Importance Masks for Revealing Occluded Objects in Augmented Reality

Erick Mendez[*]
Graz University of Technology

Dieter Schmalstieg[†]
Graz University of Technology

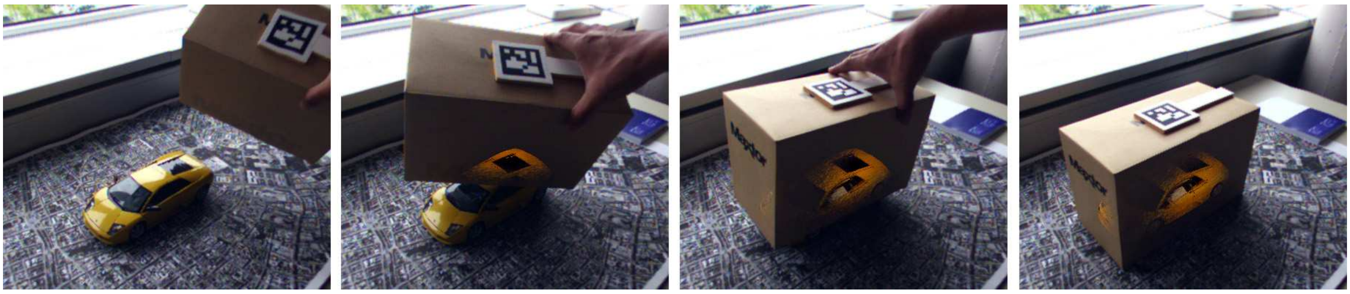

Figure 1: *Example of our technique. Notice how there is enough information of the occluder preserved even though it has little initial structure. Furthermore, the preserved information helps conveying spatial arrangements between virtual and real objects.*

Abstract

When simulating "X-ray vision" in Augmented Reality, a critical aspect is ensuring correct perception of the occluded objects position. Naïve overlay rendering of occluded objects on top of real-world occluders can lead to a misunderstanding of the visual scene and a poor perception of the depth. We present a simple technique to enhance the perception of the spatial arrangements in the scene. An importance mask associated with occluders informs the rendering what information can be overlaid and what should be preserved. This technique is independent of scene properties such as illumination and surface properties, which may be unknown. The proposed solution is computed efficiently in a single-pass fragment shaders on the GPU.

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

Keywords: augmented reality, focus and context, x-ray vision, importance masks

1 Introduction and Related Work

Occlusion management is an important topic in Virtual Reality and Augmented Reality (AR). Previous attempts at managing occluding and occluded objects include view restrictions [Bichlmeier et al. 2007], illustration techniques [Kruger et al. 2006] and context preservation [Kalkofen et al. 2009].

Interrante showed that merely overlaying the object that is about to be occluded on top of the - real or virtual - scene disrupts the perception of distance and spatial arrangements of the scene [Interrante 1996]. Such overlays merely swap the roles of the occluding and occluded objects. These swap does not preserve the occluder's contributions to perception of the overall scene, coming from its edges, texture, reflections, curvature and so on. Interrante suggested multiple strategies to preserve the information to be occluded so that it retained the perception of mutual arrangements across objects.

The topic was revisited in ClearView [Kruger et al. 2006] for VR scenes by harvesting information from occluding objects and overlaid on top of augmented objects. Bichlmeier et al. [Bichlmeier et al. 2007] used volume data obtained from CT scans and MRIs and overlaid them on their real counterparts in medical AR application. Kalkofen et al. [Kalkofen et al. 2007] used polygonal data in AR and extracted edges from both live video and polygonal objects for controlling occlusion effects. In a follow up work [Kalkofen et al. 2009], the authors computed a haloing of the image variance to better control the video information that should be preserved.

However, using a mask obtained from live video for preserving the structure of the occluder makes the result dependent on illumination and material properties of the physical environment. Reflection on the surface of occluders disturb the effect, and the result is strongly dependant on the viewing angle. Moreover, the occluding object with insufficient texture will be rendered almost completely transparent. Finally, the computation of the preserved features from the image is an expensive multi-pass operation.

In this work, we rely on a predefined importance mask - comparable to an alpha mask - to define what amount of information of the occluder should be preserved. This requires knowledge of the scene objects, but resolves all of the above shortcoming. It is therefore the recommended technique for AR applications with a detailed model of the scene, which is more or less a prerequisite for applications which intend to present meaningful augmentations in complex environments. The technique is also quite fast as it only depends in a single pass fragment shader, and the visual results empirically improve the perception of spatial arrangements of scene objects.

2 Overview Technique

The importance mask is created in a preprocessing step, through user interaction or heuristics such as mesh saliency [Lee et al. 2005]. It is assigned as a texture map to the occluder, and defines how transparent the occluder should be relative to the occludee. Figure 2, for example, shows a photograph of a real world object next to its substitute model for occlusion. This helps providing a

[*]e-mail: mendez@icg.tugraz.at
[†]e-mail: schmalstieg@icg.tugraz.at

smoother transition between occluded and visible states for the user. The importance masks is used during rendering to determine which fragments to preserve from the incoming video input and which ones to blend with the substitute model for occlusion.

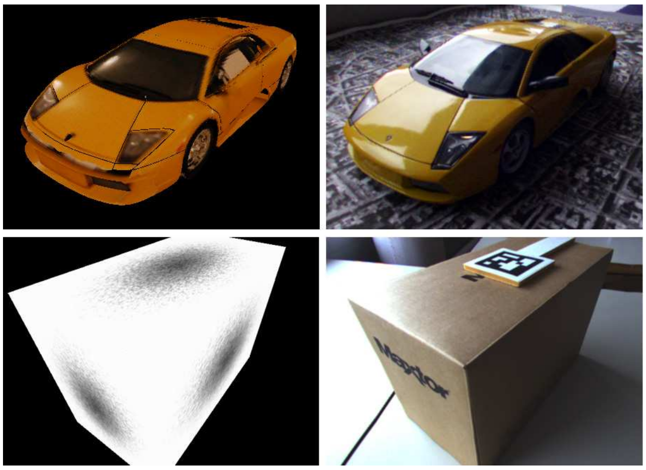

Figure 2: *(top left) Substitute model for occlusion. Notice how care has been taken to model not only the shape using a polygonal mesh, but also applying a photographed texture. (top right) Real world occludee candidate. (bottom left) Importance mask for the occluder (bottom right) Real box used as occluder.*

This technique is easily implementable in any fragment shader, such as GLSL. Given the position of occluder and occludee, we first collect the textures for compositing, for example in framebuffer objects. The substitute model is rendered to tex_0, while the occluder textured with the importance mask is rendered to tex_1. The background, usually the video feed for AR, is placed in tex_2 All those texture fragments that are not being used should have $\alpha = 0$. Finally we render a quad covering the whole are of the screen with the following fragment shader:

vec4 $T_{occluded} = texture2D(tex_0, gl\_FragCoord.xy).rgba$;
float $T_{mask} = texture2D(tex_1, gl\_FragCoord.xy).a$;
vec4 $T_{video} = texture2D(tex_2, gl\_FragCoord.xy).rgba$;
if $T_{occluded}.a == 0.0$ **then**
    $gl\_FragData[0].rgb = T_{video}.rgb$;
end
else if $T_{mask} == 0.0$ **then**
    $gl\_FragData[0].rgb = T_{video}.rgb$;
end
else
    $gl\_FragData[0].rgb =$
    $(T_{video}.rgb * T_{mask}) + (T_{occluded}.rgb * (1.0 - T_{mask}))$;
end

2.1 Results and Conclusion

Figure 1 presents a sequence, where a toy car is progressively occluded by a box. The second image in the sequence show parts of the real occlude and the substitute model at the same time. Figure 3 shows a virtual office scene inside a building (right) and a detailed view of the car interior through the body of a real car model (left).

These results nicely demonstrate the expressive capabilities of using importance masks for enhancing the augmentation of occluded objects. While it is effective and fast, it requires a to be created for all scene objects that will act as occluders. Moreover, said objects need to be tracked. This implies that precise registration will be required for convincing effects in AR scenes.

However, the technique is reasonable robust against registration errors, as small errors in registration of occluders will only render the preservation of occluders less efficient, but will rarely lead to disturbing artifacts.

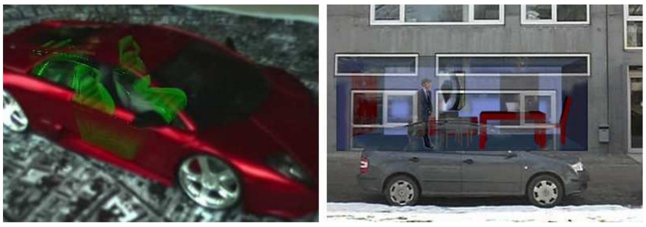

Figure 3: *Final composed result examples. The resulting images of our technique preserve information of the occluder in order to aid for a better perception of spatial arrangements. (left) The seats and engine of the car can be seen through the car body. (right) The office inside the building can be seen through.*

The presented technique has the significant practical advantage that it allows for objects without inherent texture or structure to be effectively used as occluder with appropriately preserved structural information. It also resistant to illumination view changes as the information preserved does not depend on the current viewing conditions. Moving the mask generation to a preprocessing step also which drastically cuts computation times. All images generated for this article were created on an nvidia GTX280 at more than 75fps.

Finally, the definition of the importance mask is independent of a particular application or camera setup and can come from any number of sources. It can be optimized using perceptual metrics or become part of artistic work in a production process. Our examples shows the liberty that application designers have when creating AR applications with this technique.

References

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 *IEEE International Symposium on Mixed and Augmented Reality(ISMAR)*.

INTERRANTE, V. 1996. *Illustrating Transparency: communicating the 3D shape of layered transparent surfaces via texture*. PhD thesis.

KALKOFEN, D., MENDEZ, E., AND SCHMALSTIEG, D. 2007. Interactive focus and context visualization for augmented reality. In *IEEE International Symposium on Mixed and Augmented Reality(ISMAR)*, 191–201.

KALKOFEN, D., MENDEZ, E., AND SCHMALSTIEG, D. 2009. Comprehensible visualization for augmented reality. *IEEE Transactions on Visualization and Computer Graphics 15*.

KRUGER, J., SCHNEIDER, J., AND WESTERMANN, R. 2006. Clearview: An interactive context preserving hotspot visualization technique. *Visualization and Computer Graphics, IEEE Transactions on 12*, 5 (Sept.-Oct.), 941–948.

LEE, C. H., VARSHNEY, A., AND JACOBD, D. 2005. Mesh saliency. In *ACM SIGGRAPH*.