Seeing Around Occluding Objects

Scott McCloskey, Michael Langer, and Kaleem Siddiqi
Centre for Intelligent Machines
McGill University
Montreal, Quebec H3A 2A7 Canada
{scott, langer, siddiqi}@cim.mcgill.ca

Abstract

This paper presents a novel method for the removal of unwanted image intensity due to occluding objects far from the plane of focus. Such occlusions may arise in scenes with large depth discontinuities, and result in image regions where both the occluding and background objects contribute to pixel intensities. The contribution of the occluding object's radiance is modeled by reverse projection, and can be removed from this region by a simple operation on the pixel's intensity. Experimental results demonstrate our ability to accurately recover the background's appearance despite significant occlusion. As compared with processing based a linear model of occlusion, the results show lower error and a more accurate contrast.

1. Introduction

There are a number of problems that might cause a photographer to discard an image. Researchers have studied many of these problems, such as red-eye or dust on a lens, and developed methods that remove their unwanted effects. One common photographic problem that has not been addressed in this manner arises when an object, such as a camera strap or the photographer's finger, falls within the camera's field of view. This is a common problem with cameras that have short lenses, and results in images that exhibit both partial and total occlusion. The partially occluded region results from the blurred image of the occluding object, which is far from the plane of focus.

While photographer error is a common cause of such a problem, the same effect results from scenes that have a wide range of depths, such as a landscape taken through a window. In the event that the photographer *chooses* to remove the contribution of the nearby object, we show that an accurate model of image formation allows for the removal of the occluder in the blurred region.

As the title of [1] suggests, a reverse projection model for blurring can be used to see behind nearby occluding objects. The model shows that the intensity of a pixel is a combination of light from the occluding object and the background. By understanding the proportions in which the light combines, it is possible to remove the contribution of the occluding object in order to recover the intended scene in the blurred region.

2. Previous Work

In Pentland's original work on depth from defocus [6], an algorithm was presented to find scene depth by measuring optical blur. At occlusion edges, under certain conditions, it was demonstrated that depth could be recovered by measuring blur in a single image. It was mentioned in a footnote (on p. 524) that the measured depth was the distance to the nearer object, though it was not immediately clear why, since there is also a well-defined distance to the background object. Partially in response to this, Asada *et al* introduced the reverse projection model for blur [1] and showed that blurring of an occlusion edge is consistent with blurring of an radiance discontinuity on a surface at the depth of the occluding object. The paper went on to show that finite apertures allow the camera to collect light from parts of the background that would not be visible to a pinhole camera. Schechner and Kiryati also examine partial occlusions in the context of depth from defocus in [7].

In [4], Favaro and Soatto use a reverse projection model and develop an algorithm which reconstructs the geometry and radiance of a scene from multiple images with different focal positions. While this method could be used to remove the contribution of occluding surfaces, it is not applicable to a photographer who would like to improve a *single* image.

A more distantly related technique is presented by Levoy *et al* in [5]. The paper discusses synthetic aperture imaging, and makes note of its ability to see around occluding objects. Though this ability is one of the key features of their system, they make no effort to remove partial occlusion in the images by image processing.

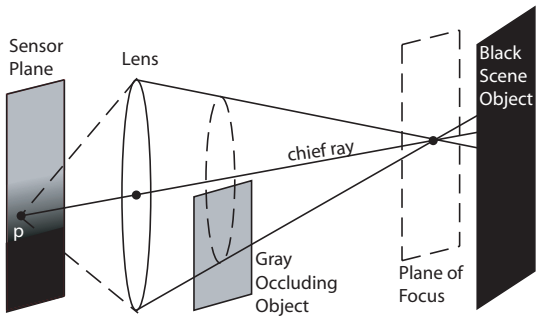


Figure 1. Imaging a black scene object with a gray object occluding. Recall that lenses invert images.

3. Focusing and Reverse Projection

All cameras with finite apertures are limited in the part of the scene that they can properly focus on a sensor. The locus of scene points that will appear well focused in an image is the *plane of focus*. For any point on the sensor, we define the *chief ray*, which is based at the sensor point and passes through the optical center of the lens (see Fig. 1). For a circular aperture, the intensity observed at a point on the sensor is related to the amount of light emitted or reflected from points within a double cone. The double cone has a base at the lens aperture and its apex at the point where the chief ray intersects the plane of focus.

An occluding object will be visible at a point on the sensor if it intersects this cone, as in Figure 1. The contribution of the occluding object to the intensity at the point will be related to its radiance and the fraction α of the cone's base that it subtends. If we assume that the occluding object has locally constant radiance R_o, and that the scene is well-focused at a point with radiance R_s, the average radiance incident at pixel p will be:

$$R(p) = \alpha(p)R_o + (1 - \alpha(p))R_s(p) \ . \tag{1}$$

Note that this is simply a blending process, such as compositing in graphics.

After tone mapping, the intensity of the pixel is a nonlinear function of $R(p)$; that is, $I(p) = f(R(p))$. We would like to recover is the intensity $I'(p)$ that would be observed in the absence of occlusion, which can be expressed as

$$I'(p) = f(R_s) = f\left(\frac{f^{-1}(I(p)) - \alpha(p)R_o}{1 - \alpha(p)}\right) \ . \tag{2}$$

Thus, in order to recover the intended intensity, we need to characterize the radiance of the occluding object R_o and the degree of occlusion α at each pixel. We also need to estimate f; for our experiments, we use the method in [2].

4. What is $\alpha(p)$?

Clearly, the function $\alpha(p)$ takes a value of 0 when the double cone defined by p's chief ray doesn't intersect the occluding object, and the value 1 when complete occlusion occurs. What are the values in the intermediate points? If we assume that the occluding contour lies in a plane parallel to the sensor, we need only look at the circular cross-section of p's cone in that plane. In such a case, α will be the fraction of the circle's area subtended by the occluding object. In the event that the occluding contour is locally linear, as in Figure 1, the area subtended by the occluding object will be

$$\begin{aligned} A(p) &= 2\int_{x_o(p)}^{\sigma} \sqrt{\sigma^2 - x^2}dx \\ &= \frac{\pi\sigma^2}{2} - x_o(p)\sqrt{\sigma^2 - x_o(p)^2} \\ &\quad -\sigma^2 \arcsin\left(\frac{x_o(p)}{\sigma}\right) \ , \end{aligned} \tag{3}$$

where $x_o(p) \in [-\sigma, \sigma]$ is the distance of the occluding contour from the chief ray of sensor point p, and σ is the radius of the circular cross-section of p's cone. Thus, α is simply

$$\alpha(p) = \frac{A(p)}{\pi\sigma^2} \ . \tag{4}$$

Restated in terms of $y(p) = \frac{x_o(p)}{\sigma} \in [-1, 1]$, equations 3 and 4 simplify to

$$\alpha(p) = \frac{1}{2} - \frac{y(p)\sqrt{1-y(p)^2} + \arcsin(y(p))}{\pi} \tag{5}$$

This function is nonlinear, particularly around the beginning and end of the occlusion. Without modeling the image formation process, one might naively assume that the α function is linear. The reader should note that this naive assumption, which will be presented for reference in the experiments section, is equivalent to a rectangular aperture.

5. Implementation Issues and User Input

Removing occluding objects from pictures can be thought of as two separate subtasks. The first is a vision task, which involves the identification and localization of blurring in the image. The second, and the subject of this work, is an image processing task of removing the unwanted contribution from the occluding object.

Any complete system for the removal of occluding objects must address both of these subtasks. A computer vision solution to the first subproblem is outside the scope of this work, so examples in this paper employ user interaction for localization. It is possible for a user to completely specify the necessary parameters for a linear occluding edge with only three mouse clicks. Two clicks identify a line L in the image at which the occluding object completely subtends the pixel's cone (i.e. when the background can no

longer be seen through the blur). With the third click, the user specifies a point p_0 at which the blurred contribution of the occluding object becomes apparent. Given L and p_0, σ is half of the distance $dist(p_0, L)$ between them. The intensity of the occluding object for a pixel is simply the intensity of the image at the nearest point on the line L.

We can estimate the value of $y(p)$ based on a pixel's distance from the line. Completely occluded pixels are given $y = -1$ by construction. Elsewhere,

$$y(p) = \min\left(2, \frac{2dist(p, L)}{dist(p_0, L)}\right) - 1 \; . \tag{6}$$

Note that pixels in the blurred region will have $y \in (-1, 1)$.

In practice, specifying the parameters in this way is inexact. An implementation of this method could include visual feedback to the user in order to assist with the positioning of the line and point p_0.

Numerically, one should place a limit on the value of α that we use in equation 2, which becomes unstable as $\alpha \to 1$. Note that any value of $\alpha > 0$, this method improves the image by removing the contribution of the occluding object. For $\alpha > .5$ this can be considered seeing behind the scene, in that the background object would not be visible in a pinhole image due to occlusion of the chief ray.

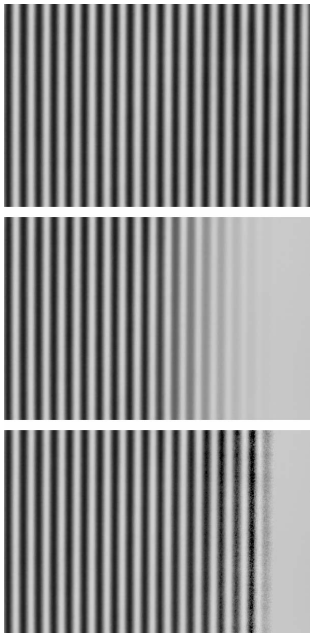


Figure 2. Validation Experiment. (Top) Background without occlusion. (Middle) Occluded image. (Bottom) Processed image.

6. Validation Experiment

For our first experiment, we construct a real scene that consists of a sinusoidal intensity pattern on a wall. The occluding object is a cardboard box. Figure 2 shows images of the unoccluded background (ground truth), occluded input, and processed result. The blur localization was performed in the manner described in Section 5. The maximum value of α used in eq. 2 for this example was 0.9. Fig. 3(a) shows 1D profiles along a row in the images from Fig. 2.

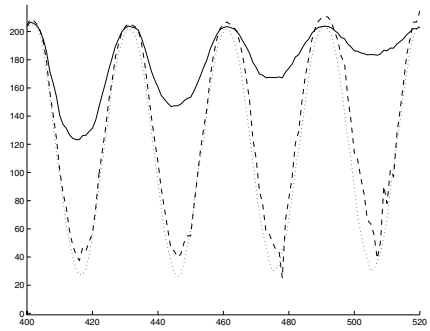
For the sake of comparison, Figure 3(b) shows the result under the naive assumption that α is a linear function. Note that, as a result, α is overestimated for slight occlusions and underestimated for occlusions of more than half of the viewing cone. Visually, this results in artificially high and low contrasts in these two regions, respectively. We measure the root mean squared error (RMSE) in the profile where $0 < \alpha < .9$, as neither method produces good results for large values of α. The RMSE of our method is 14.7 pixel code values, as compared to 28.6 for the naive method.

7. Demonstration on a Natural Scene

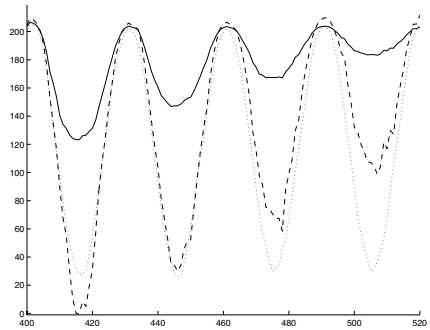
A further experiment was performed to show that, while the assumptions listed in Section 4 may seem strict, this method is applicable to interesting situations. In this example, the occluding object was a plastic fork (roughly the same shape as a finger), which was neither Lambertian nor had an occluding contour that was linear. Because the method in Section 5 is not applicable in the case of a non-linear contour, the location of the occluding object was manually specified by a mask indicating points that were completely occluded by the object. The extent of the blurring was specified in terms of the width of the blurred region, and the value of $y(p)$ was determined by the distance between p and the nearest completely occluded pixel.

Figure 4 shows the input image and the processed output, which allows the viewer to read more of the text. Viewers may note that the region around the occluding object is low in contrast which, recalling Figure 3(b), results from an underestimate of α. In the low contrast region, pixels are processed using the maximum allowed value of α (0.96 in this case) instead of the true value, which is higher. Using a higher maximum value of α results in an unpleasant image (not shown), as the contribution of noise is greater than that of the background object in this region.

The images shown in Figure 4 are 500-by-1000 pixels, cropped from 6 MP to show greater detail. The black region in the lower left of Figure 4 (bottom) indicates the mask; pixels in this region are not processed. The user observed blurring in the original image at all points within 775 pixels of the mask.



(a) Proposed processing method (dashed)



(b) Processing with linear model (dashed)

Figure 3. 1D profiles of outputs, compared to input (solid) and ground truth (dotted).

8. Summary and Conclusions

This paper has presented a novel method for the removal, from a single image, of unwanted contamination due to occluding objects near the camera. Such situations arise in photographs when, for instance, the camera strap falls in front of the lens and goes unnoticed due to a non-TTL (through the lens) viewfinder. More generally, this may arise when the photographer is unable to compose the scene without an occluding object, due to the presence of a building, tree, etc. The method proposed here uses the reverse projection model to determine the contribution of the occluding object to each pixel. Given an estimate of the occluding object's location and the extent of blurring, that contribution can be removed via a simple processing step. Experimental results have validated this model, and have demonstrated utility in a context with limited constraints.

Since this method is only applicable to pixels that exhibit *partial* occlusion, other methods such as in-painting [3] may be used to replace pixels that are completely blocked in the original image. We stress that, unlike in-painting methods that invent image content, we recover true data by modeling the image formation process.

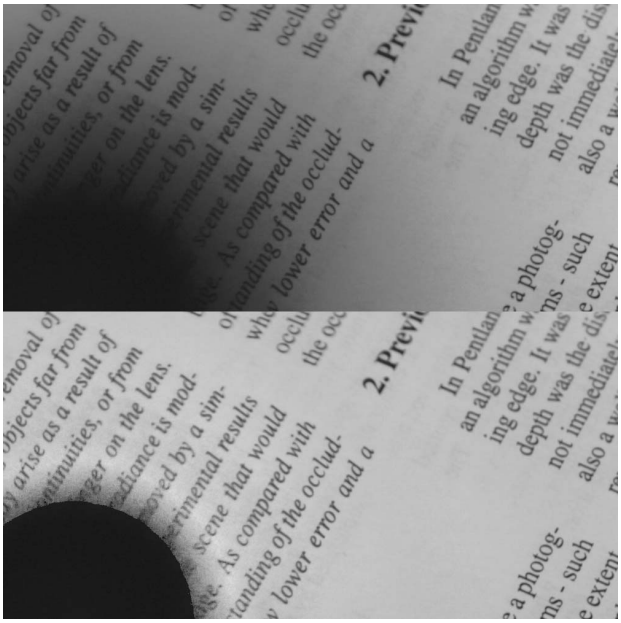


Figure 4. Camera (top) and processed (bottom) images with an occluding object.

References

[1] N. Asada, H. Fujiwara, and T. Matsuyama. *Seeing Behind the Scene: Analysis of Photometric Properties of Occluding Edges by the Reversed Projection Blurring Model*. IEEE Trans. on Patt. Anal. and Mach. Intell., Vol. 20, pp. 155-167, 1998.

[2] P. Debevec and J. Malik. *Recovering High Dynamic Range Radiance Maps from Photographs*. Proc. SIGGRAPH 1997, pp. 369-378

[3] A. Efros and T. Leung. *Texture Synthesis by Non-parametric Sampling*. Proc. Intl. Conf. on Comp. Vis. 1999, pp 1033-1038.

[4] P. Favaro and S. Soatto. *Seeing Beyond Occlusions (and other marvels of a finite lens aperture)*. Proc. of the IEEE Conf. on Comp. Vis. and Patt. Recog. 2003, pp 579-586.

[5] M. Levoy, B. Chen, V. Vaish, M. Horowitz, I. McDowall, M. Bolas. *Synthetic Aperture Confocal Imaging*. Proc. SIGGRAPH 2004, pp. 825-834.

[6] A. Pentland *A New Sense for Depth of Field*. IEEE Trans. on Patt. Anal. and Mach. Intell., vol. 9, no. 4, pp. 523-531, July 1987.

[7] Y. Schechner and N. Kiryati. *Depth from Defocus vs. Stereo: How Different Really Are They?*. Intl. J. of Comp. Vision, 39(2):141-162, 2000.