

Real Time Monocular Depth from Defocus

Jean-Vincent Leroy¹, Thierry Simon², and François Deschenes^{1,3}

¹ Centre de MOIVRE, Université de Sherbrooke, Sherbrooke (Québec) Canada J1K2R1

² Université de Toulouse ; UTM-IUT de Figeac ; LRP-mip (Laboratoire de recherche pluridisciplinaire du nord-est de Midi-Pyrénées) ; Avenue de Nayrac, F-46100 FIGEAC, France

³ Université du Québec en Outaouais, Gatineau (Québec) Canada J8X3X7

Jean-vincent.leroy@usherbrooke.ca, thierry.simon@univ-tlse2.fr, francois.deschenes@usherbrooke.ca

Abstract. The method proposed in this paper uses two blurred images, acquired from the same point of view with different focus settings, in order to estimate depth in real time. It falls under the group "Depth from defocus" methods. The blur is modelled by the convolution of a Gaussian Point Spread Function (PSF) which with the theoretical sharp image. We calculate the gradients and Laplacians of the two blurred images and according to the type of contour, step, ramp, roof of line, we compute the differences of these images or their derivative until order 2 to obtain the difference in blur, difference of the variances of the Gaussian ones, which we connect to the depth by taking account of the parameters of the optical system. We use then this difference to the depth by tacking account of the parameters of the optical system. We present a set of results on real images showing the performance of the method and its limits. The computing times which we measured make possible the use of our algorithm at video rate.

Keywords: Depth from defocus, image processing.

1 Introduction

We propose a new method to estimate depth of a scene observed by a single camera. The three-dimensional information lost by projection onto the image plane can be get back using a second image. This acquisition is done from the same point of view with different focus settings. This modification leads to variations of optical blur between the two images. The proposed method estimates defocus blur at pixel locations that correspond to intensity discontinuities, that is to say edges. It thus belongs to the category of DFD (Depth from Defocus) methods [1-4].

Our work is the continuation of a previous method proposed by Simon et al. which is based on the use a blurred image and a sharp one [5]. The difficulty to obtain a sharp image led Simon et al. to extend their method in order to use three blurred images [6]. However the acquisition of three images increases both complexity of the process and computing time.

The depth map obtained remains however sparse matrix. A group of solution using the textured areas of the image makes it possible on the contrary to obtain a dense map for example [8]. These solutions model a zone of the image by a set of functions by exploiting the successive derivatives until order n . In these methods the computing times become prohibitory for applications where the execution at video rate is necessary.

To address those issues, we propose in this paper a new and simple method that is able to compute depth from defocus at video rate using two blurred images of the same scene. This method is based on a simple use of the Laplacian magnitude of the two images.

The technique that our propose uses on the theoretical elements developed initially by Subbarao [2] then generalized by Deschênes and Al in [3] exploiting the optical blur on textured zones. Our algorithm determines the type of edges and constitutes the difference of the images or their successive derivatives to obtain the difference in blur which we know to connect to the depth. This method thus makes it possible to obtain the depth for scenes of which the image present of the zones textured or not. The textured zones are regarded as very dense zones of contours, discontinuities of luminance.

We chose to apply the methodology developed by Deschênes and al. [3] approach by textured areas, to discontinuities of luminance marking contours as well as textures.

Section 2 briefly presents the theoretical background behind our DFD method. Section 3 presents our new method. Section 4 is dedicated to experimental results with real images. Finally section 5 presents the limits of our method and the perspectives of development.

2 Theoretical Background

2.1 Relationship between Blur and Depth

The figure 1 presents an ideal optical system in the case of a thin lens of focal length f , and an aperture of diameter L . For a given object point located at a distance s_o of the lens plane corresponds a single point (i.e., sharp image) at a distance s_i . Any movement of the sensor plane and/or the object along the optical axis causes blur, that is an object point now corresponds to a circular disk on the image plane, called circle of confusion, whose size is estimate by its diameter $2R$.

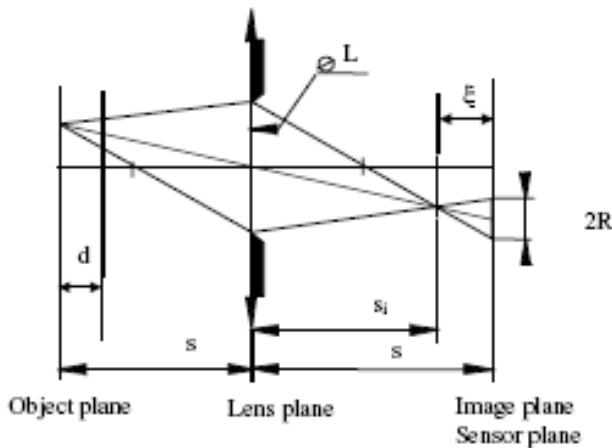


Fig. 1. Blurred image formation process for given focal length and aperture

The Gauss law of the thin lenses:

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

makes it possible to establish an explicit relationship between the diameter of the circle of confusion and the parameters of the optical system:

$$2R = Ls_i \left(\frac{1}{f} - \frac{1}{s_o} - \frac{1}{s_i} \right)$$

For $\xi \geq 0$, the sensor plane is behind the focal plane, otherwise it is in front of the focal plane. From optical relationships, the distance to an object (commonly called depth) can be expressed by the following equations:

$$s_o = \frac{fs_i}{s_i - f - 2R \frac{f}{L}} \quad \text{for } \xi < 0$$

$$s_o = \frac{fs_i}{s_i - f + 2R \frac{f}{L}} \quad \text{for } \xi \geq 0$$

These expressions are the basis of all methods that use the optical blur to compute depth.

2.2 Image Formation Process: Modeling Blur

A 2D image point of a given 3D object point can be considered as the sum of the contributions of points belonging to the surface in the vicinity of this 3D point that reflect light toward the camera. This process can be modeled by a 2D convolution [7] between an ideal image $I_s(x, y)$, that is which is sharp every where, and the point spread function (PSF), $h(x, y, s_o)$, of the optical system.

$$I_b(x, y, s_o) = I_s(x, y) \otimes_{2D} h(x, y, s_o)$$

The PSF is obviously a function of the camera parameters but is also function of the depth. It is commonly modeled by the Pillbox function or the Gaussian function [9]. In what follows, we will consider the PSF is Gaussian:

$$h(x, y, s_o) = G(\sigma_{s_o}) = \frac{1}{2\pi\sigma_{s_o}^2} \exp\left(-\frac{x^2 + y^2}{2\sigma_{s_o}^2}\right)$$

The diameter of the circle of confusion thus depends on the scale parameter σ_{s_o} of the Gaussian function. The relationship between those two parameters is given by $\sigma_{s_o} = k2R$, where k is a proportionality constant.

The camera parameters f, L, k, s_i can first be determined by calibration. Hence, being able to estimate σ_s , the remaining unknown, would allow us to estimate the depth s_o .

2.3 Edge Model

In order to process video in real time, we focus on a subset of features of the image: the edges. An edge is a discontinuity of luminance.

Specifically, we consider three edge models, $C_1(x, y)$, $C_2(x, y)$, and $C_3(x, y)$, that are respectively a step edge, a ramp and roof. For instance, in the case of vertical edges, C_1, C_2, C_3 can be represented as follows:

$$C_1(x, y) = \begin{cases} a & x < x_o, \forall y \\ a + b & x \geq x_o, \forall y \end{cases}$$

which is a function of Heaviside of amplitude (b-a) and of offset a, and

$$C_2(x, y) = \begin{cases} a & x < x_o - \frac{\varepsilon}{2}, \forall y \\ a + \frac{(b-a)}{\varepsilon} \left(x - x_o + \frac{\varepsilon}{2} \right) & x_o - \frac{\varepsilon}{2} \leq x \leq x_o + \frac{\varepsilon}{2}, \forall y \\ a + b & x > x_o + \frac{\varepsilon}{2}, \forall y \end{cases}$$

which is a ramp function of amplitude (b-a), of width ε and inflexion point x_o .

$$C_3(x, y) = \begin{cases} a & x < x_o - \varepsilon_1, \forall y \\ a + \frac{(e-a)}{\varepsilon_1} (x - x_o + \varepsilon_1) & x_o - \varepsilon_1 \leq x \leq x_o, \forall y \\ b + \frac{(e-b)}{\varepsilon_2} (x_o + \varepsilon_2 - x) & x_o \leq x \leq x_o + \varepsilon_2, \forall y \\ b & x > x_o + \varepsilon_2, \forall y \end{cases}$$

Which is roof or line model, roof for large value of ε_1 et ε_2 .

For edges that are not vertical the previous two models must simply be applied in the direction of the gradient. Particular attention must be paid to the sampling step which highly depends on the orientation. In order to make independent from the orientation one may use interpolation techniques.

We will apply a specific detection according to the type of edges previously definite starting from an analysis of the magnitude of the gradients of the image.

For example, for edges of the C_1 type, the thresholded magnitudes of gradients give the pixels of contour directly. For contours of the type C_2 or C_3 we use the zero-crossing of the Laplacian.

3 Our Method

We propose to locally model blurred edges by development in series of functions, of many authors [2][3] showed that the difference of the two blurred images can then be expressed by:

$$I_{b1}(x, y) - I_{b2}(x, y) = \frac{1}{2}(\sigma_2^2 - \sigma_1^2)\nabla^2 I_b$$

where

$$\nabla^2 I_b(x, y) = \frac{\nabla^2 I_{b1}(x, y) + \nabla^2 I_{b2}(x, y)}{2}$$

and $I_{b1}(x, y) = I_s(x, y) \otimes_{2D} G(\sigma_1)$, $I_{b2}(x, y) = I_s(x, y) \otimes_{2D} G(\sigma_2)$ and ∇^2 Laplacian operator.

We can write : $I_{b2}(x, y) = G(\beta) \otimes_{2D} I_{b1}(x, y)$, with $\beta^2 = (\sigma_2^2 - \sigma_1^2)$, $\sigma_2 > \sigma_1$.

In the same way the partial derivatives are dependent according to the same type of relation:

$$I_{b1}^{(i)(j)} - I_{b2}^{(i)(j)} = \frac{1}{2}(\sigma_2^2 - \sigma_1^2)\nabla^2 I_b^{(i)(j)} \text{ with } I_b^{(i)(j)} = \frac{\partial^{i+j} I_b}{\partial x^i \partial y^j}.$$

Moreover authors showed that the difference in blur β^2 makes it possible to calculate the depth by [8] :

$$s_0^{-1} = \frac{1}{f} - \frac{1}{s_i + s} \left(1 + \sqrt{1 + \frac{(s_i + s)k^2 \beta^2}{L^2 (s - s_i)}} \right) \tag{1}$$

The parameters of the optical system are evaluated in a phase of calibration. These relations usually used on textured zones, in our new method, are applied to edges. For a better estimate of β^2 we actually calculate:

$$\beta^2 = 2 \frac{t_{(0)(0)}\psi |D^{(0)(0)}| + t_{(1)(0)}|D^{(1)(0)}| + t_{(0)(1)}|D^{(0)(1)}|}{t_{(0)(0)}\psi |\nabla^2 I_b^{(0)(0)}| + t_{(1)(0)}|\nabla^2 I_b^{(1)(0)}| + t_{(0)(1)}|\nabla^2 I_b^{(0)(1)}|}$$

where $D^{(i)(j)} = I_{b2}^{(i)(j)} - I_{b1}^{(i)(j)}$ and if $|\nabla^2 I_{b1}^{(i)(j)}| < L_{\min}$ or $|\nabla^2 I_{b2}^{(i)(j)}| < L_{\min}$ or $|\nabla^2 I_b^{(i)(j)}| < L_{\min}$ or $\beta_{(i)(j)}^2 < \beta_{\min}$ or $\beta_{(i)(j)}^2 > \beta_{\max}$ then $t_{(i)(j)} = 0$ else $t_{(i)(j)} = 1$, with

$\beta_{(i)(j)}^2 = \frac{\nabla^2 I_b^{(i)(j)}}{2(I_{b1}^{(i)(j)} - I_{b2}^{(i)(j)})}$. Actually we compute the value of β^2 on a small surface of the image of size c x d:

$$\beta^2 = \frac{2 \sum_c \sum_d t_{(0)(0)} \psi |D^{(0)(0)}| + t_{(1)(0)} |D^{(1)(0)}| + t_{(0)(1)} |D^{(0)(1)}|}{\sum_c \sum_d t_{(0)(0)} \psi |\nabla^2 I_b^{(0)(0)}| + t_{(1)(0)} |\nabla^2 I_b^{(1)(0)}| + t_{(0)(1)} |\nabla^2 I_b^{(0)(1)}|} \quad (2)$$

The derivatives are calculated by the application of the operator of Prewitt, and the Laplacian by the successive application, twice, of $\sqrt{A_x^2 + A_y^2}$ with A_x and A_y the two directional masks of Prewitt.

The following algorithm summarizes the method:

1. Computing of the directional derivatives by the operators of Prewitt.
2. Computing of Laplacian.
3. Detection of edges on the derivative of the least blurred image
4. For each point of edge:
 - a. Computing β^2 by (2) on area with the size $c \times d$.
 - b. Computing the depth by (1).

4 Results

The results which we present here make it possible to judge the effectiveness of the method. Figure 2 presents the real image, least blurred, with three plans in-depth, of which two outsides are slowly inclined and with various textures. The depth varies between 790 and 990mm, the image has a resolution of 1700 X 1000 pixels.

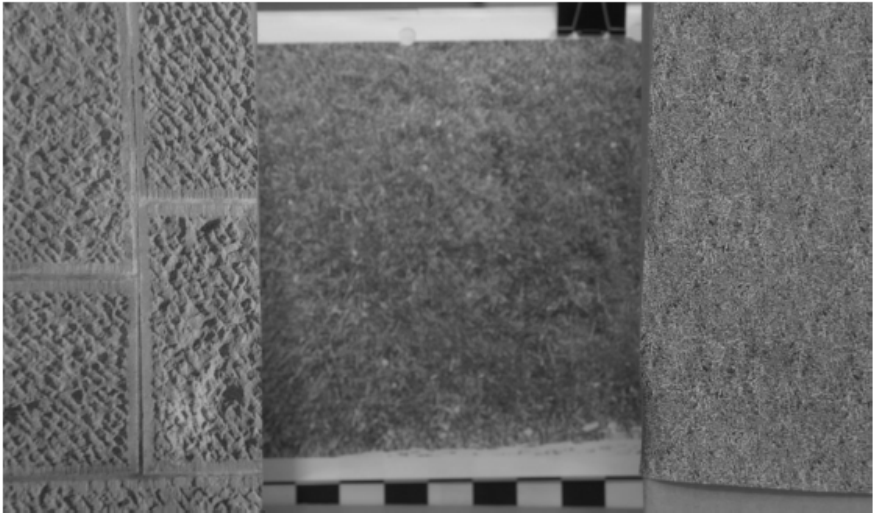


Fig. 2. The least blurred image

Figure 3 gives the results of the computing of β .

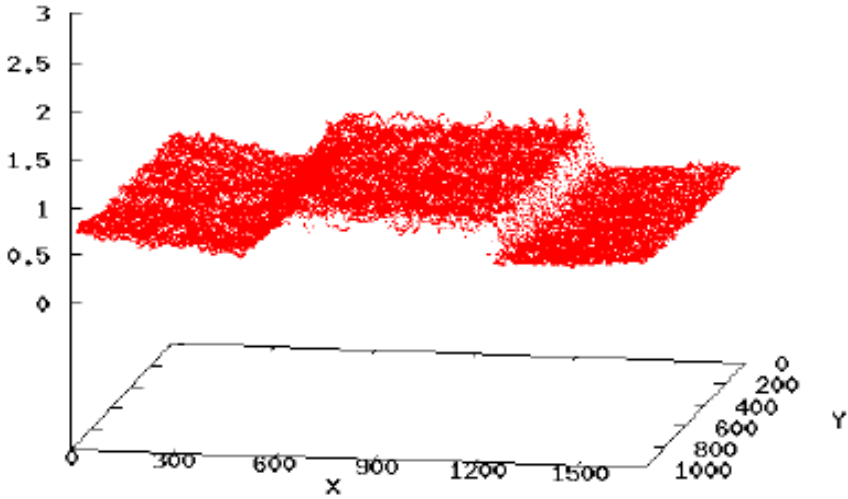


Fig. 3. β Estimations

Figure 4 shows the results obtained for the depth. The computing time is of 80ms what corresponds to approximately 23ms for a size of 800x600 pixels, traditional size for the standard cameras. This computing time makes it possible to consider a processing at the video rate. The mean errors are respectively of 11.03mm, 20.05mm and 5.44mm for the plan of right-hand side, the medium and left. It appears that according to the density of edges, characteristic of textures, the results vary. They are better when the density is high.

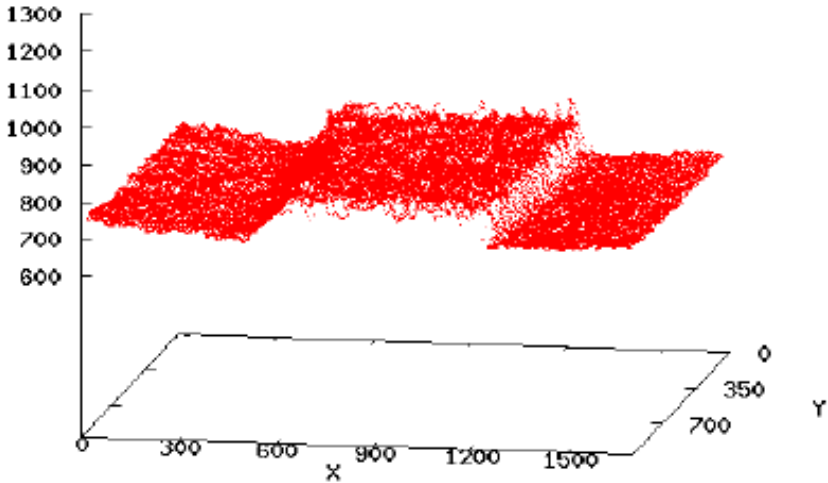


Fig. 4. Depth Estimations

Computations were carried out on APPLE Mac Pro bi-processors, Bi-hearts INTEL Xeon, at 3 GHz. Figure 5 shows a little textured scene to which we apply our method. As we could envisage it the results, figure 6, are worse, taking into account the absence of texture. The depths remain within the limits of real dimensions but they are less stable.

To carry out calculations the value $\psi = 3$ gave us the best results and the values L_{\min} , β_{\min} , β_{\max} were given in experiments. They seem independent of the type of images; we respectively fixed them at 0.5, 0.1, 6. Finally the zone of computing of the mean was selected to 7 X 7.

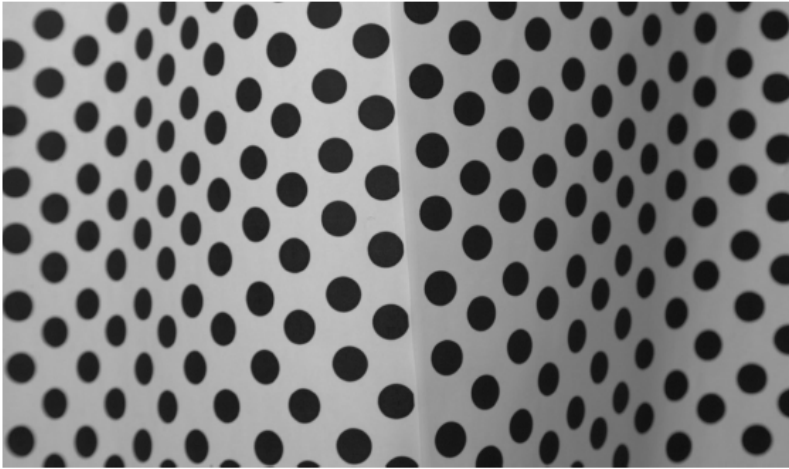


Fig. 5. The least blurred image

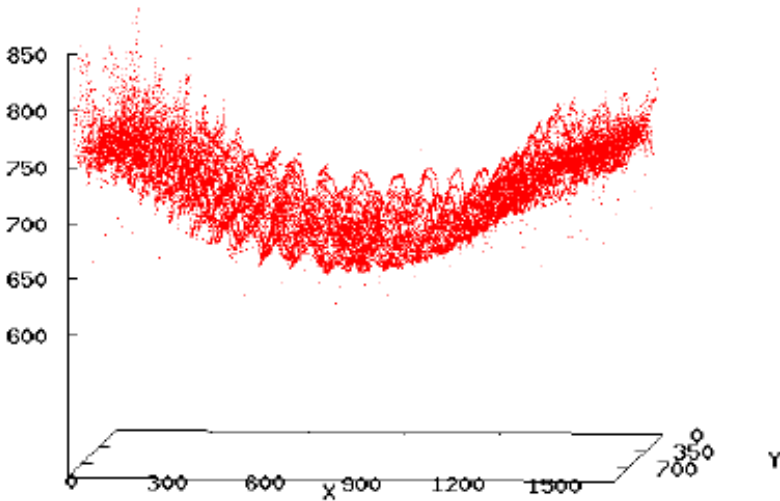


Fig. 6. Depth Estimations

5 Conclusion

We proposed in this paper a new method to estimate depth of a scene from a single camera. The three-dimensional information lost by projection on the plane sensor is obtained using defocus blur at pixel locations that correspond to edges. This method is based on the computation of the ratio of the laplacian magnitude of the images. It is thus fast. Experimental results confirm that this method lead to accurate results at video rate. In further work, we will increase the accuracy to detect in a first step the discontinuity of the depth.

Indeed the mean calculated on an area of size 7×7 makes it possible to limit the influence of the noise but led in the areas of strong variation of depth to a smoothing.

References

1. Lin, H.Y., Chang, C.H.: Depth from motion and defocus blur. *Optical engineering* 12(45), (127201)- (1-12201-10) SPIE (2006)
2. Subbarao, M., Surya, G.: Depth from defocus: a spatial domain approach. *Int. J. Comp. Vision* 13, 271–294 (1994)
3. Deschênes, F., Ziou, D., Fuchs, P.: An unified approach for simultaneous and cooperative estimation of defocus and spatial shifts. *Image and Vision Computing* n° 22, 35–57 (2004)
4. Favaro, P., Soatto, S.: 3D shape estimation and image restoration: exploiting defocus and motion blur. Springer, London (2007)
5. Simon, C., Bicking, F., Simon, T.: Depth Estimation Based on Thick Edges in Images. In: IEEE ISIE 2004, Ajaccio, France, May 3-8 (2004)
6. Simon, T., Simon, C.: Depth Perception from three blurred images. *International Electronic Congress*
7. Hopkins, H.H.: The frequency response of a defocused optical system. In: *Proc. Of the Royal Society of London series A*, pp. 91–103 (February 1955)
8. Ziou, D., Deschênes, F.: Depth from Defocus Estimation in Spatial Domain. *Computer Vision and Image Understanding* 81, 143–165 (2001)
9. Pentland, A.P.: A new sense for depth of field. *IEEE Trans. On Pattern Analysis and Machine Intelligence* 9(4), 523–531 (1987)