Image-based Environment Matting

Yonatan Wexler, Andrew W. Fitzgibbon, Andrew Zisserman*
The University of Oxford

Environment matting is a powerful technique for modelling the complex light-transport properties of real-world *optically active elements*: transparent, refractive and reflective objects. Zongker *et al* [1999] and Chuang *et al* [2000] show how environment mattes can be computed for real objects under carefully controlled laboratory conditions. However, for many objects of interest, such calibration is difficult to arrange. For example, we might wish to determine the distortion caused by filming through an ancient window where the glass has flowed; we may have access only to archive footage; or we might simply want a more convenient means of acquiring the matte.

We show in this sketch that accurate environment mattes can be computed from natural images, without the need for specialized calibration of the acquisition. The goal is to take a set of example images, containing the optical element of interest (e.g. the lens in figure 1), and transfer the element's environment matte to a new background image (example in figure 3).

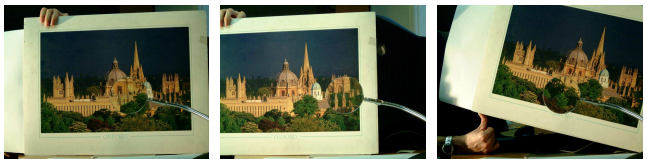


Figure 1: ***Input****: Three of a sequence of 42 images, static optical element (magnifying glass), moving background. The environment matte is computed using only the information in these images.*

The technique is best understood by working backwards from the final composite of a novel background image N and the computed environment matte. Each pixel in the output collects light from a blend of pixels in N. Let us call the set of pixels which contribute to a given output pixel p the *footprint* of p, or p's *receptive field*. Previous researchers have defined the footprint using rectangular regions [Zongker et al. 1999] or mixtures of Gaussians [Chuang et al. 2000]. In this work, we must deal with complex multimodal distributions, so we use a discrete map of source pixels, where each source pixel has an associated weight. The value of the output pixel is then computed as a weighted sum over the pixels of N. Thus if we can compute the receptive field for each pixel, we can compute the composite.

In order to compute the receptive field of a given pixel p, we need at least two images: one containing the test object (e.g. the lens in figure 1), and one containing only the background (figure 2). We note that pixels in the background which have contributed to p's colour will have similar colour to p. In fact, for each background pixel, the similarity between its colour and the query colour is a function of the amount that background pixel contributes. Thus, we can obtain a bound on p's receptive field by computing the correlation between a small (e.g. 3×3) window around p and each location in the background image. Such a bound is illustrated in figure 2c. Of course, for a single image, this bound is very weak—many pixels which accidentally share p's colour are included in the receptive field. However, with a sequence of images, as in figure 1, the receptive field is constant as the background moves, and with each new image, the footprint can be refined. Figure 2d shows the refined receptive field for the indicated foreground pixel after 8 views have been integrated. Note how the single peak corresponds to the true source pixel, indicated in figure 2b.

*e-mail:{wexler,awf,az}@robots.ox.ac.uk

Step 1: Compute clean background for each image:



Step 2: Receptive field (RF) computed for one pixel. (a) Foreground. (b) Background. (c) RF for this pair. (d) RF from all pairs.

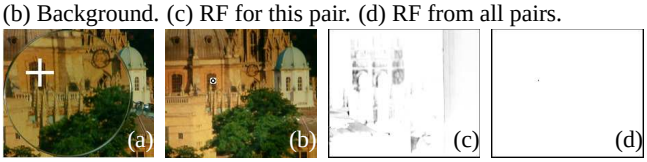


Figure 2: ***Steps in the algorithm***

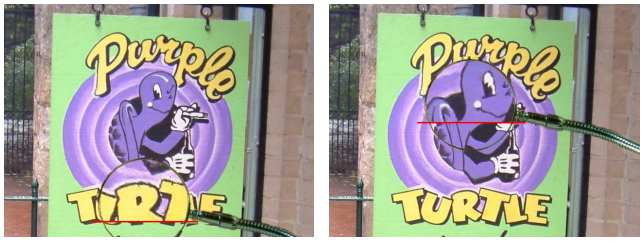


Figure 3: ***Output****: Recovered environment matte over new image. Compare the environment matte (above red line) and transparency (below red line).*

Computing the background image may be achieved by mosaicing the moving-background sequence [Irani et al. 1994] or moving the camera. Figure 4 shows an example where the camera is moved to obtain a clean view of the background. In this example, there is just one reference view, so strong regularizing constraints were employed in order to permit a solution: the receptive fields were assumed small and close to their source pixels.

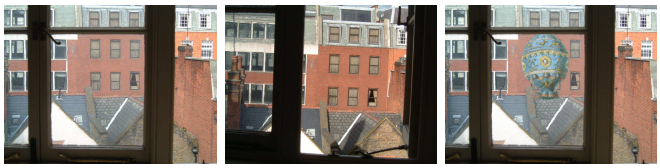


Figure 4: *Base image, single reference view of background, composite using computed environment matte.*

The examples show that, although its performance is scene-dependent, the technique can work well given sufficiently rich backgrounds, or sufficiently many images. They demonstrate that environment mattes can be captured under less stringent assumptions than have previously been described.

References

CHUANG, Y.-Y., ZONGKER, D. E., HINDORFF, J., CURLESS, B., SALESIN, D. H., AND SZELISKI, R. 2000. Environment matting extensions: Towards higher accuracy and real-time capture. In *Proceedings of ACM SIGGRAPH*, 12–130.

IRANI, M., ROUSSO, B., AND PELEG, S. 1994. Computing occluding and transparent motions. *Intl. J. Computer Vision 12*, 1, 5–16.

ZONGKER, D. E., WERNER, D. M., CURLESS, B., AND SALESIN., D. H. 1999. Environment matting and compositing. In *Proceedings of ACM SIGGRAPH*, 205–214.