Depth Compositing for Augmented Reality

Jonathan Ventura*
Department of Computer Science
University of California, Santa Barbara

Tobias Höllerer†
Department of Computer Science
University of California, Santa Barbara

Correct handling of occlusion is a significant challenge when compositing real and virtual content, whether it be for augmented reality or film. For film, the solution is often solved offline by arduously creating alpha mattes by hand. In the augmented reality context, the compositing must be real-time, so offline solutions are not possible. Occlusions are usually disregarded in augmented reality, so that virtual objects are always rendered on top of physical objects. If a highly accurate 3D model of the scene is available, then depth compositing can be performed automatically; however, such models can be difficult to create, and limit the range of the system to the extent of the model.

We have developed a method for automatic depth compositing which uses a stereo camera, without assuming static camera pose or constant illumination. The traditional approach to automatic depth compositing with a stereo camera uses the normal SAD block matching algorithm, and copies the disparity map into the z-buffer [Wloka and Anderson 1995; Schmidt et al. 2002]. These approaches result in disparity maps which have noise in texture-less regions and/or at edge boundaries. Berger's work [Berger 1997] uses snakes to find the outlines of occluded objects. However, this approach has not been proven to be real-time or sufficiently robust.

Kolmogorov et al. describe the Layered Graph Cut algorithm to fuse color and stereo likelihood for binary segmentation [Kolmogorov et al. 2005]. They learn color and depth distributions for foreground and background from hand-labelled data, and then use these distributions in a graph cut which encourages consistent labeling within smooth regions. In our work we extend the Layered Graph Cut to general depth compositing by decoupling the color and depth distributions, so that the depth distribution is determined by the disparity map of the virtual scene to be composited in.

Method

We restrict ourselves to the scenario where the interacting physical objects (e.g. hands) can be accurately described by a color histogram. In our system, color histograms are learned from video in an initialization procedure. After initialization, the histograms are updated with newly labeled pixels from subsequent frames, to provide robustness against camera movement and illumination change.

The compositing algorithm consists of two steps. First, we use a color-based graph cut to separate foreground F and background B. Second, we use a depth-based graph cut to separate F from FO, foreground pixels that are occluded by the virtual object. $\mathbf{d}'$ is the set of disparities of the virtual object, and d_{max} is the maximum disparity considered by the algorithm.

The energy E to be minimized by the graph cut is defined as:

$$E(\mathbf{z}, \mathbf{x}; \Theta) = U(\mathbf{z}, \mathbf{x} : \Theta) + V(\mathbf{z}, \mathbf{x}) \quad (1)$$

where $\mathbf{z}$ contains the color pixels in the left image, $\mathbf{x}$ is the labeling, and V is the normal color boundary term [Kolmogorov et al. 2005]. For color segmentation, Θ contains the color histograms and U is the color likelihood. For depth segmentation, Θ contains the match likelihoods (for all disparities d) between the left and right image, and $U = U^S$ as defined by Kolmogorov et al. For pixel location m, $p(d_m = d|x = F)$ is uniform between d'_m and d_{max}, and $p(d_m = d|x = FO)$ is uniform between 0 and d'_m.

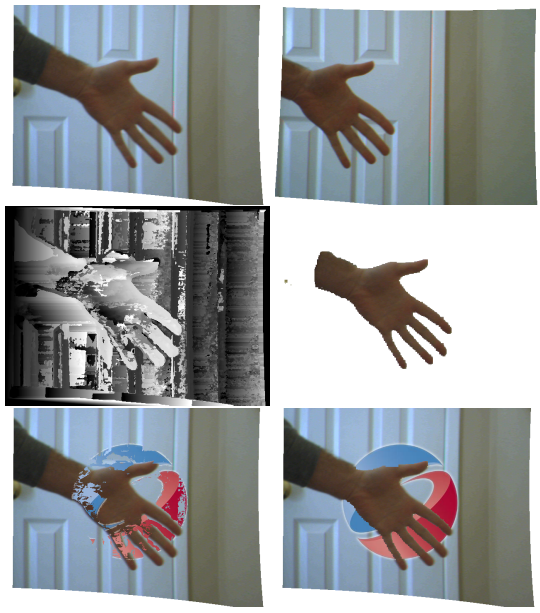

Figure 1: *Depth compositing example (left to right, top to bottom): left and right stereo pair; depth map and color segmentation; compositing based on depth alone, and compositing with graph cut.*

As Figure 1 shows, the graph cut achieves a clean composite of the hand with a virtual object, as opposed to the composite using a stereo depth map, which contains significant noise.

References

BERGER, M.-O. 1997. Resolving occlusion in augmented reality: a contour based approach without 3d reconstruction. *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, 91–96.

KOLMOGOROV, V., CRIMINISI, A., BLAKE, A., CROSS, G., AND ROTHER, C. 2005. Bi-layer segmentation of binocular stereo video. *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on 2*, 1186 vol. 2–.

SCHMIDT, J., NIEMANN, H., AND VOGT, S. 2002. Dense disparity maps in real-time with an application to augmented reality. *Applications of Computer Vision, 2002. (WACV 2002). Proceedings. Sixth IEEE Workshop on*, 225–230.

WLOKA, M. M., AND ANDERSON, B. G. 1995. Resolving occlusion in augmented reality. In *SI3D '95: Proceedings of the 1995 symposium on Interactive 3D graphics*, ACM, New York, NY, USA, 5–12.

*e-mail: jventura@cs.ucsb.edu
†e-mail:holl@cs.ucsb.edu