

# Evolving Measurement Regions for Depth from Defocus

Scott McCloskey, Michael Langer, and Kaleem Siddiqi

Centre for Intelligent Machines, McGill University  
{scott, langer, siddiqi}@cim.mcgill.ca

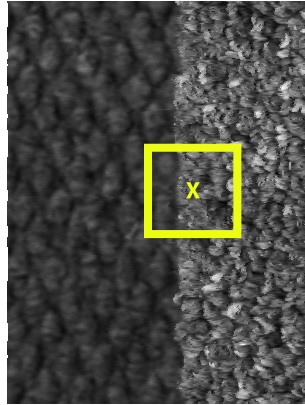
**Abstract.** Depth from defocus (DFD) is a 3D recovery method based on estimating the amount of defocus induced by finite lens apertures. Given two images with different camera settings, the problem is to measure the resulting differences in defocus across the image, and to estimate a depth based on these blur differences. Most methods assume that the scene depth map is locally smooth, and this leads to inaccurate depth estimates near discontinuities. In this paper, we propose a novel DFD method that avoids smoothing over discontinuities by iteratively modifying an elliptical image region over which defocus is estimated. Our method can be used to complement any depth from defocus method based on spatial domain measurements. In particular, this method improves the DFD accuracy near discontinuities in depth or surface orientation.

## 1 Introduction

The recovery of the 3D structure of a scene from 2D images has long been a fundamental goal of computer vision. A plethora of methods, based on many different depth cues, have been presented in the literature. Depth from defocus methods belong to class of depth estimation schemes that use optical blur as a cue to recover the 3D scene structure. Given a small number of images taken with different camera settings, depth can be found by measuring the resulting change in blur. In light of this well-known relationship, we use the terms 'depth' and 'change in blur' interchangeably in this paper.

Ideally, in order to recover the 3D structure of complicated scenes, the depth at each pixel location would be computed independently of neighboring pixels. This can be achieved through measurements of focus/defocus, though such approaches require a large number of images [9] or video with active illumination [17]. Given only a small number of observations (typically two), however, the change in blur must be measured over some region in the images. The shape of the region over which these measurements are made has, to date, been ignored in the literature. Measurements for a given pixel have typically been made over square regions centered on its location, leading to artificially smoothed depth estimates near discontinuities.

As a motivating example, consider the image in Fig. 1 of a scene with two fronto-parallel planes at different distances, separated by a depth discontinuity. Now consider an attempt to recover the depth of the point near the discontinuity



**Fig. 1.** Two fronto-parallel surfaces separated by a step edge in depth. When estimating depth at a point near the step edge, inaccuracies arise when measurements are made over a square window.

(such as the point marked with the X) by estimating blur over a square region centered there (as outlined in the figure). We would like to recover the depth of the point, which is on one plane, but this region consists of points that lie on both surfaces. Inevitably, when trying to recover a single depth value from a region that contains points at two different depths, we will get an estimate that is between the distances of the two planes. Generally, the recovered depth is an average of the two depths weighted by the number of points in the measurement region taken from each plane. As a result, the depth map that is recovered by measuring blur over square windows will show a smooth change in depth across the edge where there should be a discontinuity. (See red curve in Fig. 3.)

The extent of the unwanted smoothing depends on the size of the measurement region. In order to produce accurate estimates, DFD methods have traditionally used relatively large measurement regions. The method of [5] uses 64-by-64 pixel regions, and the method of [10] uses 51-by-51 regions. For the scene in Fig. 1, this results in a gradual transition of depth estimates over a band of more than 50 pixels where there should be a step change.

This example illustrates a failure of the standard *equifocal assumption*, which requires that all points within the measurement region be at the same depth in the scene. In order for the equifocal assumption to hold over square regions, the scene must have constant depth over that region. In the case of a depth discontinuity, as illustrated above, the violation of the equifocal assumption results in smoothed depth estimates.

In this paper, we propose a method for measuring blur over *elliptical regions* centered at each point, rather than over square regions at each point as in standard DFD [5,10,14,15]. Each elliptical region is evolved to minimize the depth variation within it (Sec. 3.1). Depths estimated over these regions (Sec. 3.2) are more accurate, particularly near discontinuities in depth and surface orientation.

Given the semantic importance of discontinuities, this improvement is important for subsequent 3D scene recovery and modeling. We demonstrate the improved accuracy of our method in several experiments with real images (Sec. 4).

## 2 Previous Work in Depth from Defocus

Broadly speaking, DFD methods fall into one of two categories: deconvolution methods based on a linear systems model, and energy minimization methods.

**Deconvolution Methods.** Pentland [12] proposed the first DFD method based on frequency domain measurements. Given one image taken through a pinhole aperture, the blur in a second (finite aperture) image is found by comparing the images' power spectra. Subbarao [15] generalized this result, allowing for changes in other camera settings, and removing the requirement that one image be taken through a pinhole aperture. Gökstorp [8] uses a set of Gabor filters to generate local frequency representations of the images. Other methods based on the frequency domain model such as by Nayar and Watanabe [11] develop filter banks which are used to measure blur (see also [16]).

Ens and Lawrence point out a number of problems with frequency domain measurements and propose a matrix-based spatial domain method that avoids them [5]. Subbarao and Surya [14] model image intensities as a polynomial and introduce the S-transform to measure the blur in the spatial domain. More recently, McCloskey *et al* use a reverse projection model, and measure blur using correlation in the spatial domain [10].

Whether they employ spatial or frequency domain schemes, square measurement regions are the *de facto* standard for DFD methods based on deconvolution. Frequency domain methods use the pixels within a square region to compute the Fast Fourier Transform and observe changes in the local power spectrum. Spatial domain methods must measure blur over some area, and take square regions as the set of nearby points over which depth is assumed to be constant. In either case, estimates made over square windows will smooth over discontinuities, either in depth or surface orientation. Despite this shortcoming, deconvolution has been a popular approach to DFD because of its elegance and accessibility.

**Energy Minimization Methods.** Several iterative DFD methods have been developed which recover *both* a surface (depth) and its radiance from pairs of defocused images. The methods presented by Favaro and Soatto [6,7] define an energy functional which is jointly minimized with respect to both shape and radiance. In [6], four blurred images are used to recover the depth and radiance, including partially occluded background regions. In [7] a regularized solution is sought and developed using Hilbert space techniques and SVD. Chaudhuri and Rajagopalan model depth and radiance as Markov Random Fields (MRF) and find a maximum *a posteriori* estimate [4]. These methods have the advantage that they don't explicitly assume the scene to have locally constant depth, though regularization and MRF models implicitly assume that depth changes slowly.

Building upon the MRF scheme with line fields, Bhasin and Chaudhuri [2] explicitly model image formation in the neighborhood of depth discontinuities in order to more accurately recover depth. However, their results are generated on synthetic data that is rendered with an incorrect model: they assert that (regardless of shadowing) a dark band arises near depth discontinuities as a result of partial occlusion. For a correct model of partial occlusion and image formation near depth discontinuities, see Asada *et al* [1].

### 3 DFD with Evolving Regions

The method we propose in this paper combines elements of both of the above categories of DFD methods. Our approach iteratively refines a depth estimate like the energy minimization methods, but it does not involve a large error function of unknown topography. Our method is conceptually and computationally straightforward like the deconvolution methods, but it will be shown to have better accuracy near discontinuities in depth and in surface orientation. To achieve this, we evolve a measurement region at each pixel location toward an equifocal configuration, using intermediate depth estimates to guide the evolution.

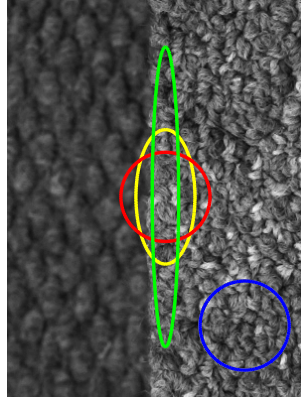
In order to vary the shape of the measurement region in a controlled fashion despite erroneous depth estimates, we impose an elliptical shape on it. The measurement region at a given location starts as a circular region centered there, and evolves with increasing eccentricity while maintaining a fixed area<sup>1</sup>  $M$ . In addition to providing a controlled evolution, elliptical regions are more general than square one in that they can provide better approximations to the scene's equifocal contours<sup>2</sup>. Furthermore, ellipses can be represented with a small number of parameters. This compactness of representation is important in light of our desire to produce dense depth maps, which requires us to maintain a separate measurement region for each pixel location in an image.

Given only two images of the scene under different focus settings as input, we have no information about depth or the location of equifocal regions. Initially, we make the equifocal assumption that the scene has locally constant depth and estimate depth over hypothesized equifocal regions (circles) based on this assumption. The key to our approach is that *this initial depth estimate is then used to update the measurement region at each pixel location*. Instead of estimating defocus over square regions that we *assume* to have constant depth, we evolve elliptical regions over which we *have evidence* that depth is nearly constant.

In order to give the reader an intuition for the process, we refer to Fig. 2, which shows the evolution of the measurement regions at two locations. In regions of the scene that have constant depth, initial measurement regions are found to contain no depth variation, and are maintained (blue circle). In regions of changing depth, as near the occluding contour, the initial region (red circle) is found to

<sup>1</sup> Maintaining a measurement region with at least  $M$  pixels is necessary to ensure a consistent level of performance in the DFD estimation.

<sup>2</sup> In regions of locally constant depth, equifocal (iso-depth) points form a 2D region. More generally, the surface slopes and equifocal points fall along a 1D contour.



**Fig. 2.** Evolving measurement regions. Near the depth discontinuity, regions grow from circles (red) through a middle stage (yellow) to an equifocal configuration (green). In areas of constant depth, initial regions (blue) do not evolve.

contain significant depth variation and is evolved through an intermediate state (yellow ellipse) to an equifocal configuration (green ellipse).

Our method can be thought of as a loop containing two high-level operations: evolution of elliptical regions and depth measurement. The method of evolving elliptical regions toward an equifocal configurations is presented in Sec. 3.1. The blur estimation method is presented in Sec. 3.2. We use the Ens and Lawrence blur estimation algorithm of [5], and expect that any DFD method based on spatial domain measurements could be used in its place.

### 3.1 Evolving Elliptical Regions

Given a circle<sup>3</sup> as our initial measurement region's perimeter, we wish to evolve that region in accordance with depth estimates. Generally, we would like the extent of the region to grow in the direction of constant scene depth while shrinking in the direction of depth variation. Using an elliptical model for the region, we would like the major axis to follow depth contours and the minor axis to be oriented in the direction of depth variation. We iteratively increase the ellipse's eccentricity in order to effect the expansion and contraction along these dimensions, provided there are depth changes within it.

Each pixel location  $p$  has its own measurement region  $\mathcal{R}_p$ , an ellipse represented by a 2-vector  $\vec{f}_p$  which expresses the location of one of the foci relative to  $p$ , and a scalar  $r_p$  which is the length of the semi-major axis. Initially  $\vec{f} = (0, 0)$  and  $r = r_c$  for all  $p$ . The value  $r_c$  comes from the region area parameter  $M = \pi r_c^2$ .

Once we have measured depth over circular regions, we establish the orientation of the ellipse by finding the direction of depth variation in the smoothed depth map. In many cases, though not all, the direction of depth variation

<sup>3</sup> An ellipse with eccentricity 0.

is the same as the direction of the depth gradient. Interesting exceptions happen along contours of local minima or maxima of depth (as in the image in Fig. 4 (left)). In order to account for such configurations, we take the angle  $\theta_v$  to be the direction of depth variation if the variance of depth values along diameters of the circular region is maximal in that direction. The set of depth values  $\mathcal{D}(\theta)$  along the diameter in direction  $\theta$  is interpolated from the depth map  $d$  at equally-spaced points about  $p$ ,

$$\mathcal{D}(\theta) = \{d(p + n(\cos \theta, \sin \theta)) | n = -r_c, -r_c + 1, \dots, r_c\}. \tag{1}$$

We calculate  $\mathcal{D}(\theta)$  over a set of orientations  $\theta \in [0, \pi)$ , and take  $\theta_v$  to be the direction that maximizes the depth variance, i.e.

$$\theta_v = \operatorname{argmax}_{\theta} \operatorname{var}(\mathcal{D}(\theta)). \tag{2}$$

The variance of  $\mathcal{D}(\theta_v)$  is compared to a threshold chosen to determine if the variance is significant. In the event that the scene has locally constant depth within the circle, the variance should be below threshold, and the measurement region does not evolve. In the event that the variance is above the threshold, we orient the minor axis in the direction  $\theta_v$  by setting  $\vec{f} = (-\sin \theta_v, \cos \theta_v)$ .

Having established the orientation of the ellipses, we estimate depth over the current measurement regions and increase their eccentricity if the depth variation along the minor axis is significant (i.e. above threshold) and the variation along the major axis is insignificant. These checks halt the evolution if the elliptical region begins to expand into a region of changing depth or if it is found to be an equifocal configuration. If these checks are passed, the ellipse at iteration  $n + 1$  is evolved from the ellipse at iteration  $n$  by increasing  $\|\vec{f}\|$  according to

$$\vec{f}_{n+1} = (\|\vec{f}_n\| + k) \frac{\vec{f}_n}{\|\vec{f}_n\|}, \tag{3}$$

where the scalar  $k$  represents the speed of the deformation. As necessary, the value of  $r$  is suitably adjusted in order to maintain a constant area despite changes in  $\vec{f}$ . Though the accuracy of depth estimation may increase by measuring blur over regions of area greater than  $M$ , we keep the area constant for our experiments, allowing us to demonstrate the improvement in depth estimates due only to changes in the shape of the measurement region.

### 3.2 Measuring Depth Over Evolving Regions

In order to be utilized in our algorithm, a depth estimation method must be able to measure blur over regions of arbitrary shape. This rules out frequency domain methods that require square regions over which a Fast Fourier Transform can be calculated. Instead, we use a spatial domain method, namely the DFD method of Ens and Lawrence [5], adapted to our elliptical regions. The choice of this particular algorithm is not integral to the method; we expect that any spatial domain method for blur measurement could be used.

The Ens and Lawrence method takes two images as input:  $i_1$  taken through a relatively small aperture, and  $i_2$  taken through a larger aperture. The integration time of the image taken through the larger aperture is reduced to keep the level of exposure constant. As is common in the DFD literature, we model each of these images to be the convolution of a theoretical pinhole aperture image  $i_0$  (which is sharply focused everywhere) with a point spread function (PSF)  $h$ ,

$$i_1 = i_0 * h_1 \quad \text{and} \quad i_2 = i_0 * h_2. \quad (4)$$

The PSFs  $h_1$  and  $h_2$  belong to a family of functions parameterized by the spread  $\sigma$ . This family of PSFs is generally taken to be either the pillbox, which is consistent with a geometric model of optics, or Gaussian, which accounts for diffraction in the optical system. We take the PSFs to be a Gaussian parameterized by its standard deviation. That is,  $h_n = G(\sigma_n)$ .

Since it was taken through a larger aperture, regions of the image  $i_2$  cannot be sharper than corresponding regions in  $i_1$ . Quantitatively,  $\sigma_1 \leq \sigma_2$ . As described in [5], depth recovery can be achieved by finding a third PSF  $h_3$  such that

$$i_1 * h_3 = i_2. \quad (5)$$

As we have chosen a Gaussian model for  $h_1$  and  $h_2$ , this unknown PSF is also a Gaussian;  $h_3 = G(\sigma_3)$ . Computationally, we take  $\sigma_3$  at pixel location  $p$  to be the value that minimizes, in the sum of squared errors sense, the difference between  $i_2$  and  $i_1 * G(\sigma)$  over  $p$ 's measurement region  $\mathcal{R}_p$ . That is,

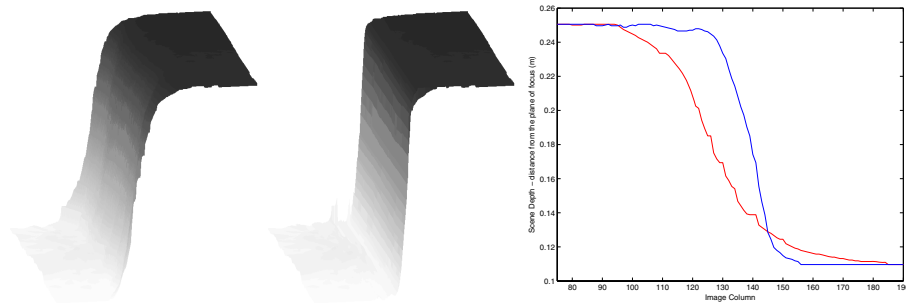
$$\sigma_3(p) = \underset{\sigma}{\operatorname{argmin}} \sum_{q \in \mathcal{R}_p} (i_2(q) - (i_1 * G(\sigma))(q))^2. \quad (6)$$

The value of  $\sigma_3$  can be converted to the depth when the camera parameters are known. We omit the details for brevity.

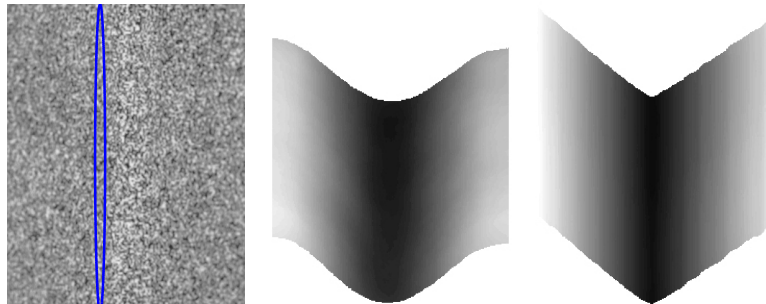
## 4 Experiments

We have conducted a number of experiments on real scenes. Our images were all acquired with a Nikon D70s digital camera and a 50mm Nikkor lens. The camera's tone mapping function was characterized and inverted by the method described in [3], effectively making the camera a linear sensor. For each scene, the lens was focused at a distance nearer than the closest object in the scene<sup>4</sup>.

**Depth Discontinuities.** Our first experiment involves the scene considered in Sec. 1, which consists of two fronto-parallel planes separated by a depth discontinuity. We have shown the sharply focused, small aperture ( $f/13$ ) input image in Figs. 1 and 2; the second, blurrier input image (not shown) was taken with a larger aperture ( $f/4.8$ ). We have worked on a relatively small (400-by-300 pixel) window of 1000-by-1500 pixel images in order to illustrate the improved accuracy in the area of the discontinuity in sufficient detail.



**Fig. 3.** (Left) Surface plot of estimates from circular regions (dark points are closer to the camera). (Center) Estimates from computed elliptical regions. (Right) Profile of depth discontinuity as measured over disks (red), and evolved elliptical regions (blue).



**Fig. 4.** (Left) Image of two planes meeting at a discontinuity in surface orientation, annotated with an evolved measurement region. (Center) Surface plot of depth recovered over circular regions (dark points are closer to the camera). (Right) Surface plot of depth recovered over elliptical regions.

Fig. 3 shows surface renderings of the recovered scene from the initial circular measurement regions (left), and the elliptical regions recovered by our method (middle). These surface renderings demonstrate the improved accuracy of the depth estimates as a result of the elliptical evolution. Fig. 3 (right) shows a plot of the depth estimates for a single row, recovered from the circular regions (red), and elliptical regions (green). As this plot shows, our method recovers an edge profile that is significantly more accurate than the initial result.

This result was obtained by evolving the ellipses 25 times with speed constant  $k = 10$ . The orientation  $\theta_v$  was determined by measuring the depth variance over 8 different angles uniformly spaced in the interval  $[0, \pi)$ . The value  $r_c = 45$  pixels, making the measurement regions comparable in size to those used in [5].

<sup>4</sup> Defocus increases in both direction away from the focused depth, and so depth from defocus suffers a sign ambiguity relative to the plane of focus [12]. This is typically avoided by having the entire scene be in front of or behind the plane of focus.





**Fig. 5.** (Left) Image of a tennis ball on a ground plane annotated with measurement regions. (Center) Surface plot of depth measured over circular regions. (Right) Surface plot of depth measured over ellipses evolved to approximate equifocal regions.

**Discontinuities in Surface Orientation.** As in the case of discontinuous depth changes, square (or circular) windows for the measurement of defocus produce inaccurate depth estimates near discontinuities in surface orientation. Consider, for example, the scene shown in Fig. 4 (left), which consists of two slanted planes that intersect at the middle column of the image. The scene is convex, so the middle column of the image depicts the closest points on the surface. Within a square window centered on one of the locations on this ridge line of minimum depth, most of the scene points will be at greater depths. As a result, the distance to that point will be overestimated. Similarly, were the ridge line at a maximum distance in the scene, the depth would be underestimated.

Fig. 4 (center and right) show surface plots of the scene recovered from circular and evolved elliptical regions, respectively. An example of an evolved measurement region is shown in Fig. 4 (left). The inaccuracies of the circular regions are apparent near the discontinuity in orientation, where the width of the measurement region results in a smoothed corner. The surface plot of estimates made over evolved elliptical regions shows the discontinuity more accurately, and has more accurate depth estimates near the perimeter of the image.

**Non-linear Discontinuities.** Because our ellipses evolve to a line segment in the limit, one may expect that our method will fail on objects with occluding contours that are non-linear. While an ellipse cannot describe a curved equifocal contour exactly, it can still provide an improvement over square or circular regions when the radius of the equifocal contour's curvature is small compared to the scale of the ellipse. Fig. 5 (left) shows an example of a spherical object, whose equifocal contours are concentric circles. The curvature of these equifocal contours is relatively low at the sphere's occluding contour, allowing our elliptical regions to increase in eccentricity while reducing the internal depth variation (blue ellipse). Near the center of the ball, the equifocal contours have higher curvature, but the surface is nearly fronto-parallel in these regions. As a result, the initial circles do not become as elongated in this region (red circle). This results in depth estimates that show a clearer distinction between the sphere and the background (see Fig 5 (center) and (right)).

The parameters for this experiment were the same as in the previous examples, except that  $\theta_v$  was found over 28 angles in the interval  $[0, \pi)$ . The algorithm was run for 25 iterations, though most of the ellipses stopped evolving much earlier, leaving the maximum value of  $\|\vec{f}\| = 100$ .

For the experiments shown in this paper, a substantial majority of the running time was spent in the depth estimation step. The amount of time spent orienting and evolving the elliptical regions depends primarily on the scene structure, and totals about 2 minutes in the worst case.

## 5 Conclusions and Future Work

We have demonstrated that the accuracy of DFD estimates can depend on the shape of the region over which that estimate is computed. The standard assumption in the DFD literature - that square regions are equifocal - is shown to be problematic around discontinuities in depth and surface orientation. Moreover, through our experiments, we have demonstrated that an elliptical model can be used to evolve measurement regions that produce more accurate depth estimates near such features. We have, for the first time, presented an algorithm that iteratively tailors the measurement region to the structure of the scene.

Future research could address both the size and shape of the measurement region. In order to illustrate the benefits of changes in its shape, we have kept the size  $M$  of the measurement region constant. In actual usage, however, we may choose to increase the size of the measurement area in regions of the scene that are found to be fronto-parallel in order to attain improved DFD accuracy.

Though we have shown that elliptical regions are more general than squares, and that this additional generality improves DFD performance, there are scene structures for which ellipses are not sufficiently general. Non-smooth equifocal contours, found near corners, will be poorly approximated by ellipses. Such structures demand a more general model for the measurement region, and would require a different evolution algorithm, which is an interesting direction for future work.

## References

1. Asada, N., Fujiwara, H., Matsuyama, T.: 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.* 20, 155–167 (1998)
2. Bhasin, S., Chaudhuri, S.: Depth from Defocus in Presence of Partial Self Occlusion. In: *Proc. Intl. Conf. on Comp. Vis.*, pp. 488–493 (2001)
3. Debevec, P., Malik, J.: Recovering High Dynamic Range Radiance Maps from Photographs. In: *Proc. SIGGRAPH*, pp. 369–378 (1997)
4. Chaudhuri, S., Rajagopalan, A.: *Depth from Defocus: A Real Aperture Imaging Approach*. Springer, Heidelberg (1999)
5. Ens, J., Lawrence, P.: Investigation of Methods for Determining Depth from Focus. *IEEE Trans. on Patt. Anal. and Mach. Intell.* 15(2), 97–108 (1993)

6. Favaro, P., Soatto, S.: Seeing beyond occlusions (and other marvels of a finite lens aperture). In: Proc. CVPR 2003, vol. 2, pp. 579–586 (June 2003)
7. Favaro, P., Soatto, S.: A Geometric Approach to Shape from Defocus. *IEEE Trans. on Patt. Anal. and Mach. Intell.* 27(3), 406–417 (2005)
8. Gökstorp, M.: Computing Depth from Out-of-Focus Blur Using a Local Frequency Representation. In: Proc. of the IAPR Conf. on Patt. Recog., pp. 153–158 (1994)
9. Hasinoff, S.W., Kutulakos, K.N.: Confocal Stereo. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 620–634. Springer, Heidelberg (2006)
10. McCloskey, S., Langer, M., Siddiqi, K.: The Reverse Projection Correlation Principle for Depth from Defocus. In: Proceedings of the 3rd International Symposium on 3D Data Processing, Visualization and Transmission (2006)
11. Nayar, S.K., Watanabe, M.: Minimal Operator Set for Passive Depth from Defocus. In: Proc. CVPR 1996, pp. 431–438 (June 1996)
12. Pentland, A.: A New Sense for Depth of Field. *IEEE Trans. on Patt. Anal. and Mach. Intell.* 9(4), 523–531 (1987)
13. Pentland, A., Scherock, S., Darrell, T., Girod, B.: Simple Range Cameras Based on Focal Error. *J. of the Optical Soc. Am.* 11(11), 2925–2935 (1994)
14. Subbarao, M., Surya, G.: Depth from Defocus: A Spatial Domain Approach. *Intl. J. of Comp. Vision* 13, 271–294 (1994)
15. Subbarao, M.: Parallel Depth Recovery by Changing Camera Parameters. In: Proc. Intl. Conf. on Comp. Vis., pp. 149–155 (1998)
16. Xiong, Y., Shafer, S.A.: Moment Filters for High Precision Computation of Focus and Stereo. In: Proc. Intl. Conf. on Robotics and Automation, pp. 108–113 (1995)
17. Zhang, L., Nayar, S.K.: Projection Defocus Analysis for Scene Capture and Image Display. In: Proc. SIGGRAPH, pp. 907–915 (2006)