A MULTI-EXPOSURE IMAGE FUSION BASED ON THE ADAPTIVE WEIGHTS REFLECTING THE RELATIVE PIXEL INTENSITY AND GLOBAL GRADIENT

Sang-hoon Lee, Jae Sung Park, Nam Ik Cho

Seoul National University
Department of Electrical and Computer Engineering
1, Gwanak-ro, Gwanak-gu, Seoul, 08826, Republic of Korea

ABSTRACT

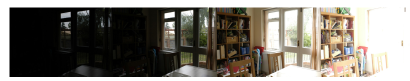
This paper presents a new multi-exposure fusion algorithm. The conventional approach is to define a weight map for each of the multi-exposure images, and then obtain the fusion image as their weighted sum. Most of existing methods focused on finding weight functions that assign larger weights to the pixels in better-exposed regions. While the conventional methods apply the same function to each of the multi-exposure images, we propose a function that considers all the multi-exposure images simultaneously to reflect the relative intensity between the images and global gradients. Specifically, we define two kinds of weight functions for this. The first is to measure the importance of a pixel value relative to the overall brightness and neighboring exposure images. The second is to reflect the importance of a pixel value when it is in a range with relatively large global gradient compared to other exposures. The proposed method needs modest computational complexity owing to the simple weight functions, and yet it achieves visually pleasing results and gets high scores according to an image quality measure.

***Index Terms*—** Image fusion, dynamic range, image enhancement

1. INTRODUCTION

The dynamic range of a camera is usually narrower than that of most of the scenes that we wish to take. Whatever the bit-depth of a camera is, it is considered to have the relatively low dynamic range (LDR) compared to the scenes with high dynamic range (HDR). Hence, for capturing such HDR scenes with the LDR camera, the most common approach is to take pictures several times while changing the exposure time from short to long [1] and merge to an HDR one.

On the other hand, for displaying the synthesized HDR image on an LDR display device, we need a tone-mapping process for compressing the HDR into the LDR [2, 3]. When the target displays are only the LDR ones, we may directly synthesize a tone-mapped-like LDR image from the multi-exposure images. The most common approach for this purpose is to define a weight map for each of the multi-exposure images and synthesize a final tone-mapped-like image as a weighted sum of the images, which is called the Multi-Exposure image Fusion (MEF) algorithm.



(a) Multi-exposure image sequence

(b) Fused image

Fig. 1. Demonstration of MEF. (a) A multi-exposure image sequence courtesy of Tom Mertens and (b) the result of our fusion method.

The conventional MEF methods are mostly pixel-wise ones, i.e., the weight maps are also images with the same size as the input, and the weight reflects the importance of the corresponding pixel in the input images. Hence, finding the appropriate weight maps is the most important task in this approach. For examples, Burt *et al.* [4] used Laplacian pyramid decomposition and computed weight maps using local efficient energy and correlation between the pyramids. Mertens *et al.* [5] defined several metrics that reflect the pixel quality, such as contrast, saturation, and well-exposedness. Raman *et al.* [6] and Zhang *et al.* [7] detected regions with rich information in an image, which can be obtained from computing gradient magnitudes [6] or a bilateral filtering process [7]. Vonikakis *et al.* [8] computed weight maps by illumination estimation. Since weight maps are often noisy, Li *et al.* [9, 10] refined weight maps using edge-preserving filters such as the recursive filter [11] or the guided filter [12]. Shen *et al.* proposed a random walk approach to fuse images. There are also some patch-wise methods, for example, Goshtasby *et al.* [13] divided the input images into non-overlapping patches and took the ones with the highest entropy from each image. Ma et al. [14] extracted patches in each image and decomposed the patches into the signal strengths, signal structures and mean intensity to synthesize a new patch. Prabhakar et al. [15] used the Convolutional Neural Network (CNN) to design a new network for the MEF.

In this paper, we introduce a new MEF algorithm which focuses on an efficient and effective weight function design. In most of the conventional pixel-wise MEF methods, the weight is obtained as a function of pixel values within an image in the set of multi-exposure images. In other words, existing methods usually apply the same rule to each image in the set whereas our method applies an adaptive rule using all the images. Specifically, we define two weight functions which may reflect the quality of pixels. The first one is to represent the pixel quality concerning the overall brightness of an input image and those of neighboring exposure images. The weight is designed to be large in bright areas of the under-exposed image and to be small in bright areas of the over-exposed one. The second weight reflects the importance of a pixel when its value is in a range with relatively large global gradient compared to other exposure images. The final weight is the combination of these two weights. The proposed method needs modest computational complexity owing to the simple weight function, and yet it achieves visually pleasing results and gets high scores according to an image quality measure.

2. DESIGN OF WEIGHT FUNCTIONS

As stated above, we design two kinds of weights: one is based on the pixel intensities and the other is based on the global gradients.

2.1. Weight design based on the pixel intensity

The conventional pixel-wise MEF generates the fusion image $I_{fused}(x,y)$ as a weighted sum of multi-exposure images, which can be expressed as

$$I_{fused}(x,y) = \sum_{n=1}^{N} W_n(x,y) I_n(x,y), \tag{1}$$

where N is the number of images in a set of multi-exposure images, $I_n(x,y)$ is the pixel intensity of the n-th image in the set and $W_n(x,y)$ denotes the weight that represents the importance of the pixel $I_n(x,y)$. In this paper, we normalize the pixel intensity into the range of $[0,1]$.

The most important part of MEF algorithm is how to design $W_n(x,y)$, which needs to be large for pixels in well-exposed regions and vice versa. As an example of most intuitive and simple weight design, Mertens *et al.* [5] introduced a quality measure named as *well-exposedness* to be used as the weight:

$$W_n(x,y) = \exp\left(-\frac{(I_n(x,y)-0.5)^2}{2\sigma^2}\right), \tag{2}$$

where σ equals 0.2. The measure favors pixels in the unsaturated regions which have intensity values closed to 0.5. It can also be seen that the function has only the n-th image as its variable. Based on this form of the weight, we add our observation about multi-exposure image sequences. First of all, the well-exposedness above cannot give large weights to the pixels in dark or bright regions that are far from 0.5 which may be important at times. Specifically, this weight cannot highlight the dark regions that are well captured in the long exposure image and/or the bright regions in the short exposure images. Hence we control the weight to be relative to the overall brightness of the image. Specifically, the dark regions are given large weights when the overall image is bright (long exposure), and vice versa. Denoting the mean of pixel intensities of the n-th image as m_n, the weight should be large when $I_n(x,y)$ is close to $1-m_n$, which can be encoded as $\exp(-(I_n(x,y)-(1-m_n))^2)$ to represent the weight in the form of Eq. (2). In addition to this, we note that there are more well-exposed pixels when the input and the neighboring exposure image have a larger difference. Hence, we give large σ_n when the mean brightness of an image (m_n) is much different from the neighboring ones (m_{n-1}, m_{n+1}). Finally, the first weight $W_{1,n}(x,y)$ that encodes the relative brightness is represented as

$$W_{1,n}(x,y) = \exp\left(-\frac{(I_n(x,y)-(1-m_n))^2}{2\sigma_n^2}\right), \tag{3}$$

where σ_n controls the weight according to the difference of m_n's as

$$\sigma_n = \begin{cases} 2\alpha(m_{n+1}-m_n) & n=1 \\ \alpha(m_{n+1}-m_{n-1}) & 1<n<N \\ 2\alpha(m_n-m_{n-1}) & n=N. \end{cases} \tag{4}$$

α is set to 0.75 in all the experiments. From the equation, we can see that when m_n is close to 1 (when the overall brightness of the image is high), the dark pixels ($I_n(x,y)$ with low values) will be given larger weights and vice versa. Also, the weight is given large value when the mean brightness is largely different between the neighboring exposure images.

2.2. Weight design based on the global gradient

In a low exposure image, the pixels in dark regions are saturated to the values close to 0, whereas the pixel values in bright regions have large variation. Thus the bright regions usually have high contrast (large gradient in pixel values). In the case of a high exposure image, the opposite relation holds.

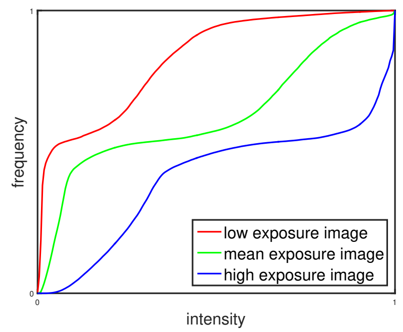


Fig. 2. Cumulative histograms of a multi-exposure image sequence.

Based on these observations, some methods give high weights to the pixels in the high contrast regions [5, 6, 7]. However, the regions without texture have low contrast or small gradient values regardless of the degree of exposures. Hence stressing only the large gradient region may fail to emphasize the pixels in a well-exposed region with small gradients.

In this respect, we propose another weight that emphasizes the well-exposed regions regardless of their local contrast. This is based on the observation of cumulative histogram of the pixel intensity (integration of histogram along the pixel intensity axis) as shown in Fig. 2. This figure is an example of cumulative histograms of low, mid and high-exposure images of "Tower" sequence in [16]. As expected, the low-exposure image has many saturated pixels close to zero so that the cumulative histogram drastically increases at the beginning. Hence, we can say that a pixel is in a well-exposed region when it is in the range that the cumulative histogram changes slowly. In other words, the pixel value in this region is a relatively rare one so that the pixel values around them have large variation. In the case of mid and high exposure images (green and blue respectively), the cumulative histograms have smaller gradients at lower pixel values than the case of low exposure image. This also means that the pixels with lower values are in the well exposed and/or high variation regions. Hence, we need to give a larger weight to the pixel when it is in the range of cumulative histogram with the smaller gradient. Formally, we can encode this observation into the weight:

$$W_{2,n}(x,y) = \frac{Grad_n(I_n(x,y))^{-1}}{\sum_{n=1}^{N} Grad_n(I_n(x,y))^{-1} + \epsilon}, \qquad (5)$$

where ϵ is a very small positive value to prevent the denominator from being zero, and $Grad_n(I_n(x,y))$ denotes the gradient of the cumulative histogram at the intensity $I_n(x,y)$. Because the gradient above is not the local gradient around the pixel, but the gradient with respect to other remote pixels that are in the similar range, we call it as the global gradient.

3. FUSION

The final weight for each image is calculated by combining two weights with normalization as

$$W_n(x,y) = \frac{W_{1,n}(x,y)^{p_1} \times W_{2,n}(x,y)^{p_2}}{\sum_{n=1}^{N} W_{1,n}(x,y)^{p_1} \times W_{2,n}(x,y)^{p_2} + \epsilon}, \qquad (6)$$

where $p_1, p_2 > 0$ are the parameters to decide which one to emphasize more. But, in our experiments, we set these parameters the same ($p_1 = p_2 = 1$).

With the weights obtained above, we can fuse the image according to Eq. (1). However, this method often yields an unnatural image with artifacts because the weight values are often discontinuous and noisy. Thus, we apply Eq. (1) in a multi-resolution fashion using the pyramidal image decomposition introduced in [5]. The MEF is processed in each pyramid and the final result is synthesized.

4. EXPERIMENTAL RESULTS

The proposed method is tested on various multi-exposure image sequences from [16]. All the test sequences are the static ones, and each sequence has a different characteristic like indoor or outdoor, with or without a lighting fixture, including many or few multi-exposures. First, we explain the effectiveness of each weight as an ablation study. Then, the proposed method is compared to the existing MEF methods subjectively and also by using an image quality measure, called MEF-SSIM [16].

Fig. 3 compares two fusion images: one is based on the conventional non-adaptive weight in Eq. (2) and the other is based on our first weight in Eq. (3) which is adaptive to the image sequence. We believe that the right one is more natural because the lamp is on the right side of the desk whereas the left result shows similar brightness on all the regions. On the other hand, our result shows that the desk on the right side is brighter than others, which coincides with our insight. In Fig. 4, we verify the effectiveness of the second weight by comparing the result with Eq. (3) only and the result with

(a) conventional weight (b) our weight

Fig. 3. Comparison of result images (a) using the conventional weight in Eq. (2) and (b) our first weight in Eq. (3). Image sequence by courtesy of Martin Cadik

Table 1. MEF-SSIM scores of the proposed method and existing methods on ten input sequences. The scores range from 0 to 1 and a higher value indicates better perceptual quality. The best and second-best results are shown in boldface.

Input Sequence	Gu12	Song12	Z.Li12	S.Li12	Zhang12	S.Li13	Mertens09	Ma15	proposed
Balloons	0.913	0.918	0.941	0.945	0.945	0.948	**0.969**	0.963	**0.971**
Cave	0.934	0.918	0.923	0.961	0.952	**0.978**	0.974	**0.980**	0.977
Chinese garden	0.927	0.967	0.951	0.982	0.978	0.984	**0.989**	0.988	**0.990**
Farmhouse	0.932	0.947	0.959	0.977	0.972	**0.985**	0.981	**0.983**	0.978
Lamp	0.871	0.829	0.933	0.937	0.932	0.934	**0.948**	0.945	**0.954**
Landscape	0.941	0.944	0.948	0.972	**0.984**	0.942	0.976	**0.991**	0.981
Madison Capitol	0.864	0.945	0.949	0.918	0.952	0.968	**0.977**	0.974	**0.978**
Office	0.900	0.961	0.954	0.972	0.968	0.967	0.984	**0.986**	**0.989**
Tower	0.932	0.939	0.950	0.984	0.974	**0.986**	0.986	0.981	**0.987**
Venice	0.889	0.942	0.937	0.952	0.962	0.954	0.966	**0.978**	**0.973**
Average	0.910	0.931	0.945	0.960	0.962	0.965	0.975	**0.977**	**0.978**

(a) without the second weight (b) with the second weight

Fig. 4. Comparison of result images by using (a) the first weight in Eq. (3) only and (b) the combined weight in Eq. (6). Image sequence by courtesy of Bartlomiej Okonek

Eq. (6). The left image includes over-exposed region, while the right one does not. We can see that the second weight helps to keep the contrast in bright areas (outside of the cave).

The proposed method is compared with eight existing MEF methods, Mertens09 [5], Gu12 [17], Zhang 12 [18] Song12 [19], Z.Li12 [18], S.Li12 [9], S.Li13 [10] and Ma15 [14] using the MEF-SSIM in [16]. This image quality assessment (IQA) model is developed for the multi-exposure image fusion, which is based on the multi-scale structural similarity (SSIM) [20] to consider not only local structure preservation but also global luminance consistency. The comparison of the proposed method and other existing methods on 10 multi-exposure image sequences is shown in Table 1, where we can see the superior performance of the proposed method. The proposed method gets comparable MEF-SSIM scores with Mertens09 [5] and Ma15 [14] which get much higher scores than others.

The visual comparison is not shown here due to limited space, and also because the recent methods show similar performance as can be expected from the Table 1. For the visual comparison of original size images, refer to https://github.com/tkd1088/multi-exposure-image-fusion where we will make our codes also available.

Table 2. Comparison of average execution time of the proposed method with existing MEF methods.

Methods	Mertens09	S.Li13	Ma15	proposed
Time(s)	0.581	1.156	2.455	0.737

To compare the computational complexity, we measure the average execution time of MATLAB codes for several fast performing algorithms as shown in Table 2. The proposed method is three times as fast as Ma15 [14] which shows comparable performance with the proposed method.

5. CONCLUSIONS AND FUTURE WORK

We have proposed a pixel-wise exposure fusion method, based on new weight function designs. The first function is to emphasize the bright areas in low-exposure images and vice versa, and also to increase the weight when the exposure change brings large brightness change. The second is to suppress the saturated areas and to emphasize the regions with highly variational pixel values. Experimental results show the effectiveness of our two weight functions which require modest computational complexity. As a future work, we are going to modify the proposed method to be applied to dynamic scenes where moving objects cause the ghost artifacts.

Acknowledgements

This research was supported by Projects for Research and Development of Police science and Technology under Center for Research and Development of Police science and Technology and Korean National Police Agency (PA-C000001).

6. REFERENCES

[1] Paul E Debevec and Jitendra Malik, "Recovering high dynamic range radiance maps from photographs," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1997, pp. 369–378.

[2] Frédo Durand and Julie Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," in *ACM transactions on graphics (TOG)*. ACM, 2002, vol. 21, pp. 257–266.

[3] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda, "Photographic tone reproduction for digital images," *ACM transactions on graphics (TOG)*, vol. 21, no. 3, pp. 267–276, 2002.

[4] Peter J Burt and Raymond J Kolczynski, "Enhanced image capture through fusion," in *Computer Vision, 1993. Proceedings., Fourth International Conference on*. IEEE, 1993, pp. 173–182.

[5] Tom Mertens, Jan Kautz, and Frank Van Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," in *Computer Graphics Forum*. Wiley Online Library, 2009, vol. 28, pp. 161–171.

[6] Shanmuganathan Raman and Subhasis Chaudhuri, "Bilateral filter based compositing for variable exposure photography.," in *Eurographics (Short Papers)*, 2009, pp. 1–4.

[7] Wei Zhang and Wai-Kuen Cham, "Gradient-directed multiexposure composition," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2318–2323, 2012.

[8] Vassilios Vonikakis, Odysseas Bouzos, and Ioannis Andreadis, "Multi-exposure image fusion based on illumination estimation," in *Proc. IASTED SIPA*, 2011, pp. 135–142.

[9] Shutao Li and Xudong Kang, "Fast multi-exposure image fusion with median filter and recursive filter," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 2, 2012.

[10] Shutao Li, Xudong Kang, and Jianwen Hu, "Image fusion with guided filtering," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2864–2875, 2013.

[11] Eduardo SL Gastal and Manuel M Oliveira, "Domain transform for edge-aware image and video processing," in *ACM Transactions on Graphics (ToG)*. ACM, 2011, vol. 30, p. 69.

[12] Kaiming He, Jian Sun, and Xiaoou Tang, "Guided image filtering," in *European conference on computer vision*. Springer, 2010, pp. 1–14.

[13] A Ardeshir Goshtasby, "Fusion of multi-exposure images," *Image and Vision Computing*, vol. 23, no. 6, pp. 611–618, 2005.

[14] Kede Ma and Zhou Wang, "Multi-exposure image fusion: A patch-wise approach," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1717–1721.

[15] K Ram Prabhakar, V Sai Srikar, and R Venkatesh Babu, "Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs," in *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2017, pp. 4724–4732.

[16] Kede Ma, Kai Zeng, and Zhou Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3345–3356, 2015.

[17] Bo Gu, Wujing Li, Jiangtao Wong, Minyun Zhu, and Minghui Wang, "Gradient field multi-exposure images fusion for high dynamic range image visualization," *Journal of Visual Communication and Image Representation*, vol. 23, no. 4, pp. 604–610, 2012.

[18] Zheng Guo Li, Jing Hong Zheng, and Susanto Rahardja, "Detail-enhanced exposure fusion," *IEEE Transactions on Image Processing*, vol. 21, no. 11, pp. 4672–4676, 2012.

[19] Mingli Song, Dacheng Tao, Chun Chen, Jiajun Bu, Jiebo Luo, and Chengqi Zhang, "Probabilistic exposure fusion," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 341–357, 2012.

[20] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.