15th IAPR International Conference on Machine Vision Applications (MVA)
Nagoya University, Nagoya, Japan, May 8-12, 2017.

DFD2.0: Motion Robustness by Amplitude Domain Approach

Takashi Kawamura, Khang Nguyen, Shunsuke Yasugi, Mitsuyoshi Okamoto, Motonori Ogura
Panasonic Corporation
1-15 Matsuo-cho, Kadoma City, Osaka 571-8504, Japan
kawamura.takash@jp.panasonic.com

Shinsaku Hiura
Hiroshima City University
3-4-1 Otsuka-higashi, Asaminami-ku, Hiroshima City, Hiroshima 731-3194, Japan
hiura@hiroshima-cu.ac.jp

Abstract

Depth from Defocus (DFD) is known as a technology which is able to estimate depth by a monocular camera without any additive devices. However, it has to get two blurred images with different focused distance sequentially so that subjects or the camera itself can move in the short interval of consecutive shots. Conventional DFD is intolerant to such image shift especially in the weak texture part. In this paper, a new evaluation function for DFD is shown in order to achieve the motion robustness in frequency domain. The Conventional DFD algorithm is modified considering the image shift, and in the derived formula both shift and phase components vanished. As for DFD technology, the formula is insensitive to the image shift because the defocus blur appears only in the amplitude part. As a result, we also confirmed that our algorithm overcomes conventional methods especially for the real life with rapidly moving subjects.

1 Introduction

Mirror-less cameras have become popular in recent years. As they have no optical devices for a view finder and Auto Focus (AF), the high quality cameras with interchangeable lenses are more compact than the digital single-lens reflex cameras. Instead, the mirror-less cameras have to estimate depth of the subjects from only the captured images.

Conventionally on Image Sensor Phase Detection AF (PD-AF) and Contrast based AF (C-AF) has been widely used. However, PD-AF needs to fabricate special image sensors and the quality of captured videos declines. C-AF is so-called Depth from Focus (DFF). It takes more time because it cannot get the depth information in the defocus area and it has to move the focus lens a long distance in order to search a wide range.

Depth from Defocus (DFD) has also been adopted as one of the fastest AF methods in consumer cameras since 2014. It has no degradation of captured videos like PD-AF sensors. However, it has to get two blurred images with different focused distances sequentially. Thus, there will be a problem when the subjects or the camera moves in the short interval. Such an image shift influences the depth measurement performance because conventional DFD algorithms are much affected by the image shift.

In this paper, we propose a new evaluation function for DFD technology to achieve motion robustness. This allows the real time robust depth measurement even if the subjects move with a hand-held camera. Therefore, our approach can contribute to a faster mirror-less AF.

2 Related Works

Conventionally various kinds of DFD were proposed. At first, Pentland[1] suggested that defocus information can be used for depth measurement. He used the Gaussian point spread function (PSF) as the lens blur model, and used two images with the different apertures.

The depth estimation is calculated in frequency domain as;

$$\frac{F_1}{F_2} = \frac{F_0 \times \mathrm{OTF}_1^d}{F_0 \times \mathrm{OTF}_2^d} = \frac{\mathrm{OTF}_1^d}{\mathrm{OTF}_2^d} \tag{1}$$

where, F_0 and F_i are the discrete Fourier transforms of the blur-less radiance and the captured images, while i is the image number. Note that you can get the ratio of blur information from captured images. In order to get the sophisticated calculation, Pentland regarded the PSF model as Gaussian and expanded the Equation(1). However, such a simple approximation is not suitable in reality because of some optical phenomenon.

Subbarao[2][3] proposed the S Transform Method (STM) as opposed to the spatial domain approach. The STM is applied to the convolution part in the space domain of Equation(1) and it is expanded to a polynomial. In this expansion, Subbarao gave an assumption that the images are smoothed out and approximated to a cubic polynomial. This means that it loses the depth accuracy because the spoiled detail texture information can contribute to the depth accuracy in the in-focus position.

The other approaches[5][6][7][8] using coded or color aperture were proposed in order to improve the depth accuracy. However, these approaches cannot be adopted to the consumer cameras because they spoil the image quality and bokeh shapes.

For the DFD calculation, there are two kind of approaches. Firstly, approaches [1] and [4] erase the subject texture information F_0 and get the ratio of two images like Equation(1).

Secondly, approach[7] estimates the subject texture information F_0 and gets the depth afterwards. As approaches [1] and [4] do not consider the noise model, it has no robustness in the noisy situations. On the contrary, the second approach utilize joint Wiener deconvolution with the multiple images [9], which prevents the noise effect when estimating the subject texture information. Minimize the following criterion function in order to estimate the depth d:

$$\hat{F_0} = \frac{F_1 \cdot \overline{\mathrm{OTF}_1^d} + F_2 \cdot \overline{\mathrm{OTF}_2^d}}{|\mathrm{OTF}_1^d|^2 + |\mathrm{OTF}_2^d|^2} \tag{2}$$

$$\hat{d}(x,y) = \arg\min \sum_{i=1}^{2} \left| \mathrm{IFFT}\big(\hat{F_0} \cdot \mathrm{OTF}_i^d - F_i\big) \right|^2 \tag{3}$$

Fundamentally in practice, almost all conventional DFD approaches have another problem. They have to get two blurred images with differently focused distances sequentially so that subjects or the camera itself can move in the short interval of consecutive shots. They are intolerant to such an image shift, especially in the weak texture part. A multi-focus camera[5] that could capture the tow different blurred images at the same time was proposed. However, the camera size itself becomes larger.

Subbarao[3] discussed this problem and he proposed the arranged STM. He used block shifts, and integrated the components of edge strength in each block. The block shift approach itself is useful, although the assumptions of STM and Gaussian PSF prevent the accurate depth measurement. As it is based on Equation (1) and ignore the noise model, the depth measurement is spoiled under the low light situation.

The other approach is using Motion Estimation (ME) and Motion Compensation (MC) in each block. It is reasonable, but the small residual of image shift affects the depth measurement especially in the weak texture part. It is a large restriction that only the depth of the strong texture parts can be measured, because in the large defocus area far from the in-focus area, only weak texture can be obtained.

3 Proposed Method

3.1 Concept

Considering the above problems; the frequency domain approach, Equations (2) and (3) are set as a start line.

Here we redefine the capturing model as:

$$f_1(x,y) = f_0(x,y) \otimes \mathrm{PSF}_1^d + \mathrm{Noise}_1(x,y) \tag{4}$$

$$f_2(x,y) = f_0(x,y) \otimes \mathrm{PSF}_2^d \otimes \mathrm{PSF}_{\mathrm{sh}} + \mathrm{Noise}_2(x,y) \tag{5}$$

where $\mathrm{PSF}_{\mathrm{sh}}$ is the image shift but unknown. Equations (2) and (3) can be arranged by this new definition.

To acquire the depth value in each point on the scene, a depth estimation for small cropped areas is necessary. However, such a block processing approach has a problem that it cannot treat the large image shift. For example, the larger the amount of shift, the smaller the overlap area between the blocks from the two images and it cannot be described like $\mathrm{PSF}_{\mathrm{sh}}$ in Equation(5). Thus, preprocessing of ME and MC is needed.

In this paper, we set as the target how to obtain motion robustness against the small residual of image shift, after ME and MC processing in frequency domain.

3.2 Theory

Along the concept of Equations (4) and (5), Equations (2) and (3) are arranged as:

$$\hat{F_0} = \frac{F_1 \cdot \overline{\mathrm{OTF}_1^d} + F_2 \cdot \overline{\mathrm{OTF}_2^d \cdot \mathrm{OTF}_{\mathrm{sh}}}}{|\mathrm{OTF}_1^d|^2 + |\mathrm{OTF}_2^d|^2} \tag{6}$$

$$\hat{d}(x,y) = \arg\min \Bigg[\left| \mathrm{IFFT}\big(\hat{F_0} \cdot \mathrm{OTF}_1^d - F_1\big) \right|^2 + \left| \mathrm{IFFT}\big(\hat{F_0} \cdot \mathrm{OTF}_2^d \cdot \mathrm{OTF}_{\mathrm{sh}} - F_2\big) \right|^2 \Bigg] \tag{7}$$

where $\mathrm{OTF}_{\mathrm{sh}}$ are the discrete Fourier transforms of $\mathrm{PSF}_{\mathrm{sh}}$. Equation(6) means that the subject texture information can be estimated considering the image shift. Equation(7) means that the depth measurement can be shifted to the correct position of each subject, matching each captured image.

Note that when OTF_i^d is symmetric, $\mathrm{OTF}_i^d = \overline{\mathrm{OTF}_i^d}$. In addition, as pixel shifts have no amplitude components, $|\mathrm{OTF}_{\mathrm{sh}}| = 1$.

Here the phase correlation is described as:

$$\mathrm{OTF}_{\mathrm{sh}} = \frac{F_2}{F_1} \cdot \frac{|F_1|}{|F_2|} \tag{8}$$

$$\overline{\mathrm{OTF}_{\mathrm{sh}}} = \frac{F_1}{F_2} \cdot \frac{|F_2|}{|F_1|} \tag{9}$$

In the partial block level, these Equations are almost true.

In addition, suppose we will get the depth as a representative value in each block, Equation(6) is assigned to Equation(7) and arranged in a final formula as:

$$\hat{d}(x,y) = \arg\min \sum_{u,v} \left(\frac{\left| \mathrm{OTF}_2^d \cdot |F_1| - \mathrm{OTF}_1^d \cdot |F_2| \right|^2}{\left|\mathrm{OTF}_1^d\right|^2 + \left|\mathrm{OTF}_2^d\right|^2} \right) \tag{10}$$

This is the new DFD calculation formula with the motion robustness.

Interestingly enough, pay attention to the numerator of Formula(10). it has the same meaning as the minimized numerator. When you get the depth, in Equation(1). However, remember Equation(1) has less robustness to noise than Equations (2) and (3), which are based on ones of the Formula(10). Hence, denominator of the Formula(10) can have the noise robustness.

In addition, numerator indicates that the captured images are refined to their amplitude components. It is considerable that these refinements might have the robustness to the motion. Note that both image shifts $\mathrm{OTF}_{\mathrm{sh}}$ and phase components of the images vanished. As for DFD technology, such formula is convenient, because the defocus blur only appears in the amplitude part in principle.

Fig.1 is the new proposed DFD processing flow chart. We use ME and MC before the DFD calculation in order to compensate the large amount of image shift. After compensating, block processing is adopted

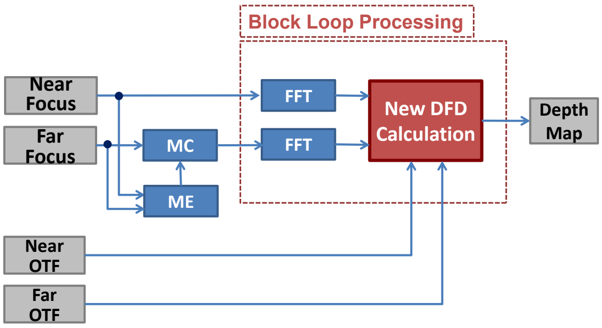

Figure 1. Flow Chart of new proposed DFD.

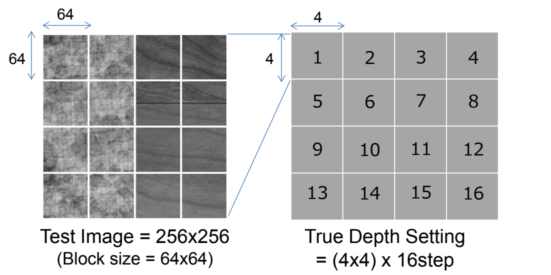

Figure 2. Left: Texture pattern for simulation, Right: True Depth Setting.

and each block of the images is transformed to the frequency domain. After that, the new DFD calculation in Formula(10) is executed which features the robustness to the small amount of image shift.

4 Experiments

4.1 Simulation

Here we show a simulation that the new formula has a robustness to the image shift.

Fig.2 is the simulation setting. The test image is 256×256pixels and the block size is 64×64pixels. Thus, the depth results are obtained as 4×4pixels at each depth step. The depth steps are 16 and the difference of the amount of blur diameter between each step is 0.65 pixel. The captured two images for DFD are focused on depth 7th and 9th, which are blurred with the simple Pillbox PSF model. The amount of image shift is within 3 pixels in each X and Y direction in the image.

The simulation results are shown in Fig.3. For estimating depth, we also use the same Pillbox PSF model. The conventional method is the approach of Equations (2) and (3).

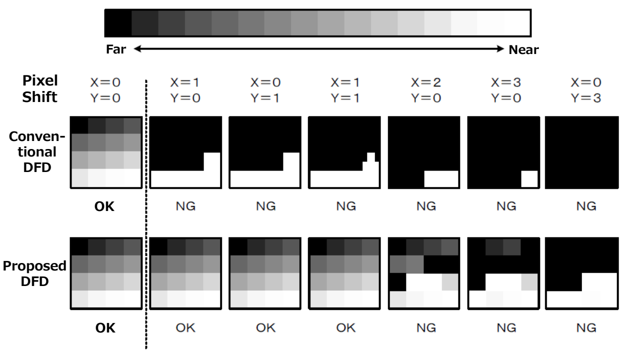

Figure 3. Simulation result.

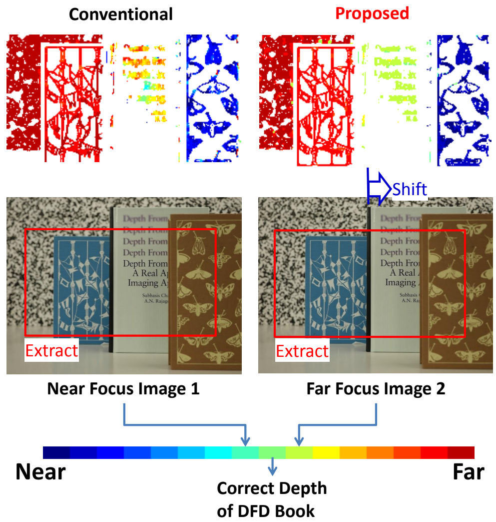

Figure 4. Simple Test: The result of moving subject.

The result indicates that the proposed method is more robust to the small image shift; within 1×1pixel shift, correct depths are measured completely. Even in NG cases, correct depths' rates are; 11/16 in (X=2,Y=0), 7/16 in (X=3,Y=0), 3/16 in (X=0,Y=3). In contrast, the conventional method do not allow no pixel shift at all. Even in (X=1, Y=0) case, correct depth rate is 2/16.

4.2 Evaluation with Actual Camera

It is confirmed how the proposed method is useful with the actual camera. We use Panasonic DMC-GH4 and H-HS12035 (LUMIX G X VARIO 12-35mm/F2.8) lens to capture two raw images in different focus positions. The PSFs are directly calculated from the lens design. Between capturing two, we only move one subject about 32pixels without changing camera angle and parameters.

Fig.4 is the simple test that we moved one book called "DFD Book" between capturing near and far focus images. We firstly focus on the "DFD Book" and shift the focus point ±1 depth amount for each capturing. Depth bar indicates that Green is correct depth of the "DFD Book". It can be confirmed that the acquired depth of the "DFD Book" is robust than the conventional one. Note that the white parts in the depth maps are masked because of the low texture and we make it with edge detection result.

Secondly, we tried to capture other various scenes mainly in outdoor with hand-held camera. After capturing the two images in a short interval by DMC-GH4, we calculate the depth afterwards on PC. As the time lag was about more than 1 second when using this prototype, we could not try to capture moving subjects because they go away from the frame during capturing 2. Thus we only evaluate the use case with natural hand shaking.

Fig.5 shows the test scenes. We set the results of DFF using tripod as the reference because it is the most accurate method.

As we show in Fig.6 and 7, we calculated depth results of each method. After obtaining the results, we calculated the depth differences between each method

Figure 5. Various Scenes' Test: The captured scenes for the experiment.

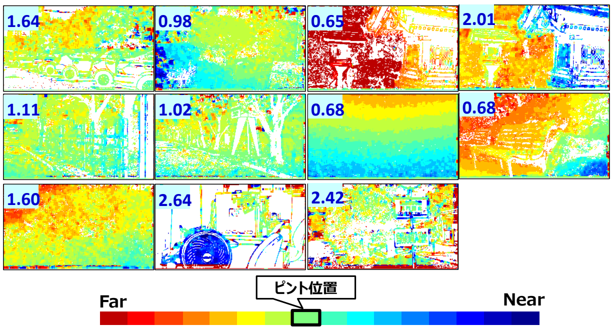

Figure 6. The depth measurement result with the Conventional DFD. The each numbers are the depth error average.

and DFF.

The numbers shown in the top left in each depth image are the depth errors, averaged in the each whole image. Here it is found that the depth errors are almost improved in the proposed method. These results have the same tendency as those of the simulation and the first simple test. In Fig.6 and 7, we can find that two samples are declined in the proposal method. However, the depth errors are lower, within 1 depth so that they are less impact. Note that no post processing of the depth maps are applied in each result.

5 Discussion

Considering the experimental results, it is found that the proposed method has a robustness to the image shift. We could not try to test the case of moving subjects, because of the prototype limitation. However, it was only the problem of implementation. Soon we will launch the product DC-GH5 in which the proposed method is implemented. With this product, we will be able to realize more than 30fps. Especially in case of photo shooting AF, 480fps will be realized. Thus, the image shift is rather decreased than the experimental situation.

This proposed method can be applied to some conventional approaches capturing two images, such as coded aperture pairs[7]. When you need more depth accuracy, such combination is effective.

By the way, we set the results of DFF as the reference in our experiment. DFF is the most accurate, passive depth measurement method. As for AF, DFF is also used as the contrast AF and conventionally combined with DFD. As I described previously, the reason why DFD is desirable is that DFF needs more than three images and DFD can improve the AF speed. In addition, as it's shown in Fig.7, the accuracy of DFD itself becomes enough. We do not have to combine with DFF in some cases, when the focus speed is especially important.

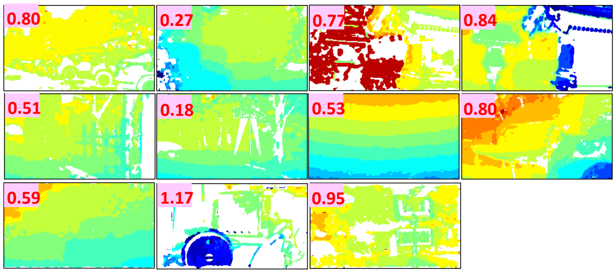

Figure 7. The depth measurement result with the Proposed DFD.

6 Conclusions

We presented the new robust DFD calculation method. We led it from the theoretical approach and obtain the better formula. Afterwards, we showed the simulation and experiments with actual camera. Experimental results indicated their effectiveness. In addition, we will launch the new camera GH5 with the higher speed AF by using this algorithm.

References

[1] A. Pentland, A New Sense of Depth of Field, Proc. IEEE Trans. Pattern Analysis and Machine Intelligence, vol.9 no.4, pp.423-430, 1987.

[2] 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.

[3] Y. Kang, X. Tue, S. Dutta and M. Subbarao, Robust Depth-from-Defocus for Autofocusing in the Presence of Image Shifts, Proc. SPIE7066, Two- and Three-Dimensional Methods for Inspection and Metrology VI, 706609, 2008. International Journal of Computer Vision, vol.13 no.3, pp.271-294, 1994.

[4] M. Watanabe and S. K. Nayar, Rational Filters for Passive Depth from Defocus, International Journal of Computer Vision, vol.27 no.3, pp.203-225, 1998.

[5] S. Hiura and T. Matsuyama, Depth Measurement by the Multi-focus Camera, Proc. IEEE International Conference on Computer Vision and Pattern Recognition, pp. 953-959, 1998.

[6] A. Levin, R. Fergus, F. Durand and W. Freeman, Image and Depth from a Conventional Camera with a Coded Aperture, Proc. ACM SIGGRAPH, pp. 953-959, 2007.

[7] C. Zhou, S. Lin and S. Nayar, Coded Aperture Pairs for Depth from Defocus, Proc. IEEE 12th International Conference on Computer Vision, pp. 325-332, 2009.

[8] Martinello, M., Wajs, A., et al., Dual Aperture Photography: Image and Depth from a Mobile Camera, International Conference on Computational Photography, pp.1-10, 2015.

[9] M. Pianna and M. Bertero, Regularized Deconvolution of Multiple Images of the Same Object, Journal of the Optical Society of America A, vol.13 no.7, pp. 1516-1523, 1996.