Gradient-directed Composition of Multi-exposure Images

Wei Zhang Wai-Kuen Cham
Department of Electronic Engineering
The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
{zhangwei,wkcham}@ee.cuhk.edu.hk

Abstract

In this paper, we present a simple yet effective method that takes advantage of the gradient information to accomplish the multi-exposure image composition in both static and dynamic scenes. Given multiple images with different exposures, the proposed approach is capable of producing a pleasant tonemapped-like high dynamic range (HDR) image by compositing them seamlessly with the guidance of gradient-based quality assessment. Especially, two novel quality measures: visibility and consistency, are developed based on the observations of gradient changes among different exposures. Experiments in various static and dynamic scenes are conducted to demonstrate the effectiveness of the proposed method.

1. Introduction

Radiance of the real world spans several orders of magnitude and its dynamic range dramatically exceeds the capability of the current electronic imaging devices. However, since each exposure can be designed to capture a certain dynamic range, HDR imaging is possible by first generating a HDR image from a stack of differently exposed images and then making it displayable by tone mapping. If the stack is captured in a static scene, the task is to recover the full dynamic range and make all present details visible in one image. In a dynamic scene that has moving objects, HDR imaging is more challenging and prone to ghosting artifacts.

In this paper, we present a novel image composition approach that is able to bypass the typical HDR process and directly yield a tonemapped-like HDR image where all parts appear well-exposed by compositing multi-exposure images with the guidance of image quality assessment. Unlike previous work (e.g. [9, 10]), the multi-exposure image composition is addressed from the perspective of gradient, and a new quality assessment system is developed to handle the composition in both static and dynamic scenes.

Specifically, the underlying idea of this work comes from the observations of gradient changes among differently exposed images. Firstly, gradient magnitude can imply pixel's exposure quality and will decrease gradually as the image is approaching over- or under-exposure. Consequently, it can be utilized as a measure on visibility to help preserve the details present in the exposure sequence. Secondly, it is also found that the gradient direction changes reveal object movement and thus can help account for the ghosting problem in dynamic scenes. More detailed, if the content in some area changes among different exposures due to object movement, the gradient direction in that area will probably have significant changes as well. Consequently, exploiting the gradient direction changes leads to a consistency measure which can get rid of the influence of moving objects and preserve the desired consistent pixels in the composite image. By combining the consistency measure and visibility measure, the proposed method is still capable of compositing all exposures gracefully in dynamic scenes and producing a pleasant well-exposed image free of ghosting artifacts.

In summary, the proposed algorithm is designed to have the following properties: first, it is easy to use and has lower computational complexity since neither radiometric camera calibration nor tone mapping is required. Second, for dynamic scenes, the proposed approach can eliminate the ghosting artifacts automatically and efficiently without resorting to any explicit complex motion detection techniques (e.g. optical flow). Third, it allows for lighting changes and can be extended naturally to other tasks such as flash and no-flash photography.

2. Related Work

To extend the dynamic range of conventional camera, some new HDR camera prototypes [1, 2] have been developed in the past years. However, they are still unavailable to consumers currently due to the expensive price and high requirements on hardware. In this paper, we are more concerned with the promising software techniques that achieve HDR imaging from a stack of images taken by a conventional camera using different exposure times. Generally speaking, the existing work can be classified into two types.

978-1-4244-6983-3/10/$26.00 ©2010 IEEE

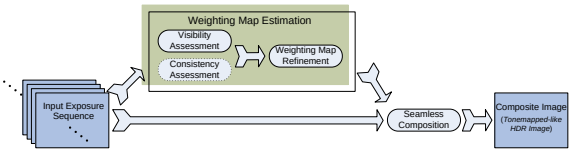


Figure 1. The proposed framework. Please note that consistency assessment is unnecessary for static scenes.

Static HDR: The standard HDR process prevalent in the current graphics software consists of two steps. First, estimate the camera response function (CRF) [3, 4] and recover the latent radiance map (HDR image). Second, apply tone mapping to make the radiance map displayable on the commonly used low dynamic range (LDR) monitors [5, 6, 7, 8]. As an alternative, the other kind of work [9, 10] attempts to produce the desired tonemapped-like HDR image directly by exposure composition in the image domain. These methods skip the typical HDR process, and no intermediate HDR image needs to be generated. Therefore, they are more efficient and do not require tone mapping. However, the major limitation of the above methods is that they require the target scene is completely still. Otherwise, the resulting image will suffer from annoying ghosting artifacts in the area where motion occurs.

Dynamic HDR: Recently, lots of efforts have been made to tackle the more challenging dynamic HDR task. The existing methods were proposed in a similar manner. They first detect the motion regions, and then produce a ghost-free HDR result by removing the contributions of these regions in the composite radiance map. For example, many different kinds of techniques such as optical flow [11], variance measurement [12], error map detection [13], entropy calculation [14] and pixel's order relation detection [15], have been adopted to find the regions where ghosting artifacts may occur due to object movement. Besides, Gallo et al. [16] and Eden et al. [17] proposed to composite the desirable radiance with the guidance of a reference image preselected automatically or manually. Some statistical tools such as kernel density estimator were employed in [18, 19] to iteratively determine the probability that a pixel belongs to the background. However, all above dynamic HDR methods were presented in the radiance domain fully or partially. Hence, they share two limitations at least. First, the performance highly relies on the success of the radiometric calibration of camera which is sensitive to image noise, lighting change and misalignment error. Second, they normally have complex working pipelines and require tone mapping for HDR reproduction. The above problems make these kinds of methods computationally expensive and restrict their applications in practice.

In addition, image gradient has been manipulated in several tasks such as tone mapping [7], image editing [20] and enhancement [21]. It is worth pointing out that our method differs essentially from [21] which aimed at removing the artifacts existing in flash photography with a gradient projection scheme. Moreover, it was proposed for static scenes and thus cannot handle the dynamic scenes.

3. Algorithm

3.1. Motivation and Overview

Since different exposures capture different dynamic range characteristics of the latent scene, taking multiple exposures and combining them together as Eq (1) may create a more informative image that captures all details of the scene.

$$H(x,y) = \sum_{i=1}^{N} W^i(x,y) I^i(x,y), \tag{1}$$

where N represents the number of the input exposures. $I^i(x,y)$ and $W^i(x,y)$ denote the intensity and weight of the pixel located at (x,y) in the i_{th} exposure respectively. H denotes the composite image. Compared to the typical HDR process, exposure composition is easier and much more efficient since neither radiometric calibration nor tone mapping is necessary. However, the composition performance relies on the weight term W and so it is crucial to develop an effective quality assessment system that can output the desired weights. In this paper, we will show that the gradient information plays well in the quality assessment and makes it possible to handle the exposure composition in both static and dynamic scenes. Our algorithm shares the same spirit with the recent work [9] and [10]. However, these methods are only applicable in static scenes and suffer from severe ghosting artifacts in dynamic scenes.

As illustrated in Figure 1, the proposed HDR process is quite simple and begins with a stack of differently exposed images. In this work, we assume all exposures are captured with the aid of a tripod or have been aligned by some registration techniques like [22]. Then, the weighting map of each exposure is estimated by a gradient-based quality assessment system. For dynamic scenes, assessments on visibility and consistency are both required, while for static scenes only the former one is necessary. Besides, the cross-bilateral filtering [23] is employed for weighting map refinement, where the standard deviations of the space and range Gaussians are normally set to 5 in the experiments. Given weighting maps, a tonemapped-like HDR image is produced eventually by compositing all exposures seamlessly with a multiresolution spline scheme [24].

3.2. Gradient-based Image Quality Assessment

In this section, we will describe how to take advantage of the gradient information to generate weighting maps for

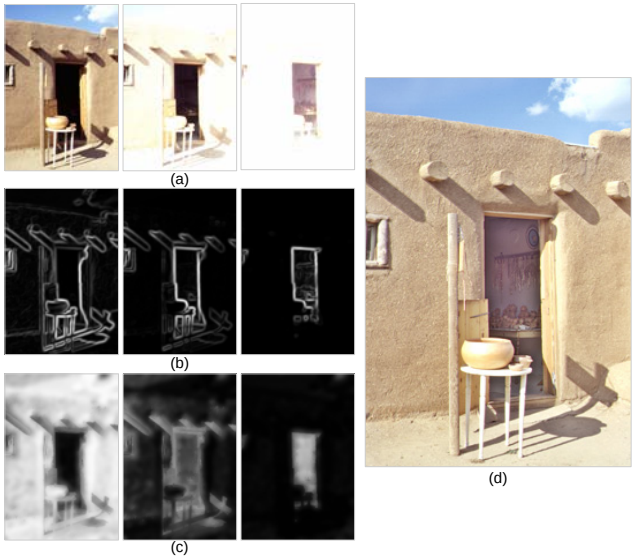


Figure 2. Static example with visibility assessment. (a) Input three exposures. (b) Gradient magnitude maps. Note that each has been normalized to $[0,\ 1]$ for display. (c) Weighting maps after refinement. (d) Composite image. Data courtesy of Shree K. Nayar.

static and dynamic scenes. For each image, the gradient information is extracted by using the first derivatives of a 2-D Gaussian kernel. $m^i(x,y)$ and $\theta^i(x,y)$ are used to denote the gradient magnitude and direction of a pixel $I^i(x,y)$.

3.2.1 Visibility Assessment

As shown in Figure 2(a), some features that are visible in one exposure disappear in the others due to over- or under-exposure. Therefore, the basic goal of composition is to preserve all features present in the exposure sequence and make them visible in one image. Gradient is associated with image features and its magnitude is an indicator of pixel's exposure quality. As illustrated in Figure 2(b), gradient magnitude becomes larger when a pixel gets better exposed. Also, it decreases gradually as the pixel approaches over- or under-exposure. Therefore, a visibility measure is developed as Eq (2) by exploiting the gradient magnitude information.

$$V^i(x,y) = \frac{m^i(x,y)}{\sum_{i=1}^{N} m^i(x,y) + \epsilon}, \tag{2}$$

where ϵ is a small value (e.g. 10^{-25}) to avoid singularity. In static scenes, the weights of Eq (1) can be obtained by setting $W^i = V^i$. As shown in Figure 2, exposure composition guided by gradient magnitude in a static scene can produce a plausible result in which all visible details are preserved.

3.2.2 Consistency Assessment

However, most scenes encountered in practice are non-static. It is hard to make all involved objects stay stationary while taking the exposure sequence. As shown in Figure 3(a), when we photography a public place, there often exist unwanted objects such as people moving around. In this case, visibility assessment only cannot avoid compositing the inconsistent content appeared in the motion area and yields an image ruined by ghosting artifacts as shown in Figure 3(c). Hence, it is necessary to seek an additional measure on consistency that can help remove the undesired moving objects and generate a ghost-free result.

Fortunately, we found gradient direction can serve in the consistency measure owing to its invariant property in different exposures as explained in Figure 3(b). In specific, it is observed that the gradient direction in the stationary regions remains stable in different exposures, provided that these regions are neither under-exposed nor over-exposed. The inherent reason of this fact is that image gradients are mainly due to the local changes in 3-D geometric shape and reflectance. If the content changes due to object movement, the gradient direction will vary accordingly (e.g. $|\varphi_1| > 0$ in Figure 3(b)). Therefore, we believe that the gradient direction information would be particularly effective to detect the inconsistency caused by motion. In this work, the measurement of gradient direction changes is accomplished in a window-based manner to make it more resistant to noise. Specifically, for each pixel located at (x,y) of the i_{th} image, its gradient direction change w.r.t that of the j_{th} image is calculated as follows.

$$d_{ij}(x,y) = \frac{\sum_{k=-l}^{l} \left|\theta^i(x+k,y+k) - \theta^j(x+k,y+k)\right|}{(2l+1)^2}, \tag{3}$$

where the size of window is $(2l+1) \times (2l+1)$ and l is normally set to 9. It is noted that $d_{ij}(x,y) = d_{ji}(x,y)$ and $d_{ij}(x,y) = 0$, when i and j are equal.

Besides, it is also observed that many exposure sequences such as Figure 3(a) have one thing in common: the moving object is only a shot for one position and appears in a relative smaller number images. This is because in most cases, the stationary parts of the scene that predominantly exist in the sequence are what the photographer is interested in. Consequently, a score S^i can be defined as Eq (4) by accumulating the gradient direction changes of each exposure to reflect its consistency in the whole sequence.

$$S^i(x,y) = \sum_{j=1}^{N} exp\left(\frac{-d_{ij}(x,y)^2}{2\sigma_s^2}\right), \tag{4}$$

where σ_s is the standard deviation and fixed at 0.2 in the experiments. Apparently, a large score implies small gradient direction change and thus the content is more frequently

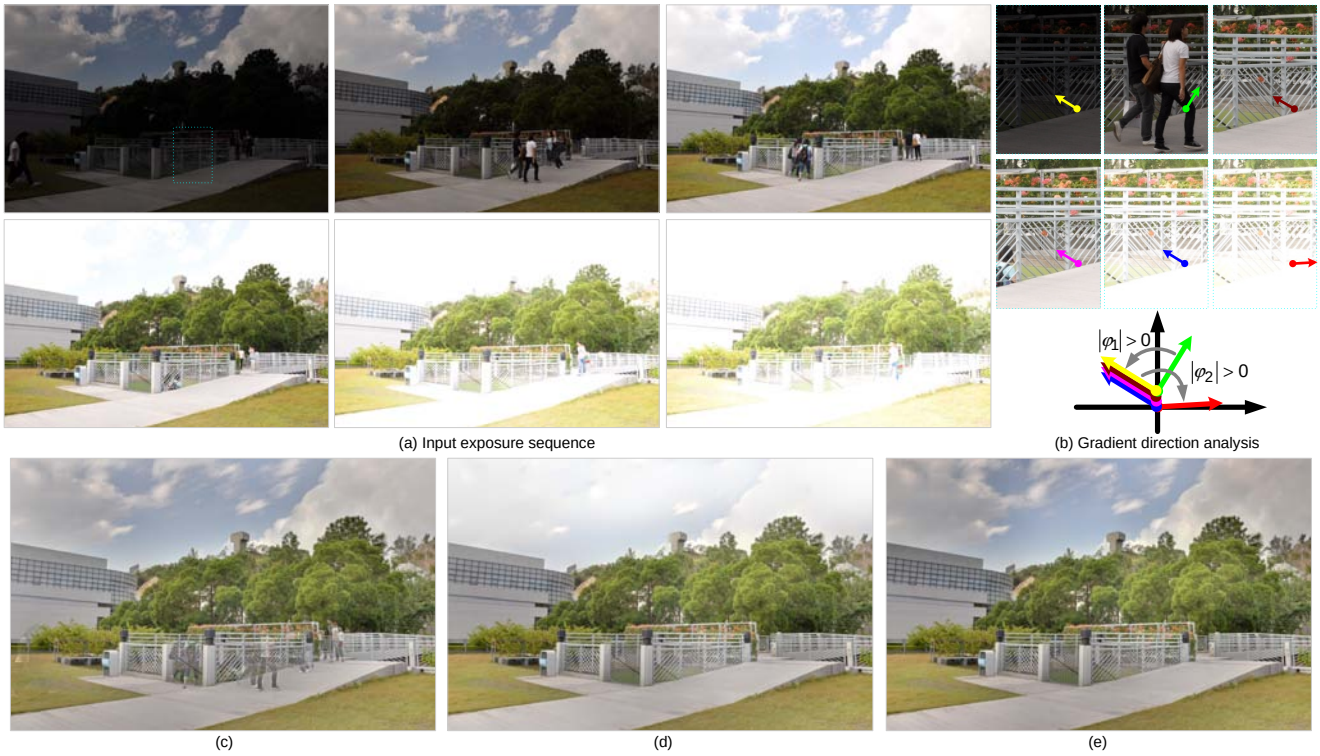


Figure 3. Dynamic example. (a) Input six exposures. (b) Analysis of the gradient direction changes among differently exposed images. Please note that the arrow here is only used to indicate the gradient direction illustratively and its length is unrelated to the magnitude. (c) Composite result with only visibility assessment. (d) Composite result without exposure correction. (e) Our final composite result.

captured in the sequence. Eq (4) can favor the stationary parts of the scene under the assumption that the exposure sequence predominantly captures the stationary parts of the latent scene, which is prevalent in the previous work [18, 15]. However, the direction changes of gradient may also be caused by over- or under-exposure (e.g. $|\varphi_2| > 0$ in Figure 3(b)). In this case, the score calculated based on $d_{ij}(x, y)$ is no longer desirable, since it may make the algorithm mistake the stationary visible objects for unwanted moving ones. Therefore, an additional term α^i which indicates the exposure quality of I^i, is introduced to jointly define the final consistency measure C^i with S^i as follows.

$$C^i(x,y) = \frac{S^i(x,y) \times \alpha^i(x,y)}{\sum_{i=1}^{N} S^i(x,y) \times \alpha^i(x,y) + \epsilon}, \tag{5}$$

where

$$\alpha^i(x,y) = \begin{cases} 1 & 1-\tau < I^i(x,y) < \tau \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Note that α^i is used to remove the invalid scores estimated in the over- or under-exposed regions. τ defines the well-exposed range and is normally fixed at 0.9 in the experiments. The final weights in dynamic scenes are calculated by combining the visibility and consistency measures as: $W^i(x,y) = \frac{V^i(x,y) \times C^i(x,y)}{\sum_{i=1}^{N} V^i(x,y) \times C^i(x,y) + \epsilon}$. As shown in Figure 3(e), they give rise to a pleasant result where all visible details are preserved and no ghosting artifact is present.

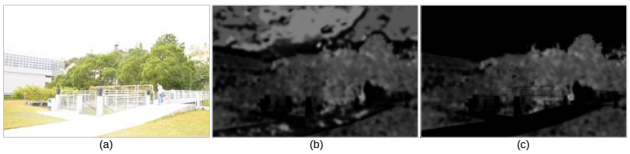
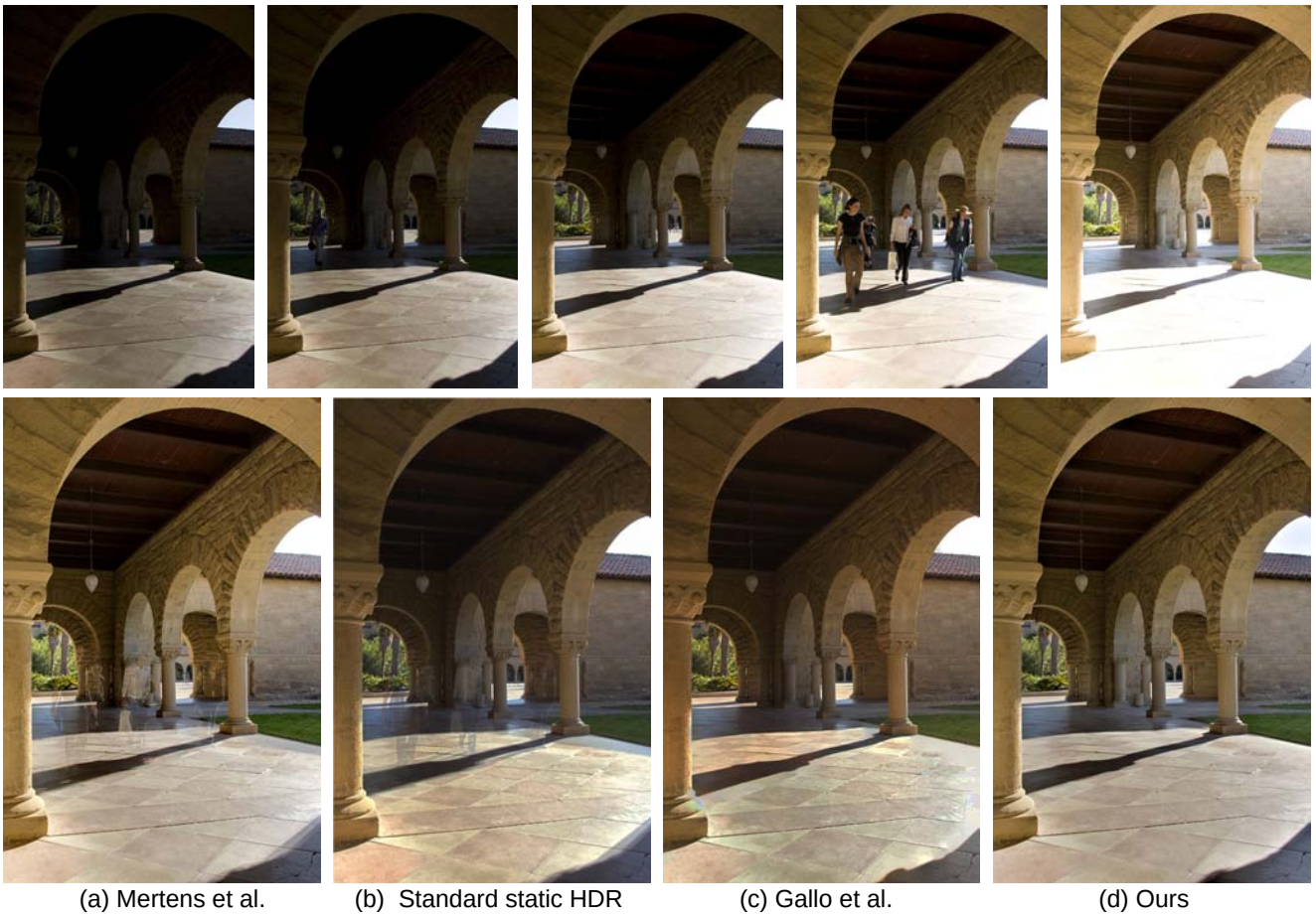


Figure 4. Effect of exposure correction on weighting map estimation. The left is the fifth input image of the sequence in Figure 3(a). The middle and right show its weighting maps before and after exposure correction respectively.

Figure 4 illustrates the effect of exposure correc in the example of Figure 3. Taking the sky region for example, since pixels in this region are over-exposed in most exposures, the weights obtained without exposure correction (i.e. removing the term $\alpha^i(x,y)$ in Eq (5)) are high as shown in Figure 4(b). The high weights favor over-exposure and suppress the occurrence of clouds in the composite result shown in Figure 3(d). After exposure correction, these weights become much lower as shown in Figure 4(c) and thus a desirable result with clouds is obtained in Figure 3(e).

(a) Mertens et al. (b) Khan et al. (c) Ours (d) Close-up comparison

Figure 5. Dynamic example (nine exposures). The top row shows four exposures of the sequence. The second row shows their weighting maps estimated by our method. The patches in (d) are cropped from (a), (b) and (c) for close-up comparison. Data courtesy of Erum Arif Khan.



(a) Mertens et al. (b) Standard static HDR (c) Gallo et al. (d) Ours

Figure 6. Dynamic example. The top row shows the input five exposures. (a) Mertens et al.'s result [9]. (b) Result obtained using standard HDR (radiometric calibration and tone mapping). (c) Result presented in Gallo et al. [16]. (d) Our result. Data courtesy of Orazio Gallo.

4. Experimental Results

In this section, the proposed algorithm is tested in various static and dynamic scenes with different types of exposure sequences. Besides, we also show its potential in flash and no-flash photography. The amazing thing about the proposed method is that it does not require much parameter tweaking. All experimental results were produced with the same parameters mentioned in the above sections. For color images, gradient extraction and cross-bilateral filtering are conducted only in the luminance channel. More results can be found in *http://www.ee.cuhk.edu.hk/∼zhangwei/GradComp.html*.

Dynamic Scene: To validate the effectiveness of the proposed method in dynamic scenes, we reproduce some results published before and make comparisons with the existing work. Figure 5 shows a scene with people moving from left to right. Apparently, the result (a) generated by Mertens et al. [9] suffers from severe ghosting artifacts. Although the deghosting result (b) presented by Khan et al. [18] is much better, some faint ghosting artifacts are still visible (see (d)). Besides, the entire image (b) looks underexposed especially around the trees and walls of the background. In contrast, our method generated a pleasant result (c) that is more informative and completely ghost-free. Figure 6 shows five differently exposed images with variable walking people. Similarly, the static methods like Mertens et al. [9] and standard HDR produced noticeable ghosting artifacts as shown in (a) and (b). Our result (d) is comparable to the result (c) presented in Gallo et al. [16] in terms of ghost removal. However, as shown in Figure 7, our result is less noisy and exhibits more details than Gallo et al.'s. It is worth noting that the performance of Gallo et al.'s method relies on the quality of the selected reference image. It cannot be used to remove the unwanted moving objects if they are present in all exposures (e.g. Figure 3) since no image is suitable for reference. Besides, the methods of Khan et al. [18] and Gallo et al. [16] are computational expensive, since camera calibration and tone mapping are both required. Moreover, Khan et al.'s method works in an iterative manner. In contrast, our method is quite simple and non-iterative. In the dynamic case, the current non-optimized Matlab implementation takes approximately 35 seconds to process four 1 megapixel images on a PC with an Intel Core2Duo 3.0GHz CPU.

Static Scene: In Figure 8, our method is compared to some prevalent tone mapping operators in a static scene. Note that our result is visually pleasing although it is generated with only six exposures of the original sequence. The overall quality is comparable to that of the tone mapping results produced from the whole radiance map.

Flash and No-flash Photography: Figure 9 shows two indoor images taken with flash and no-flash. The no-flash image is faithful to the ambient lighting, while the flash image reveals more details but suffers from hot spot artifacts. Our method with visibility assessment can combine the advantages of them and generate a desirable image free of hot spot.

Figure 7. Close-up comparison of (left) Gallo et al.'s result in Figure 6(c) and (right) ours in Figure 6(d).

5. Conclusions and Discussions

Image gradients convey important information about the latent scene. In this paper, we have shown that well utilization of image gradient makes it possible to handle the static and dynamic exposure composition in a simple but effective way. The proposed method shares some common limitations of HDR imaging. For example, it may not work well when the exposures contain severe sensor noise or blurring artifacts caused by camera shake. Besides, since it is assumed that the stationary parts of the scene are predominant in the sequence, the deghosting performance may not be satisfied if some portions of the scene change frequently such as trees blowing in the wind. Also, this assumption implies that at least three images are required for dynamic scenes. However, there is no such limitation for static scenes.

In the future, it is worthy to further improve the efficiency of the proposed method through optimized implementation. We also found that proper retouching such as color correction and sharpness adjustment can make the result more impressive, so it would be nice to add some retouching techniques to the current framework. Besides, we would like to further investigate the potential of this work in other related tasks.

Acknowledgment

We would like to thank Orazio Gallo for his kind help and suggestions. We also thank all anonymous reviewers for their valuable comments.

References

[1] M. Aggarwal and N. Ahuja. Split aperture imaging for high dynamic range. *IJCV*, v. 58, pp. 7-17, 2004.

[2] J. Tumblin, A. Agrawal and R. Raskar. Why I want a gradient camera. *CVPR*, 2005.

[3] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. *SIGGRAPH*, 1997.

[4] M. D. Grossberg and S. K. Nayar. Determining the camera response from images: what is knowable? *IEEE Trans. PAMI*, v. 25, pp. 1455-1467, 2003.

(a) Input sequence (b) Ours (c) Durand et al. (d) Reinhard et al. (e) Fattal et al.

Figure 8. Comparison with tone mapping operators. (a) Input six exposures. (b) Our composite result. (c) Durand et al. [5]. (d) Reinhard et al. [6]. (e) Fattal et al. [7]. Data courtesy of Paul Debevec.

Figure 9. Flash hot spot removal using visibility assessment. Top: flash and no-flash images. Bottom: our composite result. Data courtesy of Amit Agrawal.

[5] F. Durand and J. Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. *SIGGRAPH*, 2002.

[6] E. Reinhard, M. Stark, P. Shirley and J. Ferwerda. Photographic tone reproduction for digital images. *SIGGRAPH*, 2002.

[7] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. *SIGGRAPH*, 2002.

[8] Y. Li, L. Sharan and E. H. Adelson. Compressing and companding high dynamic range images with subband architectures. *SIGGRAPH*, 2005.

[9] T. Mertens, J. Kautz and F. V. Reeth. Exposure fusion: a simple and practical alternative to high dynamic range photography. *Computer Graphics Forum*, v. 28, pp. 161-171, 2009.

[10] R. Shanmuganathan and S. Chaudhuri. Bilateral filter based compositing for variable exposure photographs. *EUROGRAPHICS*, 2009.

[11] S. B. Kang, M. Uyttendaele, S. Winder and R. Szeliski. High dynamic range video. *SIGGRAPH*, 2003.

[12] E. Reinhard, G. Ward, S. Pattanaik and P. Debevec. *High dynamic range imaging: acquisition, display and image-based lighting*. Morgan Kaufmann Publishers, 2005.

[13] T. Grosch. Fast and robust high dynamic range image generation with camera and object movement. *Vision, Modeling and Visualization (VMV)*, 2006.

[14] K. Jacobs, C. Loscos and G. Ward. Automatic high-dynamic range image generation for dynamic scenes, *IEEE Computer Graphics and Applications*, v. 28, pp. 84-93, 2008.

[15] D. Sidibe, W. Puech and O. Strauss. Ghost detection and removal in high dynamic range images. *EUSIPCO*, 2009.

[16] O. Gallo, N. Gelfand, W. Chen, M. Tico and K. Pulli. Artifact-free high dynamic range imaging. *ICCP*, 2009.

[17] A. Eden, M. Uyttendaele and R. Szeliski. Seamless image stitching of scenes with large motions and exposure differences, *CVPR*, 2006.

[18] E. A. Khan, A. O. Akyüz and E. Reinhard. Ghost removal in high dynamic range images. *ICIP*, 2006.

[19] M. Pedone and J. Heikkilä. Constrain propagation for ghost removal in high dynamic range images. *VISAPP*, 2008.

[20] P. Pérez, M. Gangnet and A. Blake. Poisson image editing. *ACM Trans. Graph.*, v. 22, pp. 313-318, 2003.

[21] A. Agrawal, R. Raskar, S.K. Nayar, and Y. Li. Removing photography artifacts using gradient projection and flash-exposure sampling. *ACM Trans. Graph.*, v. 24, pp. 828-835, 2005.

[22] G. Ward. Fast, robust image registration for compositing high dynamic range photographs from handheld exposures. *Journal of Graphics Tools*, v. 8, pp. 17-30, 2003.

[23] S. Paris and F. Durand. A fast approximation of the bilateral filter using a signal processing approach. *ECCV*, 2006.

[24] P. J. Burt and E. H. Adelson. A multiresolution spline with application to image mosaics. *ACM Trans. Graph.*, v. 2, pp. 217-236, 1983.