CORRELATION BASED GUIDED DEPTH FILTERING IN SHAPE FROM FOCUS

Usman Ali, Muhammad Tariq Mahmood

School of Computer Science and Engineering, Korea University of Technology and Education,
1600, Chungjeol-ro, Byeongcheon-myeon, 31253, Cheonan, South Korea.

ABSTRACT

In the literature, shape-from-focus (SFF) techniques do not consider any guidance information for depth enhancement. As a result, reconstructed depth map might not accurately contain the structures of the scene. In this letter, we suggest to apply guided image filtering for the enhancement of depth map in SFF. For this, a novel guidance map for SFF has been crafted that is based on the correlation among the image sequence and focus volume along the optical axis. When this guidance map is incorporated for the guided filtering of depth map, a remarkable improvement has been achieved in the quality of reconstructed depth maps.

***Index Terms*—** Shape from focus (SFF), Focus measure, Correlation, Depth map

1. INTRODUCTION

Shape from focus (SFF) is a passive monocular technique that employs focus as a cue to reconstruct the 3D shape of the scene. At the first step of SFF, an image sequence is acquired by capturing a number of images of the scene by gradually varying the focus settings [1]. At the second step, the focus quality of each pixel in the image sequence is estimated by applying a focus measure (FM) operator and it provides an image focus volume [2]. At the next step, for each imaged point, the depth can be estimated by finding the image number in the image focus volume that gives the maximum focus measurement along the optical axis. In the literature, few methods have been proposed to improve the recovered depth map by enforcing smoothness constraints [3, 4]. However, it is to be noted that these methods try to improve the depth map without taking into consideration any other information about the geometry of the scene. As a result, reconstructed depth map might lack the accurate structures of the scene under consideration.

In this work, we not only suggest to improve the depth map in SFF through guided image filtering but also propose a guidance map. For an imaged point in the scene, we design its guidance information by measuring the correlation between its *image curve* and the *focus curve*. Image curve is the set of pixel intensities along the optical axis in the image sequence at that specific point. Similarly, the focus curve is the set of focus measurements along the optical axis in the focus volume. The correlation between image curve and focus curve provides a measure of structural similarity between them. The correlation is higher if the focus curve accurately incorporates the variations of the corresponding image curve. Consequently, the focus curve and the corresponding depth estimate at this imaged point are treated as reliable. On the other hand, if the focus curve does not adjusts well with the variations of the image curve, the correlation turns out to be insignificant. Thus, the focus curve and the corresponding depth estimate can is regarded as incorrect. The proposed guidance characterizes the correlation among the image curve and the focus curve and, when employed through guided image filtering, improves the shape reconstruction in SFF remarkably.

2. PROPOSED METHOD

The proposed SFF system has been depicted in Fig. 1 (a). At the first step, an initial depth map is estimated by applying the traditional SFF technique. Let the image sequence be represented by 3D volumetric data $I_z(x, y)$, where $x \in \{1, 2, \ldots, X\}$ and $y \in \{1, 2, \ldots, Y\}$ respectively indicate the number of rows and columns of each image and $z \in \{1, 2, \ldots, Z\}$ denotes the frame (image) number in the image sequence. For simplicity, we denote pixel coordinates by $p = (x, y)$. For a given input image sequence, focus or sharpness level for each pixel is measured by applying a focus measure operator. Though any suitable FM operator can be employed, we have applied the most popular sum-modified-Laplacian (SML) [1] which computes the focus value for a pixel through the second order derivative. As a result, an image focus volume $F_z(p)$ is obtained which is given by

$$F_z(p) = FM \otimes I_z(p) \,, \tag{1}$$

where $\otimes$ is the two-dimensional convolution operator. In the image focus volume, maximum focus measurement along the optical axis are determined for each imaged point. The image numbers corresponding to these maximum focus measurements make up the depth map. This depth map is termed as *initial depth map* and is given by,

$$D(p) = \arg\max_{z} \big(F_z(p) \big) \,. \tag{2}$$

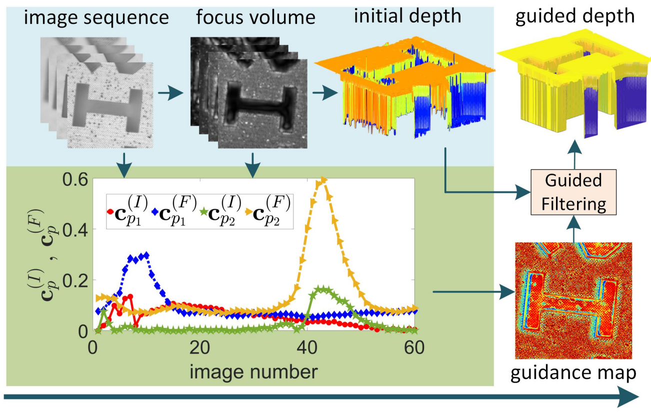

Fig. 1. Proposed SFF system. In the bottom left corner, the image and focus curves for two pixel positions $p_1 = (145, 60)$ and $p_2 = (210, 270)$ have been shown. Best viewed on screen.

Due to the limited capability of focus measure operators, image focus volume and the estimated depth map might be erroneous at certain points [5]. These errors result into large variations in the initial depth estimates of the closely spaced imaged points where the scene depth might have been same originally.

At the second step, correlation based guidance map is constructed. Image curve at p is the set of pixel intensities along the optical axis in the image sequence and is given by

$$\mathbf{c}_p^{(I)} = \{I_z(p)\}_{z=1}^{Z} . \tag{3}$$

Similarly, the focus curve at p in the focus volume is given by

$$\mathbf{c}_p^{(F)} = \{F_z(p)\}_{z=1}^{Z} . \tag{4}$$

It is to be noted that, generally, the focus curves follow the pattern of bitonic sequences, such that, $\mathbf{c}_p^{(F)} = F_1(p) \leq \cdots \leq F_z(p) \geq \cdots \geq F_Z(p)$ for some $z : 0 \leq z \leq Z$. However, each $\mathbf{c}_p^{(I)}$ roughly follows the pattern of its corresponding $\mathbf{c}_p^{(F)}$ either in the identical form or in the inverted form. The former is the case when there is a bright pixel at p in the relatively dark neighborhood, and later is the case when there is a dark pixel at p in the relatively bright neighborhood[1]. As a result, in the later case, the image curve is to be inverted first for the measurement of concurrent variations in the $\mathbf{c}_p^{(I)}$ and $\mathbf{c}_p^{(F)}$. This inversion can be achieved by,

$$\mathbf{c}{*}_p^{(I)} = |\mathbf{c}_p^{(I)} - \mathbf{c}_1^{(I)}| , \tag{5}$$

where $\mathbf{c}{*}_p^{(I)}$ is the inverted image curve. The guidance value at p is measured through the normalized cross correlation (NCC) between inverted image curve and the focus curve, and is expressed as,

$$G_p = \rho\big(\mathbf{c}{*}_p^{(I)}, \mathbf{c}_p^{(F)}\big) = \frac{cov\big(\mathbf{c}{*}_p^{(I)}, \mathbf{c}_p^{(F)}\big)}{\sigma\big(\mathbf{c}{*}_p^{(I)}\big)\, \sigma\big(\mathbf{c}_p^{(F)}\big)} , \tag{6}$$

where G is the guidance map, $cov\big(\mathbf{c}{*}_p^{(I)}, \mathbf{c}_p^{(F)}\big)$ is the covariance of $\mathbf{c}{*}_p^{(I)}$ and $\mathbf{c}_p^{(F)}$, $\sigma(.)$ denotes the standard deviation. The values of $\rho\big(\mathbf{c}{*}_p^{(I)}, \mathbf{c}_p^{(F)}\big)$ range from 0 to 1, with 0 representing no correlation, and 1 representing a direct positive correlation. When the inverted image curve and the focus curve have similar variations, even with different magnitudes, $\rho\big(\mathbf{c}{*}_p^{(I)}, \mathbf{c}_p^{(F)}\big) = 1$. $\rho\big(\mathbf{c}{*}_p^{(I)}, \mathbf{c}_p^{(F)}\big)$ is large when inverted image curve and the focus curve are similar. In Fig. 1, $\mathbf{c}{*}_p^{(I)}$ and $\mathbf{c}_p^{(F)}$ have been shown for two randomly selected pixel coordinates for letter *I* object. While viewing the curves $\mathbf{c}{*}_p^{(I)}$ and $\mathbf{c}_p^{(F)}$ (in red) at $p_1 = (145, 60)$, it has been observed that these curves have opposite trend for a considerable range of image numbers. Starting from image 40 to 60 approximately, the inverted image intensities decrease while the focus measurements increase. As a result, the correlation at this point $p_1 = (145, 60)$ is low and it turns out to be $\rho\big(\mathbf{c}{*}_p^{(I)}, \mathbf{c}_p^{(F)}\big) = 0.34$. However, the curves $\mathbf{c}{*}_p^{(I)}$ and $\mathbf{c}_p^{(F)}$ (in blue) at $p_2 = (210, 270)$ are more symmetrically aligned to each other. Consequently, a high value of 0.92 for the correlation is obtained at this point. We exploit these correlation values as the guidance values to filter the initial depth map in the next step.

At the third step, an improved depth map is obtained by filtering the initial depth map through the guidance map. For this, any of the suitable guided image filter can be used. Among the recently proposed guided image filters, the most popular is the guided image filter (GIF) [6]. This filter computes the filtering output $\hat{D}$ through the linear transform of guided image G in a window ω_k centered at the pixel k

$$\hat{D}_p = a_k G_p + b_k , \qquad \forall p \in \omega_k , \tag{7}$$

where (a_k, b_k) are two linear coefficients in the square window ω_k. The optimal values for these coefficients (a_k, b_k) are obtained by minimizing the cost function,

$$E(a_k, b_k) = \sum_{p \in \omega_k} \big[(a_k G_p + b_k - D_p)^2 + \lambda a_k^2 \big] , \tag{8}$$

where λ is a regularization parameter penalizing large a_k, D is the filtering input. To reduce the halo artifacts that may appear in GIF, effecitve guided image filter (EGIF) [7] and the related works mentioned in it have incorporated different weighting schemes for the optimization objective function. Being inspired by the achievements of these works, we introduce new weights and optimize the modified cost function

$$E(a_k, b_k) = \sum_{p \in \omega_k} \big[(a_k G_p + b_k - D_p)^2 + \lambda \sigma_k^2 a_k^2 \big] , \tag{9}$$

[1]Assuming high value is assigned to the bright pixels as compared to the dark pixels.

where σ_k^2 is the variance of the window centered at k,

$$\sigma_k^2 = \frac{1}{|\omega_k|} \sum_{p\in\omega_k} (G_p - \mu_k)^2 \,. \tag{10}$$

The optimal values of a_k and b_k are computed as,

$$a_k = \frac{\frac{1}{|\omega|} \sum_{p\in\omega_k} G_p D_p - \mu_k \bar{D}_k}{\sigma_k^2(1+\lambda)} \,, \tag{11}$$

$$b_k = \bar{D}_k - a_k \mu_k \,, \tag{12}$$

where μ_k is the mean of G in ω_k, $|\omega|$ is the number of pixels in ω_k and $\bar{D}_k$ is the mean of D in ω_k.

The final value of $\hat{D}_p$ is given as follows:

$$\hat{D}_p = \bar{a}_p G_p + \bar{b}_p \,, \tag{13}$$

where $\bar{a}_p = \frac{1}{|\omega|} \sum_{k\in\omega_p} a_k$ and $\bar{b}_p = \frac{1}{|\omega|} \sum_{k\in\omega_p} b_k$ are the mean values of a_k and b_k in the window ω_p, respectively. A modification introduced in Eq. 9 attenuates the halo artifacts and noise significantly.

3. RESULTS

Now, we present the experimental results. For this, one synthetic image sequence, named as cosine, and three real image sequences, named as cone, letter I and LCD filter, have been employed. More detail regarding the image sequences and how synthetic images are generated can be found in [8, 2]. In Fig. 2, the first column displays the guidance maps, second column is for initial depth maps and third column displays the filtered depth maps. For all these four objects, initial depth maps seem to have been contaminated with distortions and noisy spikes. However, these inaccuracies have been removed by filtering through proposed guidance map, and the resulting improved quality of depth maps can be seen in the last column. For the synthetic object cosine, the quality of reconstructed depth map has further been analyzed quantitatively. For this, in addition to SML, two more FM operators, gray level variance (GLV) and Tanengrad (TEN), have been employed to estimate their corresponding depth maps for cosine object. These initial depth maps have been improved by filtering through proposed guidance map and, for comparison, root mean square errors (RMSEs) have been computed for initial and guided depth maps. It can be seen in Fig. 3 that the RMSE decreases through guided filtering for all three FM operators. This indicates that the filtered depth maps are more closer to the ground truth as compared to the initial depth maps. As a result, accurate depth maps have been obtained through guided image filtering by employing proposed guidance map.

4. CONCLUSION

In this letter, a guidance map has been proposed for SFF. The proposed guidance map measures the correlation among the image sequence and focus volume along the optical axis. Guided image filter exploits this guidance map to improve the initial depth map. The experiments conducted on synthetic and real image sequences demonstrate the effectiveness of incorporating the proposed guidance map for the filtering of depth maps.

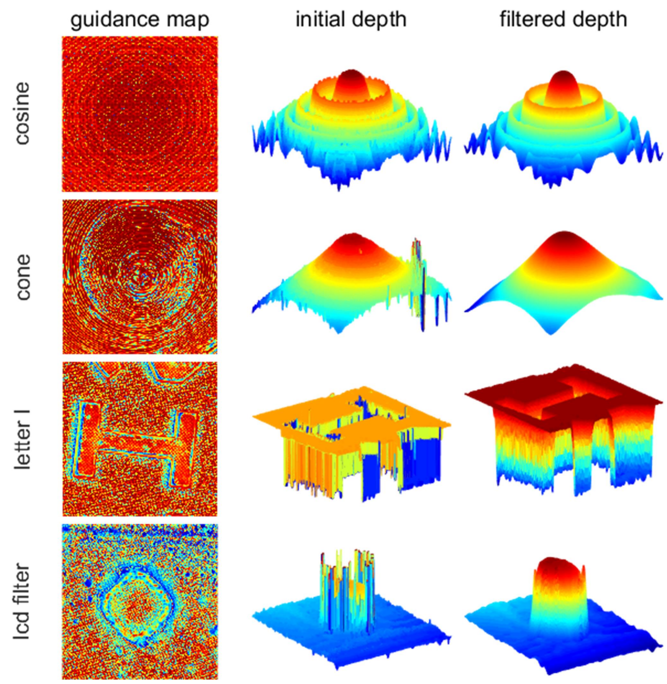

Fig. 2. Improvement in the quality of depth maps through guided image filtering for synthetic and real objects. (first column) Correlation based guidance maps, (second column) initial depth maps, (third column) guided filtered depth maps.

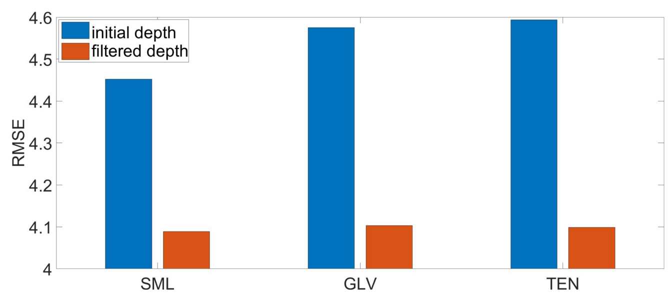

Fig. 3. Improvement in RMSE of synthetic object through guided depth filtering.

5. REFERENCES

[1] S. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 8, pp. 824–831, 1994.

[2] M. T. Mahmood and T.-S. Choi, "Nonlinear approach for

enhancement of image focus volume in shape from focus," *IEEE Transactions on image processing*, vol. 21, no. 5, pp. 2866–2873, 2012.

[3] V. Gaganov and A. Ignatenko, "Robust shape from focus via markov random fields," in *Proceedings of Graphicon Conference*, 2009, pp. 74–80.

[4] M. Moeller, M. Benning, C. Schönlieb, and D. Cremers, "Variational depth from focus reconstruction," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5369–5378, 2015.

[5] 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.

[6] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 6, pp. 1397–1409, 2013.

[7] Z. Lu, B. Long, K. Li, and F. Lu, "Effective guided image filtering for contrast enhancement," *IEEE Signal Processing Letters*, vol. 25, no. 10, pp. 1585–1589, 2018.

[8] M. Subbarao and M.-C. Lu, "Image sensing model and computer simulation for ccd camera systems," *Machine Vision and Applications*, vol. 7, no. 4, pp. 277–289, 1994.