Reconstructing Occluded Surfaces using Synthetic Apertures: Stereo, Focus and Robust Measures

Vaibhav Vaish* Richard Szeliski† C. L. Zitnick† Sing Bing Kang†
Marc Levoy*
*Computer Science Department †Microsoft Research
Stanford University, CA Redmond, WA

Abstract

Most algorithms for 3D reconstruction from images use cost functions based on SSD, which assume that the surfaces being reconstructed are visible to all cameras. This makes it difficult to reconstruct objects which are partially occluded. Recently, researchers working with large camera arrays have shown it is possible to "see through" occlusions using a technique called synthetic aperture focusing. This suggests that we can design alternative cost functions that are robust to occlusions using synthetic apertures. Our paper explores this design space. We compare classical shape from stereo with shape from synthetic aperture focus. We also describe two variants of multi-view stereo based on color medians and entropy that increase robustness to occlusions. We present an experimental comparison of these cost functions on complex light fields, measuring their accuracy against the amount of occlusion.

1. Introduction

Reconstructing the shape and appearance of objects behind partial occlusions is a challenge for current 3D reconstruction algorithms, even for Lambertian scenes. One problem is the limited number of views; as a result, we may not be able to reliably match partially occluded objects across the views. Another problem is the cost functions used by most algorithms (based on SSD, SAD, normalized cross-correlation, etc.) implicitly assume that the surfaces being reconstructed are visible in all views. Occlusions may be compensated for later in the pipeline [9] or ignored completely [10].

The first problem can be addressed by simply imaging the scene with sufficiently many cameras (e.g., a 100-camera array [19]). When we have enough cameras which span a baseline (or *synthetic aperture*) wider than the occluders in scene, we can capture enough rays that go around the foreground occluders and are incident on the partially occluded background objects. Using a technique called *synthetic aperture focusing*, researchers have used large camera arrays to image objects behind dense occlusions like foliage [7, 15] or people in crowds [16].

To address the second problem, we need to design alternative cost functions which are robust to occlusion. One approach is project all camera images on to a virtual focal plane and compute their mean. In the resulting image, objects at the depth of the focal plane will be aligned and sharp while occluders in front will be blurred. By using a sharpness measure, we could identify the objects at the current depth. This is a synthetic aperture analogue of shape from focus. A second approach is to enforce color constancy as in standard multi-baseline stereo, while being robust to outliers resulting from occlusions. This raises the question: which of the two (stereo or focus) is better ?

In this paper, we explore cost functions for reconstructing occluded surfaces from synthetic apertures. Our first contribution is a comparison of shape from stereo with shape from synthetic aperture focus. Experiments indicate that focus performs better for sufficiently textured surfaces as occlusion increases. Our second contribution is the development of two variants of stereo, which are more robust to occlusions than standard multi-view stereo. These are based on color medians and entropy.

There are two major differences between our work and previous reconstruction algorithms. First, we use many more cameras spanning a wider synthetic aperture than previous methods. Second, we seek to reconstruct objects *occluded* in a significant portion of the input images, i.e. we expect several outliers amongst pixels to be matched across views. This is a more general and harder problem than standard multi-view stereo [10], which assumes no outliers. The occluders themselves are not reconstructed explicitly; we rely on the increased number of views with a large aperture and robust cost functions to achieve good reconstruction.

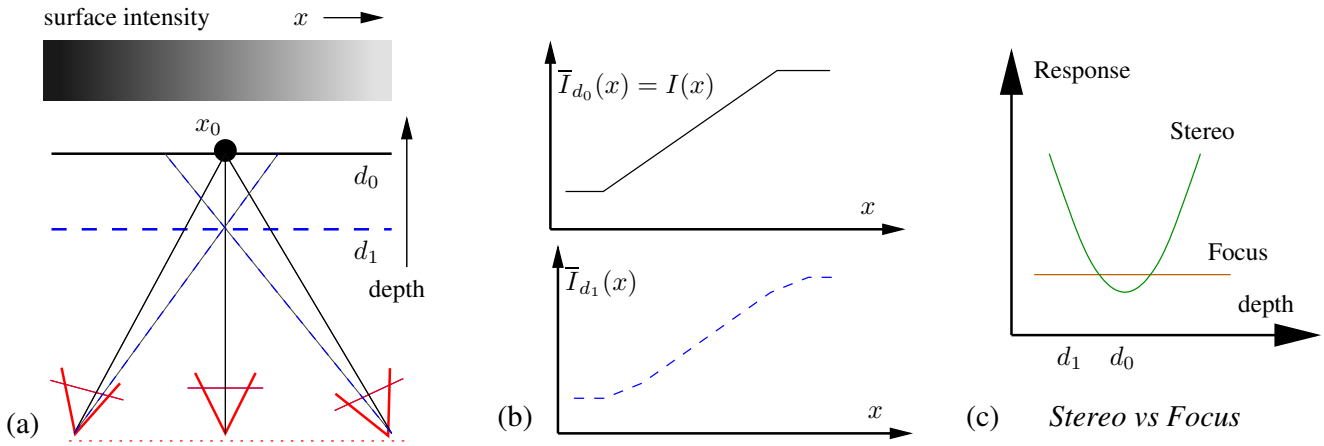


Figure 1. Stereo vs. focus for a surface at depth d_0 with a constant intensity gradient. (a) Camera layout, and rays considered for depth hypotheses $d = d_0$ and $d = d_1$. (b) The intensity profiles of the mean of camera images, projected on to depth planes $d = d_0$ (top) and $d = d_1$ (bottom). (c) Comparing the response of stereo (variance) and focus. The variance has a parabolic profile, with the minimum at the correct depth d_0. The mean image does not vary with depth, so shape from focus is not able to reconstruct depth accurately.

1.1. Related Work

In their comparison of shape from stereo versus focus [12], Schechner et al observed that focus ought be more robust than stereo in presence of occlusions while stereo ought to be more accurate due to larger baselines. This suggests that a synthetic aperture baseline may let us achieve both accuracy and robustness, and was one of the inspirations for our research. The use of a finite aperture of a single lens to see beyond occluders was studied by Favaro [5]. The limited aperture size (and hence limited depth of field) of a single lens limits the depth resolution and the size of occluders that can be seen around. Moreover, a single lens captures a (3D) stack of 2D images focused at depths fixed during acquisition, as opposed to a 4D light field captured by a synthetic aperture. This prevents us from using robust cost functions based on medians and entropy.

In [8], Kang et al introduced a cost function based on view selection to increase robustness to occlusions. Their method reconstructs the depths of only those points visible in a reference view. Our cost function based on color medians generalizes this, so that visiblity in a reference view is not required.

Several algorithms for 3D reconstruction work by partitioning the scene into layers at different depths [17, 2, 20]. Translucent layers were considered in [14]. These approaches usually begin by first estimating the foreground layer and using residuals to estimate the rest. For complex foreground occluders whose reconstruction may be error-prone, errors in foreground estimation can severely degrade the reconstruction of the background. Since we seek to develop cost functions robust to occluders, we do not require an initial reconstruction of the foreground layer. In our experiments, we restrict the search for the occluded objects to depths behind the foreground.

Algorithms based on voxel coloring [13] attempt a complete scene reconstruction using a front-to-back sweep. Since a threshold is typically used to commit to foreground occluders, erroneous assignments result in background pixels marked as foreground or foreground pixels not deleted in background reconstruction. Both cases degrade the reconstruction of the background. Thresholding effects can be ameliorated to an extent using iterative probabilistic variants [4], but these are not guaranteed to converge. In section 4, we compare our approach with voxel coloring.

2. Stereo vs. Focus

We begin by analytically comparing shape from (synthetic) focus and shape from stereo. A common framework to describe the two is the space-sweep approach of Collins [3]. This involves sweeping a plane over a range of depths in the scene. At each depth (more precisely, disparity) d we compute a cost value for every pixel (x, y) on the plane, creating a *disparity space image* (DSI) $D(x, y, d)$ [14]. The main difference between stereo and focus methods is that they use different cost functions in constructing the DSI. A depth map $d(x, y)$ is computed by finding the minimum cost surface from the DSI: $d(x, y) = \arg\min_d D(x, y, d)$.

We now define the cost functions used by stereo and focus. For simplicity, we work in flatland: the cameras are rectified line cameras and the depth of a pixel determines its horizontal disparity with respect to a reference view [10].

Let $I_{i,d}(x)$ denote the image from camera i projected on the plane corresponding to disparity d. The mean and variance of the warped images from the N cameras are given by

$$\overline{I}_d(x) = \frac{1}{N}\sum_i I_{i,d}(x), \quad (1)$$

$$v_d(x) = \frac{1}{N}\sum_i (I_{i,d}(x) - \overline{I}_d(x))^2 \quad (2)$$

Shape from stereo uses the variance of rays $v_d(x, y)$ through the 3D point (x, y, d) as the cost function. Shape from focus uses the sharpness of the synthetically focused image $\overline{I}_d(x)$ as the cost function:

$$f_d(x) = -\left[\frac{\partial \overline{I}_d(x)}{\partial x}\right]^2 \quad (3)$$

Which method is better? One straightforward observation we can make is that focus is less sensitive to sensor noise (due to averaging) and varying bias across cameras (since it computes a spatial derivative) than stereo. While both would fail for textureless surfaces, we can show that there exist textured surfaces which can be reconstructed accurately by stereo but not by focus (see Fig 1.)

Theorem 1. *Shape from stereo can reconstruct the depth of diffuse surfaces whose 2nd and higher order spatial derivatives of radiance vanish, whereas shape from focus requires the radiance to have non-zero 3rd-order spatial derivatives*

Proof. Consider a frontoparallel surface at disparity d_0 with intensity given by $I(x)$. We compute the response of stereo and focus at a point x_0 for disparity d_1. Let $\delta = d_1 - d_0$ and x_i the displacement of camera i from the central view. Given rectified images, we can compute $I_{i,d_1}(x)$ using Taylor series:

$$\begin{aligned} I_{i,d_1}(x_0) &= I(x_0 + x_i\delta) \\ &= I(x_0) + (x_i\delta)I_x(x_0) + \frac{(x_i\delta)^2}{2}I_{xx}(x_0) + \ldots \end{aligned}$$

Using eqs. (1-3) and $\sum_i x_i = 0$, we can compute the mean image, variance and focus to be

$$\begin{aligned} \overline{I}_{d_1}(x_0) &= I(x_0) + I_{xx}(x_0)\delta^2 \sum_i x_i^2/N + O(\delta^3), \\ v_d(x_0) &= \delta^2 I_x(x_0) \sum_i x_i^2/N + O(\delta^4), \text{ and} \\ f_d(x_0) &= -[I_x(x_0) + \delta^2 I_{xxx}(x_0) \sum_i x_i^2/N + O(\delta^4)]^2 \end{aligned}$$

We see that for a non-zero gradient $I_x(x_0)$, stereo has a minimum at $\delta = 0$, enabling accurate reconstruction. However, if the 3rd order gradient $I_{xxx}(x_0)$ is zero, the focus response will be approximately constant as δ varies. This completes the proof. The behavior of stereo and focus for a the special case of a constant gradient texture $I(x)$ is shown in Fig. 1. A weaker result for single lens apertures was proved in [6], they showed that shape from focus requires textures with nonzero *2nd-order* gradients.

In the analysis above, we have ignored occlusions. How would the response of stereo and focus change if the surface we are trying to reconstruct is occluded in some views? In general, the response would depend on the nature of the occluder. For shape from focus, the mean image will have a blurred image of the occluder superimposed on it at the correct depth hypothesis. If the aperture is wide, the blurred image should not contribute to the spatial derivatives. However, the focus response will be attenuated by the cameras that do not observe the surface being reconstructed. For shape from stereo, the variance will not approach zero at the correct depth and may not attain a minimum there unless the occluder is of a fairly uniform color. In Section 4, we show experimental comparisons of stereo and focus for reconstructing occluded surfaces.

3. Median and Entropy

One problem with stereo and focus is that they assign equal importance to all rays through the 3D point (x, y, d) in constructing the DSI. When reconstructing occluded surfaces, many of these rays will actually hit occluders and should therefore be considered as outliers. In this section, we introduce two variants of stereo that try to mitigate the effects of outliers due to occlusions on depth and color reconstruction.

3.1. Shape from median

This approach is inspired by the observation that amongst measures of central tendency, the median is more robust to outliers than the mean. Consider the case of grayscale images. Suppose that a surface point (x, y, d) corresponds to pixels $S = \{I_{i,d}(x, y) \ : \ 1 \leq i \leq N\}$ in the N input images. If the point is occluded in some images, the median $I_M = Median\ S$ will be a better estimator for the surface intensity than the mean. This is precisely why median colors have been used in matte extraction [18]. However, in addition to estimating the surface intensity we need a cost function to quantify the accuracy of the depth estimate. The cost function we use for constructing the DSI is the median distance of all the rays from I_M, i.e.,

$$D(x, y, d) = Median\{|I_{i,d}(x, y) - I_M| \ : \ 1 \leq i \leq N\}$$

When a surface point is visible in more than half the cameras, the median color at the correct depth should cor-

respond to the color of the occluded surface. As the number of occluded cameras increases over 50%, the estimate of the median will begin to break down. An interesting question is how to extend shape from median to color images. There are many ways to generalize the notion of median to higher dimensions [1]. We have experimented with the component-wise median and the L1 median, and observed they yield similar results. We prefer to use the component-wise median, since it can be computed in $O(N)$ time.

3.2. Shape from entropy

Let us consider the distribution of intensities of the rays through the point (x, y, d), which may be visualized by plotting a histogram. If the depth hypothesis d is incorrect, these rays will hit different points on the foreground occluders and background surface. Given textured surfaces, we would expect the histogram to be spread over a wide range of intensities. If the depth hypothesis d is correct, rays hitting the background surface will be clustered over a small range of intensities, while those hitting the occluder will still be spread out resulting in a more peaked histogram. This suggests that we can use the Shannon entropy of the intensity histogram as a cost function for constructing the DSI. The entropy is maximized for a uniform distribution and decreases as the intensities are clustered together.

For 8-bit grayscale images, we divide the intensity range into $K = 16$ bins. If the number of rays in bin i is b_i, the entropy is given by

$$H = -\sum_{i=1}^{K} \frac{b_i}{N} \log \frac{b_i}{N}.$$

(For color images, we use entropy of a color histogram where bins correspond to cubes in RGB space). The entropy cost function penalizes rays which fall into different bins. Unlike stereo, the penalty *does not increase with the distance between different ray colors*. This helps mitigate the influence of outliers on depth reconstruction.

There are two interesting properties of shape from entropy we would like to mention. First, given N rays, the entropy can vary in increments of $\frac{1}{N}$ from 0 to $\log N$. Thus, the precision of entropy is $O(\frac{1}{N})$, making it more precise as the number of views increases. Secondly, since we are binning the rays, entropy does not use the lower bits of the intensity values (for $K = 16$ bins, we use only 4 bits per channel to compute the bin index for a pixel). This could be exploited for saving storage or bandwidth, and also makes entropy less sensitive to color miscalibration.

Both entropy and shape from median retain the desirable property of shape from stereo of being sensitive to first-order texture gradients. Both of these start to fail when the occluder is of a fairly uniform color and the amount of occlusion exceeds 50% - in this case, the reconstructed color is the color of the occluder rather than the background surface.

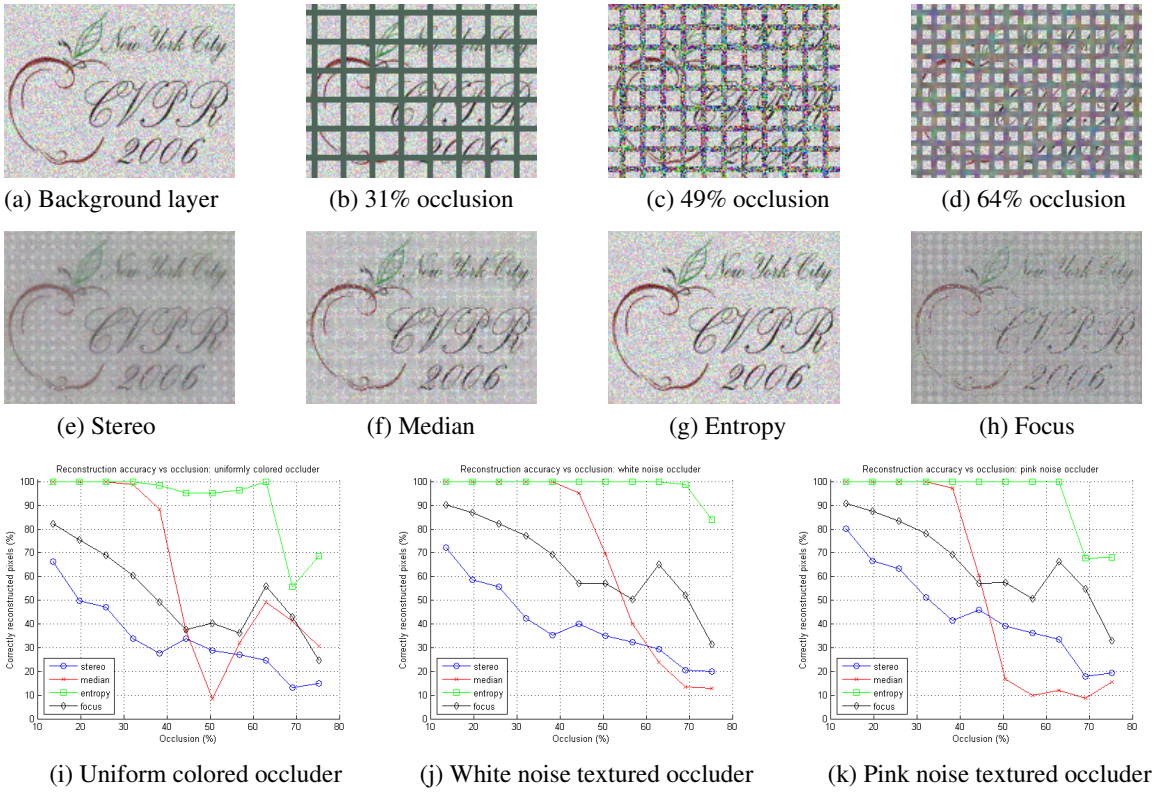
4. Experiments and Results

We have compared the performance of the four cost functions—stereo (variance), focus, median and entropy on reconstructing occluded objects in two acquired light fields of complicated scenes and one synthetic scene. As we are avoiding explicit reconstruction of the occluders, we limit our search to a range of depths *behind* the occluder. The accuracy of the depth maps computed by each method is compared with ground truth. To visualize the reconstructed appearance of the occluded surfaces, we construct *winner color images* obtained by computing the color for each point on the reconstructed surface, and projecting the surface (with color) into the central camera. In shape from focus and stereo, the mean color of rays through a surface point is used as a color estimate. The winner color image is just a synthetically focused image on a focal surface corresponding to the computed depth map. In shape from median, we use the median color rather than the mean. For entropy, we use the modal color from the color histogram.

4.1. Experiment 1: Synthetic scene

To measure the performance of our cost functions (stereo, focus, median, entropy) under varying amounts of occlusion, we experimented with a synthetic scene with precisely controlled amounts of occlusion. The scene consists of two planes. The background plane (Fig. 2a) has the CVPR logo composited with white noise to ensure there is sufficient texture at each point. The foreground occluder is composed of horizontal and vertical bars. We control the amount of occlusion (between 19% and 75%) by varying the spacing between the bars. We have experimented with three different textures on the foreground bars: white noise (strong texture), pink noise (weak texture, obtained by convolving white noise with a 5x5 box filter) and uniformly colored bars (Fig. 2 b-d).

For reconstruction, we used 81 views of this scene. The camera positions were on 9x9 planar grid perturbed by random offsets. (This is to avoid producing a strong correlation between visibility and camera positions, which rarely arises in practise and would introduce addition errors in reconstruction, contaminating our simulation). The two planes were separated by a disparity of 40 pixels across the synthetic aperture. We measured how well each of the four



(a) Background layer (b) 31% occlusion (c) 49% occlusion (d) 64% occlusion

(e) Stereo (f) Median (g) Entropy (h) Focus

(i) Uniform colored occluder (j) White noise textured occluder (k) Pink noise textured occluder

Figure 2. A synthetic two-plane scene used for performance evaluation of our four cost functions for reconstructing occluded surfaces. (a) Unoccluded view of background plane. (b-d) The background plane behind occluding bars. We experimented with the three different textures on the bars as shown. By varying the spacing between the bars, we can vary the amount of occlusion. (e)-(h) Reconstruction of the background plane (winner color images) using stereo, median, entropy and focus respectively for the setup in (c). 81 views of the scene were used. (i-k) Plot of reconstruction accuracy (percentage of pixels reconstructed correctly) vs. the amount of occlusion for the four cost functions when the occluder texture is white noise.

cost functions reconstructed the background plane.

Over the range of occlusion densities and foreground textures in our experiments, entropy does best, with near-perfect reconstructions up to 65% occlusion. Shape from median starts to fail as the occlusion density crosses 50%. For all three foreground textures, the fraction of background points correctly reconstructed by focus is about 15% higher than that for stereo. This suggests that given an adequately textured surface, focus does better than stereo in the presence of occlusions. Note that real scenes could have surface textures which stereo can reconstruct but not focus (Theorem 1); in which case stereo would obviously do better. Plots of reconstruction accuracy versus amount of occlusion for the three occluder textures are shown in Fig. 2 (i-k). The entire set of scenes and the reconstructions obtained are available on our website [11].

4.2. Experiment 2: CD case behind plants

This light field was acquired using a single camera on computer controlled gantry. The camera positions span a 21×5 grid (synthetic aperture size 60cm by 10cm). Our goal was to reconstruct a picture of the Strauss CD case behind the plants (Fig. 3 a). For ground truth, the depth of the CD case was estimated manually. To estimate the amount of occlusion for the CD case, we captured an identical light field of the scene with the occluding plants removed. a pixel in the occlusion map stores the number of views in which the corresponding pixel on the CD is occluded (Fig. 3 b). Given ground truth and an occlusion value, we can determine for each of the four cost functions the percentage of pixels reconstructed correctly (within one disparity level) for different amounts of occlusion (Fig. 3 c). The histogram indicates that stereo, median, and entropy start to perform poorly as the amount of occlusion increases, with entropy performing best. The median-based cost function starts to fail once the amount of occlusion exceeds 50%. Depth from focus does not do well on the whole, but it overtakes the stereo-based approaches as occlusion increases beyond 60%.

The winner color images from each cost function are

shown in Fig. 4 (top row). The text on the CD case is most legible for entropy and median. The bottom row shows the reconstruction of the CD obtained by voxel coloring after reconstructing and deleting the occluding plants. Voxel coloring uses a threshold to commit to foreground occluders so that pixels corresponding to the occluder are deleted prior to reconstructing the background. Note that there is no single threshold for which the CD is reconstructed properly. At low thresholds, all foreground pixels are not deleted, at higher thresholds, many background pixels are reconstructed as foreground and deleted. The reconstructed depth maps for the complete scene are available at [11].

4.3. Experiment 3: Dense Foliage

This light field (Fig. 5 a) was captured using an array of 88 cameras. The scene consists of a person and a statue behind a dense ivy wall. We computed a per-camera matte for the occluder using standard blue screen techniques. While we do not have ground truth for this scene, we can estimate visibility near the true depths using the occluder mattes. The objects behind the ivy are occluded in about 70% of the cameras. Entropy, median and voxel coloring all failed to reconstruct any part of the occluded objects accurately. This is due to the high degree of occlusion, and the relatively uniform color of the occluder. Focus does somewhat better than stereo — it is able to reconstruct the statue's torch and the person's face where stereo failed (Fig. 5 b,c).

5. Conclusions and Future Work

In this paper, we have studied cost functions for reconstructing occluded surfaces using synthetic apertures. Most existing algorithms use cost functions developed for reconstructing the (unoccluded) foreground. We have studied an alternative approach: designing cost functions that leverage a large synthetic aperture and large number of views to be robust to occlusions. This is useful when the foreground is difficult to reconstruct accurately (see, for example, Fig. 5) and in applications like surveillance of crowded areas where rreconstructing the appearance of occluded objects may be of primary interest.

While the cost functions we have studied show encouraging results, we believe we have merely scratched the surface of this design space. We would like to explore ways to combine information from different cost functions to gain better results. Our approach could also be combined with existing multi-layer estimation algorithms to improve reconstruction of *all* layers. One limitation of our work is that we have not enforced any spatial regularization. We would like to find ways of extending algorithms based on graph cuts or belief propagation to handle outliers due to dense occlusion. While our focus in this work has been on diffuse surfaces, we believe robust cost functions like median and entropy may also be useful for the reconstruction of non-Lambertian surfaces.

Finally, we would like to develop a theoretical model that explains the performance limits of synthetic aperture reconstruction algorithms. Such a model should incorporate the size, fill factor and texture of the occluders, the texture on the occluded surfaces to be reconstructed, and the size and number of views of the synthetic aperture.

Acknowledgements. We would like to thank Augusto Román, Bennett Wilburn and Eino-ville Talvala for assistance in acquisitions, and the reviewers for their suggestions. This work was begun during first author's internship at Microsoft Research, Redmond, and later supported in part by a gift from Bosch Research, grants NSF IIS-0219856-001 and DARPA NBCH 1030009.

References

[1] Greg Aloupis. On Computing Geometric Estimators of Location. *Master's thesis*, School of Computer Science, McGill University, 2001. 4

[2] S. Baker, R. Szeliski, P. Anandan. A Layered Approach to Stereo Reconstruction. In Proc. *CVPR*, 1998, 434-441. 2

[3] R. T. Collins. A space-sweep approach to true multi-image matching. In Proc. *CVPR*, 1996, 358-363. 2

[4] J. S. de Bonet and P. Viola. Roxels: Responsibility weighted 3D volume reconstruction. In Proc. *ICCV*, 1999, 418-425. 2

[5] P. Favaro and S. Soatto. Seeing beyond occlusions (and other marvels of a finite lens aperture). In Proc. *CVPR*, 2003, 579-586. 2

[6] P. Favaro, A. Mennucci and S. Soatto. Observing shape from defocused images. In *IJCV*: 52(1), 25-43, 2003. 3

[7] A. Isaksen, L. McMillan and S. Gortler. Dynamically Reparametrized Light Fields. In Proc. *SIGGRAPH*, 2000, 297-306. 1

[8] S. B. Kang, R. Szeliski, J. Chai. Handling occlusions in dense multi-view stereo. In Proc. *CVPR*, 2001, 103-110. 2

[9] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In Proc. *ECCV* 2002, 82-96. 1

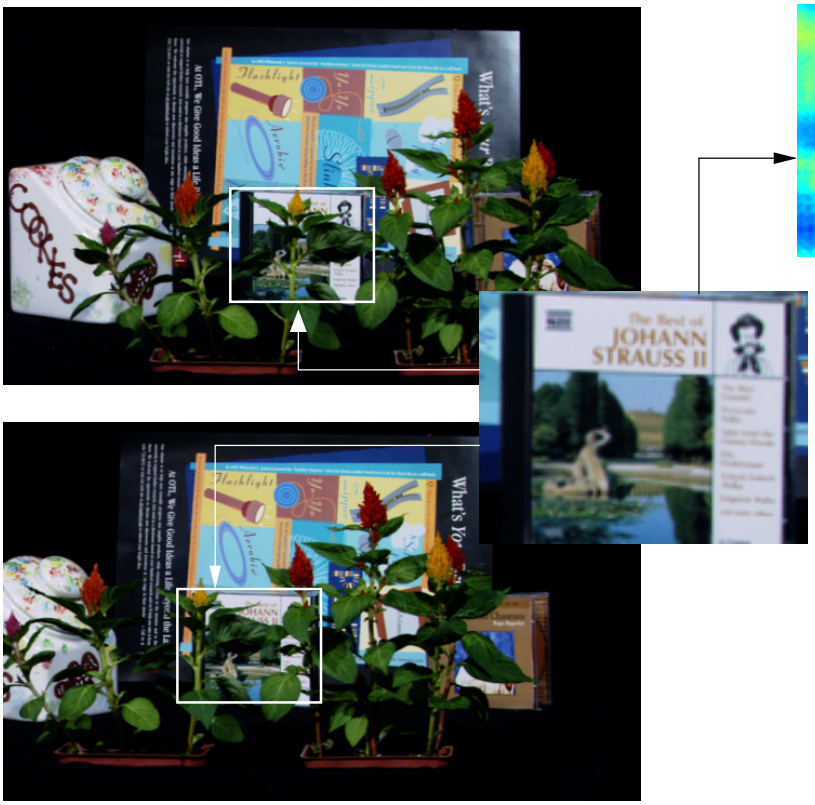
[10] M. Okutomi, T. Kanade. A multiple-baseline stereo. In *IEEE Trans. on PAMI*, 15(4):353-363, 1993. 1, 2

[11] Reconstructing Occluded Surfaces using Synthetic Apertures. Project website, *http://graphics.stanford.edu/papers/sap-recons/*. 5, 6

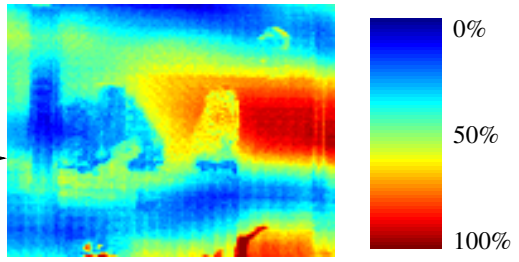
[12] Y. Schechner, N. Kiryati. Depth from defocus vs. stereo: How different really are they? In *IJCV*, 39(2):141-162. 2

[13] S. Seitz, C. Dyer. Photorealistic scene reconstruction by Voxel Coloring. In Proc. *CVPR*, 1997, 1067-73. 2

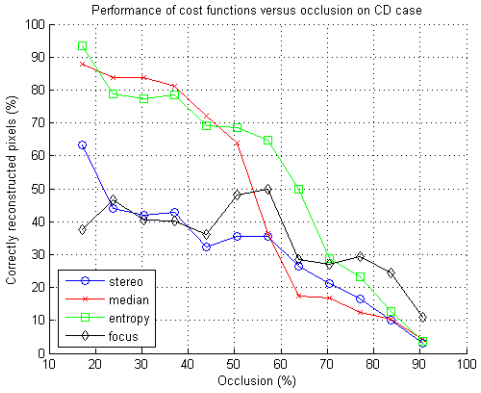
[14] Stereo Matching with Transparency and Matting. R. Szeliski, P. Golland. In Proc. *ICCV*, 1998, 517-524. 2



(a) Two images from the light field. We wish to reconstruct the Strauss CD case behind the plants (inset)



(b) Occlusion map on the CD case, showing the percentage of views each point is occluded in



(c) Performance of the four cost functions on the the Strauss CD case.

Figure 3. Experiment 2: Comparison of cost functions for reconstruction of the CD case behind the plants using a 105 image light field.

[15] V. Vaish, B. Wilburn, N. Joshi, M. Levoy. Using plane + parallax to calibrate dense camera arrays. In Proc. *CVPR*, 2004, 2-9. 1

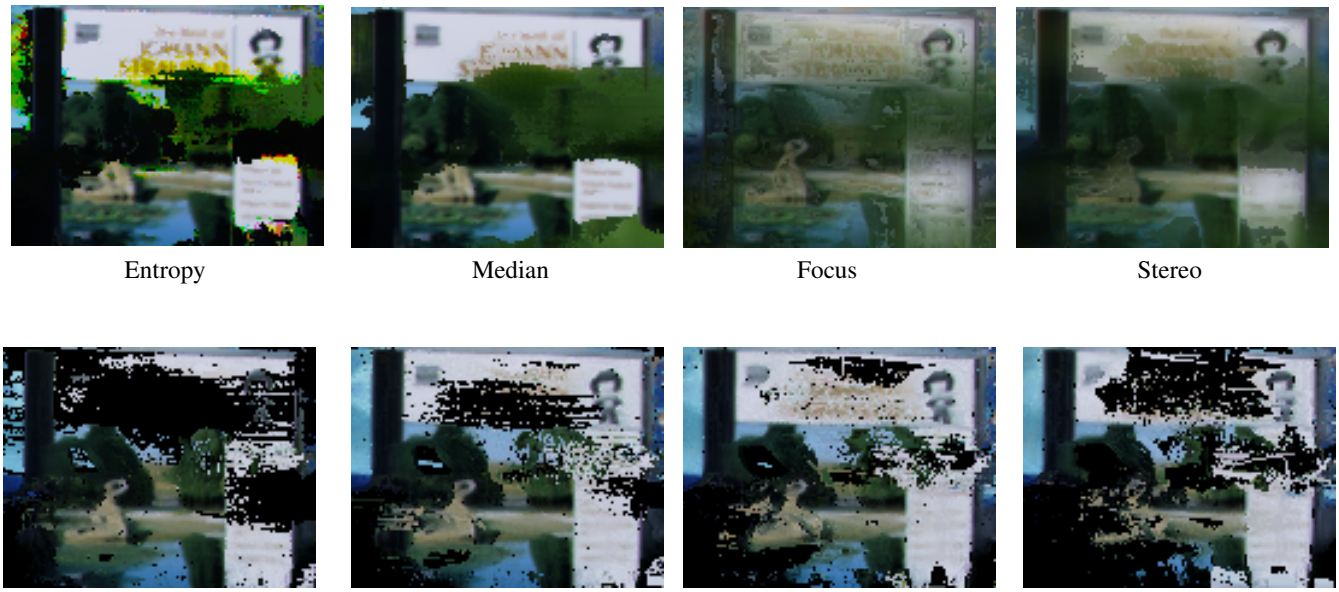
[16] V. Vaish, G. Garg, E. Talvala, E. Antunez, B. Wilburn. M. Horowitz, M. Levoy. Synthetic Aperture Focusing using a Shear-Warp Factorization of the Viewing Transform. In *Proc. of IEEE Workshop on A3DISS, CVPR,* 2005. 1

[17] J. Wang, E. Adelson. Layered Representation for Motion Analysis. In Proc. *CVPR*, 1993, 361-366. 2

[18] Y. Wexler, A. Fitzgibbon, A. Zisserman. Bayesian estimation of layers from multiple images. In Proc. *ECCV*, 2002, 487-501. 3

[19] B. Wilburn, N. Joshi, V. Vaish, E. Talvala, E. Antunez, A. Barth, A. Adams, M. Levoy, M. Horowitz. High Performance Imaging Using Large Camera Arrays. In Proc. *SIGGRAPH*, 2005, 765-776. 1

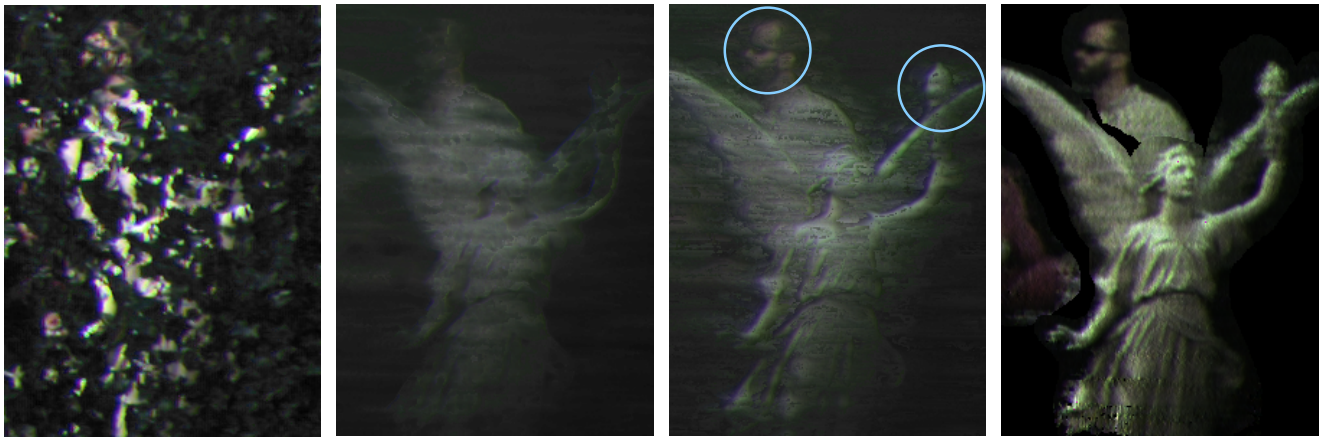
[20] C. L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High-quality video view interpolation using a layered representation. In Proc. *SIGGRAPH*, 2004, 600-608. 2



Entropy Median Focus Stereo

Voxel coloring (threshold increases left to right)

Figure 4. Experiment 2: Winner color images for reconstructing the Strauss CD case behind plants (see Fig. 3). The top row shows the reconstructed surface and color using entropy, median, focus and stereo. The bottom row shows the reconstruction using voxel coloring, these images are obtained by deleting all voxels up to the depth of the foreground occluders, and projecting the rest into the reference camera. For low thresholds, some points on the CD surface are not reconstructed at all. At high thresholds, the computed depth for points on the CD case can be much closer to the cameras than the true depth, causing them to be deleted with the occluders. These two effects cause the holes visible in the reconstruction.



(a) Image from one camera (b) Stereo reconstruction (c) Focus reconstruction (d) Unnoccluded objects

Figure 5. Experiment 3: Reconstruction of a person and statue behind a dense wall of artifical ivy. The high occlusion (about 70%) and uniform color of the occluder makes this a failure case for median and entropy. Focus does somewhat better than stereo on the right part of the statue and the person's face (shown in blue circles). Image (d) was created by manually compositing layers after deleting the occluded pixels using blue screening.