Multi-focus Image Fusion for Extended Depth of Field

Wisarut Chantara and Yo-Sung Ho
School of Electrical Engineering and Computer Science
Gwangju Institute of Science and Technology (GIST)
{wisarut,hoyo}@gist.ac.kr

ABSTRACT
In this paper, we address the extended depth of field method, which appropriately applicable to a set of stack images. The proposed fusion algorithm improved a conventional discrete wavelet transform using spatial frequency and sum-modified-Laplacian. Both methods are applied to the approximation coefficient and the detail coefficients (horizontal, vertical and diagonal) for calculating in-focused regions. Finally, the inverse wavelet transform is utilized to obtain a final result image. The performance of the proposed method is conducted and compared with conventional fusion methods. Experiment results can demonstrate that the proposed method outperforms other reference methods.

CCS CONCEPTS
• Computing methodologies → Image processing;

KEYWORDS
Depth of field, discrete wavelet transform, spatial frequency, sum-modified-Laplacian

ACM Reference Format:

1 INTRODUCTION
The depth of field (DOF), also called focus range or effective focus range, is the distance between the nearest and farthest objects in a scene that appears acceptably sharp in an image. Extended DOF creates an image with an extended focal range from a series of images focused at a different depth, particularly useful for computer vision, digital photography, and macro photography. So that multi-focus image fusion is the process which combines two or more different focal image with the same scene into a single image, maintaining the significant information from each of the stack images. Therefore, all important objects in the image are in-focused. Generally, the digital camera can set only one distance focus at a time. Then, to acquire a set of stack images with the same scene by captured in a single shot, a light field camera such as Lytro camera [8] and Raytrix camera [10] provide a set of the stack images with different focal display in the same view.

In recent year, many depth of field extension techniques have been developed, e.g., Li et al. [6], introduced a method based on the selection of image blocks from source images to construct the fused image using spatial frequency. While Desale et al. [2], proposed a study and analysis of image fusion technique by way of PCA, DCT, and DWT. Moreover, Wang et al. [11], presented multisource image fusion using spatial frequency and simplified pulse coupled neural network. In this paper, we proposed an improved discrete wavelet transform using spatial frequency and sum-modified-Laplacian to detect focal regions and reconstruct in-focused regions to an all-in-focused image.

2 PROPOSED METHOD
The following techniques are involved in the proposed methods implementation. Light field camera [8] is utilized as a capture device. Light field picture splitter [9] is applied to split the raw file into different focus level images. This process provided a set of different focal images which display the same position as shown in Fig. 1.

Figure 1: A set of different focal images.
2.1 Discrete wavelet transform (DWT)

Discrete wavelet transform uses a cascade of special low-pass and high-pass filter and a sub-sampling operation. The output from first-order of DWT contains four decomposition parts as shown in Fig. 2. Those are CA, CH, CV and CD, where CA is the approximation coefficient, which is sensitive to human eyes [5]. While CH, CV, and CD are the detail coefficient (horizontal, vertical and diagonal) which have more detail information more than CA. Since DWT of image signals produces a nonredundant image representation, it can provide better spatial and spectral localization of image information.

2.2 Spatial frequency (SF)

Spatial frequency, which originated from the human visual system, indicates the overall active level in an image. It is difficult to completely comprehend the human visual system with current physiologic means. While the use of SF has led to an effective contrast criterion for image fusion [7]. SF is defined as:

\[ SF = \sqrt{(CF)^2 + (RF)^2} \]  

(1)

where RF and CF are the row frequency and column frequency respectively:

\[ RF = \sqrt{\frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=2}^{N} [I(x, y) - I(x, y - 1)]^2} \]  

(2)

\[ CF = \sqrt{\frac{1}{M \times N} \sum_{j=1}^{M} \sum_{i=2}^{N} [I(x, y) - I(x - 1, y)]^2} \]  

(3)

for \( M \times N \) image with a grayscale \( I(x, y) \) at \( (x, y) \) coordinate.

2.3 Sum-modified-Laplacian

Sum-modified-Laplacian is developed to compute a local measure of the quality of image focus. SML can provide better performance in focus measurement criterion [4]. SML is defined as:

\[ SML(x, y) = \sum_{i=x-N}^{i=x+N} \sum_{j=y-N}^{j=y+N} \nabla^2_{ML} I(i, j) \]  

(4)

for \( \nabla^2_{ML} I(i, j) \geq T_{SML} \)

where \( \nabla^2_{ML} I(i, j) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \)

2.4 Enhancement SML (eSML)

In the homogeneous region, the focus measure can be affected by the pixel noise [1]. In order to decrease this effect, the SML values are computed in a small window to determine the eSML focus measure of the center pixel \( (x, y) \) as:

\[ eSML = \sum_{i=x-N}^{i=x+N} \sum_{j=y-N}^{j=y+N} SML(i, j) \]  

(5)

2.5 Image fusion

The wavelet decomposition coefficient structures (approximation and detail coefficients) are combined as:

\[ F_{i,j} = D_{i,j}^m \]  

(6)

where \( m = \arg \max_t C_{i,j}^t, t = 1, 2, ..., N \)

where \( F_{i,j} \) is a final coefficient in an all-focused image at \( (i, j) \). While \( D_{i,j}^m \) is a maximum coefficient information at \( (i, j) \). \( C_{i,j}^t \) is a coefficient information at \( t \) stack image. \( N \) is a set of input stack images.

2.6 The proposed method procedure

A summary of the proposed method is provided as follow:

1. Apply the DWT on each image of the input stack images,
2. Apply the SF by using Eq. (1) to (3) for the approximation coefficients (CA),
3. Apply the SML by using Eq. (4) for the detail coefficients (CH, CV, and CD),
4. Apply the eSML by using Eq. (5) for the noise reduction improvement in the detail coefficients,
5. Combine each coefficient of the approximation and detail coefficients by using Eq. (6), and
6. Apply an inverse discrete wavelet transform to obtain a multi-focus image with extended depth of field as a final image.

3 EXPERIMENT RESULTS

The experiments are conducted to compare the performance of the proposed method with those of popular widely conventional methods (such as pixel averaging method, DWT-averaging method [2], DWT-maximum method [2], and the spatial frequency method). We conducted experiments on 512 x 512 pixels sample images. The test images are a set of cup and stuff images, which contain multiple objects at different focus level images capturing by light filed camera [8]. In all our experiments, we set \( T_{SML} \) to be 10. The experiment results are presented in Fig. 3 and 4. From the fusion results, we can easily observe that the results of the pixel averaging and DWT methods have a lower contrast.
than those of the spatial frequency method and the proposed method. However, it is hard to tell the difference between the results of the spatial frequency method and the proposed method by subjective evaluation. Hence, the paper applies some non-reference fusion metrics such as Feature Mutual Information (FMI) [3] and Petrovics metric ($Q^{AB/F}$) [12] are then introduced and employed. These evaluation criteria metrics are calculated without respect to the reference images. FMI measures the amount of information that the fused image contains the source images, while $Q^{AB/F}$ measures the relative amount of edge information that is transferred from the source into the fused image. The higher the FMI or $Q^{AB/F}$ value, the better the fused image performance. The comparison results are summarized in Table 1.

The above two evaluation criteria are then applied to evaluate the four fusion methods in Fig. 3 and 4, the detailed quantitative results are given in Table 1. From Table 1, we can observe that the values of all quality indices of the proposed method are larger than those of pixel averaging, DWT-averaging, DWT-Maximum, and the conventional spatial frequency methods, which means the proposed algorithm can effectively combine sharp parts of the original image to the fused image, and yield superior quality than conventional methods.

### REFERENCES


### ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIP) (No. 2011-0030079).
Figure 3: Comparison of "Cup": (a) Pixel Avg., (b) DWT-Avg., (c) DWT-Max, (d) SF, (e) Proposed method.

Figure 4: Comparison of "Stuff": (a) Pixel Avg., (b) DWT-Avg., (c) DWT-Max, (d) SF, (e) Proposed method.