Depth estimation based on thick oriented edges in images

Christophe Simon, Frédérique Bicking, Thierry Simon

Estimation of depth on thick edges images

C. SIMON[1], F. BICKING[2], and T. SIMON[2]

[1]ESSTIN, 2 rue J. Lamour, Vandoeuvre, France, e-mail : (Christophe.Simon [1], Frederique.Bicking [2])@esstin.uhp-nancy.fr
[3] IUT de Figeac, Département GMP, Avenue de Nayrac, Figeac, France, e-mail: thierry.simon@univ-tlse2.fr

***Abstract*—This article deals with a spatial approach to depth estimation by analysis of edges in images. A Depth from Defocus method is explained and the physical process is described. Theoretical developments are made to apply it on thin or thick edges and the conditions of use are pointed out. Some results on images are presented to illustrate the efficiency and the influence of the conditions of use.**

***Index Terms*— Blur estimation, depth from defocus, thick edge, 3D primitives.**

I. INTRODUCTION

The depth, distance between the visible surface of objects in a scene and the sensor of the camera, is a useful piece of information for the computation of the coordinates of the points belonging to this surface in a 3D-space of reference. Several methods such as stereopsis, depth from focus or defocus and shape from shading or structure for motion have been proposed to obtain these coordinates.

There are many works with multiocular stereovision approaches on depth perception. The well known triangulation principle is used and good efficiency can be obtained. Nevertheless, the problem of stereo correspondence has to be handled. Monocular stereovision approaches have also been developed and overcome this correspondence problem. Depth From Focus (DFF) techniques ([1], [2], [3], [4]), look for patches of sharpness in the image and link them to the depth by classic relationships of the geometrical optic. One major drawback is the long processing time to obtain a complete depth map because of the movement of the lens. So, this technique is reserved to low time process. By using at least two images of the same scene with only one point of view, Depth From Defocus (DFD) methods speed up the processing time by using the perceptible optical blur on heterogeneous image patch corresponding to edges [5] or textures [6]. These images are acquired with different well-known camera parameters settings. In comparison with multiocular stereovision, the correspondence problem does not arise in DFD. However, the main problems associated with the DFD are the requirement of proper modelling of defocusing in terms of camera parameters and the need for precise camera calibration.

In the literature on DFD, Pentland [7] compares two images of the scene, one formed with a very small (pin-hole) aperture and the other image formed with a normal aperture. By approximating the defocus blur by a Gaussian function, the depth was recovered through inverse filtering. Subbarao [8] proposes a frequency domain method in which the constraint of one image being formed with a pin-hole aperture is removed and allowed several camera parameters to be varied simultaneously. Subbarao and Surya propose an approach in [9] called the S-transform where the inverse filtering is done directly in the spatial domain. Ziou and Deschenes [10] improve this method with the Hermite polynoms. Pham and Astlantas [12] employ a multi-layer perceptron network trained by backpropagation to compute distances from derivative images of blurred edges. For each approach, the actual problem is the blur identification from observations. In recent works [13], the original image is modelled as a 2D autoregressive process and the identification problem is formulated as a maximum likelihood estimation problem.

As in the major works on blur identification, our method deals with the impulse response of the optical system. Our technique uses only sharp and blurred images of the same scene. The optical blur, characteristic of the depth, is observed on gray level discontinuities in the blurred image.

This article deals with the generalization of a local depth estimation on edges. First of all the estimation method of the amount of blur linked with the depth using a couple of focused and defocused images is explained. More details can be found in [14]. The theoretical relations are exposed and improvement with introduction of a generalized form is described. Therefore, a new relation to compute the spread parameter on several points belonging to the edges is defined and allows to reduce the noise sensitivity of the method. Conditions of application are pointed out and lead to new consideration such as constraints for further development. Then, some experiments are carried out to highlight improvements of noise sensitivity, importance of application constraints and performances. Results on synthetic and real images are finally presented.

II. CAMERA DEFOCUS BLUR AND IMAGE MODEL

In DFD methods, a relationship between depth, the parameters of the camera and optical blur in images is searched. The physical effect produced by the modification of the aperture of the diaphragm on images characteristics is used in order to establish this relationship. In the image formation process, for a scene containing several depth planes, only one plane gives a sharp or a focused image corresponding to an image plane with a fixed distance. Points of objects belonging to the other planes will form spots more or less blurred according to their distance to the image plan.

The formation of the optical blur is linked to the optical transfer function of the system [15] in the spatial domain represented by its Point Spread Function (PSF). Thus, a con-

volution relationship is established between the sharp image $I_s(i,j)$, the blurred image $I_b(i,j)$ and the PSF $h(i,j)$ given by:

$$I_b(i,j) = I_s(i,j) \overset{\otimes}{}_{2D} PSF(i,j) \quad (1)$$

where i and j are the coordinates of a pixel and $\overset{\otimes}{}_{2D}$ is the 2D convolution operator.

The PSF depends on the properties of the optic materials (indication of refraction) and on the geometrical form of the lens (focal distance) as well as on the parameter shot (distance of the object plan to the main plan, aperture, lighting). A realistic model taking into account both the aspects of the geometrical optic effects of the diffraction and the non-idealities of lenses does not exist. Several models have already been proposed. In order to deal with different forms of point spread function h, we use the spread parameter σ_h to characterize them where σ_h is the standard deviation of the distribution of any function h.

The ideal model or pillbox model doesn't take any optical aberrations into account and directly translates the geometrical optic relation. This model is usually used if there is a large amount of blur because in this case effects of defocus are predominant on those due to diffraction. Under this assumption, this model appears as a good approximation of the physical point spread function.

A *real* model that takes diffraction of light into account was defined by Hopkins [16] and approximated by Stokseth [17]. This model uses Bessel functions with a wavelength as a parameter. In our application, we use only intensity images and are only able to give an estimated mean value for the wavelength. The accuracy of the model is lost in the definition of the value of the wavelength.

A good approximation for the intensity distribution has been suggested by a 2D gaussian model h. This model is a good approximation of Stokseth model in spatial domain but it is different over the first lobe in frequency domain that means small differences appear only for large values of (i,j) for which the value of $h(i,j)$ becomes lower. However, these differences can be neglected. So, we have selected this model with a corresponding spread parameter σ_h.

$$h(i,j) = \frac{1}{2\pi\sigma_h^2} e^{-\frac{i^2+j^2}{2\sigma_h^2}} \quad (2)$$

For each PSF model, the spread parameter σ is directly linked with the depth s_o by:

$$\sigma_{s_o} = \frac{m}{s_o} + c \quad (3)$$

where m and c are characteristic of a set of camera tuning parameters [18]. These parameters are based on the ideal thin lens model and can never be measured precisely for any camera since real imaging systems have several lenses. That's why a calibration procedure is often used to determine m and c. Experimental determination of m and c has to be carefully done because uncertainty in these parameters is source of error evaluation of the depth s_o. .

Relation (3) shows that depths s_o can be computed if values of σ_{s_o} are evaluated, that means the depth is directly linked to the spread parameter of the space variant point spread function.

III. Depth perception method

Classic depth estimation approaches use either the spatial content of the image by geometrical characteristics ([19], [7]) as well as the form of objects in the scene [1], or the frequency information by a Fourier analysis [8].

Discontinuities of luminance expressed on the edges of observed objects for which the blur effects are easily perceptible, are used. The initial used method is similar to those proposed by Pentland 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. An estimation of the optical blur is obtained from the module of the gradients of each gray level image. With the ratio of gradient magnitude of sharp and blurred images, it is possible to identify the spread parameter σ_{s_o} and thus to estimate the depth s_o with the relation (3). Figure 1 shows the synopsis of the method.

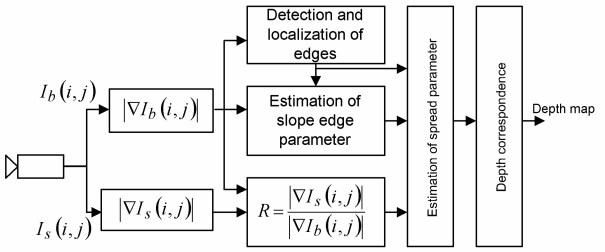

Fig. 1. Synopsis of the method

$|\bigtriangledown I_s(i,j)|$ and $|\bigtriangledown I_b(i,j)|$ are respectively the sharp image gradient magnitude and the blurred image gradient magnitude used to compute the ratio:

$$R(i,j) = \frac{|\bigtriangledown I_s(i,j)|}{|\bigtriangledown I_b(i,j)|} \quad (4)$$

As $\bigtriangledown I_s(i,j)$ can take different values, computing this ratio allows us to normalize the level of luminance of each edge point.

A. Method description

Lets the sharp image $I_s(i,j)$ presenting gray level discontinuities in slope with a magnitude $(b-a)$ and a length ε under a single direction θ. In other directions, the sharp image does not present gray level variations. The gray level function for the sharp edge profile $c_s(x)$ can be expressed by the following relation where x is expressed in pixels and represents one line or one column of this image:

$$c_s(x) = \begin{cases} a & x < x_0 \\ a + \left(\frac{b-a}{\varepsilon}\right)(x - x_0) & x_0 \leq x \leq x_0 + \varepsilon \\ b & x > x_0 + \varepsilon \end{cases} \quad (5)$$

In order to compute $\bigtriangledown I_s(i,j)$ from the digital image $I_s(i,j)$, the gradient Prewitt operator $(\bigtriangledown_p)$ is used for its simplicity in image real time implementation. The Sobel gradient operator can also be applied for its efficiency in noisy context but it introduces an higher computation time. One may use optimal edge detectors like Canny-Deriche, however as IIR filter it introduces contributions of all edges in the image in the gradient computation . Its application will be more suitable in frequency domain approaches of DFD.

The gradient magnitude of the sharp edge profile is computed using a 2D-1D correspondence in case of edge directions $\theta = \pm k\pi/2$:

$$|\bigtriangledown_p c_s\,(x)| = \begin{cases} 0 & x < x_0 \\ \left|\frac{b-a}{\varepsilon}\right| & x = x_0 \\ 2\left|\frac{b-a}{\varepsilon}\right| & x_0 < x < x_0 + \varepsilon \\ \left|\frac{b-a}{\varepsilon}\right| & x_0 = x_0 + \varepsilon \\ 0 & x > x_0 + \varepsilon \end{cases} \quad (6)$$

The same analysis can be done with the blurred image. The blurred edge profile can be expressed by:

$$|c_b(x)| = |c_s(x) \otimes LSF_{s_o}(x)| \quad (7)$$

where $LSF_{s_o}(x)$ is the line spread function with a spread parameter $\sigma_h = \sigma_{s_o}$ expressed by:

$$LSF_{s_o}\,(x) = \sum_y h\,(x,y)\,\Delta y \cong \frac{1}{\sigma_{s_o}\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_{s_o}^2}} \quad (8)$$

The subscript s_o is introduced here to specify the depth dependency of spread parameter.

Sharp and blurred edge profiles are represented on figure 2.

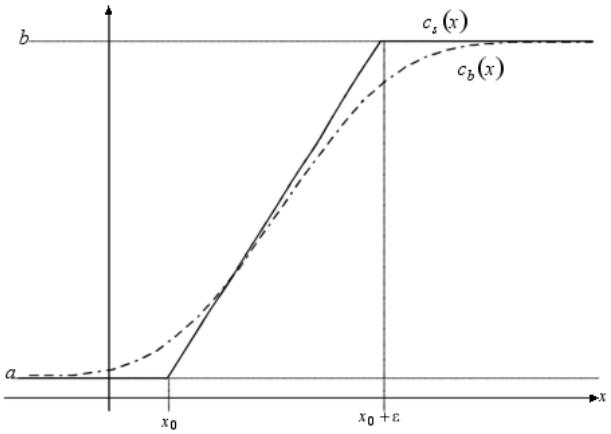

Fig. 2. Sharp and blurred edge profiles

The ratio given by (4) is computed for different values of ε at two particular points x_0 and $x_0 + \varepsilon$ with $R(x_0) = R(x_0 + \varepsilon)$.

The gradient of the sharp edge presented on figure 3 is:

$$|\bigtriangledown_p c_s(x_0)| = \left|\frac{b-a}{\varepsilon}\right| \quad (9)$$

The gradient of the blurred edge presented on figure 3 is:

$$|\bigtriangledown_p c_b(x_0)| = \left|\left(\frac{b-a}{\varepsilon}\right)(LSF_{s_o}\,(0) + LSF_{s_o}(\varepsilon))\right| \quad (10)$$

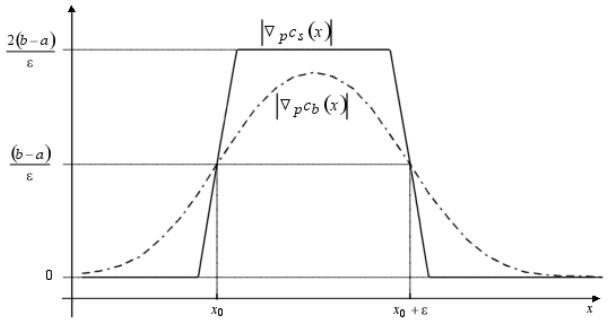

Fig. 3. Sharp and blurred edge gradient profiles

Previous works ([21],[14]) use a maximum length ε set to 3, considering that no larger values occur with high quality optics. Using standard quality optics, edges of length ε higher than 3 can be found. Thus, a general expression of $R(x_0)$ available for thin and thick edges that means for all ε was defined.

The gradient of the sharp edge is still expressed by (9) and the blurred one becomes:

$$|\bigtriangledown_p c_b(x_0)| = \left|\left(\tfrac{b-a}{\varepsilon}\right)(LSF_{s_o}\,(0) + LSF_{s_o}\,(\varepsilon) + 2\sum_{u=1}^{\varepsilon-1} LSF_{s_o}\,(u)\,)\right| \quad (11)$$

So, the general form of the ratio $\forall\varepsilon$ can be expressed by:

$$R\,(x_o,\varepsilon) = \frac{1}{LSF_{s_o}\,(0) + LSF_{s_o}\,(\varepsilon) + 2\sum_{u=1}^{\varepsilon-1} LSF_{s_o}\,(u)} \quad (12)$$

This relation allows to estimate the spread parameter σ_{s_o} for the points x_0 and $x_0 + \varepsilon$ from the values of the gradient magnitudes of the sharp and blurred images at these two points and for all ε.

In presence of noisy images, estimation will be corrupted. Thus, the use of each point belonging to the edge to give several estimated values of σ_{s_o} was defined and the estimation results will appear more robust. Using the relations (6) and (8), the gradient magnitude of the blurred edge profile for each point x with $x_0 < x < x_0 + \varepsilon$ is given by:

$$\begin{aligned} |\bigtriangledown_p c_b\,(x)| \;=\; & \left|\frac{b-a}{\varepsilon}\right|(LSF_{s_o}\,(x-x_0) + \\ & LSF_{s_o}\,(x-(x_0+\varepsilon)) + \\ & 2\sum_{u=1}^{\varepsilon-1} LSF_{s_o}\,(x-(x_0+u))\,) \end{aligned} \quad (13)$$

Using the relations (6) and (13), the generalization of the ratio (12) for each point between x_0 and $x_0 + \varepsilon$ and $\forall\varepsilon$ is given by:

$$R\,(x,\varepsilon) = \begin{cases} \dfrac{2}{LSF_{s_o}(x-x_0)+LSF_{s_o}(x-(x_0+\varepsilon))+2\sum_{u=1}^{\varepsilon-1} LSF_{s_o}(x-(x_0+u))} \\ \text{for } x_0 < x < x_0+\varepsilon \\ \dfrac{1}{LSF_{s_o}(x-x_0)+LSF_{s_o}(x-(x_0+\varepsilon))+2\sum_{u=1}^{\varepsilon-1} LSF_{s_o}(x-(x_0+u))} \\ \text{for } x = x_0 \text{ or } x = x_0+\varepsilon \end{cases} \quad (14)$$

To reduce noise sensitivity, and thus to improve the quality of the solution, $\varepsilon + 1$ estimations of the spread parameter σ_{s_o} are computed and a statistical value is attributed to the point x_0 in the depth map. That implies more computing time but the estimation obtained is better.

B. Generalization with other edge orientations

The ratio 14 is valid for edge orientation $\theta = \pm k\pi/2$. In order to generalize the depth perception method to all potential edges in images, the same study was made for directions $\theta = \pi/4 \pm k\pi/2$. The study made on image lights out the anisotropy of Prewitt operator. The analysis of the gradient magnitude of the sharp image and those of the blur one leads to the definition to new expressions for the direction $\theta = \pi/4 \pm k\pi/2$:

The gradient magnitude of the sharp edge computed for all ε at points x_0 and $x_0+\varepsilon$ is:

$$|\bigtriangledown_p c_s(x_0)| = \begin{cases} \frac{2\sqrt{2}}{3}|a-b| & \text{for } \varepsilon = 1 \\ \frac{\sqrt{2}}{\varepsilon}|a-b| & \text{for } \varepsilon > 1 \end{cases} \quad (15)$$

The gradient magnitude of the blur edge computed for all ε at points x_0 and $x_0+\varepsilon$ is:

$$|\bigtriangledown_p c_b(x_0)| = \begin{cases} \frac{\sqrt{2}}{3}|a-b| \,.\, [2LSF_{d_{s_o}}(x_0) + \\ 3LSF_{d_{s_o}}(x_0+1) + \\ LSF_{d_{s_o}}(x_0+2)\,] \\ \text{for } \varepsilon = 1 \\ \frac{\sqrt{2}}{\varepsilon}|a-b| \,.\, [LSF_{d_{s_o}}(x_0) + \\ LSF_{d_{s_o}}(x_0+\varepsilon) + \\ \frac{1}{3}LSF_{d_{s_o}}(x_0+\varepsilon+1) + \\ \frac{5}{3}LSF_{d_{s_o}}(x_0+\varepsilon-1) + \\ 2\sum_{i=1}^{\varepsilon-2} LSF_{d_{s_o}}(x_0+i)\,] \\ \text{for } \varepsilon > 1 \end{cases} \quad (16)$$

where $LSF_{d_{s_o}}(x)$ is the line spread function for direction $\theta = \pi/4 \pm k\pi/2$ defined by:

$$LSF_{d_{s_o}}(x) \cong \frac{1}{2\sigma_{s_o}\sqrt{\pi}} e^{-\frac{x^2}{4\sigma_{s_o}^2}} \quad (17)$$

The ratio defined by 14 becomes:

$$R_d(x_0,\varepsilon) = \frac{\bigtriangledown_p c_s(x_0)}{\bigtriangledown_p c_b(x_0)} = \begin{cases} \frac{2}{2LSF_{d_{s_o}}(x_0)+3LSF_{d_{s_o}}(x_0+1)+LSF_{d_{s_o}}(x_0+2)} \\ \text{for } \varepsilon = 1 \\ \frac{1}{LSF_{d_{s_o}}(x_0)+LSF_{d_{s_o}}(x_0+\varepsilon)+\frac{1}{3}LSF_{d_{s_o}}(\varepsilon+1)} \\ +\frac{5}{3}LSF_{d_{s_o}}(\varepsilon-1)+2\sum_{i=1}^{\varepsilon-2}LSF_{d_{s_o}}(i) \\ \text{for } \varepsilon > 1 \end{cases} \quad (18)$$

As previously, the ratio for all points belonging the edge was defined. Its expression is :

suite

REFERENCES

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

[2] P. Grossman, "Depth from focus," *Pattern Recognition Letters*, vol. 5, pp. 63–69, 1987.

[3] E. Krotkov, "Focusing," *International Journal of Computer Vision*, vol. 1, pp. 223–237, 1987.

[4] M. Subbarao, T. Choi, and A. Nikzad, "Focusing techniques," *Journal of Optical Engineering*, vol. 32 (11), pp. 2824–2836, 1993.

[5] G. SURYA, *Three-Dimensional scene recovery from image defocus*, Ph.D. thesis, State University of New York, Stony Brook, Dept.of Electrical Engineering, June 1994.

[6] M. Léard, M. Khoudeir, J. Brochard, and J. Bernard, "Depth measurement of textured surface in a 3d scene," in *IEEE International Conference on Electronics Circuits and Systems*, Le caire, Egypte, December 1997, pp. 435–438.

[7] A.P. Pentland, "A new sense of depth of field," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 4, pp. 523–531, July 1987.

[8] M. Subbarao, "Parallel depth recovery by changing camera parameters," in *Proc. Second International IEEE Conference on Computer Vision*, Tampa, Florida, USA, 1988, pp. 149–155.

[9] M. Subbarao and G. Surya, "Depth from defocus : A spatial domain approach," *International Journal of Computer Vision*, vol. 13, no. 3, pp. 271–294, 1994.

[10] D. Ziou and F. Deschenes, "Depth from defocus in spatial domain," *Computer vision and image understanding*, vol. 81, pp. 143–165, 2001.

[11] D.T. Pham and V. Aslantas, "Depth from defocus using a neural network," *Pattern Recognition*, vol. 32, pp. 715–727, 1999.

[12] A.N. Rajagopalan and S. Chaudhuri, "Performance analysis of maximum likelihood estimator for recovery of depth from defocused images and optimal selection of camera parameters," *International Journal of Computer Vision*, vol. 30, no. 3, pp. 175–190, 1998.

[13] T. Simon, B. Heit, and J. Brémont, "The optical blur for depth mapping in computer vision," in *Proc. 1st Japanese-French Congres of Mechatronic*, Besançon, France, 1992.

[14] C.S. Williams and A.A. Becklund, *Introduction to the OTF*, Wiley Series in Pare and Applied optics, 1989.

[15] H.H. Hopkins, "The frequency response of a defocused optical system," in *Proc. Royal Soc.*, 1955, vol. 231, pp. 91–103.

[16] A. Stokseth, "Properties of a defocused optical system," *Journal of the Optical Society of America*, vol. 59, no. 10, pp. 1314–1321, 1969.

[17] M. Subbarao, "Determining distance from defocused images of simple objects," Tech. Rep. 89-07-20, Computer vision laboratory, Dpt of Electrical Engineering, State University of New-York, Stony Brook, NY 11794-2350, USA, 1989.

[18] A.P. Pentland, "Depth of scene from depth of field," in *Proc. Image Understanding Workshop*, 1982, pp. 253–259.

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