Low Complexity Detail Preserving Multi-Exposure Image Fusion for Images with Balanced Exposure

Ashish V. Vanmali, Sanket S. Deshmukh, and Vikram M. Gadre
Department of Electrical Engineering, IIT Bombay, Mumbai 400076, India
Email: ashish@ee.iitb.ac.in, sanelec@ee.iitb.ac.in, vmgadre@ee.iitb.ac.in

***Abstract*—Multi-exposure image fusion has undergone considerable growth in the last few years, but the challenge still remains to design an algorithm which is efficient in fusion yet low in complexity, such that it can be implemented for real time applications on embedded platforms which have low computational power. A simple, low complexity algorithm for multi-exposure image fusion is proposed in this paper to cater to this need. The algorithm is designed for the image sequences that spread evenly over the exposure range, without any bias towards under or overexposedness. The fusion algorithm assigns weights to pixels of images to be fused, based on the exposure of the input images. The fusion algorithm is free from filtering and transformation, and is carried out on per pixel basis, thus making it highly efficient in terms of computational complexity. The experimental results show that our algorithm produces visually comparable results with some of the commonly used algorithms, which are computationally much more complex. Also, the proposed algorithm is immune to ghosting artifacts caused due to spatial misalignment of the input images.**

I. Introduction

Real world scenes are of high dynamic range (HDR) as opposed to the scenes captured by cameras, which can capture only a limited dynamic range (typically 8 bits per color channel). Images captured by cameras have a low dynamic range (LDR), though the actual scene may have an extended range of intensities. Under such circumstances the captured scenes are either underexposed or overexposed, thereby missing out information in the dark or bright regions. To tackle this issue, one can capture a series of multi-exposure LDR images (exposure bracketed images) by varying the exposure time or the aperture size, or by neutral density filters; each highlighting different details of a scene. These images can then be fused to form an HDR representation of the scene capturing all details.

Several techniques are proposed in the past to combine multi-exposure images. One of the approaches is to create an HDR image from multi-exposure images. This process is carried out in three steps. We review the steps from the text of Szeliski [1]: estimating the radiometric response function, estimating a radiance map by selecting or blending pixels from different exposures and tone mapping the resulting HDR images back into displayable gamut.

Readers can refer to the work of many previous researchers [1], [2], [3], [4] for details regarding the estimation of response function and radiance map. The resultant HDR image cannot however be displayed on normal display devices, as these devices have a low dynamic range. The HDR image is therefore mapped to a LDR image by an operation called tone mapping. Details of various tone mapping operators can be found in a number of references [1], [5], [6]. The major limitation of this approach is that it requires camera parameters (exposure times) to calculate the radiometric response function, which might not be known to the user. Secondly, it is required to tone map the HDR image back to LDR for practical purposes.

Another approach is to fuse the LDR images directly, without obtaining the intermediate HDR image by using appropriate fusion technique.

Goshtasby [7] proposed a region based fusion technique. In his approach, the images are partitioned into sub blocks. The final image is constructed by selecting the sub blocks having the maximum entropy. A blending function is used to remove the blocky artifacts. It uses gradient-ascent algorithm to optimize block size and blending function. This makes the algorithm computationally complex and slow.

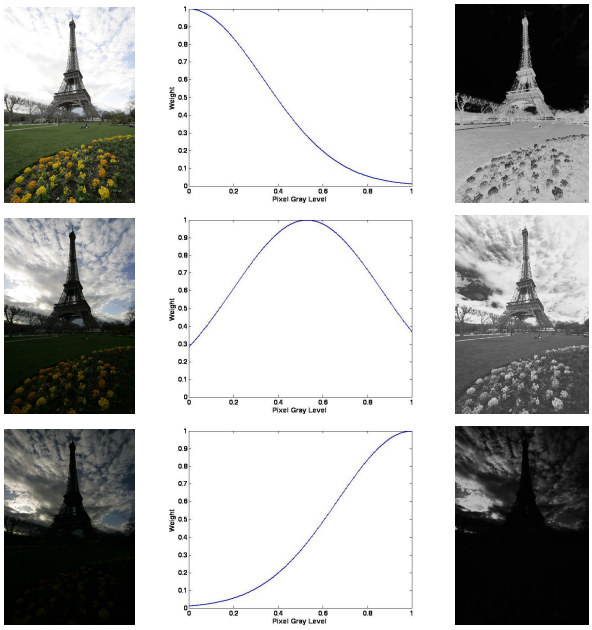
Mertens et al. [8] presented a multi-resolution pixel wise fusion algorithm. Each pixel in the input image is assigned a weight based on contrast, saturation and well exposedness. The fusion is carried out using Laplacian decomposition of the images and Gaussian pyramid of the weight maps. The fused image is vibrant in color and contrast. However, the technique cannot capture details in extreme bright and dark regions of image sequence. Also for proper blending, the pyramidal decomposition has to be carried out up to the level of one pixel, which makes the algorithm slow as the number of images and the image size increases.

Malik et al. [9] have extended the work of Merten et al. in the Wavelet domain. Li et al. have used a symmetric analysis-synthesis filter bank with local gain control to the sub bands [10].

Raman et al. have proposed a matte based technique for fusion using edge preserving bilateral filters for creating matte [11]. The fusion process is then guided by normalized matte functions.

Most recently Kotwal et al. [12] proposed an optimization based approach for multi exposure image fusion. The algorithm defines a multi-objective cost function based on the desired characteristics of the image such as high entropy and high contrast. An iterative solution is obtained using a variational method. The algorithm has excellent detail preserving capability, but at the same time cannot preserve color information and is computationally expensive. The iterative solution makes it very slow and unsuitable for implementation

978-1-4673-5952-8/13/$31.00



(a) (b) (c)

Fig. 1: Images and their weight maps
(a) Eiffel tower image sequence (Images courtesy: Jacques Joffre), (b) Weight map functions, (c) Weight maps

on an embedded platform.

Kartalov et al. [13] gave a simple approach for fusion using luminance transformation and selection of hue and saturation, having a very low computational complexity. The only limitation in this work was that the algorithm is developed only for two input images, one underexposed and the other overexposed.

In this paper, we present a low complexity detail preserving algorithm for the fusion of multi-exposure images. The fusion is performed pixel-by-pixel and does not involve any filtering or transformation. This makes the algorithm very fast and computationally efficient. The experiments show that the resulting image quality is comparable with the popular algorithms. Also, the algorithm is successful in getting a ghost free output while dealing with input images having a slight spatial misalignment.

II. Proposed Algorithm

For the images which are under exposed, dark regions do not convey any information. Similarly, for over exposed images, the saturated regions do not convey any information. Hence, when images are weighted, such regions should have minimum weight and bright and dark regions in under and overexposed images respectively, should have higher weights. This fact governs our algorithm.

Consider $I_k, k = 1, 2, ..., N$ to be the set of N color images with multiple exposures with $I_{Mono,k}$ representing their gray scale versions. The pixel values are assumed to be normalized in the range of 0 to 1. We use a Gaussian function to construct the weight map of these images. Let m_k denote the average gray level of the k^{th} image. The weight map function for the k^{th} image is then given as:

$$W_k(x,y) = e^{-\left(\frac{(I_{Mono,k}(x,y)-\mu_k)^2}{2\sigma^2}\right)} \tag{1}$$

where,

$$\mu_k = 1 - \frac{m_k - \min\{m_{k,k=1,...,N}\}}{\max\{m_{k,k=1,...,N}\} - \min\{m_{k,k=1,...,N}\}} \tag{2}$$

Here, $I_{Mono,k}(x,y)$ denotes the gray level of the k^{th} image at pixel location (x,y) and $W_k(x,y)$ represents the weight assigned to that pixel. The parameter μ_k decides the location of the peak of the Gaussian weight map function depending on the image exposure. For the most overexposed image, the peak will be located at 0, and as the exposure decreases, the peak will move towards 1. For the most under exposed image, the peak will be located at 1. The parameter σ controls the spread of Gaussian weight map function and is given as $\sigma = 1/N$. For five or more images, we freeze the value of σ to 0.2 to ensure that there is significant overlap between successive functions for consistent results. The weight maps so obtained are normalized by:

$$W_{Norm,k}(x,y) = \frac{W_k(x,y)}{\sum_{k=1}^{N} W_k(x,y)} \tag{3}$$

This ensures that the weight maps sum to one at each pixel location (x,y).

An example of weight maps is shown in Figure 1. Here, Figure 1(a) shows the input sequence. Figure 1(b) shows the corresponding weight map functions based on the exposure of the images. Figure 1(c) shows the weight maps according to equation (3). One can clearly observe that these weight maps are consistent, and they do not have any spurious transitions. This enables the fusion process without any need of multi-resolution blending, thereby reducing computations.

The final fused image F is obtained by a pixel-by-pixel weighted sum of the input images:

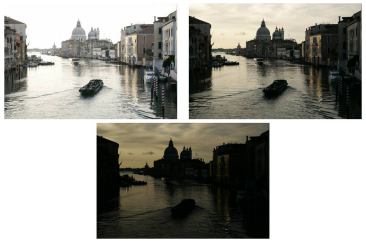
$$F(x,y) = \sum_{k=1}^{N} W_{Norm,k}(x,y) \cdot I_k(x,y) \tag{4}$$

For color images, equation (4) is performed separately on R, G and B color channels to obtain the fused RGB image.

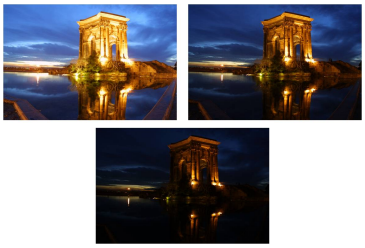
III. Experimental Results And Analysis

We have implemented the proposed algorithm in Matlab and have tested for sixteen different datasets. For the sake of brevity, we present the results for five datasets shown in Figure 1(a) and Figure 2. For comparison of quality of fusion, the results are compared against the results of Mertens et al. [8] and Kotwal et al. [12].

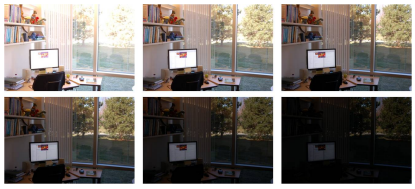
Figure 3, 4 and 5 show the results for the multi-exposure image sequence of 'Eiffel tower', 'Grandcanal' and 'Office' respectively. It can be observed that the results generated by the algorithm of Mertens et al.'s exhibit the best visual



(a) Grandcanal (Images courtesy: Jacques Joffre)



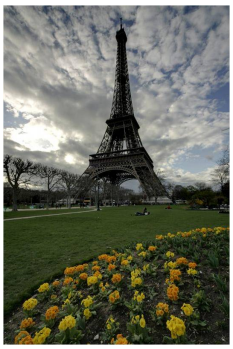
(b) Peyrou (Images courtesy: Jacques Joffre)



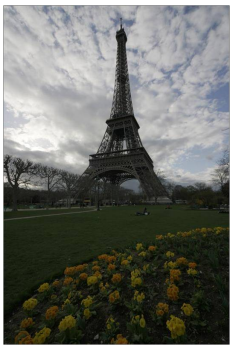
(c) Office (Images courtesy: Matlab Toolbox)

(d) Cup (Images courtesy: Ron Rubinstein)

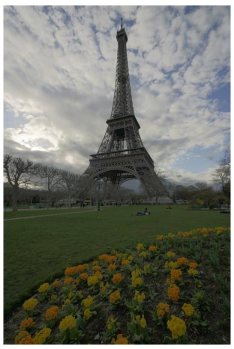
Fig. 2: Test image sequences



(a) Result of Mertens et al.



(b) Result of Kotwal et al.



(c) Result of our algorithm

Fig. 3: Results for 'Eiffel tower' image sequence

appearance with excellent color and contrast, but at the same time the details in extreme dark and bright are not captured completely. The result of Kotwal et al.'s algorithm achieves good details specifically, but the color reproduction is poor compared to the results of Mertens et al. In our algorithm, we find a balance between detail and color reproduction. The color reproduction is better than the algorithm of Kotwal et al. and is close to the results of Mertens et al. Also, the details are captured more clearly in dark and bright regions. For example, observe the details of Eiffel tower and the bright cloud for Figure 3, the details of the boat at the center and other dark areas in Figure 4, and the details of the chair in Figure 5.

It is often possible that there is a movement of either the camera or some objects in the scene, while capturing a multi-exposure image sequence. Such images will not be spatially aligned and hence will produce ghosting artifacts in the fused image. The subsequent results show that our algorithm gives negligible ghosting artifacts when images are not spatially aligned.

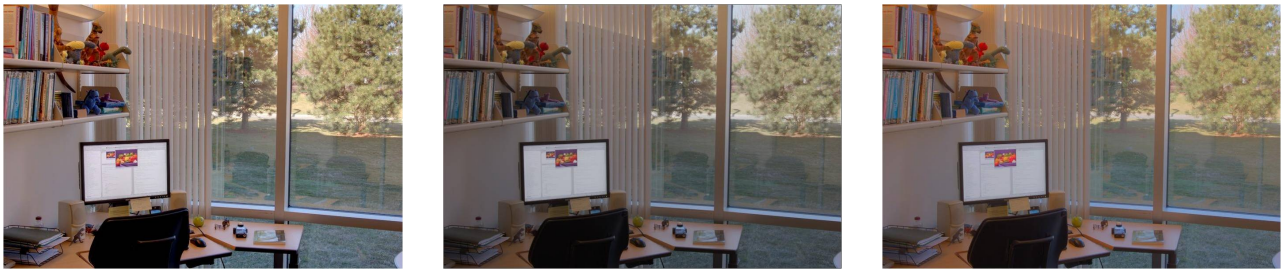
Figure 6 and 7 shows results for 'Peyrou' and 'Cup' multi-exposure image sequences respectively. Ghosting artifacts can be seen around the central structure for results of both Mertens et al. and Kotwal et al., whereas in our results, no such artifacts are seen in Figure 6. Similarly, there are almost negligible ghosting artifacts in our result as opposed to results of Mertens et al. and Kotwal et al. in Figure 7. Also, the color and details are preserved better in our algorithm than that in the algorithm of Mertens et al. and Kotwal et al.

The quantitative analysis in presented in Table I for three quality factors viz., Saturation, RMS Contrast and Entropy. Saturation represents the color richness of the image. For this, we convert the resultant image to HSI format. The mean value of the S component represents the overall saturation of the image. Contrast represents the dynamic range of an image. The RMS contrast is calculated using the I component as :

$$Contrast_{RMS} = \sqrt{\frac{1}{MN}\sum_{x=0}^{N-1}\sum_{y=0}^{M-1}(I(x,y) - I_{mean})^2} \quad (5)$$

(a) Result of Mertens et al. (b) Result of Kotwal et al. (c) Result of our algorithm

Fig. 4: Results for 'Grandcanal' image sequence



(a) Result of Mertens et al. (b) Result of Kotwal et al. (c) Result of our algorithm

Fig. 5: Results for 'Office' image sequence

TABLE I: Quantitative analysis of fusion algorithms

Test Sequence	RMS Contrast			Saturation			Entropy		
	EF	OPT	GF	EF	OPT	GF	EF	OPT	GF
Eiffel Tower	0.2398	**0.2922**	0.2450	**0.2835**	0.2377	0.2392	**7.6137**	7.3043	7.3387
Grandcanal	0.2556	**0.2695**	0.2395	**0.2296**	0.1315	0.1632	**7.7456**	7.6530	7.3575
Office	**0.2000**	0.1701	0.1553	**0.2548**	0.1972	0.2442	**7.3964**	7.3141	7.1884
Peyrou	**0.2196**	0.2079	0.1802	0.6646	0.6555	**0.6738**	**7.6388**	7.5500	7.3719
Cup	0.2035	**0.2198**	0.1904	0.4694	0.4677	**0.4821**	7.2284	**7.2642**	7.0733

Note - EF: Result of Mertens et al., OPT: Result of Kotwal et al., GF: Result of our algorithm

TABLE II: Execution time of fusion algorithms

Image sequence	Number of images	Image size	Execution Time (sec)		
			Mertens et al.	Kotwal et al.	Our algorithm
Eiffel Tower	3	530 x 795	1.303491	769.175253	0.211894
Grandcanal	3	1200 x 800	2.219605	1818.791429	0.428439
Office	6	903 x 600	2.502115	996.283538	0.509703
Peyrou	3	1200 x 800	2.289247	1749.222667	0.434574
Cup	5	672 x 504	1.55583	616.894309	0.277677

where, $I(x, y)$ denotes the I component of an $M \times N$ image at location (x, y) and I_{mean} is the average value of the I component. The information content in the fused image can be measured using entropy. The entropy can be calculated based on I component as:

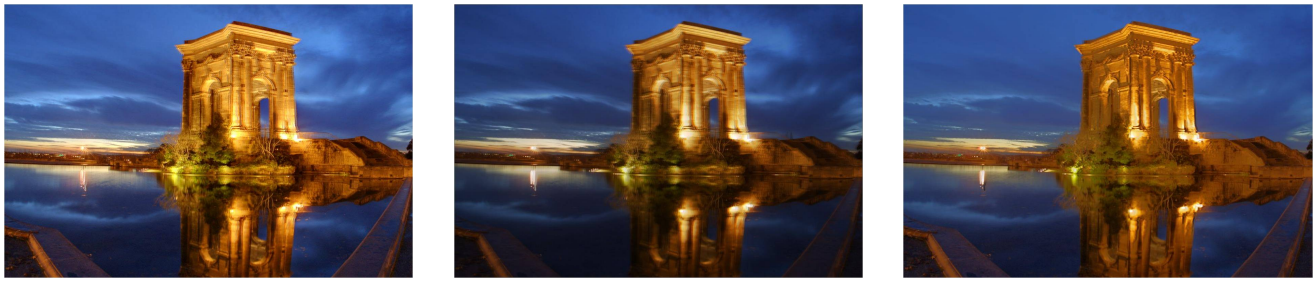
$$H = -\sum p(x_i) \log p(x_i) \tag{6}$$

where, $p(x_i)$ is the probability of occurrence of $x_i{}^{\text{th}}$ gray level.

From the Table I, we can also note that the results of Mertens et al. are best in terms of entropy and color reproduction, whereas, results of Kotwal et al. give an advantage in RMS contrast. Our algorithm tries to balance both these factors. Our algorithm certainly performs better in color quality, compared to results of Kotwal et al. In two cases it even performs better than results of Mertens et al. For RMS contrast and entropy, our algorithm gives a closer approximation of the best results. For the sequence of 'Peyrou' and 'Cup' judging based on the contrast and entropy value may not be absolutely correct because of the ghosting artifacts. The subjective comparison indicates that, our algorithm has a better contrast and correct information content compared to the other two results.

To check the real time implementation, we recorded execution time of these algorithms in Matlab. These are presented in Table II. It is evident that algorithm of Kotwal et al. runs in minutes and hence is not suitable for real time implementation. The algorithm of Metrens et al. is fast and takes few seconds, but our algorithm is almost 5 times faster than Metrens et al. and takes less than half second for almost all test cases, making it most suitable for real time applications. Also, our algorithm does not involve any filtering or transformation, making it very low in computational complexity, which is required for embedded platforms in mobile applications.

The simplicity of the algorithm comes with two limitations. Firstly, the image sequence should spread evenly over the exposure range, without any bias towards under or overexposedness. When the image set is biased to under or



(a) Result of Mertens et al. (b) Result of Kotwal et al. (c) Result of our algorithm

Fig. 6: Results for 'Peyrou' image sequence

(a) Result of Mertens et al. (b) Result of Kotwal et al. (c) Result of our algorithm

Fig. 7: Results for 'Cup' image sequence

overexposedness, the results will be either washed out or dark. Secondly, the number of images to be fused should be limited to 6 or 7. This is because, as the number of images increases, the weight map functions will be placed very closely, overlapping the same range of gray values. This results in degradation of the fusion quality and fused image looks washed out.

IV. Conclusions

This paper presented an algorithm for fusion of multi-exposure images, which has very low computational complexity. The proposed algorithm is driven by a pixel-by-pixel fusion, governed by a simple weight map, calculated on the basis of exposure of the image. The proposed algorithm is fully automatic and does not require any knowledge of camera parameters or any sort of human intervention. The experimental results show that our algorithm produces good color while preserving details even in very dark or very bright regions. When the images are not spatially aligned, our algorithm gives negligible ghosting artifacts. The simplicity of the algorithm makes it suitable for real time implementation even on embedded platforms with low computational power.

References

[1] R. Szeliski. *Computer Vision: Algorithms and Applications*. Texts in Computer Science. Springer, 2010.

[2] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 classes*, SIGGRAPH '08, pages 31:1–31:10, New York, NY, USA, 2008. ACM.

[3] S. Mann and R. W. Picard. Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. Technical Report 323, M. I. T. Media Lab Perceptual Computing Section, Bosten, Massachusetts, 1994. Also appears, IS&T's 48th annual conferance, Cambridge, Massachusetts, May 1995.

[4] T. Mitsunaga and S. Nayar. Radiometric Self Calibration. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 374–380, Jun 1999.

[5] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. *ACM Trans. Graph.*, 21(3):249–256, July 2002.

[6] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda. Photographic tone reproduction for digital images. *ACM Trans. Graph.*, 21(3):267–276, July 2002.

[7] A. A. Goshtasby. Fusion of multi-exposure images. *Image Vision Comput.*, 23(6):611–618, June 2005.

[8] T. Mertens, J. Kautz, and F. V. Reeth. Exposure fusion. *Computer Graphics and Applications, Pacific Conference on*, 0:382–390, 2007.

[9] M. H. Malik, S. A. M. Gilani, and A. ul Haq. Wavelet based exposure fusion. In *Proceedings of the World Congress on Engineering 2008 Vol I, WCE '08, July 2 - 4, 2008, London, U.K.*, pages 688–693. International Association of Engineers, 2008.

[10] Y. Li, L. Sharan, and E. H. Adelson. Compressing and companding high dynamic range images with subband architectures. *ACM Trans. Graph*, pages 836–844, 2005.

[11] S. Raman and S. Chaudhuri. Bilateral filter based compositing for variable exposure photography. In *Short Papers*, pages 1–4, Munich, Germany, 2009. Eurographics.

[12] K. Kotwal and S. Chaudhuri. An optimization-based approach to fusion of multi-exposure, low dynamic range images. In *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, pages 1–7, July 2011.

[13] T. Kartalov, A. Petrov, Z. Ivanovski, and L. Panovski. A real time algorithm for exposure fusion of digital images. In *MELECON 2010 - 2010 15th IEEE Mediterranean Electrotechnical Conference*, pages 641 –646, April 2010.