Defocus Difference Matting (sketches_0456)

Morgan McGuire
Brown University

Wojciech Matusik
MERL

Introduction: Matting is a classic problem in both computer graphics and vision. Practical solutions that yield high quality results are important for special effects in the movie industry. In our defocus difference matting (DDM) approach, we assume that the background is known, typically because it is static and pre-recorded. In this respect, our work is related to blue-screen matting. Two common problems with blue-screen matting are the limitations imposed on the color of the foreground, e.g., the actor cannot wear a blue shirt, and more importantly a color spill of the background on the foreground, which considerably changes the lighting of the scene. Our method alleviates both of these problems.

DDM is different from background subtraction because it operates on defocus (color derivative) differences instead of image differences, which allows it to pull mattes even at pixels where the foreground and background colors are identical. When the background is known, pulling the matte is still an underconstrained problem. Smith and Blinn's triangulation method [1996] used two different backgrounds, which is not practical for live subjects. We instead constrain the solution by using two video streams that share a common center of projection but different focus characteristics.

A related recent method *defocus video matting* [McGuire et al. 2005] uses three video streams with different depth-of-field and focus that share the same center of projection to pull mattes for scenes with unconstrained, dynamic backgrounds. DDM can be seen as a special case of this.. Compared to the original defocus matting, DDM requires a static background with high frequencies but achieves higher quality results, is very simple to implement, and can operate in real-time (defocus matting takes several minutes per frame). Also, we use only two video cameras and a single beam splitter, which provides twice as much light.

We have obtained excellent results for synthetic images with both patterned and natural backgrounds. For real images our results are acceptable but not ideal. This is because of the primary drawback of DDM; it is very sensitive to color and alignment calibration of the cameras, which we currently achieve only by physical adjustment. As future work we intend to make the method more robust by adding software correction via optical flow estimation, which has been applied successfully in other matting algorithms.

Algorithm: Assume a scene containing a foreground object whose image is αF and background whose image is B. The image formation equations for a pinhole camera (I_1) and a narrow-depth-of-field camera focussed on the foreground (I_2) are well-approximated by:

$$I_1 = \alpha(F - B_1) + B_1 \qquad (1)$$

$$I_2 = \alpha(F - B_2) + B_2 \qquad (2)$$

The foreground appears identical in these images and the background differs only by defocus. If we have previously captured images of the background by both cameras, then B_1 and B_2 are known and we can solve directly for the matte and matted foreground:

$$\alpha = 1 - \frac{I_1 - I_2}{B_1 - B_2} \qquad (3)$$

$$\alpha F = I_1 + (\alpha - 1) B_1 \qquad (4)$$

When the difference between the two background images is small, these equations are ill-conditioned. We detect pixels where the result is ill-conditioned and replace the matte at those pixels with a value interpolated from well-conditioned neighbors.

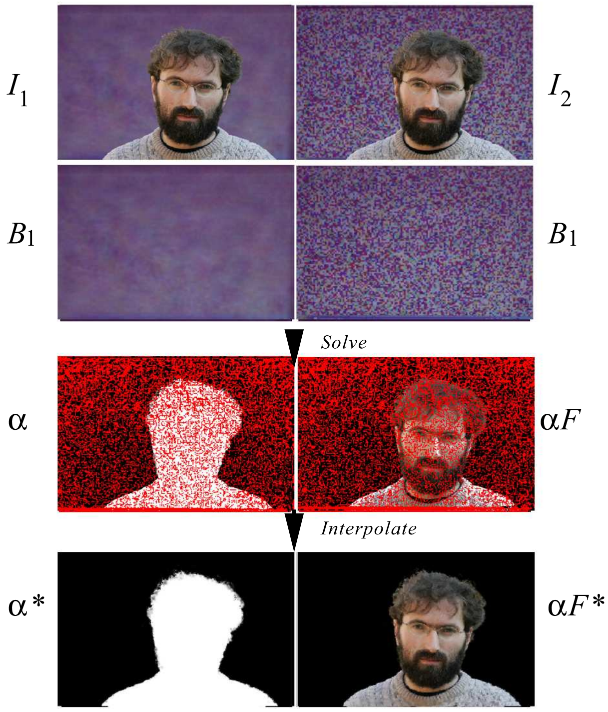

Figure 1: Top: Input images and backgrounds. Center: Recovered matte and foreground with ill-conditioned pixels marked red. Bottom: Matte and foreground reconstructed from well-conditioned samples.

Figure 2: Input and result for a synthetic image with a natural background and for a real image with a striped background.

For perfectly calibrated data we obtain excellent results, e.g., Figure 1 shows results on synthetic data composed from real photographs (and the defocus is from the real camera). For real data miscalibration can introduce noise into the $\alpha \approx 1$ regions, so we post-process the results to improve region coherence. The $\alpha \approx 0$ areas are stable because we choose camera 2's background circle of confusion to be twice the highest background frequency, which itself should ideally be less than half the sampling frequency (resolution) of camera 1.

References

McGuire, M., Matusik, W., Pfister, H., Durand, F., and Hughes, J. 2005. Defocus Video Matting. *ACM Transactions on Graphics 24*, 3 (Aug.).

Smith, A. R., and Blinn, J. F. 1996. Blue screen matting. In *Proc. Computer graphics and interactive techniques*, ACM Press, 259–268.