An Augmented Reality X-Ray System Based on Visual Saliency

Christian Sandor*
University of South Australia
Magic Vision Lab

Andrew Cunningham†
University of South Australia
Magic Vision Lab

Arindam Dey‡
University of South Australia
Magic Vision Lab

Ville-Veikko Mattila§
Nokia
Nokia Research Center

(a) (b) (c)

Figure 1: Our novel Augmented Reality X-ray system (a) provides users with more high-level context of the occluder object than our previous system (b). In this example, important visual landmarks such as the white umbrellas are preserved. We achieve this effect by determining important visual landmarks through saliency maps (c).

ABSTRACT

In the past, several systems have been presented that enable users to view occluded points of interest using Augmented Reality X-ray visualizations. It is challenging to design a visualization that provides correct occlusions between occluder and occluded objects while maximizing legibility. We have previously published an Augmented Reality X-ray visualization that renders edges of the occluder region over the occluded region to facilitate correct occlusions while providing foreground context. While this approach is simple and works in a wide range of situations, it provides only minimal context of the occluder object.

In this paper, we present the background, design, and implementation of our novel visualization technique that aims at providing users with richer context of the occluder object. While our previous visualization only employed one salient feature (edges) to determine which parts of the occluder to display, our novel visualization technique is an initial attempt to explore the design space of employing multiple salient features for this task. The prototype presented in this paper employs three additional salient features: hue, luminosity, and motion.

We have conducted two evaluations with human participants to investigate the benefits and limitations of our prototype compared to our previous system. The first evaluation showed that although our novel visualization provides a richer context of the occluder object, it does not impede users to select objects in the occluded area; but, it also indicated problems in our prototype. In the second evaluation, we have investigated these problems through an online survey with systematically varied occluder and occluded scenes, focussing on the qualitative aspects of our visualizations. The results were encouraging, but pointed out that our novel visualization needs a higher level of adaptiveness.

*e-mail: chris.sandor@gmail.com
†e-mail: andrewcunningham@mac.com
‡e-mail: arindam.dey@postgrads.unisa.edu.au
§e-mail: ville-veikko.mattila@nokia.com

Keywords: augmented reality, visualization, evaluation, augmented reality X-ray, saliency

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities H.1.2 [Models and Principles]: User/Machine Systems—Human factors

1 INTRODUCTION

Throughout the last decade, mobile information browsing has become a widely-adopted practice. Most of today's mobile internet devices contain facilities to display maps of the user's surroundings with points of interest (POIs) embedded into the map. Other researchers have already explored complementary, egocentric visualizations of these POIs using mobile Augmented Reality (AR), since AR can enhance a user's perception of their environment by showing information that they cannot directly sense[6]. While early systems have relied on symbolic representations of the occluded POIs, the most recent systems aim at providing users with photorealistic views by rendering a "superman"-like X-ray, seeing through objects to the *occluded regions* (throughout this paper, we refer to objects occluding the POIs as *occluder regions*). This rendering represents the user's current view and the occluded region in context on the display, reducing cognitive load.

Rendering AR X-ray has many unanswered questions. Rendering the occluded region naively over the real world image can cause the occluded region to appear to float in front of the real world and lose the contextual information of the occluder region. AR X-ray must be rendered carefully to address these issues to improve cognition of the occluded and occluder regions. We have previously published an AR X-ray technique[1] that renders edges of the occluder region over the occluded region to help maintain context. While this improved spatial perception, it is not ideal, as important

information about the occluder gets lost; for example, color.

Novel visualization techniques are required to optimally render AR X-ray. Enough visual information of the occluder region should be provided to aid depth perception, recognition of the relationship between the occluder and occluded region, and the contents of occluder and occluded area. Conversely, too much visual information from the occluder can cause distracting visual noise and hide essential regions of the occluded area.

An understanding of human perception can aid the design of AR X-ray. Contrasts in the visual features of the image, including color, luminosity, orientation and motion, determine *salient regions*[16]. Salient regions can be understood as the regions in an image, which are most likely to attract the viewer's gaze. Our previous prototype preserved edges in the foreground, which are one type of salient regions. In this work, we have conducted an initial exploration of the design space of using multiple salient features.

Contribution The core contribution of this paper is the design, implementation, and evaluation of a novel AR X-ray system. While our previous AR X-ray visualization technique only employed one salient feature (edges) to determine which parts of the occluder to display, our novel visualization technique is an initial attempt to explore the design space of employing multiple salient features for this task. The prototype presented in this paper employs three additional salient features: hue, luminosity, and motion. The goal of this approach is to provide users with richer context of the occluder object.

We have conducted two evaluations with human participants to investigate the benefits and limitations of our prototype compared to our previous system. The first evaluation showed that although our novel visualization provides a richer context of the occluder object, it does not impede users to select objects in the occluded area; but, it also indicated problems in our prototype. In the second evaluation, we have investigated these problems through an online survey with systematically varied occluder and occluded scenes, focusing on the qualitative aspects of our visualizations. The results were encouraging, but pointed out that our novel visualization needs a higher level of adaptiveness.

The rest of this paper is structured as follows: in Section 2, we discuss related work for AR X-ray and visual saliency. In Section 3, we describe the visualization design for our prototype. In Section 4, we present two evaluations that we have performed. Finally, Section 5 concludes by discussing the benefits and limitations of our prototype in the context of the evaluation results and by pointing out directions for future work.

2 Related Work

In this section we review the theory behind visual saliency, the concept we apply to AR X-ray, summarize previous AR X-ray approaches, and discuss work that combines both visual saliency and AR.

2.1 Human Perception and Visual Saliency

In a modern view of visual perception, it is recognized that the eye receives comparatively little information compared to that which our perception system infers[20]). Our perception system builds a collage of visual information through visual working memory, knowledge, and visual attention. Perception combines both a bottom-up process, driven by the raw visual information received by the eye, and a top-down process, driven by the task at hand, to direct attention. In this paper, we focus upon a bottom-up model.

The bottom-up process is driven by the low-level visual stimuli received by the eye. This is preceded by a preattentive process, originally proposed by Treisman and Gelade[16] in their Feature Integration Theory. The preattentive process registers particular visual features to determine which regions of view require our conscious attention. These visual features, or cues, are registered in parallel, and include properties such as color hue, shape, color intensity, and motion. A more complete list of cues was presented by Healey[3].

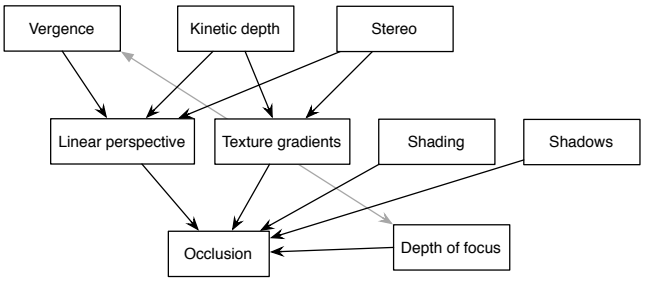


Figure 2: A dependency graph for depth cues. Arrows indicate how depth cues depend on each other for undistorted appearance. Note how occlusion is the most fundamental depth cue.

Given these preattentive visual features, Koch and Ullman[10] propose a model of conspicuous or salient regions within view. Their model relies on a saliency map; a topographic map that does not code for peaks in a particular feature, but instead codes for contrasts in features. A red ball on a green background, for example, would be encoded as a peak in this saliency map. A popular toolkit for implementing saliency is the iLab Neuromorphic Vision C++ Toolkit[17]. Although it is very similar in spirit to our approach, it is not designed for realtime processing.

Several systems have been presented that employ visual saliency for adapting graphics output in a variety of applications: volume rendering[8], mesh simplification[13], image resizing[15], and lighting[12]. However, only very few systems have been published that employ realtime saliency-based adaption of graphics in an AR context; we discuss those in the next section.

2.2 Augmented Reality X-ray

AR X-ray can represent an occluded region in-situ with the user's current view, improving cognition by representing both as a single, unified visual event instead of disparate events. This is recognized as requiring greater cognitive effort[5]. Naively rendering AR X-ray, such as rendering the occluded region on top of the occluder can be problematic. Bajura et al. first observed the issues that occur when rendering occluded regions[2], noting that "the [occluded] images did not appear to be inside the subject, so much as pasted on top of her". Furthermore, such a rendering simply swaps the roles of the occluder and occluded scene, loosing any perceptive contribution the occluder makes to the overall scene[14]. This is primarily an issue of depth perception, and several works have attempted to address this by rendering portions of the occluder over the occluded region.

While it may seem counter intuitive, occlusion is the most important depth-cue. Figure 2 (adapted from[20]) highlights this fact, as all other depth cues depend on it. Therefore, it is very difficult to have incorrect occlusions and a perceptually coherent scene. Our approach aims at preserving correct occlusions as much as possible.

Krüger et al.[11] render the occluder over the occluded object with volume based rendering techniques. The authors treat X-ray as a focus+context visualization technique, where the occluded region is the focus and the occluder is the context. Portions of the occluder are transparent based on their proximity to the focus within the user's view. While this is a compelling approach, they do not apply their visualization technique to an AR environment and it requires volumetric data of the scene. Similar to Krüger et al., Kalkofen et al.[7] describe an X-ray focus+context technique, but apply it in an AR environment. Their technique also expands upon Krüger et al.'s by applying several stylized renderings to the occluder region, such as an edge overlay and a haloing technique. In previous work,

(a) (b) (c) (d) (e) (f) (g)

Figure 3: Examples of how salient features affect the final composition results. (a,b) *Hue*: strong colors (red, yellow) are preserved both in the occluder and in the occluded area. (c-g) *Motion*: A moving person (c,d,f,g) is displayed, whereas a standing person (e) is suppressed (note: only motion saliency has been applied).

we have applied Kalkofen et al.'s visualization technique to outdoor AR X-ray[1], and found that, while the edge overlay did contribute to depth perception and spatial understanding, a more careful visualization design is required to reduce visual noise. Furmanski et al.[6] clarify this point by observing two main issues in AR X-ray (and related) visualizations: depth ambiguity, as we have noted, and visual complexity.

The work presented in this paper aims at providing a heuristic for automatically selecting important regions of the occluder that should be preserved. While our approach has the advantage of being applicable without any advance preparation or user intervention, there exist two complementary approaches that enable preservation of foreground regions pertinent to a specific task.

The first approach, presented by Mendez and Schmalstieg[14], uses predefined importance maps to determine the opacity of the occluder region. A predefined map is computationally inexpensive to render, can encode prior semantic knowledge, and is not affected by changing lighting conditions. However, predefined importance maps can be expensive to create, especially for outdoor environments, as the designers of these maps must consider what is salient to a particular task. The second approach, presented by Zollmann and colleagues[21], gives users interactive control over which parts of the foreground to preserve.

3 Visualization Design

As we have discussed in the background section, an AR X-ray visualization design must selectively show portions of the occluder to provide depth cues while minimizing visual complexity. Our approach to this problem is an AR X-ray system utilizing visual saliency. Salient features in both the occluded and occluder regions should be emphasized. Our visualization design takes a bottom-up approach to visual saliency, based purely on visual data.

Our novel AR X-ray visualization technique uses saliency maps to determine the relative contribution of the occluder and occluded region to the final composed image. We first compute the saliency of both the occluded and occluder regions. Second, we perform a saliency based image composition. Salient regions in the occluder that do not cover salient regions in the occluded region are made opaque, while non-salient regions are transparent. When salient features overlap, we blend them according to their strength. With this method, we reduce visual noise while still maintaining strong depth cues. Before we give details for the two stages (Saliency Map Computation (Section 3.1) and Composition (Section 3.2), we describe the core effects of our novel X-ray technique.

Compared to our old AR X-ray system, which only used edges as salient feature, our novel system additionally employs luminosity, hue, and motion. Figure 1(a) illustrates the effect for luminosity, as the white umbrellas of the occluder are preserved. Figure 3(a,b) shows how hue is employed to preserve strong colors in both occluder and occluded region. Finally, Figure 3(c-g) highlights how motion affects the visualization: while a person who is standing still and does not contain strong salient features is suppressed in the foreground, they are not suppressed anymore when they are moving (due to the saliency of moving objects). The rationale for these effects is that navigational landmarks often have strong hue and luminosity. For example, the sign in Figure 3(b) provides useful navigational information. The motion feature is motivated differently. Moving objects are seldom navigational landmarks; instead, moving objects can be dangerous to the user if they overlook it. Therefore, we decided to add motion as a salient feature mainly for safety reasons.

3.1 Saliency Map Computation

Our saliency computational model is based on Walther's[19], as illustrated in Figure 4. The sensory properties of the human eye are recognized to form a hierarchy of receptive cells that respond to contrast between different levels to identify regions that stand out from their surroundings. This hierarchy is modeled by subsampling an input image I into a dyadic pyramid of $\sigma = [0 \dots 8]$, such that the resolution of level σ is $1/2^{\sigma}$ the resolution of the original image. From this image pyramid, P_{σ}, we extract the visual features of luminosity l, color hue opponency c, and motion t.

Luminosity is the brightness of the color component, and is defined as:

$$M_l = \frac{r+g+b}{3}$$

Color hue opponency mimics the visual system's ability to distinguish opposing color hues. Illumination independent Red-Green and Blue-Yellow opponency maps are defined as:

$$M_{rg} = \frac{r-g}{max(r,g,b)}$$

$$M_{by} = \frac{b-min(r,g)}{max(r,g,b)}$$

These maps M_{rg} and M_{by} are combined into a single map M_c.

Motion is defined as observed changes in the luminosity channel over time.

Contrasts in the dyadic feature pyramids are modeled as across scale subtraction $\ominus$ between fine and coarse scaled levels of the pyramid. For each of the features, a set of feature maps are generated as:

$$F_{f,p,s} = P_p \ominus P_s$$

where f represents the visual feature $f \in \{l,c,m\}$. p and s refer to pyramid levels and are applied as $p \in \{2,3,4\}$, $s = p+S$, and

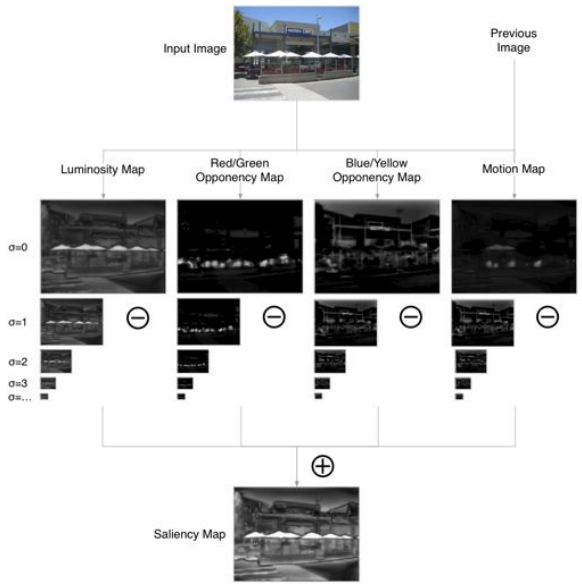


Figure 4: Saliency map computation: an input image is split into feature maps which are across-scale subtracted to mimic the receptive fields of the human eye. The features maps are combined to yield the final saliency map.
$\oplus$ and $\ominus$ denote across scale addition and across-scale subtraction.

$S \in \{3,4\}$. Features maps are combined using across-scale addition $\oplus$ to yield conspicuity maps:

$$C = \bigoplus_{p=2}^{4} \bigoplus_{s=p+3}^{p+4} F_{p,s}$$

Finally, all conspicuity maps are combined to form the saliency map:

$$S = \frac{1}{3} \sum_{k \in \{I,c,t\}} C_k$$

At this point, a saliency map has been created for an image, combining the hue, luminosity and motion features. In the next stage, occluded and occluder regions are composed using their saliency information to create the final AR X-ray image.

3.2 Composition

Figure 5 illustrates our composition method. Saliency maps S_o and S_d are generated for both the occluder I_o and occluded I_d images respectively. Further to this, we highlight edges in the occluder to emphasize structure. An edge map E is generated from the occluder region and weighted with the occluder saliency map:

$$E = \gamma(I_o) \times S_o \times \varepsilon$$

Where γ is a Sobel edge function and ε is a weighting constant. This edge map is combined with the occluder saliency map as an addition, $S_{o'} = S_o + E$. We combine $S_{o'}$ and S_d to create the final saliency map indicating the transparency of the occluder. We assume that salient regions of the occluder should take precedence over salient regions of the occluded.

A mask M and inverse mask M' is generated to reveal only the portion of the occluded region we are concerned with. Given this, we create the final image composition as:

$$I_c = S_{o'} \times M + P_o \times M + P_d \times M'$$

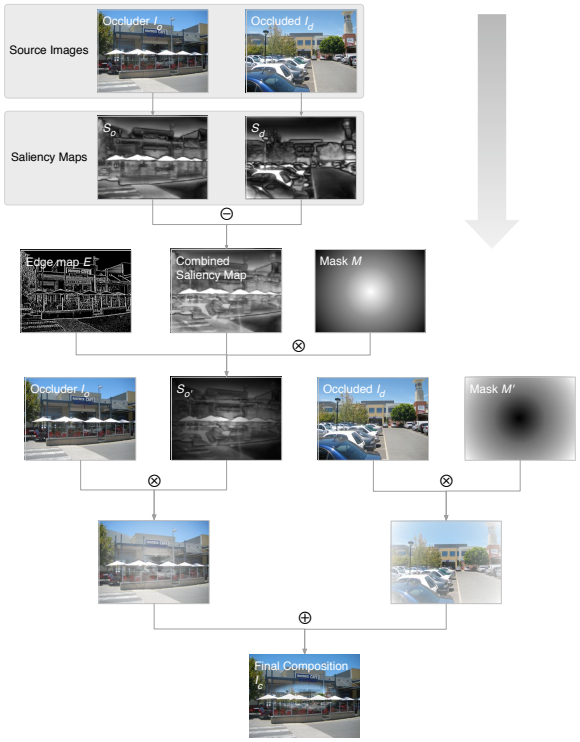


Figure 5: Composition: source images are processed through a series of filters and combinations to produce the final output image.
$\oplus$, $\ominus$, and $\otimes$ denote addition, subtraction, and multiplication of pixel values.

This final composition contains the salient features of the occluded and occluder images as well as revealing structure of the occluder with emphasized edges.

4 Evaluation

We have performed two empirical studies to compare our new saliency-based AR X-ray with our previous edge-overlay X-ray: a target acquisition task (Section 4.1) and an online survey to investigate visual information representation capability of these two X-ray techniques (Section 4.2).

Our design goal for these studies was to evaluate the top-down capabilities of our system; for example, the overall appearance and how it supports users in top-down tasks, such as target acquisition. Although the underlying algorithm of our system works in a bottom-up fashion, mimicking the human visual system, we ultimately aim to support users in natural, goal-driven tasks.

4.1 Study 1: Target Acquisition Task

The purpose of this evaluation was to compare the two AR X-ray variants. We have decided to use a target acquisition task in the background of the scene, since typically users want to study the background when employing AR X-ray. Based on the algorithm of our saliency-based X-ray, it is clear that this variant shows less background information. However, when designing the saliency-based AR X-ray system, we aimed at improving the overall appearance of the visualization. We had two hypotheses before running this experiment:

H1: Quantitatively, saliency-based X-ray will perform as well as edge-overlay X-ray.

H2: Qualitatively, participants will prefer saliency-based X-ray over edge-overlay X-ray.

While we did not find enough statistical evidence to confirm or refute our first hypothesis, our second hypothesis was refuted. These results points to problems in our prototype, which we have investigated in more detail in the second study. After presenting the first study in detail, we discuss our findings in Section 4.1.7.

4.1.1 Experimental Platform

The experimental platform was implemented on a 2.4 GHz Intel Core 2 Duo laptop with 2 GB RAM, an NVIDIA GeForce 8600M GT, and an external FireWire camera. The FireWire camera was a PYRO webcam running at a resolution of 640 $\times$ 480 and was attached to a handheld screen. The implementation of our visualization design was carried out using our AR platform TINT[4]. TINT is written in Python with a minimum set of C components. Our visual saliency system is implemented as GLSL shader programs. GLSL allows our saliency system to be run in realtime, with little effect on the performance of the system. While TINT is capable of rendering complex AR scenes, the visualizations presented in our experiments render the occluded area as a static image instead of a three-dimensional scene.

The handheld screen was mounted on a tripod to avoid the adverse effects of registration error. Participants were not allowed to move the screen but just to see through it and perform the experimental task.

4.1.2 Participants

Sixteen voluntary participants (15 male, 1 female) with ages ranging from 22 to 33 years *(*mean=25.75, SD=3.48) were recruited from the student population of the university and randomly distributed into two matched groups. All participants had no known vision impairment like color-blindness or impaired acuity . Three of the participants had experienced our edge-overlay X-ray in a previous experiments that was conducted seven months ago. Participants were provided with some refreshments for their effort and time.

4.1.3 Task and Procedure

There were two different tasks in this experiment: a target acquisition task and a questionnaire.

In the target acquisition task, participants were asked to search and select a red target circle from a scene of the occluded region that was revealed using two different types of AR X-ray techniques with four different surfaces (see Figure 7). Note that the red target circle was not included in the saliency computation for the background in order not to bias the experiment towards saliceny-based X-ray. After viewing each scene, participants had to select the target object using the stylus provided to them. Once a target object was successfully selected the next scene was presented to the participant with the target object randomly placed at a different position on the screen and the same process was followed. One set of trials consisted of ten repetitions for each participant (five times for each type of X-ray). Each participant performed four sets of trials. The target acquisition task took around 20 minutes to complete per participant.

We then showed the experimental scenes again and asked participants to complete a subjective questionnaire where they had to rate them on a seven-level Likert scale. We separated the questionnaire from the target acquisition task to avoid confounding effects between these two tasks. The whole experiment took around 30 minutes to complete per participant and was conducted over a period of three days.

4.1.4 Variables

The experiment was of mixed-factorial design where all of the participants experienced all levels of X-ray visualizations and occluder surface. It contained five independent and three dependent variables. We have randomized all independent variables with the exception of the level of target object size, which was a between-subject variable. We did not cross occluder surface with other variables during this experiment, as moving the camera after every trial could produce calibration errors and would make the procedure lengthy. The first group of eight participants worked on 16 pixel target object first, then we ran other group of participants with 9 pixel target object. Next, we briefly describe the dependent variables, followed by the independent variables.

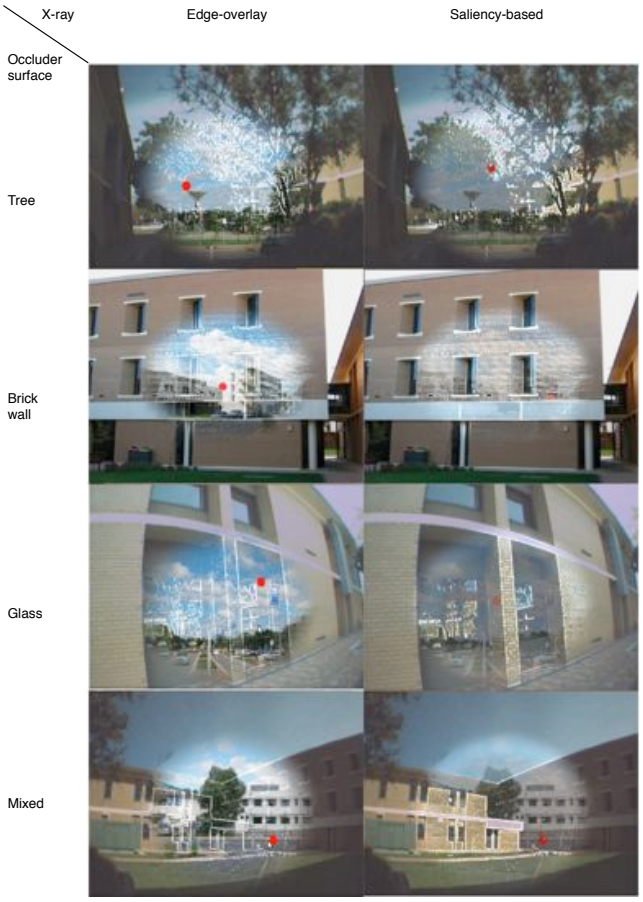


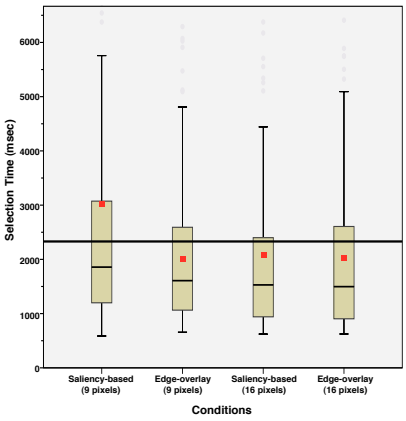
Figure 7: Typical screenshots for Study 1: Target Acquisition Task.

Dependent Variables We measured the time taken in milliseconds to select the target object successfully. We also collected qualitative responses from participants.

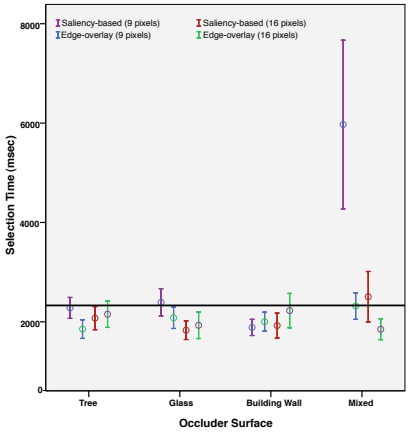
Independent Variables

X-ray $\in$ { Saliency-based, Edge-overlay} *within subjects*
We have evaluated two different X-ray visualizations: saliency-based X-ray and edge-overlay X-ray. X-ray was selected as a within-subject variable in this experiment which means every participant experienced both of the X-ray visualizations.

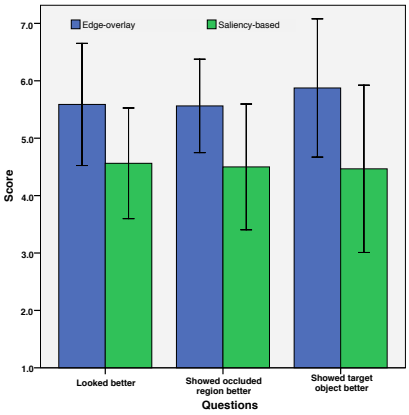
Occluder Surface $\in$ {Brick wall, Tree, Glass, Mixed} *within subjects*
As our new saliency-based X-ray is based on a combination of saliency maps of occluder and occluded regions, it is important to find the effect it has on various surfaces. At the same time the edge-overlay X-ray also produces different levels of edge overlay for different surfaces. Therefore, we have



(a) Mean selection time of the two X-ray variants crossed with the size of the target object with ±1 standard error. The red squares indicate means and the thick black line indicates overall mean.



(b) Mean selection time of two X-ray visualizations crossed with the size of the target object at four different occluder surfaces with ±1 standard error. The thick black line shows the overall mean selection time across all conditions.



(c) Mean score provided by the participants on a seven-level Likert scale with ±1 standard deviation.

Figure 6: Results for Study 1: Target Acquisition Task.

selected four different occluder surfaces. The mixed surface had a combination of tree (grass), building and sky.

Target Object $\in \{16 \text{ pixels}, 9 \text{ pixels}\}$ *between subjects*
After a pilot study, we selected the radius of the target to be 16 pixels and 9 pixels. Target object were implemented in this experiment as a between-subject variable. We used two different levels of target object to investigate the effect of size of target object on the X-ray visualizations. As the physical size of the target object will vary depending on the resolution of the display used, we are reporting the radius in pixels.

Participant Group $\in \{A, B\}$ *between subjects*
Sixteen participants were randomly distributed into two equal groups of eight in this experiment. Group A performed the task with the 16 pixel target and group B with the 9 pixel target. In each group, we randomized the order in which the two X-ray variants were presented.

Trial $\in \{1 \ldots 40\}$ *within subjects*
Each participant performed the selection task 10 times (5 for each level of X-ray) in every set of trials performed from each level of Occluder Surface. Hence, each participant performed four sets of trials and twenty trials for each X-ray type. Overall, there were 10 (trials per set) $\times$ 4 (levels of occluder surface) $\times$ 16 (participants) = 640 trials performed by all the participants in this experiment. Every set of trials was randomized by alternating both levels of X-ray in successive trials. The target object was placed randomly in all trials.

4.1.5 Quantitative Results

Using the statistical software package SPSS, we analyzed the effect of the three main independent variables (X-ray, Target Object, Occluder Surface) on selection time.

Effect of X-ray There was no significant difference when comparing the two overall means of the two X-ray variants using a two-tailed t-test after adjusting the outliers.

Effect of Target Object We ran two sets of t-tests to investigate the main effects of X-ray and Target Object while averaging the results for the different Occluder Surfaces (see Figure 6(a)).

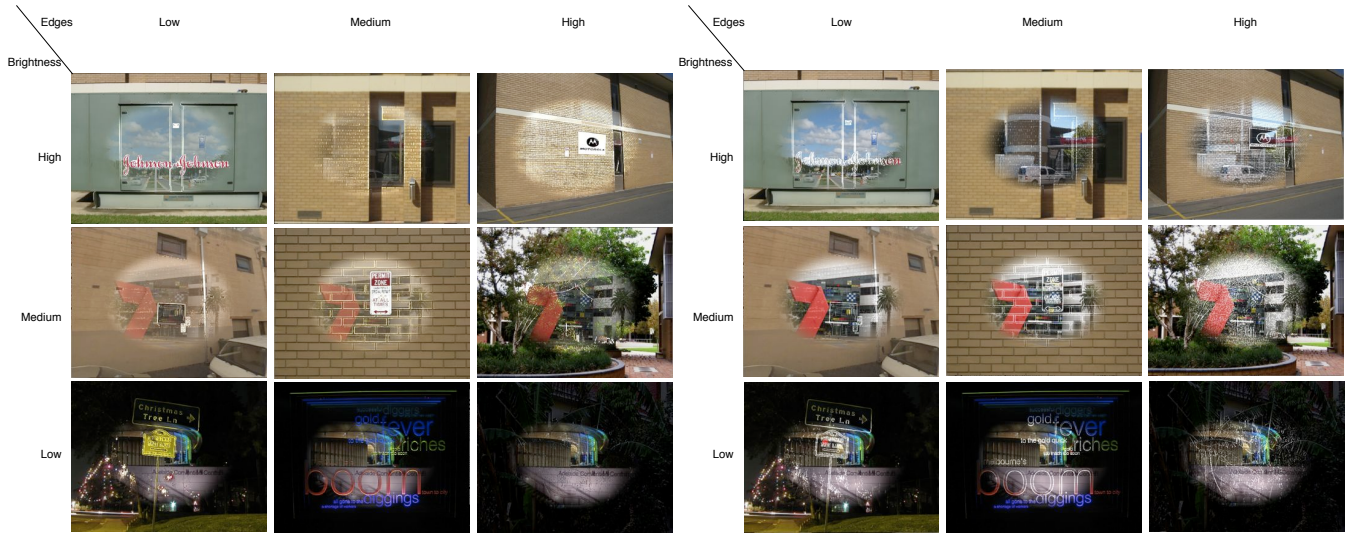
First, we ran two paired t-test to analyze the difference between the two levels of X-ray separately for the two levels of Target Object. We found no significant difference in the case of 16 pixel target objects. However, in the case of 9 pixel target objects, edge-overlay X-ray (M=2063.6) was significantly faster than saliency-based X-ray (M=3132.7) with $t(159) = 2.326, p = 0.021$.

Second, we ran two t-tests to analyze the difference between the two levels of Target Object separately for the two levels of X-ray. Although we did not find any significant difference in the case of edge-overlay X-ray, we did find a significant difference between 16 pixel (M=2084.27) and 9 pixel (M=3132.72) with $t(318) = 2.195, p = 0.029$, for saliency-based X-ray.

Effect of Occluder Surface We ran two factorial ANOVAs for each level of Target Object to investigate the effect of Occluder Surface (see Figure 6(b)); X-ray and Occluder Surface being independent factors. In the case of 16 pixel target objects, we did not find any significant difference between the means of selection time of the four different occluder surfaces. However, in the case of 9 pixel target objects we found a significant difference between the occluder surfaces $F(1,3) = 5.35, p = 0.001$. A Tukey's HSD post-hoc test revealed that Mixed Surface was significantly slower than all other surfaces. There was also a significant interaction between X-ray and Occluder Surface (p=0.011).

4.1.6 Qualitative Results

Our second hypothesis was refuted. In the subjective questionnaire participants reported a consistent preference of edge-overlay X-ray over saliency-based X-ray (see Figure 6(c)). However, there were no significant differences in the scores. In case of the tree occluder surface, most participants preferred saliency-based X-ray as it had a reduced visual noise in comparison with edge-overlay X-ray. P10 commented "I could not see the target object clearly behind the trees with edges on the screen but it was very clear in saliency-based X-ray". Two participants pointed out that two different layers of saliency-based X-ray provided more information about the relationship between foreground and background whereas, edge-overlay X-ray mainly provided better information about the background only.



(a) Saliency-based (b) Edge-overlay

Figure 8: Images used in Study 2: Online Survey.

4.1.7 Discussion

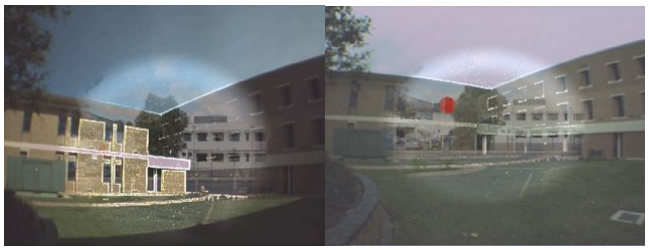
We had two hypotheses when running this experiment. Our first hypothesis was that saliency-based X-ray would perform as well as edge-overlay X-ray. Our second hypothesis was that users would prefer our novel AR X-ray system. The first hypothesis was neither refuted nor confirmed, although edge-overlay performed slightly and non-significantly better. The second hypothesis was refuted: participants did not prefer saliency-based X-ray; instead, they expressed a slight and non-significant preference for edge-overlay. These results point to potential problems in our prototype.

The quantitative analysis showed that smaller targets were significantly more difficult to select with saliency-based X-ray. When investigating this finding more deeply, we have found that the main problem was selecting the small target in the Mixed Surface condition (see Figure 6(b)). In the following, we will give a potential explanation for this problem, which has dictated the design of the second study as presented in the next section.

A confounding variable in this experiment were the weather conditions. Coincidentally, the first day of our experiment was cloudy, but the rest of the days were sunny. This in turn confounded our experiment as we ran six of the eight participants with 16 pixel target object on the first day, whereas, rest of the participants worked on a sunny day. As our saliency-based X-ray operates based on saliency maps of foreground and background, on a sunny day the saliency map of the foreground was assigned more importance than the saliency map of the background. Hence, it was tough to see the occluded region along with the target object on a bright sunny day.

In the questionnaire part of the study, we asked participants about the visibility of the target object. Most of the participants who performed on a sunny day reported problems. P13 reported that viewing the Mixed Surface on a sunny day was "...one problematic situation encountered with saliency-based visualization". Similarly, P16 reported: "...saliency blended the circle too much and it did not stand out like in edge-overlay".

On the cloudy day it was easy to see the occluded region and the target object (see Figure 9). No participant reported any problems in that case. We assume that the effect in Figure 6(b) was not only due to a different surface, but that brightness also played a major role in it.



(a) Sunny weather (b) Cloudy weather

Figure 9: Problem in the saliency-based X-ray on mixed surface condition: weather conditions influence the quality of the composition result significantly. The occluded region and selection target are much more visible in cloudy weather conditions (b), compared to sunny weather conditions (a).

4.2 Study 2: Online Survey

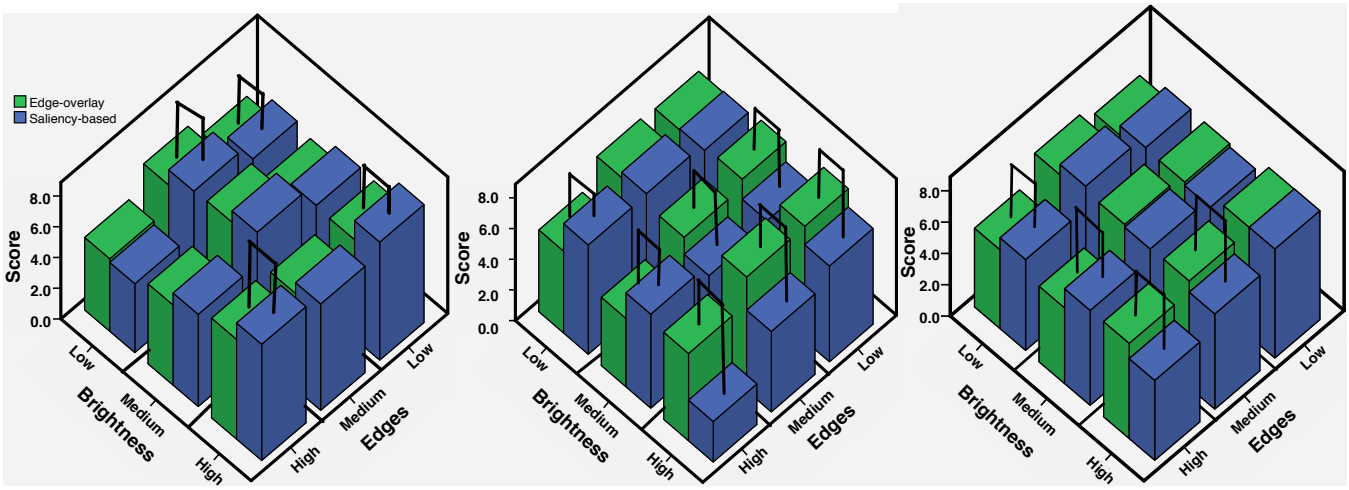
Our second study, an online survey, was inspired by the findings of the first user study, which pointed to problems for saliency-based X-ray on bright surfaces. Similarly, the edge-overlay X-ray is problematic when there are too many edges in the foreground. Therefore, we decided to systematically vary edges and brightness of the foreground and let respondents evaluate the effectiveness of composition. Before running the experiment, we hypothesized that:

H1: Overall, saliency-based X-ray will be rated higher.

H2: High levels of brightness will have a negative effect on the background legibility of saliency-based X-ray

H3: High levels of edges will have a negative effect on the background legibility of edge-overlay X-ray

It is important to note that H1 in Study 2 is different from H1 in Study 1. The reason for this change is based on the different aims of the studies. Study 1 aimed at proving that target acquisition speed for background objects is not impeded by the richer foreground information. Study 2 aimed at confirming our intuition



(a) Foreground (b) Background (c) Total

Figure 10: Results for Study 2: Online Survey. The bars represent mean scores provided by respondents. Connected bars represent significant differences between means.

that the overall visual results are more compelling using our novel method.

Hypotheses 2 and 3 were confirmed in this study. Hypothesis 1 was not confirmed, which points again to problems in our prototype. After presenting the details of this study, we discuss its significance in Section 4.2.5.

4.2.1 Experimental Platform

We created an online survey using a free online survey portal. We invited respondents through e-mails and social networking sites. The images for the survey were generated using a desktop version of the platform used in Study 1.

4.2.2 Respondents

27 voluntary respondents responded to our online invitation (ages: 18–62 years). We discarded responses of two respondents for random answering. Each respondent was allowed to take the survey only once. There was no overlap between the respondents in Study 2 and the participants in Study 1.

4.2.3 Task and Procedure

We instructed respondents to do a *see-perceive-score* task. The images were presented to respondents one at a time. After observing the image carefully, respondents had to score on a scale of 1 (worst) to 10 (best) how well the image conveyed information for foreground and background. There were 18 images of different X-ray conditions presented to each respondent and the survey took 30 minutes on average per respondent. We decided on this design over a design which would just ask respondents to select the better visualization out of a pair of two images (edge-overlay vs. saliency-based) in order to get statistically significant results while saving time, as the pairwise comparison would require a much larger set of sample images.

This within-subject survey was based on three independent variables: Brightness, Edge, and X-ray. We selected three levels for Brightness and Edge: High, Medium, and Low. An expert panel, consisting of eight members of our research group, carefully selected nine foreground/background combinations from a set of 25 random combinations to represent the different levels of brightness and edges. We then executed our X-ray visualizations to create 18 different images (see Figure 8). As a dependent variable, we recorded the scores provided by respondents to these images. The whole survey was based on 18 (images) $\times$ 2 (questions per image) $\times$ 25 (respondents) = 900 data points.

4.2.4 Results

Using the statistical software package SPSS, we have analyzed the collected data. The results were mostly consistent with our hypotheses (see Figure 10 for a summary of results; the Appendix of this paper contains the numerical scores). We ran one-tailed t-tests and found:

Foreground information Saliency-based X-ray provides better foreground information than edge-overlay X-ray in all of the experimental conditions. In the case of high brightness-low edge, high brightness-high edge, low brightness-low edge, and low brightness-medium edge conditions the differences were significant.

Background information In five of the experimental conditions (high brightness-all edges, medium brightness-low and medium edges) edge-overlay X-ray provided significantly better information about the background than saliency-based X-ray. However, in all other conditions saliency-based X-ray provides better background information than edge-overlay X-ray. These differences were significant in the case of medium and low brightness-high edge conditions.

Overall performance Considering both scores for foreground and background information, we found that in the case of high brightness-medium and high edge surfaces, edge-overlay X-ray performed significantly better. For medium and low brightness-high edge conditions saliency-based X-ray was significantly better than edge-overlay X-ray. Across all of the eighteen images saliency-based X-ray (M=156.9, SD=32.6) scored higher than edge-overlay X-ray (M=153.4, SD=23.4). This difference was not statistically significant with the number of respondents in this survey.

4.2.5 Discussion

Hypotheses 2 and 3 were confirmed in this study. High levels of edges cause problems in the edge-overlay X-ray, whereas high levels of brightness cause problems in the saliency-based X-ray. This result is consistent with the way we have designed the saliency-based and edge-overlay X-ray algorithms. Hypothesis 1 was refuted. Although saliency-based X-ray was preferred in 10 out of

the 15 statistically significant differences and the foreground appearance was always scored higher (see Figure 10), this was not overall significant.

This result points once again at the core problem of our prototype: background legibility for high brightness foregrounds. In the following section, we point out how we intend to fix this in the future.

5 Future Work and Conclusions

There is an obvious need for an AR X-ray technique that reduces visual complexity while still conveying depth. In this paper, we have presented an AR X-ray technique that leverages visual saliency to effectively render occluded regions in the user's display. Visual saliency leverages an understanding of the human perception system to determine the most conspicuous or salient features of an image. Salient regions of the occluder are emphasized in our technique, as these provide the most structure and depth information to the user without undue visual complexity. While our previous AR X-ray visualization technique only employed one salient feature (edges) to determine which parts of the occluder to display, our novel visualization employs three additional salient features: hue, luminosity, and motion. We have evaluated two of the three additional salient features in two distinct evaluations.

Overall, we believe that the evaluation results are encouraging. Our main design goal of providing users with richer context of the occluder object (Study 2) while not impeding background visibility significantly (Study 1) has been achieved. However, we could not show a clear superiority of saliency-based X-ray. Both evaluations have pointed at problems in handling bright foregrounds. We have identified three main areas for future work: first, and most importantly, to improve our prototype in order to be able to deal with bright foregrounds. Second, to refine and evaluate motion saliency. Third, to port our system to mobile phones in order to do more realistic evaluations.

The problem of bright foregrounds for saliency-based X-ray is very similar to the problem of edge-overlay X-ray when too many edges are present. On an abstract level, they both point towards the need for an adaptive classification of salient features. The most straightforward way to implement this would be to scale the saliency maps in a way that the overall amount of salient features is below a fixed threshold. While this would certainly improve the overall appearance, it is still not ideal, as the semantics of foreground objects are not evaluated. Therefore, we intend to add an object recognition algorithm to our system, that can detect typical landmarks. Such a system would be able to scale the saliency maps non-uniformly and according to importance.

Motion saliency could be a key element for making AR X-ray applicable in real-world environments, such as a typical pedestrian zone, because they typically include many moving objects. Therefore, we intend to evaluate the motion aspect of our saliency-based X-ray in a pedestrian zone. Our main rationale for including motion into our saliency system was for safety reasons, since moving objects can be harmful to the user when they overlook them. We also consider to treat the motion aspects separately. While edges, luminosity, and hue all give a good indication of typical landmarks, motion does the opposite. Moving objects are typically not landmarks. Therefore, we intend to treat motion saliency separately. One option would be to only overlay the overall outline of moving objects. This would fulfill the safety requirement, but would not obstruct the user's view to real landmarks.

In order to perform an evaluation in a pedestrian zone, we are currently porting our saliency-based X-ray system to the Nokia N900. Since our visualizations are almost completely implemented in GLSL shaders, this port is quite straightforward. However, we have to find ways to address the tracking problem on mobile phones. In that regards, two promising approaches have already been presented: the panorama tracker by Wagner and colleagues[18] and the low-footprint SLAM system that was presented running on an iPhone by Klein and Murray[9]. We are convinced that performing a realistic evaluation on current mobile phones will be a key stepping stone to introducing our visualization technique on the consumer market.

Acknowledgements

The authors wish to thank: the participants of the two evaluations for their voluntary participation and Helen Ftanos and Brad Cameron for helping with the accompanying movie. Thanks to Graeme Jarvis, Thanh Nguyen, and Rhys Moyne for providing last-minute technical support. This research was funded by Nokia Research Center.

References

[1] B. Avery, C. Sandor, and B. Thomas. Improving spatial perception for augmented reality x-ray vision. In *Proceedings of the IEEE Virtual Reality Conference*, pages 79–82. IEEE, 2009.

[2] M. Bajura, H. Fuchs, and R. Ohbuchi. Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. *ACM SIGGRAPH Computer Graphics*, 26(2):210, 1992.

[3] Christopher G. Healey. Perception in visualization. Last accessed on 25 May 2010. `http://www.csc.ncsu.edu/faculty/healey/PP/#Table_1`.

[4] U. Eck and C. Sandor. Tint: Towards a pure python augmented reality framework. In *Proceedings of the Third Workshop on Software Engineering and Architectures for Realtime Interactive Systems*, Waltham, MA, USA, March 2010.

[5] B. Fisher and Z. Pylyshyn. The cognitive architecture of bimodal event perception: A commentary and addendum to Radeau. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, 13(1):92–96, 1994.

[6] C. Furmanski, R. Azuma, and M. Daily. Augmented-reality visualizations guided by cognition: Perceptual heuristics for combining visible and obscured information. In *Proceedings of the 1st International Symposium on Mixed and Augmented Reality*. IEEE Computer Society Washington, DC, USA, 2002.

[7] D. Kalkofen, E. Mendez, and D. Schmalstieg. Comprehensible visualization for augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 15(2):193–204, 2009.

[8] Y. Kim and A. Varshney. Saliency-guided enhancement for volume visualization. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):925–932, 2006.

[9] G. Klein and D. W. Murray. Parallel tracking and mapping on a camera phone. In *Proceedings of the 8th IEEE International Symposium on Mixed and Augmented Reality*, pages 83–86, 2009.

[10] C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Human neurobiology*, 4(4):219, 1985.

[11] J. Kruger, J. Schneider, and R. Westermann. Clearview: An interactive context preserving hotspot visualization technique. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):941–948, 2006.

[12] C. H. Lee, Y. Kim, and A. Varshney. Saliency-guided lighting. *IEICE Transactions*, 92-D(2):369–373, 2009.

[13] C. H. Lee, A. Varshney, and D. W. Jacobs. Mesh saliency. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, pages 659–666, New York, NY, USA, 2005. ACM.

[14] 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*, pages 247–248. ACM, 2009.

[15] V. Setlur, T. Lechner, M. Nienhaus, and B. Gooch. Retargeting images and video for preserving information saliency. *IEEE Computer Graphics and Applications*, 27(5):80–88, 2007.

[16] A. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136, 1980.

[17] University of Southern California. ilab neuromorphic vision c++ toolkit. Last accessed on 25 May 2010. `http://ilab.usc.edu/toolkit/`.

[18] D. Wagner, A. Mulloni, T. Langlotz, and D. Schmalstieg. Real-time panoramic mapping and tracking on mobile phones. In *Proceedings of the IEEE Virtual Reality Conference*. IEEE, 2010.

[19] D. Walther. *Interactions of visual attention and object recognition : computational modeling, algorithms, and psychophysics*. PhD thesis, California Institute of Technology, February 2006.

[20] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.

[21] S. Zollmann, D. Kalkofen, E. Mendez, and G. Reitmayr. Image-based ghostings for single layer occlusions in augmented reality. In *Proceedings of the 9th IEEE International Symposium on Mixed and Augmented Reality*, 2010.

APPENDIX

Results for Study 2: Online Survey. The following tables list mean scores of respondents. Statistically significant difference at $p < 0.05$ are highlighted in **bold** when saliency-based X-ray (SB) is superior and in *italics* where edge-overlay (EO) X-ray is superior.

Table 1: Foreground

Amount of Edges						Brightness
Low		Medium		High		
EO	SB	EO	SB	EO	SB	
6.44	**7.68**	6.20	6.92	**6.48**	**7.56**	High
5.68	6.56	7.32	8.08	5.20	6.00	Medium
4.92	**6.08**	**6.24**	**7.24**	4.72	4.48	Low

Table 2: Background

Amount of Edges						Brightness
Low		Medium		High		
EO	SB	EO	SB	EO	SB	
7.44	*6.24*	*7.4*	*5.24*	*5.68*	*2.64*	High
7.00	*5.68*	*6.48*	*5.36*	**4.44**	**6.12**	Medium
6.68	6.76	6.84	7.20	**5.32**	**7.12**	Low

Table 3: Total

Amount of Edges						Brightness
Low		Medium		High		
EO	SB	EO	SB	EO	SB	
7.00	7.00	*6.80*	*6.10*	*6.10*	*5.10*	High
6.40	6.10	6.90	6.70	**4.80**	**6.10**	Medium
5.80	6.40	6.50	7.20	**5.00**	**5.80**	Low