al., 2004] Martinez, K., Hart, J. K., and Ong, R. (2004). Environmental sensor networks. *Computer*, 37(8):50–56.

[McCleary, 1983] McCleary, G. (1983). An effective graphic ?vocabulary? *IEEE Computer Graphics and Applications*, 3:46–53.

[McCormick, 1988] McCormick, B. H. (1988). Visualization in scientific computing. *SIGBIO Newsl.*, 10(1):15–21.

[Meiguins et al., 2006] Meiguins, B. S., Casseb do Carmo, R. M., Goncalves, A. S., Alves Godinho, P. I., and de Brito Garcia, M. (2006). Using augmented reality for multidimensional data visualization. In *Proceedings of the conference on Information Visualization*, IV '06, pages 529–534, Washington, DC, USA. IEEE Computer Society.

[Mendez, 2011] Mendez, E. (2011). *On the Usage of Context for Augmented Reality Visualization*. Phd, Graz University of Technology, Graz, Austria.

[Mendez et al., 2010] Mendez, E., Feiner, S. K., and Schmalstieg, D. (2010). Focus and Context in Mixed Reality by Modulating First Order Salient Features. In *Int. Symp. on Smart Graphics*, pages 232–243.

[Mendez and Schmalstieg, 2009] Mendez, E. and Schmalstieg, D. (2009). Importance masks for revealing occluded objects in augmented reality. *16th ACM Symposium on Virtual Reality*, pages 247—-248.

[Milgram and Kishino, 1994] Milgram, P. and Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems*, E77-D(12).

[Mitas and Mitasova, 1998] Mitas, L. and Mitasova, H. (1998). Distributed soil erosion simulation for effective erosion prevention. *Water Resources Research*, 34(3):505–516.

[Mitterer et al., 2011] Mitterer, C., Hirashima, H., and Schweizer, J. (2011). Wet-snow instabilities: comparison of measured and modelled liquid water content and snow stratigraphy. *Annals of Glaciology*, 52(58):201–208.

[Mitterer et al., 2009] Mitterer, C., Mott, R., and Schweizer, J. (2009). Observations and analysis of two wet-snow avalanche cycles. In *International Snow Science Workshop ISSW*, number April 2008, pages 262–266. Swiss Federal Institute for Forest, Snow and Landscape Research WSL.

[Neumann et al., 2003] Neumann, U., You, S., Hu, J., Jiang, B., and Lee, J. (2003). Augmented virtual environments (ave): Dynamic fusion of imagery and 3d models. In *IEEE Virtual Reality (VR)*, pages 61–67.

[Nikishkov and Tsuchimoto, 2007] Nikishkov, G. and Tsuchimoto, T. (2007). Using augmented reality for real-time visualization of tactile health examination. In Braz, J., Vázquez, P.-P., and Pereira, J. M., editors, *GRAPP (AS/IE)*, pages 91–97. INSTICC - Institute for Systems and Technologies of Information, Control and Communication.

[Nocke et al., 2008] Nocke, T., Sterzel, T., and Wrobel, M. (2008). Visualization of Climate and Climate Change Data: An Overview. In Eds, E. E. A., editor, *Digital Earth Summit on Geoinformatics 2008 Tools for Global Change Research ISDE08*, volume 1999.

[Norman, 1990] Norman, D. (1990). *The Design of Everyday Things*. Doubleday, New York.

[Ouerhani et al., 2004] Ouerhani, N., Wartburg, R. V., Hugli, H., and Muri, R. (2004). Empirical validation of the saliency-based model of visual attention. *Electronic Letters on Computer Vision and Image Analysis*, 3(1):13–24.

[Paczkowski et al., 2011] Paczkowski, P., Kim, M. H., Morvan, Y., Dorsey, J., Rushmeier, H., and O'Sullivan, C. (2011). Insitu: Sketching architectural designs in context. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia 2011)*, 30(6):182:1–10.

[Pasewaldt et al., 2011] Pasewaldt, S., Trapp, M., and Döllner, J. (2011). Multiscale Visualization of 3D Geovirtual Environments Using View-Dependent Multi-Perspective Views. *Journal of WSCG*, 19(3):111–118.

[Pawson and Greenberg, 2009] Pawson, M. and Greenberg, S. (2009). Extremely Rapid Usability Testing. *Journal of Usability Studies*, 4(3):124–135.

[Paxton and Benford, 2009] Paxton, M. and Benford, S. (2009). Experiences of participatory sensing in the wild. *Proceedings of the 11th international conference on Ubiquitous computing Ubicomp 09*, page 265.

[Peterson, 2009] Peterson, S. D. (2009). *Stereoscopic Label Placement: Reducing Distraction and Ambiguity in Visually Cluttered Displays*. Phd, Linköping University, Norrköping, Sweden.

[Piekarski and Thomas, 2001] Piekarski, W. and Thomas, B. (2001). Tinmith-Metro: New Outdoor Techniques for Creating City Models with an Augmented Reality Wearable Computer. In *IEEE International Symposium on Wearable Computers (ISWC)*, pages 31–38, Zurich, Switzerland.

[Pinelle and Gutwin, 2005] Pinelle, D. and Gutwin, C. (2005). A Groupware Design Framework for Loosely Coupled Workgroups. In *European Conference on Computer-Supported Cooperative Work*, number September, pages 119–139.

[Rauhala et al., 2006] Rauhala, M., Gunnarsson, A.-S., and Henrysson, A. (2006). A novel interface to sensor networks using handheld augmented reality. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*, MobileHCI '06, pages 145–148, New York, NY, USA. ACM.

[Reitmayr and Drummond, 2006] Reitmayr, G. and Drummond, T. W. (2006). Going out: Robust Tracking for Outdoor Augmented Reality. In *IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 109–118, Santa Barbara, CA, USA. IEEE Computer Society.

[Reitmayr and Schmalstieg, 2001] Reitmayr, G. and Schmalstieg, D. (2001). Mobile collaborative augmented reality. In *Proceedings IEEE and ACM International Symposium on Augmented Reality*, pages 114–123. IEEE Comput. Soc.

[Rekimoto, 1995] Rekimoto, J. (1995). The Magnifying Glass Approach to Augmented Reality Systems. In *International Conference on Artificial Reality and TeleExistence*, volume 95, pages 123–132.

[Rekimoto, 2001] Rekimoto, J. (2001). NaviCam: A Palmtop Device Approach to Augmented Reality. In *Fundamentals of Wearable Computing and Augmenred Reality*. Laurence Erlbaum Associates.

[Rekimoto and Nagao, 1995] Rekimoto, J. and Nagao, K. (1995). The World through the Computer: Computer Augmented Interaction with Real World Environments. *Proc 8th Ann ACM Symp User Interface and Software Technology UIST ACM Press*, pages:29–36.

[Rensink et al., 2000] Rensink, R. A., O'Regan, J. K., and Clark, J. J. (2000). On the Failure to Detect Changes in Scenes Across Brief Interruptions. *Visual Cognition*, 7(1-3):127–145.

[Sandor et al., 2010] Sandor, C., Cunningham, A., Dey, A., and Mattila, V.-V. (2010). An Augmented Reality X-Ray system based on visual saliency. *2010 IEEE International Symposium on Mixed and Augmented Reality*, pages 27–36.

[Sandor et al., 2009] Sandor, C., Cunningham, A., Eck, U., Urquhart, D., Jarvis, G., Dey, A., Barbier, S., Marner, M. R., and Rhee, S. (2009). Egocentric space-distorting visualizations for rapid environment exploration in mobile mixed reality. *2009 8th IEEE International Symposium on Mixed and Augmented Reality*, pages 211–212.

[Santella and DeCarlo, 2004] Santella, A. and DeCarlo, D. (2004). Visual interest and NPR: an evaluation and manifesto. In *Int. Symp. on Non-photorealistic Animation and Rendering (NPAR)*.

[Sawhney et al., 2002] Sawhney, H. S., Arpa, A., Kumar, R., Samarasekera, S., Aggarwal, M., Hsu, S., Nister, D., and Hanna, K. (2002). Video flashlights: real time rendering of multiple videos for immersive model visualization. *Eurographics Workshop on Rendering*, pages 157–168.

[Schall et al., 2008a] Schall, G., Grabner, H., Grabner, M., Wohlhart, P., Schmalstieg, D., and H, B. (2008a). 3d tracking using on-line keypoint learning for mobile augmented reality. In *Workshop on Visual Localization for Mobile Platforms in conj. with IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008)*, Anchorage, Alaska, USA.

[Schall et al., 2010] Schall, G., Junghanns, S., and Schmalstieg, D. (2010). VIDENTE - 3D Visualization of Underground Infrastructure using Handheld Augmented Reality. In Anand, S., Ware, M., and Jackson, M., editors, *GeoHydroinformatics: Integrating GIS and Water Engineering*. Taylor & Francis Inc.

[Schall et al., 2007] Schall, G., Mendez, E., Junghanns, S., Schmalstieg, D., and Reitinger, B. (2007). Handheld Geospatial Augmented Reality using Urban 3D Models. In *Workshop on Mobile Spatial Interaction, ACM International Conference on Human Factors in Computing Systems (CHI'07)*.

[Schall et al., 2008b] Schall, G., Mendez, E., Kruijff, E., Veas, E., Junghanns, S., Reitinger, B., and Schmalstieg, D. (2008b). Handheld Augmented Reality for underground infrastructure visualization. *Personal and Ubiquitous Computing*, 13(4):281–291.

[Schall and Schmalstieg, 2008] Schall, G. and Schmalstieg, D. (2008). Interactive urban models generated from context-preserving transcoding of real-wold data. In *Proc. of the 5th International Conference on GIScience, abstracts volume, Park City*.

[Schall et al., 2009] Schall, G., Wagner, D., Reitmayr, G., Taichmann, E., Wieser, M., Schmalstieg, D., and Hofmann-Wellenhof, B. (2009). Global pose estimation using multi-sensor fusion for outdoor augmented reality. In *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*, pages 153–162. IEEE.

[Schmalstieg et al., 2002] Schmalstieg, D., Fuhrmann, A., Hesina, G., Szalavári, Z., Encarnacao, L. M., Gervautz, M., and Purgathofer, W. (2002). The studierstube augmented reality project. *Presence: Teleoperators and Virtual Environments*, 11(1):33–54.

[Schmalstieg and Reitmayr, 2005] Schmalstieg, D. and Reitmayr, G. (2005). The World as a User Interface: Augmented Reality for Ubiquitous Computing. In *Central European Multimedia and Virtual Reality Conference 2005 (CEMVRC 2005)*, Prague, Czech Republic.

[Schuler and Namioka, 1993] Schuler, D. and Namioka, A. (1993). *Participatory design: principles and practices*. Routledge.

[Schweizer et al., 2003] Schweizer, J., Jamieson, J. B., and Schneebeli, M. (2003). Snow avalanche formation. *Reviews of Geophysics*, 41(4):1016.

[Selavo et al., 2007] Selavo, L., Wood, A., Cao, Q., Sookoor, T., Liu, H., Srinivasan, A., Wu, Y., Kang, W., Stankovic, J., Young, D., and Porter, J. (2007). Luster: wireless sensor network for environmental research. In *Proceedings of the 5th international conference on Embedded networked sensor systems*, SenSys '07, pages 103–116, New York, NY, USA. ACM.

[Shepard and Metzler, 1971] Shepard, R. N. and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(972):701–703.

[Shin et al., 2007] Shin, I., Kim, J., and Lee, S.-G. (2007). Development of an internet-based water-level monitoring and measuring system using CCD camera. In *ICMIT The International Society for Optical Engineering*. SPIE.

[Shneiderman, 1996] Shneiderman, B. (1996). The eyes have it: a task by data type taxonomy for information visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages*, 0(July):336–343.

[Simoni et al., 2008] Simoni, S., Zanotti, F., Bertoldi, G., and Rigon, R. (2008). Modelling the probability of occurrence of shallow landslides and channelized debris flows using geotop-fs. *Hydrological Processes*, 22(4):532–545.

[Stutzman et al., 2009] Stutzman, B., Neilson, D., Broderick, T., and Neubert, J. (2009). MARTI: Mobile augmented reality tool for industry. In *Proceedings of the IEEE World Congress on Computer Science and Information Engineering*.

[Sukan and Feiner, 2010] Sukan, M. and Feiner, S. K. (2010). SnapAR: Storing snapshots for quick viewpoint switching in hand-held augmented reality. In *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium*, pages 273–274.

[Sutherland, 1965] Sutherland, I. E. (1965). The ultimate display. In *Proceedings of the IFIP Congress*, pages 506–508.

[Sutherland, 1968] Sutherland, I. E. (1968). A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, AFIPS '68 (Fall, part I), pages 757–764, New York, NY, USA. ACM.

[Swan et al., 2007] Swan, J. E., Jones, A., Kolstad, E., Livingston, M. A., and Smallman, H. S. (2007). Egocentric Depth Judgments in Optical, See-Through Augmented Reality. *IEEE Trans. Vis. Comput. Graph.*, 13(3):429–442.

[Szewczyk et al., 2004] Szewczyk, R., Polastre, J., Mainwaring, A. M., and Culler, D. E. (2004). Lessons from a sensor network expedition. In *EWSN*, pages 307–322.

[Techel et al., 2011] Techel, F., Pielmeier, C., and Schneebeli, M. (2011). Microstructural resistance of snow following first wetting. *Cold Regions Science and Technology*, 65(3):382–391.

[Thomas et al., 2000] Thomas, B., Close, B., Donoghue, J., Squires, J., de Bondi, P., Morris, M., and Piekarski, W. (2000). {ARQuake:} An {Outdoor/Indoor} Augmented Reality First Person Application. In *Proceedings of the 4th {IEEE} International Symposium on Wearable Computers*, {ISWC} '00, pages 139–, Washington, {DC}, {USA}. {IEEE} Computer Society.

[Thomas et al., 1998] Thomas, B., Demczuk, V., Piekarski, W., Hepworth, D., and Gunter, B. (1998). A wearable computer system with augmented reality to support terrestrial navigation. In *In proceedings of ISWC'98*.

[Thomas and Cook, 2005] Thomas, J. J. and Cook, K. A. (2005). *Illuminating the path: The research and development agenda for visual analytics*, volume 54. IEEE.

[Treisman and Gelade, 1980] Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, 12(1).

[Tsang and Vidulich, 2006] Tsang, P. S. and Vidulich, M. A. (2006). *Mental workload and situation awareness*, chapter Handbook o. John Wiley & Sons.

[Tversky, 2000] Tversky, B. (2000). *Cognitive mapping: past, present, and future*, chapter Levels and. Routledge Frontiers of Cognitive Science No.4. Routledge.

[Veas et al., 2012a] Veas, E., Grasset, R., Ferencik, I., Grünewald, T., and Schmalstieg, D. (2012a). Mobile Augmented Reality for Environmental Monitoring. *Submitted for review to Personal and Ubiquitous Computing*, 0(0):0–16.

[Veas et al., 2012b] Veas, E., Grasset, R., Kruijff, E., and Schmalstieg, D. (2012b). Extended overview techniques for outdoor augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 18:565–572.

[Veas and Kruijff, 2008] Veas, E. and Kruijff, E. (2008). Vesp'R: design and evaluation of a handheld AR device. In *ISMAR '08: Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 43–52, Washington, DC, USA. IEEE Computer Society.

[Veas et al., 2009] Veas, E., Kruijff, E., and Mendez, E. (2009). HYDROSYS - first approaches towards on-site monitoring and management with handhelds. In J. Hřebíček, J. Hradec, E. Pelikán, O. Mírovský, W. Pillmann, I.Holoubek, T. B., editor, *Towards e-environment, EENVI2009*, Prague, Czech Republic.

[Veas et al., 2011] Veas, E., Mendez, E., Feiner, S., and Schmalstieg, D. (2011). Directing Attention and Influencing Memory with Visual Saliency Modulation. In *to appear in Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI2011)*, Vancouver, British Columbia, Canada.

[Veas and Kruijff, 2010] Veas, E. E. and Kruijff, E. (2010). Handheld Devices for Mobile Augmented Reality. In *MUM '10 Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia*, Limassol, Cyprus.

[Veas et al., 2010] Veas, E. E., Mulloni, A., Kruijff, E., Schmalstieg, D., and Regenbrecht, H. (2010). Techniques for View Transition in Multi-Camera Outdoor Environments. In *Graphics Interface 2010 (GI '10)*, pages 193–200, Toronto, Ont., Canada. Canadian Information Processing Society.

[Wagner, 2007] Wagner, D. (2007). *Handheld Augmented Reality*. Phd, Graz University of Technology, Graz, Austria.

[Wagner and Schmalstieg, 2003] Wagner, D. and Schmalstieg, D. (2003). First steps towards handheld augmented reality. In *In Proc. of International Conference on Wearable Computers (ISWC'03)*, page 127.

[Wang et al., 2007] Wang, Y., Krum, D. M., Coelho, E. M., and Bowman, D. A. (2007). Contextualized videos: Combining videos with environment models to support situational understanding. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 13(6):1568–1575.

[Werner-Allen et al., 2005] Werner-Allen, G., Johnson, J., Ruiz, M., Lees, J., and Welsh, M. (2005). Monitoring volcanic eruptions with a wireless sensor network. In *in Proceedings of the Second European Workshop on Wireless Sensor Networks (EWSN'05)*.

[White, 2009] White, S. (2009). SiteLens : Situated Visualization Techniques for Urban Site Visits. *Evaluation*, pages 1117–1120.

[Wolfe et al., 2007] Wolfe, J. M., Horowitz, T. S., and Michod, K. O. (2007). Is visual attention required for robust picture memory? *Vision research*, 47(7):955–64.

[Yang et al., 2009] Yang, J., Zhang, C., Li, X., Huang, Y., Fu, S., and Acevedo, M. F. (2009). Integration of wireless sensor networks in environmental monitoring cyber infrastructure. *Wireless Networks*, 16(4):1091–1108.

[Zacks et al., 2001] Zacks, J. M., Mires, J., Tversky, B., and Hazeltine, E. (2001). Mental spatial transformations of objects and perspective. *Spatial Cognition and Computation*, 2(4):315–332.

[Zijstra, 1993] Zijstra, F. R. H. (1993). *Efficiency in Work Behaviour; A Design Approach for Modern Tools*. Dissertation, Industrial Design Engineering, Delft University.