International Journal of Computer Vision 39(2), 141–162, 2000

Depth from Defocus vs. Stereo: How Different Really Are They?

YOAV Y. SCHECHNER
Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel
yoavs@tx.technion.ac.il

NAHUM KIRYATI
Department of Electrical Engineering-Systems, Faculty of Engineering, Tel-Aviv University, Ramat Aviv 69978, Israel
nk@eng.tau.ac.il

Abstract. Depth from Focus (DFF) and Depth from Defocus (DFD) methods are theoretically unified with the geometric triangulation principle. Fundamentally, the depth sensitivities of DFF and DFD are not different than those of stereo (or motion) based systems having the same physical dimensions. Contrary to common belief, DFD does not inherently avoid the matching (correspondence) problem. Basically, DFD and DFF do not avoid the occlusion problem any more than triangulation techniques, but they are more stable in the presence of such disruptions. The fundamental advantage of DFF and DFD methods is the two-dimensionality of the aperture, allowing more robust estimation. We analyze the effect of noise in different spatial frequencies, and derive the optimal changes of the focus settings in DFD. These results elucidate the limitations of methods based on depth of field and provide a foundation for fair performance comparison between DFF/DFD and shape from stereo (or motion) algorithms.

Keywords: Defocus, depth from focus, depth of field, depth sensing, range imaging, shape from X, stereo, triangulation

1. Introduction

In recent years range imaging based on the limited depth of field (DOF) of lenses has been gaining popularity. Methods based on this principle are normally considered to be a separate class, distinguished from triangulation techniques such as depth from stereo, vergence or motion (Besl, 1988; Engelhardt and Hausler, 1988; Jarvis, 1983; Krotkov and Bajcsy, 1993; Marapane and Trivedi, 1993; Scherock, 1991; Stewart and Nair, 1989; Subbarao et al., 1997). Cooperation between depth from *focus*, stereo and vergence procedures has been studied in Abbott and Ahuja (1988, 1993), Dias et al. (1992), Kristensen et al. (1993), Krotkov and Bajcsy (1993), Stewart and Nair (1989), and Subbarao et al. (1997). Cooperation of depth from *defocus* with stereo was considered in Darwish (1994), Klarquist et al. (1995), and Subbarao et al. (1997).

Successful application of computer vision algorithms requires sound performance evaluation and comparison of the various approaches available. The comparison of range sensing systems that rely on different principles of operation and have a wide range of physical parameters is not easy (Besl, 1988; Jarvis, 1983). In particular, in such cases it is difficult to distinguish between limitations of *algorithms* to those arising from fundamental physical bounds.

The following observations and statements are common in the literature:

1. The resolution and sensitivity of Depth from Defocus (DFD) methods are limited in comparison to triangulation based techniques (Besl, 1988; Hwang et al., 1989; Pentland, 1987; Pentland et al., 1989, 1994; Rajagopalan and Chaudhuri, 1997; Stewart and Nair, 1989; Subbarao, 1988; Subbarao and Surya, 1994; Subbarao and Wei,

1992; Subbarao et al., 1997; Surya and Subbarao, 1993).
2. Unlike triangulation methods, DFD avoids the missing-parts (occlusion) problem (Bove, 1989, 1993; Ens and Lawrence, 1993; Nayar et al., 1995; Pentland et al., 1994; Saadat and Fahimi, 1995; Scherock, 1991; Subbarao and Liu, 1996; Subbarao and Surya, 1994; Subbarao and Wei, 1992; Subbarao et al., 1997; Surya and Subbarao, 1993; Watanabe and Nayar, 1996).
3. Unlike triangulation methods, DFD avoids matching (correspondence) ambiguity problems (Bove, 1989, 1993; Darwish, 1994; Ens and Lawrence, 1993; Hwang et al., 1889; Nayar et al., 1995; Pentland, 1987; Pentland et al., 1989, 1994; Rajagopalan and Chaudhuri, 1995a; Saadat and Fahimi, 1995; Scherock, 1991; Simoncelli and Farid, 1996; Subbarao and Liu, 1996; Subbarao and Surya, 1994; Subbarao and Wei, 1992; Subbarao et al., 1997; Surya and Subbarao, 1993; Swain et al., 1994; Watanabe and Nayar, 1996; Xiong and Shafer, 1993).
4. DFD is reliable (Nayar et al., 1995; Pentland, 1987; Pentland et al., 1989; Subbarao and Wei, 1992).

Similar statements were made with regard to Depth from Focus (DFF) (Abbott and Ahuja, 1993; Darrell and Wohn, 1988; Dias et al., 1992; Engelhardt and Hausler, 1988; Krotkov and Bajcsy, 1993; Marapane and Trivedi, 1993; Subbarao and Liu, 1996). There have been several attempts to explain these observations. For example, the limited sensitivity of DFD was associated with suboptimal selection of parameters (Rajagopalan and Chaudhuri, 1997), leading to interest in optimizing the changes in imaging system parameters. A major step towards understanding the relations between triangulation and DOF has been recently taken in Adelson and Wang (1992), Farid (1997), and Farid and Simoncelli (1998). A large aperture lens was utilized to build a "monocular stereo" system, with sensitivity that has the same functional dependence on parameters as in a stereo system (without vergence).

We show that the difference between methods that rely on the limited depth of field of the optical system (DFD and DFF) and "classic" triangulation techniques (stereo, vergence, motion) is mainly due to technical reasons, and is hardly a fundamental one. In fact, DFD and DFF can be regarded as ways to achieve triangulation. We study the fundamental characteristics of the above mentioned methods and the differences between them in a formal and quantitative manner. The first statement above claims superiority of stereo over DFD with regard to sensitivity. However, the origins of this observation are primarily in the physical size difference between common implementations of focus and triangulation based systems, not in the fundamentals. Generally, this statement does not hold.

As to the second and third statements (that unlike stereo, the occlusion and matching problems are avoided in DFD), they again follow mainly from physical size differences in the common implementations. As they are expressed, these two statements do not hold. Actually, we note a fundamental matching problem in DFD, analogous to the problem in stereo. There are, however, some differences between DFD, stereo, and DFF with respect to matching ambiguity and occlusion that can be expressed quantitatively.

In contrast, the fourth observation (reliability of DFD) has a solid foundation. DFF and DFD rely on more data than common discrete triangulation methods, and are thus potentially more reliable. Note that an approach and algorithm similar to DFD can also be applied in *Depth from Motion Blur* (*smear*) (Fox, 1988), leading to improved robustness. Still, unlike motion smear which is one dimensional (1D), DFF and DFD rely on two dimensional (2D) blur and thus have an important advantage.

In order to study the influence of noise on the various ranging methods considered in this paper, we analyze its effect in each spatial frequency of which the image is composed. We show that some frequencies are more useful for range estimation, while others do not make a significant or reliable contribution. Our analysis leads to a new property of depth of field: it is the optimal interval between focus-settings in depth-from-*defocus* for robustness to perturbations. We also show that in DFD, if the step used is larger by a factor of 2 or higher, the estimation process may be very unstable. We thus obtain the limits on the interval between focus settings that ensures stable operation of DFD. Some preliminary results were presented in Schechner and Kiryati (1998, 1999).

2. Sensitivity

2.1. DFD

Consider the imaging system sketched in Fig. 1. The sensor at distance $\tilde{v}$ behind the lens can image in-focus a point at distance $\tilde{u}$ in front of the lens. An object point

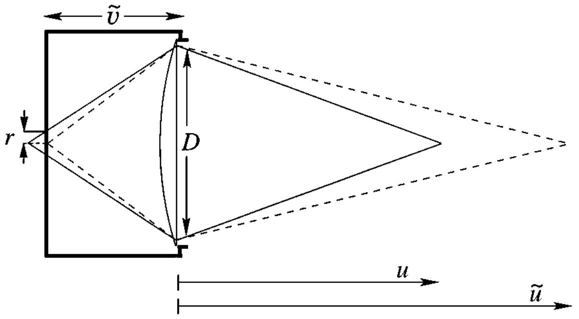

Figure 1. The imaging system with an aperture D is tuned to view in focus object points at distance $\tilde{u}$. The image of an object point at distance u is a blur circle of radius r in the sensor plane.

at distance u is defocused, and its image is a blur-circle of radius r in the sensor plane.

In this system the blur radius is (Scherock, 1991)

$$r = \frac{D}{2}\frac{|uF - \tilde{v}u + F\tilde{v}|}{Fu} \tag{1}$$

where F is the focal length and D is the aperture of the lens. For simplicity we adopt the common assumption that the system is invariant to transversal shift. This is approximately true for paraxial systems, where the angles between light rays and the optical axis are small.

Suppose now that the entire lens is blocked, except for two pinholes on its perimeter, on opposite ends of some diameter (Adelson and Wang, 1992; Hiura et al., 1998), as shown in Fig. 2. Only two rays pass the lens. The geometrical point spread function (PSF) thus consists of only two points, x_L and x_R. The distance between the points is

$$|x_R - x_L| = 2r. \tag{2}$$

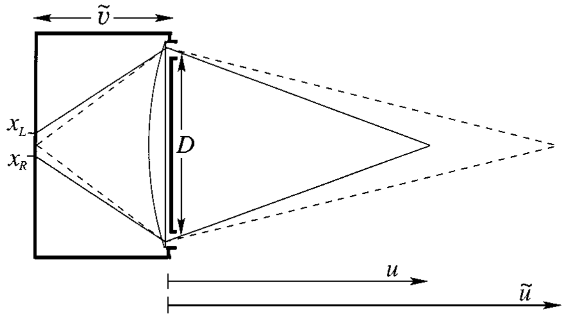

Figure 2. An imaging system similar to that of Fig. 1, with its lens blocked except for two pinholes on its perimeter, on opposite ends of some diameter. The image of an out-of-focus object point is two points, with disparity equal to the diameter of the blur circle that would have appeared had the blocking been removed.

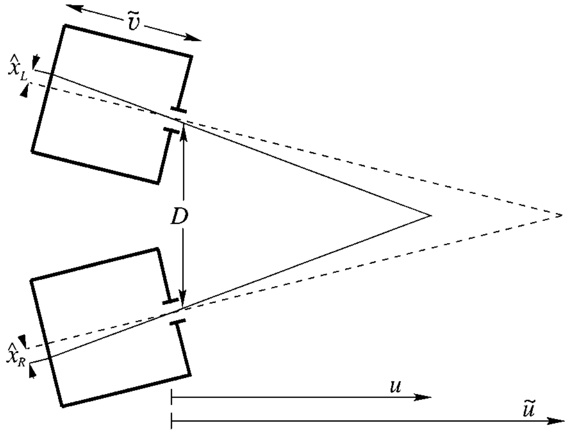

Figure 3. A stereo system with a baseline D equal to the lens diameter in Fig. 1. The distance $\tilde{v}$ from the entrance pupil to the sensor is also the same. The vergence eliminates the disparity for the object point at distance $\tilde{u}$. The resulting disparity caused by the object point at u is equal to the diameter of the blur kernel formed by the system of Fig. 1.

The fact that the image of each object point consists of two points, separated by a distance that depends on the depth of the object, gives rise to the analogy to stereo. Note that for an object point at a distance $\tilde{u}$, the image points coincide, i.e. have no disparity. To accommodate this in the analogy, we incorporate *vergence* into the stereo system. Now, consider the stereo & vergence system shown in Fig. 3 that consists of two pinhole cameras. It has *the same physical dimensions* as the system shown in Fig. 1, i.e., the baseline between the pinholes is equal to the width of the large aperture lens, and the sensors are at the same distance $\tilde{v}$ behind the pinholes. The image of an object point at u is again two points, now one on each sensor. Since the angles are small (e.g., $D \ll u$) the disparity can be well approximated by

$$d = \hat{x}_R - \hat{x}_L = D\frac{uF - \tilde{v}u + F\tilde{v}}{Fu} = Df(u). \tag{3}$$

Comparing this result to Eqs. (1) and (2) we see that

$$|\hat{x}_R - \hat{x}_L| = |x_R - x_L| = 2r. \tag{4}$$

The same result is also obtained for $u > \tilde{u}$. Thus, *for a triangulation system with the same physical dimensions as a DFD system, the disparity is equal to the size of the blur kernel.* An alternative interpretation is to consider the stereo baseline as a *synthetic aperture* of an imaging system. A proportion between the disparity and blur-diameter in a system as Fig. 2 (with the holes on the diameter having a finite support) was noticed in Adelson and Wang (1992).

The sensitivity (and resolution) of the triangulation systems are equivalent to those of DFD systems and are related to the disparity/PSF-support size (Eq. (4)): Depth deviation from focus is sensed if this value is larger than the pixel period[1] Δx (See Refs. (Abbott and Ahuja, 1993; Engelhardt and Hausler, 1988) and Subsection 5.5). The conclusion is that *methods that rely on the depth of field are not inherently less sensitive than stereo or motion*. In particular the rate of decrease of the resolution with object distance is fundamentally the same. In practice, however, the typical lens apertures used (Adelson and Wang, 1992) are merely in the order of ~1 cm while stereo baselines are usually one or two orders of magnitude larger, leading to a proportional increase of the sensitivity.

It is interesting to note that the common limits on lens apertures can be broken by the use of holographic optical elements (HOE). Holographic "lenses" are very thin, yet allow the deviation of rays by large angles. The design of such elements for imaging purposes is non-trivial, but HOE are actually in use in wide-angle head-up and helmet displays for aircraft (Amitai et al., 1989).

Consider depth from motion, that can be regarded as a "classic" triangulation approach. We shall see that it provides an effect analogous to 1D defocus blur. If discrete images are taken, the baseline between the initial and final frames dictates the depth resolution. Most DFD and motion approaches differ in the algorithms used: In DFD the support of the blur kernel is calculated by comparison to a small-aperture (reference) image, while motion based analysis relies on matching. However, the principle of operation of *Depth from Motion Blur* (DFMB) (Fox, 1988), is similar to DFD: A fast-shutter photograph is compared to an image blurred by the camera motion (slow shutter), to estimate the motion extent (Chen et al., 1996), from which depth is extracted (Fig. 4).

The analogy between DFD and DFMB can be enhanced by demonstrating the equivalent to a focused point in motion blur. Consider the system shown in Fig. 4. The camera moves along an arc of radius $\tilde{u}$, with its optical axis pointing towards the center of the circle. While the scene is generally motion blurred, a point at a distance $\tilde{u}$ remains unblurred! The analogous DFD system is constructed by removing part of the blocking shown in Fig. 2, exposing a thin line on the lens, between the former pinholes (thus the system can still be analyzed as having a single transversal dimension). Thus, the analysis of the spread is not based only on the two marginal points, but on a 1D continuum of points.

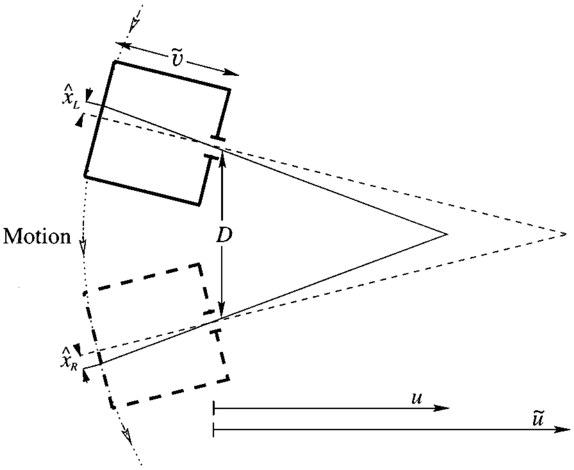

Figure 4. While the shutter is open, the camera moves along an arc, pointing to the arc axis at $\tilde{u}$. This point is sharply imaged while closer or farther points are motion blurred, in a manner analogous to defocus.

2.2. DFF

In DFF, depth is estimated by searching for the state of the imaging system for which the object is in-focus. Referring to Fig. 1, this may be achieved by changing either $\tilde{v}$ (the lens to sensor distance), F (the focal length) or u (the object distance), or any combination of them. Images are taken for each incremental change of these parameters. The state of the set-up for which the best-focused image was taken indicates the depth by the relation

$$\frac{1}{\tilde{u}} = \frac{1}{F} - \frac{1}{\tilde{v}}. \tag{5}$$

The process of changing the camera parameters to achieve a focused state is analogous to changing the convergence angle between two cameras in a typical triangulation system. This qualitative analogy has been stated before (Abbott and Ahuja, 1988; Pentland, 1987; Pentland et al., 1989). This can be seen clearly in Figs. 1–3. For example, focusing the system of Fig. 1 by axial movement towards/away from the object point changes u, to have $u \rightarrow \tilde{u}$, until the blur-radius is zero (or undetectable) has the same effect as moving the stereo system of Fig. 3 in that direction. Alternatively, focusing by changing the focal length F does not induce magnification, but shifts v so that $v \rightarrow \tilde{v}$ by changing the refraction angles of the light-rays in Figs.1 and 2. This has the same effect as changing the convergence angle in Fig. 3. Focusing by axially moving the sensor changes $\tilde{v}$ so that $\tilde{v} \rightarrow v$. This changes the magnification as well as the angles of the light-rays which hit the sensor at focus. This has the same effect as changing both $\tilde{v}$ and the convergence angle in Fig. 3. We note that magnification corrections (Darrell and Wohn, 1988; Nair and Stewart, 1992; Subbarao, 1988), which are usually insignificant (Stewart and Nair, 1989; Subbarao and Wei, 1992), enable focusing when the settings change is accompanied with magnification change.

The sensitivity to changes in parameters in DFF is related to the smallest detectable blur-diameter, while the sensitivity in stereo & vergence is related to the smallest detectable disparity. Both the disparity and the blur-diameter are sensed if they are larger than the pixel period. Since for the same system dimensions the blur-diameter and the disparity are the same, the sensitivity of DFF is similar to that of depth from convergence.

In Stewart and Nair (1989) the disparity in a stereo image pair was found empirically to be approximately linearly related to the focused state setting of a DFF system. We can now explain this result analytically. Suppose the system is initially focused at infinity. In order to focus on the object at u, the sensor has to be moved by

$$\Delta\tilde{v} = v - F, \tag{6}$$

which according to Eq. (5) is

$$\Delta\tilde{v} = \frac{Fv}{u}. \tag{7}$$

The sensor position $\tilde{v}$, or its distance $\Delta\tilde{v}$ from the focal point, indicate the focus setting. The stereo baseline is D_{stereo}. In the system of Stewart and Nair (1989), the stereo system was fixated at infinity thus the disparity was

$$d = D_{\text{stereo}} \cdot \frac{\tilde{v}}{u} = D_{\text{stereo}} \cdot \frac{v}{u}, \tag{8}$$

where in the right hand side of Eq. (8) we assumed that the disparity was measured at the state for which the object was focused, in that cooperative system. Combining Eqs. (7) and (8) we get

$$d = \frac{D_{\text{stereo}}}{F}\Delta\tilde{v} \tag{9}$$

which is a linear relation between the focus setting and the disparity. If focusing is achieved differently (e.g. moving the lens but keeping the sensor position fixed), there are higher order terms in the relation between focus-setting and disparity, but in practice they are negligible compared to the linear dependence.

3. Occlusion

3.1. DFD

The observation that monocular methods are not prone to the missing parts (occlusion) problem is mostly a consequence of the small "baseline" associated with the lens. The small angles involved reduce the number of points that will be visible to a part of the lens while being occluded at another part (vignetting caused by the scene). However, such incidents may occur (Adelson and Wang, 1992; Asada et al., 1998; Farid, 1997; Marshall et al., 1996).

Note that the same applies to stereo (Simoncelli and Farid, 1996) (or motion) with the same baseline! Although mechanical constraints usually complicate the construction of stereo systems with a small baseline, such systems can be made. An example is the "monocular stereo" system presented in Farid and Simoncelli (1998), whose principle of operation is similar to that shown in Fig. 2. Another possibility is to position a beam-splitter in front of the triangulation system. There is, of course, no "free lunch": the avoidance of the occlusion problem (and also the correspondence problem as will be discussed in Section 4) by decreasing the baseline leads to a reduction in sensitivity (Pentland et al., 1994).

The main differences between DFD and common triangulation methods arise when we consider the 2D nature of the image. It turns out that for the same system dimensions, *the chance of occurrence of the*

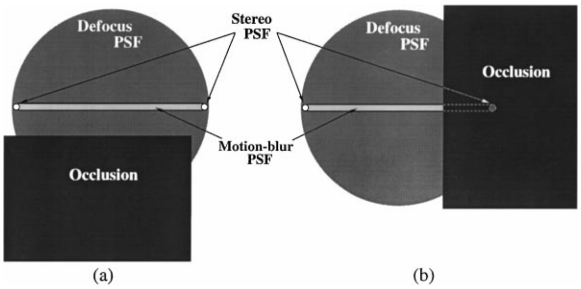

Figure 5. The stereo PSF consists of two distinct impulse functions. The line segment that defines the disparity between them is the support of the motion blur kernel, and the diameter of the defocus blur kernel for the same system dimensions. An occluding object is in-focus (and in perfect convergence in the stereo case) in this diagram, hence has sharp boundaries. In (a) the object is occluding only a DFD setup but not the stereo/motion setups. In (b) occlusion makes stereo matching impossible, and an error occurs in DFMB and DFD. In DFD, the diameter parallel to the occluding edge makes error-free recovery possible.

occlusion phenomenon is higher for DFD than for stereo (Fig. 5(a)). This is due to the fact that the defocus point-spread is much larger than for stereo. That is, there may be many situations in which occlusion occurs for the DFD system, and not for the stereo system.

Nevertheless, there is a difference in the consequences of occlusion. In stereo, the fact that one of the rays is blocked makes matching and depth estimation impossible (Fig. 5(b)). In contrast, DFD relies on a continuum of rays, thus allowing estimation, although with an error. If the occluded part is small compared to the support of the blur-kernel, and its depth is close to that of the occluding object, the error will be small. Depth from motion blur, acquired as described in Fig. 4 (or even a discrete sequence of images acquired as the camera is in motion) will have a similar stable behavior (Fig. 5(b)).

Consider small occlusions, covering less than half the blur PSF. In these cases the chief ray (the light ray that would have passed through a pinhole camera and marks the center of the PSF) is not occluded.[2] As seen in Fig. 6 the relative error in the support of the defocus blur is smaller than that of motion blur. This is an advantage of DFD over DFMB. Moreover, from Fig. 5 one can notice that with DFD it is also possible (although not by the current algorithms known to us) to *fully* recover the true blur diameter using a line in the PSF that is parallel to the occluding edge.

Evidence of problems near occlusion boundaries in a “monocular stereo” system is reported in Adelson and Wang (1992). These problems occur since some points in the scene were occluded to certain parts of the lens aperture. Had that system been used for DFF/DFD, similar occlusions would have taken place. Ref. (Adelson and Wang, 1992) reported that the occlusion effect is small. This is due to the small baseline associated with that system. Experimental evidence of the phenomenon is also reported in Asada et al. (1998).

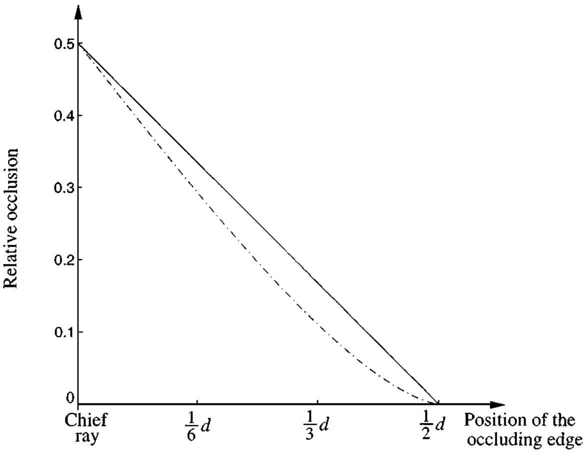

Figure 6. The occluding edge of Fig. 5b is at a certain distance to the right of the chief ray. For small occlusions the chief ray is visible and the relative part of the PSF that is occluded is smaller for DFD [dashed line] than for motion [solid line].

To conclude, DFD does not avoid the occlusion problem anymore than stereo/motion methods (on the contrary). It is, however more stable to such disruptions. In principle, with DFD it is possible to fully recover the depth as long as the occlusion is small.

3.2. DFF

From the discussion in Subsection 3.1, it follows that occlusion is present also in DFF. In a stereo system with a baseline that is as small as the aperture of typical DFF systems, the occlusion phenomenon would be much less noticeable than in a stereo system with a large baseline. Moreover, as described in Fig. 5(a), for systems of the same physical dimensions *the chance of occlusion is higher in DFF than in stereo* due to the 2D nature of the PSF.

The imaging of occluded objects by finite aperture lenses was analyzed in Marshall et al. (1996). Since the occluding object is out of focus, it is blurred. However, this object causes vignetting to the objects behind it. Thus, the occluded object *fades* into the occluder. If the occluded object is left of the occluding edge (in the image space), the image obtained using an aperture D is

$$g_D = \text{Occluded} \cdot (1 - h_D * \text{Step}(x_0)) + \text{Occluder} * h_D, \qquad (10)$$

where $\text{Step}(x_0)$ is the step-function at the occluding edge position x_0. In Eq. (10) the blur kernel h_D of the occluding object has a radius r while the occluded object (for which we seek focus) is assumed to be focused.

Inspecting Figs. 7 and 8, there are four classes of image points:

1. $x < x_0 - r$. The point is not occluded. Depth at the point is unambiguous.
2. $x_0 - r \leq x < x_0$. The point is slightly occluded (See Fig. 7(a)). The chief ray from this object point reaches the lens. The point may appear focused but the disturbance of the blurred occluder may shift the estimation of the plane of best focus in DFF. In a stereo & vergence system of the same physical dimensions, each of the two pinholes will see a different object, either the occluder or the occluded one. Thus fixation is ill posed (no solution).
3. $x_0 \leq x \leq x_0 + r$. The point is severely occluded. The chief ray from this object point does not reach the lens. The point may appear focused but during the focus search the same point x will indicate a focused state also when the occluder is focused (see Fig. 8). The solution is not unique (double valued). Simple DFF is thus ambiguous. Nevertheless, the depth at the point may be resolved if the possibility of a layered scene is taken into account (See (Schechner et al., 1998) for a proposed method for DFF with double valued depth).

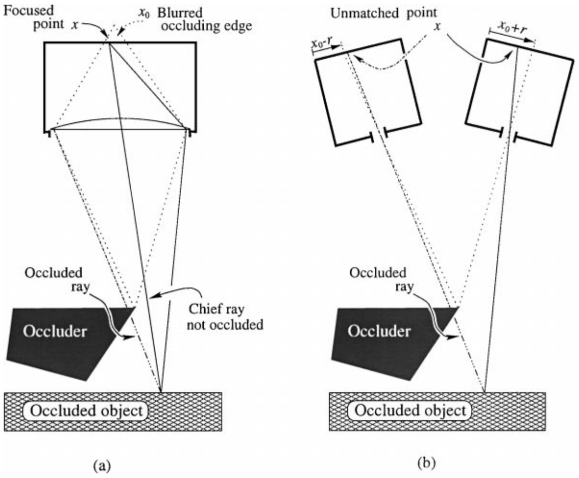

Figure 7. (a) If the chief ray is not occluded but resides within the blurred image of the occluding edge (slight occlusion) focusing is possible but may be erroneous. (b) For the same system dimensions matching the occluded object point in the stereo/vergence images is not possible.

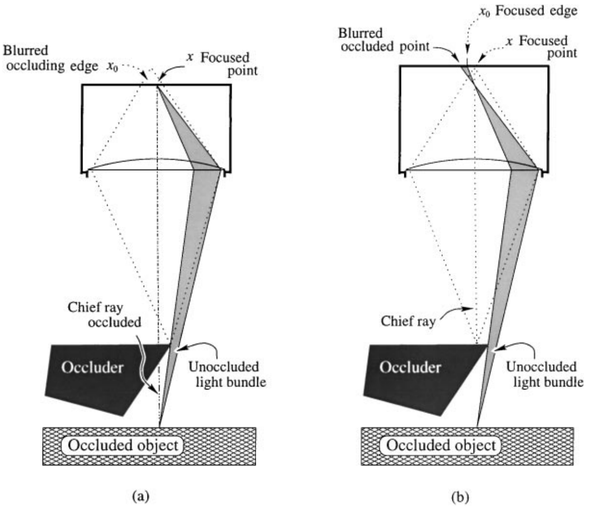

Figure 8. (a) If light emanating from the object point reaches the sensor but the chief ray is occluded (severe occlusion) focusing on this occluded point is possible. (b) The same transversal image point is also in focus if the system is tuned on the occluder. Thus, the depth at the point x is double valued. Matching stereo/vergence points is possible only in case (b) (see Fig. 7).

   The occluder at that point is seen to both pinholes in the stereo & vergence system. Thus convergence is possible and the correct depth of the occluder will be the solution at point x. This is a unique solution since matching the occluded point is impossible, for the same reason detailed in the case of slight occlusion.
4. $x > x_0 + r$. The focusing (DFF) and fixation (convergence) are done on the close (possibly occluding) object. Depth at the point is unambiguous.

Occlusion is present in cases 2 and 3 above, and a correct and unique matching is not guaranteed. However, if the occlusion is small (i.e. the chief ray is visible) the situation is similar to that described in Subsection 3.1: the stereo/vergence system cannot yield the solution while DFF yields a depth value that approaches the correct one for smaller and smaller occlusions. On the other hand, if the occlusion is severe (the chief ray is occluded) DFF yields an ambiguous estimation (which can be resolved if a *layered* scene is admitted, as in Schechner et al. (1998)) while depth from convergence yields a correct and unique depth estimation.

4. Matching (Correspondence) Ambiguity

Defocus measurement is not a point operation in the sense that in order to estimate depth at given image coordinates it is *not* sufficient to compare the points having those coordinates in the acquired images. In DFD, depth is extracted by matching a spot (sharp or blurred) in one image with a corresponding blurred spot in another image. Even if the center of the blurred spot is known, its support is unknown—unless the scene consists of sparse points. It is possible to estimate the support of the blur kernel for piecewise planar scenes (Scherock, 1991), or scenes with slowly varying depth, as long as the support of the blur-kernel is sufficiently small to ensure that the disturbance from points of different depths is negligible. The estimation of the blur kernel support is generally difficult, though not impossible, if large depth deviations can take place within small neighborhoods. Note that in stereo too the disparity should be approximately constant over the patches (which are segments along the epipolar lines) to ease their registration between the images (Abbott and Ahuja, 1988; Abbott and Ahuja, 1993).

The neighborhoods used for the estimation of the kernel need to be larger than the support of the PSF. A good demonstration for this aspect is given in Rioux and Blais (1986). In that work, the object was illuminated with sparse points and the PSF was a ring. The depth was estimated by the ring-diameter.[3] This seems like an easy task since the points are sparse. However, this task would have been much more complicated if adjacent rings had overlapped. Thus *to avoid ambiguity, the 'image patches' had to be larger than the largest possible blur kernel*.

In natural scenes, if a significant feature is outside the neighborhood used in the estimation, and its distance from the patch is about the extent of the point-spread, *edge bleeding* (Jarvis, 1983; Nair and Stewart, 1992; Stewart and Nair, 1989) occurs, spoiling the solution. This demonstrates that DFD *is not a pointwise measurement* (but rather a point-to-patch or a patch-to-patch comparison). Thus the assumption that in DFD each image point corresponds simply to the point with the same coordinates in the other image is erroneous. This wrong assumption cannot be used to overrule the possibility of matching (correspondence) problems.

Image patches that contain the support of the blur kernel (or the disparity) are needed in DFD as well as in stereo, when trying to resolve the disparity/blur-diameter. However the implications are much less significant in stereo/motion, since there the search for the matching is done only along the epipolar lines so the "patches" are 1D (very narrow). Usually, the correspondence problem in stereo *is* solvable, but its existence complicates the derivation of the solution. We claim that a similar problem exists also in DFD, and it also may complicate the estimation. We now concentrate on the simple situation where the patches are sufficiently large and depth-homogeneous. Then, analysis in the spatial-frequency domain is possible.

4.1. Stereo

One of the disadvantages attributed to stereo/motion is the correspondence problem. Adelson and Wang (1992) interpreted this problem as a manifestation of aliasing. Let the left image be $g_L(x, y)$ while the right image is $g_R(x, y) = g_L(x - d, y)$. We postpone the effect of noise to Section 5. Having the two images, we wish to estimate the disparity, for example by minimizing the square error

$$E^2(\hat{d}) = |g_R(x, y) - g_L(x - \hat{d}, y)|^2, \qquad (11)$$

where the baseline is along the x-axis. We denote a spatial frequency by $\vec{\nu} = (\nu \cos\phi, \nu \sin\phi)$. In case the image is periodic (Abbott and Ahuja, 1993; Marapane and Trivedi, 1993; Pentland et al., 1994; Stewart and Nair, 1989), for example, if the image contains a single frequency component $g_L(x, y) = Ae^{j2\pi\nu(x\cos\phi + y\sin\phi)}$, the solution is not unique:

$$\hat{d} = d + \frac{k}{\nu\cos\phi} \qquad k = \ldots -2, -1, 0, 1, 2, 3 \ldots \qquad (12)$$

This difficulty arises from the fact that the transfer function between the images,

$$H(\vec{\nu}) = e^{-j2\pi\nu d\cos\phi}, \qquad (13)$$

is not one-to-one. The problem is dealt with by restricting the estimation to be in frequency bands for which the transfer function is one-to-one, for example by demanding

$$|\nu d \cos\phi| < \frac{1}{2} \quad \text{or} \quad 0 < \nu d \cos\phi < 1. \tag{14}$$

Subject to these restrictions, the registration of the two images is easy and unique. Thus, the correspondence problem is greatly reduced if the disparity is small. If the frequency or disparity are too high (larger than the limitation posed by Eq. (14)), the ambiguity is analogous to aliasing (Adelson and Wang, 1992). If the stereo system is built with a small baseline (Adelson and Wang, 1992; Farid and Simoncelli, 1998) as in common monocular systems, the correspondence problem will be avoided (Adelson and Wang, 1992).

The raw images are usually not restricted to the cut-off frequencies dictated by Eq. (14), when d is larger than a pixel. Thus the images should be blurred before the estimation is done, either digitally as in Bergen et al. (1992) or by having the sensor placed out of focus as in Adelson and Wang (1992), and Simoncelli and Farid (1996). In this process information is lost, leading to a rough estimate of the disparity (as will be indicated by the results in Section 5). This coarse estimate can be used to resolve ambiguity in the band $0 < \nu d \cos\phi < 2$, and thus the estimation can be refined. This in turn allows further refinement by using even higher frequency bands. This is the basis of the *coarse-to-fine* estimation of the disparity (Bergen et al., 1992). The larger the product νd, the more calculations are needed to establish the correct matching. This is compatible with the observations that the complexity of stereo matching increases as disparities grow (Klarquist et al., 1995; Marapane and Trivedi, 1993) and that edgel-based stereo (which relies on high frequency components) is more complex than region based matching (Marapane and Trivedi, 1993). The source of the coarse estimate is not necessarily achieved by the same stereo system, but is nevertheless needed (Dias et al., 1992; Klarquist et al., 1995; Marapane and Trivedi, 1993; Subbarao et al. 1997).

4.2. DFD by Aperture Change and DFMB

Does DFD avoid the matching ambiguity problem at all? We shall now show that the answer is, generally, no. We consider in the following the pillbox model (Nayar et al., 1995; Watanabe and Nayar, 1996) which is a simple geometrical optics model for the PSF. In this model the intensity is spread uniformly within the blur kernel. In 1D blurring, the pillbox kernel is simply the window function $h_D = D/d$ for $|x| < d/2$. The total light energy collected by the aperture (and spread on the sensor) is proportional to its width D in this 1D system. This system is analogous to DFMB. The transfer function is

$$H_D(\vec{\nu}) = D\frac{\sin(\pi\nu d\cos\phi)}{\pi\nu d\cos\phi} = D\,\text{sinc}(\nu d\cos\phi), \tag{15}$$

where the blur diameter d is given by Eq. (3). Inserting Eq. (1) into Eq. (15) and taking the limit of small D, the transfer function of the pinhole (reference) aperture is

$$H_0(\vec{\nu}) = D_0 \tag{16}$$

for all ν, where D_0 is the width of the pinhole. Having the pinhole image g_0 and the large-aperture image g_D, we wish to estimate the blur diameter, for example by minimizing an error criterion (Hiura et al., 1998), like

$$E^2(\hat{d}) = |g_D * h_0 - g_0 * \hat{h}_D|^2. \tag{17}$$

In the case where the image is periodic and consists of a single frequency component,

$$g_0(x, y) = D_0 G e^{j2\pi\nu(x\cos\phi + y\sin\phi)}, \tag{18}$$

the solution is again not unique since the transfer function between the images

$$H(\vec{\nu}) = \frac{H_D(\vec{\nu})}{H_0(\vec{\nu})} = \frac{D}{D_0}\text{sinc}(\nu d\cos\phi), \tag{19}$$

is not one-to-one (The DFMB transfer function is proportional to the one in Eq. (19), where the aperture dimensions ratio is replaced by the ratio of exposure times.). Since the transfer function is not one-to-one, a measured attenuation is the possible outcome of several blur kernel diameters.

As done in stereo (Adelson and Wang, 1992), we may restrict the estimation to frequency bands for which the transfer function is one-to-one. For DFMB this dictates that

$$0 < \nu d\cos\phi < 1.43, \tag{20}$$

where 1.43 is the location of the first minimum of expression (19). So, we can use a wider frequency band than that used in stereo systems (14) having the same physical dimensions, before needing a coarse to fine approach.

In the 2D pillbox model (Nayar et al., 1995; Watanabe and Nayar, 1996), the PSF is $h_D = D^2/d^2$ for $\sqrt{x^2+y^2} < (d/2)^2$. The defocus transfer function is

$$H_D(\vec{\nu}) = \frac{\pi D^2}{2} \frac{\mathrm{J}_1(\pi \nu d)}{\pi \nu d} \tag{21}$$

while

$$H_0(\vec{\nu}) = \pi \frac{D_0^2}{4}. \tag{22}$$

Thus

$$H(\vec{\nu}) = 2\frac{D^2}{D_0^2} \frac{\mathrm{J}_1(\pi \nu d)}{\pi \nu d} \tag{23}$$

is also not one-to-one. Thus, the ambiguity (correspondence) problem also occurs in the DFD approach, and finite-aperture monocular systems do not guarantee uniqueness of the solution for periodic patterns. *There are scenes for which the solution of DFD (i.e. matching blur kernels in image pairs) is not unique.*

The defocus transfer function in Eq. (21) is monotonically decreasing in the range

$$0 < \nu d < 1.63. \tag{24}$$

Eq. (24) appears as if it enables unique matching in a wider band than can be used in stereo (Eq. (14)). However, note that very high spatial frequencies may be used in the stereo process without matching ambiguity, as long as the component along the baseline has a sufficiently small frequency. On the other hand, Eq. (24) does not allow that. Hence, in contrast to common belief, common triangulation techniques (as stereo) may be *less prone* to matching ambiguity than 2D-DFD.

The above discussion is relevant not only for periodic functions. Integrating Eq. (11) or Eq. (17) over a patch is equivalent to integrating the square errors in all frequencies. Furthermore, disparity/blur estimation by fitting a curve or a model to data obtained in several frequencies has been used (Bove, 1989; Hiura et al., 1998; Pentland, 1987; Pentland et al., 1994; Watanabe and Nayar, 1996).

The conclusion that the ambiguity problem is present in DFD is not restricted to the pillbox model, but to all transfer functions which are not one-to-one, particularly those having side lobes (see Castleman (1979), FitzGerrell et al. (1997), Hopkins (1955), Lee (1990), and Schneider et al. (1994) for theoretical functions). Hopkins (1955) explicitly referred to the phenomenon of increase of contrast at large defocus due to un-monotonicity of the transfer function at the high frequencies. In other words, although the two acquired images and the laws of geometric optics impose constraints on the spread parameter (blur-diameter) (Subbarao, 1988; Subbarao and Liu, 1996), there may be several 'intersections' between these constraints, leading to ambiguous solutions.

Empirical evidence for the possibility of this phenomenon can be found by studying the results reported in Klarquist et al. (1995). In that work, flat objects textured with a single spatial frequency were imaged at various focus settings. The graphs given in Klarquist et al. (1995) show that, especially at high spatial frequency inputs, the attenuation as a function of focus setting (i.e., the blur diameter) is not monotonous, potentially leading to ambiguous depth estimation.

The common assumption in DFD that the PSF is a Gaussian simplifies calculations (Subbarao, 1988; Surya and Subbarao, 1993) but generally is incorrect (Bove, 1993). This assumption should not be taken as a basis for believing that the actual transfer function is one-to one (using the wrong transfer function will lead to a wrong estimation of d). If, however, the actual transfer function is one-to-one for all frequencies (Lee, 1990), the ambiguity phenomenon does not exist, and there is a unique match. However, as will be discussed in Section 5, in that situation the problem is still ill conditioned in the high frequencies.

4.3. *DFD by Change of Focus-Settings*

The change in the blur-diameter between the input images may be achieved by changing the focus settings rather than changing the aperture size. For example, the sensor array may move axially between image acquisitions. We shall show that this leads to the same limitation as when DFD is done by changing the aperture size (Eq. (24)). We assume that geometric changes in magnification are compensated or do not take place (e.g. by the use of a telecentric system (Nayar et al., 1995; Watanabe and Nayar, 1996), depicted in Fig. 9). The aperture size D is constant, so in this Subsection we parameterize the transfer function by the blur diameter d.

Let the two images be $g_1 = g_0 * h_d$ and $g_2 = g_0 * h_{d+\Delta d}$. Δd is the change in the blur-diameter due

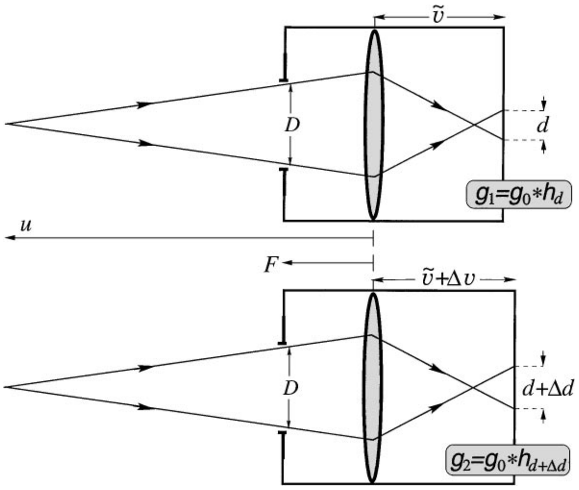

Figure 9. In a telecentric system, the aperture stop is at the front focal plane. Such a system attenuates the magnification change while defocusing. Shifting the sensor position by Δv causes a change of Δd in the blur diameter.

to the known shift Δv in the sensor position (Fig. 9). This change is invariant to the focus settings and the object depth in telecentric systems (Nayar et al., 1995; Schechner et al., 1998). The transfer function between the images is now

$$H(\vec{\nu}) = \frac{H_{d+\Delta d}(\vec{\nu})}{H_d(\vec{\nu})}. \qquad (25)$$

At frequencies for which $|H_d(\vec{\nu})| \ll |H_{d+\Delta d}(\vec{\nu})|$ we can take the reciprocal of Eq. (25) as the transfer function between the images (in reversed order).

In Subsection 4.2 we showed that if $H(\vec{\nu})$ is not one-to-one in d, the estimation may be ambiguous. Fig. 10 plots the response to a specific frequency ν of the 2D pillbox model (21) as a function of the blur-diameter. The figure also plots the response at the axially-displaced image ($\Delta d = 1/(2\nu)$ in this example), which is the same as the former response, but shifted along the d axis. Each ratio between these responses can be yielded by many diameters d. To illustrate, view Fig. 11, which plots the ratio between the frequency responses in Fig. 10. The ratio is indeed not one-to-one. The lowest band for which the ratio is one-to-one in this figure is $0 < \nu d < 1.46$. However, if the axial increments of the sensor position are smaller, this bandwidth broadens. As Δd is decreased, the responses shown in Fig. 10 converge. Convergence is fastest near the local extrema of $H_d(\nu)$. Hence, as $\Delta d \to 0$ the lowest band in which the matching (correspondence) ambiguity is avoided is between the two first local extrema, i.e.,

$$0 < \nu d < 1.63, \qquad (26)$$

which is the same as Eq. (24).

Simulation and experimental results reported in Watanabe and Nayar (1996) support this theoretical result. In the DFD method suggested in Watanabe and

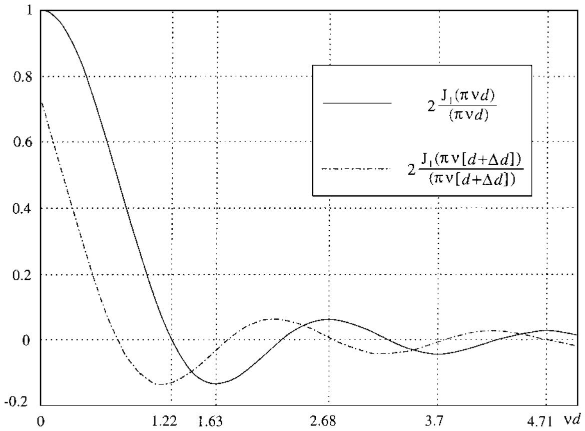

Figure 10. [Solid line] The attenuation of a frequency component ν between a focused and a defocused image as a function of the diameter of the blur kernel d. The horizontal axis is scaled by ν. [Dashed line] The attenuation of the same frequency component when the focus settings are changed so that the blur diameter is $d + \Delta d$, for the case $\Delta d = 1/(2\nu)$.

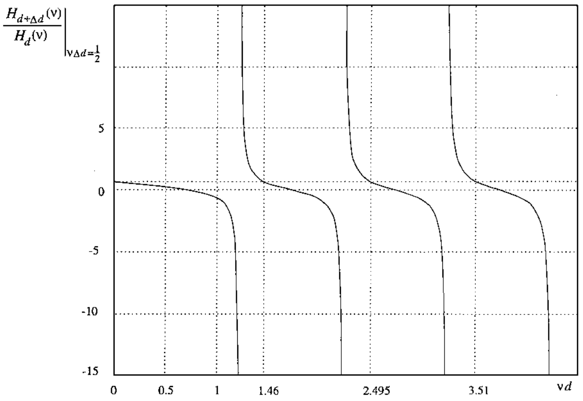

Figure 11. Two images are acquired with different focus settings. The transfer function between the images is the ratio between their individual frequency responses, plotted in Fig. 10. In the DOF threshold (see Subsection 5.5) $\Delta d = 1/(2\nu)$, for which the width of the band without ambiguities satisfies $\nu d \approx 1.46$. For infinitesimal Δd this width satisfies $\nu d \approx 1.63$. For high frequencies or large diameters the width of each band is $\nu d \approx 1$ as in stereo.

Nayar (1996), the defocus change between acquired images was obtained by changing the focus settings. The images were then filtered by several band pass operators, and the ratios of their outputs were used to fit a model. Watanabe and Nayar (1996) noticed that the solution may be ambiguous due to the unmonotonicity of the ratios, as a function of the frequency and the blur diameter. However, the relation to correspondence, which was related there only to stereo, was not noticed. To avoid the ambiguity they limited the band used to the first zero crossing of the pillbox model (21) which occurs at $\nu d = 1.22$. However, their tests revealed that the frequency band can be extended by about 30%, i.e., to $\nu d \approx 1.6$, in agreement with Eq. (26). The ratio computed in Watanabe and Nayar (1996) is actually a function of the transfer function defined in Eq. (25) between the images. Thus, the possibility of extending the frequency band beyond the zero crossing is not unique to the rational filter method; it is a general property of DFD.

High frequencies were not used in Watanabe and Nayar (1996) for depth estimation since they are beyond the monotonicity cutoff. It seems that these 'lost' frequencies can be used in a manner similar to the coarse-to-fine approach in stereo (i.e., using the estimation based on low frequencies to resolve the ambiguity in the high frequencies).

4.4. DFF

We believe that by using a sufficiently large evaluation patch and some depth homogeneity within the patch, DFF is freed of the matching problem. Contrary to common statements in the literature, the avoidance of the matching problem in DFF is not trivial.

Focus measurement (like defocus and disparity measurements) is not a point operation. It must be calculated (Jarvis, 1983; Nair and Stewart, 1992; Stewart and Nair, 1989; Subbarao and Liu, 1996) over a small patch implicitly assuming that the depth of the scene is constant (or moderately changing) within the patch (Dias et al., 1992; Marapane and Trivedi, 1993). The state of focus is detected by comparison of focus ("sharpness") measurements in the same patch over several focus settings. To have a correct depth estimation, the focus measure in the patch should be largest in the focused state. The patch must be at least as large as the support of the widest blur kernel expected in the setup, otherwise errors due to *edge bleeding* (Nair and Stewart, 1992; Stewart and Nair, 1989) could occur (Fig. 12.). Assuming the patch to be sufficiently large, we can make some observations in the frequency domain.

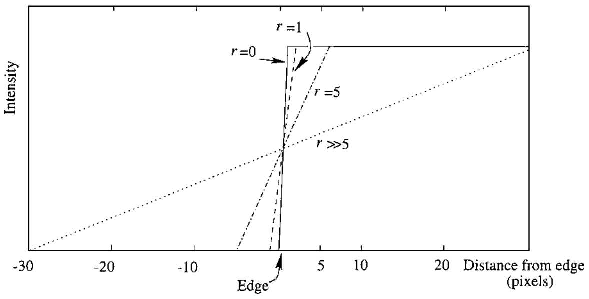

Figure 12. Edge bleeding. The solid line shows an intensity edge. The dashed and dotted lines show the edge in pillbox-blurred images with several blur radii. The gradient at location 5^- is maximal when the radius is 5 rather than 0, misleading focus detection.

Periodic images make depth from stereo ambiguous (Subsec. 4.1). They do the same to depth from vergence. As the vergence angles are changed, several vergence states yield perfect matching. On the other hand, DFF seems indeed to be immune to ambiguity due to periodic input (Abbott and Ahuja, 1993; Marapane and Trivedi, 1993; Stewart and Nair, 1989). Since the blur transfer function is a LPF, the energy at any spatial frequency composing the image is largest at the state of focus. As the image is defocused the high-frequencies response quickly decreases (Hopkins, 1955), and decrease in the response to other frequencies (except DC) follows. As the image is further defocused there may be local risings of the frequency response (side lobes in the response at some frequency, as a function of d). However, no local maximum is as high as the response at focus in reasonable physical systems. Thus, the determination of the focused state is unambiguous in each of the frequency components (except DC).

5. Robustness and Response to Perturbations

In some previous works, it has been empirically observed that DFD/DFF methods are more robust than stereo. In this section we analyze the responses of DFD, stereo and motion to perturbations, in a unified framework. Some of the results depend on the characteristics of the specific model of the optical transfer function (OTF), like monotonicity and the existence of zero-crossings. For defocus we use the pillbox model (Nayar et al., 1995; Noguchi and Nayar, 1994; Watanabe and Nayar, 1996), since it is valid for aberration-free geometric optics, and has been shown to be a good approximation for large defocus (Hopkins, 1955; Lee, 1990; Schneider et al., 1994). The effects of physical optics and aberrations influence the results but one must remember that these affect also stereo and motion. Since the literature on stereo and motion neglects these effects, we maintain this assumption so as to have a common basis for comparison between stereo/motion and DFD. Nevertheless, the procedure used in this chapter is general and can serve as a guideline in the analysis of other models.

5.1. *General Error Propagation*

Let us analyze the effect of a perturbation in some spatial frequency component of the image. The perturbation affects the estimated transfer function between the images, which in turn causes an error in the estimated blur-diameter (DFD) or disparity (stereo). This leads to an error in the depth estimation. As in Section 4 we note that studying the behavior of each spectral component has an algorithmic ground: there are several methods (Bove, 1989; Hiura et al., 1998; Pentland, 1987; Pentland et al., 1994; Watanabe and Nayar, 1996) which rely directly on the frequency components or on frequency bands (Pentland et al., 1994) for depth estimation. Since stereo, DFD or DFMB are based on comparison of two acquired images, we shall check the influence of a perturbation in any of the two. The problem is illustrated in Fig. 13.

The transfer function $H(\vec{\nu})$ between the image G_D (in the frequency domain) to a reference image G_0 is parameterized by the disparity/blur-diameter. We wish to estimate this parameter, for example by looking for the transfer function $\hat{H}$ that will satisfy

$$G_D(\vec{\nu}) = G_0(\vec{\nu})\hat{H}(\vec{\nu}). \tag{27}$$

Let a perturbation occur at the reference image g_0. The images are thus related by

$$G_D(\vec{\nu}) = [G_0(\vec{\nu}) - N_0(\vec{\nu})]H(\vec{\nu}), \tag{28}$$

where $H(\vec{\nu})$ is the true transfer function and N_0 is the perturbation. Eqs. (27,28) yield

$$\begin{aligned}\hat{H}(\vec{\nu}) &= H(\vec{\nu}) - N_0(\vec{\nu})H(\vec{\nu})/G_0(\vec{\nu})\\ &= H(\vec{\nu}) - \frac{|N_0(\vec{\nu})|}{|G_0(\vec{\nu})|}e^{j\vartheta(\vec{\nu})}H(\vec{\nu}),\end{aligned} \tag{29}$$

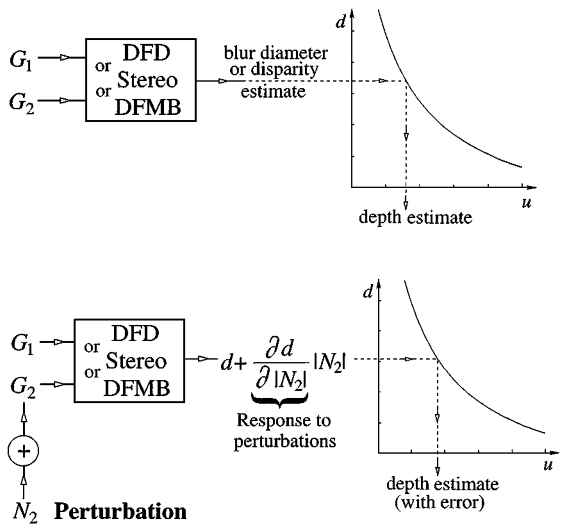

Figure 13. [Top]: In either of the depth estimation methods, two images are compared, where G_1, G_2 may be G_0 and G_D, respectively, or vice versa. The comparison yields and estimate of the blur diameter/disparity, leading to the depth estimate. The relation between d and u is similar for DFD/stereo/DFMB for the same system dimensions. [Bottom] A perturbation added to one of the images leads to a deviation in the estimation of d, leading to an error in the depth estimate.

where $\vartheta(\vec{\nu})$ is the phase of the perturbation relative to the signal component $G_0(\vec{\nu})$. Usually both constraints (27,28) cannot be satisfied simultaneously at all frequencies, hence a common method is to minimize the MSE

$$\begin{aligned}E^2 &= \int_{\vec{\nu}} |G_D(\vec{\nu}) - \hat{H}(\vec{\nu})G_0(\vec{\nu})|^2\, \mathrm{d}\vec{\nu}\\ &= \int_{\vec{\nu}} |G_0(\vec{\nu})|^2 |[H(\vec{\nu}) - \hat{H}(\vec{\nu})] - N_0 H(\vec{\nu})/G_0|^2\, \mathrm{d}\vec{\nu}.\end{aligned} \tag{30}$$

This is achieved by looking for the extremum points

$$\begin{aligned}\frac{\partial(E^2)}{\partial\hat{d}} = -2Re \int_{\vec{\nu}} |G_0(\vec{\nu})|^2 \biggl[H(\vec{\nu}) - \hat{H}(\vec{\nu})\\ - \frac{N_0 H}{G_0}\biggr] \frac{\partial \hat{H}^*(\vec{\nu})}{\partial \hat{d}}\, \mathrm{d}\vec{\nu} = 0.\end{aligned} \tag{31}$$

Local minima of E^2 may appear at different estimates $\hat{d}$, for different signals and perturbations, depending on their spectral content.

Attempting to analyze in a systematic way, let us assume that the signal is made of a single frequency $\vec{\nu}$, thus

$$G_0(\vec{\nu}') = D_0^2 G(\vec{\nu})\delta(\vec{\nu} - \vec{\nu}'). \tag{32}$$

If at that frequency $\partial\hat{H}^*(\vec{\nu})/\partial\hat{d} = 0$, the estimation of $\hat{d}$ is ill posed (or very ill conditioned). Otherwise, nulling the integrand yields Eq. (29), which shows how the estimated frequency response changes with the influence of the perturbation. From $\hat{H}(\vec{\nu})$ the parameter $\hat{d}$ and the depth $\hat{u}$ are derived (3). The response of the depth estimation to perturbations is

$$\frac{\partial\hat{u}(\vec{\nu})}{\partial|N_0(\vec{\nu})|} = \frac{\partial\hat{u}}{\partial f(\hat{u})}\frac{\partial f(\hat{u})}{\partial|N_0(\vec{\nu})|}, \tag{33}$$

where $f(u) = d/D$ is as defined in Eq. (3). As we showed in Section 2, $f(u)$ is the same for stereo and DFD systems having the same physical dimensions, thus the factor $\partial u/\partial f(u)$ is common for both systems. Hence, in the coming comparison between these approaches we omit this factor and use $\partial f(u)/\partial|N_0|$ as a measure for the response to perturbations. Since the estimation will be frequency-dependent, we write

$$\begin{aligned}\frac{\partial f(\hat{u},\vec{\nu})}{\partial|N_0(\vec{\nu})|} &= \frac{\partial f(\hat{u},\vec{\nu})}{\partial\hat{H}(\vec{\nu})}\frac{\partial\hat{H}(\vec{\nu})}{\partial|N_0|}\\ &= -\frac{e^{j\vartheta(\vec{\nu})}H(\vec{\nu})}{|G_0(\vec{\nu})|}\left[\left.\frac{\partial H(\vec{\nu})}{\partial f(u)}\right|_{\hat{u}}\right]^{-1},\end{aligned} \tag{34}$$

where G_0 is given by Eq. (32).

Suppose now that the perturbation occurs in the transformed (shifted, or blurred) image. Eq. (28) takes the form

$$G_D(\vec{\nu}) = G_0(\vec{\nu})H(\vec{\nu}) + N_D(\vec{\nu}), \tag{35}$$

while Eq. (30) changes to

$$\begin{aligned}E^2 &= \int_{\vec{\nu}} |G_D(\vec{\nu}) - \hat{H}(\vec{\nu})G_0(\vec{\nu})|^2 \mathrm{d}\vec{\nu}\\ &= \int_{\vec{\nu}} |G_0(\vec{\nu})|^2 |[H(\vec{\nu}) - \hat{H}(\vec{\nu})] + N_D/G_0|^2\, \mathrm{d}\vec{\nu}.\end{aligned} \tag{36}$$

Reasoning similar to Eqs. (29) and (31) yields

$$\hat{H}(\vec{\nu}) = H(\vec{\nu}) + \frac{|N_D(\vec{\nu})|}{|G_0(\vec{\nu})|}e^{j\vartheta(\vec{\nu})}. \tag{37}$$

The response of the depth estimation to the perturbation is

$$\begin{aligned}\frac{\partial f(\hat{u},\vec{\nu})}{\partial|N_D(\vec{\nu})|} &= \frac{\partial f(\hat{u},\vec{\nu})}{\partial\hat{H}(\vec{\nu})}\frac{\partial\hat{H}(\vec{\nu})}{\partial|N_D|}\\ &= \frac{e^{j\vartheta(\vec{\nu})}}{|G_0(\vec{\nu})|}\left[\left.\frac{\partial H(\vec{\nu})}{\partial f(u)}\right|_{\hat{u}}\right]^{-1}.\end{aligned} \tag{38}$$

5.2. Stereo—The Aperture Problem

For stereo, the transfer function $H(\vec{\nu})$ is given by Eq. (13), so

$$\frac{\partial f_{\text{stereo}}(\hat{u},\vec{\nu})}{\partial|N_0(\vec{\nu})|} = \frac{e^{j[\vartheta(\vec{\nu})-\pi/2]}}{|G(\vec{\nu})|}\frac{1}{2\pi D_0^2 D}\frac{1}{\nu\cos\phi}, \tag{39}$$

$$\frac{\partial f_{\text{stereo}}(\hat{u},\vec{\nu})}{\partial|N_D(\vec{\nu})|} = \frac{e^{j[\vartheta(\vec{\nu})+\pi/2+2\pi\nu d\cos\phi]}}{|G(\vec{\nu})|}\frac{1}{2\pi D_0^2 D}\frac{1}{\nu\cos\phi}. \tag{40}$$

The terms in these equations express in a quantitative manner intuitive characteristics: the stronger the signal $G(\vec{\nu})$, the smaller is the response to the perturbation; the DC component ($\nu = 0$) contribution to the disparity estimation is ill-posed; estimation by the low frequencies is ill-conditioned. The instability at the low frequencies stems from the fact that much larger deviations in $\hat{d}$ are needed to compensate for the perturbation, while trying to maintain Eq. (29), than in the higher frequencies. Thus, Eq. (39) expresses mathematically the weakness of stereo in scenes lacking high-frequency content.

These equations also express mathematically the *aperture problem* in stereo. The smaller the component of the periodic signal along the baseline (Adelson and Wang, 1992), the larger the error is. As $|\phi| \to \pi/2$ we need to have $D \to \infty$ to keep the error finite.

5.3. Motion and 1D Blur

For DFMB (analogous to 1D-DFD) the transfer function is proportional to expression (19), which has zero crossings. Perturbations in the reference image at frequencies/diameters for which $H(\vec{\nu}) = 0$ influence neither the error (30) nor the depth estimation (34). Thus, if the transfer function has zero crossings (as in Castleman (1979), FitzGerrell et al. (1957), Hopkins (1955), Lee (1990), and Schneider et al. (1994), the

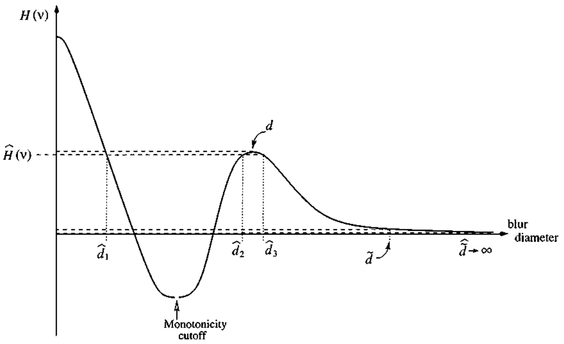

Figure 14. For a specific $\vec{\nu}$ the transfer function H depends on the blur-diameter. A true diameter d (larger than the monotonicity cutoff) has several solutions (e.g. $\hat{d}_1, \hat{d}_2$). Close to a peak or trough a small deviation in the estimated $\hat{H}$ causes a significant but bounded error (see $\hat{d}_2, \hat{d}_3$). At the high frequencies or defocus blur the transfer function is indifferent to changes in d, thus the error may be infinite (see $\tilde{d}$ vs. $\hat{\tilde{d}}$). Hence such frequencies would better be discarded.

estimation based on the zero-crossing frequencies is completely immune to noise added to the reference image, i.e.,

$$\left.\frac{\partial f(\hat{u},\vec{\nu})}{\partial|N_0(\vec{\nu})|}\right|_{H(\vec{\nu})=0,\ \frac{\partial H}{\partial d}\neq 0} = 0. \tag{41}$$

As for perturbation in the blurred image,

$$\left.\frac{\partial f_{\text{DFMB}}(\hat{u},\vec{\nu})}{\partial|N_D(\nu)|}\right|_{H(\vec{\nu})=0} = \pm\frac{e^{j\vartheta(\vec{\nu})}}{|G(\vec{\nu})|}\frac{1}{D}f(\hat{u}). \tag{42}$$

Thus close to the zero crossings the results are stable even when the frequency is high.

Nevertheless, if the transfer function has zero-crossings it is not monotonous, having peaks and troughs. In these situations $\partial\hat{H}/\partial\hat{d}$ is locally zero, yielding an ill conditioned estimation (see Fig. 14). Note that these are exactly the limits between the bands well posed for matching (Section 4). Assuming that a change of defocus/motion blur diameter mainly causes a scale change in $H(\vec{\nu})$, as in the case of the pillbox model (19), this phenomenon means that some frequencies will yield an unreliable contribution to the estimation. Still, a perturbation about a peak or trough will usually yield a bounded error since locally, the range of frequencies in which $\partial\hat{H}/\partial\hat{d} \approx 0$ is small.

Consider for example the peak about the DC. Substituting Eq. (32) in Eq. (34), and expanding H (Eq. (19)) in a Taylor series we obtain that

$$\left.\frac{\partial f_{\text{DFMB}}(\hat{u},\vec{\nu})}{\partial|N_0(\vec{\nu})|}\right|_{\nu d\cos\phi<1} \sim \frac{1}{D_0^2 D^2}\frac{1}{(\nu\cos\phi)^2}, \tag{43}$$

$$\left.\frac{\partial f_{\text{DFMB}}(\hat{u},\vec{\nu})}{\partial|N_D(\vec{\nu})|}\right|_{\nu d\cos\phi<1} \sim \frac{1}{D_0 D^3}\frac{1}{(\nu\cos\phi)^2}. \tag{44}$$

Eqs. (43) and (44) indicate that the estimation is very unstable in the low frequencies (Subbarao et al., 1997): the response to perturbations in 1D-DFD and DFMB behaves as ν^{-2} in the low frequencies, and thus these methods are more sensitive to noise in this regime than stereo (39), for which the response behaves as ν^{-1}. This is due to the fact that DFD/DFMB use summation of rays, and wide spatial perturbations affect them most. However, since enlarging the 1D aperture enables more light to reach the sensor, the signal is stronger and thus the estimation is more stable. Thus, DFD/DFMB may outperform stereo when the aperture (baseline) is large (compared to the pinhole reference). This is due to the fact that DFD relies on numerous rays for estimation. This additional data makes the estimation potentially more robust than simple discrete triangulation. Note that according to Eq. (41) there are certain frequencies for which a perturbation does not influence the estimation by DFD/DFMB. As with stereo, the response to perturbation of DFMB depends on the orientation ϕ of the spatial frequency, since the *aperture problem* exists also in motion.

5.4. 2D DFD

In comparison to stereo, motion and motion-blur systems of the same physical dimensions, 2D-DFD relies on much more points in the estimation of depth (Pentland, 1987; Pentland et al., 1989; Subbarao and Wei, 1992) and is thus potentially more reliable (Farid and Simoncelli, 1998) and robust. First of all, the amount of light gathered through the large aperture is proportional to D^2 (compared with $D_0 D$ for DFMB, and D_0^2 through a pinhole) making the signal to noise ratio (Farid and Simoncelli, 1998) much higher for large apertures. Eqs. (42) and (44) take the form

$$\begin{aligned}\frac{\partial f_{\text{DFD}}(\hat{u},\vec{\nu})}{\partial|N_D(\vec{\nu})|} &= -\frac{e^{j\vartheta(\vec{\nu})}}{|G(\vec{\nu})|}\frac{1}{D^2}\frac{f(\hat{u})}{\mathrm{J}_2[\pi\nu D f(\hat{u})]}\\ &\xrightarrow{\nu\to\infty} \frac{e^{j\vartheta(\vec{\nu})}}{|G(\vec{\nu})|}\frac{\pi f^{1.5}(\hat{u})}{\sqrt{2}}\frac{\sqrt{\nu}}{D^{1.5}\cos[(\pi\nu\hat{d})-\pi/4]},\\ &\xrightarrow{H(\nu)=0} \pm\frac{e^{j\vartheta(\vec{\nu})}}{|G(\vec{\nu})|}\frac{\pi f^{1.5}(\hat{u})}{\sqrt{2}}\frac{1}{D^{1.5}}\sqrt{\nu},\end{aligned} \tag{45}$$

$$\left.\frac{\partial f_{\mathrm{DFD}}(\hat{u},\vec{\nu})}{\partial|N_D(\vec{\nu})|}\right|_{\nu d<1} \approx \frac{e^{j\vartheta(\vec{\nu})}}{|G(\vec{\nu})|}\frac{4}{f(\hat{u})\pi^2 D^4}\frac{1}{\nu^2}. \tag{46}$$

To derive these relations we used

$$\frac{\partial[\mathrm{J}_1(\xi)/\xi]}{\partial\xi} = -\frac{\mathrm{J}_2(\xi)}{\xi} \tag{47}$$

and

$$\mathrm{J}_k(\xi) \overset{\xi\to\infty}{\longrightarrow} \sqrt{2/(\pi\xi)}\cos[\xi - k(\pi/2) - (\pi/4)], \tag{48}$$

for the circular pillbox model (21). As can be seen, in this model the error due to perturbations decreases faster with the aperture size, compared to the 1D triangulation methods. Here too there is instability at the very low frequencies. Indeed, to avoid the ill-posedness at the DC, in Watanabe and Nayar (1996) this component was nulled by band-pass filtering, and as a by-product the unstable contribution of the low-frequencies was suppressed.

For a circularly symmetric lens-aperture, the response is indifferent to the orientation of the frequency component. Hence the *aperture problem* does not exist.[4] This characteristic is valid also if the lens-aperture is not circularly symmetric, as long as it is sufficiently wide along both axes (the usual case). Hence, more frequencies (components of the images) may participate in the estimation by DFD and contribute stable and reliable information to the estimator. Therefore *DFD is potentially more robust than classic triangulation methods* if the system dimensions are the same.

The indifference of the transfer function to the orientation of the frequency components was utilized in Pentland (1987) and Pentland et al. (1989). In that work, DFD was implemented by comparing an image acquired via a circularly symmetric large aperture to a small ("pinhole") aperture image. Results were averaged over all orientations in the frequency domain, thus increasing the reliability of the estimation.

An example for the better robustness of DFD is the "monocular stereo" system presented in Simoncelli and Farid (1996), whose principle of operation is similar to that shown in Fig. 2. This was demonstrated in Farid (1997) and Farid and Simoncelli (1998). There, the same system was used for depth sensing once by differential DFD and once by differential stereo. The empirical results indeed show that the estimated depth fluctuations were significantly smaller in DFD than in stereo.

Note, that at high frequencies the estimation becomes unstable, at a moderate rate ($\sim\sqrt{\nu}$). However, for other models of the OTF, it might be much more severe. Consider for example a Gaussian kernel for DFD (Nayar, 1992; Rajagopalan and Chaudhuri, 1995; Subbarao, 1988; Surya and Subbarao, 1993). Accounting for the total light energy (as in Eqs. (15) and (16), the frequency response behaves like

$$\frac{H_D(\vec{\nu})}{H_0} = \frac{D^2}{D_0^2}e^{-[\kappa\nu D f(u)]^2}, \tag{49}$$

where κ is a constant (real). The response to the perturbation (38) is

$$\frac{\partial f_{\mathrm{gauss}}(\hat{u},\vec{\nu})}{\partial|N_D(\vec{\nu})|} = -\frac{e^{j\vartheta(\vec{\nu})}}{|G(\vec{\nu})|2f(\hat{u})\kappa^2 D^4}\frac{1}{\nu^2}e^{[\kappa\nu D f(\hat{u})]^2}, \tag{50}$$

which is very ill conditioned in the high frequencies. This situation is also schematically described in Fig. 14: if the slope of the frequency response from $\tilde{d}$ to ∞ is very small, the estimation error is unbounded.

5.5. The Optimal Axial Interval in DFD

In this subsection we refer to the method considered in Subsection 4.3, where the change between the two images is achieved by changing the focus settings, in particular the axial position of the sensor. Since the aperture D is the same for all images, we parameterize the transfer function by the blur diameter d in the equations to follow. Since the system has circular symmetry we use $H(\nu)$ instead of $H(\vec{\nu})$. Let one image be (in the frequency domain)

$$G_1(\nu) = G_0(\nu)H_d(\nu) + N_1(\nu), \tag{51}$$

where $N_1(\nu)$ is a perturbation while the other image is

$$G_2(\nu) = G_0(\nu)H_{d+\Delta d}(\nu). \tag{52}$$

If there is no perturbation, the two images should satisfy the constraint

$$G_2(\nu)H_d(\nu) - G_1(\nu)H_{d+\Delta d}(\nu) = 0. \tag{53}$$

We wish to estimate $\hat{d}$ by searching for the value that will satisfy

$$G_2(\nu)H_{\hat{d}}(\nu) - G_1(\nu)H_{\hat{d}+\Delta d}(\nu) = 0. \tag{54}$$

Similar to the discussion in Subsection 5.1, this can be satisfied for a single frequency signal. For other signals an error can be defined and minimized. Substituting Eqs. (51) and (52) into Eq. (54) yields

$$\begin{aligned} &H_{\hat{d}+\Delta d}(\nu)H_d(\nu) \\ &\quad = H_{\hat{d}}(\nu)H_{d+\Delta d}(\nu) - \frac{N_1(\nu)}{G_0(\nu)}H_{\hat{d}+\Delta d}(\nu). \end{aligned} \tag{55}$$

Assume for a moment that $H_d(\nu) \neq 0$, and define (as in Eq. (25))

$$H(\nu) = \frac{H_{d+\Delta d}(\nu)}{H_d(\nu)}, \qquad \hat{H}(\nu) = \frac{H_{\hat{d}+\Delta d}(\nu)}{H_{\hat{d}}(\nu)}. \tag{56}$$

Eq. (55) can be written as

$$\hat{H}(\nu) = H(\nu)\left[1 + \frac{N_1(\nu)}{G_0(\nu)H_d(\nu)}\right]^{-1}. \tag{57}$$

The perturbation causes the estimated transfer function to change:

$$\begin{aligned} \frac{\partial \hat{H}(\nu)}{\partial|N_1(\nu)|} &= -\frac{1}{\left[1 + \frac{N_1(\nu)}{G_0(\nu)H_d(\nu)}\right]^2}\frac{e^{j\vartheta(\nu)}}{|G_0(\nu)|}\frac{H_{d+\Delta d}(\nu)}{H_d^2(\nu)} \\ &\approx -\frac{e^{j\vartheta(\nu)}}{|G_0(\nu)|}\frac{H_{d+\Delta d}(\nu)}{H_d^2(\nu)}, \end{aligned} \tag{58}$$

where the approximation in the right hand side of Eq. (58) is for the case that $|N_1(\nu)|$ is small compared to $|G_0(\nu)H_d(\nu)|$. Similarly to Eq. (34) we seek the error induced by the perturbation on the depth estimation. For small perturbations we assume that $\hat{H}(\nu) \approx H(\nu)$, so

$$\begin{aligned} &\frac{\partial f(\hat{u},\nu)}{\partial|N_1(\nu)|} = \frac{\partial \hat{H}(\nu)}{\partial|N_1(\nu)|} \cdot \left[\frac{\partial \hat{H}(\nu)}{\partial f(\hat{u})}\right]^{-1} \\ &\approx -\frac{e^{j\vartheta(\nu)}}{|G(\nu)|D_0^2}\frac{H_{d+\Delta d}(\nu)}{D\frac{\partial H_{d+\Delta d}(\nu)}{\partial d}H_d(\nu) - D\frac{\partial H_d(\nu)}{\partial d}H_{d+\Delta d}(\nu)}. \end{aligned} \tag{59}$$

According to Eqs. (58) and (59), *if* $H_{d+\Delta d}(\nu) = 0$ *for this frequency, a perturbation* N_1 *does not affect the estimation*.

If $|H_d(\nu)| \ll |H_{d+\Delta d}(\nu)|$ we define the transfer function between the images as the reciprocal of Eq. (56):

$$H^{-1}(\nu) = \frac{H_d(\nu)}{H_{d+\Delta d}(\nu)}, \qquad \widehat{H^{-1}}(\nu) = \frac{H_{\hat{d}}(\nu)}{H_{\hat{d}+\Delta d}(\nu)}. \tag{60}$$

This takes care of the cases in which $H_d(\nu) = 0$ but $H_{d+\Delta d}(\nu) \neq 0$. Eq. (55) can be written as

$$\widehat{H^{-1}}(\nu) = H^{-1}(\nu) + \frac{N_1(\nu)}{G_0(\nu)H_{d+\Delta d}(\nu)}. \tag{61}$$

The perturbation causes the estimated transfer function to change:

$$\frac{\partial \widehat{H^{-1}}(\nu)}{\partial|N_1(\nu)|} = \frac{e^{j\vartheta(\nu)}}{|G_0(\nu)|H_{d+\Delta d}(\nu)}. \tag{62}$$

Calculating the influence on the depth estimation based on this transfer function, we arrive at the same relation as Eq. (59). Thus, we do not need to assume that $|N_1(\nu)|$ is small compared to $|G_0(\nu)H_d(\nu)|$.

In the pillbox model we use Eq. (23), and Eq. (59) takes a relatively simple form,

$$\begin{aligned} &\approx \frac{e^{j\vartheta(\nu)}}{2|G(\nu)|}\frac{f(u)}{D^2} \\ &\quad\times \frac{\mathrm{J}_1[\pi\nu(d+\Delta d)]}{\mathrm{J}_2[\pi\nu(d+\Delta d)]\mathrm{J}_1(\pi\nu d) - \mathrm{J}_2(\pi\nu d)\mathrm{J}_1[\pi\nu(d+\Delta d)]} \end{aligned} \tag{63}$$

which at the high frequencies (or defocus) becomes (48)

$$\begin{aligned} &\frac{\partial f(\hat{u},\nu)}{\partial|N_1(\nu)|} \\ &\quad \approx \frac{e^{j\vartheta(\nu)}}{|G(\nu)|}\frac{\pi d\sqrt{\nu d}}{D^3 2\sqrt{2}}\frac{\sin[\pi\nu(d+\Delta d) - (\pi/4)]}{\sin(\pi\nu\Delta d)}. \end{aligned} \tag{64}$$

A similar relation is obtained in case a perturbation N_2 is present in G_2 rather than in G_1:

$$\begin{aligned} \frac{\partial f(\hat{u},\nu)}{\partial|N_2(\nu)|} &\approx -\frac{e^{j\vartheta(\nu)}}{|G(\nu)|}\frac{\pi(d+\Delta d)\sqrt{\nu(d+\Delta d)}}{D^3 2\sqrt{2}} \\ &\quad \times \frac{\sin[\pi\nu d - (\pi/4)]}{\sin(\pi\nu\Delta d)}. \end{aligned} \tag{65}$$

To appreciate the significance of Eqs. (64) and (65), observe that the reliability of the defocus estimation at high frequencies is optimized (for unknown u, hence for unknown d) if

$$|\nu \Delta d| = 0.5, 1.5, 2.5 \ldots \tag{66}$$

Then, the magnitude of the term $\sin(\pi \nu \Delta d)$ in the denominator is maximal, minimizing the effect of the perturbation on the estimation $\hat{d} = Df(\hat{u}, \nu)$. Thus, if DFD is achieved by changing the focus settings, the change (e.g. the axial movement of the sensor) is optimized if it causes the blur-diameter to change according to Eq. (66), where ν is the high frequency of choice. Alternatively, *if Δd is given, Eq. (66) indicates the optimal frequencies around which the depth estimation would be done.*

On the other hand, if

$$|\nu \Delta d| = 1, 2, 3 \ldots \tag{67}$$

the denominator of Eqs. (64) and (65) is nulled. In this situation the estimation is highly ill-conditioned. Note that as the axial interval is increased, hence Δd is increased, for a given scene, the number of problematic components that satisfy Eq. (67) is increased (as well as the number of useful frequency components that satisfy Eq. (66)).

The optimal Δd was used in Figs. 10 and 11. Note that at high frequencies Bessel functions resemble a cosine function, and the two functions (Fig. 10) are out of phase by $\pi/2$. Hence, in this situation extrema of the H_d are at zero-crossings of $H_{d+\Delta d}$, and vice-versa, yielding the maximum changes in the ratio between these functions. On the other hand, if Eq. (67) is satisfied, at the high frequencies the functions of Fig. 10 have a ratio of $\approx \pm 1$ for all blur-diameters, except for the zero crossings where the ratio is not defined. Thus, the transfer function between the images is "indifferent" to the exact blur diameter, and thus does not provide a good estimation.

In Subsection 4.3 we noted that if $\nu \Delta d$ is small, the lowest band without ambiguities is $0 < \nu d < 1.63$ but that this band becomes narrower if Δd increases. If we use the guideline[5] of Eq. (66), Fig. 11 (where $\nu \Delta d = 0.5$) shows that for unambiguous estimation

$$0 < \nu d < 1.46. \tag{68}$$

This result too is supported by the tests performed in Watanabe and Nayer (1996). Although the authors noticed that range of unambiguous solutions can be extended to $\nu d = 1.6$, for reasons of numerical stability (measured by the behavior of the Newton-Raphson algorithm that was used for estimation), the frequency band limit was actually set in Watanabe and Nayer (1996) to $\nu d = 1.46$ (i.e., $\nu r = 0.73$). Within this band the results came out to be unique and stable, while beyond it the range estimation became unstable. Note that this is in excellent agreement with Eq. (68)!

An important application of Eq. (66) is to show a new aspect of depth of field. Suppose that the highest frequency in the image is $\nu_{\max} = 1/(2\Delta x)$ where Δx is the inter-pixel period of the sensor (the Nyquist rate). Since $\nu \le \nu_{\max}$, Eq. (66) yields

$$\Delta d \ge \frac{1}{2}\left(\frac{1}{2\Delta x}\right)^{-1} = \Delta x. \tag{69}$$

So, in order to obtain reliable results, it is preferable to sample the axial position so that the change in the blur-diameter is at least one inter-pixel spacing. However, to avoid instability at any frequency, we should avoid Eq. (67) and thus require that $\nu_{\max}\Delta d < 1$. Hence *the safe and optimal range of change of the focus settings is such that the blur diameter change is bounded by*

$$\Delta x \le \Delta d < 2\Delta x. \tag{70}$$

With the threshold $\Delta d = \Delta x$, if one of the images is in focus (having $d = 0$), the blur kernel at the other image will have a diameter of $d_{\text{th}} = 0 + \Delta d = \Delta x$. This threshold diameter determines the depth of field of the system (the threshold of $\tilde{u} - u$) by the geometric relation (1). Thus, using a Δd which is smaller than the threshold given in Eq. (69) is an attempt to sense defocus or change of defocus smaller than the uncertainty imposed by the DOF. As noted above, optimality with respect to noise sensitivity is achieved only above the threshold. Hence, *sampling the axial position in DOF intervals (for which $\Delta d = \Delta x$) is optimal with respect to robustness to perturbations at the Nyquist frequency.* Changing the focus settings in a smaller interval means that no frequency in the image will satisfy the optimality condition (66). Changing the focus settings in a larger interval will be sub-optimal for the Nyquist frequency, but will be optimal for some lower frequency. If the interval of the axial position is twice than the DOF or more, estimation based on some frequencies will be very unstable (67).

5.6. DFF

d_{th} determines the DOF of the DFF system (see the geometric relation (1)). Note that in the same manner we can define the "DOF for vergence", which is the amount of axial displacement without detectable disparity. The latter DOF is related to d_{th} by Eq. (3), and is thus the same as the DOF of the DFF system, for the same system dimensions. To sample the depth efficiently[6] the image slices should be taken at DOF intervals (Abbott and Ahuja, 1988; Abbott and Ahuja, 1993). In this situation, the highest frequencies in the image are detectably affected by defocus.

For transfer functions (between an image and a reference image) which change scale with d (including stereo, the pillbox model and the Gaussian model), the least detectable blur-diameter/disparity satisfies

$$d_{\text{th}}(\nu) \propto \frac{1}{\nu}, \tag{71}$$

for 1D images. It is clear that for low frequencies the blur-diameter/disparity has to be larger in order to be detected ($d_{\text{th}}(\nu) > d_{\text{th}}$). Thus if we sample the scene efficiently, the frequencies below $\nu_{\max}$ will yield results which are within the inherent uncertainty of the system and are thus ineffective.

For 2D images the DOF of the DFF system is rotation-invariant. For all ϕ

$$d_{\text{th}}^{\text{DFF}}(\vec{\nu}) = d_{\text{th}} \frac{\nu_{\max}}{\nu}, \tag{72}$$

In stereo only the frequency component along the baseline changes between the frames:

$$d_{\text{th}}^{\text{stereo}}(\vec{\nu}) = d_{\text{th}} \frac{\nu_{\max}}{\nu \cos\phi}. \tag{73}$$

Thus, for frequency orientations not parallel to the baseline, the "DOF for vergence" (as defined above) is larger than that of DFF (the *aperture problem*).

In critical sampling, the only frequency components for which defocus/disparity will be detected are those with $\nu = \nu_{\max}$. However, comparing Eqs. (72) and (73), in stereo, all the frequencies yield results which are within the inherent uncertainty of the measurement and are thus ineffective, except for $\cos\phi = \pm 1$. For DFF, all ϕ yield reliable results. Hence, DFF allows more frequencies $\vec{\nu}$ to reliably participate in the detection of depth deviation, leading to a more reliable depth estimation.

6. Conclusions

We have shown that, in principle, the sensitivities of Depth from Focus and Defocus techniques are not inferior but similar to those of stereo and motion based methods. The apparent differences are primarily due to the difference in the size of the physical setups. This also accounts for the fact that matching (correspondence) problems are uncommon in DFD and DFF. The "absence" of the occlusion problem in DFD and DFF is not a fundamental feature and is mostly a consequence of the small aperture ("baseline") that is normally used. Stereo systems having a similar level of immunity can be constructed.

The observation that physical size (baseline in stereo, aperture size in DFD/DFF) determines the characteristics of various range imaging approaches in a similar manner is important in performance evaluation of depth sensing algorithms. It indicates that performance results should be scaled according to setup dimensions. As long as enlarging the baseline is cheaper than enlarging the lens aperture (beyond a few centimeters), stereo will remain the superior approach in terms of resolution/cost. Improvements of DFD/DFF by algorithmic developments is limited in common implementations by the small aperture size.

The monocular structure of DFD/DFF systems does not ensure the avoidance of occlusion and matching problems. Adelson and Wang (1992) formalized the correspondence problem in the frequency domain. They have shown that in stereo it is a manifestation of aliasing, since the transfer function between the stereo images is not one-to-one. Matching problems in DFD arise due to the same reason. There are scenes for which the solution of depth estimation by DFD (i.e. matching blur kernels to image pairs) is not unique. Moreover, for the same system dimensions, common triangulation techniques, such as stereo, may be less prone to matching ambiguity than DFD. A coarse to fine approach may resolve the matching problem in a way analogous to a method used in stereo and motion (Irani et al., 1994). In this way frequencies that are "lost" (Watanabe and Nayar, 1996) can be used. Unlike DFD (and stereo), DFF seems indeed to be immune to matching ambiguities, if the evaluation patch of the focus measure is larger than the support of the widest blur-kernel expected, and if the depth is homogeneous in that patch.

In contrast to common belief, for the same system dimensions the chance of occurrence of the occlusion phenomenon is higher in DFD/DFF than in stereo or motion. However, DFD/DFF are more stable in the presence of such disruptions. Note that in the presence of severe occlusion, straightforward DFF may yield double valued depth. A layered scene model resolves this ambiguity.

We analyzed the effect of additive perturbations by examining their influence in each spatial frequency component of the images. An estimation that relies on some frequency components yields stable results, while the contribution of other frequencies is very sensitive to perturbations. A possible future research may be on algorithms that rely on a coarse estimate of the disparity/blur-diameter to select the optimal spatial frequencies (for which the response to perturbations is very small) to obtain a better estimate. In DFD, if the frequency selected for the estimation is ν, the axial movement of the sensor is optimal if it causes the change Δd in the blur diameter to satisfy $|\nu \Delta d| = 0.5, 1.5, 2.5 \ldots$. Sampling the axial position in DOF intervals is optimal with respect to robustness to perturbations. Using an interval which is twice or more than that, may yield unstable results.

Our analysis of the response to perturbations is deterministic and is based on the assumption that a perturbation exists only in a single frequency. In order to extend this analysis to the general case, and obtain the response to noise, a stochastic analysis, based on the deterministic results derived here, is needed.

The two dimensionality of the aperture is the principal difference between DFD/DFF and conventional triangulation methods. It allows much more image points to contribute to the depth estimation and the higher light energy that passes the large-aperture lens leads to a higher signal to noise ratio. This difference accounts for the inherent robustness of methods that rely on depth of field. In this respect DFF and DFD methods are also superior to Depth from Motion Blur. Specifically, the insensitivity to the orientation of features in DFD/DFF provides higher flexibility in the depth estimation process. Another advantage of DFD that follows from the two dimensionality of the PSF is that full depth recovery may be possible in the presence of slight occlusion. A practical implication of the advantages of methods that are based on DOF is that if the full resolution potential of stereo imaging is not needed, and the resolution obtainable with common DFD/DFF implementations is sufficient, DFD/DFF should be preferred over small baseline stereo.

The analysis of the depth estimation methods done in this work was based solely on geometrical optics, and is thus valid for setups (i.e., objects and systems) in which diffraction effects are not dominant. In particular, it does not apply to microscopic DFF. A more rigorous analysis requires the consideration of physical optics (e.g., diffraction). Doing the analysis in systems based on depth of field is straightforward. However, in comparison to stereo or motion, we should note that geometric triangulation methods have traditionally been based on the geometric optics approximation. Therefore, for a full derivation of the relations between DFD/DFF and stereo, a model for the diffraction effects in triangulation has to be developed. Note also that the comparison was based on the assumption of small angles (paraxial optics) in the imaging setup. It would be beneficial to extend this work to the general case. In particular, the characteristics of the epipolar geometry, and the space-varying transfer function between the images may provide new points of view in the comparison between DFD and stereo. Another possible generalization is to analyze DFD when the two images are taken with a fixed focus setting, but with different apertures of which none is a pinhole.

Acknowledgments

The authors wish to thank Rafael Piestun and Joseph Shamir for stimulating discussions. This research was supported in part by the Eshkol Fellowship of the Israeli Ministry of Science, by the Ollendorff Center of the Electrical Engineering Department, Technion, and by the Tel Aviv University Internal Research Fund.

Notes

1. Some improvement can be achieved by super-resolution techniques.
2. Cases of severe occlusions, where the chief ray is occluded, are ignored, since in this case the object point is not seen in the pinhole image, thus the depth of the occluder will be measured.
3. A system based on *circular* motion blur (Kawasue et al., 1998) was recently presented. When the object po ints are sparse, this method is analogous to the ring defocus PSF of Rioux and Blais (1986).
4. This immunity is also shared by “stereo” systems having vertical parallax as well as a horizontal one. However, these require at least three images to be acquired and processed, in contrast to DFD which requires two images. We therefore do not deal with such systems. Nevertheless, this problem can be avoided by nonlinear camera trajectory (in DFMB), as used in Kawasue et al. (1998).
5. This guideline is approximate for the low frequencies and exact for the high ones.
6. Note that according to the conclusion in Subsection 5.5, these intervals do not only make the sampling efficient for DFF but also best for reliable estimation in DFD.

References

Abbott, A.L. and Ahuja, N. 1988. Surface reconstruction by dynamic integration of focus, camera vergence, and stereo. In *Proc. ICCV*, Tarpon Springs, Florida, pp. 532–543.

Abbott, A.L. and Ahuja, N. 1993. Active stereo: Integrating disparity, vergence, focus, aperture and calibration for surface estimation. *IEEE Trans. PAMI*, 15:1007–1029.

Adelson, E.H. and Wang, J.Y.A. 1992. Single lens stereo with a plenoptic camera. *IEEE Trans. PAMI*, 14:99–106.

Amitai, Y.Y., Friesem, A.A., and Weiss, V. 1989. Holographic elements with high efficiency and low aberrations for helmet displays. *App. Opt.*, 28:3405–3417.

Asada, N., Fujiwara, H., and Matsuyama, T. 1998. Seeing behind the scene: Analysis of photometric properties of occluding edges by the reversed projection blurring model. *IEEE Trans. PAMI*, 20:155–167.

Bergen, J.R., Burt, P.J., Hingorani, R., and Peleg, S. 1992. A three-frame algorithm for estimating two-component image motion. *IEEE Trans. PAMI*, 14:886–895.

Besl, P.J. 1988. Active, optical range imaging sensors. *Machine Vision and Applications*, 1:127–152.

Bove Jr., V.M. 1989. Discrete fourier transform based depth-from-focus, Image Understanding and Machine Vision 1989. *Technical Digest Series*, 14, Conference ed., 118–121.

Bove Jr., V.M. 1993. Entropy-based depth from focus. *J. Opt. Soc. Amer. A*, 10:561–566.

Castleman, K.R. 1979. *Digital Image Processing*. Prentice-Hall: New Jersey, pp. 357–360.

Chen, W.G., Nandhakumar, N., and Martin, W.N. 1996. Image motion estimation from motion smear—a new computational model. *IEEE Trans. PAMI*, 18:412–425.

Darrell T. and Wohn, K. 1988. Pyramid based depth from focus. In *Proc. CVPR*, Ann Arbor, pp. 504–509.

Darwish, A.M. 1994. 3D from focus and light stripes. In *Proc. SPIE Sensors and Control for Automation*, Vol. 2247, Frankfurt, pp. 194–201.

Dias, J., de Araujo, H., Batista, J., and de Almeida, A. 1992. Stereo and focus to improve depth perception. In *Proc. 2nd Int. Conf. on Automation, Robotics and Comp. Vis.*, Vol. 1, Singapore, pp. cv-5.7/1–5.

Engelhardt, K. and Hausler, G. 1988. Acquisition of 3-D data by focus sensing. *App. Opt.*, 27:4684–4689.

Ens J. and Lawrence, P. 1993. An investigation of methods for determining depth from focus. *IEEE Trans. PAMI*, 15:97–108.

Farid, H. 1997. Range estimation by optical differentiation. Ph.D Thesis, University of Pennsylvania.

Farid, H. and Simoncelli, E.P. 1998. Range estimation by optical differentiation. *JOSA A*, 15:1777–1786.

FitzGerrell, A.R., Dowski, Jr., E.R., and Cathey, T. 1997. Defocus transfer function for circularly symmetric pupils. *App. Opt.*, 36:5796–5804.

Fox, J.S. 1988. Range from translational motion blurring. In *Proc. CVPR*, Ann Arbor, pp. 360–365.

Girod, B. and Scherock, S. 1989. Depth from defocus of structured light. In *Proc. SPIE Optics, Illumination and Image Sensing for Machine Vision IV*, Vol. 1194, pp. 209–215. TR-141, Media-Lab, MIT.

Hiura, S., Takemura G., and Matsuyama, T. 1998. Depth measurement by multi-focus camera. In *Proc. of Model-Based 3D Image Analysis*, Mumbai, pp. 35–44.

Hopkins, H.H. 1955. The frequency response of a defocused optical system. *Proc. R. Soc. London Ser. A*, 231:91–103.

Hwang, T., Clark, J.J., and Yuille, A.L. 1989. A depth recovery algorithm using defocus information. In *Proc. IEEE CVPR*, San-Diego, pp. 476–482.

Irani, M., Rousso, B., and Peleg, S. 1994. Computing occluding and transparent motions. *Int. J. Comp. Vis.*, 12:5–16.

Jarvis, R. 1983. A perspective on range-finding techniques for computer vision. *IEEE Trans. PAMI*, 3:122–139.

Kawasue, K., Shiku, O., and Ishimatsu, T. 1998. Range finder using circular dynamic stereo. In *Proc. ICPR*, Vol. I, Brisbane, pp. 774–776.

Klarquist, W.N., Geisler, W.S., and Bovic, A.C. 1995. Maximum-likelihood depth-from-defocus for active vision. In *Proc. Inter. Conf. Intell. Robots and Systems: Human Robot Interaction and Cooperative Robots*, Vol. 3, Pittsburgh, pp. 374–379.

Kristensen, S., Nielsen, H.M., and Christensen, H.I. 1993. Cooperative depth extraction. In *Proc. Scandinavian Conf. Image Analys*, Vol. 1, Tromso, Norway, pp. 321–328.

Krotkov, E. and Bajcsy, R. 1993. Active vision for reliable ranging: Cooperating focus, stereo, and vergence. *Int. J. Comp. Vis.*, 11:187–203.

Lee, H.C. 1990. Review of image-blur models in a photographic system using the principles of optics. *Opt. Eng.*, 29:405–421.

Marapane, S.B. and Trivedi, M.M. 1993. An active vision system for depth extraction using multi-primitive hierarchical stereo analysis and multiple depth cues. In *Proc. SPIE Sensor Fusion and Aerospace Applications*, Vol. 1956, Orlando, pp. 250–262.

Marshall, J.A., Burbeck, C.A., Ariely, D., Rolland, J.P. and Martin, K.E. 1996. Occlusion edge blur: A cue to relative visual depth. *J. Opt. Soc. Amer. A*, 13:681–688.

Nair, H.N. and Stewart, C.V. 1992. Robust focus ranging. In *Proc. CVPR*, Champaign, pp. 309–314.

Nayar, S.K. 1992. Shape from focus system. In *Proc. CVPR*, pp. 302–308.

Nayar, S.K., Watanabe, M., and Nogouchi, M. 1995. Real time focus range sensor. In *Proc. ICCV*, Cambridge, pp. 995–1001.

Noguchi M. and Nayar, S.K. 1994. Microscopic shape from focus using active illumination. In *Proc. ICPR-A*, Jerusalem, pp. 147–152.

Pentland, A.P. 1987. A new sense for depth of field. *IEEE Trans. PAMI*, 9:523–531.

Pentland, A.P., Darrell, T., Turk, M., and Huang, W. 1989. A simple, real-time range camera. In *Proc. CVPR*, San Diego, pp. 256–261.

Pentland, A., Scherock, S., Darrell T., and Girod, B. 1994 Simple range camera based on focal error. *J. Opt. Soc. Amer. A*, 11:2925–2934.

Rajagopalan, A.N. and Chaudhuri, S. 1995a. A block shift-invariant blur model for recovering depth from defocused images. In *Proc. ICIP*, Washington DC, pp. 636–639.

Rajagopalan, A.N. and Chaudhuri, S. 1995b. Recovery of depth from defocused images. In *Proc. 1st National Conference on Communications*, pp. 155–160.

Rajagopalan, A.N. and Chaudhuri, S. 1997. Optimal selection of camera parameters for recovery of depth from defocused images. In *Proc. CVPR*, San Juan, pp. 219–224.

Rioux, M. and Blais, F., 1986. Compact three-dimensional camera for robotic applications. *JOSA A*, 3:1518–1521.

Saadat, A. and Fahimi, H. 1995. A simple general and mathematically tractable way to sense depth in a single image. In *Proc.*

SPIE Applications of Digital Image Processing XVIII, Vol. 2564, San Diego, pp. 355–363.

Schechner, Y.Y. and Kiryati, N., 1998. Depth from defocus vs. Stereo: How different really are they? In *Proc. ICPR*, Brisbane, pp. 1784–1786.

Schechner, Y.Y. and Kiryati, N. 1999. The optimal axial interval in estimating depth from defocus. In *Proc. IEEE ICCV*, Vol. II, Kerkyra, pp. 843–848.

Schechner, Y.Y., Kiryati, N., and Basri, R. 1998. Separation of transparent layers using focus. In *Proc. ICCV*, Mumbai, pp. 1061–1066.

Scherock, S. 1991. Depth from defocus of structured light. TR-167, Media-Lab, MIT.

Schneider, G., Heit, B., Honig, J., and Bremont, J. 1994. Monocular depth perception by evaluation of the blur in defocused images. In *Proc. ICIP*, Vol. 2, Austin, pp. 116–119.

Simoncelli, E.P. and Farid, H. 1996. Direct differential range estimation using optical masks. In *Proc. ECCV*, Vol. 2, Cambridge, pp. 82–93.

Stewart, C.V. and Nair, H. 1989. New results in automatic focusing and a new method for combining focus and stereo. In *Proc. SPIE Sensor Fusion II: Human and Machine Strategies*, Vol. 1198, Philadelphia, pp. 102–113.

Subbarao, M. 1988. Parallel depth recovery by changing camera parameters. In *Proc. ICCV*, Tarpon Springs, Florida, pp. 149–155.

Subbarao, M. and Liu, Y.F. 1996. Accurate reconstruction of three-dimensional shape and focused image from a sequence of noisy defocused images. In *Proc. SPIE Three dimensional imaging and Laser-Based Systems For Metrology and Inspection II*, Vol. 2909, Boston, pp. 178–191.

Subbarao M. and Surya, G. 1994. Depth from defocus: A spatial domain approach. *Int. J. Comp. Vis.*, 13:271–294.

Subbarao, M. and Wei, T.C. 1992. Depth from defocus and rapid autofocusing: A practical approach. In *Proc. CVPR*, Champaign, pp. 773–776.

Subbarao, M., Yuan, T., and Tyan, J.K., 1997. Integration of defocus and focus analysis with stereo for 3D shape recovery. In *Proc. SPIE Three Dimensional Imaging and Laser-Based Systems for Metrology and Inspection III*, Vol. 3204, Pittsburgh, pp. 11–23.

Surya, G. and Subbarao, M. 1993. Depth from defocus by changing camera aperture: A spatial domain approach. In *Proc. CVPR*, New York, pp. 61–67 .

Swain, C., Peters, A., and Kawamura, K. 1994. Depth estimation from image defocus using fuzzy logic. In Proc. 3rd International Fuzzy Systems Conference, Orlando, pp. 94–99.

Watanabe, M. and Nayar, S.K. 1996. Minimal operator set for passive depth from defocus. In *Proc. CVPR*, San Francisco, pp. 431–438.

Xiong, Y. and Shafer, S.A. 1993. Depth from focusing and defocusing. In *Proc. CVPR*, New York, pp. 68–73.