

SCENE CREATION AND EXPLORATION IN OUTDOOR AUGMENTED REALITY

by Mustafa Tolga Eren

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Sabanci University

August, 2013

SCENE CREATION AND EXPLORATION IN OUTDOOR AUGMENTED REALITY

APPROVED BY:

Doç. Dr. Selim Balcısoy
(Dissertation Supervisor)



Doç. Dr. Tolga Çapın



Doç. Dr. Albert Levi



Doç. Dr. Gözde Ünal



Doç. Dr. Berrin Yanıkoğlu



DATE OF APPROVAL:

13.08.2013

©Mustafa Tolga Eren, 2013

ALL RIGHTS RESERVED

SCENE CREATION AND EXPLORATION IN OUTDOOR AUGMENTED REALITY

Mustafa Tolga Eren

Computer Science and Engineering, PhD Thesis, 2013

Thesis Supervisor: Assoc. Prof. Dr. Selim Balcısoy

Keywords Outdoor Augmented Reality, Modelling, Annotations, Scene Exploration, X-Ray Visualization

This thesis investigates Outdoor Augmented Reality (AR) especially for scene creation and exploration aspects. We decompose a scene into several components: a) Device, b) Target Object(s), c) Task, and discuss their interrelations. Based on those relations we outline use-cases and workflows. The main contribution of this thesis is providing AR oriented workflows for selected professional fields specifically for scene creation and exploration purposes, through case studies as well as analyzing the relations between AR scene components. Our contributions include, but not limited to: i) analysis of scene components and factoring inherently available errors, to create a transitional hybrid tracking scheme for multiple targets, ii) a novel image-based approach that uses building block analogy for modelling and introduces volumetric and temporal labeling for annotations, iii) an evaluation of the state of the art X-Ray visualization methods as well as our proposed multi-view method. AR technology and capabilities tend to change rapidly, however we believe the relation between scene components and the practical advantages their analysis provide are valuable. Moreover, we have chosen case studies as diverse as possible in order to cover a wide range of professional field studies. We believe our research is extendible to a variety of field studies for disciplines including but not limited to: Archaeology, architecture, cultural heritage, tourism, stratigraphy, civil engineering, and urban maintenance.

AÇIKHAVA ARTIRILMIŞ GERÇEKLIK ORTAMLARINDA SAHNE YARATILMASI VE İNCELENMESİ

Mustafa Tolga Eren

Bilgisayar Bilimleri ve Mühendisliği, Doktora Tezi, 2013

Supervisor: Doç. Dr. Selim Balcısoy

Anahtar Kelimeler: Açık hava Artırılmış Gerçeklik, Modelleme, Artırılmış Gerçeklik Ortamında Bilgi Notları, Sahne İncelenmesi, X-Işını Görselleştirme

Bu tez Artırılmış Gerçeklik (AG) ortamında sahne yaratılması ve incelenmesi konularını, özellikle açık hava ortamında araştırma amacını güder. Bu amaçla AG sahneleri üç temel bileşende incelenip; a) Cihaz, b) Hedef Obje(ler), c) Görev, ve bu maddelerin kendi içlerindeki ilişkileri tartışılmıştır. Bu ilişkiler doğrultusunda kullanım senaryoları ve iş akışları tanımlanmıştır. Tezin literatüre ana katkısı, profesyonel çalışmalara odaklı olarak sağlanan açık hava AG iş akışları ve bu iş akışlarının sahne bileşenleriye olan ilişkilerinin incelenmesinden kaynaklıdır. Diğer katkı noktaları ise şöyle sıralanabilir: i) Sahne bileşenlerinin içsel hatalarının tespit edilmesi ve incelenmesi. Bu inceleme doğrultusunda ortaya çıkan, sahnedeki hedefleri takip etmeye yarayan, geçişlere uygun şekilde hazırlanmış, melez bir izleme methodu. ii) Resim tabanlı, blokların birbirine bağlanması ile işleyen bir tekniğin modelleme amacıyla tanıtılması. Ayrıca, sahnelere eklenen bilgi notlarının, hacimsel ve zamansal olarak da incelenmesi ve uyarlanması. iii) Güncel X-Işını görselleştirme tekniklerinin deneysel bir metod ile karşılıklı incelenmesiyle çıkan sonuçlar ve bu sonuçlar doğrultusunda tasarlanan çok yüzeyli yeni bir görselleştirme tekniği. AG teknolojisi ve getirileri hızlı bir şekilde değişmekte olsa bile, sahne bileşenlerinin kendileriyle ve kullanıcıyla olan ilişkisinin incelenmesinden doğan pratik getirilerin değerli ve kalıcı olduğu kesindir. Bu tez içerisinde yer alan fikir ve çalışmaların şu çeşitli alanlara da uyarlanabileceğini düşünmekteyiz: Arkeoloji, mimari, kültürel miras, turizm, stratigrafi, inşaat ve şehircilik.

To my dear grandmother Emine Konak

'Ananem' Emine Konak'a

Table of Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 2 | Related Work | 5 |
| 2.1 | Brief Introduction to Augmented Reality | 5 |
| 2.2 | Tracking | 9 |
| 2.2.1 | Visual Tracking | 9 |
| 2.2.2 | Sensor Based Tracking | 10 |
| 2.2.3 | Hybrid Tracking | 10 |
| 2.3 | Modelling and Annotations | 11 |
| 2.3.1 | Modeling | 11 |
| 2.3.2 | Annotations | 12 |
| 2.3.3 | In Field Studies | 12 |
| 2.4 | Exploration and Measurements | 13 |
| 2.4.1 | Perception | 13 |
| 2.4.2 | X-Ray Visualization | 13 |
| 2.4.3 | Focus and Context | 15 |
| 3 | Devices and Target(s) | 17 |
| 3.1 | Orientation Sensors on Mobile Devices | 17 |
| 3.1.1 | Accelerometer | 18 |
| 3.1.2 | Gyroscope | 18 |
| 3.1.3 | Magnetometer | 19 |
| 3.1.4 | Summary of Orientation Sensors' Capabilities | 19 |
| 3.2 | GPS sensors on Mobile Devices | 20 |
| 3.3 | Standard Localization Workflow Using Inertial Sensors and GPS | 22 |
| 3.4 | Computational Experiments to Evaluate Perceived Errors | 23 |
| 3.4.1 | Scene Definition and Overview of Experiments | 24 |
| 3.4.2 | Experiment 1: | 24 |
| 3.4.3 | Experiment 2: | 26 |

| | | |
|----------|--|-----------|
| 3.4.4 | Experiment 3 and 4: | 27 |
| 3.5 | Hybrid Localization | 31 |
| 3.6 | Case Study | 33 |
| 3.6.1 | Transition Model | 34 |
| 3.6.1.1 | Transition Criteria | 34 |
| 3.6.1.2 | Transition function | 35 |
| 3.6.2 | Visualization | 35 |
| 3.6.2.1 | Visualization Modes | 37 |
| 3.6.3 | Tracking | 38 |
| 3.6.4 | Discussion | 39 |
| 4 | Modeling and Annotations | 41 |
| 4.1 | Block Based Mobile Modeling | 43 |
| 4.2 | Temporal and Volumetric Annotations | 45 |
| 4.2.1 | Spatial Component | 46 |
| 4.2.2 | Semantic Component | 46 |
| 4.3 | Case Study | 49 |
| 4.4 | User Study and Discussion | 52 |
| 5 | Exploration and Measurement | 53 |
| 5.1 | X-Ray Visualization | 55 |
| 5.2 | Absolute Vertical Depth Judgments | 56 |
| 5.2.1 | Absolute Vertical Depth Judgment Experiment Setup and Task | 57 |
| 5.2.2 | Absolute Vertical Depth Judgment Experiment Results | 59 |
| 5.3 | Multi-view visualization | 61 |
| 5.3.1 | Spherical and Screen Space Clipping | 61 |
| 5.3.2 | Orthographic-View | 62 |
| 5.4 | Case Study | 63 |
| 5.4.1 | AR Application | 63 |
| 5.4.2 | Interaction | 67 |
| 5.5 | User Study | 67 |
| 5.5.1 | Comparative Vertical Depth Judgments | 69 |
| 5.5.1.1 | Experiment Setup and Task | 69 |
| 5.5.1.2 | Experiment Results | 70 |
| 5.6 | Discussion | 72 |
| 5.6.1 | Estimation Tendency | 72 |
| 5.6.2 | Precision Scenarios | 73 |
| 5.6.3 | Multi View Technique | 73 |

| | |
|---------------------|-----------|
| 6 Discussion | 75 |
| Bibliography | 76 |

List of Figures

| | | |
|------|---|----|
| 2.1 | A rear view mirror found in almost every automobile since 1914 | 5 |
| 2.2 | Gyro gunsight display system from Spitfire Mk V fighter plane | 6 |
| 2.3 | Heads-Up-Display system of an F/A-18C | 6 |
| 2.4 | Sutherland et al.'s Sword of Damocles setup. | 7 |
| 2.5 | MARS is one of the earliest examples of mobile AR. | 8 |
| 2.6 | Nokia's Mara application is one of the earliest AR applications on a camera phone. | 8 |
| 2.7 | Left: Wagner et al.'s method can be used to guide users in taking photos of the environment to be later stitched with a desktop application. Right: An Annotated panorama. | 9 |
| 2.8 | Layar and Wikitude applications on Android OS are showed side by side. | 10 |
| 2.9 | A hybrid tracking flow for outdoor Augmented Reality | 10 |
| 2.10 | VideoTrace application allows video-based rapid modelling via user interaction. | 12 |
| 2.11 | Edge overlay technique as used by Avery et al. | 14 |
| 2.12 | Left side shows a naïve approach and right side demonstrates excavation box approach for visualizing underground structures. | 14 |
| 2.13 | Ghosting effect is utilized in purpose of X-Ray visualization by Zollman et al. | 15 |
| 2.14 | In order to improve the depth perception, magic lens approach can be utilized. | 16 |
| 2.15 | Occluding augmentations. A typical AR visualization to draw the user's attention with overlaid semitransparent geometry, occluding the object of interest. | 16 |
| 3.1 | Calculation of true heading from magnetometer and accelerometer readings. | 19 |
| 3.2 | GPS inaccuracy visualized. | 21 |
| 3.3 | A sample handheld device is shown. Red dot denotes the ideal placement of target object without any errors induced. Blue dots are deviated from the result due to the induced errors. | 24 |

| | | |
|------|---|----|
| 3.4 | Red and blue dots denote the viewer and object positions, respectively. Red line visualizes a correct measurement for user's viewing direction. The green lines denotes typical errors for a magnetometer measurement. | 25 |
| 3.5 | Perceived error due to magnetometer inaccuracies is visualized in \pm percentage of screen width. | 26 |
| 3.6 | Average misplacement due to GPS error per distance is plotted | 27 |
| 3.7 | Red dot denotes the exact location for the user. Blue dots are the deviated user positions. Green dot visualizes the target object which is 10 m away. Red dot denotes the exact location for the user. Blue dots are the deviated user positions. Green dots visualize the target object which is 10 m (left) and 100m (right) away. | 28 |
| 3.8 | A sphere and a building model's correct projection ratio with respect to distance is plotted. | 29 |
| 3.9 | Object sizes relative to screen space are plotted. | 29 |
| 3.10 | To inspect viewing angle's effect with respect to GPS and orientation errors, several virtual cameras are placed around the object on a grid. | 30 |
| 3.11 | A Building model is shown with respect to its color coded correctness map. | 31 |
| 3.12 | A color coded correctness map for the sphere model is shown. | 31 |
| 3.13 | Threshold values for orientation sensor tracking and vision tracking is visualized for the building model. | 32 |
| 3.14 | Threshold values for orientation sensor tracking and vision tracking is visualized for the sphere model. | 32 |
| 3.15 | Tracking thresholds for multiple viewing angles and distances are visualized. The red region favors sensor based tracking where the green region favors vision based tracking. | 33 |
| 3.16 | General transition diagram for tracking mode switching for each target object. | 34 |
| 3.17 | Transition function is illustrated. b^n and f^n denotes cutoff values for forward and backwards movement respectively. v^n denotes visualization modes. | 35 |
| 3.18 | Transition Diagram with four different visualization modes. | 36 |
| 3.19 | Sematic relation between visualization states is demonstrated. | 36 |
| 3.20 | Four prototypical visualization modes are shown. Transition between states is performed via the transition function in Figure 3.18, based on a transition criterion. | 37 |

| | | |
|------|--|----|
| 3.21 | Transition function is illustrated as a hysteresis curve. f^n and b^n denote cutoff values for forward and backwards movement respectively; for switching between two consecutive visualization modes, namely v^m and v^{m+1} | 40 |
| 4.1 | An urban scene is (a) photographed. Using these images, two objects are (b) modelled and (c) annotated using our workflow. Annotations are color coded; a legend is shown in the canvas for identification. | 42 |
| 4.2 | Example The green polygon is the initial polygon. The red polygon is defined by user clicks. Green and red edges are supplied to Delaunay triangulation. The output is the combination of green, red and black edges. | 44 |
| 4.3 | The user observes a 3D model ready to be annotated. | 47 |
| 4.4 | Red squares denote user clicked 3D positions. Using these two points and the position of the virtual camera, a clipping plane is calculated. With this clipping plane the 3D model is divided into two 3D volumetric regions. Green line is the contact region of these two regions. | 47 |
| 4.5 | A new volumetric region is generated using the same approach in Figure 4 4. The user clicked points, do not have to be on the same face. As long as they are located on the model geometry, a new clipping plane is calculated. | 48 |
| 4.6 | A final region is added. The created volumetric regions are associated with semantic components to create annotations. The annotations are presented in different colors and superimposed over the model. | 48 |
| 4.7 | Our workflow is summarized in three steps. A modelled object can be annotated more than once. | 49 |
| 4.8 | A building is photographed from four different angles, two of these are shown here. | 49 |
| 4.9 | Modelling process starts with creating and adjusting a reference block. This block has the same orientation with the building. | 49 |
| 4.10 | Completed model is shown; in this example 6 blocks are used to model the entire building. | 50 |
| 4.11 | After generating volumetric regions as spatial components, four different annotations are created. These are, from top to down; 2nd Floor, 1st Floor, Ground and Basement. | 50 |

| | | |
|------|---|----|
| 4.12 | The real world image(a), is annotated using our workflow(c). The sketch(b) is provided to subjects as a guideline for annotation task of the user study. Subjects were expected to label four different layers, namely; <i>steel support</i> , <i>first restoration</i> , <i>second restoration</i> and <i>new base</i> | 51 |
| 5.1 | Visualization of underground pipe networks using different techniques. a) Careless overlay b) Edge overlay c) Excavation box d) Our proposed multi view technique. | 54 |
| 5.2 | Absolute Vertical Depth Judgment experiment: Each participant performs six estimations for vertical position of the pipe B_i , where $i \in 1..6$ (25 to 100cm). | 57 |
| 5.3 | Absolute vertical depth judgments were plotted against actual distances. Each technique has a vertical offset in the plot for clarification. | 58 |
| 5.4 | From left to right, sample scenes are visualized via careless overlay, edge overlay and excavation box techniques for absolute vertical depth judgment experiment. | 59 |
| 5.5 | “above” and “under” planes are placed in an empty scene. An edge overlay is drawn to denote ground plane. a) front view, b) “above” and “under” touching, c) “under” plane at a depth of four meters directly beneath the “above” plane. | 60 |
| 5.6 | Clipping sphere used for focus preservation through information filtering. a) Focused on red pipe layer. b) Focused on the area between the layers. c) Focused on blue pipe layer. | 60 |
| 5.7 | 2D clipping for edge overlay. Clipping circle’s size is determined via the anchor object’s position in screen space. Focus region is larger when the anchor is closer to the user. | 61 |
| 5.8 | Perpendicular blue and red pipes are viewed at a close range. Using a-b) the spatial relations are ambiguous; c) our method clearly identifies relative positioning via orthographic view. | 64 |
| 5.9 | Two parallel pipes are visualized at a distance. Using a) edge overlay method, both pipes seem underground, however relational positioning information is lost. b) Blue pipe is fully occluded by the excavation box. c) Using our method both pipes with their spatial relation are visualized via multi views. | 64 |

| | | |
|------|--|----|
| 5.10 | Parallel red and blue pipes are visualized using three methods. Due to perspective projection, the red pipe occludes the blue pipe that is behind. In a-b) blue pipe cannot be seen, c) our method is able to visualize both pipes in each view. | 65 |
| 5.11 | Touch based interactions translates the anchor a) along the viewing direction, b) through the ground. | 66 |
| 5.12 | Absolute vertical depth judgments were plotted against actual distances for digbox and multi-view techniques. | 68 |
| 5.13 | Comparative Vertical Depth Judgment experiment: Each participant is asked to identify the relative vertical distance between the red pipe b and the blue pipe D_i , where $i \in 1..6$ | 68 |
| 5.14 | From left to right, sample scenes are visualized via careless overlay, edge overlay, excavation box and multi-view techniques for comparative vertical depth judgment experiment. | 69 |
| 5.15 | Plots for each technique's average results are shown for exocentric experiment. | 71 |

List of Tables

| | | |
|-----|--|----|
| 3.1 | A summary of inertial sensors and their capabilities. L and R denote, Linear and Rotational, respectively. C denotes complementary support. . . | 20 |
| 3.2 | GPS error causes and their effects are reported | 21 |
| 3.3 | Overview of simulations. | 23 |
| 3.4 | b^m and f^m are the values for a transition model with four visualization modes. Threshold values are given in meters. The tracked object has 8m width and depth and 25m of height. | 39 |
| 4.1 | Spatial components of an annotation are summarized. | 45 |
| 4.2 | Semantic Components of an annotation is visualized. | 46 |
| 5.1 | Participants are asked to order two non-intersecting underground pipes. Values represent the percentage of correct ordering for each technique. . . | 70 |
| 5.2 | Percentages of participants' distance estimations over techniques are given. Careless and edge overlay techniques are dominated with overestimated answers where dig box answers show underestimation. Majority of participants in multi view technique gave | 71 |
| 5.3 | Percentages of participants' distance estimations over techniques are given. Careless overlay, edge overlay and multi view techniques tend to have overestimated results in general. On the other hand, dig box answers are more likely to be underestimated. | 72 |
| 5.4 | Sample use cases are shown for X-ray visualization techniques. | 72 |

List of Abbreviations

| | |
|------|--------------------------------|
| AR | Augmented Reality |
| F+C | Focus and Context |
| GPS | Global Positioning System |
| GUI | Graphical User Interface |
| HMD | Head Mounted Display |
| HUD | Heads-Up-Display |
| UMPC | Ultra Mobile Personal Computer |
| VR | Virtual Reality |

Chapter 1

Introduction

At first, Augmented Reality (AR) was an exclusive research area for military and academic studies. In these early days, largely due to the equipment availability and cost, AR was only investigated through these channels for almost three decades. However, by the end of 2012, the number of smartphones in circulation has grown to roughly one billion, meaning there are at least one billion readily available AR capable devices [1]. This should have formed an AR industry for commercial as well as professional applications. Nonetheless, if we look into currently available AR applications in two dominant market spaces, namely Apple's App Store and Google's Play Store, we find the total number of AR applications to be around 7500 out of 1.7 million total [2, 3].

One question to consider is; "Why has AR not achieved its potential traction within smartphone community?" One answer to this question is the range of available applications. Currently, most of the available applications are either for advertorial or gaming purposes. Even though these applications are popular, they are very short lived since they offer limited replay value or lose impact factor after a few iterations. We are interested in AR approaches that help users to perform their daily duties faster and better. For this purpose we focus on professional fields where there exist tasks that can be improved through AR.

Another possible answer is the technology is not well investigated towards user experience and needs. Especially if we consider aiding professionals in the field. Only four out of 85 academic publications in the last year were targeted around aiding professionals, in major AR related venues [4, 5].

We believe careful investigation of AR capabilities and applicable fields can produce

optimal experiences that may help AR to gain traction within professional community. To further investigate this topic, we employ a bottom-up approach and start with a simple question:

What kind of AR technology is most suitable for a given scene?

In outdoor AR context, a scene can be decomposed into several components:

1. a) Device
2. b) Target Object(s)
3. c) Task

As a) devices we consider current generation smartphones, tablets and ultra-mobile personal computers (UMPC). The first two are readily available for the majority of population in developed countries. The latter offers more extendibility and the availability to include external hardware.

b) Targets are abstract concepts for objects of interest. A target can represent anything from a building to a pipe that is buried underground.

In Chapter 3, we look into the relation between a) Device and b) Target(s) through intrinsic errors included in the localization process. Specifically, inertial sensor and Global Positioning System (GPS) errors are studied. The investigation and discussion in this chapter aim to shed some light onto underlying behaviors of Outdoor AR challenge.

The error analysis for the localization process is heavily investigated in the literature [6, 7]. However we are interested in the visual impact of these errors and their acceptable ranges. Our findings point out that there may be an optimal region for competing tracking approaches. In example, visual tracking is most suitable in ranges from 5 meters to 45-60 m depending on the geometry of targets, on the contrary sensor based tracking becomes more viable as the distance between the object and user increases, in our findings the ideal region for sensor based tracking is for distances over 75-100m.

Also in Chapter 3, through a case study we propose a platform that can localize and track targets using either visual or sensor based approaches. Multiple objects exist in the same scene and tracked with optimal approach utilizing a transitional model.

Arguably c) Task is the essential component for defining requirement sets for the outdoor AR. We hypothesize AR technology should be task oriented and now through the analysis in Chapter 3, we have some inner knowledge of the related factors. In order to investigate these findings further we define two diverse tasks for aiding professionals in the field:

1. i. Modelling and Annotation
2. ii. Exploration and Measurements

Task i) is designed for building-sized objects that are 50-75m away and detailed in Chapter 0. Through this task we first analyze rapid modeling approaches for buildings. Then we focus on annotation creation and visualization. Due to the nature of the task and distances involved, we opted to use a sensor based localization approach.

We propose a novel rapid image-based modeling approach that uses building block analogy. We also introduce volumetric and temporal annotations. By a preliminary user study we confirm our approach is rapid and takes less than 15 minutes from scratch to an annotated object.

On the other hand, task ii) is designed for smaller objects, specifically underground pipe networks where the pipes are 10-50cm in diameter and discussed throughout Chapter 5. This task is aimed at visualizing obstructed objects that cannot be seen with the naked eye. Upon locating these objects, we discuss how to explore and measure targets through a multi-view visualization approach.

For this task we have implemented several state of the art X-Ray visualization methods as well as our proposed multi-view method. Via a user study we define requirement sets for the optimal usage of these techniques. To best of our knowledge, this is the first user study that investigates the vertical depth judgments for X-ray visualization. Our user study reveals that there are situations even the simplest visualization technique can be useful, and provide insight when to use more complex methods.

The main contribution of this thesis is providing AR oriented workflows for selected professional fields specifically for scene creation and exploration purposes, through case studies as well as analyzing the relations between AR scene components. Each case study is targeted towards varying real life requirements and outdoor AR challenges. We have chosen the case studies as diverse as possible in order to cover a wide range of professional tasks.

Modeling objects, viewing and editing annotations, exploring existing scenes and making measurements of hidden objects are tasks applicable to a wide variety of professional fields such as: Archaeology, architecture, cultural heritage, tourism, stratigraphy, civil engineering, and urban maintenance. We believe AR has a place in the professional workflow, however this can only be achieved with careful analysis of scene components.

Chapter 2

Related Work

2.1 Brief Introduction to Augmented Reality

Augmented reality (AR) is a live, direct or indirect, view of a physical, real-world environment whose elements are augmented by computer-generated sensory input such as sound, video, graphics or GPS. It is related to a more general concept called mediated reality, in which a view of reality is modified [8].

The AR technology functions by enhancing one's current perception of reality. Some researchers insist definition of AR should also involve interactivity of the user [9]. However this feature is present in the most recent applications and studies of the field. In this section we will first examine the earliest and primitive examples of Augmented Reality and make our way into the modern definition. AR related studies up to 2001 are examined in Azuma et al.'s work [10, 11].



Figure 2.1: A rear view mirror found in almost every automobile since 1914

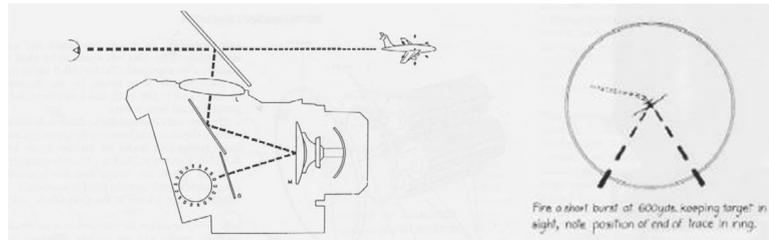


Figure 2.2: Gyro gunsight display system from Spitfire Mk V fighter plane



Figure 2.3: Heads-Up-Display system of an F/A-18C

Figure 2.1 shows an image of a rear view mirror found in almost every terrestrial vehicle since its introduction into manufacturing process in 1914 [12]. A rear view mirror exist in user's, in this case driver's, viewport and supplies additional information about the real world that the driver normally cannot see, improving road safety conditions and possibly avoiding accidents. Although simple, the device is capable of enhancing one's field of view with additional information, hence can be seen as of the earliest usage of this technology in commercial hardware. However the additional view is not generated by a computer.

In order to find first digital images used in augmented reality we have to look into military research field. One of the earliest usage in this sense can be found in military aircrafts from 1950s. Figure 2.2 demonstrates a gyro gunsight display system mounted on a Spitfire Mk V [13]. A gyro gunsight is a reflector sight that visualizes the amount of aim-off and bullet drop due to plane's angular rotation. These reflector sights allowed gunners to see the actual paths the bullets would follow as can be seen in the right side of Figure 2.2. Gyro gunsights generally used analog technology and utilized an electrically controlled camera and projector setup.

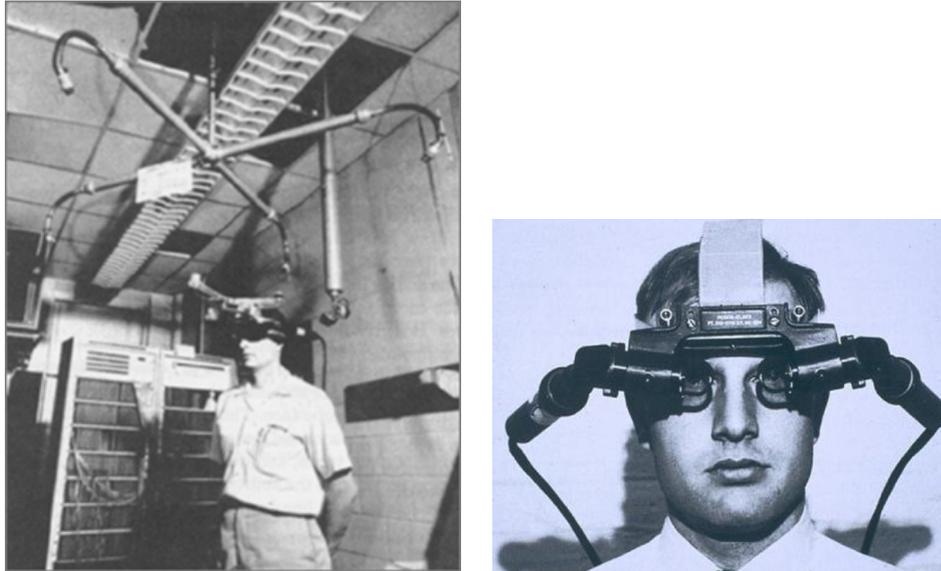


Figure 2.4: Sutherland et al.'s Sword of Damocles setup.

Next iteration of gyro gunsights were heads-up-display (HUD) systems. A HUD is a transparent visualization system that overlays computer generated images. In military aircraft sense these contain the information of gyro gunsights as well as additional aviation related statistics. In Figure 2.3 a modern HUD system can be seen. Although first found in military aircraft, by 2000s almost every commercial aircraft contains HUD systems [14].

In academics Sutherland's early work opened up several research possibilities and can be cited as the earliest Virtual Reality (VR), AR and Head Mounted Display (HMD) system in academia [15], photographs from the device can be seen in Figure 2.4. Though the 'Augmented Reality' term was not coined until 1992 by Caudell et al [16].

Following Sutherland's work Feiner et al. introduced the first mobile AR system [17]. Figure 2.5 shows Mobile Augmented Reality System (MARS) in the field. The system contained an HMD, a laptop computer (in a backpack) and several sensors. This study also marked another important milestone as making the system mobile, Feiner et al. allowed AR systems to be used outdoors.

In 2000s, AR hardware had experienced a shift of focus with the availability of camera phones. Since camera phones became widely available in a very short period of time, this development led to wide spread awareness and deployability of AR. In Figure 2.6, one of the earliest examples of an AR application on a smartphone can be seen. This setup featured an additional inertial measurement unit attached to the back of the phone, as can be seen in Figure 2.6a.



Figure 2.5: MARS is one of the earliest examples of mobile AR.

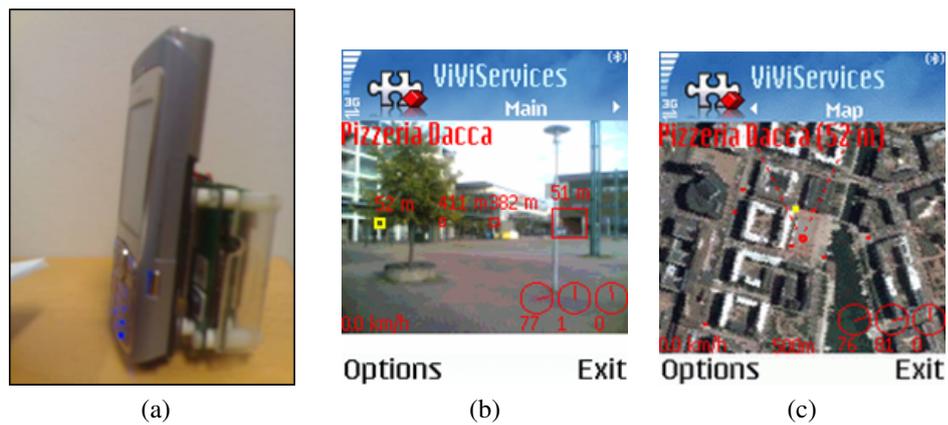


Figure 2.6: Nokia's Mara application is one of the earliest AR applications on a camera phone.



Figure 2.7: Left: Wagner et al.'s method can be used to guide users in taking photos of the environment to be later stitched with a desktop application. Right: An Annotated panorama.

Today, many smartphones and tablets are capable of performing AR tasks without additional hardware. Not only have manufacturers included inertial sensors such as accelerometers and gyroscopes as well as magnetometers, current devices also have enough CPU power to process and generate images in real time.

2.2 Tracking

In Chapter 3, we discuss localization and tracking approaches, recent studies in these and related fields can be separated into three; visual tracking, sensor-based tracking and hybrid tracking.

2.2.1 Visual Tracking

Tracking algorithms have been a research area of interest for almost 25 years. There are several fundamentally different techniques, with respective advantages and disadvantages. Some vision based systems can offer up to millimeter accuracy when tracking. However most of these algorithms require careful setup of environment, such as deploying markers and artificial light sources for better registration and tracking [18].

There are also natural feature based tracking systems that can operate without fiducial markers [19, 20]; in this case the algorithms require a priori knowledge of the tracked object, such as a wireframe model or a texture. Additionally, a detailed survey on visual tracking can be found by Yilmaz et al [21].



Figure 2.8: Layar and Wikitude applications on Android OS are showed side by side.

Figure 2.9: A hybrid tracking flow for outdoor Augmented Reality

Panoramic tracking has become an area of interest in the recent years. In this approach, first a 360° panoramic view is created in a predetermined location [22]. Then users can view annotations on live video tracked over features extracted from original panoramic images. Panoramic image creation and annotation editing can be seen in Figure 2.7. More recently, Langlotz et al. introduced an improvement to map annotations created in one panoramic image to another, reducing the limitation of predetermined locations [23].

2.2.2 Sensor Based Tracking

Fully sensor dependent systems utilize accelerometer and magnetometer sensors, also referred as IMUs. Recently gyroscopes have been made available in mass production for current smartphones. Several algorithms can take advantage of this sensor [24].

Many commercial applications opted to use sensor based tracking for AR visualization [25, 26] (see Figure 2.8). Sensor based-tracking does not require an organized environment, thus allowing large scale applications. These applications acquire targets' GPS coordinates from existing databases and places them into viewport. Sensor based systems are reliable as they are generally not effected by environmental conditions; however they lack the precision of vision-based tracking.

2.2.3 Hybrid Tracking

Hybrid systems that combine inertial sensor data with vision input use complicated filters to assist vision tracking [24, 27, 28], as supplying control signals such as “fast rotation”

to tracking algorithm [29]. The algorithm accepts this input and behaves accordingly. In many studies, the vision algorithm is only activated when rotational speed is below a predetermined threshold as seen in Figure 2.9.

2.3 Modelling and Annotations

Chapter 4 discusses a novel modeling and annotation workflow. Recent studies in these fields are as the following:

2.3.1 Modeling

Modelling of objects is a well-researched topic of both computer graphics and vision. Geometric models can be created from scratch or sampled from real objects using a number of techniques.

Many commercial 3D modelling packages support image based modelling tools, such as Blender and Maya [30, 31, 32]. These packages often support using top, side and front photograph views as superimposed over the model. There are also fully automated solutions based on computer vision techniques for creating models out of sets of images [33]. However these are prone to artifacts caused by vision algorithms when fed with noisy or under exposed images. In order to deal with these artifacts researchers adopted semi-automated processes such as PFTRACK and Voodoo [34, 35, 36]. These approaches allow some user interaction; i.e. letting users to manually mark corresponding features.

VideoTrace by van den Hengel et al. [37] is an improvement over semi-automated processes as it supports user interacted geometry creation, however it requires users to work within the VideoTrace environment. Like VideoTrace, Sinha et al.'s system makes use of the underlying sparse reconstruction, moreover they utilize vanishing directions [38]. Recently Thormählen and Seidel presented an ortho-image based solution for creating high quality models without forcing modellers to leave their desired modelling environment [39]. Other vision-based methods use large geo-tagged photo sets to generate textured 3D models of buildings [40, 41, 42].



Figure 2.10: VideoTrace application allows video-based rapid modelling via user interaction.

2.3.2 Annotations

Annotating real objects is heavily investigated under Augmented Reality (AR). Feiner et al. [17] and Rekimoto and Nagao [43] were early works used AR to annotate the real world with overlaid textual labels. Although a 3D model is generally used to place annotations, Snavely et al. [41] used a system to transfer annotations from one image to another. Recently Wither et al investigated annotations in outdoor augmented reality domain [44, 45]. Another outdoor AR work; by Schall et al. [46], introduced an annotation authoring tool which creates 2D information labels in 3D coordinates.

Visualization of annotations is also a popular research topic. Annotations can be associated with a 2D point [47] or a 3D position [48] depending on the application. Generally if the virtual camera is mobile, the 3D approach is preferred.

2.3.3 In Field Studies

Our modelling approach is inspired by image-based methods. Similar approaches have been utilized by Piekarski [49] to create object models in the field using a backpack based system known as Tinmith-Endavour. MARS is another backpack based system which also includes a hand-held device to annotate and view merged environments [50]. To author physical models, Baillot et al. [51] used mobile computers by generating 3D models from floor plans via user interaction. Backpack-based approaches offer computing power as well as centimeter accurate GPS sensors. Although a backpack-based computer was required for these tasks in the past, currently hand-held computers are capable of perform-

ing even more complicated tasks [52]. A recent work by Schall et al. [46], focuses on displaying pre-defined 3D models to aid civil engineers using hand-held mobile devices. For on-site archaeological studies Benko et al. [53] provided collaborative mixed reality visualization following data recording and archiving principles defined by Harris [54].

2.4 Exploration and Measurements

Chapter 5 discusses exploration and measurements in mobile AR context as perceived by users. Recent studies in these fields are as the following:

2.4.1 Perception

Perception is recognition and interpretation of visual sensory stimuli to understand depth [55]. The human visual system utilizes multiple depth cues to derive a vivid three-dimensional perceptual world from two-dimensional retinal images of a scene [56]. Landy et al. describe this procedure as cue theory and explain how depth cues interact and combine with each other [57]. Lappin et al. explains the influence of context to perceived distances by experimenting in different indoor and outdoor settings [58].

The notion of depth perception is studied extensively in AR and VR [59, 60]. Jones et al. provide a comparative analysis of egocentric depth perception between real world, VR and AR [61]. They report that conventional under estimation problem is considerably low in AR. Livingston et al. compares AR depth perception in outdoor and indoor settings and analyze the effects of supplying user with linear virtual depth cues [62]. They report that although they found evidence for conventional under estimation problem in indoor, subjects over estimate depth values at outdoors.

2.4.2 X-Ray Visualization

X-ray visualization techniques are used for viewing occluded objects while preserving important features in an AR scene. Exploding diagrams, ghosting and cutaways are examples of such techniques. Bane et al. propose several tools of X-ray vision to be used in AR context [9].



Figure 2.11: Edge overlay technique as used by Avery et al.

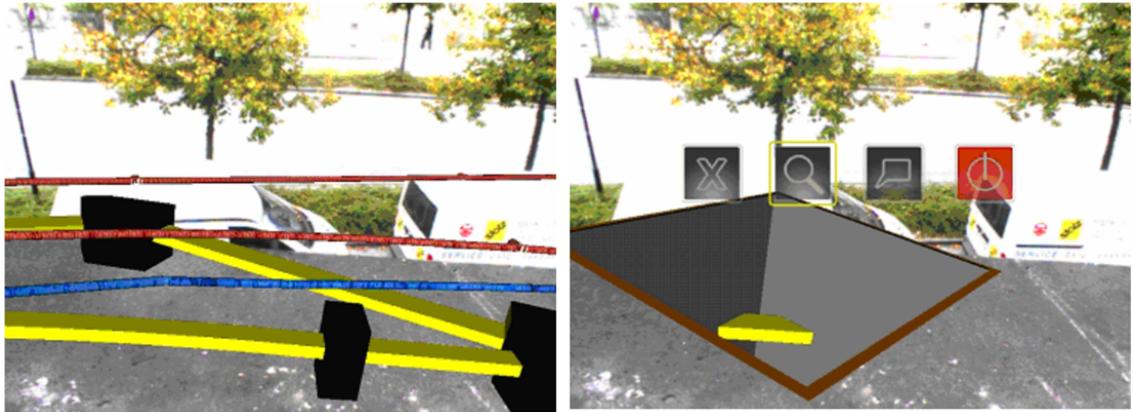


Figure 2.12: Left side shows a naïve approach and right side demonstrates excavation box approach for visualizing underground structures.

Avery et al. discusses how overlaying edge features of the occluding structure would give better depth cues to the viewer and describes three tools for further improving spatial perception [63]. In our proposed multi-view technique, we used a similar approach for promoting sense of occlusion for subterranean structures. There are a number of X-ray visualization techniques addressing subsurface occlusion problems.

Shall et al. introduces an excavation tool inspired from magic lens techniques [64] that virtually digs the ground letting viewer to see underground pipes [65]. This technique requires viewer to be close to the location to effectively perceive the hidden structure (see Figure 2.12). In other words it suffers from the long-flat view problem described in [66].

Zollman et al. employs ghosting techniques for solving single layer occlusion problems between the surface and the infrastructure system [67] (see Figure 2.13). Panoramic images are used from the viewed site for calculating a ghostmap, then use features on this map to preserve the above ground context. Although they demonstrate occlusion clearly for a single layer of subsurface system, in the real world subsurface systems may consist of multiple layers that are occluding each other.

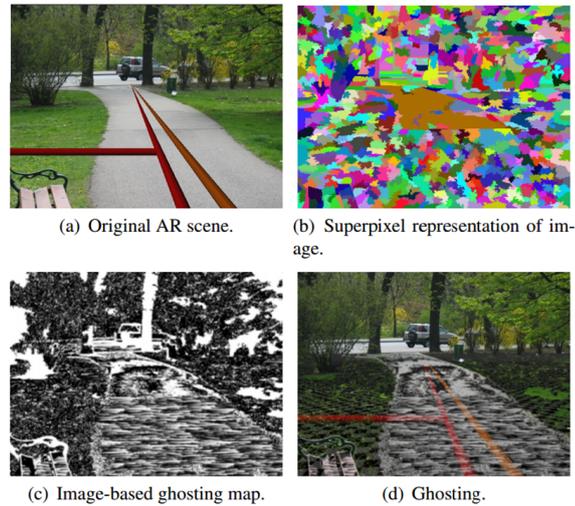


Figure 2.13: Ghosting effect is utilized in purpose of X-Ray visualization by Zollman et al.

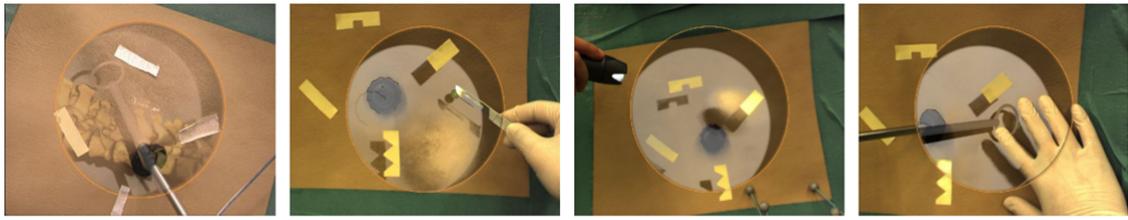
In addition, Livingston et al. proposed an algorithm that solves multi-layer occlusion problem on Z axis by changing the opacity values of virtual objects [68]. The attacked problem is similar to ours but on a different domain of distinguishing occluded buildings. In our work instead of modulating opacity values of virtual objects, we employed a second view to explicitly indicate separate layers.

Furthermore, Dey et al.'s work on X-Ray vision for navigation discusses this technique in outdoor AR context [69].

2.4.3 Focus and Context

Focus-Context (F+C) paradigm is described as using visual tools to separate the center of attention (focus) from the surroundings (context). X-Ray vision illustrations coming from Superman Comics also cherish this paradigm [70]. Kruger et al. describes some X-ray techniques that human artist uses to give shape and depth clues in their technical drawings of hotspots and discuss motivations behind their practices. Furthermore, they present a computer graphics technique called ClearView that makes use of the curvature, distance and view distance features of a model to achieve similar results to human artists [71].

In a similar work Bichlmeier et al. describes ways to improve depth perception in Medical Augmented Reality [72, 73]. A sample image from this technique can be seen in Figure 2.14. They make use of the magic lens techniques for information filtering and



(a) Deformation of the skin (b) Extracted instruments cast shadow (c) Flashlight causes context information. (d) Hand causes shadow feedback.

Figure 2.14: In order to improve the depth perception, magic lens approach can be utilized.



Figure 2.15: Occluding augmentations. A typical AR visualization to draw the user's attention with overlaid semitransparent geometry, occluding the object of interest.

viewing relevant parts of the inner body, which is described to be in-situ visualization, while viewer may still keep track of the real body [74]. They are using curvature, angle of incidence factor and distance falloff features of a model to attack floating effect problem similar to [71]. Methods described in these works require detailed models for calculating curvature values and very precise tracking. These methods are not applicable for our case, since models are not always present in Outdoor Augmented Reality and tracking may not be as accurate.

Kalkofen et al. uses edge features to give occlusion clues and discusses the F+C paradigm's importance in Augmented Reality scenes (see Figure 2.15). They discuss techniques used to render occluded objects and demonstrate how information filtering for Augmented Reality can be achieved through magic-lens techniques [75]. Similarly our perspective-view uses a clipping sphere and orthographic-view showing features within the defined frustum, for filtering unnecessary information.

Chapter 3

Devices and Target(s)

In order to perform the simplest AR visualization, we are required to perform localization both for the user and the target object. In this chapter we will examine possible solutions and discuss their advantages and shortcomings. A recent survey on outdoor AR gives an introduction to the challenges in the field [76]. For outdoor AR context we will specifically examine inertial sensor and vision based approaches.

3.1 Orientation Sensors on Mobile Devices

With the rise of smartphones and tablets the outdoor Augmented Reality Field had been blessed with variety of hardware choices. Many commercial devices are capable of performing complicated outdoor AR tasks without additional hardware. With every passing year the hardware becomes more powerful and capable. However there are some critical components that does not get upgrades either due to their cost or due to the technical limitations of their underlying systems. Additionally, when utilizing sensor based tracking, hardware is significantly important for localization in the field. For this purpose we are specifically interested in orientation and GPS sensors. We will first examine the orientation sensors.

An orientation sensor generally refers to a combination of inertial sensors that are present in mobile devices. These sensors can be broken into an accelerometer, a gyroscope and a magnetometer.

3.1.1 Accelerometer

As the name applies the accelerometer is responsible for registering devices' relative orientation with respect to earth's gravity represented as a 3D vector of acceleration. A device with its default orientation on the surface of earth would register an upward force of 9.81m/s^2 . This upward force is the counterforce due to its weight as response to the earth's gravity. Some devices may produce normalized results. Using a multi-axis accelerometer it is possible to sense magnitude and direction of proper acceleration as a vector quantity and this vector can be utilized to represent an orientation for the device.

3.1.2 Gyroscope

Digital gyroscopes can be found in current high end smartphones. Where accelerometers record linear acceleration, gyroscopes record the angular rate of motion. Currently, MEMS (Microelectromechanical System) gyroscopes are utilized in commercial hardware. The underlying principle is that a vibrating object will tend to continue vibrating in the same plane as its support rotates.

Consider two proof masses vibrating in plane (as in the MEMS gyroscope) at frequency ω . Recall that the Coriolis Effect induces an acceleration on the proof masses equal to $a_c = -2(v \times \Omega)$ where v is a velocity and Ω is an angular rate of rotation. The in-plane velocity of the proof masses is given by: $X_{ip}\omega_r\cos(\omega_r t)$, if the in-plane position is given by $X_{ip}\omega_r\cos(\omega_r t)$. The out-of-plane motion y_{op} , induced by rotation, is given by:

$$y_{op} = \frac{F_c}{k_{op}} = \frac{(2m\Omega X_{ip}\omega_r\cos(\omega_r t))}{k_{op}} \quad (3.1)$$

where m is a mass of the proof mass, k_{op} is a spring constant in the out of plane direction, and Ω is a magnitude of a rotation vector in the plane of and perpendicular to the driven proof mass motion.

Contemporary mobile device APIs [77, 78] do not allow direct access to gyroscope readings. These measurements are in fact used to correct drift errors and support magnetometer readings in hardware level.

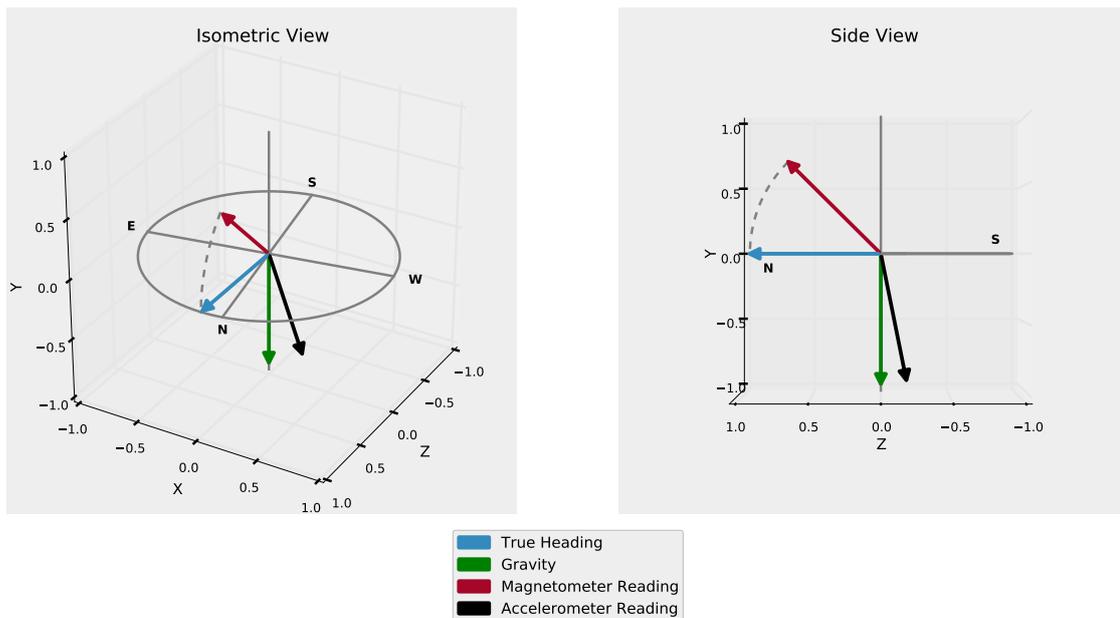


Figure 3.1: Calculation of true heading from magnetometer and accelerometer readings.

3.1.3 Magnetometer

A magnetometer is a measuring instrument used to measure the strength and, in some cases, the direction of magnetic fields. Vector magnetometers have the capability to measure the component of the magnetic field in a particular direction, relative to the spatial orientation of the device.

Magnets, metallic objects and metal ores in the ground cause interference in magnetometer readings. This issue almost makes acquiring a correct heading indoors impossible.

Magnetometers are critically important for localization purposes since they provide a heading of viewer with respect to earth's magnetic field. Magnetometers provide a 3-axis vector reading as can be seen in Figure 3.1. To compute a true heading from this vector, we also have to utilize the accelerometer reading. The true heading is this vector's projection to the plane that is perpendicular to the accelerometer vector.

3.1.4 Summary of Orientation Sensors' Capabilities

Capabilities of orientation sensors that can be found in contemporary mobile devices are summarized in Table 3.1. For localization purposes we are interested in sensors that pro-

| Sensor | Capabilities | | | | |
|---------------|-------------------|-----|-------------------------|-------------------------|--------------|
| | Relative to Earth | L/R | Sensitivity | | Availability |
| | | | Specs. | Measured | |
| Accelerometer | ✓ | L | $\pm 0.2 \frac{m}{s^2}$ | $\pm 0.5 \frac{m}{s^2}$ | ✓ |
| Gyroscope | | R | $\pm 1^\circ$ | - | C |
| Magnetometer | ✓ | R | $\pm 1.4^\circ$ | $\pm 15^\circ$ | ✓ |

Table 3.1: A summary of inertial sensors and their capabilities. L and R denote, Linear and Rotational, respectively. C denotes complementary support.

duce readings relative to earth, i.e. accelerometers and magnetometers. We specifically use magnetometer readings with correctional and complementary support from others.

3.2 GPS sensors on Mobile Devices

The Global Positioning System (GPS) is a space-based satellite navigation system that provides location information anywhere on the Earth. GPS devices require an unobstructed line of sight to at least four GPS satellites. The more satellites the device can communicate, the better the accuracy becomes. It is maintained by the United States government and is freely accessible to anyone with a GPS receiver [79]. The network can also provide precise time readings.

Smartphones or other network capable mobile devices generally utilize Assisted GPS in order to reduce fix times. In Assisted GPS, the device first contacts a nearby base station in order to acquire recent satellite information and then communicates with the GPS satellites, enabling the receiver to lock to the satellites more rapidly.

Before January 2000, for civilian usage the system had been reported to have an average error rate of around 100m [80]. This error mainly caused by a currently disabled feature called Selective Availability. The feature introduced intentional, time varying errors to disable enemy usage. Since its discontinuation in 2000, the reported error rate dropped to ~10 m on first fix and measured in real world scenarios to be around 15m [6].

The overall error as seen in Figure 3.2, can be deconstructed to several contributing effects such as:

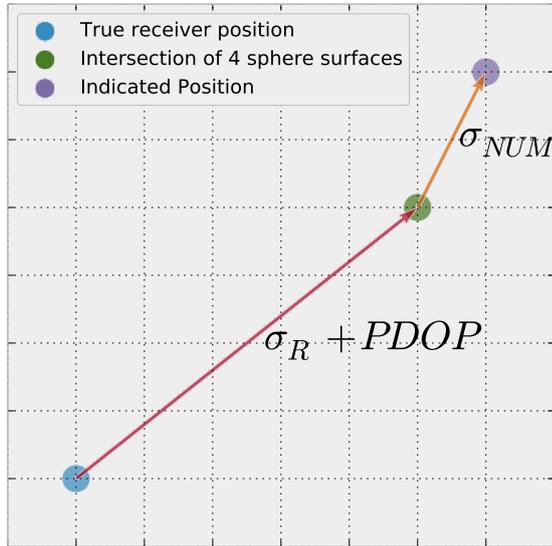


Figure 3.2: GPS inaccuracy visualized.

| Cause | Effect |
|------------------------|------------|
| Signal arrival | $\pm 3m$ |
| Ionospheric effects | $\pm 5m$ |
| Ephemeris errors | $\pm 2.5m$ |
| Satellite clock errors | $\pm 2m$ |
| Multipath distortion | $\pm 1m$ |
| Tropospheric effects | $\pm 0.5m$ |

Table 3.2: GPS error causes and their effects are reported

The standard deviation for a receiver can be computed using the following formula:

$$\sigma_{rc} = \sqrt{(PDOP^2) + \sigma_R^2 + \sigma_{NUM}^2} \quad (3.2)$$

where *PDOP* is the *Position Dilution of Precision*, a value to measure the geometrical dilutions dependent on user location and satellite positions. σ_R is the standard deviation for errors shown in Table 3 2, and can be calculated as:

$$\sigma_R = \sqrt{3^2 + 5^2 + 2^2 + 1^2 + 0.5^2} = 6.7m \quad (3.3)$$

σ_{NUM} is the standard deviation of numerical errors and is assumed to be around $\pm 1m$. The effects in Table 3.2 are given as their standard deviation, and reported as \pm values having an unbiased 0 mean. The analysis and underlying sources for these errors fall out of the scope of this manuscript, and can be found in detail in the related studies[81, 82, 83]. However one important topic to mention is the largest cause for the error is from atmospheric effects. Although the network had been developed for all weather usage, bad weather as well as the ionospheric conditions affect the system's precision. Because these causes are mostly naturally occurring and random, we will assume they have a normal distribution.

GPS sensors require a direct line of sight with several GPS capable satellites. Indoor usage is almost impossible and outdoor usage requires an initial fix step generally taking up to a minute.

3.3 Standard Localization Workflow Using Inertial Sensors and GPS

To project an object into correct screen coordinates, we first need a GPS position for the target object. This information can either be entered manually or can be acquired through an online service such as Google Maps [84]. Secondly, we need to get a lock on GPS satellites to provide a position for the user. Third, using heading calculation devices local orientation as well as true heading is acquired. We are also required to have a camera calibration or field of view values calculated. Then using these values the target object is projected onto the screen via desired visualization technique.

| Target | Simulated Error Type | |
|----------------|----------------------|--------------|
| | Magnetometer | GPS |
| Point | Simulation 1 | Simulation 2 |
| Sphere | Simulation 3 | |
| Building Model | Simulation 4 | |

Table 3.3: Overview of simulations.

3.4 Computational Experiments to Evaluate Perceived Errors

Error analysis for GPS and inertial sensors have been thoroughly investigated by researchers. However, the actual visual artifacts caused by these errors have not specifically been a topic of interest. On the contrary, in an augmented reality context we are more interested in the perceived errors. Through these computational experiments we do not aim to model these errors, however we aim to analyze them in order to make suggestions to handle them correctly.

Specifically, we are interested in the following questions:

- If GPS error is $\pm x$ m, and the user is y m away, how much the projected geometry will be misplaced on screen?
- If orientation sensor error is $\pm x^\circ$, and the user is y m away, how much the projected geometry will be misplaced on screen?
- Considering GPS and orientation sensor errors, does the geometry of the projected object, have an effect on misplacement on screen?

Since we would like to examine localization errors, we will look specifically into errors caused by magnetometer and GPS in accuracies. It is also important to note that we will examine them first separately and then look at the combined effect.

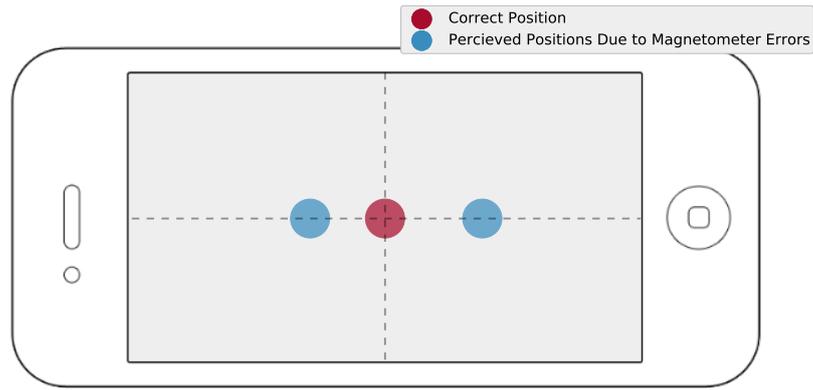


Figure 3.3: A sample handheld device is shown. Red dot denotes the ideal placement of target object without any errors induced. Blue dots are deviated from the result due to the induced errors.

3.4.1 Scene Definition and Overview of Experiments

In order to evaluate the perceived error values we will examine four different scene setups. In the first and the second experiments a virtual point is projected onto screen in order to evaluate the effects of magnetometer and GPS errors separately. Third and fourth experiments are geometrical studies where 3D models of a sphere and a real-life building is being projected onto the screen in order to evaluate the effects of both magnetometer and GPS sensors.

In our experiments we assume the mobile device is held in landscape mode and we report findings in terms of percentage of the screen size. For the first and second experiments, we will analyze only horizontal misplacement for simplification.

3.4.2 Experiment 1:

For this experiment we assume the location of the user and the object is exactly known without any errors. Given this assumption, we try to examine the deviation caused by the magnetometer alone. Average magnetometers in contemporary mobile devices has an error value of $\pm 15^\circ$. The deviation caused by this error can be seen in Figure 3.3 and 3.4.

Since we are interested in the error as perceived by the user. In order to estimate this deviation in projected screen space as in percentage of screen width the following formula can be used.

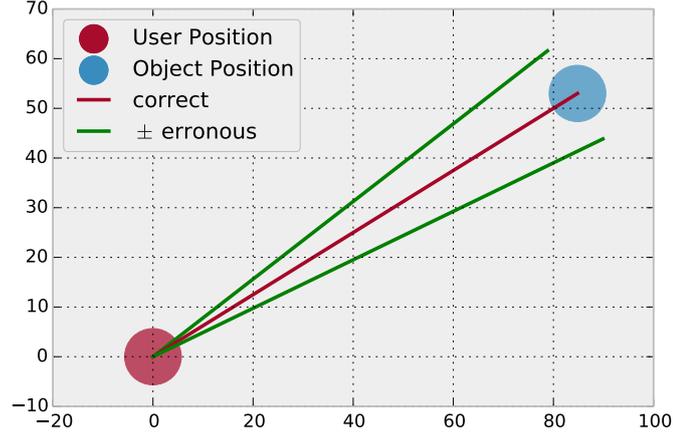


Figure 3.4: Red and blue dots denote the viewer and object positions, respectively. Red line visualizes a correct measurement for user’s viewing direction. The green lines denotes typical errors for a magnetometer measurement.

$$e_p = \frac{e_m}{fov_h} \times 100 \quad (3.4)$$

where e_p is the perceived error, e_m is the magnetometer error and fov_h is the horizontal field of view of the mobile device camera. Field of view for a live image from a mobile device can be computed by the following formula:

$$fov_h = 2 \times atan\left(\frac{s_w}{2 \times f}\right) \quad (3.5)$$

where s_w is sensor width in mm and f is focal length. Sensor sizes and focal lengths can be obtained from technical specifications provided by the manufacturers. As a sample, Samsung Galaxy S2 smartphone has a fov_h of 59.1498° [85]. Figure 3.4 demonstrates the perceived error for this device. Other than a few exceptions many contemporary devices uses similar sensors, this value is valid for most of the smartphones currently sold.

Figure 3.5 shows magnetometer error with respect to distance from object. We inspect these values for two kinds of users, one stationary and one moving. Red line denotes an error along X direction (to the left or right of the screen). The estimated error is around $\pm 10\%$ of screen width, meaning a single dot may be misplaced 10% at most to where it should have been. Similarly the blue line demonstrates the perceived error for a user in motion.

As we can observe from the chart, error values do not change with distance. We

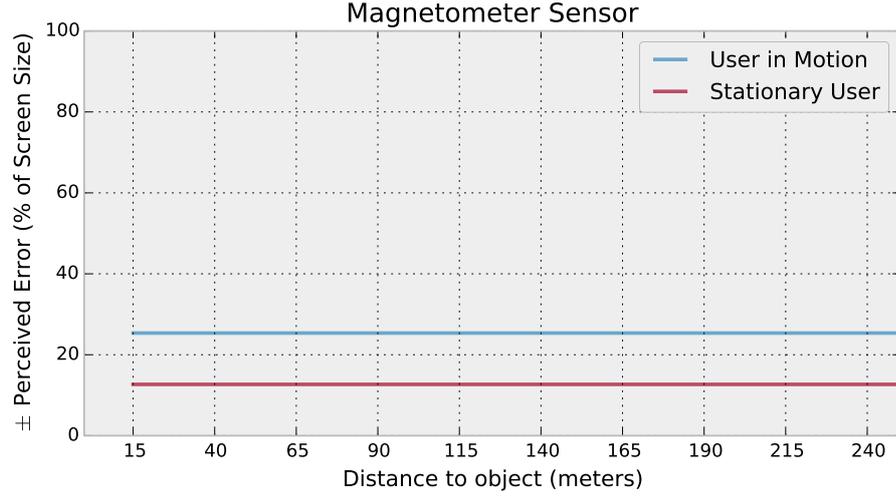


Figure 3.5: Perceived error due to magnetometer inaccuracies is visualized in \pm percentage of screen width.

calculate the angle between the user and the object, using their respective geo coordinates. Then we position them on screen by comparing this angle to user's heading. In other words, we add the erroneous user heading data after calculating the angle between the object and the user. Thus we end up with constant inaccuracy with respect to distance.

3.4.3 Experiment 2:

In the second phase of the experiments, we try to visualize the perceived effects caused by GPS inaccuracy. When the user in motion the average error for GPS is measured to be around 15 meters. For stationary users standard deviation of the error is reported to be around 10m [83]. For each distance (discretized in 1 meter intervals) we project the virtual point on the screen by misplacing the user position with the average error. The following formula can be used for this purpose:

$$e_p = \frac{e_{gps}}{fov_h} \times 100 \quad (3.6)$$

where e_p is the perceived error, e_{gps} is the angular error caused by GPS and fov_h is the horizontal field of view of the mobile device camera. e_{gps} can be calculated as follows:

$$e_{gps} = \text{acos} \left(\frac{v_u \cdot v'_u}{\|v_u\| \|v'_u\|} \right) \quad (3.7)$$

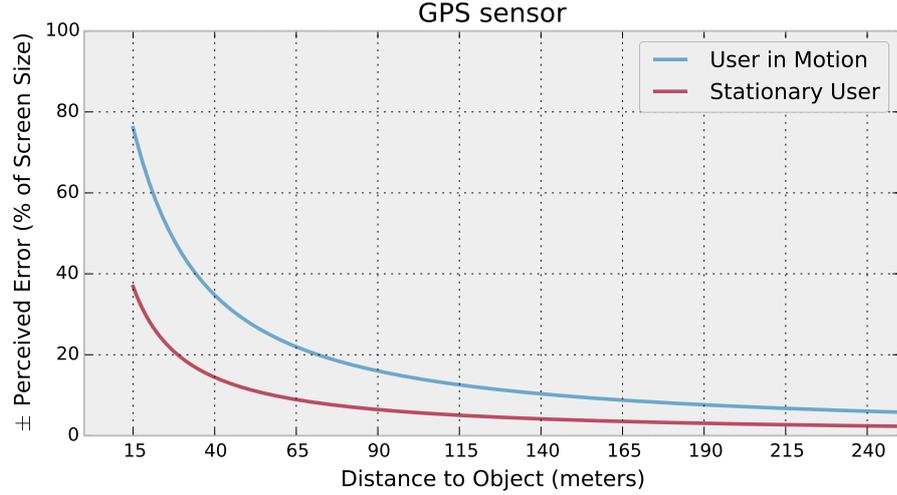


Figure 3.6: Average misplacement due to GPS error per distance is plotted

where e_{gps} is angle between v_u and v'_u , and v_u is the vector between the user and the object and v'_u is the vector between erroneous user position and the object.

Figure 3.6 demonstrates the average misplacement of target object on screen for a user in motion and a stationary user. Unlike magnetometer, perceived misplacements due to GPS errors reduce as the distance between the object and the user grows. At 15m the average misplacements are around 75% and 37% for a user in motion and a stationary user, respectively. Both of these numbers cast localization through GPS almost unusable at 15m. However at around 150m the numbers drop to 10% and 6% respectively. It is possible to visually confirm these results via comparing Figure 3.7a and Figure 3.7b.

3.4.4 Experiment 3 and 4:

Experiment 3 and 4 are different from the first two experiments in the way we examine perceived errors. For the first two experiments, errors were analyzed using a virtual point as a target. For the latter experiments we also want to analyze effect of the geometric properties of target objects. In this purpose we have chosen a sphere and a building model where the sphere acts as a control group.

To observe the visual effects caused by GPS error, we generate a number of deviated user positions using a multivariate normal distribution as:

$$x \sim \mathcal{N}_2(\mu, \Sigma) \quad (3.8)$$

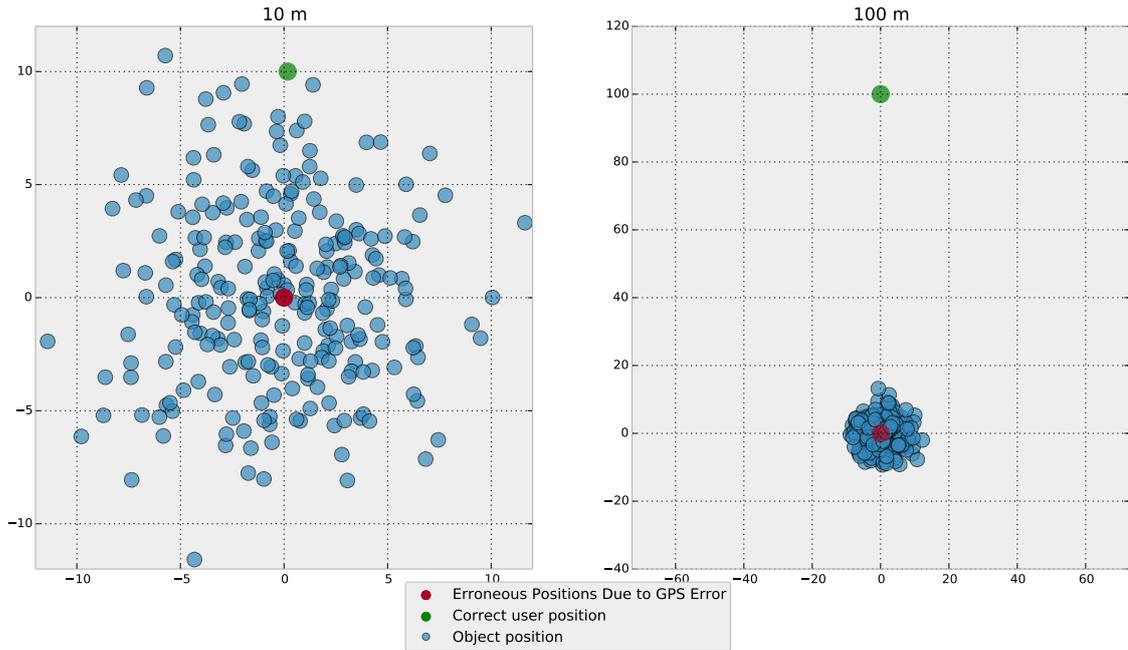


Figure 3.7: Red dot denotes the exact location for the user. Blue dots are the deviated user positions. Green dot visualizes the target object which is 10 m away. Red dot denotes the exact location for the user. Blue dots are the deviated user positions. Green dots visualize the target object which is 10 m (left) and 100m (right) away.

Where μ is a two dimensional mean vector and Σ is a 2x2 covariance matrix used to simulate GPS errors. A sample of generated points can be seen in Figure 3.7a and Figure 3.7b. Figure 3.7a plots deviated user positions where the target object is 10m away from the viewing user and Figure 3.7b plots the same scenario where the object is 100m away.

To project the geometries on screen, we use a virtual camera calibrated to a physical mobile device camera, in this case Samsung Galaxy S2. Using OpenGL, 3D geometries were drawn on screen once for correct position and once for deviated user position. Figure 3.8 simulates a user walking on a straight line towards the object from 255m to 15m. In this figure, we plotted the percentage of projected pixels that are inside the original projection area. For the building model distances further than 75m contain over 60% or more correctly placed pixels. As for the sphere after 75m, the percentage of correctly placed pixels are 82 or more. We suspect the better results of sphere simulation is due to the uniform geometry of the object.

Furthermore, we have analyzed the screen space occupied by the object for each distance step. Figure 3.9 demonstrate these values. In order to visualize the sphere as at least 30% of screen area the user must be at 27m or closer, on the other hand the user must stand closer than 73m for the same projection area.

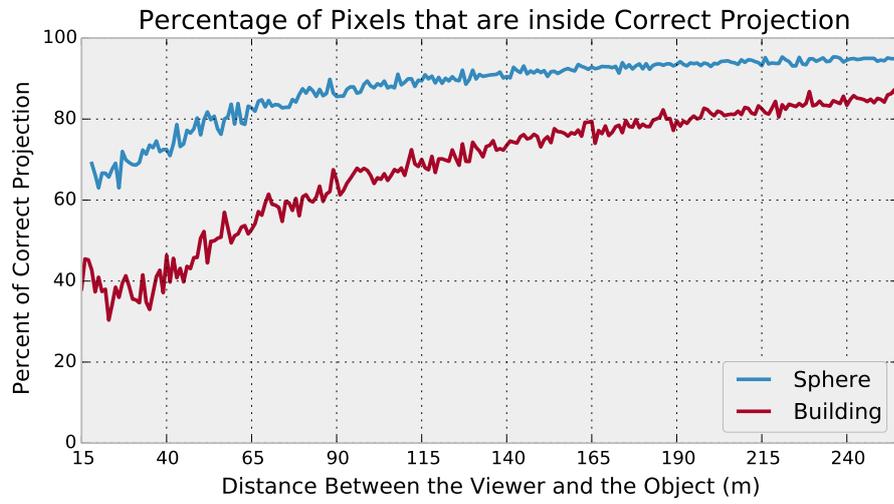


Figure 3.8: A sphere and a building model's correct projection ratio with respect to distance is plotted.

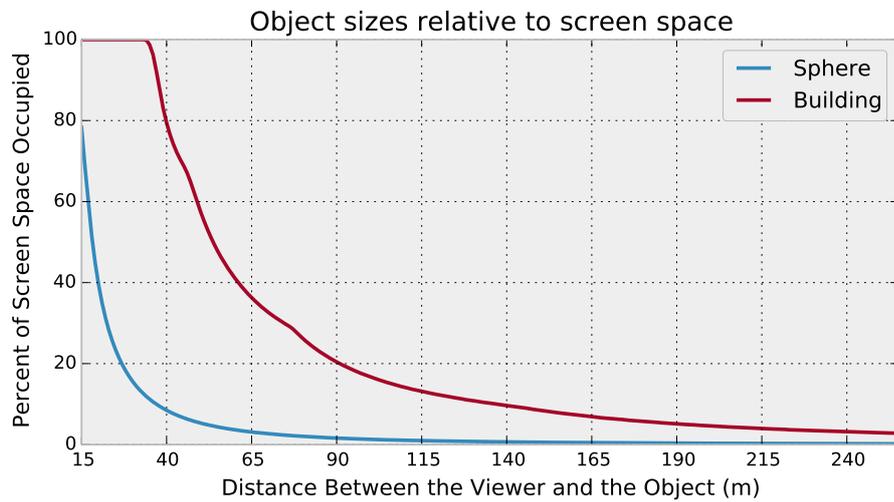


Figure 3.9: Object sizes relative to screen space are plotted.

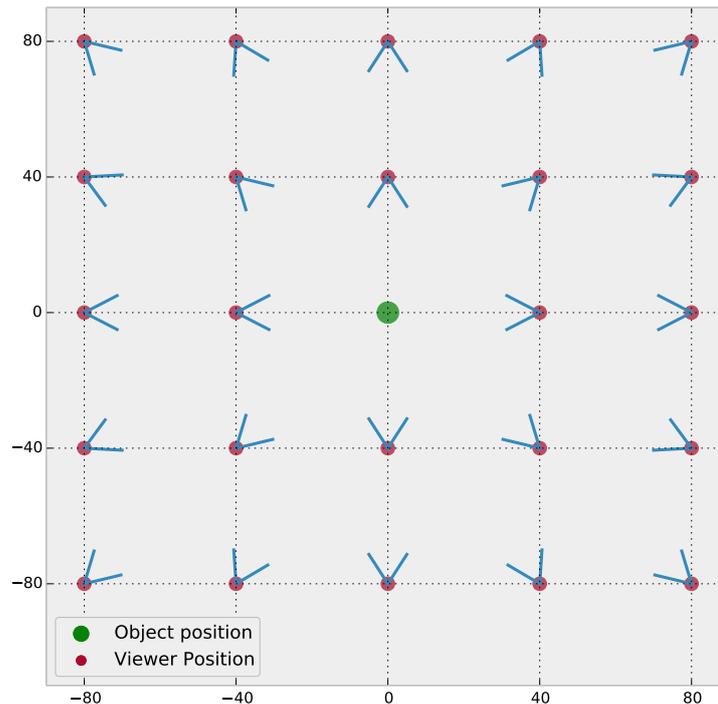


Figure 3.10: To inspect viewing angle’s effect with respect to GPS and orientation errors, several virtual cameras are placed around the object on a grid.

The figures so far demonstrate user moving on a straight line. Due to the geometry of the object the viewing angle as well as the distance may be important. In order to evaluate views from different positions we created a grid around the object and placed virtual viewers on these points as can be seen in Figure 3.10. For each of these positions we generated sample error deviations as we did before, similar to Figure 3.7. For a grid of 512 by 512, ranging from -255 to +255 m in both X and Z axis, our simulation produced Figure 3.11 for the building model. The values are drawn using a color map where outer red region shows highest accuracy and inner red the lowest. The color coded values represent the correctness measure we discussed earlier. When the similar approach is applied to the sphere model, the result is shown in Figure 3.12. The map produced from the sphere model is only affected by distance. Since a spherical object has no distinguishing geometrical factors for different angles, this result was expected. However the building model produced a map with its geometrical features embedded.

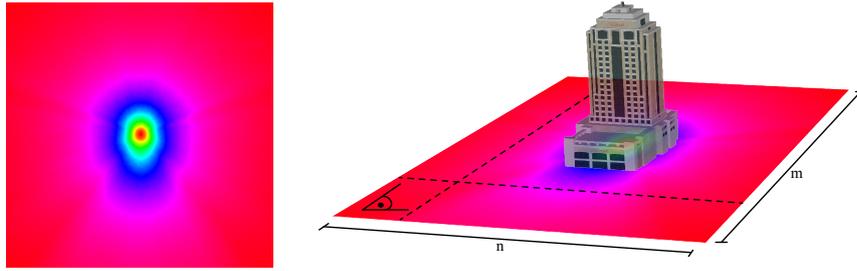


Figure 3.11: A Building model is shown with respect to its color coded correctness map.

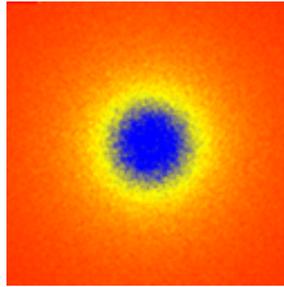


Figure 3.12: A color coded correctness map for the sphere model is shown.

3.5 Hybrid Localization

From user's point of view, the most significant aspect is to be able to observe the desired information. When tracking fails, the user cannot observe nor interact with the information. Robust systems utilizing hybrid tracking were proposed to avoid this problem. In many hybrid tracking systems sensors complement each other to correct noisy or erroneous input. However these algorithms examine the whole scene when deciding to employ different sensors. There may be multiple objects in a scene, and each one may be tracked optimally by a different algorithm.

In case of vision tracking 30 percent screen area is recommended for proper feature extraction [18]. For the GPS/orientation sensor tracking we can set a limit to at least 60 percent correctness. The vertical lines Figure 3.13 visualizes these threshold values. Left of the green line, vision tracking would be most suitable and right of the purple line sensor tracking would produce optimal results. As for the sphere model these threshold values can be seen in Figure 3.14. Since these objects have drastically different geometries and physical sizes the threshold values are also very different. The region between the lines are suited with tracking with both algorithms, in this case we suggest vision based tracking due to higher precision and falling back to sensor based tracking.

By imposing these simple threshold values, it is possible to track multiple objects in a

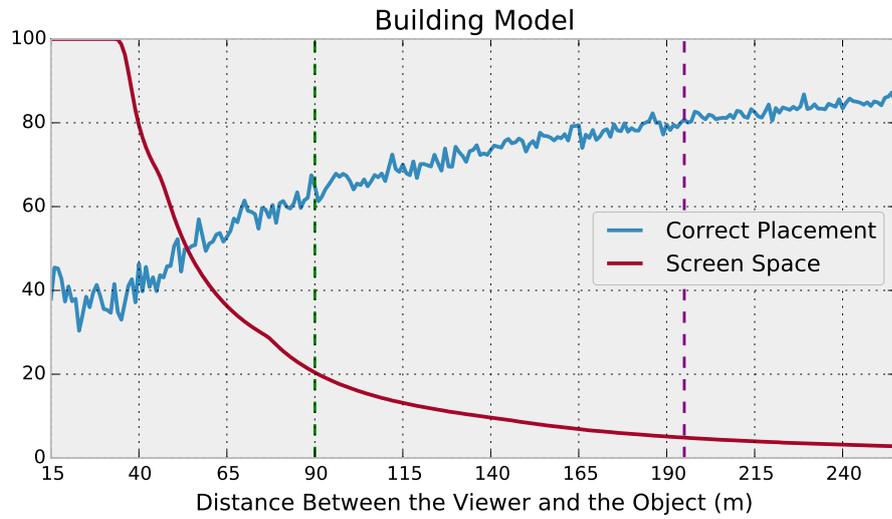


Figure 3.13: Threshold values for orientation sensor tracking and vision tracking is visualized for the building model.

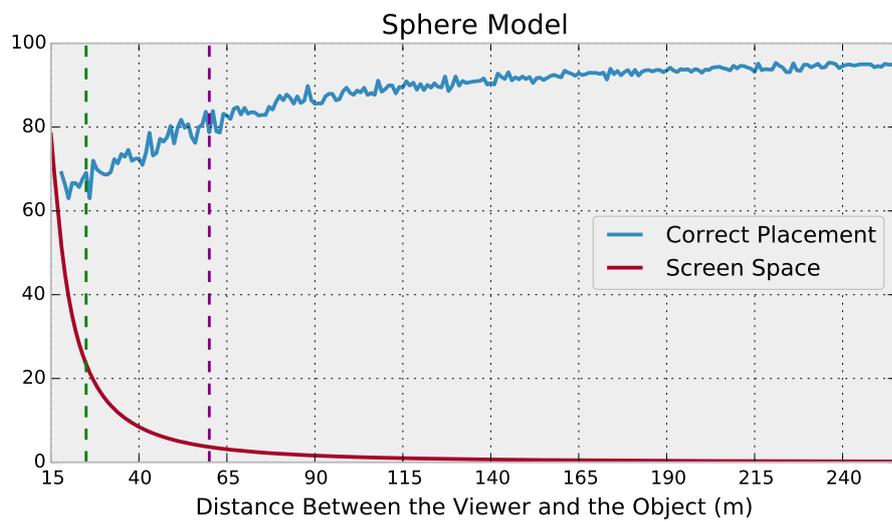


Figure 3.14: Threshold values for orientation sensor tracking and vision tracking is visualized for the sphere model.

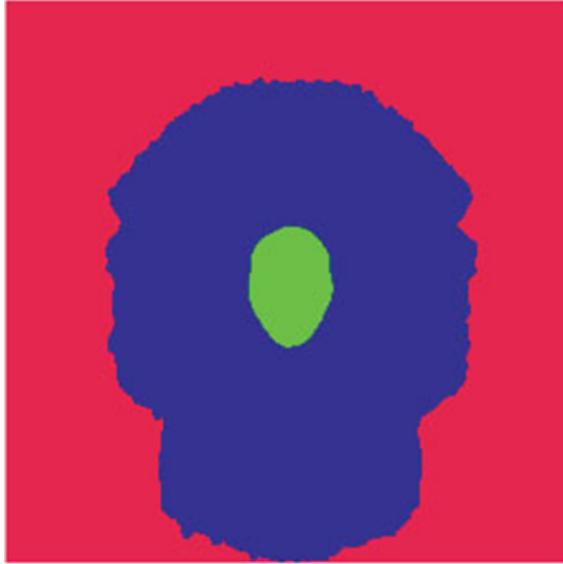


Figure 3.15: Tracking thresholds for multiple viewing angles and distances are visualized. The red region favors sensor based tracking where the green region favors vision based tracking.

scene with the most suited algorithm. We can also impose this threshold values into our color coded views as seen in Figure 3.15.

3.6 Case Study

In order to visualize and exemplify our findings and calculations we have designed a case study that features an outdoor augmented reality application with multiple tracking and visualization modes.

We propose a novel approach to allow both vision and sensor tracked objects to co-exist in the same scene. One of our main goals is to create an object dependent fallback mechanism that can produce visualization at any condition. The proposed hybrid tracking system provides visualization based on viewing conditions. Whenever possible the most precise and detailed visualization is preferred. The proposed transition model works on top of tracking algorithms and provides cues for each object's tracker. These cues include tracking algorithm as well as visualization mode suggestions.

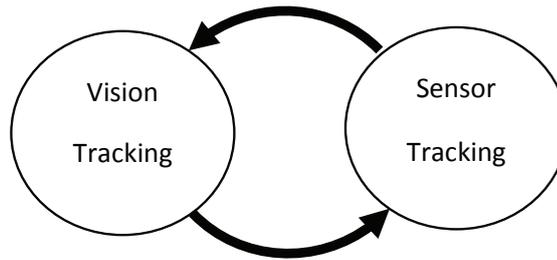


Figure 3.16: General transition diagram for tracking mode switching for each target object.

3.6.1 Transition Model

We are envisioning an AR environment that allows an object to be tracked by separate algorithms and visualized by different methods; In this case; there must be a model to coordinate when and which algorithms should be utilized.

Figure 3.16 demonstrates a general transition diagram between tracking algorithms. The visualization methods we define are bound to some transition criteria as well as to a tracking algorithm. Our transition model is object dependent, rather than scene dependent, as each object has its own transition states and function.

3.6.1.1 Transition Criteria

Transition criteria help us to decide when to switch between tracking and visualization modes. A criterion may be dependent on a specific measurement or a combination of several signals.

Different criteria can be selected to emphasize each tracking method's advantages. Scene lighting conditions acquired by analyzing an image can act as an input to switch from vision to sensor based tracking. As in a poorly lit environment vision tracking would be less likely to find enough image features to perform tracking [86].

Another criterion can be defined as detecting shadows in an image. This criterion would also serve as an input to switch from vision to sensor based tracking [87].

The distance between the target object and the user is also a viable candidate as a transition criterion. There is an inversely proportional relation between the distance and object's projection on screen space. As we move closer to the object, it will cover more

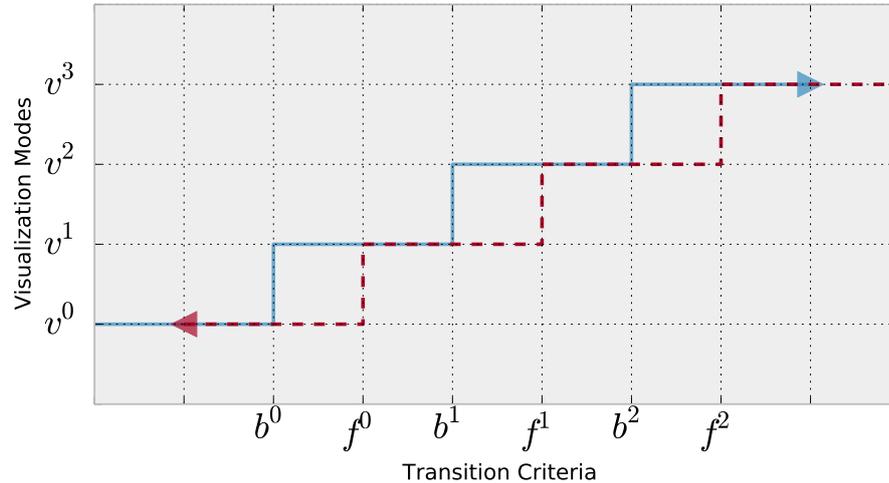


Figure 3.17: Transition function is illustrated. b^n and f^n denotes cutoff values for forward and backwards movement respectively. v^n denotes visualization modes.

screen area, thus allowing vision-algorithm to be able find more features on the object. Conversely, when we are further away from the target object, sensor based tracking starts to work better, as sensor data noise loses its significance. In this study we have selected distance between the user and the object as transition criterion to demonstrate our findings. As we move closer to the object we gradually perform a transition from sensor based tracking to vision based tracking through visualization modes.

3.6.1.2 Transition function

In order to create a seamless user interface we opted to use a transition function that utilizes hysteresis phenomena. As seen in Figure 3.17, we define different cutoff values for moving forward or backwards on the criterion dimension. If we consider our criterion to be distance, moving forward means getting closer to the object. Instead of using hysteresis curves directly, we have defined Schmitt triggers for transitions between consecutive states. Figure 3.17 demonstrates a transition function with four different visualization modes.

3.6.2 Visualization

As different sensors provide vastly diverse precision for tracking, different visualization modes can be suitable for each tracking mode. Even in the same tracking mode, vision or sensor based, there can be defined several distinct visualization modes in order to serve

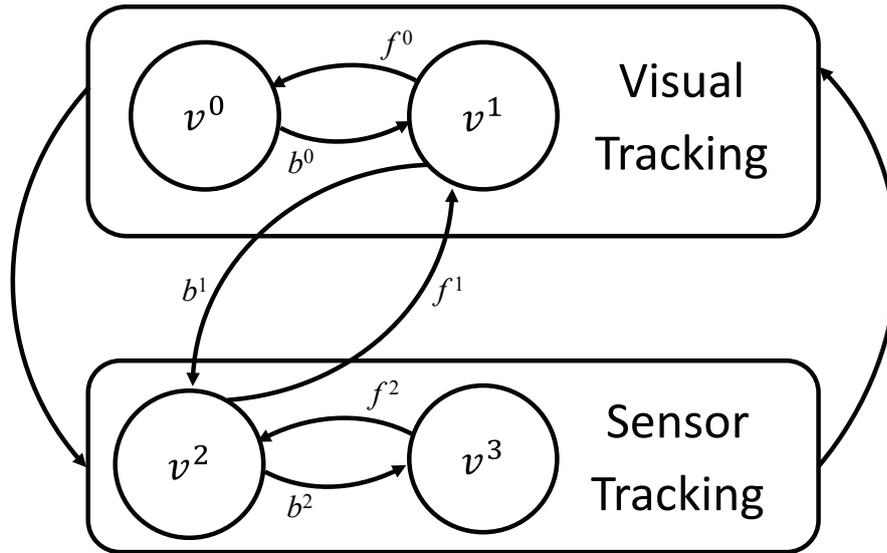


Figure 3.18: Transition Diagram with four different visualization modes.

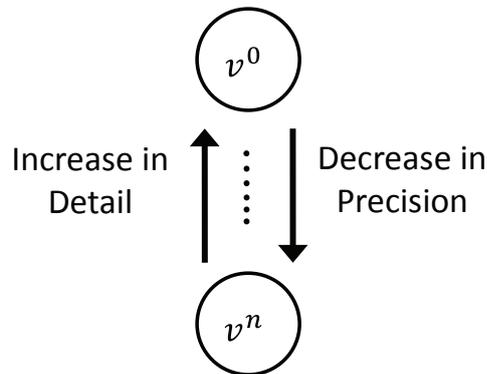


Figure 3.19: Sematic relation between visualization states is demonstrated.

the desired information more appropriately. Figure 3.18 demonstrates a state diagram with four different visualization modes. First two of these modes require vision-based tracking modes. The latter two are only active when tracking is done via sensors. Transitions between consecutive visualization modes or states are allowed. We also allow a nonconsecutive transition between states, when a switch between tracking algorithms is required.

If our transition criterion is distance; when farthest away from the object, v^3 is activated. As the distance gap between the user and the tracked object is reduced, state transitions occur as $v^3 \rightarrow v^2$, $v^2 \rightarrow v^1$ and $v^1 \rightarrow v^0$.

If we have another transition criterion such as scene lighting condition, it may also cause a transitions such as $v^3 \rightarrow v^1$ or $v^2 \rightarrow v^0$.

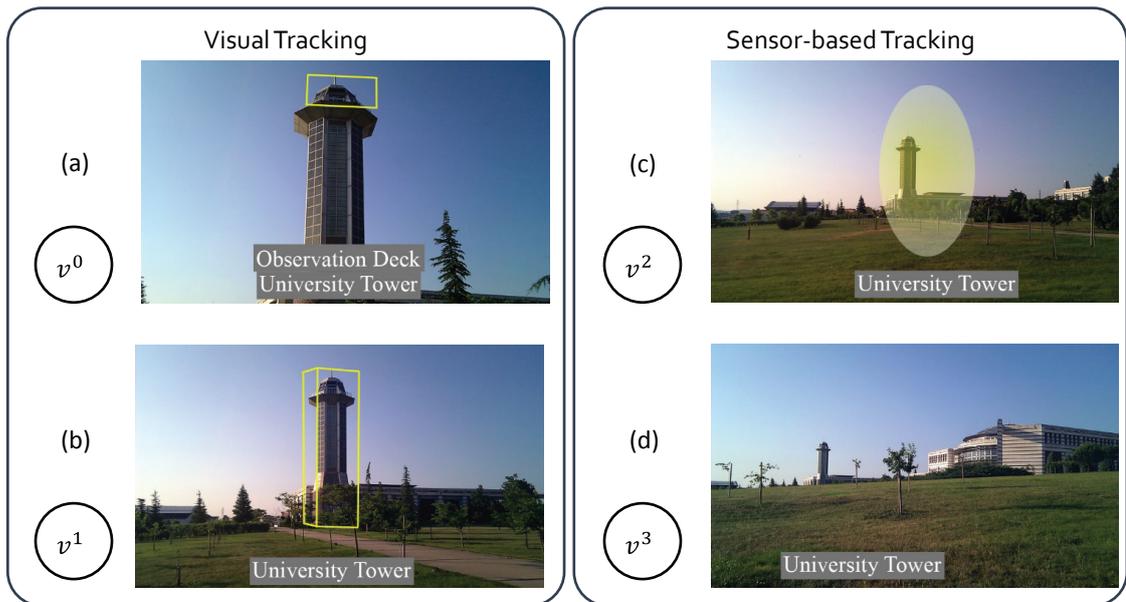


Figure 3.20: Four prototypical visualization modes are shown. Transition between states is performed via the transition function in Figure 3.18, based on a transition criterion.

3.6.2.1 Visualization Modes

Theoretically we can define any finite amount of visualization modes. However we must have a semantic relation between these modes. Figure 3.19 demonstrates this relation.

As we move from $v^{n \geq 1}$ towards v^0 , we expect to see an increase in detail, such as where v^n gives only a general direction of a building, as v^0 identifies a particular room inside a building. Conversely when moving from v^0 to v^n we observe a decrease in tracking precision. In order to identify a room inside a building we require very high accuracy, however to provide a general direction less precision is adequate.

To demonstrate our system, we defined four prototypical visualization modes. These modes are shown in Figure 3.20 and explained here:

v^0 : This mode is the most detailed visualization and requires the highest tracking precision. It can only be activated when vision based tracking is possible. As seen in Figure 3.20a, we can identify a particular room's orientation inside a building.

v^1 : As v^0 , this mode can only be activated when vision based tracking is possible. Figure 3.20b demonstrates an object visualized in this mode. A tight bounding box is drawn around the tracked object.

v^2 : Unlike previous modes, v^2 is used when sensor-based tracking is active. This mode

denotes the rough position of the tracked object on the screen. Unlike vision-based tracking we cannot find tight bounds for object's geometry due to noisy sensor data, therefore a colored gradient is rendered to visualize the objects' probabilistic location, as seen in Figure 3.20c.

v^3 : This mode is the least detailed visualization and requires the least tracking precision. Sensor-based tracking is also used for this visualization mode. In v^3 mode, we only provide the general direction of the object via a label, located at the bottom side of our user-interface, as in Figure 3.20d.

Along our transition criterion as distance analogy, v^0 is when the user is relatively close to the tracked object. This means the object covers a certain amount of screen space and we can easily identify a sub-segment of the whole object. As the user moves further away from the object, its projection on the screen becomes smaller, making it hard to differentiate sub-segments visually. So we switch to v^1 mode, where we only show a tight bounding box. As distance increased, the number of pixels covered by object's projection becomes so small that vision tracking may not be possible. When the distance is greater than a threshold value, v^2 mode is activated and sensor-based tracking takes over.

3.6.3 Tracking

Our visualization modes require objects to be tracked either by vision or sensor-based methods. Our transition model is compatible with most of the available tracking methods.

In our experiments we have utilized Vuforia SDK from Qualcomm as our vision-based tracker [88]. A previously taken photograph of the object is provided as an input to the registration algorithm. Then a feature set is generated from the input image and fed to the tracking algorithm. The objects are tracked in 3D.

For sensor based tracking we utilize GPS, accelerometer and magnetometer sensors. GPS provides location, magnetometer supplies heading data and accelerometer is used for pitch and roll compensation. We use a pre low pass filter for accelerometer data and a post filter for magnetometer data as described in [89].

| Transition Direction | <i>m</i> values | | |
|----------------------|-----------------|-----|-----|
| | 0 | 1 | 2 |
| b^m | 40 | 180 | 480 |
| f^m | 50 | 200 | 500 |

Table 3.4: b^m and f^m are the values for a transition model with four visualization modes. Threshold values are given in meters. The tracked object has 8m width and depth and 25m of height.

3.6.4 Discussion

We have implemented our proposed system on Android 2.2 platform [78]. Tracking algorithms combined, run roughly around 25-30 fps on ARM Cortex A7 processors commonly found in recent smartphones.

One challenge was choosing appropriate threshold values for our transition function. Chosen values based on experiments are shown in Table 3.4. If a priori knowledge exist about the target object, it may be possible to perform some analysis and statistically determine the optimal values. The priori knowledge may be the geometry of our target, analyzing it we could determine the best distance to switch from sensor to vision based tracking.

To that effect, transition criteria also hold great importance. We tried to leave it as generic as possible. Distance to the object, scene lighting conditions can be used as a criterion. It is also possible to record and compare time of image capture and AR viewing. In example a feature set prepared by examining an object’s photo which has been taken in daylight, would probably be a little of use when trying to register and track an object in night time. In this case it would make sense to switch to sensor based tracking which provides time-independent tracking.

In our transition function, we have used Schmitt triggers. It is also possible to use hysteresis curves as well for smoother transitions, as seen in Figure 3.21.

Our sensor based tracking method has 100ms delay, due to use of low pass filter. One way to improve this would be to implement a predictive filter such as a Kalman filter. By combining accelerometer and magnetometer input, it is possible to design an Unscented Kalman Filter with dynamic noise correction, as in [89].

In this study we proposed a transitional AR system, which can work with multiple

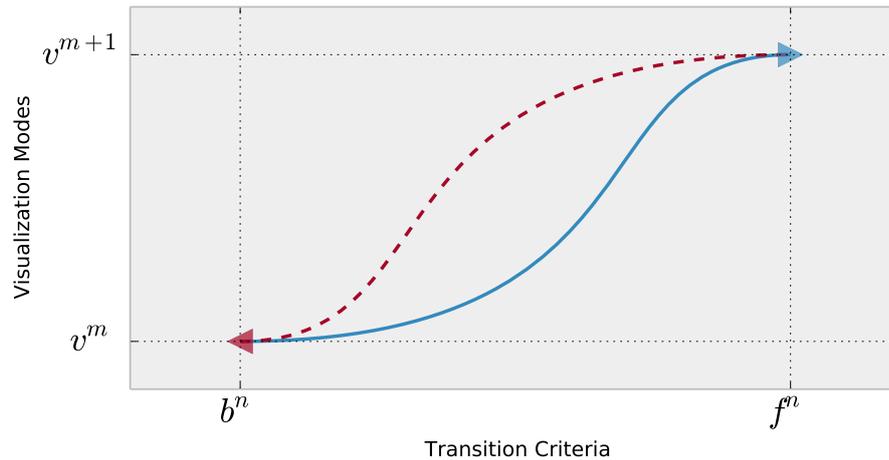


Figure 3.21: Transition function is illustrated as a hysteresis curve. f^n and b^n denote cutoff values for forward and backwards movement respectively; for switching between two consecutive visualization modes, namely v^m and v^{m+1} .

objects. The transition states are identified by their visualization styles. These states are also bound to different tracking algorithms, such as vision and sensor based. Transition function is acquired through examining several signals that we call transition criteria. We also modify this transition function via Schmitt triggers in order to create a non-intrusive user experience.

We propose to enable a hybrid tracking system with consistent transitions between visualization states. The system provides visualization based on viewing conditions, whenever possible the most precise and detailed visualization is preferred. However due to sensor noise or other external phenomena we provide fallback visualizations requiring less precision. One of our main goals is to create an object dependent fallback mechanism that can produce visualization at any condition, rather than a scene dependent mechanism.

Our system runs on current smartphones in real time and tries to produce the best available output of information for a seamless user experience.

An accompanying video demonstrating our work can be found at [90].

Chapter 4

Modeling and Annotations

Outdoor field work such as geographic or archaeological surveying require editing and processing of semantic information on spatial data. Currently these studies are performed manually using pen and printed maps or a laptop with a GPS receiver and digital two dimensional (2D) maps. In many cases such as rescue excavations for urban archeology or site surveys after a flooding; there is limited time before construction work starts and traditional techniques do not suffice.

In practice annotations are used to mark different layers and regions in civil engineering or stratigraphy studies. The processing time consists of manual work in the field and digitization of the annotations at the office afterwards. A laptop allows users to process digital data in the field but hinders walking around freely and requires constant switching between the laptop screen and the real world. This mental mapping process may lead to high error rates. In this workflow we demonstrate improving the limitations of this procedure using hand-held mobile computers.

Hand-held mobile computers already have started replacing notebooks and desktops for many computing tasks in the field. These devices have several shortcomings, such as: limited battery life, small display area and limited user interaction. Solutions which have been optimized for desktop environment need to be carefully re-designed and extended in line with the requirements of the mobile work environment. The main goal of this workflow is to let the professionals perform the annotation task in the field successfully using a mobile device in minimum time.

We propose a workflow featuring a simple modelling and annotation authoring process. There are two major issues need to be taken into consideration; i) how to create

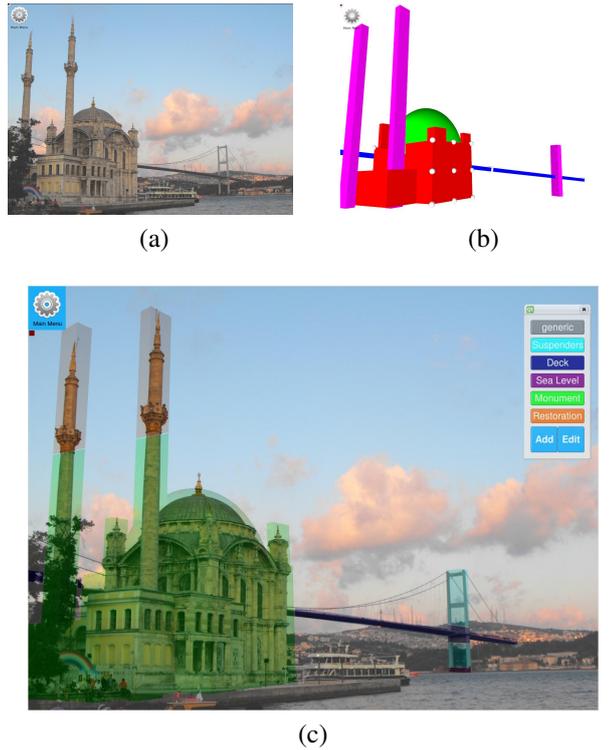


Figure 4.1: An urban scene is (a) photographed. Using these images, two objects are (b) modelled and (c) annotated using our workflow. Annotations are color coded; a legend is shown in the canvas for identification.

three dimensional (3D) annotations and ii) how to visualize these annotations in a mobile context. An annotation can be defined as adding extra virtual information over a real object [44]. We extend this definition and employ a variety of annotation types ranging from a single point to four dimensional (4D) annotations, an annotation of a volume over time. In the context of this workflow, the main goal of annotations is to identify the primary building blocks or layers of an object. Labels and text may not be enough for complete annotation authoring. Archaeologists and civil engineers are interested in layer based studies such as stratigraphy. In order to annotate a layer of a 3D object correctly, a 2D label is not the best choice. A layer represents a volume of the object, so we propose a volume based annotation authoring process.

To visualize annotations, we utilize user generated 3D models as an underlying structure. When dealing with static images, where the user only observes the scene from a single point, annotation authoring and management can be achieved using 2D constructs such as labels and floating text [91]. On the other hand, in mobile context the user can freely move in the scene, thus effectively changes the eye position. When the user moves, 2D constructs may start overlap or even become distorted and very hard to read. In or-

der to handle annotation visualization for mobile users, an underlying 3D structure is preferred, especially to handle occlusions by utilizing depth information [92]. Kopf et al. used high quality models and accompanying textures to visualize and annotate large scenes, such as Manhattan Island. Although the results are visually impressive, editing and processing of dense models on a mobile device may not be feasible. In addition, our use cases require fast-generated and simple models that allow editing for annotation authoring. To overcome these issues, we propose a simple and semi-automated image-based modelling process, where the user combines several building blocks in order to create a model. Annotations are then presented over these user generated models.

Our contribution includes i) a novel annotation technique based on 3D geometric regions and ii) a modelling workflow based on building blocks. To best of our knowledge, this is the first method that utilizes volumetric annotations. We also introduce a simple and intuitive interface for modelling and annotation editing processes. In addition, we conducted a user study and observed that the proposed solution is suitable for mobile field work.

4.1 Block Based Mobile Modeling

Our modelling process utilizes a “construction toy” analogy. The output of our modelling process is a combination of interlocked primitives. In order to create a complex model, user attaches 3D geometric primitives to each other, one at a time. These 3D geometric primitives are referred as “building blocks” in the rest of the text. For simplicity the variety of building blocks are kept at minimum, i.e. cube, column, dome and cone. However for each building block the user is able to define an independent transformations, translation, rotation and scaling. By utilizing these individual transformations it is possible to create many required primitives to model a building.

Semi-automated approaches have long been examined for image-based modelling processes. In many of these approaches, the user is asked to match exact features in several images [93]. More recently, VideoTrace [37], allowed users to define polygons on video frames and these polygons are auto transformed with respect to camera positioning. Our approach lies in between; rather than letting the user match exact positions in several images, we ask the user to adjust an initial building block, incrementally fixing the orientation over several images. Additional building blocks inherit the orientation of this reference block. The final orientation is saved in real world coordinate system. Using

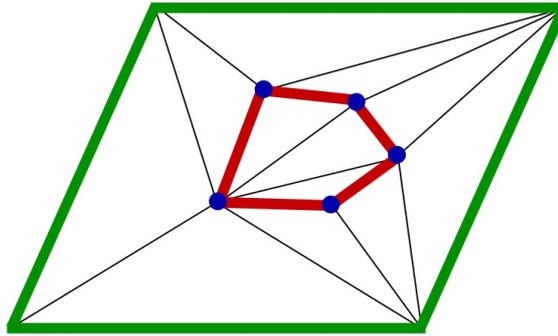


Figure 4.2: Example The green polygon is the initial polygon. The red polygon is defined by user clicks. Green and red edges are supplied to Delaunay triangulation. The output is the combination of green, red and black edges.

GPS and digital compass data associated with every reference image, we triangulate and find the estimated position for the real world objects.

The modelling process starts with inserting an initial building block to our scene. This block is translated, rotated and scaled by the user, to match a primitive of the object that is being modelled. Then user is able to drag and drop next desired building block to the scene. The new block is attached to the model when the user drops this building block on to any previous block. In this case the newly added block is automatically transformed and inserted to the scene hierarchy as a child of that previous block. The user may adjust the transformation via a simple graphical user interface (GUI). This process is repeated until the object is completely modelled.

The Building blocks can interlock each other at 26 different locations. These locations lay on the bounding box of each block. They consist of 8 corner points, 12 points in the middle of each corner pair and 6 face middles.

The interlocking process takes source and target blocks' scale and an interlocking vector as input. For example if the user inserts a new block to right side of a previous block, then the interlocking vector should have a positive value along the X axis, in particular this vector is $v(1,0,0)$.

To adjust the scale of new the block, the axes with 0 value is considered. Corresponding scale values on these axes are used to find the maximum ratio in between. The inverse of this ratio is used to scale down the new block. After auto-scaling, the new block is translated to the edge of the previous block to make the blocks look like interlocked at each other. A newly added block carries the rotation of its parent.

Table 4.1: Spatial components of an annotation are summarized.

To create holes and extrusions on the model, the user is provided with fine tuning tools such as slicing and extrusion. Slicing is achieved via adding user defined vertices on to a plane on the model. The newly added vertices along with the initial vertices of the plane, then fed to a constrained Delaunay Triangulation [94]. This process is demonstrated in Figure 4 2. The output of this triangulation is the same plane with more polygons in it. The user is able to delete any of these polygons to create holes, or extrude them to create additional extrusions.

During the modelling process photographs of the modelled object is shown as background images. The virtual camera is translated to relevant position for each corresponding image. By utilizing a pre-computed camera calibration, the model is ensured to superimpose the object correctly for each image. When modelling is completed it is possible to export geometric data into a Collada [95] supported format.

4.2 Temporal and Volumetric Annotations

Wither et al. explains that every annotation should have two parts; a spatially dependent component that links to the object and a spatially independent component that contains the information that is to be annotated over the object [44].

Our annotation system is based on the definition of Wither et.al. We extend this definition by adding specific items for spatial and semantic components. These components vary as detailed in Table 4 1 and Table 4 2. Spatial component of an annotation is defined as one of the following; vertex-based, planar or volumetric. Semantic Component can have all of the values described in Table 4 2. Using a combination of these components it is possible to create any annotation ranging from a label to a 4D annotation, an annotation of a volume over time.

In order to create an annotation the user first defines a spatial component and assigns a semantic component to it. The semantic component can be previously defined or can be created from scratch on-the-fly.

Table 4.2: Semantic Components of an annotation is visualized.

4.2.1 Spatial Component

There are three different geometric options for defining the spatial component using the previously generated model. In case of vertex-based spatial component user simply defines a point in the scene by clicking to the desired location. A label is created in this location, representing semantic component of this annotation.

For a planar spatial component, the user is able to select a face of any building block. It is also possible to adjust this selection by adding arbitrary points on the face to create a more detailed polygon on the model. This is achieved by inserting user defined vertices on the face and computing a constrained Delaunay triangulation.

In order to create a volumetric spatial component, the user needs to define a volume on the model. This process is simplified by utilizing clipping planes. The user creates desired number of clipping planes to create a sub-section of the 3D model. The volume which resides in between the clipping planes becomes the volumetric spatial component. Figure 4 3 through Figure 4 6 illustrate this process.

4.2.2 Semantic Component

A semantic component must have an ID and a color, other fields are optional. When assigned to a vertex-based spatial Component, an annotation is created as a label. The ID of the semantic component is displayed on this label with the appropriate color. When assigned to a planar or volumetric spatial component, the geometric region defined by the spatial component is colored accordingly to create an annotation as seen in Figure 4 6. A legend is also displayed to identify colored components on the canvas separately.

We utilize time as a variable to visualize annotations, in a chronologically ordered scene. As shown in Figure 4 1, many urban settings contain visible objects from different eras, the user is able to observe annotations of these objects in chronological order by moving a time slider. As time progresses, relevant annotations simply fade in to the scene to superimpose the real world images. The annotation is active and visible only for the interval defined in the associated semantic component. A descriptive text is shown when the user clicks a specific annotation.

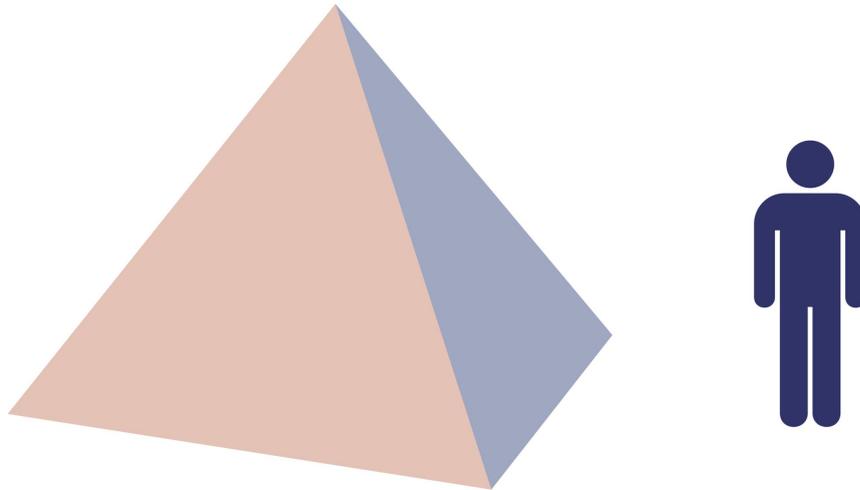


Figure 4.3: The user observes a 3D model ready to be annotated.

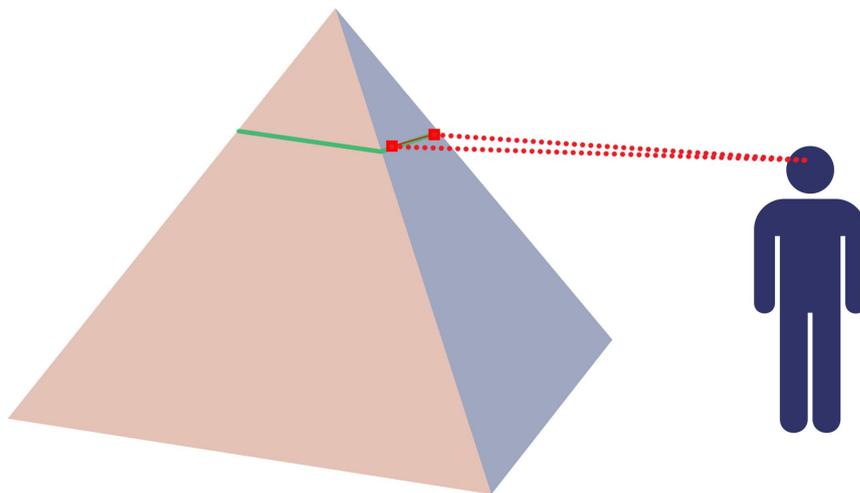


Figure 4.4: Red squares denote user clicked 3D positions. Using these two points and the position of the virtual camera, a clipping plane is calculated. With this clipping plane the 3D model is divided into two 3D volumetric regions. Green line is the contact region of these two regions.

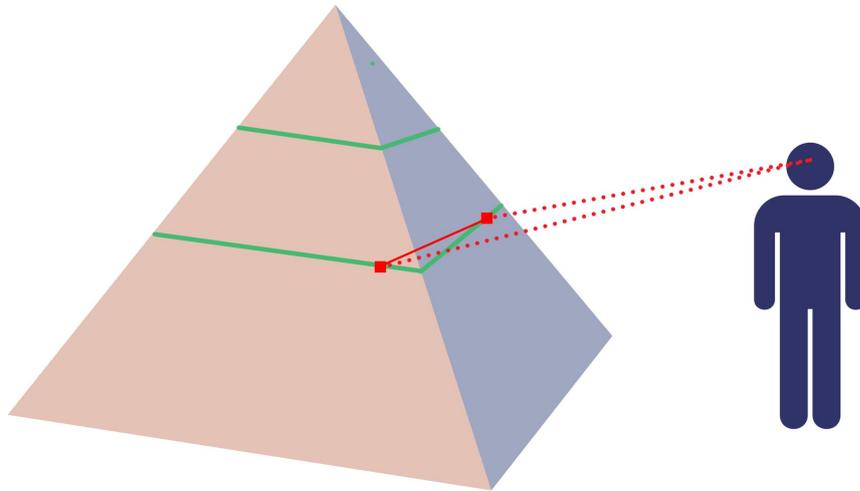


Figure 4.5: A new volumetric region is generated using the same approach in Figure 4.4. The user clicked points, do not have to be on the same face. As long as they are located on the model geometry, a new clipping plane is calculated.

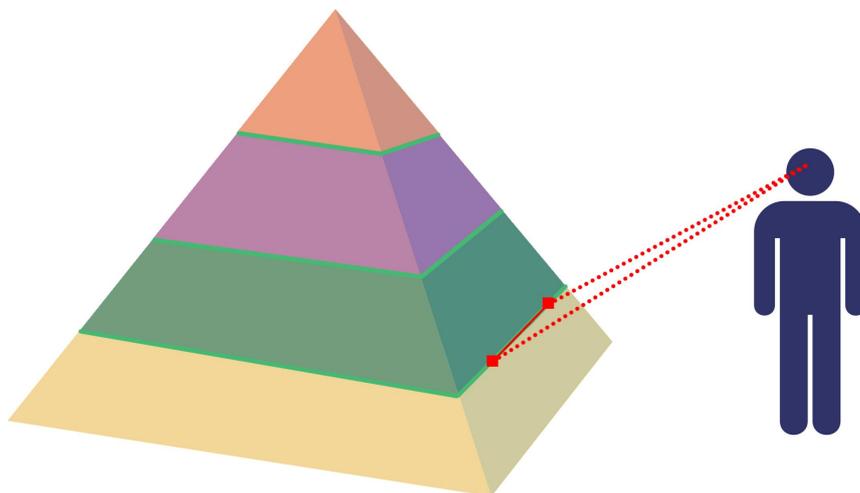


Figure 4.6: A final region is added. The created volumetric regions are associated with semantic components to create annotations. The annotations are presented in different colors and superimposed over the model.

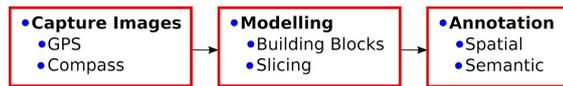


Figure 4.7: Our workflow is summarized in three steps. A modelled object can be annotated more than once.



Figure 4.8: A building is photographed from four different angles, two of these are shown here.

4.3 Case Study

This chapter elaborates on the design choices we have taken. A flow chart demonstrating our approach can be seen in Figure 4.7. Figure 4.8 through Figure 4.10 demonstrate the workflow with specific examples.

The first step of our process is capturing and placing images in our scene coordinate system. This requires GPS and heading data. The minimum required number of images is one, however capturing 2-5 images from different viewing angles produces better results. These images will be used as reference images in the application.

The next step is modelling. Reference images serve as background and virtual camera is translated in order to represent the position of the real camera. The very first building block for each new object establishes a mapping from scene coordinates to object coordinates.



Figure 4.9: Modelling process starts with creating and adjusting a reference block. This block has the same orientation with the building.

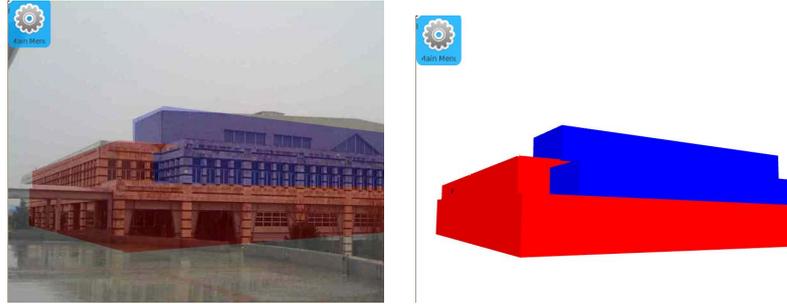


Figure 4.10: Completed model is shown; in this example 6 blocks are used to model the entire building.



Figure 4.11: After generating volumetric regions as spatial components, four different annotations are created. These are, from top to down; 2nd Floor, 1st Floor, Ground and Basement.

dinates. We call this the reference block of the object. This reference block is the root of the object building block hierarchy. We save the orientation information and use it to place our model in world coordinates by a simple triangulation process.

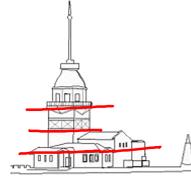
The user can also translate the virtual camera to a pre-defined position such as top view. This is similar to the approaches used by [96]. The user is now able to add new blocks to the scene using point and shoot analogy. The new block is attached to the user selected block along the interlocking vector. The interlocking vector is selected by clicking directly on the building block's related area. Alternatively a pre-defined vector can be selected from the GUI.

Further in the workflow, the user starts to annotate the model by first identifying the geometric region of the annotation. The spatial component can be a point in the scene, or a polygon on the model. Both of these components can be defined via clicking on the desired location.

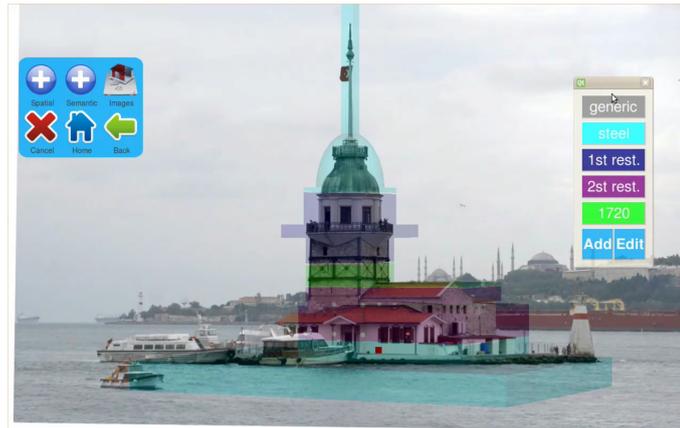
It is also possible to define a volume of the model as spatial component. In order to define a sub-section of the model as a volume, the user facilitates clipping planes. She clicks two different points on screen to form a clipping plane. This plane divides the model into two different volumetric regions. Any number of regions can be created by



(a) A photograph of a historical building



(b) Hand drawn lines over the sketch



(c) Annotated model

Figure 4.12: The real world image(a), is annotated using our workflow(c). The sketch(b) is provided to subjects as a guideline for annotation task of the user study. Subjects were expected to label four different layers, namely; *steel support*, *first restoration*, *second restoration* and *new base*.

defining additional clipping planes as shown through Figure 4 8 to Figure 4 11. The volumetric region that lays in between consecutive clipping planes becomes the spatial component. This process is especially useful for defining layers in stratigraphic studies.

After identifying a spatial component the user assigns a semantic component to complete the annotation process. It is possible to use a pre-defined semantic component or create a new one. An ID and a color is required for each semantic component. A dialog window is used for creating and editing semantic components. This dialog window contains a color picker in RGB color space and input fields for related text and sliders for time.

A modelled object can be annotated more than once; i.e. for several users or may be updated to reflect recent changes.

4.4 User Study and Discussion

We conducted a preliminary user study to test the ease of use our framework. The users were asked to model and annotate a historical building, as seen in Figure 4 12. In these tests, the average task completion time for modelling task is 647 seconds. In this period of time, users interacted with the touch screen approximately 320 times to model the building given previously captured images. All users produced usable models which can be correctly annotated, with 8.3 building blocks on the average.

The average task completion time for annotation is 168 seconds. Users interacted with the touch screen approximately 36 times to create and label four different volumetric annotations. Only one, out of eight users, failed to generate these layers correctly. Users found our framework generally to be user friendly, a score of 4.1 out of 5, 5 being very user friendly, is received from qualitative questions.

For an average user it takes about 15 minutes from scratch to model and accurately annotate a building with approximately 10 blocks. 65 percent of all the touch screen interaction is navigation through images and menus. Reducing this ratio would result in even faster task completion times. As a future work, we are looking into multi-modal interfaces; voice and sensor-based interaction for simplifying navigation will be investigated.

Each modelling tool has its strengths, some create highly accurate visuals [39] and some others emphasize on fast modelling [52, 97]. Modelling is essential in our workflow in order to visualize annotations in a meaningful way. Our modelling flow has a simple and intuitive interface for modelling in real-time and in the field.

Our volumetric annotation system is most applicable to layer based identification. This identification method is mainly used in stratigraphy and archaeological studies. It is possible to include different annotation schemes by simply registering extra clipping planes for regions.

Chapter 5

Exploration and Measurement

Outdoor Augmented Reality is a wide research area with a large set of application areas spanning from defense to entertainment. A key issue is visualization of occluded objects with highest possible accuracy and preservation of spatial relationships between visible and rendered objects. A large set of research activities are focused on displaying information hidden behind other surfaces such as walls, buildings, and mountains based on X-ray visualization techniques [63, 98, 68]. We are interested in evaluation of different visualization techniques providing information on what is hidden beneath other objects such as floors, streets and terrain.

There are several techniques on exploring existing urban infrastructure and archaeological artifacts such as ground-penetrating radar, radio frequency or electrical-resistance-tomography and robotic sensors [99, 100, 101, 102]. New and existing pipe networks and other geo-referenced subterranean data are documented using geographical information systems. Hence there is a need for in-situ visualization of what is documented on a mobile device such as a smartphone or tablet in AR fashion.

The visualization of the occluded objects should be done in a comprehensive manner. A straight forward approach is to utilize careless overlay of virtual objects on top of real world imagery as in Figure 5.1a. inadvertently this results in a floating effect [67]. On a similar approach, uniform transparency modulation will result in misleading perception of the depth order [55]. The viewer needs depth and shape cues as well as clearly visualized spatial relationships between occluded and the occluding objects to have spatial awareness.

A recent work on subterranean visualization is the Smart Vidente project [103], which

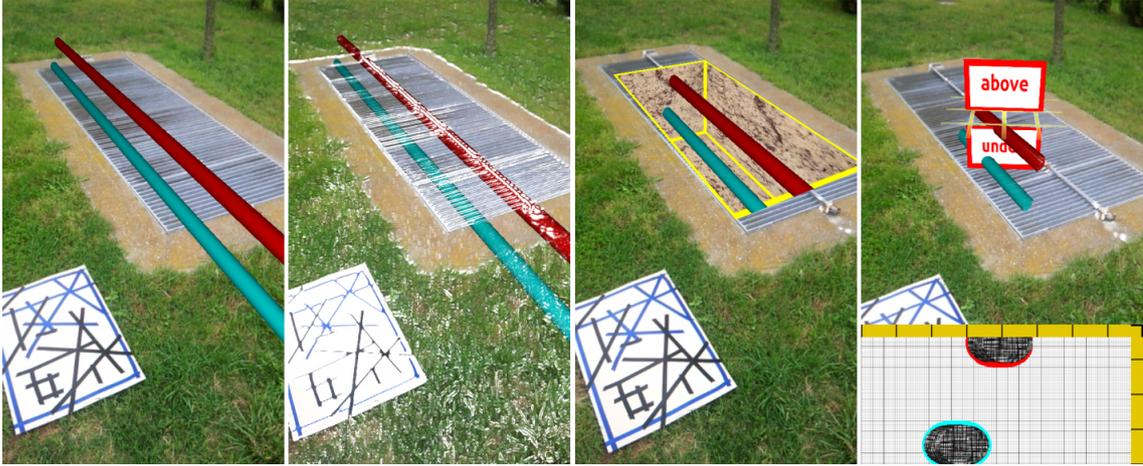


Figure 5.1: Visualization of underground pipe networks using different techniques. a) Careless overlay b) Edge overlay c) Excavation box d) Our proposed multi view technique.

employs a single view perspective display technique on a handheld device. However Roberts et al. reports that utilizing 3D single view techniques are not sufficient for visualizing complex data [104]. Single view based techniques are weaker for conveying relative orientations of different layers, when compared to correlating multi views.

In order to create a comprehensible scene, we propose a multi view approach. The first view is a calibrated 3D virtual scene overlaid on top of camera imagery. This view can be seen on the upper side of Figure 5.1d, and is referred as the perspective view in our context. Similar visualizations can be found in several studies in AR domain [105, 106].

Secondly, we generate a cutaway view which visualizes the subterranean layers as seen in the lower part of Figure 5.1d. In technical illustrations 2D visualization techniques are favored [71, 107, 108]. Cutaway visualization of complex 3D objects are also investigated in AR context [109]. We perform an orthogonal projection to render a thin slice of the subterranean data. There are several use cases that form severely different categories of requirements for X-ray visualization. Some of these categories require highly precise measurements such as medical imaging or sensitive archaeological excavations; where others are relatively more relaxed on precision constraints such as maintenance of urban infrastructure or geocaching.

We have designed a user study to examine and analyze state of the art X-ray visualization techniques' performance for identifying and comparing vertical depth distances at medium range. The evaluation of three existing and a proposed X-ray visualization technique by a comprehensive user study on perceived vertical depths, is our main contribution

for scene exploration. A further contribution lies in a novel multi view visualization technique, which allows effortless interaction with subterranean data and tries to maximize spatial perception while minimally cluttering the view.

5.1 X-Ray Visualization

X-ray visualization techniques are used for viewing occluded objects while preserving important features in an AR scene. Exploding diagrams, ghosting and cutaways are examples of such techniques [110, 111]. Bane et al. propose several tools of X-ray vision to be used in AR context [9]. Avery et al. discusses how overlaying edge features of the occluding structure would give better depth cues to the viewer and describes three tools for further improving spatial perception [63]. In our proposed multi-view technique, we used a similar approach for promoting sense of occlusion for subterranean structures.

There are a number of X-ray visualization techniques addressing subsurface occlusion problems. Shall et al. introduces an excavation tool inspired from magic lens techniques [64] that virtually digs the ground letting viewer to see underground pipes [106]. This technique requires viewer to be close to the location to effectively perceive the hidden structure (see Figure 5.1). In other words it suffers from the long-flat view problem described in [66]. Zollman et al. employs ghosting techniques for solving single layer occlusion problems between surface and the infrastructure system [67]. Panoramic images are used from the viewed site for calculating a ghostmap, then use features on this map to preserve the above ground context. Although they demonstrate occlusion clearly for a single layer of subsurface system, in the real world subsurface systems may consist of multiple layers that are occluding each other.

Occluded geometry visualization is studied extensively in AR domain. A careless overlay of occluded geometry as seen in Figure 5.1a, is not sufficient for visualizing these objects [67]. Previous studies enhance the scene by employing X-ray visualization via using Ghostmaps or simple edge-overlay techniques to give the sense of occlusion while visualizing hidden geometry [63, 55]. Figure 5.1b presents an edge overlay technique. Detected edges of the background image are overlaid on top of infrastructure pipes. This technique provides user a depth cue for occluded objects. However the user is not supplied with any focus cues. In Smart Vidente project, researchers utilized a dig box, to present focused visualization for an excavation site. We implemented a similar visualization that can be seen in Figure 5.1c. The rendering of underground objects is restricted to the

volume covered by this rectangular excavation box. In this work flow, the excavation box is created and fixed to user defined geo-location. This technique is tailored for examining a specific excavation location [106].

5.2 Absolute Vertical Depth Judgments

Distance judgments have been investigated heavily both in AR and VR domain [62, 56]. There are also studies that report X-Ray visualization's effect on depth perception [63]. One of the common focus of these studies is the experiments were done to investigate horizontal distance judgments; "how far away the objects from the observer?" As an object moves along the Z axis, two main difficulties may arise, namely perspective distortion and long flat view problems. The perspective distortion problem can be defined as the optical illusion of projecting a distant large object and nearby smaller object to similar screen areas. Long flat view problem is caused when viewing flat virtual objects at a distance [66, 56].

On the other hand, vertical depth distance is a fairly under-investigated topic, especially for fully occluded geometries. We believe, the main issue worthy of further investigation in underground X-ray visualization domain is vertical depth judgments:

- Are state of the art methods, perceptually and numerically accurate for estimating vertical depths? Moreover, do they have estimation tendencies, such as over or under estimation? Furthermore, are these tendencies in line with horizontal distance estimations reported in recent studies?

Another key issue is to examine observers' ability to compare multiple vertical distances. We believe depth cues provided by current methods are not designed for this purpose. And there is a need for careful design and implementation process focused on this topic. Hence, we propose a multi-view technique with depth cues that allows the observer to investigate both absolute and relative vertical depths by focusing on the following questions:

- Are state of the art methods, perceptually and numerically accurate for comparing two vertical distances? Can a multi-view approach improve comparative vertical depth judgments? Is a multi-view method more suitable for applications requiring exact judgments?

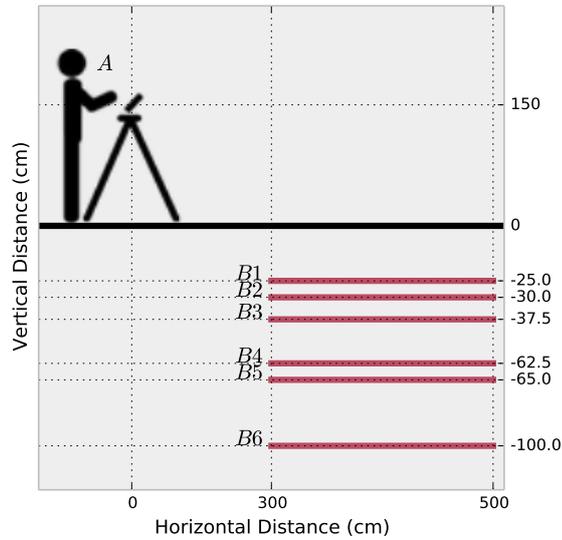


Figure 5.2: Absolute Vertical Depth Judgment experiment: Each participant performs six estimations for vertical position of the pipe B_i , where $i \in 1 \dots 6$ (25 to 100cm).

In order to answer these questions, we have conducted two experiments: An absolute (non-comparative distance judgments) and a comparative judgment experiments.

5.2.1 Absolute Vertical Depth Judgment Experiment Setup and Task

The experiment took place outdoors where users were shown underground pipes via a mobile phone mounted on a tripod. Underground pipes were placed about horizontally 3 meters away from the participant. The pipes had varying vertical depth distances to the ground between 25cm and 100cm. The pipes with 20cm diameters were shaded with red color. The localization was done via marker that was placed on the ground. Independent values for this experiment are as the following:

- Observers: We have experimented with 60 undergrad or graduate university students of which 28 were female and 32 were male. Height difference between observers was an issue of concern when they were asked to judge vertical distances. To overcome height variance and disparities in viewing angles between observers, we utilized a tripod in the experiment setup; fixing the viewing position and angle.
- Viewing conditions: Observers were presented with one of the three visualization techniques; careless overlay, edge overlay, dig-box. Each visualization technique had been observed by 20 participants. For each of the techniques, we supplied some verbal clues: For careless overlay, we have acknowledged that the pipes are

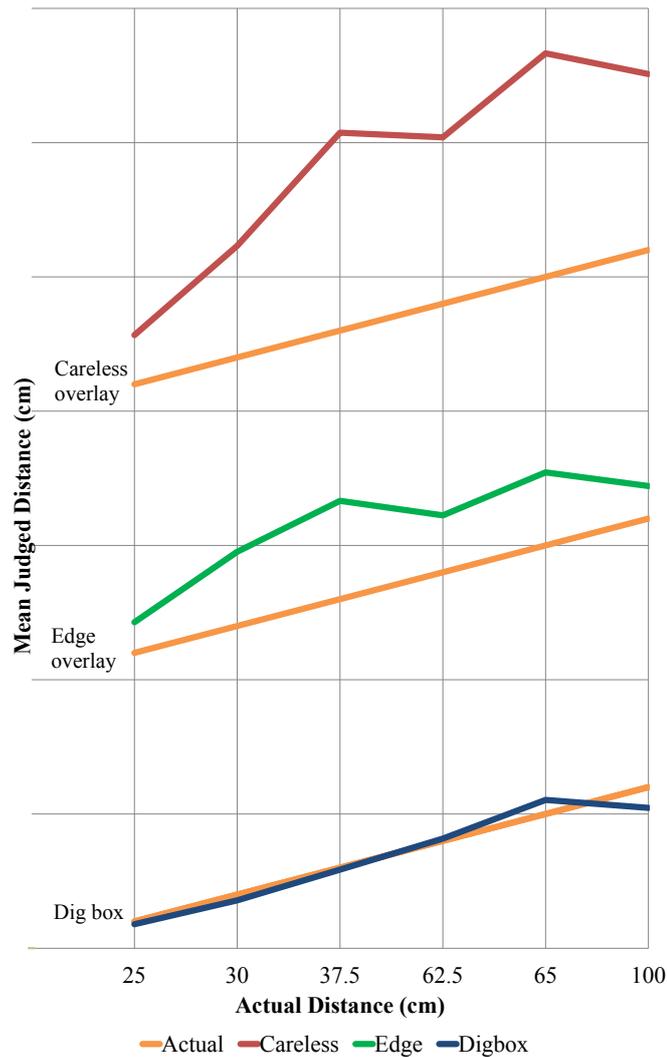


Figure 5.3: Absolute vertical depth judgments were plotted against actual distances. Each technique has a vertical offset in the plot for clarification.

underground objects. For dig box technique, participants were supplied with the real world dimensions of the virtual dig box (2 by 5 meters). No additional training was given other than these clues.

- Absolute Vertical Depth: For each observation, participants were shown a single underground pipe and were asked to evaluate the pipe’s vertical distance to the ground as in “how deep is this pipe buried?” We define this distance as the egocentric vertical distance. 6 vertical depth values; 25, 30, 37.5, 62.5, 65 and 100 cm were shown to the participant, one at a time.

An illustration of this experiment’s setup can be seen in Figure 5.2. The vertical depths of the pipes were selected manually, to span medium range values. To prevent pre-



Figure 5.4: From left to right, sample scenes are visualized via careless overlay, edge overlay and excavation box techniques for absolute vertical depth judgment experiment.

sentation order effects, we have designed a between subject experiment [61]. Observers reported judged distances verbally. Each observation was normalized as judged distance / realdistance . We have also calculated absolute errors as $\| \text{judged distance} - \text{realdistance} \|$ to be used in ANOVA.

5.2.2 Absolute Vertical Depth Judgment Experiment Results

We have run two-way ANOVA allowing us to analyze effects of each technique over different settings and questions, using absolute error values. The visualization techniques have significant effect over judged distance and each technique is statistically different than others ANOVA produced $p\text{-value} = 1 \cdot 10^{-6}$, $F = 20.44$.

In Figure 5.3 each technique's performance is visualized and compared against actual values. We have normalized the answers as judged distance / real distance then removed outliers via a z-test. As expected, the careless overlay was the least performer with a mean value of 2.77. Next, edge overlay group ended up with a 1.65 mean. Dig box participants performed significantly better than two previous techniques with a mean of 0.98. Variance within groups also reveals significant data, careless overlay and edge overlay groups produced 2.52 and 1.12 standard deviation, hinting in high variance. Dig box group showed the least value for standard deviation as 0.26. It is worth mentioning that careless and edge overlay observations were extremely overestimated by almost a factor of 2. The only perceptually sound technique from the three was dig-box. This was expected as the other two methods did not provide any depth cues. Although, overlaid edges made the

pipes look underground, they did not hint in how deep the objects were. Participants in dig-box group, utilized the back side of the box, where the pipes cut through and used it to estimate better results. In light of these findings, it is possible to report dig-box method is suitable for absolute depth judgments. However the question of comparative depth judgments still remains open. In order to test our claim that a multi-view technique would be suitable for comparative analysis, we designed the following method.

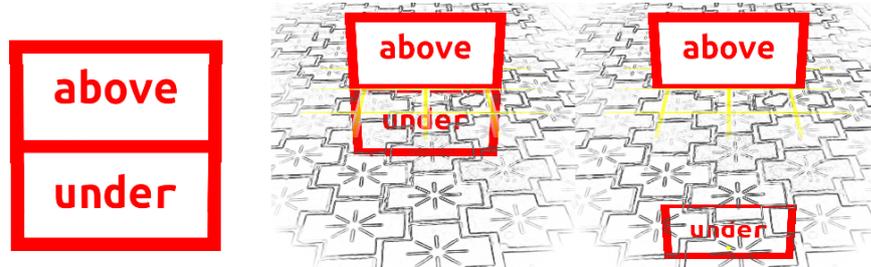


Figure 5.5: “above” and “under” planes are placed in an empty scene. An edge overlay is drawn to denote ground plane. a) front view, b) “above” and “under” touching, c) “under” plane at a depth of four meters directly beneath the “above” plane.

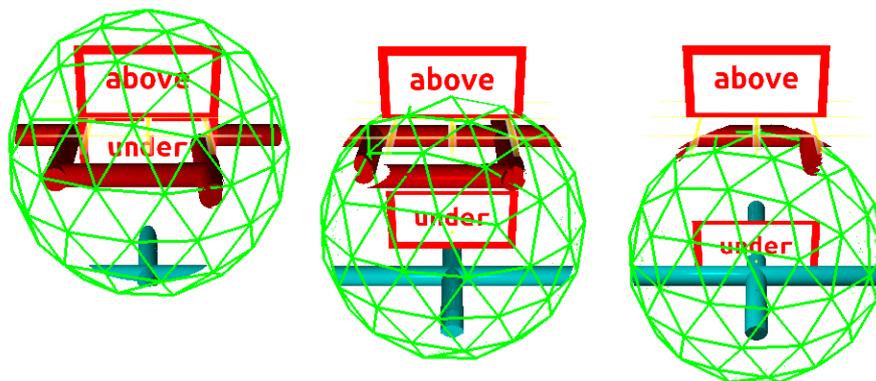


Figure 5.6: Clipping sphere used for focus preservation through information filtering. a) Focused on red pipe layer. b) Focused on the area between the layers. c) Focused on blue pipe layer.

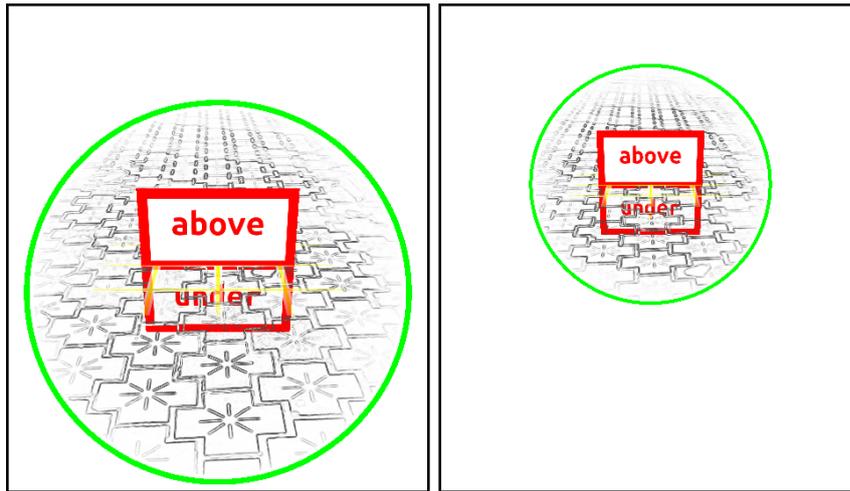


Figure 5.7: 2D clipping for edge overlay. Clipping circle's size is determined via the anchor object's position in screen space. Focus region is larger when the anchor is closer to the user.

5.3 Multi-view visualization

We utilize a two view approach; a perspective overlay with focused context preservation and an orthographic slice view. These views consistently visualize the same data with shape cues correlating with each other. During the exploration process, single view approach often leads to misinterpretation of the underlying data [104]. By displaying the same data using different techniques and from different angles; viewers are encouraged to match correlated elements. This approach may be used to overcome the depth perception shortcomings of three point perspective visualizations.

A cutaway slice of the infrastructure that lay immediately below this anchor is displayed on the orthographical view. When anchor position changes, both views change in a consistent manner, in order to keep user engaged to the system. Orthographic visualization is a supplementary view to perspective-view for providing user with additional depth and shape cues which are consistent to the perspective view.

5.3.1 Spherical and Screen Space Clipping

The size of the clipping sphere is significant in preserving focus-context relation. A bigger sphere enables user to view more information but lose focus, vice-versa, using a smaller

sphere results in increased focus but loses context. The size of the clipping sphere should be proportional to the size of the anchor. In our tests we have found double the width of “under” plane (4 meters) to be a reasonable value.

We perform spherical clipping using a fragment shader. This shader takes the center of the anchor’s world coordinates as an input. The distance between this position and the fragment’s computed coordinates is calculated. Then a test compares the distance to a pre-defined clipping radius. If the distance is smaller the lighting for this fragment is computed via Phong shading model, otherwise the fragment is discarded as seen in Figure 5.6 .

Similarly, for the edge overlay visualization another fragment shader is utilized. Without clipping, this shader is a full screen edge detector using Sobel convolution [112]. To enable screen space edge overlay clipping, this shader requires 2D screen coordinates of the anchor as input parameters. If the fragment falls in to the circular area, the edge detection is computed; otherwise the fragment is again discarded.

Utilizing the focused edge overlay technique; edge detection is only performed on a specific region as seen in Figure 5.7a and Figure 5.7b. Another upside of this approach is avoidance of unnecessary computation when compared to full screen edge detection.

5.3.2 Orthographic-View

The orthographic view is generated with an orthographic camera that is positioned facing the “under” plane. This projection has a very narrow near to far-field range. We employ this narrow range in order to mimic volumetric cutaway visualization.

Preserving spatial and F+C relationships are crucial for improving user’s perception of the scene. Users need more clues to understand this pairing. Firstly the existing color schemes are used to color wire frame representations of the subsurface objects. This technique helps users understand that the object being cut by the “under” plane is being visualized in orthographic-view via color matching. Secondly the pulse animation separates the focused object from rest of the scene similar to [113].

The geometry on this view is rendered in three passes. In the first pass only a wire frame representation of the objects is rendered. For a pulsing animation effect that emphasizes each object’s current slice, the line width is oscillated from one to twenty pixels via a sinus curve.

Second and third passes renders the object with a hatching shader [114]. Second pass draws back faces with flipped normals. Finally third pass renders the front faces. A similar multi-pass non photo realistic (NPR) technique is discussed in [115].

This NPR technique is aimed for the use of professional field workers and civil engineers that are familiar with technical illustrations. Considering this fact, we decided to use a hatching shader [114] for acquiring charcoal sketch like cutaway illustrations. Another advantage of using a hatching based shader, by drawing hatched lines along the normals of the object we preserve the sense of geometry on the final image.

5.4 Case Study

5.4.1 AR Application

We employ a mobile AR setup where users explore the underground objects through a smartphone. Perspective view in our technique is similar to classical mobile AR; video images from a calibrated camera are used to register and track predefined markers. A custom marker placed on the ground level is utilized in our experiments and demonstrations. The virtual objects are then rendered relative to this marker's position and orientation. For registration and tracking purposes we have used the Vuforia SDK [88].

We implemented two existing techniques in addition to our proposed one into our test environment to conduct a comparative analysis. One of them is a commonly used X-Ray vision technique; edge overlay and the other one is a directly related subterranean hand held mobile visualization technique similar to excavation box method used in Smart Vidente project [103, 65]. From Figure 5.8 through Figure 5.10 our sample scenes are depicted.

In Figure 5.8, a scene with a perpendicular set of pipes, colored red and blue are placed approximately three meters away from the user. The user views the scene by an angle of approximately 50 degrees; between the viewing direction and the ground grid. The red and blue pipes are located 1 and 1.5 meters below the ground respectively. Pipe networks can be clearly identified with each visualization method; edge overlay (Figure 5.8a), excavation box (Figure 5.8b) and our method (Figure 5.8c). Since edge overlay method does not have an anchor in the scene, user focus may shift to irrelevant regions. On the other hand, excavation box allows the user to focus on a specific location. However

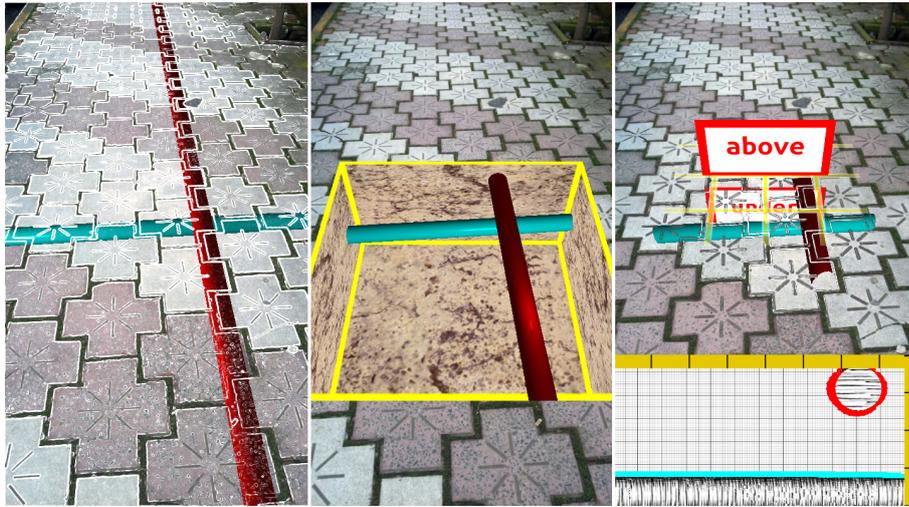


Figure 5.8: Perpendicular blue and red pipes are viewed at a close range. Using a-b) the spatial relations are ambiguous; c) our method clearly identifies relative positioning via orthographic view.

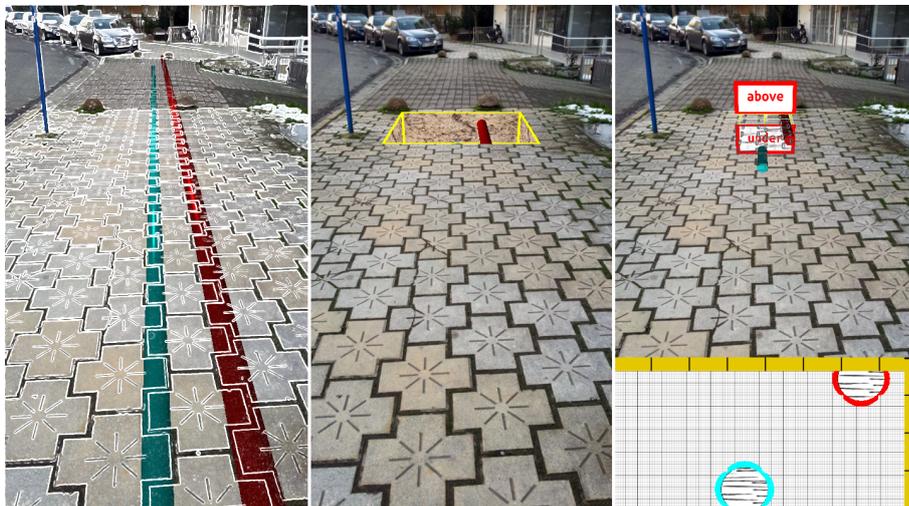


Figure 5.9: Two parallel pipes are visualized at a distance. Using a) edge overlay method, both pipes seem underground, however relational positioning information is lost. b) Blue pipe is fully occluded by the excavation box. c) Using our method both pipes with their spatial relation are visualized via multi views.

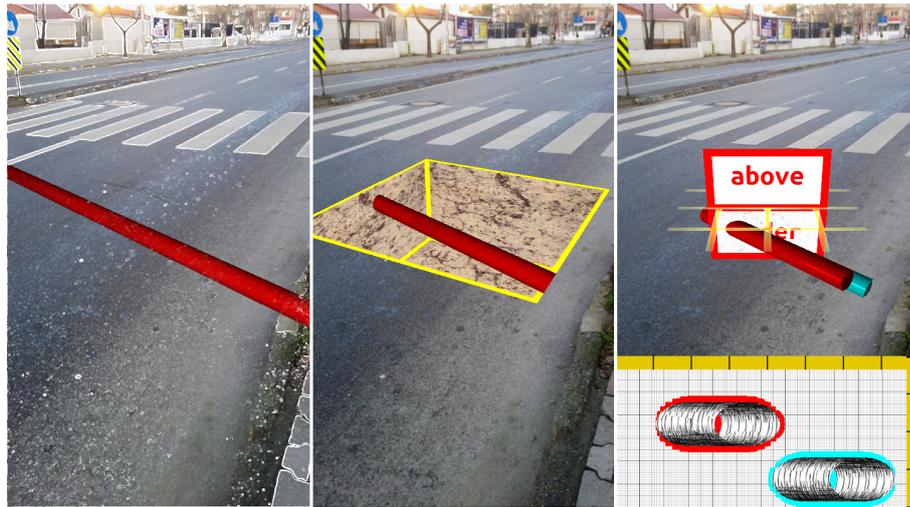


Figure 5.10: Parallel red and blue pipes are visualized using three methods. Due to perspective projection, the red pipe occludes the blue pipe that is behind. In a-b) blue pipe cannot be seen, c) our method is able to visualize both pipes in each view.

the box occupies a large portion of the screen and shrouds the camera feed. Our method preserves focus through selective geometry clipping and selective surface highlighting in perspective view while minimally cluttering the camera feed.

By examining any one of the images in Figure 5.8, it is possible to deduce that red pipe is above the blue one. However, the perceived distance between them may be ambiguous. As seen in Figure 5.8, the orthographic view is used to convey spatial relation information using our method.

Figure 5.9 demonstrates another sample scene where the focused area is farther away from the user. In this scenario the explored infrastructure is located roughly 12 meters away and the viewing angle is 30 degrees with respect to the ground grid. In Figure 5.9a, our interested region covers a small portion of the screen. Similarly, in Figure 5.9b, excavation box technique allows only the red pipe to be seen. We can also see the long flat view effect on this image as the box is located at a distance. This issue makes the exploration process harder to perform effectively. In our method, orthographic view's fixed size ensures that the pipes have a constant size regardless of their distance. User can perceive the positional relations of the pipes by examining the orthographic view (Figure 5.9c).

Lastly, two sets of pipes are displayed in Figure 5.10. In Figure 5.10a and Figure 5.10b the blue pipe is hidden behind the red one. The pipes are located 8 meters away from the user and the viewing angle is around 30 degrees. Our method provides the user

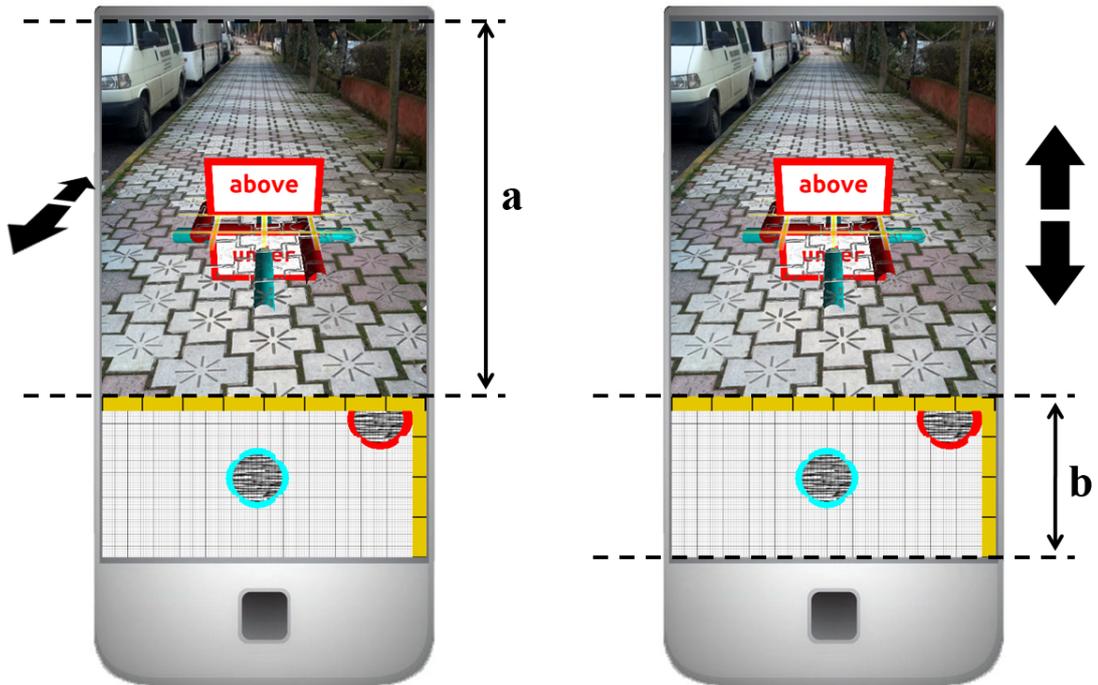


Figure 5.11: Touch based interactions translates the anchor a) along the viewing direction, b) through the ground.

the ability to explore both sets of pipes in perspective as well as the orthographic view. In perspective view spherical clipping removes part of the red pipe geometry on the left side of the image revealing underlying blue pipe. Orthographic view is also able to visualize the two distinct pipe sets (Figure 5.10c).

The effectiveness of edge overlay visualizations depends on input camera image features. Almost no edges are present at the lower part of Figure 5.10a. Thus, sense of occlusion is lost. In perspective view of our method the surface occlusion sense is preserved through overlaying a ground grid as a fall back mechanism.

Through these sample scenes, we have investigated the advantages of our multi view technique over previous methods where they may fail to visualize spatial relations of underground infrastructure. Moreover, our technique is not limited to underground infrastructure visualization and applicable to many other layered structure visualization. For instance, electricity cables, water pipes or foundation inside a building can be explored. Providing a slice view allows the viewer inspect the inner geometry that otherwise would be hidden.

5.4.2 Interaction

One to one motion mapping from user actions to objects in 3D space can be achieved using 3D input devices such as VR gloves. For the desktop environment, these interactions are mimicked using keyboard and mouse paradigm. However, on modern mobile devices the user interfaces are mostly limited to touch based interaction. Thus, deriving 3D translations from 2D touch interactions may be problematic.

While user's position and orientation are being tracked via a marker, she can also trigger two separate touch based interactions. In order to create an egocentric visualization, the anchor's position is fixed on the X axis in camera space. This assures the anchor will always be in the center of the screen horizontally.

When the viewer rotates the device around the Y axis, the anchor will still be in front of the user on the perspective-view. However since the underlying scene is also rotated accordingly, the user will be exploring the desired location.

For second and third interactions the screen space is divided into two. If the user touches and drags her finger on the perspective view, the anchor pair is moved closer or further away along the viewing direction (see Figure 5.11).

Similarly user interaction on top of the orthographic view, translates bottom part of the anchor, along the Y axis through the ground. This interaction causes the "under" plane to shift into the ground. The "above" plane is left on the surface to reflect "under" plane's relative position on the ground (see Figure 5.5).

5.5 User Study

In order to test our multi-view technique's perceptual validity, we included it in our absolute vertical depth experiment with an additional 20 subjects. Before the experiment, we informed the participants of "above" and "under" planes' functionality, as well as their corresponding real world sizes (1 by 2 meters each). The results can be seen in Figure 5.12 where we compare multi-view and dig box techniques. Multi view group participants had a mean of 1.03 and a standard deviation of 0.18. Participants' mean was similar to dig box technique; however the variance between users were smaller. Similar to dig box, multi-view technique is suitable for non-comparative vertical depth analysis since it provides a depth cue for the occluded geometries in the scene.

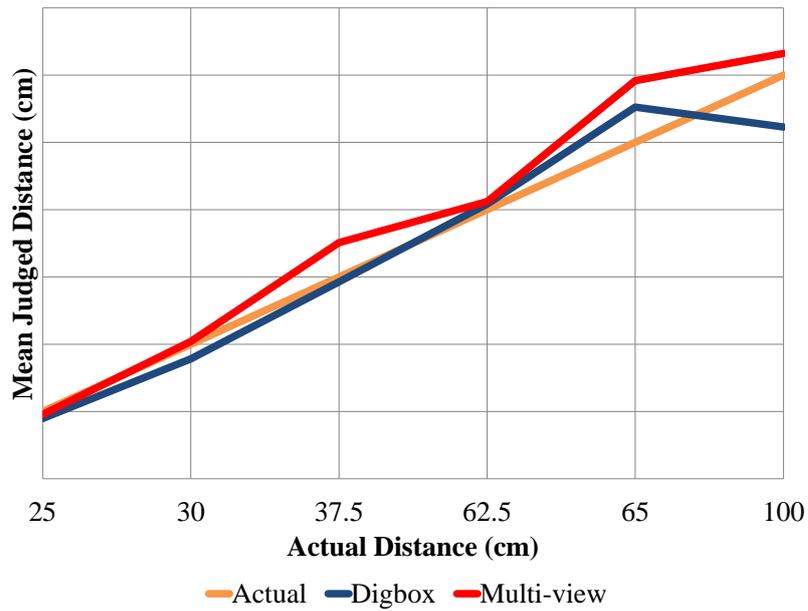


Figure 5.12: Absolute vertical depth judgments were plotted against actual distances for digbox and multi-view techniques.

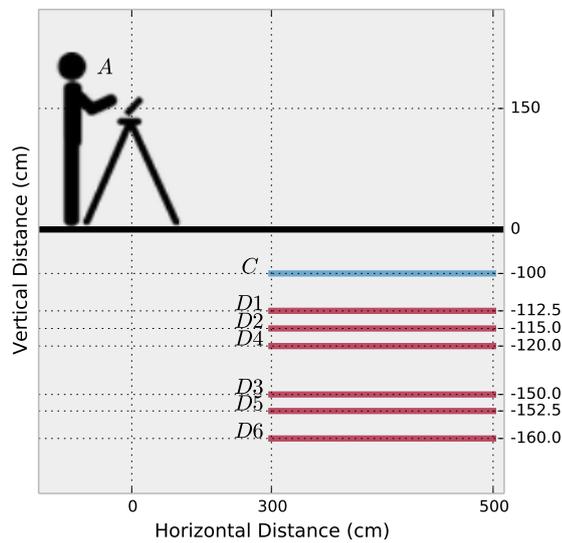


Figure 5.13: Comparative Vertical Depth Judgment experiment: Each participant is asked to identify the relative vertical distance between the red pipe b and the blue pipe D_i , where $i \in 1 \dots 6$.



Figure 5.14: From left to right, sample scenes are visualized via careless overlay, edge overlay, excavation box and multi-view techniques for comparative vertical depth judgment experiment.

5.5.1 Comparative Vertical Depth Judgments

5.5.1.1 Experiment Setup and Task

The exocentric experiment was designed to evaluate the relative spatial perceptiveness of each given technique. Underground pipes were placed about horizontally 3 meters away from the participant. The pipes had varying vertical depth differences between 12.5cm and 50cm. The pipes with 20cm diameter were shaded with red and blue colors. This experiment was done right after the absolute vertical depth one using the same physical setup, however this time the observers were asked to provide verbal judgments on exocentric vertical depths.

Between subjects design was also used for this experiment. Experiment's setup and sample images can be seen in Figures 5.13 and 5.14, respectively.

- Observers: All the observers from the absolute vertical depth experiment (60) plus the multi view technique's observers (20) were used in this experiment.
- Viewing conditions: Observers were presented with one of the four visualization techniques; careless overlay, edge overlay, dig-box and multi view. Each visualization technique had been observed by 20 participants.

The experiment task can be described in two steps:

Correct % Careless overlay 85.96 Edge overlay 78.94 Dig box 96.29 Multi view 100

Table 5.1: Participants are asked to order two non-intersecting underground pipes. Values represent the percentage of correct ordering for each technique.

1. Vertical Ordering: For each observation, participants were shown a pair of underground pipes. For 3 different pairs, shown one pair at a time, participants were asked to identify the shallower pipe.
2. Exocentric Vertical Depth: In the second step, participants were asked to judge exocentric vertical depths. Exocentric vertical depth is defined as the relative vertical distance between the two pipes. The first pipe's depth was always kept at 100 cm. The second pipe was placed at varying depths of 112.5, 115 and 150 cm. The depth values were chosen manually to span medium range values.

5.5.1.2 Experiment Results

Similar to absolute vertical depth experiment, two-way ANOVA over absolute error values for exocentric experiment shows visualization techniques have significant effect over judged distance and each technique is statistically different than others. ANOVA produced $p\text{-value} = 1 \cdot 10^{-7}$, $F = 13.21$.

Results for the first step of our exocentric experiment are summarized in Table 5.1. Edge overlay was the worst performer with 78.94% correct answers, followed by careless overlay with 85.96%. Dig box and multi view techniques produced highest percentages for correct answers, where dig box technique's percentage was 96.29; none of the participants in multi view technique group gave a wrong ordering for the pipes, resulting a 100% of correct answers.

For the second step of the exocentric experiment, results are presented in Figure 5.15. Careless overlay is again the least performer with a mean of 2.2 and standard deviation of 1.51. Participants in edge overlay group produced 1.35 and 0.72 as mean and standard deviation respectively. Similar to previous experiments dig box and multi view technique turned out to be the best results. Dig box participants averaged 1.07 where multi view technique participants produced 0.99 as mean value. Their standard deviations are 0.34 and 0.19 respectively.

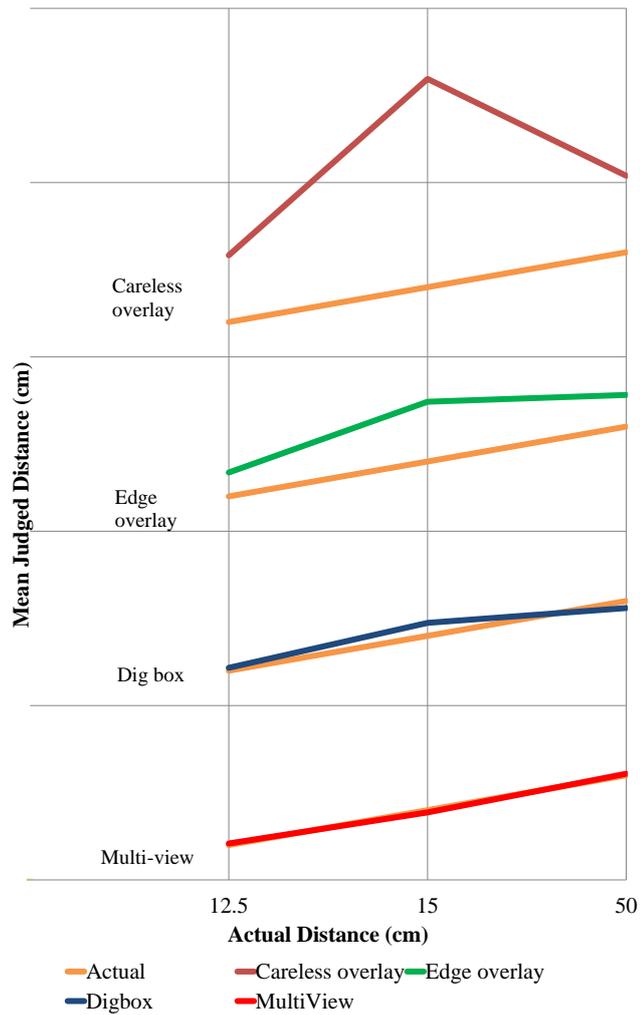


Figure 5.15: Plots for each technique’s average results are shown for exocentric experiment.

Under % Exact % Over % Careless overlay 31 3 65 Edge overlay 28 7 63 Dig box 40 24 35 Multi view 25 25 49

Table 5.2: Percentages of participants’ distance estimations over techniques are given. Careless and edge overlay techniques are dominated with overestimated answers where dig box answers show underestimation. Majority of participants in multi view technique gave

Under % Exact % Over % Careless overlay 10 23 66 Edge overlay 27 30 42 Dig box 37 29 33 Multi view 16 54 29 Table 5 3:

Table 5.3: Percentages of participants' distance estimations over techniques are given. Careless overlay, edge overlay and multi view techniques tend to have overestimated results in general. On the other hand, dig box answers are more likely to be underestimated.

Binary Questions Occluded Perception Absolute Vertical Comparative Vertical Careless overlay Edge overlay Dig box Multi view

Table 5.4: Sample use cases are shown for X-ray visualization techniques.

5.6 Discussion

5.6.1 Estimation Tendency

We have compared each technique's estimation tendencies, divided into three categories; underestimation, exact result and overestimation. All observations were normalized as judged distance / real distance . Values between 0.98 and 1.02 were accepted as exact results. Values smaller than 0.98 were analyzed as underestimated and values larger than 1.02 were analyzed as overestimated.

Table 5.3 and Table 5.2 summarizes estimation results as percentages of all answers for comparative and non-comparative experiments. In general, participants in careless overlay and edge overlay groups overestimated the distances in both experiments. On the other hand, dig box participants gave underestimated results over-all. In non-comparative experiment multi view technique participants overestimated the distances generally, however in exocentric experiment majority of the participants gave an exact answer.

Ellis reported the X-ray condition caused a bias of depth judgments towards the user [116]. In our experiments when comparing the number of overestimated answers to underestimated ones, only dig box technique produced results showing underestimation. In our opinion, this is due to the fact that dig box removes the occlusion with the ground plane, replacing it with its own geometry. Thus results show classic underestimation in augmented reality rather than overestimation in X-ray visualization techniques.

5.6.2 Precision Scenarios

We can group use cases according to their precision requirements. Medical applications and sensitive archeological excavations demand high precision. Other use cases such as urban infrastructure maintenance require medium precision. We can also mention a further category that requires significantly lower precision where only the general area of object is relevant. In example the answer to the question “is there any underground fiber optics cable passing through a given street?” falls in to the last category. For these kinds of tasks edge overlay or even careless overlay prove to be useful. Another use case can be described as occlusion perception, where only requirement is to indicate objects are hidden behind a surface. Techniques other than careless overlay are found to fulfill this requirement. Our user study shows that for categories requiring medium or high precision dig box or multi view technique is more suitable. When relative distances are significant with high precision, our exocentric experiment shows that the multi view technique should be preferred, supplying the highest percentage of exact guesses, 54.54%. Table 5.4 summaries proposed use cases for each technique.

5.6.3 Multi View Technique

Multi view technique produced 1.03 and 0.99 mean values for non-comparative and comparative experiments, also in the exocentric ordering case it lead to 100% correct answers. Accordingly, it is possible to state the multi view technique is perceptually sound. These findings are also in line with our hypothesis that supplying multiple views with correlating visualization improves spatial perception. We can also argue that high number of exact guesses for multi view technique is due to the secondary orthographic view. As this view transforms the correlating top perspective view to a 2D-solvable visualization, participants were able to judge relative distances exceptionally well, resulting in high spatial perception.

In general, participants judged distances with higher precision in comparative than non-comparative experiment. For each group exocentric experiment answers resulted in smaller variance and better or similar means. Dig box participants performed better when the underground pipe was placed closer to the top, middle and the bottom of the box. In these cases participants were able to judge distances more precisely utilizing the box’s own dimension as a depth cue. The proposed multi view technique participants took advantage of multi view based approach and utilized the fixed size of under plane as

well as the grids in secondary view as depth cues for judging distances more accurately. The multi view technique was able to pull away from other techniques by producing the highest number of exact guesses as can be seen in Table 5.3.

One of the limitations of multi view technique is the fixed panel size (1 by 2 meters). For precise exocentric comparisons the relative distance should be smaller than panel size, such as the values used in our exocentric experiment. This limitation can be overcome by providing a panel size that fits the physical environment that is being inspected. It is also possible to let the user change the panel size in runtime via specific user interaction. In this case the orthographic view would also be scaled to match the panel dimensions, allowing precise comparisons.

Chapter 6

Discussion

We have decomposed AR scenes into key components (device, target(s) and task) and first discussed their interrelations and then presented several case studies for aiding professionals in the field.

The relation between devices and targets can be used to gather insight for using appropriate localization approaches. We have proposed a basic transitional framework that allows annotating multiple objects through optimal tracking techniques and using a transitional approach, we provide smooth switches between states.

Through rapid modelling case study, we aim to provide an image based work flow for modelling geographical sites, architectural area, and cultural heritage sites.

Annotation editing is a general concept and found in many subfields of AR. Using underlying 3D models, it is possible to generate, modify and append semantic information in a volumetric manner. Especially in cultural heritage cases, augmenting these volumetric regions with semantic information can produce temporal annotations that is important for recognizing the history of given scene.

Our modeling and annotation workflow is also practical in the sense that it provides compartmentalization for AR field studies. There may be several users only tasked with acquiring initial images for localization process. Then an expert can use the framework to model the objects in the scene. Another expert can be tasked with creating annotations. Finally the deployed scene can be viewed by any user or expert in the field.

Environmental factors have also importance in AR scenes. Growing vegetation, changing seasons, weather conditions even time of day can deteriorate localization and viewing

performance. We have shown that our workflow is robust in this sense as annotations created in daylight can be viewed in night time.

Scene exploration is another crucial task where professionals can examine a scene before the work is even undertaken. We demonstrate an egocentric visualization that leaves the control to the user in the field and allows viewing subterranean geometry that is normally hidden from the naked eye.

This visualization paired with a secondary cutaway view allows users to make precise measurement of vertical distances between objects in the scene. Our user study suggest that this secondary view enhances the measurement accuracy. This user study has also given insight for practical usage scenarios for each of the X-Ray visualization techniques.

AR technology and capabilities tend to change rapidly, however we believe the relation between scene components and the practical advantages their analysis provide are valuable. Moreover, we have chosen case studies as diverse as possible in order to cover a wide range of professional field studies. We believe our research is extendible to a variety of field studies for disciplines including but not limited to: Archaeology, architecture, cultural heritage, tourism, stratigraphy, civil engineering, and urban maintenance

Bibliography

- [1] Strategy Analytics, “Global smartphone installed base forecast by operating system for 88 countries: 2007 to 2017,” October 2012. [Online]. Available: <https://www.strategyanalytics.com/default.aspx?mod=reportabstractviewerI&a0=7834>
- [2] Apple Inc., “App store,” July 2013. [Online]. Available: <https://itunes.apple.com/en/genre/ios/id36?mt=8>
- [3] Google Inc., “Google play store,” July 2013. [Online]. Available: <https://play.google.com/store/apps?hl=en>
- [4] B. MacIntyre, Ed., *Proceedings of 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2012.
- [5] S. Coquillart, Ed., *Proceedings of 2013 IEEE Virtual Reality Conference*, 2013.
- [6] J. R. Blum, D. Greencorn, and J. R. Cooperstock, “Smartphone sensor reliability for augmented reality applications,” in *Proc. MobiQuitous*, 2012.
- [7] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, “A survey of mobile phone sensing,” *Communications Magazine, IEEE*, vol. 48, no. 9, pp. 140–150, 2010.
- [8] M. Graham, M. Zook, and A. Boulton, “Augmented reality in urban places: contested content and the duplicity of code,” *Transactions of the Institute of British Geographers*, vol. 38-3, pp. 464–479, 2012.
- [9] R. Bane and T. Höllerer, “Interactive tools for virtual x-ray vision in mobile augmented reality,” in *Mixed and Augmented Reality, 2004. ISMAR 2004. Third IEEE and ACM International Symposium on*. IEEE, 2004, pp. 231–239.
- [10] R. T. Azuma, “A survey of augmented reality,” *Presence*, vol. 6, no. 4, pp. 355–385, 1997.

- [11] R. Azuma, Y. Baillet, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," *Computer Graphics and Applications, IEEE*, vol. 21, no. 6, pp. 34–47, 2001.
- [12] Volvo Inc., "All female team create award-winning concept car," August 2006.
- [13] R. Clarke, *British Aircraft Armament: Raf Guns and Gunsights from 1914 to the Present Day*, ser. British Aircraft Armament. Patrick Stephens, 1995.
- [14] Airbus Inc., "Airbus a318 approved for head up display," December 2007. [Online]. Available: <http://web.archive.org/web/20071207164904/http://www.airbus.com/en/presscentre/pressreleases/pressreleases-items/07-12-03-a318-hud.html>
- [15] I. E. Sutherland, "A head-mounted three dimensional display," in *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*. ACM, 1968, pp. 757–764.
- [16] T. P. Caudell and D. W. Mizell, "Augmented reality: An application of heads-up display technology to manual manufacturing processes," in *System Sciences, 1992. Proceedings of the Twenty-Fifth Hawaii International Conference on*, vol. 2. IEEE, 1992, pp. 659–669.
- [17] S. Feiner, B. MacIntyre, T. Höllerer, and A. Webster, "A touring machine: Prototyping 3d mobile augmented reality systems for exploring the urban environment," *Personal Technologies*, vol. 1, no. 4, pp. 208–217, 1997.
- [18] V. Lepetit and P. Fua, *Monocular model-based 3d tracking of rigid objects: A survey*. Now Publishers Inc, 2005.
- [19] T. Drummond and R. Cipolla, "Real-time visual tracking of complex structures," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 932–946, 2002.
- [20] G. Klein, *Visual Tracking for Augmented Reality: Edge-based tracking techniques for AR applications*. VDM Publishing, 2009.
- [21] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *Acm Computing Surveys (CSUR)*, vol. 38, no. 4, p. 13, 2006.
- [22] D. Wagner, A. Mulloni, T. Langlotz, and D. Schmalstieg, "Real-time panoramic mapping and tracking on mobile phones," in *Virtual Reality Conference (VR), 2010 IEEE*. IEEE, 2010, pp. 211–218.

- [23] T. Langlotz, D. Wagner, A. Mulloni, and D. Schmalstieg, “Online creation of panoramic augmented reality annotations on mobile phones,” *Pervasive Computing, IEEE*, vol. 11, no. 2, pp. 56–63, 2012.
- [24] B. Jiang, U. Neumann, and S. You, “A robust hybrid tracking system for outdoor augmented reality,” in *Virtual Reality, 2004. Proceedings. IEEE*. IEEE, 2004, pp. 3–275.
- [25] Layar Inc., “Layar augmented reality browser,” July 2011. [Online]. Available: <http://www.layar.com/>
- [26] Wikitude Org., “Wikitude augmented reality browser,” July 2011. [Online]. Available: <http://www.wikitude.me/>
- [27] T. Lee and T. Höllerer, “Hybrid feature tracking and user interaction for markerless augmented reality,” in *Virtual Reality Conference, 2008. VR’08. IEEE*. IEEE, 2008, pp. 145–152.
- [28] G. Schall, D. Wagner, G. Reitmayr, E. Taichmann, M. Wieser, D. Schmalstieg, and B. Hofmann-Wellenhof, “Global pose estimation using multi-sensor fusion for outdoor augmented reality,” in *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*. IEEE, 2009, pp. 153–162.
- [29] G. Bleser and D. Stricker, “Advanced tracking through efficient image processing and visual–inertial sensor fusion,” *Computers and Graphics*, vol. 33, no. 1, pp. 59–72, 2009.
- [30] Autodesk Inc., “Autodesk maya,” April 2012. [Online]. Available: <http://usa.autodesk.com/maya/>
- [31] Blender Foundation, “Blender,” April 2013. [Online]. Available: <http://www.blender.org>
- [32] Eos Systems, “Photomodeler: A commercial photogram-metry product,” April 2005. [Online]. Available: <http://www.photomodeler.com>
- [33] M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch, “Visual modeling with a hand-held camera,” *International Journal of Computer Vision*, vol. 59, no. 3, pp. 207–232, 2004.
- [34] P. E. Debevec, C. J. Taylor, and J. Malik, “Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach,” in *Proceedings*

of the 23rd annual conference on Computer graphics and interactive techniques. ACM, 1996, pp. 11–20.

- [35] PFTRACK, “Thepixelfarm,” 2010, a commercial camera tracking and image based modelling product. [Online]. Available: <http://www.thepixelfarm.co.uk>
- [36] T. Thormählen and H. Broszio, “Voodoo camera tracker,” 2010. [Online]. Available: www.digilab.uni-hannover.de
- [37] A. van den Hengel, A. Dick, T. Thormählen, B. Ward, and P. H. Torr, “Videotrace: rapid interactive scene modelling from video,” in *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3. ACM, 2007, p. 86.
- [38] S. N. Sinha, D. Steedly, R. Szeliski, M. Agrawala, and M. Pollefeys, “Interactive 3d architectural modeling from unordered photo collections,” in *ACM Transactions on Graphics (TOG)*, vol. 27, no. 5. ACM, 2008, p. 159.
- [39] T. Thormählen and H.-P. Seidel, “3d-modeling by ortho-image generation from image sequences,” in *SIGGRAPH '08: ACM SIGGRAPH 2008 papers*. ACM, 2008, pp. 1–5.
- [40] R. Grzeszczuk, J. Kosecka, R. Vedantham, and H. Hile, “Creating compact architectural models by geo-registering image collections,” in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 1718–1725.
- [41] N. Snavely, S. M. Seitz, and R. Szeliski, “Photo tourism: exploring photo collections in 3d,” in *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*. ACM, 2006, pp. 835–846.
- [42] J. Xiao, T. Fang, P. Tan, P. Zhao, E. Ofek, and L. Quan, “Image-based façade modeling,” in *SIGGRAPH Asia '08: ACM SIGGRAPH Asia 2008 papers*. ACM, 2008, pp. 1–10.
- [43] J. Rekimoto and K. Nagao, “The world through the computer: Computer augmented interaction with real world environments,” in *Proceedings of the 8th annual ACM symposium on User interface and software technology*. ACM, 1995, pp. 29–36.
- [44] J. Wither, S. DiVerdi, and T. Höllerer, “Annotation in outdoor augmented reality,” *Computers & Graphics*, vol. 33, no. 6, pp. 679–689, 2009.

- [45] J. Wither and T. Höllerer, "Pictorial depth cues for outdoor augmented reality," in *ISWC '05: Proceedings of the Ninth IEEE International Symposium on Wearable Computers*. IEEE Computer Society, 2005, pp. 92–99.
- [46] G. Schall, E. Mendez, and D. Schmalstieg, "Virtual redlining for civil engineering in real environments," in *Mixed and Augmented Reality, 2008. ISMAR 2008. 7th IEEE/ACM International Symposium on*. IEEE, 2008, pp. 95–98.
- [47] R. Azuma and C. Furmanski, "Evaluating label placement for augmented reality view management," in *Proceedings of the 2nd IEEE/ACM international Symposium on Mixed and Augmented Reality*. IEEE Computer Society, 2003, p. 66.
- [48] S. Henderson and S. Feiner, "Opportunistic tangible user interfaces for augmented reality," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 16, no. 1, pp. 4–16, 2010.
- [49] W. Piekarski, "3d modeling with the tinmith mobile outdoor augmented reality system," *Computer Graphics and Applications, IEEE*, vol. 26, no. 1, pp. 14–17, 2006.
- [50] T. Höllerer, S. Feiner, T. Terauchi, G. Rashid, and D. Hallaway, "Exploring mars: developing indoor and outdoor user interfaces to a mobile augmented reality system," *Computers and Graphics*, vol. 23, no. 6, pp. 779–785, 1999.
- [51] Y. Baillet, D. Brown, and S. Julier, "Authoring of physical models using mobile computers," in *Wearable Computers, 2001. Proceedings. Fifth International Symposium on*. IEEE, 2001, pp. 39–46.
- [52] A. van den Hengel, "Image-based modelling for augmenting reality," in *Ubiquitous Virtual Reality (ISUVR), 2010 International Symposium on*. IEEE, 2010, pp. 1–4.
- [53] H. Benko, E. W. Ishak, and S. Feiner, "Collaborative mixed reality visualization of an archaeological excavation," in *Mixed and Augmented Reality, 2004. ISMAR 2004. Third IEEE and ACM International Symposium on*. IEEE, 2004, pp. 132–140.
- [54] E. C. Harris, *Principles of archaeological stratigraphy*. Academic, 1989.
- [55] E. B. Goldstein, *Sensation and perception*. Cengage Learning, 2013.

- [56] J. Swan, M. A. Livingston, H. S. Smallman, D. Brown, Y. Baillot, J. L. Gabbard, and D. Hix, "A perceptual matching technique for depth judgments in optical, see-through augmented reality," in *Virtual Reality Conference, 2006*. IEEE, 2006, pp. 19–26.
- [57] M. S. Landy, L. T. Maloney, E. B. Johnston, and M. Young, "Measurement and modeling of depth cue combination: In defense of weak fusion," *Vision research*, vol. 35, no. 3, pp. 389–412, 1995.
- [58] J. S. Lappin, A. L. Shelton, and J. J. Rieser, "Environmental context influences visually perceived distance," *Perception and Psychophysics*, vol. 68, no. 4, pp. 571–581, 2006.
- [59] A. Dey, A. Cunningham, and C. Sandor, "Evaluating depth perception of photorealistic mixed reality visualizations for occluded objects in outdoor environments," in *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*. ACM, 2010, pp. 211–218.
- [60] E. Kruijff, J. Swan, and S. Feiner, "Perceptual issues in augmented reality revisited," in *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*. IEEE, 2010, pp. 3–12.
- [61] J. A. Jones, J. E. Swan II, G. Singh, E. Kolstad, and S. R. Ellis, "The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception," in *Proceedings of the 5th symposium on Applied perception in graphics and visualization*. ACM, 2008, pp. 9–14.
- [62] M. A. Livingston, Z. Ai, J. E. Swan, and H. S. Smallman, "Indoor vs. outdoor depth perception for mobile augmented reality," in *Virtual Reality Conference, 2009. VR 2009. IEEE*. IEEE, 2009, pp. 55–62.
- [63] B. Avery, C. Sandor, and B. H. Thomas, "Improving spatial perception for augmented reality x-ray vision," in *Virtual Reality Conference, 2009. VR 2009. IEEE*. IEEE, 2009, pp. 79–82.
- [64] E. A. Bier, M. C. Stone, K. Pier, W. Buxton, and T. D. DeRose, "Toolglass and magic lenses: the see-through interface," in *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*. ACM, 1993, pp. 73–80.
- [65] G. Schall, E. Mendez, E. Kruijff, E. Veas, S. Junghanns, B. Reitinger, and D. Schmalstieg, "Handheld augmented reality for underground infrastructure visualization," *Personal and ubiquitous computing*, vol. 13, no. 4, pp. 281–291, 2009.

- [66] G. R. King, W. Piekarski, and B. H. Thomas, "Arvino-outdoor augmented reality visualisation of viticulture gis data," in *Mixed and Augmented Reality, 2005. Proceedings. Fourth IEEE and ACM International Symposium on*. IEEE, 2005, pp. 52–55.
- [67] S. Zollmann, D. Kalkofen, E. Mendez, and G. Reitmayr, "Image-based ghostings for single layer occlusions in augmented reality," in *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*. IEEE, 2010, pp. 19–26.
- [68] M. A. Livingston, J. E. Swan II, J. L. Gabbard, T. H. Höllerer, D. Hix, S. J. Julier, Y. Baillot, and D. Brown, "Resolving multiple occluded layers in augmented reality," in *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE Computer Society, 2003, p. 56.
- [69] A. Dey, G. Jarvis, C. Sandor, A. K. Wibowo, and V.-V. Mattila, "An evaluation of augmented reality x-ray vision for outdoor navigation," Ph.D. dissertation, Virtual Reality Society of Japan, 2011.
- [70] O'Neil et al., "Star Light, Star Bright - Death Star I See Tonight!" pp. 28–32, 1969.
- [71] J. Kruger, J. Schneider, and R. Westermann, "Clearview: An interactive context preserving hotspot visualization technique," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 12, no. 5, pp. 941–948, 2006.
- [72] C. Bichlmeier, F. Wimme, S. M. Heining, and N. Navab, "Contextual anatomic mimesis hybrid in-situ visualization method for improving multi-sensory depth perception in medical augmented reality," in *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*. IEEE, 2007, pp. 129–138.
- [73] C. Bichlmeier, M. Kipot, S. Holdstock, S. M. Heining, E. Euler, and N. Navab, "A practical approach for intraoperative contextual in-situ visualization," in *Proceedings of 5th Workshop on Augmented Environments for Medical Imaging including Augmented Reality in Computer-Aided Surgery*. AMIARCS, 2009.
- [74] S. Julier, M. Lanzagorta, Y. Baillot, L. Rosenblum, S. Feiner, T. Hollerer, and S. Sestito, "Information filtering for mobile augmented reality," in *Augmented Reality, 2000.(ISAR 2000). Proceedings. IEEE and ACM International Symposium on*. IEEE, 2000, pp. 3–11.

- [75] D. Kalkofen, E. Mendez, and D. Schmalstieg, “Comprehensible visualization for augmented reality,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 15, no. 2, pp. 193–204, 2009.
- [76] I. Zedjebil, F.-e. Ababsa, J.-Y. Didier, J. Vairon, L. Frauciel, M. Hachet, P. Guitton, R. Delmont *et al.*, “Outdoor augmented reality: State of the art and issues,” in *10th ACM/IEEE Virtual Reality International Conference (VRIC 2008)*, 2008, pp. 177–187.
- [77] Apple Inc., “ios sdk 6.1.3,” July 2013. [Online]. Available: <https://developer.apple.com/devcenter/ios/index.action>
- [78] Google Inc., “Android sdk rev. 18,” July 2013. [Online]. Available: <http://developer.android.com/sdk/index.html>
- [79] United States Government, “Global positioning system official webpage,” July 2013. [Online]. Available: <http://www.gps.gov/>
- [80] United States Air Force, “Gps modernization fact sheet,” July 2013. [Online]. Available: <http://www.pnt.gov/public/docs/2006/modernization.pdf>
- [81] M. S. Grewal, A. P. Andrews, and C. G. Bartone, *Global Navigation Satellite Systems, Inertial Navigation, and Integration*. John Wiley and Sons, 2013.
- [82] A. Leick, *GPS satellite surveying*. Wiley. com, 2004.
- [83] P. Misra and P. Enge, “Global positioning system: Signals, measurements and,” 2011.
- [84] Google Inc., “Google maps,” July 2013. [Online]. Available: <http://maps.google.com>
- [85] Samsung Inc., “Samsung galaxy s2,” July 2013. [Online]. Available: <http://www.samsung.com/global/microsite/galaxys2/html/specification.html>
- [86] G. Silveira and E. Malis, “Real-time visual tracking under arbitrary illumination changes,” in *Computer Vision and Pattern Recognition, 2007. CVPR’07. IEEE Conference on*. IEEE, 2007, pp. 1–6.
- [87] R. Guo, Q. Dai, and D. Hoiem, “Single-image shadow detection and removal using paired regions,” in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 2033–2040.

- [88] Qualcomm Inc., “Vuforia sdk 2.5.2,” June 2013. [Online]. Available: www.vuforia.com
- [89] B. Huyghe, J. Doutreloigne, and J. Vanfleteren, “3d orientation tracking based on unscented kalman filtering of accelerometer and magnetometer data,” in *Sensors Applications Symposium, 2009. SAS 2009. IEEE*. IEEE, 2009, pp. 148–152.
- [90] M. T. Eren and S. Balcisoy, “Transitional outdoor augmented reality accompanying video,” June 2013. [Online]. Available: <http://dl.dropbox.com/u/871904/paper18.mp4>
- [91] Q. Luan, S. M. Drucker, J. Kopf, Y.-Q. Xu, and M. F. Cohen, “Annotating gigapixel images,” in *Proceedings of the 21st annual ACM symposium on User interface software and technology*. ACM, 2008, pp. 33–36.
- [92] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, and D. Lischinski, “Deep photo: Model-based photograph enhancement and viewing,” in *ACM Transactions on Graphics (TOG)*, vol. 27, no. 5. ACM, 2008, p. 116.
- [93] R. I. Hartley, “In defense of the eight-point algorithm,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 6, pp. 580–593, 1997.
- [94] J.-D. Boissonnat, O. Devillers, M. Teillaud, and M. Yvinec, “Triangulations in cgal,” in *Proceedings of the sixteenth annual symposium on Computational geometry*. ACM, 2000, pp. 11–18.
- [95] M. Barnes, “Collada,” in *SIGGRAPH ’06: ACM SIGGRAPH 2006 Courses*. ACM, 2006, p. 8.
- [96] W. Piekarski and B. H. Thomas, “Interactive augmented reality techniques for construction at a distance of 3d geometry,” in *Proceedings of the workshop on Virtual environments 2003*. ACM, 2003, pp. 19–28.
- [97] D. H. Kim and M.-J. Kim, “A new modeling interface for the pen-input displays,” *Computer-Aided Design*, vol. 38, no. 3, pp. 210–223, 2006.
- [98] Google Inc., “Google sky map,” April 2012. [Online]. Available: <http://www.google.com/mobile/skymap/>
- [99] N. Boubaki, A. Saintenoy, and P. Tucholka, “Gpr profiling and electrical resistivity tomography for buried cavity detection: A test site at the abbaye de l’ouye

- (france),” in *Advanced Ground Penetrating Radar (IWAGPR), 2011 6th International Workshop on*. IEEE, 2011, pp. 1–5.
- [100] L. Lo Monte, D. Erricolo, V. Picco, F. Soldovieri, and M. C. Wicks, “Distributed rf tomography for tunnel detection: Suitable inversion schemes,” in *Aerospace and Electronics Conference (NAECON), Proceedings of the IEEE 2009 National*. IEEE, 2009, pp. 182–189.
- [101] L. Lo Monte, D. Erricolo, F. Soldovieri, and M. C. Wicks, “Radio frequency tomography for tunnel detection,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, no. 3, pp. 1128–1137, 2010.
- [102] A. Simi, S. Bracciali, and G. Manacorda, “Hough transform based automatic pipe detection for array gpr: algorithm development and on-site tests,” in *Radar Conference, 2008. RADAR’08. IEEE*. IEEE, 2008, pp. 1–6.
- [103] G. Schall, D. Schmalstieg, and S. Junghanns, “Vidente-3d visualization of underground infrastructure using handheld augmented reality,” *Geohydroinformatics-Integrating GIS and Water Engineering" CRC Press/Taylor and Francis Publisher: CRC*, vol. 1, pp. 1–17, 2010.
- [104] J. C. Roberts, “On encouraging multiple views for visualization,” in *Information Visualization, 1998. Proceedings. 1998 IEEE Conference on*. IEEE, 1998, pp. 8–14.
- [105] A. H. Behzadan and V. R. Kamat, “Visualization of construction graphics in outdoor augmented reality,” in *Proceedings of the 37th conference on Winter simulation*. Winter Simulation Conference, 2005, pp. 1914–1920.
- [106] G. Schall, E. Mendez, S. Junghanns, and D. Schmalstieg, “Urban 3d models: what’s underneath? handheld augmented reality for subsurface infrastructure visualization,” in *Proceedings of Ubicomp, 2007*.
- [107] F. Robert, R. Brommecker, B. T. Bourne, P. J. Dobak, C. J. McEwan, R. R. Rowe, and X. Zhou, “Models and exploration methods for major gold deposit types,” in *Proceedings of Exploration 07: Fifth Decennial International Conference on Mineral Exploration, 2007*, pp. 691–711.
- [108] M. Smith, I. Sheppard, G. Karunatillake, and F. Camisani-Calzolari, “Using mip for strategic life-of-mine planning of the lead/zinc stream at mount isa mines,” in *APCOM 2003 Conference, 2003*, pp. 1–10.

- [109] W. Li, L. Ritter, M. Agrawala, B. Curless, and D. Salesin, “Interactive cutaway illustrations of complex 3d models,” *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3, p. 31, 2007.
- [110] D. Kalkofen, M. Tatzgern, and D. Schmalstieg, “Explosion diagrams in augmented reality,” in *Virtual Reality Conference, 2009. VR 2009. IEEE*. IEEE, 2009, pp. 71–78.
- [111] E. Mendez and D. Schmalstieg, “Importance masks for revealing occluded objects in augmented reality,” in *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*. ACM, 2009, pp. 247–248.
- [112] N. Senthilkumaran and R. Rajesh, “Edge detection techniques for image segmentation—a survey of soft computing approaches,” *International Journal of Recent Trends in Engineering*, vol. 1, no. 2, pp. 250–254, 2009.
- [113] S. Bruckner and M. Groller, “Enhancing depth-perception with flexible volumetric halos,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 13, no. 6, pp. 1344–1351, 2007.
- [114] B. Freudenberg, M. Masuch, and T. Strothotte, “Real-time halftoning: a primitive for non-photorealistic shading,” in *Proceedings of the 13th Eurographics workshop on Rendering*. Eurographics Association, 2002, pp. 227–232.
- [115] A. Gooch, B. Gooch, P. Shirley, and E. Cohen, “A non-photorealistic lighting model for automatic technical illustration,” in *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*. ACM, 1998, pp. 447–452.
- [116] S. R. Ellis and B. M. Menges, “Localization of virtual objects in the near visual field,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 40, no. 3, pp. 415–431, 1998.