A depth estimation algorithm with a single image

V. Aslantas

Erciyes University, Engineering Faculty, Computer Engineering Division 38039 Kayseri, Turkey.
aslantas@erciyes.edu.tr

Abstract: This paper explains the use of a sharpening filter to calculate the depth of an object from a blurred image of it. It presents a technique which is independent of edge orientation. The technique is based on the assumption that a defocused image of an object is the convolution of a sharp image of the same object with a two-dimensional Gaussian function whose spread parameter (SP) is related to the object depth. A sharp image of an object is obtained from a defocused image of the same object by applying sharpening filters. The defocused and sharp images of the object are used to calculate the SP which is then related to the object depth. The paper gives experimental results which show the feasibility of employing sharpening filters for depth computation.

References and links

1. M. Hebert, "Active and passive range sensing for robotics," in *Proceedings of IEEE Conference on Robotics and Automation*, (Institute of Electrical and Electronics Engineers, San Francisco, CA, 2000), pp. 102-110.
2. P. Grossmann, "Depth from focus," Pattern Recogn. Lett. **5**, 63-69 (1987).
3. A. P. Pentland, "A new sense for depth of field," IEEE Trans. Pattern Anal. Mach. Intell. **9**, 523-531 (1987).
4. M. Subbarao and N. Gurumoorthy, "Depth recovery from blurred edges," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, (Institute of Electrical and Electronics Engineers, Ann Arbor, MI, 1988), pp.498-503.
5. M. Subbarao, "Efficient depth recovery through inverse optics," *Machine Vision Inspection and Measurement*, H. Freeman ed., (Academic, Boston, 1989).
6. C. Cardillo and M. A. Sid-Ahmed, "3-D position sensing using passive monocular vision system," IEEE Trans. Pattern Anal. Mach. Intell. **13**, 809-813 (1991).
7. R. V. Dantu, N. J. Dimopoulos, R. V. Patel, and A. J. Al-Khalili, "Depth perception using blurring and its application in VLSI wafer probing," Mach. Vision Appl. **5**, 35-45 (1992).
8. S. H. Lai, C. W. Fu, and S. Chang, "A generalised depth estimation algorithm with a single image," IEEE Trans. Pattern Anal. Mach. Intell. **14**, 405-411 (1992).
9. D. T. Pham and V. Aslantas, "Depth from defocusing using a neural network," J. Pattern Recogn. **32**, 715-727 (1999).
10. V. Aslantas, "Estimation of depth from defocusing using a neural network," in *International Conference on Signal Processing*, (Canakkale, Turkey, 2003), pp. 305-309.
11. B. K. P. Horn, *Robot vision* (McGraw-Hill, New York, 1986).
12. V. Aslantas and D. T. Pham, "Depth from automatic defocusing" Opt. Express **15**, 1011-1023 (2007).
13. R. C. Gonzalez, and R. E. Woods, *Digital image processing* (Addison-Wesley, Reading, MA 1992).
14. A. Rosenfeld and C. Kak, *Digital picture processing* Second Edition, (Academic Press, New York 1982).

1. Introduction

The depth of a visible surface of a scene is the distance between the surface and the sensor. Recovering depth information from two-dimensional images of a scene is an important task in computer vision that can assist numerous applications such as object recognition, scene interpretation, obstacle avoidance, inspection and assembly.

Various passive depth computation techniques have been developed for computer vision applications [1]. They can be classified into two groups. The first group operates using just

one image. The second group requires more than one image which can be acquired using either multiple cameras or a camera whose parameters and positioning can be changed.

With no prior knowledge of the scene under analysis, depth estimation cannot be carried out using a single image of that scene. Therefore, single image based depth cues such as texture gradients, surface shading, etc. require heuristic assumptions. Hence, they cannot be used to recover absolute depth.

Several monocular depth from defocusing (MDFD) techniques have been developed and demonstrated [2-10]. The techniques depend on the blur information in the image. Without having any idea about the scene, blur can not be measured using a single image of that scene and consequently its distance. This is because soft edges in the images may be either defocused step edges or focused soft edges in the scene. Therefore, MDFD techniques had to assume that the scene contains either sharp edges (step edges) or edges with a known form perpendicular to line of sight and the defocused image is the result of convolving the focused image of those edges with the point-spread function (PSF) of the camera. Most of these techniques are also based on the PSF of the camera being either a Gaussian or a circularly symmetric function to obtain a relation between the SP of the PSF and depth.

Edge orientation is critical to the estimation of depth in most monocular MDFD techniques. This paper describes a more general technique, which is independent of edge orientation. A defocused image of an object is defined as the convolution of a sharp image of the same object with a two-dimensional Gaussian function whose SP is related to the object depth. The sharp image of an object is estimated from the defocused image of the same object by applying the sharpening filter. The defocused and sharp images of the object are used to estimate the SP of the Gaussian function. The parameter is then related to the object depth.

The technique described in this paper does not require special scene illumination and needs only a single camera. Therefore, there are no correspondence and occlusion problems as found in stereo vision and or intrusive emissions as with active depth computation techniques.

2. Problem formulation

The transformation effected by an optical system can be modeled as a convolution operation [11]. The image of a blurred object may then be written as:

$$i(x, y) = \int\int h(x - \xi, y - \eta, d(\xi, \eta)) s(\xi, \eta) \, d\xi d\eta \tag{1}$$

where x and y are image coordinates, ξ and η are two spatial variables, $s(x,y)$ and $i(x,y)$ are the sharp and defocused images of the source object respectively, $d(x,y)$ is the distance from the object to the plane of best focus (PBF) and $h(x,y,d)$ is the PSF. If the distance from the object to the PBF is constant, then the PSF $h(x,y,d)$ can be written as $h(x,y)$ and the defocusing process is defined as a convolution integral:

$$i(x, y) = \int\int h(x - \xi, y - \eta) s(\xi, \eta) \, d\xi d\eta \tag{2}$$

The convolution operation is usually denoted by the symbol $\otimes$. Therefore, Eq. (2) can be abbreviated as:

$$i(x, y) = h(x, y) \otimes s(x, y) \tag{3}$$

In the Fourier domain, Eq. (3) can be expressed as:

$$I(u, v) = H(u, v) S(u, v) \tag{4}$$

where $\{i(x,y), I(u,v)\}$, $\{h(x,y), H(u,v)\}$ and $\{s(x,y), S(u,v)\}$ are Fourier pairs. Most of the focus based techniques assume that the distance function $d(x,y)$ is slowly varying, so that it is almost constant over local regions. The defocus is then modeled by the convolution integral over these regions.

2.1 Form of point-spread function

Figure 1 shows the basic geometry of image formation. Each point in a scene is projected onto a single point on the focal plane, causing a focused image to be formed on it. However, if the sensor plane does not coincide with the focal plane, the image formed on the sensor plane will be a circular disk known as a "*circle of confusion*" or "*blur circle*" with diameter 2*R*, provided that the aperture of the lens is also circular. According to geometrical optics, the intensity distribution within the blur circle is assumed to be approximately uniform i.e., the PSF is a circular "pillbox". In reality, however, diffraction effects and characteristics of the system play a major role in forming the intensity distribution within the blur circle. After examining the net distribution of several wavelengths and considering the effects of lens aberrations the net PSF is best described by a 2D Gaussian function [3, 4]:
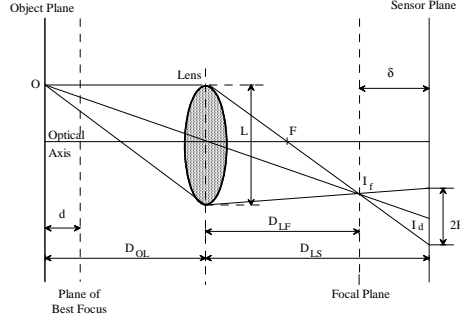

Fig. 1. Basic image formation geometry

$$h(x, y) = \exp\left(-(x^2 + y^2)/(2\sigma^2)\right)/(2\pi\sigma^2) \tag{5}$$

where σ is the SP which is proportional to the radius *R* of the blur circle.

$$R = k\sigma \tag{6}$$

The proportionality constant *k* depends on the system and can be determined through calibration. The optical transfer function $H(u,v)$ for geometrical optics can be obtained by taking the Fourier transform (FT) of $h(x,y)$:

$$H(u,v) = \exp\left(-(u^2 + v^2)\sigma^2/2\right) \tag{7}$$

By substituting Eq. (7) into Eq. (4) and solving for σ, Eq. (4) can be rewritten as:

$$\sigma^2 = -2\ln\left(I(u,v)/S(u,v)\right)/(u^2 + v^2) \tag{8}$$

It is sufficient to calculate σ at a single point (u,v) by employing Eq. (8). However, a more accurate value can be obtained by averaging σ over some domain in the frequency space:

$$\sigma^2 = \frac{-2}{A} \iint_R \frac{1}{(u^2 + v^2)} \ln\left(I(u,v)/S(u,v)\right) du dv \tag{9}$$

where *P* is a region in the (u,v) space containing points where $(I(u,v)/S(u,v)>0$ and *A* is the area of *P* [5].

2.2 Relating depth to camera parameters and defocus

The object may be either in front of or behind the PBF on which points are sharply focused on the focal plane. From Fig. 1, by using similar triangles, a formula for a camera with a thin convex lens of focal length *F* can be derived to establish the relationship between the radius *R* of the blur circle and the distance D_{OL} from a point in a scene to the lens [3, 12]:

$$D_{OL} = \frac{FD_{LS}}{D_{LS} - F - 2fk\sigma} \tag{10}$$

where D_{LS} is the distance between the lens and the sensor plane, f is the f-number of a given lens. When the object is in front of the PBF, Eq. (10) becomes:

$$D_{OL} = \frac{FD_{LS}}{D_{LS} - F + 2fk\sigma} \tag{11}$$

Equations (10) and (11) relate the object distance D_{OL} to σ.

3. Obtaining sharp images

As can be observed from Eq. (9), a blurred image and a sharp image of the same scene are needed to compute σ. A sharp image of a scene can be achieved in either an optical or a computational way. These two methods will be explained in the following sections.

3.1 Optical method for obtaining sharp images

A sharp image of an object can be obtained optically by setting the diameter of the aperture to a very small value. In this case, the camera effectively acts like a pin-hole camera. σ is proportional to aperture diameter L. Therefore, when L is very small, σ is also very small. Then the PSF may be approximated by an impulse function. Consequently, the captured image closely resembles the focused image. However, setting the aperture diameter to a very small value causes some serious practical problems. First, as the aperture diameter decreases, the diffraction effects increase. Therefore, the observed image is distorted. Second, a small aperture gathers only a small amount of light. Consequently, the period of exposure of the sensor has to be lengthened and the light intensity must be increased to take advantage of the sensor's full dynamic range. This increases the time required. Also, the scene must be stationary not only while each of the two images (one obtained with a large aperture and the other with a small aperture) is captured but also in the interval between the acquisitions of those images.

In the computational method of obtaining a sharp image, there is no need to take more than one image of the scene. A sharp image of a scene can be obtained from a blurred image of the same scene by employing sharpening filters. Therefore, in this study the second approach has been utilized to obtain the sharp image from its blurred version. The employed sharpening filter is explained in the following section.

3.2 The Laplacian sharpening filter

Averaging of pixels over an area blurs detail in an image. As the averaging or blurring operation is similar to the integration operation, the differentiation operation can be expected to have the opposite effect. Therefore, a blurred image can be sharpened by performing differentiation operations [13].

Because blurred features which are to be sharpened (such as lines and edges) can have any orientation in an image, it is important to employ a derivative operator whose output is not biased by a particular feature orientation. Therefore, the operator should be isotropic, i.e. rotation invariant. The Laplacian is a linear derivative operator that is rotationally invariant. The Laplacian of an image is a second-order spatial derivative defined as:

$$\nabla^2 i = \frac{\partial^2 i}{\partial x^2} + \frac{\partial^2 i}{\partial y^2} \tag{12}$$

How the Laplacian is used for sharpening a blurred image can be shown by assuming that the blur in the image is the result of a diffusion process which satisfies the well-known partial differential equation:

$$\frac{\partial i}{\partial t} = c\nabla^2 i \tag{13}$$

where c is a constant and i is a function of x, y and t (time). $i(x,y,0)$ is the sharp image $s(x,y)$ at $t = 0$. The blurred image $i(x,y,t)$ is obtained at some $t=\tau >0$. Then, $i(x,y,t)$ is approximated at $t=\tau$ by the following Taylor polynomial:

$$i(x, y,0) = i(x, y, \tau) - \tau \frac{\partial i}{\partial t}(x, y, \tau) + \frac{\tau^2}{2} \frac{\partial^2 i}{\partial t^2}(x, y, \tau) - \dots \tag{14}$$

By ignoring the quadratic and higher-order terms and substituting s for $i(x,y,0)$, $i(x,y)$ for $i(x,y,\tau)$ and $c\nabla^2 i$ for $\partial i/\partial t$, a mathematical expression can be derived for $s(x, y)$ as:

$$s(x, y) = i(x, y) - c\,\tau\nabla^2 i(x, y) \tag{15}$$

The above equation indicates that the sharp image s can be obtained by subtracting from the blurred image i a positive multiple of its Laplacian. If higher-order approximations based on the Taylor series expansion are used, better results can be achieved. However, this will increase the computational cost. The aim of this paper is to find a relation between blur and depth rather than restoring the exact sharp image and the above first-order approximation is sufficient to derive that relation.

Although diffusion may not be an appropriate model for image blur, it is possible that the sharp image can be computed by a subtractive combination of the blurred image and its Laplacian. According to the diffusion model, a point source blurs into a spot with a brightness distribution whose SP is proportional to c. Therefore, c can be estimated by fitting a Gaussian to the PSF [14]. By convolving both sides of Eq. (15) with the PSF $h(x,y)$ and substituting σ for $c\tau$, the following formula is obtained:

$$s(x, y) \otimes h(x, y) = \left(i(x, y) - \sigma\nabla^2 i(x, y)\right) \otimes h(x, y) \tag{16}$$

Substituting Eq. (3) into Eq. (16) gives:

$$i(x, y) = \left(i(x, y) - \sigma\nabla^2 i(x, y)\right) \otimes h(x, y) \tag{17}$$

$h(x,y)$ can be searched iteratively to minimize the difference between the left and right hand sides of Eq. (17) over a region P, namely:

$$\min \sum_{P} \sum \left[i(x, y) \otimes (1 - h(x, y)) + \sigma\nabla^2 i(x, y) \otimes h(x, y)\right]^2 \tag{18}$$

As stated in Section 2, $h(x,y)$ is the unique indicator of the depth of a scene. Thus, when the $h(x,y)$ that minimizes the above expression is obtained, the depth can be computed using the SP of that $h(x,y)$. By taking the Laplacian of the blurred edge and subtracting the result from the blurred edge ($c = 1$), the sharpened edge is obtained. However, it also produces overshoot or "ringing" on either side of the edge. This problem can be solved by "clipping" the extreme low and high grey level values.

4. Data collection

Eight 64x64 images of a step edge with different orientations were obtained for 31 distances ranging from 150 to 450mm at intervals of 10mm from the CCD camera (three of the pictures taken at distances of 150, 300 and 450mm are shown in Fig. 2). A PULNIX TMC 76S RGB camera was used. Images from the camera were acquired by a frame grabber board (model DT2871 from Data Translation). A color camera was employed because of its availability although all the computations were performed on the intensity buffer of the frame grabber board. The focal length and f-number of the lens used were set to 20mm and 2.8 respectively. The distance of the step edge from the camera varied for different images but the camera

parameters were the same for all images. The camera was set such that the PBF was at infinity (thus the object would always be between the PBF and the lens). For each distance, fifteen images were employed for averaging to minimize the effects of noise.
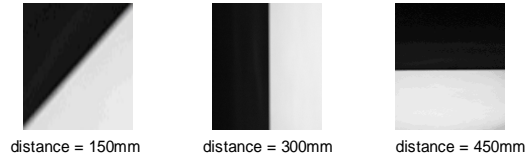

distance = 150mm   distance = 300mm   distance = 450mm

Fig. 2. Pictures of blurred step edge

5. Results

Laplacian mask was applied to a blurred image to compute its Laplacian image. Because each evaluation of Eq. (18) is a relatively expensive operation, it is important to minimize the number of evaluations required. Since Eq. (18) is a unimodal function of one variable, the Fibonacci search technique was used for finding the extrema for each image. The SPs of the computed PSFs are shown in Fig. 3(a). The estimated SPs are inexact and yielded incorrect depth values when used with Eq. (10). However, as can be seen from Fig. 3(a), there is a relation between the depth and the estimated SP. A polynomial curve was fitted to the data which is given by:

$$D_{OL} = -35.1\sigma^2 + 100.24\sigma + 562.61 \qquad (19)$$

Equation (19) can be used for edges with different heights provided that they have been taken using the same camera parameters as given in the previous section. Figure 3(b) shows the depth values estimated from Eq. (19). The percentage error achieved was 1.64%.
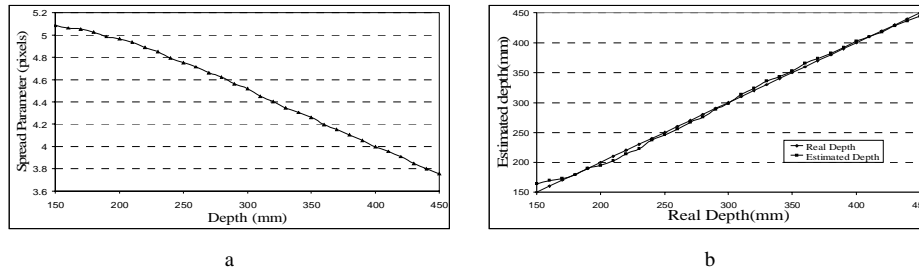

a   b

Fig. 3. Estimated a) σ b) depth

6. Conclusion

The work developed in this paper has shown that it is possible to use a sharpening filter for computing depth using a single defocused image. If the images are acquired with camera parameters other than those used in the experiments, a new polynomial curve should be fitted to the data obtained with the new camera parameters to express the correct relation between depth and SP. The technique does not require special scene illumination and needs only a single camera. Therefore, there are no correspondence and occlusion problems as found in stereo vision and motion parallax or intrusive emissions as with active depth computation techniques. The technique presented in the paper inherently have the advantage of being able to handle edges running in any direction due to the rotationally invariant property of the filter used. The technique can also be applied to the edges with known form such as slope edges provided that a new polynomial curve should be fitted to the obtained data. The technique can be used under different illumination. However, if extremely low and high illumination is used, the depth computations can be very erroneous. For example, if the illumination is low, a contrast between the dark and bright side of the edge is small and it is difficult to detect the edge. On the other hand, if the illumination is very high, the camera saturates. In this case, the shape of the edge may not be obtained completely.