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

Yoav Y. Schechner and Nahum Kiryati
Department of Electrical Engineering, Technion, Haifa 32000, ISRAEL
yoavs@tx.technion.ac.il nk@ee.technion.ac.il

Abstract

Depth from Focus (DFF) and Depth from Defocus (DFD) methods are shown to be realizations of 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. 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.

1. Introduction

In recent years range estimation based on the limited depth of field (DOF) of lenses has been gaining popularity. These methods are normally considered to be in a separate class, distinguished from triangulation techniques such as depth from stereo, vergence or motion [3, 4, 7, 14]. Successful application of computer vision algorithms requires sound performance evaluation and comparison of various approaches. The comparison of range sensing systems that rely on different principles of operation and have a wide range of physical parameters is not easy [3]. 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 [3, 12].

2. Unlike triangulation methods, DFD avoids the matching (correspondence) ambiguity and occlusion problems [10, 14, 15].

3. DFD is reliable [10, 11, 12, 15].

Similar statements were made with regard to Depth from Focus (DFF) [4, 7]. A major step towards understanding the relations between triangulation and DOF has been recently taken in [5]. A wide lens was utilized to build a "monocular stereo" system, with sensitivity that has the same functional dependence on system parameters as in stereo.

We show that the difference between DFD/DFF and "classic" triangulation techniques (stereo, vergence, motion) is not a fundamental one. In fact, we claim that DFD and DFF can be regarded as ways to achieve triangulation. Our analysis indicates that the origins of the observations in first two statements above are primarily in the physical size differences between the common implementations of focus and triangulation based approaches, not in the fundamentals. Generally, these statements do not hold. In contrast, the third observation has a solid foundation. DFF and DFD rely on more data than common discrete triangulation methods, and are thus potentially more reliable. Ref. [13] is a full version of this paper.

2. Sensitivity

Consider the imaging system sketched in Fig. 1(a). The sensor at distance $\tilde{v}$ behind the lens can image in-focus a point at distance $\tilde{u}$ in front the lens. An object point at distance u is defocused, and its image in the sensor plane is a blur-circle of diameter d. For this system [14]

$$d = 2r = D\frac{|uF - \tilde{v}u + F\tilde{v}|}{Fu} \quad , \qquad (1)$$

where F is the focal length 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.

If the lens is blocked except for two pinholes on its perimeter, on opposite ends of some diameter [1], only two rays pass the lens (Fig. 1(b)). The geometrical point spread

This research was carried out in the Ollendorff Center of the Department of Electrical Engineering, Technion.

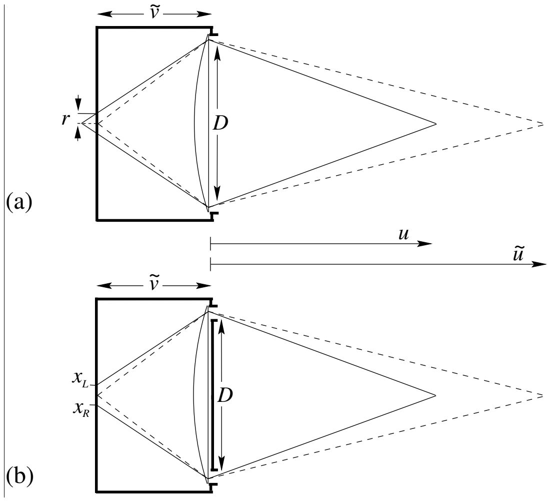

Figure 1. (a) The imaging system is tuned to view in focus object points at distance $\tilde{u}$. (b) The lens is blocked except for two pinholes on opposite ends of its diameter. The image of a defocused object point is two points.

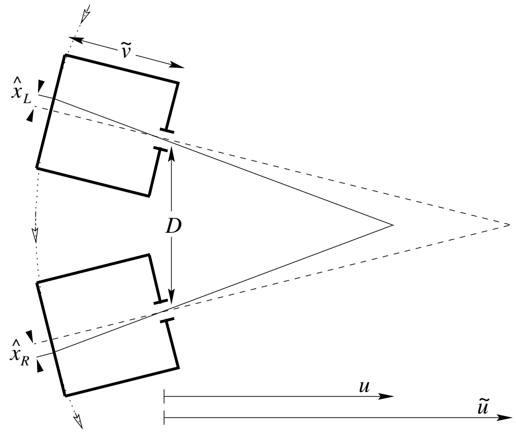

Figure 2. A stereo system with a baseline D equal to the lens diameter in Fig. 1, and distance $\tilde{v}$ from the entrance pupil to the sensor that is also the same. Motion along the arc is analogous to defocus blur.

function (PSF) now consists of only two points, x_L and x_R. The distance between the points is $|x_R - x_L| = 2r$.

Consider now the stereo & vergence system shown in Fig. 2 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 diameter of the large aperture lens, and the sensors are at the same distance $\tilde{v}$ behind the pinholes. The vergence eliminates the disparity for the object point at distance $\tilde{u}$. The image of an object point at u is again two points, now one on each sensor. It can be shown [13] that the disparity is

$$|\hat{x}_R - \hat{x}_L| = |x_R - x_L| = 2r = d \ , \qquad (2)$$

where we exploited the fact that the angles are small. 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.

The sensitivity (and resolution) of the triangulation and DFD systems are equivalent and related to the disparity/PSF-support size (2): Depth deviation from focus is sensed if this value is larger than the pixel period [4]. Hence, *DFD and DFF are not inherently less sensitive than stereo or motion.* In practice, typical lens apertures used are merely in the order of $\approx 1cm$ while stereo baselines are usually one or two orders of magnitude larger, leading to a proportional increase in sensitivity. However, using holographic optical elements [2] rather than conventional optics, can increase the aperture size (and performance) of DFD/DFF by an order of magnitude.

3. Occlusion and matching problems

The common observation that monocular methods avoid the missing parts (occlusion) problem is mostly a consequence of the small "baseline" associated with the lens. The small angles involved reduce the chance that a point will be visible to a part of the lens while being occluded at another part (vignetting caused by the scene [8]). Note that the same applies to stereo (or motion) with the same baseline! The small aperture also accounts for the fact that matching problems are uncommon in DFD. If the stereo system is built with a small baseline as in common monocular systems, the correspondence problem will mostly be avoided [1]. An example is the "monocular stereo" system presented in [5], whose principle of operation is similar to that shown in Fig. 1(b). There is, of course, no "free lunch": The avoidance of the correspondence and occlusion problems by decreasing the baseline leads to a reduction in sensitivity.

To analyze cases of occlusion in DFD, let us first note that the principle of operation of Depth from Motion Blur (DFMB) [6] 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, from which depth is extracted. If a camera moves along an arc of radius $\tilde{u}$, with its optical axis pointing towards the center of the circle (Fig. 2), a point at a distance $\tilde{u}$ remains unblurred (analogous to being focused), while the scene is generally blurred.

The stereo PSF consists of two distinct impulses (Figs. 1(b),2), separated by d. The PSF of the DFMB system is a 1D pillbox, stretched between those impulses. The defocus blur PSF is a disc of diameter d. It thus has a much larger support than the PSFs of stereo or motion, and a larger chance of being partially occluded. *For the same* *system dimensions, the chance of occlusion in DFD/DFF is larger than in stereo or motion*. However, due to this large support, DFD/DFF are more stable to such disruptions [13].

DFD and DFF are based on point-to-patch or patch-to-patch comparisons. To estimate depth at given image coordinates, comparing just the points having those coordinates in the images is insufficient. It is possible to estimate the support of the blur kernel for piecewise planar scenes [14] 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. Thus depth should be homogeneous within patches which are at least as large as the widest blur kernel expected, otherwise *edge-bleeding* may occur [9]. In stereo too, patches used for correspondence establishment should be larger than the disparity and of homogeneous depth, to enable registration.

With sufficiently large patches, frequency domain analysis of stereo and DFD is possible. Adelson and Wang [1] showed that the stereo matching problem is a manifestation of aliasing. This is since the transfer function between the stereo images, $\exp(j2\pi\nu d)$, is not one-to-one outside the band $0 < d\nu < 1$. The transfer function of the 2D pillbox model, $\mathrm{J}_1(\pi d\nu)/(\pi d\nu)$ is not one-to-one beyond the band $0 < d\nu < 1.63$, i.e., a measured attenuation in DFD may be the possible outcome of several blur diameters [13]. *There are scenes for which the solution of DFD (matching blur kernels in image pairs) is not unique.* Thus, matching ambiguity occurs in DFD in a similar manner to stereo.

4. Robustness

In contrast to stereo, the (implied) triangulation in DFMB, DFD and DFF is not done solely with the two marginal points (rays), but with a continuum of points. This additional data makes the estimation potentially more robust than simple discrete triangulation. Due to the two dimensionality of the lens aperture (and blur), DFD/DFF rely on even more points (further increasing the robustness potential) than motion and motion-blur (let alone stereo). Since the 2D aperture gathers more light than the stereo "pinholes", the signal to noise ratio in the raw images is increased. Moreover, only the projection of spatial frequencies onto the baseline is affected by stereo or motion disparity. This is the source of the *aperture problem*. On the other hand, defocus blur attenuates in the same manner all spatial frequencies, regardless of their orientation. Thus more frequency components of the images yield stable contribution to the estimation by DFD/DFF than by stereo or motion [13].

In [5] depth was estimated once by DFD and once by differential stereo using the same system dimensions (specifically, the baseline was equal to the lens aperture size). The empirical results indeed show that the estimated depth fluctuations were significantly smaller in DFD than in stereo.

5. Discussion

Physical size (stereo baseline / DFD aperture) determines the characteristics of DFD/DFF in the same manner as in stereo. Thus, when evaluating the performance of depth sensing algorithms, results should be scaled according to setup dimensions. Matching (correspondence) ambiguity and occlusion appear in DFD similarly to stereo, for the same system dimensions. However, DFD/DFF are more robust due to the 2D nature of the aperture, and should thus be preferred over small baseline stereo, if the resolution obtainable with common DFD implementations suffices.

Our analysis is based solely on geometrical optics, and is thus valid for objects and systems in which diffraction effects are not dominant, (e.g., it does not apply to microscopic DFF). Nevertheless, the geometrical triangulation methods (such as stereo) rely on this approximation as well (most notably be the extensive use of the pinhole camera model). The relation between DFF/DFD and depth from stereo, taking diffraction into account, has yet to be studied.

References

[1] E. M. Adelson and J. Y. A. Wang. Single lens stereo with plenoptic camera. *IEEE Trans. on PAMI*, 14:99–106, 1992.

[2] A. A. Amitai, Y. Y. Friesem and V. Weiss. Holographic elements with high efficiency and low aberrations for helmet displays. *App. Opt.*, 28:3405–3416, 1989.

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

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

[5] H. Farid and E. P. Simoncelli. Range estimation by optical differentiation. *To be published in JOSA A*, 1998; PhD thesis of H. Farid, University of Pennsylvania, 1997.

[6] J. S. Fox. Range from translational motion blurring. In *Proc. CVPR*, pp. 360–365, 1988.

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

[8] J. A. Marshall, C. A. Burbeck, D. Ariely, J. P. Rolland and K. E. Martin. Occlusion edge blur: a cue to relative visual depth. *JOSA A*, 13:681–688, 1996.

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

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

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

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

[13] Y. Y. Schechner and N. Kiryati. Depth from Defocus vs. stereo: How different really are they? Technical Report EE-PUB-1155, Technion - Israel Inst. Tech., May 1998.

[14] S. Scherock. Depth from defocus of structured light. Technical Report TR-167, Media-Lab, MIT, 1991.

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