Depth Estimation Based on 3D Focus Measurement

Seong-O. Shim

***Abstract*—The technique to estimate depths of every object points from multiple images of same object acquired at different focus settings is called depth (or shape) from focus. The focus measurement at each pixel in image stack is important part for accurate depth map computation. The traditional focus measure operators compute focus (sharpness) value of each pixel from its neighbourhood. In this paper, we propose a focus measure based on three-dimensional gradients. Experimental results demonstrate that the accuracy of final depth map can be improved by applying the proposed method.**

***Index Terms*—Depth from focus, focus measure, depth map, three-dimensional gradient.**

I. Introduction

Computing depth map of an object from a stack of defocused images has been increasingly studied in computer vision field [1]-[4] and industrial applications [5]-[7]. In optical system based on small depth of field, the acquired images have both focused and defocused area. The technique to compute depths of every object points from stack of images obtained by gradually changing the level of focus is called depth (or shape) from focus. Focus level change is done either by changing the image sensor location in camera or by changing the distance of the object from the camera lens. The obtained images constitute three dimensional image volume where x and y axis are first and second dimensions of image and z axis is the optical axis. In this image stack, each pixel $(x, y, \cdot)$ corresponds to an certain point in real object and is slowly sharpened until it reaches to maximum sharpness and then slowly unsharpened along the z axis if magnification effect is assumed to be corrected or minimal. The basic image formation geometry is shown in Fig. 1 for convex lens camera with a lens of focal length F. We need to calculate D, i.e., depth of an object from the lens. Depth map is made by computing the depth D of every object point P. Pentland [4] has shown that the depth D is given by:

$$D = \frac{Fv_0}{v_0 - F - 2\sigma f} \tag{1}$$

where f is the f-number ($F/2r$) of lens system, and v_0 is the distance between the lens and the image detector (ID) plane.

Depth from focus techniques try to locate the image detector position where the blur circle radius σ becomes zero (maximum focus), thus the depth D can be computed as:

$$D = \frac{Fv_0}{v_0 - F} \tag{2}$$

where the frame number corresponding to the maximum focus gives the information about v_0 and F is lens focal length which is already known. To determine the maximum focus position, a focus measure operator is applied on the small regions of every pixel. Then, at each point (x, y) in image detector, which correspond to certain object point, the image frame which exhibits maximum sharpness is determined by comparing focus values at $(x, y, \cdot)$ along the z direction. This image frame provides the information about v_0 and final depth D at (x, y) is determined in (2).

In this paper, we propose a new focus measure based on three-dimensional gradients. Instead of apply focus measure on 2D neighbourhood of each pixel, we take three slices on 3D neighbourhood and apply gradient operators on these slices. In addition, we propose a new noise filtering technique to the focus measure volume to improve the final depth map quality.

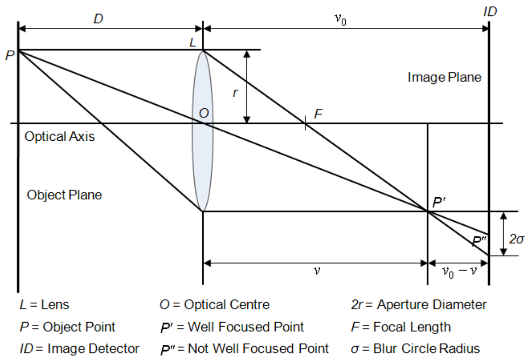


Fig. 1. Image formation of 3D object point in convex lens.

II. Focus Measure

Focus measure is defined as a quantity to locally evaluate the sharpness of a pixel. It takes small local neighborhood and computes the sharpness of a chosen center pixel. Since each object point has different surface characteristic and geometry, the focus measure values of the same object point from different optical settings are compared. A variety of focus measures have been proposed in the spatial domain and the transformed domains [8]-[12]. Some of the commonly used focus measures are given in Table I.

III. Proposed Method

A. Focus Measure Computation

We propose a focus measure based on three dimensional

Manuscript received May 15, 2015; revised July 16, 2015.

Seong-O. Shim is with the Faculty of Computing and IT, University of Jeddah, Jeddah, Saudi Arabia (e-mail: seongo@gmail.com).

DOI: 10.7763/IJMO.2015.V5.473

gradients. First, at each pixel (x,y,z) in image volume $I(x,y,z)$, we make three image patches l_{xyz}^1, l_{xyz}^2, l_{xyz}^3 as:

TABLE I: COMMONLY USED FOCUS MEASURES

Focus Measure	Mathematical Expression	Remarks
Sum Modified Laplacian [10]	$F_{SML}(x,y)=\sum_{(\xi,\eta)\in w(x,y)} F_{ML}(\xi,\eta)$, where $F_{ML}(x,y)=\begin{bmatrix} \lvert I(\xi+1,\eta)+I(\xi-1,\eta)-2I(\xi,\eta)\rvert + \\ \lvert I(\xi,\eta+1)+I(\xi,\eta-1)-2I(\xi,\eta)\rvert \end{bmatrix}$	where $w(x,y)$ denotes small window centered at pixel $w(x,y)$.
Tenenbaum Gradient [9]	$F_{TEN}(x,y)=\sum_{(\xi,\eta)\in w(x,y)} \left[\left(S_x * I(\xi,\eta)\right)^2 + \left(S_y * I(\xi,\eta)\right)^2 \right]$	S_x and S_y are Sobel operators along x and y axis respectively.
Gray Level Variance [8]	$F_{GLV}(x,y)=\frac{1}{N-1}\sum_{(\xi,\eta)\in w(x,y)} \left[I(\xi,\eta)-\mu(x,y)\right]^2$	where $\mu(x,y)$ and N are mean gray level value and total number of pixels within the window $w(x,y)$.
Energy Ratio in DCT [13]	$F_{DCT}(x,y)=\frac{E_{AC}}{E_{DC}}$	where E_{AC} and E_{DC} are energies of the AC and DC parts in Discrete Cosine Transform (DCT) of an image patch.
Energy Ratio in DWT [14]	$F_{DWT}(x,y)=\frac{M_H^2}{M_L^2}$	where M_H^2 and M_L^2 are energies of the high frequency and low frequency components in Discrete Wavelet Transform (DWT) of an image patch.

$$l_{xyz}^1=\begin{pmatrix} I(x-1,y-1,z) & I(x,y-1,z) & I(x+1,y-1,z) \\ I(x-1,y,z) & I(x,y,z) & I(x+1,y,z) \\ I(x-1,y+1,z) & I(x,y+1,z) & I(x+1,y+1,z) \end{pmatrix}$$

$$l_{xyz}^2=\begin{pmatrix} I(x-1,y,z+1) & I(x,y,z+1) & I(x+1,y,z+1) \\ I(x-1,y,z) & I(x,y,z) & I(x+1,y,z) \\ I(x-1,y,z-1) & I(x,y,z-1) & I(x+1,y,z-1) \end{pmatrix}$$

$$l_{xyz}^3=\begin{pmatrix} I(x,y-1,z+1) & I(x,y,z+1) & I(x,y+1,z+1) \\ I(x,y-1,z) & I(x,y,z) & I(x,y+1,z) \\ I(x,y-1,z-1) & I(x,y,z-1) & I(x,y+1,z-1) \end{pmatrix}$$

Then, the partials i_x, i_y along the x and y directions are estimated by Sobel operator with the convolution masks as:

$$i_x=\frac{1}{4}\begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \qquad i_y=\frac{1}{4}\begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix} \tag{3}$$

Then, the gradient magnitude is computed as:

$$G(x,y,z)=\sqrt{\begin{array}{l} \left[i_x * l_{xyz}^1\right]^2+\left[i_y * l_{xyz}^1\right]^2 \\ +\left[i_x * l_{xyz}^2\right]^2+\left[i_y * l_{xyz}^2\right]^2 \\ +\left[i_x * l_{xyz}^3\right]^2+\left[i_y * l_{xyz}^3\right]^2 \end{array}} \tag{4}$$

The final proposed focus measure at pixel (x,y,z) is defined as:

$$F_0(x,y,z)=\sum_{N(x,y,z)} G(x,y,z)^2 \tag{5}$$

where $N(x,y,z)$ is $3\times3\times3$ neighbourhood centered at (x,y,z). The computed focus measure volume F_0 has some noisy values caused from the noises both from image acquisition process and from focus value computation.

To improve the focus measure F_0, we propose a filtering technique. First, for each pixel (x,y,z), we take small three dimensional neighborhood $N(x,y,z)$ and count the total number of pixels n_{xyz} within $N(x,y,z)$ that has close value to the focus value of center pixel (x,y,z) as:

$$n_{xyz}=\sum_{(x',y',z')\in N(x,y,z)} C(x',y',z') \tag{6}$$

$$C(x',y',z')=\begin{cases} 1 & \text{if } \lvert F_0(x,y,z)-F_0(x',y',z')\rvert \le T_f \\ 0 & \text{otherwise} \end{cases}$$

where T_f is predefined threshold value. Then, the mask image $M(x,y,z)$ is defined to flag the focus value at (x,y,z) as noisy or noise free. If n_{xyz} is less then predefined value T_n, the focus value at (x,y,z) is considered as noisy and $M(x,y,z)$ is set to 0 and otherwise 1.

$$M(x,y,z)=\begin{cases} 1 & \text{if } n_{xyz}\le T_n \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

The updated focus measure $F_1(x,y,z)$ is computed as:

$$F_1(x,y,z)=\sum_{(x',y',z')\in N(x,y,z)} M(x',y',z')\cdot F_0(x',y',z') \tag{8}$$

B. Depth Map Estimation

Then, for each pixel (x, y), the temporary depth map $D_{temp}(x, y)$ is computed by locating the frame number that has maximum focus value along the z axis from focus measure volume $F_1(x, y, z)$.

$$D_{temp}(x, y) = \arg\max_z F_1(x, y, z) \tag{9}$$

The temporary depth map $D_{temp}(x, y)$ has information about v_0 in Fig. 1 where the blur circle radius σ becomes zero (maximum focus). If F is the focal length of the lens, real depth $D(x, y)$ of the object point corresponding to the pixel (x, y) is computed from (2).

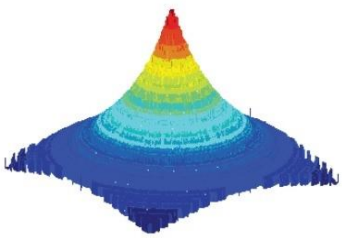
IV. Results

The performance of the proposed method is evaluated by using synthetic and real image sequences. The synthetic object is computer simulated cone from camera simulation software. A total of 97 images corresponding to 97 different lens positions was produced. Another object is a real cone. A total of 87 images was acquired by changing CCD camera lens position. The proposed method was compared with the most widely used traditional methods - SML, TEN, and GLV discussed in Section I.

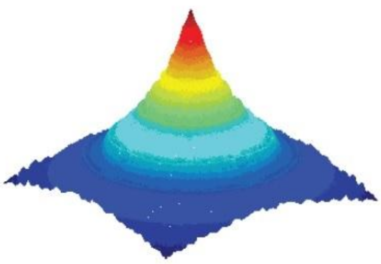
For quantitative analysis, the Root Mean Square Error (RMSE) values between the actual depth map of the synthetic object and the computed depth map based on various focus measures were compared in Table II. From the table, we can observe that applying the proposed technique could produce lower RMSE values in comparison to other methods.

TABLE II: RMSE Comparison on Simulated Cone

Focus Measure	SML	TEN	GLV	Proposed
RMSE	8.0941	8.1232	8.0886	8.0121

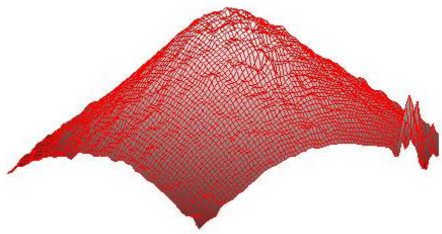


(a) Shape from SML

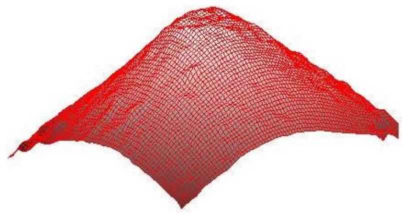


(b) Shape from the proposed method

Fig. 2. Depth map results on simulated cone based on SML focus measure and the proposed focus measure.



(a) Shape from GLV



(b) Shape from the proposed method

Fig. 3. Depth map results on real cone based on GLV focus measure and the proposed focus measure.

In Fig. 2, the depth map results on simulated cone object were compared. Compared to the results from SML, the proposed method generated sharper tip shape and noises on the slanted side is less conspicuous. In Fig. 3, the depth map results on real cone based on GLV focus measure and the proposed focus measure were compared. The shape from GLV has more noises in overall surface. However, the shape from the proposed technique produces smoother surface, and the corrupted part in the right side of the shape generated from GLV method was considerably improved.

V. Conclusions

In this paper, we proposed a new focus measure based on three-dimensional gradients. First, three image slices are taken from 3D neighbourhood of each pixel, and two partial gradient operators are applied on these images. Results are squared and summed together to form a final focus value. In addition, we propose a new noise filtering technique to improve the focus value. Experiments were conducted on both synthetic and real cone objects, and the results shows that the proposed method produces improved RMSE values on synthetic object and showed better reconstructed shape for real cone object in comparison to previous methods.

References

[1] S. K. Nayara and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 16, no. 8, pp. 824-831, 1994.

[2] M. Subbarao and T.-S. Choi, "Accurate recovery of three dimensional shape from image focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 3, pp. 266-274, 1995.

[3] S. -O. Shim and T. -S. Choi, "A novel iterative shape from focus algorithm based on combinatorial optimization," *Pattern Recognition*, vol. 43, no. 10, pp. 3338-3347, 2010.

[4] A. Pentland, "A new sense for depth of field," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 4, pp. 523-531, 1987.

[5] S. -O. Shim, M. T. Mahmood, and T.-S. Choi, "Searching surface orientation of microscopic objects for accurate 3D shape recovery," *Microscopy Research and Technique*, vol. 75, no. 5, pp. 561-565, 2012.

[6] A. S. Malik and T. S. Choi, "Application of passive techniques for three dimensional cameras," *IEEE Trans. Consumer Electronics*, vol. 53, no. 2, pp. 258-264, 2007.

[7] M. T. Mahmood, S. -O. Shim, S. Alshomrani, and T.-S. Choi, "Depth from image focus methods for micro-manufacturing," *The International Journal of Advanced Manufacturing Technology*, vol. 6, no. 5, pp. 1701-1709, 2013.

[8] E. Krotkov, "Focusing," *Int'l J. Computer Vision*, vol. 1, no. 3, pp. 223-237, 1987.

[9] J. M. Tenenbaum, "Accomodation in computer vision," Ph.D. dissertation, Standford Univ., Standford, CA, 1970.

[10] S. K. Nayara and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 16, no. 8, pp. 824-831, 1994.

[11] Y. Sun, S. Duthaler, and B. J. Nelson, "Autofocusing in computer microscopy: Selecting the optimal focus algorithm," *Microscopy Research and Technique*, vol. 65, no. 3, pp. 139-149, 2004.

[12] I. -H. Lee, S. -O. Shim, and T. -S. Choi, "Improving focus measurement via variable window shape on surface radiance distribution for 3D shape reconstruction," *Optics and Lasers in Engineering*, vol. 51, no. 5, pp. 520-526, 2013.

[13] S. Pertuz, D. Puig, and M. A. Garcia, "Analysis of focus measure operators for shape-from-focus," *Pattern Recognition*, vol. 46, no. 5, pp. 1415-1432, 2013.

[14] H. Xie, W. Rong, and L. Sun, "Construction and evaluation of a wavelet-based focus measure for microscopy imaging," *Microscopy Research Technique*, vol. 70, no. 11, pp. 987-995, 2007.

Seong-O. Shim received his BS degree in electronics engineering from Ajou University, Korea, in 1999; his M.S degree in mechatronics from Gwangju Institute of Science and Technology, Korea, in 2001; and his PhD degree in information and mechatronics from Gwangju Institute of Science and Technology, Korea, in 2011.

From 2003 to 2007, he was with the LG electronics DTV labs, Seoul, Korea, and working on research and development of digital TV. He is currently working as an assistant professor in Faculty of Computing and Information Technology, University of Jeddah, Saudi Arabia. His research interests include computer vision, data mining, 3D shape recovery, medical imaging.