Influence of mathematic models used on the quality of estimation of the depth in images

Christophe Simon, Frédérique Bicking, Thierry Simon

Influence of mathematic models used on the quality of estimation of the depth in images

Christophe SIMON
Université Henri Poincaré Nancy 1, ESSTIN
2 Rue Jean Lamour, 54519 Vandœuvre-lès-Nancy Cedex, France
Tél : +33 (0)383 685 134 Fax : +33 (0)383 685 001
email: Christophe.Simon@esstin.uhp-nancy.fr
Frédérique BICKING
Université Henri Poincaré Nancy 1, ESSTIN
2 Rue Jean Lamour, 54519 Vandœuvre-lès-Nancy Cedex, France
Tél : +33 (0)383 685 134 Fax : +33 (0)383 685 001
email: Frederique.Bicking@esstin.uhp-nancy.fr
Thierry SIMON
IUT de Figeac, Avenue de Nayrac, 46100 Figeac
Tél : +33 (0)565 503 068 Fax : +33 (0)565 503 061
email: Thierry.Simon@univ-tlse2.fr

Index Terms

depth map, depth from defocus, point spread function, gradient operators

I. INTRODUCTION

The field of contact free 3D measure is particularly open since current measuring devices use lines or points. In this domain, cameras are the preferential sensors to develop global 3D measurement system. Many methods have been developed to obtain the 3D coordinates of objects using images and all exploit the variations of acquisition parameters. Acquisition parameters of the system or of the luminous environment, controlled or not, are the essential information to establish a

relationship between the image and the real scene. Two categories of approaches can be defined. In multi-ocular approaches, a single scene is acquired with several viewpoints. In monocular approaches, the scene is acquired with the same viewpoint. The first category is extremely developed but these techniques present the major drawback of a high processing time for point matching. Nevertheless, a good feature correspondence will provide an accurate measure. The second category gathers techniques using the optical blur as information of depth. The matching problem does not arise, but the accuracy depends strongly on acquisition conditions. Among these techniques, some exploit the perceptible optical blur on heterogeneous zones of the image like edges or textures. These techniques, called Depth From Defocus or DFD ([5]; [10]; [6]), use at least two images acquired with different camera parameter settings. Classical DFD techniques can use the spatial content of the image by geometrical characteristics ([5]) as well as the form of the objects ([3]) in the scene, or even frequency information ([2]). Frequency analysis is time consuming. That's why we have preferred a spatial domain approach on object characteristics for its ability in real time applications.

In this article, we review the theoretical background of our depth perception method by defining the relation between the depth and the optical blur and by specifying the mathematic models used. We propose different image processing operators, optical system models and define their numerical application in our DFD method. With some chosen experiments, we show their influence on the accuracy of the measure, their appropriateness in noisy context and their application with imaging systems.

II. THEORETICAL DEVELOPMENT

DFD methods allow the determination of an object depth by using at least two images acquired with only one optical system but with different acquisition conditions. If a scene containing several depth planes is acquired by a real lens, points situated at a particular distance from the lens will

be focused and will appear sharp on the image. Points of objects belonging to the other planes will form spots which will be more or less defocused (blurred) according to their distances from the camera plane. The depth information inherent in defocus is the basis of our developments. The DFD method exploits the physical effect produced by the modification of the focal length or the diaphragm aperture on image characteristics. Thus, the relationships between depth, camera parameters and optical blur amount can be defined. For a camera with a focal length f, the relation between an object point at a distance s_o and the distance of its focused image s_i is given by the well known lens formula under Gauss conditions:

$$\frac{1}{s_o} + \frac{1}{s_i} = \frac{1}{f} \tag{1}$$

Figure 1 shows the geometrical image formation process. All light rays radiated by the object O and intercepted by the lens are refracted by it to converge at a point on the focal plane. Each point in a scene is projected onto a single point on the focal plan causing a focused image if the sensor plan is on the focus plan. If the sensor plane does not coincide with the focal plan, the corresponding image of the point becomes a blurred circular patch of radius r_b assuming that the diaphragm aperture is also circular.

If the distance between the lens and the sensor plane is s and the diameter of the circular lens aperture is L, then $r_b = \frac{L}{2} s(\frac{1}{f} - \frac{1}{s_o} - \frac{1}{s})$. For $\delta > 0$, the sensor plane is behind the focal plane, otherwise it is in front of the focal plane. From optical relations, the distance to an object or the depth in a scene can be expressed by equation (2):

$$\begin{aligned} s_o &= \frac{fs}{s_i - f - 2R\frac{f}{L}} \text{ pour } \delta < 0 \\ s_o &= \frac{fs}{s_i - f + 2R\frac{f}{L}} \text{ pour } \delta \geq 0 \end{aligned} \tag{2}$$

This expression is the basis of all methods that use the optical blur amount to compute depth.

Usually, radius r_b is not directly measured because it is not a perfect disc owing to diffraction

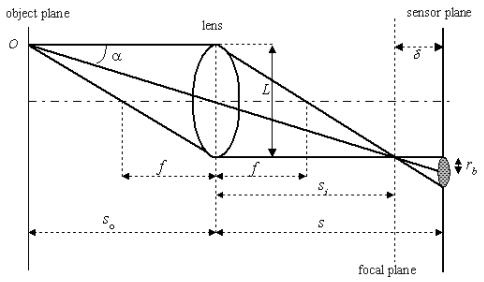

Fig. 1. Image formation process

effects. The model used to represent this patch will be a characteristic of the optical transfer function (OTF) that links r_b to the spread of the image patch. This OTF is thus a characteristic of a depth and allows to define the relation between a focused plane and a defocused one. In the spatial domain, this relation is $i_b(i,j) = i_s(i,j) \overset{\otimes}{}_{2D} h(i,j)$ where $\overset{\otimes}{}_{2D}$ is the 2D convolution operator, $h(i,j)$ the impulse response of the optical system, $i_s(i,j)$ the sharp image and $i_b(i,j)$ the blurred one. The knowledge of the impulse response of the optical system called PSF (Point Spread Function) allows to obtain spread parameters for different depths. The link between the spread parameter Pe_{s_o} and the depth s_o is then immediate with relationship (3):

$$\frac{1}{s_o} = \frac{Pe_{s_o}}{m} - \frac{c}{m} \tag{3}$$

where constants c and m are characteristics of a set of camera tuning parameters. They are determined by an appropriate calibration procedure.

The general principle of the method presented on figure 2 is similar to those proposed by Pentland ([4]; [8]) where the acquisition of a sharp image with a closed aperture and a blurred image with an open aperture is retained. The position of edges is detected with a gradient operator and a threshold. An estimation of the blur is obtained from the gradient magnitudes. With the ratio of sharp and blurred gradient image magnitudes, the spread parameter of the PSF is estimated.

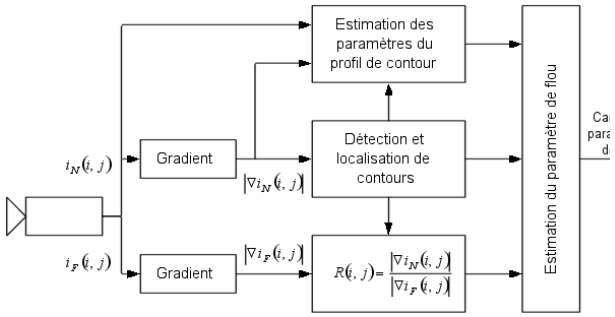

Fig. 2. Image formation process

By (3) the depth is obtained with a spread parameter estimation. If the spread parameter is determined for each pixel belonging to object edges, we obtain a depth map. The ratio of the two gradient image magnitudes can be expressed by (4):

$$R(i,j) = \frac{|\nabla i_s(i,j)|}{|\nabla i_b(i,j)|} = \frac{|\nabla i_s(i,j)|}{\left|\nabla\left(i_s(i,j) \underset{2D}{\otimes} h(i,j)\right)\right|} \tag{4}$$

where ∇ represents the gradient operator. The originality of this approach is to consider the spatial discontinuity under a one-dimensional form rather than a bi-dimensional one. For this purpose, we have to use a general model of edges applicable whatever the sharp image quality is. The model of edges in slope that is among the most used is retained. A gradient operator and a threshold allow the location of edges in the sharp image. Consider a sharp edge profile $c_s(x)$ taken in the direction of the axis x perpendicular to the edge of the object founded by the threshold. $c_s(x)$ is defined as a slope of magnitude $b-a$ and length ε ([8]). The edge profile of gradient magnitude in the blurred image is given by relation (5):

$$\nabla c_b(x) = \nabla c_s(x) \otimes h(x) \tag{5}$$

This relation uses the line spread function (LSF) defined by (6):

$$h(x) = \int_{-\infty}^{\infty} h(x,y)\,dy \tag{6}$$

Relation (6) is valid only if the PSF is circularly symmetric.

The ratio of gradient magnitudes becomes (7):

$$R(x) = \frac{|\nabla c_s(x)|}{|\nabla c_b(x)|} = \frac{|\nabla c_s(x)|}{|\nabla c_s(x) \otimes h(x)|} \tag{7}$$

The value of $R(x)$ is obtained by measuring the values of pixels in sharp and blurred gradient images magnitudes. Then, we obtain the spread parameter Pe_{s_o} of the LSF (or the PSF) by numerical resolution.

III. INFLUENCE OF MODELS

Two elements of the method have a significant influence on depth estimation results. It concerns the gradient operator used whose behavior in noisy context and in quality of edge location can be different and the model of PSF used to represent the optical system.

A. Image processing operators

In numerical image processing, many approximations of the gradient operator allow to obtain a more or less simple analytic form of the equation (7). Very often, the Prewitt operator is used for its simplicity of application, notably in real time context during of an implantation on DSP. In ([8]), we had used this operator and written the relationship (8) as an analytic expression of relationship (7):

$$R_\varepsilon(x) = \begin{cases} \dfrac{1}{h(x-x_0)+h(x-(x_0+\varepsilon))+2\sum_{u=1}^{\varepsilon-1} h(x-(x_0+u))} \\ \text{for } x = x_0 \text{ et } x = x_0+\varepsilon \\ \dfrac{2}{h(x-x_0)+h(x-(x_0+\varepsilon))+2\sum_{u=1}^{\varepsilon-1} h(x-(x_0+u))} \\ \text{for } x \in \,]x_0, x_0+\varepsilon[\end{cases} \tag{8}$$

where x_0 is the position of the first edge point along axis x and ε the length of the slope edge. The subscript ε signifies that the expression of ratio $R(x)$ depends on the value of ε estimated on the sharp edge profile.

This expression is valid only for edge orientations $\theta = \pm k\pi/2$ with $k \in N^+$ because of the anisotropy of the Prewitt operator which implies a different expression of the gradient magnitude of the sharp edge profile. The expression of the ratio of gradient magnitudes (8) allows us to estimate the spread parameter values for all $x \in [x_0, x_0 + \varepsilon]$. One obtains a final value by the estimated value mean ([8]). Even if Prewitt operator operates a low pass filtering on the image, its behavior in noisy environment is not performant ([1]). Therefore, the Sobel operator is often more suitable ([1]).

In addition, if we define the correspondence 2D / 1D for this operator in the orientation $\theta = \pm k\pi/2$, the theoretical expression of the gradient magnitudes ratio is equivalent to expression (8). On the other hand, the computing cost is more important than this of Prewitt operator since coefficients introduced in convolution masks are not equal to 1.

The optimal operator of Canny-Deriche is a reference operator in gradient calculation. Contrary to the two preceding operators, it is a IIR filter. So, the gradient of each edge in the image is going to bring a contribution on other edge gradients and entails a more complex expression of denominators in (7). A solution to this problem can be obtained by staking the kernel of the Canny-Deriche operator to construct a pseudo - Gaussian FIR filter. This solution will be less optimal and thus becoming less suitable. Moreover, the calculation cost dramatically increases and reduces the appropriateness of a real time application for smart sensor development. Furthermore, this gradient operator is of fundamental interest in frequency approach because its spread parameter is added to those of PSF in the frequency domain. So, it will be reserved to this kind of approaches.

B. Optical models

The PSF depends on the properties of the optic materials (indication of refraction) and on the geometrical form of the lens (focal length) as well as on the parameter shot (distance of the object, aperture and lighting). A realistic model taking defocus, diffraction effects, the wave behavior of light and lens aberration into account does not exist. The three main impulse response models usually encountered are the model of Born and Wolf ([4]), the gaussian model ([9]) and the pillbox model ([7]), each model ignoring lens aberrations.

The most usual model is the gaussian one that presents the advantage of being continuously derivable and with a circular symmetry. The expression of the corresponding LSF is:

$$h\left(x\right)=\sum_{y=-\infty}^{\infty}\frac{1}{2\pi\sigma^{2}}e^{-\frac{\left(x^{2}+y^{2}\right)}{2\sigma^{2}}}\approx\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^{2}}{2\sigma^{2}}} \tag{9}$$

The pillbox model is based on geometric considerations and is simpler since it does not take the diffraction effects into account. So it only deals with the defocus effects. Its expression is given by (10):

$$h\left(i,j\right)=\begin{cases}\frac{1}{\pi R^{2}}\sqrt{i^{2}+j^{2}}\\ 0\text{ elsewhere}\end{cases} \tag{10}$$

To apply the pillbox model in (8), one should find the expression of the pillbox LSF. By integrating over y direction at point (i,j) one obtains the following expression:

$$h\left(x\right)=\begin{cases}\frac{2}{\pi R^{2}}\sqrt{R^{2}-x^{2}}\text{ for }x\leq R\\ 0\text{ elsewhere}\end{cases} \tag{11}$$

The model of Born and Wolf takes the wave nature of light into account and is given considering one wavelength λ. The expression of the PSF is given by the following equation:

$$h\left(i,j,\lambda\right)=\left(\frac{2}{a}\right)^{2}\left[U_{1}^{2}\left(a,b\right)+U_{2}^{2}\left(a,b\right)\right]I \tag{12}$$

where U_n are Lommel function, λ the wavelength, I_0 the emitted light intensity. As for a, b they depend on camera parameters and depth. In monochromatic incoherent illumination, one should take all the wavelength into account by integrating $h(i, j, \lambda)$ over λ. In most cases, the resulting model is not so different from the gaussian one ([9]).

Note that each of these models has a circular symmetry which allows us to exploit directly expression (8) by calculating the LSF of each model. Moreover, these models are significant if the optical system of the camera is isoplanetic that is to say linear shift invariant. The Pillbox model should be used when the amount of blur owing to defocus becomes more important than the one owing to diffraction. On the contrary, the gaussian model can be used. The model of Born and Wolf remains too complicated to be used in our method because one should take all the wavelength into account. We should reserve it for measurement under one wavelength illumination or short wavelength spectrum.

IV. EXPERIMENTAL EVALUATION

- Estimation accuracy.

  We have tested different models with our method on synthetic images where we control noise, spread parameter and complexity. Figure 3 shows the estimation results for an image with four rectangles noted 1 to 4 with different gray levels. Rectangle 3 shows an occlusion that characterises a classic complexity in the depth perception problem. These rectangles have been blurred with three values of spread related to their depths (for $L = 5$mm) $\sigma_{s_{o_1}} = 1.2$, $\sigma_{s_{o_2}} = \sigma_{s_{o_3}} = 1.8$, $\sigma_{s_{o_4}} = 2.6$ et $s_{o_1} = 590$mm, $s_{o_2} = s_{o_3} = 740$mm and $s_{o_4} = 1117$mm.

  The estimation is globally correct even when there is an occlusion form and different gray levels for the same depth values. However, with close edges, estimation errors of the spread parameter are important. That is due to a mutual blur contribution of these edges.

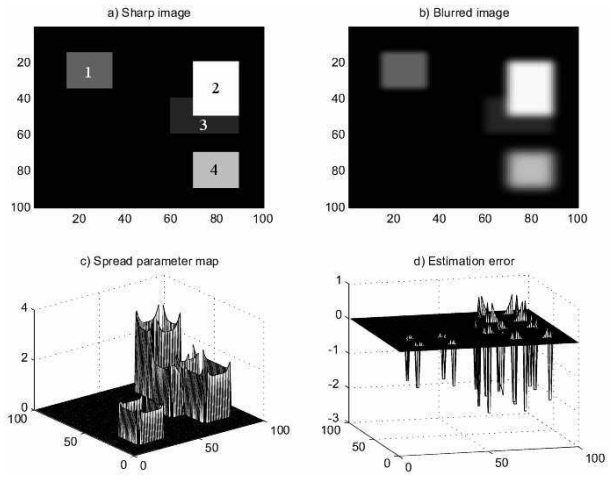

Fig. 3. Complexity test with Prewitt or Sobel

- Accuracy in noisy context.

  In a noisy context, we evaluate the influence of different gradient operators in our method when images have various signal to noise ratios. These experiments allow us to conclude to prefer the Sobel operator with noisy images.

- Influence of optical models.

  We apply our method with different PSF models and Prewitt or Sobel operators to real images. On the one hand, we evaluate the depth measurement accuracy and on the other hand the correspondence between models and the amount of blur.

V. CONCLUSION

In this article, we have shown a method of global 3D measure. First of all, we have specified theoretical aspects of this method. Then, we have shown the influence of numerical image processing operators on theoretical relationships and results. Finally, we have shown the problem and the influence of the choice of the optical impulse response model. These choices can be of great importance on accuracy and computing cost if we wish to develop a smart sensor for global 3D measure with a DSP solution.

REFERENCES

[1] J. Canny. A computational approach to edge detection. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 8(6):679–698, November 1986.

[2] S. Chaudhuri and A. Rajagopalan. *Depth from defocus : A real aperture imaging approach*. Springer-Verlag, 1998.

[3] T. Darell and K. Wohn. Depth from focus using a pyramid architecture. *Pattern Recognition Letters*, 11 (12):787–796, 1990.

[4] H. Hopkins. The frequency response of a defocused optical system. In *Proc. Royal Soc.*, volume 231, pages 91–103, 1955.

[5] A. Pentland. A new sense of depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):523–531, July 1987.

[6] D. Pham and V. Aslantas. Depth from defocus using a neural network. *Pattern Recognition*, 32:715–727, 1999.

[7] G. Schneider. *Numerical Analysis of Monocular Depth Sensing Principles and their Applications in Robotics*. PhD thesis, IAR-Université de Nancy 1, CRAN, Université de Nancy 1, july 1995.

[8] C. Simon, F. Bicking, and T. Simon. Estimation of depth on thick edges from sharp and blurred images. In *IEEE Instrumentation and Measurement Technology Conference*, pages 323–328, Anchorage, USA, May 21-23 2002. IEEE Instrumentation and Mesurement Society.

[9] M. Subbarao. Direct recovery of depth-map I: Differential methods. In *Proc. IEEE Comput. Soc. Workshop Comput. Vision*, pages 58–65, 1987.

[10] M. Subbarao and G. Surya. Depth from defocus : A spatial domain approach. *International Journal of Computer Vision*, 13(3):271–294, 1994.