Image-based X-ray visualization techniques for spatial understanding in Outdoor Augmented Reality

Stefanie Zollmann[1,2], Raphael Grasset[3], Gerhard Reitmayr[1], and Tobias Langlotz[4]
[1]Institute for Computer Graphics and Vision
Graz University of Technology, Graz, Austria
[2]Animation Research Ltd., Dunedin, New Zealand
[3]Qualcomm Austria Research Center, Vienna, Austria
[4]Department of Information Science, University of Otago, Dunedin, New Zealand
stefanie@arl.co.nz, reitmayr@icg.tugraz.at, tobias.langlotz@otago.ac.nz

ABSTRACT

This paper evaluates different state-of-the-art approaches for implementing an X-ray view in Augmented Reality (AR). Our focus is on approaches supporting a better scene understanding and in particular a better sense of depth order between physical objects and digital objects. One of the main goals of this work is to provide effective X-ray visualization techniques that work in unprepared outdoor environments. In order to achieve this goal, we focus on methods that automatically extract depth cues from video images. The extracted depth cues are combined in ghosting maps that are used to assign each video image pixel a transparency value to control the overlay in the AR view. Within our study, we analyze three different types of ghosting maps, 1) alpha-blending which uses a uniform alpha value within the ghosting map, 2) edge-based ghosting which is based on edge extraction and 3) image-based ghosting which incorporates perceptual grouping, saliency information, edges and texture details. Our study results demonstrate that the latter technique helps the user to understand the subsurface location of virtual objects better than using alpha-blending or the edge-based ghosting.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*Artificial, augmented, and virtual realities*

Keywords

Augmented Reality, X-ray, Visualization, Evaluation

1. INTRODUCTION

X-ray visualization is a visualization technique traditionally often used in medical visualization that reveals otherwise hidden information to the user simulating the output of X-ray imaging. X-ray visualization is nowadays also commonly used in Augmented Reality (AR). Normally, AR provides an interface that allows an integration of digital data into our physical environment using visual overlays. Combined with X-ray visualization techniques we can blend the visible with the normally hidden information in an AR view of our environment and thereby create new types of applications. These applications range from applications in the surveying domain where the X-ray visualization in AR are used to reveal hidden subsurface infrastructure such as underground pipes [20], or medical applications visualizing minimally invasive procedures in AR giving the surgeon the feeling to have the ability to see through the skin or body [2].

For achieving a convincing Augmented Reality experience we need a seamless integration of digital data into the physical world. This is a big challenge for X-ray visualizations in AR. In particular the problem of missing depth cues (e.g., relative size, occlusion, shadows) affect the quality of visual integration (Figure 1). Humans are naturally not used to having the visual ability of an X-ray view and to looking inside objects. This raises several questions such as: How do we communicate occlusions and the order of objects in X-ray AR visualizations?

Figure 1: Naïve AR overlay of digital assets for civil engineering. Augmented digital subsurface information seems to float over the ground due to missing depth cues making it hard to get a reasonable scene understanding.

OZCHI '14, Dec 2-5, 2014, Sydney, Australia

Figure 2: AR views using different approaches of extracting depth cues from a video image. (Left) Random occlusion cues randomly preserve image information but can not transport the depth order. (Middle) Only edges are preserved to provide depth cues. (Right) Using important image regions based on a saliency computation as depth cues creates the impression of subsurface objects.

1.1 Motivation

The main goal of this paper is to compare different visualization techniques aiming for a better spatial understanding of depth order for occluded objects in X-ray AR visualizations. While some techniques require full scene knowledge (e.g., geometry and texture information of the scene), we focus on techniques that create depth cues solely from information in the camera image (e.g., edges). We thereby address the issue of achieving a seamless integration of digital content into the physical world by detecting and maintaining physical pictorial cues from the camera image. If natural pictorial cues are not sufficiently presented in an AR visualization, the scene will either look unnatural or produce a wrong perception of the order of objects in the scene. Missing occlusion cues in an X-ray view may also lead for instance to perceiving virtual subsurface objects as to be floating over the ground (Figure 1). Besides order of objects in the X-ray visualization, we are also interested in communicating the shape of the objects visualized in the X-ray view.

Our research is driven by industrial demands and focused to develop AR visualization techniques that support professionals working within the civil engineering industries. Our work should later help workers in the field to plan excavations without damaging underlying infrastructure. The civil engineering industry requests a better general understanding of the subsurface infrastructure without using paper maps and before starting with the actual excavation. One of our main interest in this context is to answer the question of how to support a comprehensible visualization of subsurface infrastructure. In the course of our research, we found that there is no existing research investigating if AR visualization techniques help users to understand the depth ordering of occluded objects in a better way.

There are a few works in X-ray AR investigating the effect of visualization techniques on the performance and depth understanding of users. For instance, Livingston et al. investigated the effect of different X-ray visualization techniques on the ability of users to map a digital object into a set of depth zones [16, 15] with the main focus on the depth order of digital objects and multiple layers of occlusions. The works of Sandor et al. studied the effect of different X-ray visualization techniques, such as saliency-based and edges-based ghostings, and view distortion technique on accuracy of absolute depth estimation and task completion time [19, 4]. In general, absolute depth estimation in X-ray AR is a highly complex tasks, since humans are not used to this kind of vision and some natural depth cues are not working, such as occlusion and height in visual field [3]. Thus, some of these experiments showed no significant effect of using X-ray visualisation techniques for absolute depth estimation [4]. Consequently in our work, we reduce the complexity of depth estimation to ordinal estimations. For many applications, it is already beneficial if the user understands the depth order of objects, e.g. if an object is in front or behind another object.

There are other works that compare X-ray visualisation techniques to non-AR interfaces, such as the work of Dey et al. [5], where the authors analyzed the effectiveness of saliency-based X-ray AR visualization techniques for navigation in comparison to map interfaces. The results showed that navigation hints displayed with an X-ray AR system have a positive effect on reducing context switches compared to map interfaces [5].

Some works focus on the perceptual effects of "visible" AR visualization techniques, but not on X-ray AR visualization, such as supporting the depth understanding using additional depth cues [14], shadow planes or color markings [21]. Since the focus of our research is the visualization of occluded objects, we will not explain these techniques in more detail.

To our knowledge none of the existing works investigates if X-ray visualization techniques helps users to understand the order between physical and digital content in scenes with occluded objects in a single-layer setup [22].

1.2 Contribution

The main focus of this paper is the important aspect of the effect of image-based X-ray AR visualization techniques on the understanding of depth order while also maintaining shape information about occluded objects. The contribution of this paper is the evaluation of different X-ray visualization techniques in AR with respect to identifying order and communicating shape information.

2. BACKGROUND

Related works can be roughly categorized into two categories. Firstly, works that address depth perception in X-ray AR and secondly, works that specifically address occlusions in X-ray AR.

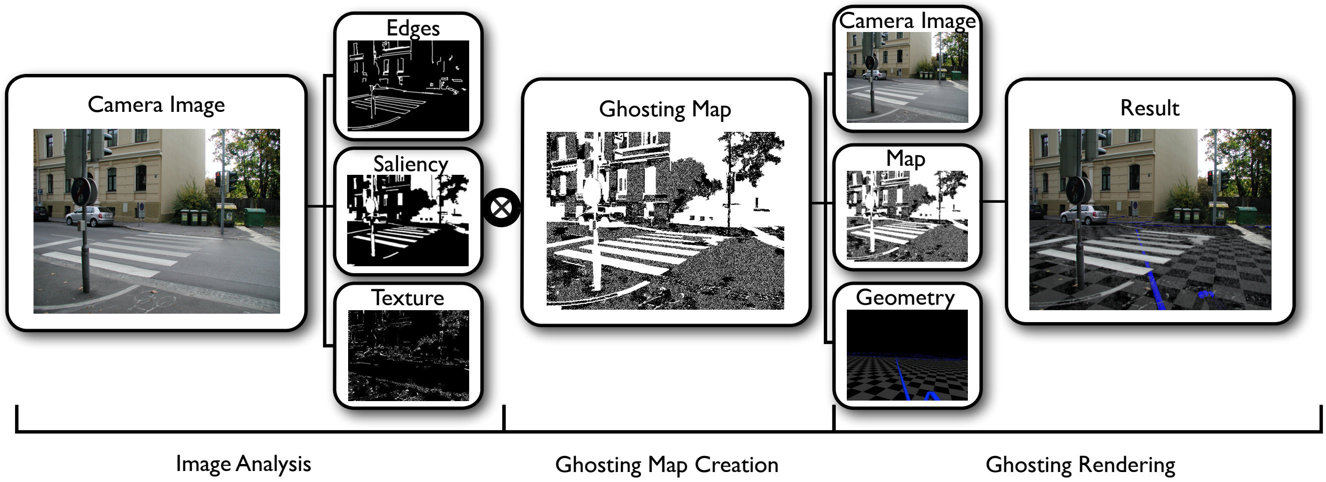

Figure 3: Image-based ghostings. (Left) Input camera image of an urban scene. (Middle) Extracted important image regions, such as edges, saliency and texture details are integrated into a ghosting map. (Right) Ghosting preserves important image parts in the final rendering and provides a convincing integration of the virtual pipes into the street scene.

2.1 Depth Perception in X-ray AR

Cutting presented a set of different cues that are important for for depth perception in general [3]. They comprise occlusion, relative size, height in visual field, convergence, aerial perspective and relative density. Since humans use these cues for building a mental model of a scene, these cues are also important to maintain for a seamless integration of digital and physical content in AR. If some depth cues are not available, spatial information is not coherent and the user will note that something is wrong in the scene. Some of these depth cues are already provided by the general AR rendering pipeline, such as the relative size: a virtual perspective camera will render more distanced objects smaller. But some cues, such as the occlusion between virtual and physical elements have to be generated additionally.

Occlusion cues give an ordinal measurement of objects, and are assumed to be the strongest of all pictorial depth cues, and to work at all distances [3]. Unfortunately, they are not automatically provided by a naïve AR overlay.

There are two main problems that have to be addressed to provide an adequate occlusion management in AR. Firstly, if there is no accurate 3D representation of the scene available (which is often the case), the depth order of virtual and physical objects has to be estimated from the camera image. Secondly, in X-ray AR it has to be decided which information of a physical occluding object is preserved in the final rendering. This decision depends on the required visibility of the occluded objects, the required visibility of occluding objects, but also on the minimum number of occlusion cues that has to be preserved to achieve a seamless scene integration. A convincing occlusion management finds the best compromise between sufficient number of occlusion cues, preserving the occluder's structure, and object visibility. Thereby, it is important that the cues preserve not only the appearance of the occluding object but also its structure. In Figure 2, we show the problem that appears when only random image elements are used as occlusion cues to preserve the appearance of the street, but not the structure. In AR exist several occlusion management techniques that address this challenge in different ways. While cutaway techniques provide additional depth cues they often replace a large amount of the physical scene with virtual cutout geometries [23]. Ghosting techniques address the occlusion problem differently as they try to find a compromise between visibility of occluder and occluded object [6]. The basic idea of ghosting techniques in AR is to preserve selected information from the physical world by rendering it on top of the digital object. This raises the questions what information should be preserved and in what amount (opacity). Former work describes model-based approaches that use a 3D representation of the occluder to determine important scene structures [11]. However, model-based ghostings techniques have a big disadvantage since they rely on the availability of a 3D model of the physical world. In practical situations, this kind of 3D model does often not exist or the 3D model does not precisely align with the physical world as the model is not accurate enough or the tracking of an AR system is too imprecise. This is in particular a problem in outdoor environments where we usually do not have a precise representation of the environment at hand or where a precisely scanning of the scene information is a complex and expensive task. Even if a perfectly registered 3D model of the occluder exists, the exact texture of the model might be missing. This makes the computation of the adequate amount of preservation difficult. For the case of exact registered data Mendez et al. proposed a method that is based on using pre-defined masks [17]. Other researchers addressed the problem of having no precise 3D models of the physical environment available, by using information extracted from camera images, such as edges [12], salient regions [19] or a combination of salient regions, edges and texture details [22].

In this paper, we compare techniques that extracts occlusion cues solely from video images. These methods are based on the assumption that the depth order between virtual objects and the physical world is known. This assumption is valid for scenes where all digital objects would be normally occluded as they are located under or behind the physical objects seen by the camera (Figure 1). We refer to this assumption as *single layer occlusions*, which applies, for example, for underground infrastructure visualizations.

Image-based ghosting techniques decide which information of the physical environment should be preserved by analyzing camera images and heuristically extracting key information. For this purpose, these techniques analyze edges, salient locations and texture details from the camera stream. These features are then used as input for the ghostings.

Figure 1 and Figure 2 (Right) contrast a naïve augmentation with an image-based ghosting in an outdoor AR application. The first image presents the problem of augmenting virtual data without considering the underlying camera image. The second image shows that the ghosting approach helps to infer the spatial positions of objects.

2.2 Addressing Occlusion in X-ray AR

Image-based techniques achieve visual coherence by extracting physical cues from video images. They are the first choice for creating physical cues in situations where the depth order of virtual and physical world is known (e.g. through a semantic meaning as it is the case when visualizing subsurface infrastructure) and no accurate and precisely registered 3D model of the occluding physical world object is available. Important elements from the camera image are extracted (Figure 3, Left) and mapped to a ghosting map (Figure 3, Middle). The ghosting map is then used to combine the camera image and the digital geometries to create the final AR visualization (Figure 3, Right).

Such an image-based approach has been introduced by Kalkofen et al. [12]. In their work, they propose to extract edges from a camera image and use them to create edge-based ghostings. The edges are rendered on top of the virtual content (Figure 2, Middle). Bichlmeier et al. extended this approach by using a combination of edges and bright pixels as physical depth cues [2]. Another approach that uses edges as input to create physical cues is the method of Avery et al. [1]. They apply edges to improve their X-ray vision system in outdoor environments. Based on this work, Sandor et al. later on proposed to us saliency information as depth cues. Their approach computes saliency masks from the camera image and the layer of virtual content to decided which information should be preserved in the final rendering [19]. At the same time, Zollmann et al. introduced a method that incorporates visual saliency, edges and texture details to compute a ghosting map [22]. Furthermore, this method preserves perceptual grouping by using a super-pixels representation of the camera image.

3. X-RAY VISUALIZATION TECHNIQUES

The simplest form of implementing X-ray visualization in AR is based on alpha-blending. We will refer to this technique as G_A in the following. Such an naïve ghosting approach preserves both digital content and video content in equal measure by using half transparent objects. However, this approach disregards the fact that each image region may require a different amount of preservation due to properties and importance of each region.

In order to address this problem, sophisticated ghosting techniques focus on the question what has to be preserved in each image region and in which amount. The idea is to analyze the physical scene to be augmented and calculate a transfer function that maps the video image into a *ghosting map* [22]. The ghosting map indicates the importance of each pixel in the scene and whether it should be preserved in the final rendering or not. The ghosting map calculation can be based on 2D image information such as edges or salient regions, but also on a 3D model [13]. In the following, we will focus on image-based techniques as they are required in outdoor environments. For outdoor environments there are usually no highly textured 3D models available that can be used as input for model-based techniques. But even if 3D models exist (e.g., from Google Earth) it is highly unlikely that they display up-to-date information or that they are accurately registered.

Image-based methods extract all information for creating a ghosting map from the camera stream. For instance, an edge-based ghosting technique maps each pixel that is part of an edge detected in the current camera image to a fully or nearly fully opaque rendering and pixels that are not part of an edge to a transparent value. In the following, we will refer to this technique as edge-based ghostings G_E.

Edge-based ghosting methods work often well in scenes with a decent amount of dominant edges. However, these methods work on a per-pixel basis and in scenes that contain a large amount of regions with similar visual characteristics, such per-pixel based method stand in contrast to the perceptual theories such as the Gestalt laws. The Gestalt laws of grouping state that humans do not only use single entities to process their perception but use the complete structure [8]. The Gestalt law of similarity states that entities with similar characteristics are more likely to be perceived as part of one group. If the applied visualization technique only manipulates selected entities such as pixels, this grouping may be destroyed reducing the comprehensibility.

In order to support perceptual grouping, we analyze a third method that incorporates both per-pixel image features and features computed from larger regions [22]. These regions are computed as superpixels to preserve perceptual grouping [18]. Based on this over-segmentation, the method computes visual characteristics for each region and uses them as input for creating a super pixel based ghosting map. Visual characteristics for this method include edge information, saliency, and texture details for each region and and are combined to obtain the amount of preservation for each super pixel. Additionally, if a region is found to be less important and lacks important structures, this ghosting method uses synthetic region-dependent structures to preserve a sketch-like representation of the region. According to Zollmann et al., we will refer to this method as *image-based ghostings* G_I in the following description [22].

4. USER EVALUATION

In order to understand how existing ghosting techniques perform, we investigate the effect of different X-ray techniques on depth perception in a user study. Within the study we compared an image-based ghosting method (G_I) with alpha-blending (G_A) and an edge-based ghosting technique using edges for preserving image features (G_E) [1]. The goal is to investigate if the image-based ghostings perform better than alpha-blending and edges in terms of depth perception. Furthermore, we analyze if the user is still able to understand the shapes of the hidden virtual objects in the scene, even if these objects are partially occluded by the extracted image information.

It is important to note that in this work, we on focus on

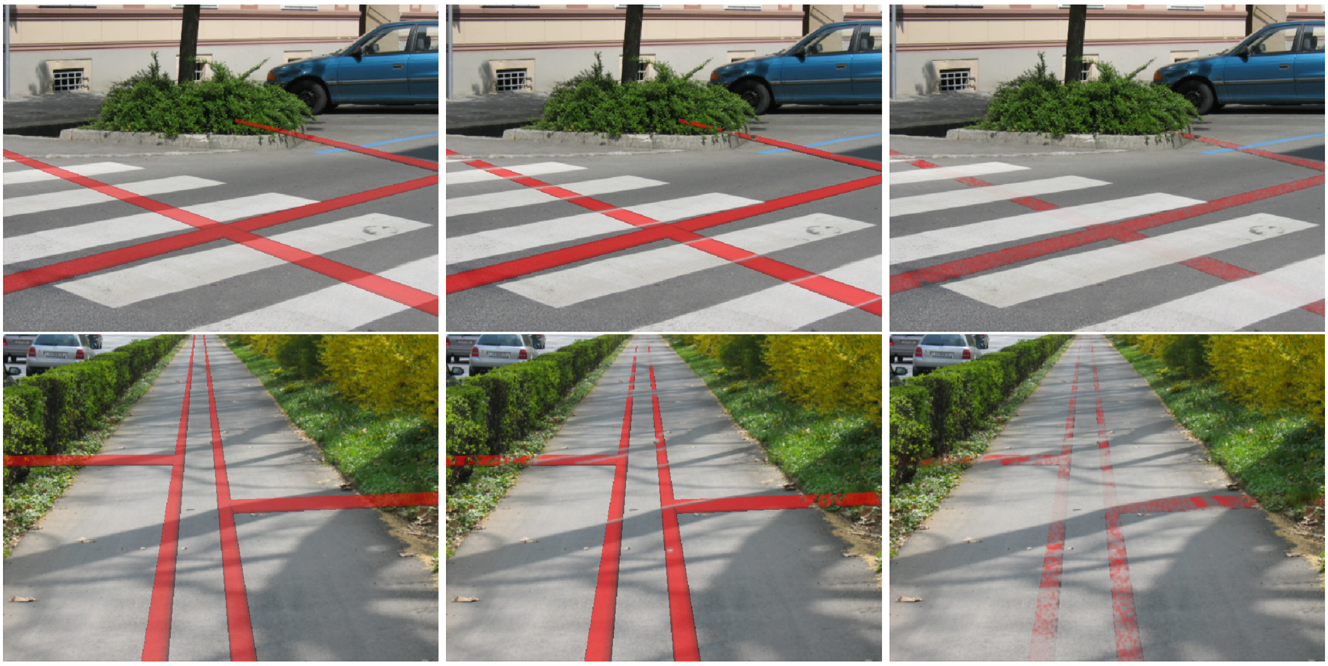

Figure 4: Different test scenes for the evaluation. (Left) Condition Alpha-Blending G_A, (Middle) Condition Edge-based Ghostings G_E, (Right) Condition Image-based Ghostings G_I. The test scenes using G_A and G_E demonstrate the effect of missing depth cues, pipes seem to float over the ground. In contrast G_I provides additional depth cues and provides a information about the depth order between virtual pipes and street.

effect of the discussed three visualization techniques on the spatial understanding. There are several aspects within each of the visualization techniques that could be varied as additional variables, such as the color of the digital objects, the complexity of objects or different levels transparency. For instance, the adaptation of color of digital assets could contribute to readability as shown by previous works on active rendering styles in AR [7, 9]. However, we decide to not use color adaption as an additional variable since there are several application fields where such an adaption is not desired. For instance, for the visualisation of underground infrastructure color patterns are usually fixed since they transport semantic information for the user.

Another parameter that we did not include into our study design are multiple transparency levels. There is some research in X-ray AR that suggest to use different transparency levels for encoding depth information of virtual objects (e.g. [16]). However, we decided against using transparency as additional variable, since other research has shown that higher levels of transparency could have an impact on readability [10]. In order to avoid an impact on the shape understanding, we used a fixed transparency level for all visualization techniques to provide a fair comparison between the techniques.

4.1 Hypotheses

For our user evaluation on X-ray techniques, we state the following hypotheses. We hypothesize that participants understand the subsurface location of virtual objects better using image-based ghostings (G_I) than using alpha-blending (G_A) or the edge-based ghosting technique (G_E). Furthermore, we hypothesize that the used visualization technique does not affect shape perception.

- H1: Image-based ghostings will outperform edge-based ghostings and simple alpha-blending in terms of a convincing depth perception. Using image-based ghostings, the user will have a stronger perception that objects are located subsurface.

- H2: The choice of visualization technique has no influence on the correctness of perceived shapes. The human visual system of the users will complete shapes automatically.

4.2 Experimental Platform

The comparability between the test scenes and the possibility to perform the study on a set of different test scenes with different characteristics had a high priority during the design of our study. Furthermore, we wanted to preclude external influences such as an unstable tracking or the influence of sensor noise from our study. In order to achieve these goals, we decided to prepare a set of static AR scenes in advance using the three different visualization techniques instead of letting participants use our outdoor X-ray AR system [20]. Another advantage of this design decision is that we are able to focus on occlusion cues. Depth cues resulting from motion are excluded when using the static scenes.

All scenes contain urban scenarios that are common when inspecting subsurface infrastructures, the application scenario we aim for. We differentiate between street scenes that contain a lot of important information such as cross walks and scenes containing less important information such as plain streets or grass. In addition, we use two different types of content. Content that belongs to a scenario of inspecting subsurface infrastructure (pipes) and abstract content (red spheres with different sizes). For all tested scenes we ap-

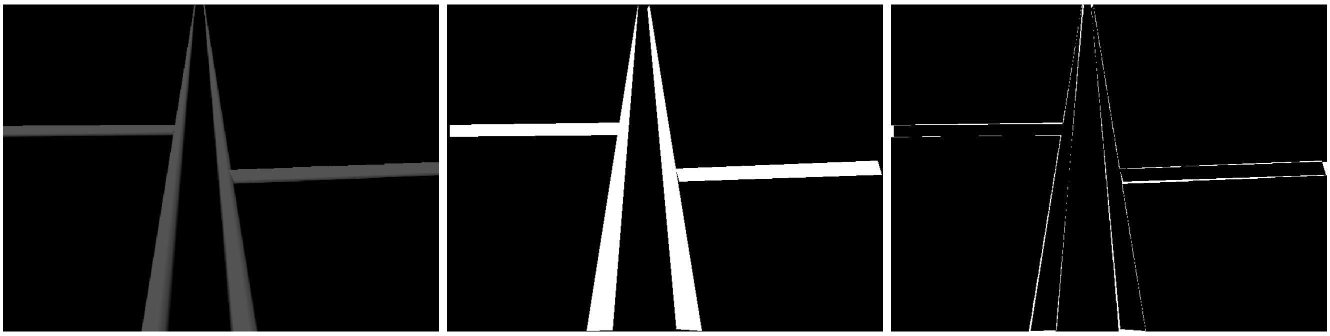

Figure 5: Contour difference analysis. (Left) Ground truth mask. (Middle) Mask from user input. (Right) Difference between ground truth mask and user input.

ply Phong shading to provide additional depth information about the virtual objects.

The settings for the visualization techniques are fixed for all scenes. For the alpha-blending the composition is computed based on the virtual content (V) and physical scene (P) as

$$C_A = \alpha V + (1 - \alpha)P. \tag{1}$$

We set the value for α to a fixed value of 0.5 for the study. As discussed before, an additional option would be to use different values for alpha as another variable in the study. However, in order to make the alpha-blending technique better comparable to the other ghosting techniques we decided to use one fixed value for α.

For computing the composition for G_E we extract edge information and map it to a ghosting map ($\alpha(x, y)$). The composition for G_E as well as for G_I is then given by

$$C_G = \alpha(x, y)V + (1 - \alpha(x, y))P. \tag{2}$$

4.3 Task and Procedure

We divided the study in two tasks. At first, a participant had to inspect each scene and to provide a rating about the depth perception. Thereby, the ordinal depth perception was based on a Likert scale ranging from 1 = strongly underground, 2 = underground, 3 = rather underground, 4 = undecided, 5 = rather overground, 6 = overground and 7 = strongly overground. We told the participants that the scenes may contain subsurface as well as overground objects. Nevertheless, all scenes contain subsurface objects. We decided to do so in order to give the user no previous knowledge about the scene configuration and to have complete freedom when choosing the spatial location of the virtual objects.

After completing this task, the participant was asked to draw an outline of the virtual objects for scenes that contained virtual pipes. For this task, the user interface for the study provided an input functionality for drawing polygonal objects on the AR scene by mouse clicks. The output of the drawing process for each scene is the outline of the virtual pipes perceived by the participants. We compared the filled outlines ((Figure 5, Middle)) with a binary ground truth mask of the virtual objects (Figure 5, Left). The difference between both masks resulted in a contour difference measurement ((Figure 5, Right), which was used to determine the ability of users to correctly understand the shape of the object. To compute the contour difference D, we used the amount of pixels that differ from the ground truth mask n_D and divided it by the amount of pixels n_{GT} from the ground truth mask

$$D = \frac{n_D}{n_{GT}}. \tag{3}$$

This task was repeated for 12 different scenes using the same visualization technique, but showing different content. After finishing these scenes, participants were asked about their experience with the applied visualization technique using a questionnaire. Afterwards, the technique was changed and the new technique was used for the same scenes as before. The order of the visualization techniques was randomized using Latin Squares. In a final questionnaire, we asked the participants to give a rating on their preferences according depth perception, coherence and general comprehension. The overall study duration for each participant was approximately thirty-five minutes.

4.4 Pilot Study

Before we started with the main study, we conducted a pilot with five users to find out if our experimental design is sound and to understand if the test is too exhausting for the participants. From the user feedback during the pilot study, we learned that we should remove the abstract shape condition for the contour drawing since the participants reported that these shapes were too simple and to easy to complete and on the other hand quite exhausting to draw due to the sphere shape. Overall, the pilot study showed that the participants seem to perceive the subsurface objects more being located underground when using the image-based ghosting G_I (compare with Figure 6 Left, average rating 2.43). Contrarily, for G_E they seem to be rather undecided (average 4.33) and for G_A they seem to more likely to rate the location being overground (average 5.39). The pilot study also showed that there seems to be only a small difference in shape understanding as the contour difference for all three conditions are similar (Figure 6, Right). These findings encouraged us to proceed with the given study design with the described minimal adjustments.

4.5 Participants

We invited 15 people with various backgrounds to take part in the final experiment (5 female, 10 male, age ranging from 22 to 35). The experience with AR of the participants ranged from not familiar at all to very familiar. We used a repeated measure design for the study. Each participant

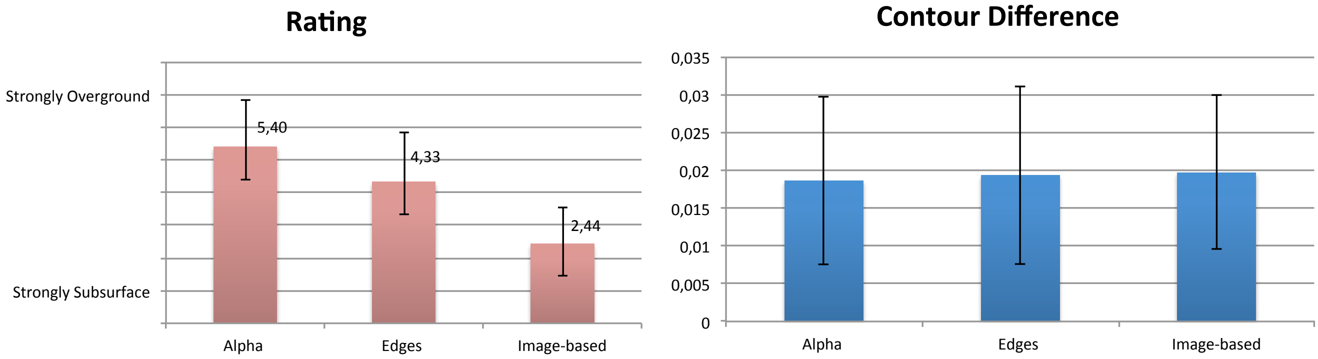

Figure 6: Results of the pilot study. (Left) Ratings of the subsurface location of the virtual content. (Right) Contour difference measurement in %.

performed all three visualization techniques G_A,G_E and G_I for all 12 scenes in randomized order.

5. RESULTS

For each participant we averaged the depth perception rating and the contour difference for each technique, resulting in an overall depth perception rating and an overall contour difference. We performed a repeated measure ANOVA on this data in order to analyze the effect of technique on overall depth perception rating and overall contour difference.

5.1 Quantitative Results

In the following we report on the achieved quantitative results.

Depth Perception.

The output of the ANOVA for overall depth perception rating shows that the F-statistics is 71.685 with a p-value $\approx$ 0.0. We can clearly reject the null hypothesis of equal means for the overall depth perception rating of all three visualization techniques. This shows that there is an significant effect of technique on depth perception.

In order to find the significant differences between the single techniques, we used a post-hoc test. The pairwise T-Test (P value adjustment method: bonferroni) showed that there are significant differences between all three methods. G_I showed a significantly better perception ($M = 2.95$, compare with Figure 7, Left) of the subsurface location of the virtual objects than the simple blending G_A ($M = 5.37$, $G_A - G_I$: $p \approx 0.0$) and state-of-the-art ghosting G_E (M=3.79, $G_I - G_E$: p = 0.004). G_E also performs better than G_A($G_A - G_E$: p $\approx$ 0.0). This confirms hypothesis H1 that the image-based ghostings are outperforming edges and alpha-blending in terms of transferring the subsurface location of objects. Users have a stronger perception that objects are located subsurface.

Object understanding.

The output of the ANOVA for the overall contour difference shows that F= 1.204 and has a p-value p=0.315. We cannot reject the null hypothesis of equal means for the accuracy of outlines during usage of the three visualization techniques. Consequently, there is no difference between the techniques, which confirms hypothesis H2 that the visualization technique has no influence on the shape perception and users can find the outline with the same accuracy.

5.2 Qualitative Results

We were able to confirm the findings on depth perception from the quantitative test with results from the questionnaires of the study. After working with each technique the participants were asked to rate this technique according to the questions "A: The subsurface visualizations using the X-ray technique was confusing", B: "The subsurface location of virtual objects in the scene was hard to understand" and C: "The shape of the virtual objects was complicated to understand.". For question A and B, participants rated the image-based ghosting technique G_I better than the other techniques (Figure 8). G_I was rated with with M=2.3 for question A and M=2.3 for question B. The rating reflects an average rating between "disagree" and "rather disagree" for the image-based ghosting technique. In contrast, alpha-blending G_A was rated with M=4.2 for question A and M=4.9 for question B, which reflects a value between undecided and rather agree. For the method G_E, the participants seemed to be rather undecided, since they rated the technique with M=3.2 for question A and M=3.4 for question B, a value between rather disagree and undecided. We found an significant effect between technique and both questions (ANOVA for question A: $F = 7.188, p = 0.003$ and B: $F = 10.334, p = 0.0004$). The pairwise T-Test shows only significance between image-based ghosting technique G_I and the naïve overlay G_A (question A: $p = 0.004$ and question B: $p \approx 0.0$). This means on the one hand, that G_E performs not significantly better then the naïve overlay. On the other hand, it shows that the image-based ghosting technique does perform significantly better.

The third question shows that even if the participants showed similar performance on understanding the shape in the quantitative part of the study, they rated the comprehension of the shape of the virtual objects slightly more complicated than with the other techniques (question C: $G_A = 1.4, G_E = 1.6, G_I = 2.5$). This measured difference between image-based ghostings and the alpha-blending techniques was significant (ANOVA F=10.334, p=0.0004, T-Test $G_I - G_A$: p=0.034). Nevertheless, the rating still indicates that they disagreed or rather disagreed that the shape was complicated to understand. Together with the quantitative measurements, it seems that it was more complicated compared to the other techniques, but still possible.

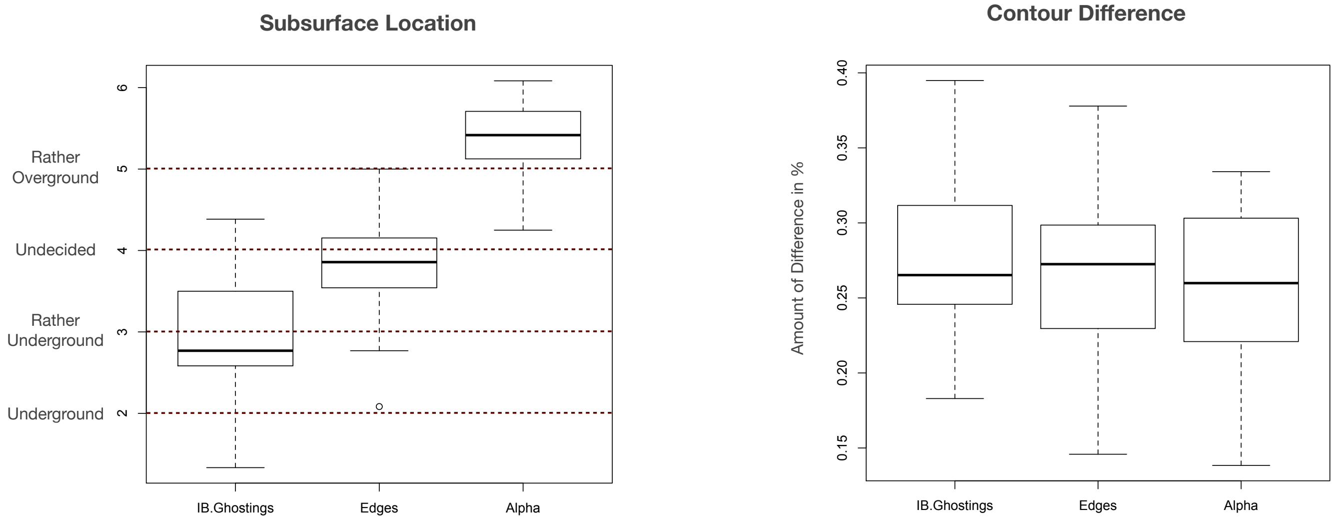

Figure 7: User study results: Evaluating ghosting techniques. (Left) Results of the depth perception task. (Right) Results of the accuracy test.

In a final questionnaire, we asked the participants to rate their preferences in terms of depth perception, coherence and general comprehension. They were asked to give their ratings according to their preferences starting with one for the favorite. As shown in Figure 9, the users preferred image-based ghostings for all the questions over the other techniques (G_I: depth perception M=1.2, coherence M=1.4, general M=1.4, G_A: depth perception M=2.7, coherence M=2.7, general M=2.6, G_E: depth perception M=2.1, coherence M=1.9, general M=2.0). The ANOVA showed significance for the effect of technique on the rating for (depth perception: F=20.24 $p \approx 0.0$, coherence: F= 11.82 p=0.0002 and general comprehension F=7.875, p=0.002).

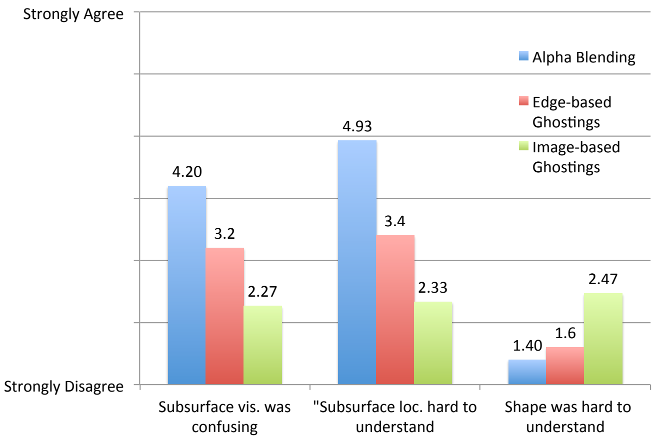

Figure 8: Results user study: User ratings for each technique for the following questions. "A: The subsurface visualizations using the X-ray technique was confusing", B: "The subsurface location of virtual objects in the scene was hard to understand" and C: "The shape of the virtual objects was complicated to understand."

For the depth perception the pairwise T-Test shows a significance between the ratings for all three technique ($G_I - G_A : p \approx 0.0, G_I - G_E : p = 0.0002$, and $G_A - G_E : p = 0.0047$). This means that the participants clearly prefer image-ghostings for depth perception. The pairwise T-Test for the question asking for the coherence of the presented X-ray techniques shows significant effects between both ghosting techniques G_E and G_I against the naïve overlay G_A ($G_I - G_A$: $p \approx 0.0$ and $G_E - G_A$: p= 0.0013). Although G_I was rated with a higher preference and shows better performance during the depth estimation tests, the user ratings have no significant difference in the perceived coherence between G_E and G_I ($p = 0.1388$). Finally, for the rating on general comprehension the pairwise T-Test (P value adjustment method: bonferroni) indicates that there is only a significant difference between G_I and alpha-blending G_A (G_I-G_A:$p \approx 0.0$).

5.3 Discussion

Overall, the results confirm our initial hypothesis that image-based ghostings outperform edge-based ghostings and alpha-blending in terms of conveying the subsurface location of digital below-surface objects, as well as our second initial hypothesis that the visualisation technique has no negative influence on the correctness of shape understanding.

Furthermore, the qualitative results show that there is a significant effect of visualization techniques on the user ratings. It seems that the participants preferred the image-based ghostings over the alpha-blending. The edge-based technique was in general not rated as being significantly better than the naïve overlay. This is interesting, since it confirms our assumption that for these urban scenes with an AR overlay showing subsurface object, the edge information provides not enough depth cues to improve the comprehension in comparison to a naïve overlay as given by the alpha-blending.

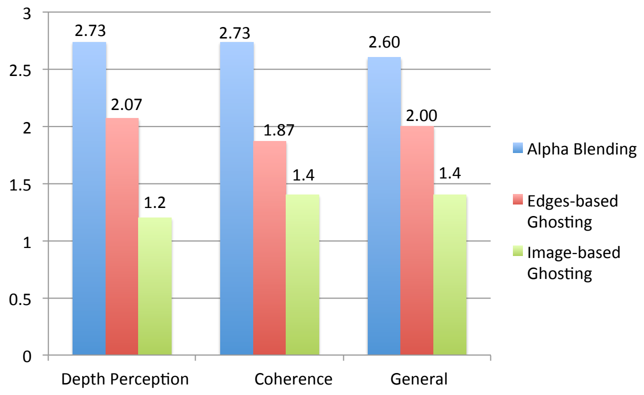

Figure 9: Results user study: User preferences on depth perception, coherence and general comprehension.

6. CONCLUSION AND FUTURE WORKS

In this paper, we presented an evaluation of a set of state-of-the art X-ray AR visualization techniques targeting the visualization of subsurface infrastructure in outdoor environments. X-ray AR visualization in outdoor applications are often highly challenging. In contrast to indoor AR application that often come with a smaller restricted workspace, in outdoor application there is often no 3D representation of the physical environment available. This makes it highly challenging to preserve sufficient depth cues.

In order to address this challenge, image-based ghosting methods were introduced that extract information that is important to scene understanding from the video stream and preserve them in the final AR composition. The goal of this paper was to investigate the effect of a set of state-of-the-art X-Ray visualization methods on the understanding of depth order between digital objects and the physical world. Our study revealed that image-based ghostings have a positive effect on the user's understanding of the subsurface location of virtual objects compared to alpha-blending and edge-based ghostings.

In our study we used a set of scenes with a large variety in urban scenes, such as street with markings, park areas, but also plain streets. Although we did not analyze the effect of the visual characteristics of the background to the spatial understanding more in depth, user feedback let us assume that the background has an influence on the reliability of techniques. We leave this question open for future research. Findings about such correlations could also be beneficial for the development of new X-ray visualization techniques or for the improvement of existing ones.

In general, existing image-based ghosting methods only extract and use a small subset of information that could improve the depth perception in X-ray AR visualization. While current approaches rely on the analysis of bottom-up features of images, more high-level concepts such as object recognition could improve image-based ghosting techniques.

7. ACKNOWLEDGMENTS

The authors would like to thank all participants of the pilot study as well as the final user study. We further would like to thank our industry partners.

8. REFERENCES

[1] B. Avery, C. Sandor, and B. H. Thomas. Improving Spatial Perception for Augmented Reality X-Ray Vision. In *IEEE Virtual Reality Conference (VR 2009)*, pages 79–82. IEEE, Mar. 2009.

[2] C. Bichlmeier, M. Kipot, S. Holdstock, S. M. Heining, E. Euler, and N. Navab. A Practical Approach for Intraoperative Contextual In-Situ Visualization. In *AMIARCS '09*, 2009.

[3] J. E. Cutting. How the eye measures reality and virtual reality. *Behavior Research Methods, Instruments, & Computers*, 29(1):27–36, 1997.

[4] A. Dey, G. Jarvis, C. Sandor, and G. Reitmayr. Tablet versus phone: Depth perception in handheld augmented reality. In *EEE International Symposium on Mixed and Augmented Reality (ISMAR 2012)*, pages 187–196, Nov 2012.

[5] A. Dey, G. Jarvis, C. Sandor, A. Wibowo, and V.-V. Mattila. An evaluation of augmented reality x-ray vision for outdoor navigation. In *In Proceedings of International Conference on Artificial Reality and Telexistence*, pages 28–32, Osaka, Japan, November 2011.

[6] S. K. Feiner and D. Duncan Seligmann. Cutaways And Ghosting: Satisfying Visibility Constraints In Dynamic 3d Illustrations. *The Visual Computer*, 8:292–302, 1992.

[7] J. Gabbard, J. Swan, D. Hix, S.-J. Kim, and G. Fitch. Active text drawing styles for outdoor augmented reality: A user-based study and design implications. In *IEEE Virtual Reality Conference (VR 2007)*, pages 35–42. IEEE, March 2007.

[8] E. B. Goldstein. *Sensation and Perception*. Wadsworth Publishing Company, 6 edition, 2001.

[9] R. Grasset, T. Langlotz, D. Kalkofen, M. Tatzgern, and D. Schmalstieg. Image-Driven View Management for Augmented Reality Browsers. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2012)*, 2012.

[10] B. L. Harrison, G. Kurtenbach, and K. J. Vicente. An experimental evaluation of transparent user interface tools and information content. In *Proceedings of the 8th annual ACM symposium on User interface and software technology - UIST '95*, pages 81–90, 1995.

[11] D. Kalkofen, E. Mendez, and D. Schmalstieg. Interactive Focus and Context Visualization for Augmented Reality. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2007)*, pages 191–200, Nov. 2007.

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

[13] D. Kalkofen, E. Veas, S. Zollmann, M. Steinberger, and D. Schmalstieg. Adaptive Ghosted Views for Augmented Reality. *IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2013)*, 2013.

[14] M. Livingston and Z. Ai. Indoor vs. outdoor depth perception for mobile augmented reality. In *IEEE Virtual Reality Conference (VR 2009)*, pages 55–62, 2009.

[15] M. Livingston, Z. Ai, K. Karsch, and G. O. Gibson. User interface design for military AR applications. In *IEEE Virtual Reality Conference (VR 2011)*, pages 175–184, Dec. 2011.

[16] 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. *Symposium on Mixed and Augmented Reality*, 2003.

[17] E. Mendez and D. Schmalstieg. Importance masks for revealing occluded objects in augmented reality. In *Proceedings of the ACM symposium on Virtual reality software and technology - VRST '09*, pages 247—-248, New York, New York, USA, 2009. ACM Press.

[18] X. Ren and J. Malik. Learning a classification model for segmentation. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 10–17, 2003.

[19] C. Sandor, A. Cunningham, A. Dey, and V. Mattila. An Augmented Reality X-Ray system based on visual saliency. In *Proceedings of the 9th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 27–36, Nov. 2010.

[20] G. Schall, S. Zollmann, and G. Reitmayr. Smart Vidente: advances in mobile augmented reality for interactive visualization of underground infrastructure. *Personal and Ubiquitous Computing*, pages 1–17, 2012.

[21] J. Wither and T. Hollerer. Pictorial depth cues for outdoor augmented reality. In *Proceedings of the Ninth IEEE International Symposium on Wearable Computers (ISWC '05)*, 2005.

[22] S. Zollmann, D. Kalkofen, E. Mendez, and G. Reitmayr. Image-based ghostings for single layer occlusions in augmented reality. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2010)*, pages 19–26. IEEE, Oct. 2010.

[23] S. Zollmann, G. Schall, S. Junghanns, and G. Reitmayr. Comprehensible and Interactive Visualizations of GIS Data in Augmented Reality. *Advances in Visual Computing*, pages 675–685, 2012.