

Literature Survey for Fusion of Multi-Exposure Images

R. Ganga

*PG Scholar, Department of ECE, PET
Engineering College, Tirunelveli, India.*

T. Agnes Ramena

*Assistant Professor, Department of ECE,
PET Engineering College, Tirunelveli,
India*

Abstract

A single digital photo is insufficient to clearly record all the details in the scene because, some areas in the photo may appear so bright that details are washed out, i.e., over-exposed and other portions may appear so dark that details can hardly be seen, i.e., under-exposed. These limitations motivate the development of fusion techniques for multi-exposure images such as generalized random walks approach. But the existing fusion methods may cause unnatural appearance in the fusion results. This literature survey discusses all the existing image fusion techniques and their performance.

Keywords - Image fusion, local contrast, multi-exposure fusion, random walks.

1. Introduction

A natural scene usually contains a wide range of intensity levels that is beyond what a common digital camera is able to capture and also beyond the display capability of a common digital screen. This contradiction between the high dynamic range (HDR) nature of a real-world scene and the low dynamic range (LDR) limitation of current capture and display devices motivates the development of fusion techniques for multi-exposure images. The HDR images usually have higher fidelity than LDR images, which benefits many applications, such as physically-based rendering and remote sensing [1]. Although cameras with spatially varying pixel exposures [6], cameras that automatically adjust exposure for different parts of a scene [15], [16], and displays that directly display HDR images [17] have been developed by previous researchers, their technologies are only at a prototyping stage and unavailable to ordinary users. In ordinary displays,

an HDR image is compressed into an LDR image using tone mapping methods [2], [3]. This two phase workflow, HDR-R+TM has several advantages: no specialized hardware is required; various operations can be performed on the High Dynamic Range images, such as virtual exposure; and user interactions are allowed in the TM phase to generate a tone-mapped image with desired appearance.

However, this workflow is usually not as efficient as image fusion (IF) [4], [5], which directly combine the captured multi-exposure images into a single LDR image without involving HDR-Reconstruction. Another advantage of IF is that IF does not need the calibration of the camera response function (CRF), which is required in HDR-R if the CRF is not linear. IF is preferred for quickly generating a well-exposed image from an input set of multi-exposure images, especially when the number of input images is small and speed is crucial.

Previous multi-exposure fusion methods [1], [4] usually define the fusion weights locally without adequate consideration of consistency in a large neighborhood, which may cause unnatural appearance in the fusion results. Some methods partition the input images into different regions, either using uniform blocks or by segmentation techniques, and then try to maximize a certain quality measure within each region. These methods tend to cause artifacts at object/region boundaries, because inter-region information is not effectively exploited. Multi-resolution fusion methods normally work better at region boundaries and are good at enhancing main image features by blending fusion weights at different scales. However, the weights are still mainly determined locally without considering large neighborhood information. This may cause some inconsistencies in the results. IF has been employed in various applications such as multi-sensor fusion [9], [10], multi-focus fusion [8], [11] and multi-exposure fusion [5], [12]. Some

general fusion approaches [13], [14] proposed earlier are not optimized for individual applications and have only been applied to gray-level images.

2. Literature review

2.1. High Dynamic Range Imaging

S.K.Nayar and T.Mitsunaga [6] proposed a very simple method for significantly enhancing the dynamic range of virtually any imaging system. The basic principle is to simultaneously sample the spatial and exposure dimensions of image irradiance. One of the several ways to achieve this is by placing an optimal mask adjacent to a conventional image detector array. The mask has a pattern with spatially varying transmittance, thereby giving adjacent pixels on the detector different exposures to the scene. The captured image is mapped to a high dynamic range image using an efficient image reconstruction algorithm. The end result is an imaging system that can measure a very wide range of scene irradiances and produce a substantially larger number of brightness levels, with a slight reduction in spatial resolution.

This technique deals with spatially varying pixel sensitivities for high dynamic range imaging. In an array of pixels the brightness level associated with each pixel represent its sensitivity, such that, the brighter pixels have greater exposure to image irradiance and the darker ones have lower exposure. When a pixel is saturated in the acquired image, it is likely to have a neighbor that produces non-zero brightness. The availability of extra bits of data at each image pixel is expected to enhance the robustness of vision algorithms

2.2. Multi-exposure image fusion

A.Goshtasby [4] used a method for fusing multi-exposure images of a static scene taken by a stationary camera into an image with maximum information content. The method partitions the image domain into uniform blocks and for each block selects the image that contains the most information within that block. The selected images are then blended together using monotonically decreasing blending functions that are centered at the blocks and have a sum of 1 everywhere in the image domain. The optimal block size and width of the blending functions are determined using a gradient-ascent algorithm to maximize information content in the fused image.

The main problem to be solved here is to identify the image that contains the most information within a particular local area. An

image that is over or under exposed within an area does not carry as much information as an image that is well-exposed in that area. Image analysis techniques rely on critical image information, which may not be available in image areas that are over or under-exposed. In situations where images at multiple exposure levels of a scene are taken, image fusion is used to combine the images into an image that is well-exposed everywhere and provides the critical information needed in a particular vision task.

The fusion method preserves scene highlights if color information within a highlight area is quite high. A characteristic of the method is that it does not have a side effect and will not change the local color and contrast in the best-exposed image. For further contrast enhancement, traditional methods such as inverse filtering are used. An improvement to this method is to use entropy as a measure for optimization when fusing the images.

2.3. Exposure fusion

T. Mertens, J. Kautz, and F.Van Reeth [5] proposed a technique for fusing a bracketed exposure sequence into a high quality image, without converting to HDR. Skipping the physically-based HDR assembly simplifies the acquisition pipeline. This avoids camera response curve calibration and is computationally efficient. It also allows including flash images in the sequence. This technique blends multiple exposures, guided by simple quality measures like saturation and contrast. This is done in a multi-resolution fashion to account for the brightness variation in the sequence.

Exposure fusion computes the desired image sequence. This process is guided by a set of quality measures, which is consolidated into a scalar-valued weight map. It is useful to think of the input sequence as a stack of images. The final image is then obtained by collapsing the stack using weighted blending. Quality is comparable to existing tone mapping operators. Compared with several tone mapping techniques, this algorithm exhibits high contrast and good color reproduction. However, it cannot extend the dynamic range of the original pictures.

Table 1 Computational time for various size images

w × h × N	init. (s)	update (s)	total (s)
864 × 576 × 3	.75	.82	1.6
1227 × 818 × 3	1.5	1.6	3.2
1728 × 1152 × 3	3.0	3.2	6.2
864 × 576 × 7	1.5	1.5	3.0
1227 × 818 × 7	3.0	3.1	6.1
1728 × 1152 × 7	6.0	6.0	12.0

From the Table 1, it is seen that when the size of the image increases, the computational time increases. Here, N represents the number of pixels, 'w' represents the width of the image and 'h' represents the height of the image

2.4. Fusing images using support vector machines

Shutao Li, James Tin-Yau Kwok, Ivor Wai-Hung Tsang, and Yaonan Wang proposed a method to improve the fusion procedure by applying the discrete wavelet frame transform (DWFT) and the support vector machines (SVM). Unlike DWT, DWFT yields a translation-invariant signal representation. Using features extracted from the DWFT coefficients, a SVM is trained to select the source image that has the best focus at each pixel location, and the corresponding DWFT coefficients are then incorporated into the composite wavelet representation.

The basic idea is to perform multi-resolution decomposition on each source image, and then integrate all these decompositions to obtain one composite representation, from which the fused image can be recovered by performing the corresponding inverse transform. However, many of these multi-resolution decompositions are not translation-invariant because of an underlying down-sampling process. Hence, in practice, their performance quickly deteriorates when there is slight object movement or when the source images cannot be perfectly registered. One way to alleviate this problem is by using the discrete wavelet frame transform.

2.5. Random walks for multi-exposure image fusion

Rui Shen, Irene Cheng, Jianbo Shi and Anup Basu introduced a fresh view of the multi-exposure image fusion problem. A probabilistic method to achieve an optimal balance between two quality measures, i.e., local contrast and color consistency is considered. The probabilities that a pixel in the fused image comes from different input images are estimated based on these two measures and then used as fusion weights in the final composition step.

The pixel in the fused image is represented as,

$$p_i = \sum_{k=1}^K P^k(x_i) p_i^k \quad (1)$$

The local contrast measure is applied in order to select those input pixels containing more details. Because high local contrast is usually associated with high local variation, for a given input pixel, the local variation around it is computed and then modified using a sigmoid-shaped function to

indicate local contrast. The color consistency measure imposes not only consistency in a large neighborhood but also consistency with the natural scene. This measure is based on the assumptions that adjacent pixels having similar colors in most input images will indicate similar colors in the fused image and that similar colors at the same pixel location in different input images under proper exposures will indicate the true color of the scene. Therefore, for two given adjacent pixels, their similarity is evaluated based on their color differences. These two locally defined quality measures are integrated in generalized random walks framework as compatibility functions to obtain optimal fusion weights.

The fused image is organized as an undirected graph, where each scene node corresponds to a pixel in the fused image. Within this graph, there is also a set of label nodes, each of which indicates an input image. The proportions of contributions from each input pixel to their corresponding pixels in the fused image are considered as the probabilities of each scene node being assigned different labels. This probability estimation is further formulated as transition probability calculation in GRW. This can be efficiently computed by solving a set of linear systems.

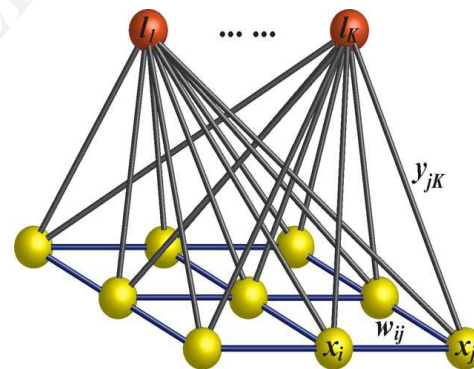


Fig. 1. Graph used in GRW

Let γ be arranged in a way that the first K nodes are label nodes, i.e., $\{v_1, \dots, v_k\} = L$, and the rest N nodes are scene nodes, i.e., $\{v_{K+1}, \dots, v_{K+N}\} = X$. With two positive coefficients γ_1 and γ_2 introduced to balance the weights between $y(\dots)$ and $w(\dots)$, define a node compatibility function $c(\dots)$ on ϵ with the following form:

$$c_{ij} \triangleq c(v_i, v_j) = \begin{cases} \gamma_1 y_{i-K, j}, & (v_i, v_j) \in \epsilon_L \quad v_j \in L \\ \gamma_2 w_{i-K, j-K}, & (v_i, v_j) \in \epsilon_X \end{cases} \quad (2)$$

Because the graph is undirected, consider $c_{ij} = c_{ji}$. Let $u(v_i)$ denote the potential associated with v_i . Based on the relationship between RW and electrical networks, the total energy of the system given in Fig 2 is

$$E = \frac{1}{2} \sum_{(v_i, v_j) \in \varepsilon} c_{ij} (u(v_i)u(v_j))^2 \quad (3)$$

The harmonic function $u(\cdot)$ can be computed efficiently using matrix operations. A Laplacian matrix can be constructed following (3); L here contains both label nodes and the scene nodes and becomes a $(K+N) \times (K+N)$ matrix

$$L_{ij} = \begin{cases} d_i, & i = j \\ -c_{ij}, & (v_i, v_j) \in \varepsilon \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where, $d_i = \sum_{v_j \in N_i} c_{ij}$ is the degree of the node v_i on its immediate neighborhood N_i . Then (2) can be rewritten in matrix form as,

$$E = \begin{pmatrix} u_L \\ u_X \end{pmatrix}^T L \begin{pmatrix} u_L \\ u_X \end{pmatrix} = \begin{pmatrix} u_L \\ u_X \end{pmatrix}^T \begin{pmatrix} L_L & B \\ B^T & L_X \end{pmatrix} \begin{pmatrix} u_L \\ u_X \end{pmatrix} \quad (5)$$

The minimum energy solution can be obtained by setting $\Delta E = 0$ with respect to u_X , i.e., solving the following equation:

$$L_X u_X = -B^T u_L \quad (6)$$

In some cases, part of X may be already labeled. These pre-labeled nodes can also be represented naturally in the current framework without altering the structure of the graph. Suppose x_i is one of the pre-labeled nodes and is assigned label l_k . Then, simply a sufficiently large value is assigned to y_{ik} and solve (6) for the unlabeled scene nodes.

The compatibility functions $y(\cdot, \cdot)$ and $w(\cdot, \cdot)$ are defined to represent respectively the two quality measures used in the fusion algorithm, i.e., local contrast and color consistency. The local contrast measure should be biased towards pixels from the images that provide more local variations in luminance. Let g_i^k denote the second order partial derivative computed in the luminance channel at the i th pixel in the image I^k , which is an indicator of local contrast. The higher the magnitude of g_i^k is, the more variations occur near the pixel p_i^k , which may indicate more local contrast. If the frequency, i.e., number of occurrences of a value $|g_i^k|$ in I^k is very low, the associated pixel may be noise. Hence, taking into account both the magnitude and the

frequency of the contrast indicator g_i^k , the compatibility between a pixel and a label is computed as,

$$y_{ik} = \theta_{ik} \left[\text{erf} \left(\frac{|g_i^k|}{\sigma_y} \right) \right]^K \quad (7)$$

where θ_{ik} represents the frequency of the value $|g_i^k|$ in I^k ; $\text{erf}(\cdot)$ is the Gaussian error function, which is monotonically increasing and sigmoid shaped; the exponent K is equal to the number of input images and controls the shape of $\text{erf}(\cdot)$ by giving less emphasis on the difference in high contrast regions as the number of input images increases; and σ_y is a weighting coefficient which is taken as the variance of all g_i^k 's.

Table 2 Computational time comparison table for Multi-exposure image fusion and Exposure fusion

Size	initializ e	Compute compatib ilities	Opti mize	Fuse	Total	EF
226x 341x 5	0.06	0.03	0.02	0.05	0.17	0.68
343x 231x 5	0.06	0.04	0.02	0.05	0.17	0.68
348x 222x 6	0.07	0.05	0.02	0.06	0.20	0.80
236x 341x 6	0.07	0.13	0.03	0.05	0.20	0.85
752x 500x 4	0.24	0.46	0.14	0.21	0.72	2.64
512x 768x 16	0.75	0.40	0.20	0.78	2.18	9.82
1500 x644 x5	0.78	0.51	0.57	0.76	2.51	9.73

From the table 2 it is found that the computational time for multi-exposure image fusion technique is less when compared to exposure fusion.

The following equation is used to evaluate the similarity/compatibility between adjacent pixels in the input image set using all three channels in the RGB color space:

$$w_{ij} = \prod_k \exp \left(- \frac{\|p_i^k - p_j^k\|}{\sigma_w} \right) \approx \exp \left(- \frac{\|\bar{p}_i - \bar{p}_j\|}{\bar{\sigma}_w} \right) \quad (8)$$

where, p_i^k and p_j^k are adjacent pixels in image I^k , $\exp(\cdot)$ is the exponential function; $\|\cdot\|$ denotes Euclidean distance, $\bar{p}_i = (1/K) \sum_k p_i^k$ denotes the average pixel, and σ_w and $\bar{\sigma}_w = (\sigma_w/K)$ are free parameters. Although the two quality measures are

defined locally, a global optimization using GRW is carried out to produce a fused image that maximizes contrast and details, as well as imposing color consistency. Once $y(\dots)$ and $w(\dots)$ are defined using (7) and (8), the probabilities $P^k(x_i)$'s are calculated using (2)- (6).

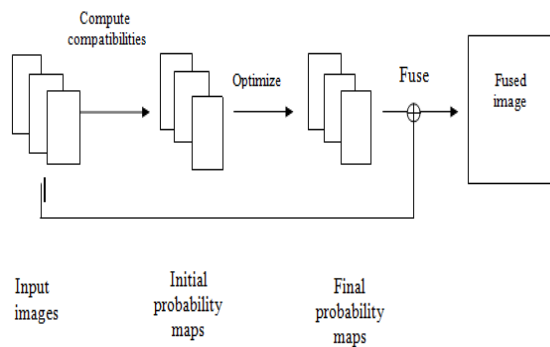


Fig.2.Processing procedure of fusion algorithm

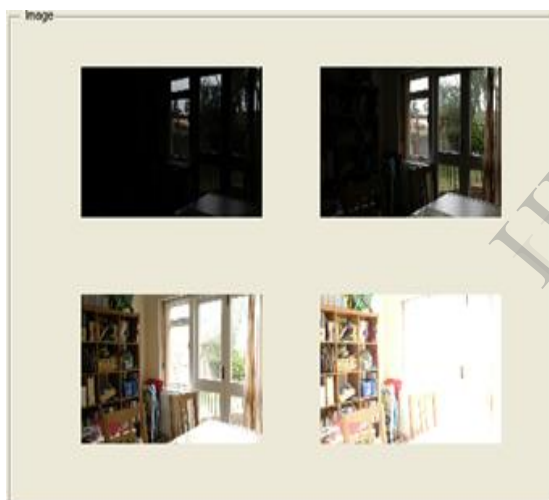


Fig. 3. Multi-exposure images

This figure illustrates an image with four exposure settings. From each image only the best pixels will be selected which will be then fused to obtain the quality image.

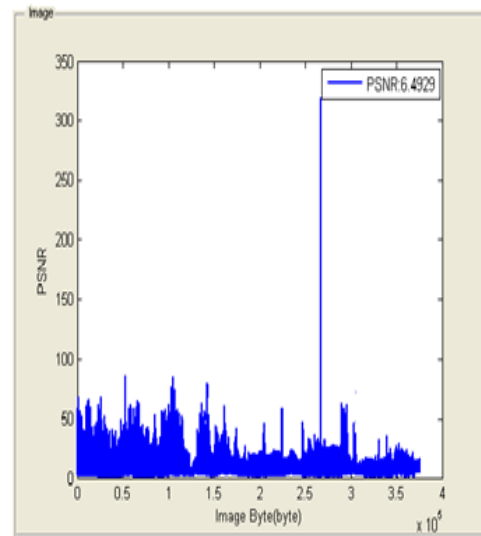


Fig. 4. PSNR performance of fused image

The graph shows the PSNR performance of the fused image. Here higher PSNR is obtained for higher quality pixels.

3. Conclusion

In this paper, a brief literature survey for multi-exposure image fusion methods are discussed elaborately. Some of these methods suffer unnatural appearance in the fusion result. All these limitations are overcome by using multi-exposure image fusion based on generalized random walks approach. Here four different exposures for an image are considered in order to obtain the high quality fused image and the PSNR performance illustrates the image quality.

Acknowledgement

I have taken efforts in this paper. However, it would not have been possible without the kind support and help of many individuals. I would like to extend my sincere thanks to all of them. I am highly indebted to my guide for her guidance and constant supervision as well as for providing necessary information regarding the project and also for her support in completing the paper. I owe a sincere prayer to the LORD ALMIGHTY for his kind blessings and giving me full support to do this work, without which it would have not been

possible. My thanks and appreciations also go to my colleague in developing the paper and people who have willingly helped me out with their abilities.

References

- [1] E. Reinhard, G. Ward, S. Pattanaik, P. Debevec, W. Heidrich and K. Myszkowski, (2010), "High Dynamic Range Imaging: Acquisition, Display, and Image based Lighting", Waltham, MA: Morgan Kaufmann, 2nd ed.
- [2] G. Krawczyk, K. Myszkowski, and H.P. Seidel (2005) 'Lightness perception in tone reproduction for high dynamic range images', *Comput. Graph. Forum*, vol.24, no.3, pp.635-645.
- [3] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, (2002), "Photographic tone reproduction for digital images", in *Proc. ACM SIGGRAPH*, pp. 267-276.
- [4] A. Goshtasby (2005) 'Fusion of multi-exposure images', *Image Vision Comput.*, vol.23, no.6, pp. 611-618.
- [5] T. Mertens, J. Kautz and F. Van Reeth (2007) 'Exposure fusion', in *Proc. Pacific Graphics*, pp. 382-390.
- [6] S.K. Nayar and T. Mitsunaga (2000) 'High dynamic range imaging: Spatially varying pixel exposures', in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol.1, pp.472-479.
- [7] L. Grady (2006) 'Random walks for image segmentation', *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp.1768-1783.
- [8] Li S., Kwok J.T.Y., Tsang W.H., and Wang Y. (2004) 'Fusing images with different focuses using support vector machines', *IEEE Trans. Neural Netw.*, vol. 15,no.6, pp.1555-1561.
- [9] M. Kumar and S. Dass (2009) 'A total variation based algorithm for pixel level image fusion', *IEEE Trans. Image Process.*, vol. 18, no.9,pp. 2137-2143.
- [10] S. Zheng, W.Z. Shi, J. Liu, G. X. Zhu G, and J.W. Tian (2007) 'Multisource image fusion method using support value transform', *IEEE Trans.. Image Process.*, vol.16, no.7, pp. 1831-1839.
- [11] H. Zhao, Q. Li and H. Feng (2008) 'Multi-focus color image fusion in the HSI space using the sum modified laplacian and a coarse edge map', *Image Vis. Comput.*, vol.26, no.9, pp.1285-1295.
- [12] L. Bogoni and M. Hansen (2001) 'Pattern-selective color image fusion', *Pattern Recognit.*, vol. 34, no. 8,pp. 1515-1526.
- [13] G. Piella, (2009), "Image fusion for enhanced visualization:A variational approach", vol. 83, no. 1, pp. 1-11.
- [14] V.S. Petrovic and C.S. Xydeas (2004) 'Gradient based multiresolution image fusion', *IEEE Trans. Image Process.*, vol.13, no.2, pp.228-237.
- [15] H. Mannami, R. Sagawa, Y. Mukaigawa, T. Echigo and Y. Yagi (2007) ' High dynamic range camera using reflective liquid crystal', in *Proc. Int. Conf. Comput. Vis.*, pp. 1-8.
- [16] J. Tumblin, A. Agrawal and R. Raskar, (2005), "Why I want a gradient camera", in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, pp. 103-110.
- [17] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Gosh, and A. Vorozcovs, (2004) "High dynamic range display systems", in *Proc. ACM SIGGRAPH*, pp. 760-768.
- [18] P.E. Debevec and J. Malik (1997) 'Recovering high dynamic range radiance maps from photographs', in *Proc. ACM SIGGRAPH*, pp.369-378.