On-Line Visualization of Underground Structures using Context Features

Jiazhou Chen, Xavier Granier, Naiyang Lin, Qunsheng Peng

HAL Id: inria-00524818

https://hal.inria.fr/inria-00524818v2

Submitted on 25 Jan 2011

On-Line Visualization of Underground Structures using Context Features

Jiazhou Chen [1,2]　　　Xavier Granier [2,1]　　　Naiyang Lin [1]　　　Qunsheng Peng [1]

[1] *State Key Lab of CAD&CG - Zhejiang University*　　　　[2] *IPARLA - INRIA Bordeaux University*
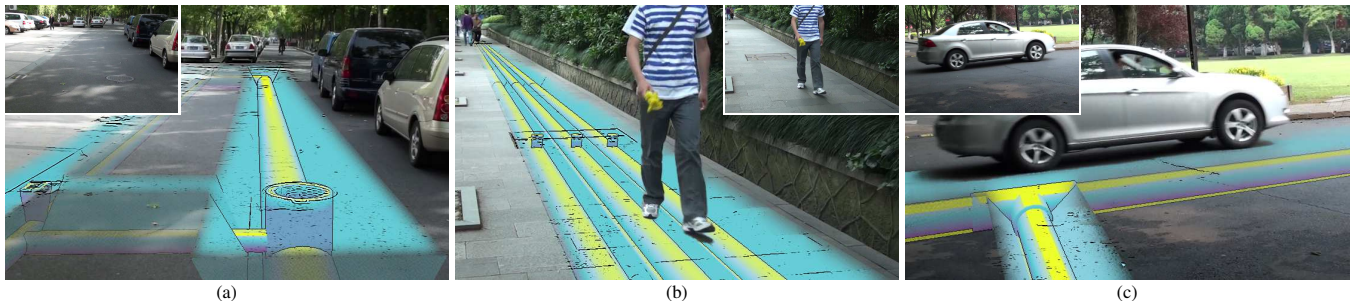
Figure 1: On-line Visualization of underground structures on an image (a) and videos with moving objects (b and c). The depth order is suggested by displaying features from images such as edges over the virtual objects, by introducing image-depend smooth transition for transparency, and by preserving the moving object in front. The framerate ranging from 12 to 18 fps depends mostly on the size of filters for edge extraction and transparency smoothing. The original frame is shown in the upper-left or upper-right corner.

Abstract

We introduce an on-line framework for the visualizing of underground structures that improves X-Ray vision and Focus and Context Rendering for Augmented Reality. Our approach does not require an accurate reconstruction of the 3D environment and runs on-line on modern hardwares. For these purposes, we extract characteristic features from video frames and create visual cues to reveal occlusion relationships. To enhance the perception of occluding order, the extracted features are either directly rendered, or used to create hybrid blending masks: we thus ensure that the resulting cues are clearly noticeable.

CR Categories: I.3.7 [Three-Dimensional Graphics and Realism]: Virtual reality—Augmented Reality;

Keywords: Augmented Reality, X-Ray Vision, Context+Focus Rendering, On-Line Video Processing

1 Introduction

Augmented Reality (AR) is classically described as the process to seamlessly integrate virtual objects into real images or videos. With up-to-date techniques, reaching photorealism still rely on a choice between visual accuracy or on-line rendering. Non-photorealism is more suited when visual efficiency is required for conveying information [Fischer et al. 2008]. Moreover, in this paper, we are interested in on-line visualization of invisible objects that are hidden by real occluders issued from an image or a video (e.g., underground pipes, organs in a body [Kutter et al. 2008], engine in a car [Kalkofen et al. 2009]): they do not share the same lighting conditions as their occluders. Such an augmentation, close to classical illustration, should focus on conveying information such as depth, characteristic structures and features, and spatial relationships.

Making "invisible" objects "visible" requires to (i) ensure that some characteristic features are legible in the final image in order to be easily identified the objects, and (ii) provide some visual cues to create a visually correct spatial relationships. This problem is identified as the Focus and Context (F+C) rendering [Kalkofen et al. 2009]. In theory, (i) is trivially achieved by rendering the virtual object on top of the real image. Unfortunately, such an approach does not provide any cue about its relative positioning in the 3D scene, not fulfilling (ii). Simple transparency provides occlusion cues [Livingston et al. 2003] but needs other cues to not confuse the user [Furmanski et al. 2002] and to understand the relative order. One successful metaphor is X-Ray Vision [Avery et al. 2009]. More generally, the transparency is efficiently replaced by compositing different layers with different styles and characteristic features.

In this paper, we present three main contributions to improve such an approach for on-line video-based AR, and illustrate them on the visualization of underground infrastructures. The first one is a depth-ordered frame segmentation for scenes with moving objects (Section 3). The second and core contribution is our process to create the masks (Section 4) that take into account information from both the virtual object and the videos, allowing the preservation of video features as cues for the occluding order. Finally, we show that well-tuned contrast enhancement participates to the legibility of such cues. We demonstrate our on-line framework and discuss its limitations and potential improvements in Section 5.

2 Previous work

One of the main problems for the visualization of hidden objects is to convey the difference between visible and hidden components [Furmanski et al. 2002]. Some techniques similar to cut-away use a simple binary mask between the image and the virtual object (e.g., [Schall et al. 2009]), but this may result in a loss of context. Some researchers have improved these techniques to complex AR task. Mendez et al. [2006] have introduced an interaction tool that extends global information filtering [Julier et al. 2002] using Context Sensitive Magic Lenses, which works with highly complex AR scenes such as geo-data models. Such an approach efficiently allows the visualization of depth cues on 3D components, but failed in preserving the context issued from the videos.

A simple solution is to use transparency (e.g., [Kameda et al. 2004]). Despite being efficient to convey occlusion events [Livingston et al. 2003], it can be confusing when the number of layers increase [Furmanski et al. 2002]. Moreover, it cannot convey
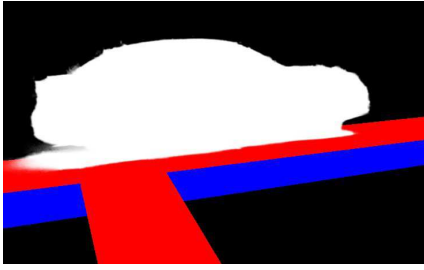
Figure 2: Depth-ordered segmentation for Figure 1-(c): The blue and red regions are respectively to the second focus (cut geometry in front of the virtual object) and main focus (virtual object not hidden by the cut geometry). Region of the video outside the white region (the moving objects) is the "static" scene.

alone the layers' order. Indeed, Furmanski et al. [2002] have shown that it is only one of five depth-dependent perceptual cues. Some works in 3D and volumetric rendering share the same goals: for suggesting depth-order, some of them use transparency modulation by the curvature of the occluding surface [Krüger et al. 2006] or by the view angle and distance falloff [Bichlmeier et al. 2007]; Viola et al. [2005] have introduced the screen-door transparency. Unfortunately, these solutions are tied to a knowledge of the 3D environment and thus not directly usable in our on-line context.

The most successful techniques is the X-Ray Vision [Bane and Hollerer 2004]. Their *Tunnel Tool* renders the invisible components which span in a frustum-cut (the tunnel) and the context (geometry of the occluders) is rendered using wires. But the lack of accuracy in the camera tracking introduce incoherent motions between real and virtual objects and thus lower down the perception of depth order. Kalkofen et al. [2009] have proposed an interactive context preserving Focus and Context (F+C) visualizations. Similarly, their approach requires a 3D reconstruction of the occluders and an accurate camera tracking. To overcome this problem, they propose to use simple edges from videos: but they are not as clearly defined as ones from 3D objects and they can be either too cluttered or not noticeable due to lack of contrast with hidden objects. However, recent results in X-Ray Vision for AR [Avery et al. 2009] show how edge overlay to render the occluders and tunnel cut-out can efficiently convey the distance between the user and hidden objects.

3 Overview and Frame Organization

Our system targets the following scenario: the camera is fixed and the geometry of the real environment is mostly unknown. This last assumption is very important for out-door scenes that are constantly evolving where on-line reconstruction of 3D environment is still not a mature technology. To display hidden objects, we identify four depth-ordered regions (cf. Figure 2) as detailed below.

Focus. The focus is the virtual object, located behind real scenes. We also add the geometry corresponding to the cut-away. This is easily done since, contrary to the real scene, the object's geometry is known. They are rendered using X-toon [Barla et al. 2006] combined with lines extracted from the depth-map discontinuities. We divide the focus region in two. The first region (red in Figure 2) corresponds to the potentially visible parts of the object if a real hole was created to reveal it: we call it the **main focus**. The second region (blue in Figure 2) corresponds to parts of the cut that occlude the object: we call it the **second focus**. In order to suggest that this layer is in-between the main focus and the scene, we use a lower transparency in this region.

Context. Similarly to the focus region, we subdivide the context region (the video frame) in two. The first (white in Figure 2) con-

tains the moving objects that are segmented out: the **moving context**. As proof of concept for future on-line processing, we use an off-line version of [Zhong et al. 2009]. Since the moving objects are mostly in front of the scene and the virtual object, we render this layer "as-is" on the top of all the others: the occlusion events resulting from the movements reinforce depth-order. The frame region that does not contain the moving objects is the **static context**. The virtual object is cut in it. We use this region to extract characteristic features (such as edges, gradients, second order derivatives, ...) in order to re-introduce them as depth-order cues in front of the 3D object.

4 Feature-based Masks and Layers

Three layers from the videos provide visual information about the context. The main layer I_{static} contains the "static" part of the video: it represents the real scene and the main occluders of the virtual object and is rendered "as-is" in regions not covered by the virtual object and the cut geometry. The second layer I_{moving} contains the moving objects of the video and is rendered directly in front of all the other ones. The third layer I_{edge} corresponds to image edges (GPU implementation of [Kang et al. 2007]) as abstracted visualization of the static context. From the virtual object and its corresponding cut geometry, two layers and one mask are issued: the shaded object $O_{shading}$ and its 3D edges O_{edge} without the silhouette of the cut geometry since they visually conflict with the smooth transition that we introduce later (Figure 5).

We compute one mask from the virtual object: α_{edge} identifies the main focus area ($\alpha_{edge} = 1$) and is used to blend the video edges. The second mask α_{color} and core contribution is issued both from the video and the virtual object. It also identifies the main focus ($\alpha_{color} \neq 0$) but is used to blend the video with the virtual object.

Color Mask. Inspired by [Mendez and Schmalstieg 2009], we propose a solution for video-based AR that does not require any explicit 3D reconstruction of the real environment. As shown in Figure 3, the α_{color} is created as follows. We extract two binary masks α_1 and α_2 that identify the main and second focus areas. We smooth inside α_1 (to create a smooth transition from 0 to 1 where $\alpha_1 = 1$ as shown in the second row of Figure 4) to obtain $\overline{\alpha}_1$. This results in a progressive removal of the context image on top of the virtual object. To distinguish between the main and second focus (that is in front of the main focus), α_2 is simply multiply by a user-selected constant value $\gamma \in [0,1]$ (we generally use 0.5). Finally, in order to reinforce the order cue with the real scene and inspired by [Krüger et al. 2006], we modulate the transparency using the curvature-like information from video (Figure 4, third row):

$$\alpha_{color} = (\overline{\alpha}_1 + \alpha_2\,\gamma)^{1+\beta\,(|L_{xx}|+|L_{yy}|)}.$$

$\beta > 0$ is a user parameter, L_{xx} and L_{yy} are the second-order derivatives of the luminance L of the static context I_{static}. This approach preserves image areas with strong features (high second derivatives) on top of the virtual object.

To preserve also uniform areas, we use cross bilateral filtering (e.g., [Eisemann and Durand 2004]) with I_{static} instead of the Gaussian filter to compute $\overline{\alpha}_1$ (Figure 4, last row).

Edge Mask. We use the *Edge Overlay* of Avery et al. [2009], requiring the use a smoothed mask. But contrary to this work and in order to distinguish the effect from the color transition, we use a Gaussian filtering outside of α_1 (where $\alpha_1 = 0$, Figure 6).

Final Compositing. Finally, a pixel color is I_{moving} in areas containing moving objects, and is resulting from the use of the different masks and layers elsewhere:

$$(\alpha_{edge}I_{edge} + 1 - \alpha_{edge})\left[\alpha_{color}\,O_{shading}\,O_{edge} + (1-\alpha_{color})\,I_{static}\right].$$
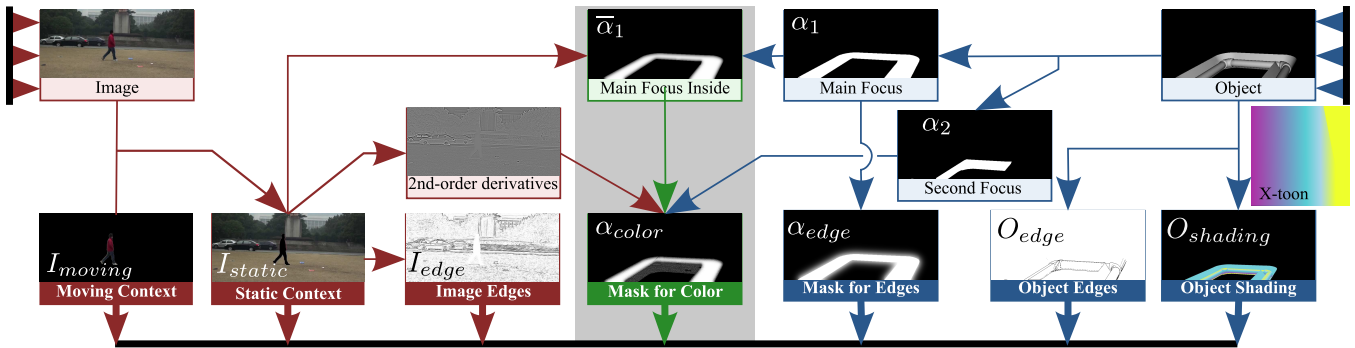
Figure 3: Extraction of the different layers and masks: From the video (red paths), we extract three layers $I_{dynamic}$, I_{static} and I_{edge}. From the virtual object (blue paths), we extract one mask α_{edge} and two layers O_{edge} and $O_{shading}$. The last mask α_{color} is a hybrid one (green paths) that combines both video and object information to create the final cues.
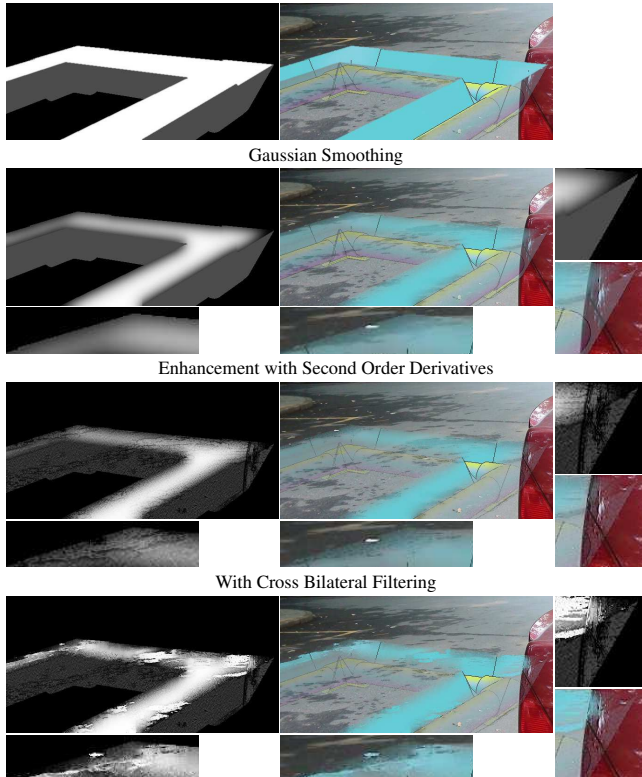

Figure 4: Mask for color blending. From upper row to bottom one: uniform transparency, smoothing the edge inside, with modulation by image second order derivative, with using cross bilateral filtering. Note that modulation by curvature tends to preserve discontinuities and that cross bilateral filtering tends to preserve uniform regions (like the upper-left dark mark on the road).
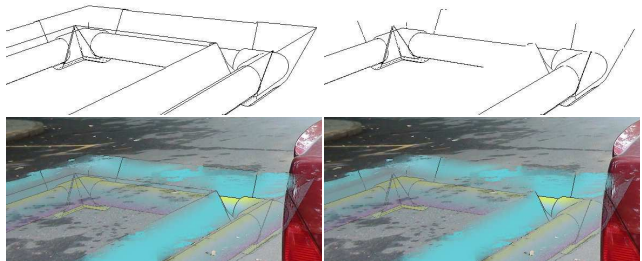

Figure 5: Computation of 3D edges from the virtual object: the silhouettes of the cut may introduce unwanted discontinuity in the smooth transition created by the masks and connect un-wanted areas (such as the back of the car and the ground).
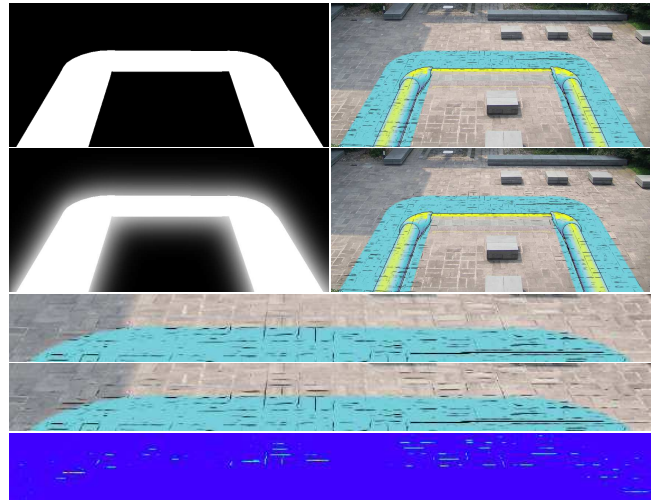

Figure 6: Edge Blending. This figure shows both the mask without and with smoothing outside, and their results on the final compositing. The three last rows show a zoom on the upper part with a Lab difference.
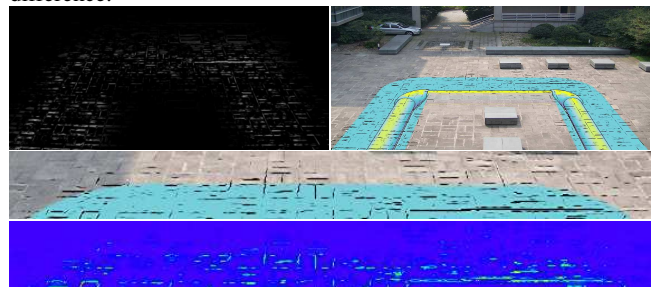

Figure 7: Local Unsharp Masking. Unsharp masking is applied around the edges I_{edges} (in white areas in the left image) to enhance their legibility. The three last rows show a zoom on the upper part with a Lab difference.

Contrast Enhancement. Alpha-blending does not ensure that the image edges are clearly visible in the resulting image. To improve their legibility and emphasize the depth-order cues, we use the unsharp masking operator [Badamchizadeh and Aghagolzadeh 2004] locally around the edges (few pixels around, Figure 7).

5 Results

All the results are computed on a PC workstation with a NVIDIA GeForce 8800 GT, an Intel(R) Core(TM)2 Duo E6550 @ 2.33 GHz and 4 GB of memory. Due to its pixel processing nature, the fram-

erate depends on the frame resolution and the size of the different filters. The most costly step is the cross bilateral smoothing (between 22 and 63 ms per frame with filter size ranging from 90 to 105 pixels in Figure 1). Furthermore, the robust flow-based technique [Kang et al. 2007] to compute image edges is also quite costly (between 13 and 23 ms) but may be replaced by a simpler operator.

We have experimented our approach on scenes with easily-extractable edges (Figures 1-(b) and 6) and some others with less strong patterns (Figures 1-(a), 1-(c) and 4). The use of image edges help in understanding the positioning where the hidden object is. In Figures 1-(a) and (b), the edges show the real place of the pipe exits. This is also reinforced in video by introducing the moving objects on the top (Figures 1-(b) and (c)). Thanks to the color mask and the use of both bilateral filtering and second order derivative modulation, we preserve the existing discontinuities in the original image (there can be some occlusion events) and uniform areas without the need of any 3D reconstruction (Figure 4).

Discussion and Limitations

Depsite that we assume a fixed camera, our framework can be used for moving cameras, and since we do not require any 3D reconstruction of the real environment, the classical misalignment between reality and virtuality is less important than in systems that rely on such reconstructions. This un-needed 3D reconstruction leads to an easy to implement on-line solution since there is no need in preprocessing to get the real 3D scenes. Unfortunately, this is also one of the limitations: better coherency and more accurate geometry of the cut might be achieved with a better knowledge of the 3D environment.

One of the main limitations of our approach is the lack of accuracy in feature extraction and video segmentation. An improved segmentation will help users to select some static objects and introduce them as new ordered layers to improve the depth order visualization.

6 Conclusion

In this paper, we have introduced a new framework for visualization of underground structure that improves X-Ray Vision. By extracting information such as edges, second order derivatives from images, and by creating some masks using both the virtual object and video frames, we manage to create visual cues that suggest the depth order while revealing the virtual object hidden behind the scene. Thanks to this approach, we achieve near real-time performances and good quality without the requirement of 3D reconstruction. We believe that such a framework may be easily adapted to reveal other hidden objects and will take benefit from future work in on-line video processing.

Acknowledgements

References

AVERY, B., SANDOR, C., AND THOMAS, B. H. 2009. Improving Spatial Perception for Augmented Reality X-Ray Vision. In *IEEE Conf. Virtual Reality*, 79–82.

BADAMCHIZADEH, M. A., AND AGHAGOLZADEH, A. 2004. Comparative Study of Unsharp Masking Methods for Image Enhancement. In *Int. Conf. on Image and Graph.*, IEEE, 27–30.

BANE, R., AND HOLLERER, T. 2004. Interactive Tools for Virtual X-Ray Vision in Mobile Augmented Reality. In *Int. Symp. Mixed and Augmented Reality*, IEEE, 231–239.

BARLA, P., THOLLOT, J., AND MARKOSIAN, L. 2006. X-toon: an extended toon shader. In *symp. Non-Photorealistic Animation and Rendering*, ACM, 127–132.

BICHLMEIER, C., WIMMER, F., HEINING, S. M., AND NAVAB, N. 2007. Contextual anatomic mimesis hybrid in-situ visualization method for improving multi-sensory depth perception in medical augmented reality. In *Int. Symp. Mixed and Augmented Reality*, IEEE, 1–10.

EISEMANN, E., AND DURAND, F. 2004. Flash photography enhancement via intrinsic relighting. *ACM Trans. Graph. 23*, 3, 673–678.

FISCHER, J., HALLER, M., AND THOMAS, B. H. 2008. Stylized Depiction in Mixed Reality. *The Int. Journal of Virtual Reality 7*, 4, 71–79.

FURMANSKI, C., AZUMA, R., AND DAILY, M. 2002. Augmented-Reality Visualizations Guided by Cognition: Perceptual Heuristics for Combining Visible and Obscured Information. In *Int. Symp. Mixed and Augmented Reality*, IEEE, 215.

JULIER, S., BAILLOT, Y., BROWN, D., AND LANZAGORTA, M. 2002. Information Filtering for Mobile Augmented Reality. *IEEE Comput. Graph. Appl. 22*, 5, 12–15.

KALKOFEN, D., MENDEZ, E., AND SCHMALSTIEG, D. 2009. Comprehensible visualization for augmented reality. *IEEE Trans. Visualization and Comput. Graph. 15*, 2, 193–204.

KAMEDA, Y., TAKEMASA, T., AND OHTA, Y. 2004. Outdoor See-Through Vision Utilizing Surveillance Cameras. In *Int. Symp. Mixed and Augmented Reality*, IEEE, 151 – 160.

KANG, H., LEE, S., AND CHUI, C. K. 2007. Coherent Line Drawing. In *ACM Symp. Non-Photorealistic Animation and Rendering*, 43–50.

KRÜGER, J., SCHNEIDER, J., AND WESTERMANN, R. 2006. ClearView: An interactive context preserving hotspot visualization technique. *IEEE Trans. Visualization and Comput. Graph. 12*, 5.

KUTTER, O., BICHLMEIER, C., MICHAEL, S., OCKERT, B., EULER, E., AND NAVAB, N. 2008. Real-time Volume Rendering for High Quality Visualization in Augmented Reality. In *int. workshop on Augmented environments for Medical Imaging including Augmented Reality in Computer-aided Surgery*.

LIVINGSTON, M. A., SWAN II, J. E., GABBARD, J. L., HÖLLERER, T. H., HIX, D., JULIER, S. J., BAILLOT, Y., AND BROWN, D. 2003. Resolving Multiple Occluded Layers in Augmented Reality. In *Int. Symp. Mixed and Augmented Reality*, IEEE, 7–10.

MENDEZ, E., AND SCHMALSTIEG, D. 2009. Importance Masks for Revealing Occluded Objects in Augmented Reality. In *ACM Symp. Virtual Reality Software and Technology*, ACM, 247–248.

MENDEZ, E., KALKOFEN, D., AND SCHMALSTIEG, D. 2006. Interactive context-driven visualization tools for augmented reality. In *Int. Symp. Mixed and Augmented Reality*, IEEE, 209–218.

SCHALL, G., MENDEZ, E., KRUIJFF, E., VEAS, E., JUNGHANNS, S., REITINGER, B., AND SCHMALSTIEG, D. 2009. Handheld augmented reality for underground infrastructure visualization. *Personal Ubiquitous Comput. 13*, 4, 281–291.

VIOLA, I., KANITSAR, A., AND GRÖLLER, M. E. 2005. Importance-driven feature enhancement in volume visualization. *IEEE Trans. Visualization and Comput. Graph. 11*, 4, 408–418.

ZHONG, F., QIN, X., CHEN, J., HUA, W., AND PENG, Q. 2009. Confidence-based Color Modeling for Online Video Segmentation. In *Asian Conf. Comput. Vision*, Springer, 697–706.