A Multi-exposure Fusion Method Based on Locality Properties

Yuanchao Bai[1], Huizhu Jia[1,*], Hengjin Liu[1], Guoqing Xiang[1], Xiaodong Xie[1], Ming Jiang[2], and Wen Gao[1]

[1] National Engineering Laboratory for Video Technology,
Peking University, Beijing 100871, China
{yuanchao.bai,hzjia,hjliu,donxie,wgao}@pku.edu.cn
[2] Department of Information Sciences, School of Mathematical Sciences,
Peking University, Beijing 100871, China
ming-jiang@ieee.org

Abstract. A new method is proposed for fusing a multi-exposure sequence of images into a high quality image based on the locality properties of the sequence. We divide the images into uniform blocks and use variance to represent the information of blocks. The richest information (RI) image is computed by piecing together blocks with largest variances. We assume that images in the sequence are high dimensional data points lying on the same neighbourhood and borrow the idea from locally linear embedding (LLE) to fuse a result image which is closest to the RI image. The result is comparable to the state-of-art tone mapping operators and other exposure fusion methods.

Keywords: Exposure fusion, High dynamic range, Locality properties, LLE.

1 Introduction

Digital cameras are developing quickly in producing higher-resolution images, but the limited dynamic range is still a challenge. Currently, the vast majority of images are represented in 8 bits per pixel for each colour channel, which is called low dynamic range (LDR) image, in contrast with the high dynamic range (HDR) in real world scenes. As a result, images captured by cameras in HDR scenes are either under- or over-exposed, missing parts of the information in dark or high-light regions. In order to achieve the full dynamic range, a bracketed exposure sequence is often made and is turned into a single image for convenient display.

There are several techniques to create such multi-exposure images. One of these techniques, reviewed in details by Erik Reinhard et al. [8], is to create a kind of HDR format images that uses floating point number to represent pixel channels and then to tone map the HDR format images into displayable gamut.

* Corresponding author.

W.T. Ooi et al. (Eds.): PCM 2014, LNCS 8879, pp. 333–342, 2014.

However, the obvious drawback of this approach is that it needs to create the HDR format image first, which is complex and costs lots of memory for the users who only want displayable multi-exposure results. Besides, the camera parameters that are required to calculate the radiometric response function [1] are often unknown by common users.

Another kind of techniques is called exposure fusion, which allocates a normalized weight to every image and then fuses the bracketed sequence into a multi-exposure image.

Tom Mertens et al. [7] proposed a technique for fusing images. The technique blends multiple exposures, guided by simple quality measures like saturation and contrast. Images are fused in a multiresolution fashion to deal with the seams.

Goshtasby [4] presented an block-wise fusion techniques. The technique first calculates the entropy of each block in each image and selects the blocks that contains the most information. Then the blocks are fused to a high quality image by a monotonically decreasing blending function. An algorithm to determine the optimal block size is also proposed. However, the optimization algorithm is slow and may be not necessary because the entropy is changed by uniform block size and is not the larger, the better.

Kotwal et al. [6] defined exposure fusion as a multi-objective optimization problem based on desired characteristics of the output and provided the solution using an Euler-Lagrange technique. The proposed technique yields visually appealing fused images with a high value of contrast. The slow iteration is the inevitable drawback of the algorithm.

Ashish et al. [11] proposed a low complexity detail preserving multi-exposure image fusion method. The fusion is performed pixel-by-pixel and does not involve any filtering or transformation. However, the saturation and the contrast of the multi-exposure method is not that good.

In this paper, we take advantage of the locality properties, the locality of pixels among their neighbourhoods and the locality of images with different exposures. Based on these properties, we make a model to optimize the information in the final image and also take the visual quality into consideration. The proposed algorithm is block-wised and each block can run parallel. The experiments show that quality of the results is comparable with tone mapping operators and other exposure fusion methods.

The rest of the paper is organized as follows. In section 2, we will introduce our algorithm in details. In section 3, many experiments and comparisons are done. In section 4, we make a conclusion and the future expectation.

2 Exposure Fusion Based on Locality Properties

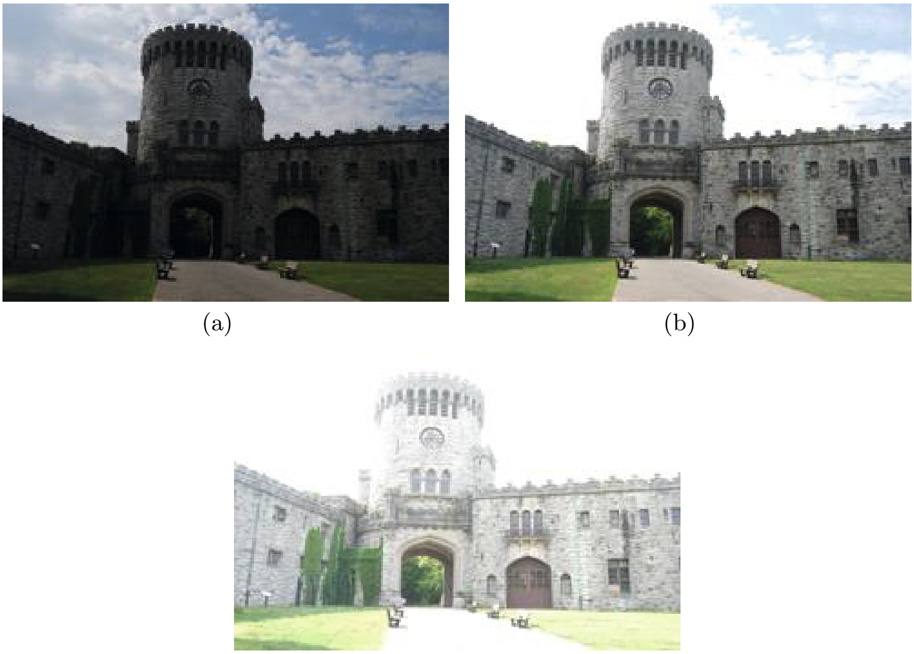
Our exposure fusion method aims at allocating a normalized weight to every pixel of every image in the multi-exposure image sequence and then fuses these images to a final image. To achieve this goal, we choose the parts with rich information in the multi-exposure image sequence and use the locality properties to obtain the weights. In our method, we assume that the images are well aligned, possibly using a registration algorithm [13],[5].

2.1 Linear Fusion Model

Most of the exposure fusion method use the Linear Fusion Model (LFM), which means that the result is computed by a weighted average at each pixel in N images. To make sure that the fused pixels is in the 0 to 255 display range, the weights of N images at each pixel (i, j) are normalized:

$$F_{ij} = \sum_{k=1}^{N} W_{ij}^{k} I_{ij}^{k}, \quad \text{s.t.} \quad \sum_{k=1}^{N} W_{ij}^{k} = 1 \tag{1}$$

where i, j, k mean the pixel (i, j) of k-th input image in the sequence. I is the pixel value, W is the weight and F is the result. An example of a multi-exposure sequence is shown in Fig. 1. According to this model, we only need to compute the weight for each pixel, but it is not straightforward when rich information and visual quality are taken into consideration. Our method is based on LFM, which takes advantage of locality properties to guide the fusion.



(a) (b) (c)

Fig. 1. Multi-exposure image sequence. The photos are taken in different exposure time.

2.2 Information Measure with Locality Property

Each area of a scene has a corresponding best exposure. In the other word, one of the images in the exposure sequence contains the richest information of the area [4]. Inspired by this idea, we find the bridge between the visual quality and countable quantity. Although the fusion process is pixel-wised, one pixel cannot represent the quantity of information. We consider one pixel and its neighbourhood. When the neighbourhood is not very large, the relation between the visual quality and information quantity is monotonic, which means that the better the visual quality is, the more information the neighbourhood has. However, the monotonicity does not hold when the neighbourhood is large, because the neighbourhood may contain areas that have different best exposures. We conclude this result as the first locality property that guide our algorithm.

We use the variance to measure the information, which is a widely used concept in data analysis. If an area is well-exposed, the variance will reach its largest. When an area is under- or over-exposed, the area will have less details and the variance will be small. To simplify the calculation, we compute the variance on the $m \times m$ uniform blocks:

$$\mathrm{Var}(\mathbf{b}) = \frac{1}{m^2}\sum_{i=1}^{m^2}(b_i - \bar{b})^2 \tag{2}$$

where the $\mathrm{Var}(\mathbf{b})$ is the variance of a block $\mathbf{b}$. b_i is the i-th pixel value and $\bar{b}$ is the mean of pixels in the block. We divide the images into $m \times m$ sized blocks and compare the variances of blocks at the same place in different images in the sequence. The blocks with largest variances are extracted and pieced together to form an initial starting-point image. This image contains the richest information but is not comfortable for viewing, as shown in Fig. 2. The richest information image is the important variable of our next fusion step.

(a) (b)

Fig. 2. The richest information image. The grayscale one is shown in (a) with the block size 16×16. The colored one is shown in (b).

2.3 Fusion with Locality Property

As the richest information (RI) image has been obtained, we fuse the multi-exposure sequence according to this image. We could directly smooth the RI image, but undesirable halos around the edge or shades are present, as demonstrated in Fig. 3(a).

(a) (b)

Fig. 3. (a) is the result from directly smoothing RI image with halos. (b) is our algorithm with one weight per image.

To address the halos-shades problem and keep the information as much as possible, we borrow the idea from locally linear embedding (LLE) [10], which is an important algorithm in manifold learning. The LLE algorithm is based on simple geometric intuitions. Suppose there are N points in a high dimensional data space of dimensionality D. The main concept of LLE is that each data point and its neighbours are expected to lie on or close to locally linear patch of the manifold, so the data point can be reconstructed from its neighbours. In exposure fusion, the images in exposure sequence and the RI image are the high dimensional data points. As these images represent the same scene, these points are assumed to lie on a locally linear patch and can be reconstructed by neighbours. We propose a fusion model according to this assumption:

$$\hat{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmin}} \left\| \sum_{k=1}^{N} W^k \mathbf{I}^k - \mathbf{R} \right\|, \quad \text{s.t.} \quad \sum_{k=1}^{N} W^k = 1 \tag{3}$$

where $\mathbf{R}$ is the RI image, $\mathbf{I}^k$ is the k-th image in the exposure sequence, which is the neighbour of RI image. Our goal is to optimize the weights $\hat{\mathbf{W}}$ to fuse the sequence and make the fused image closest to the RI image. The solution to (3) can be derived below:

$$\hat{\mathbf{W}} = \left(\mathbf{I}^T \mathbf{I}\right)^{-1} \cdot \left[\mathbf{I}^T \mathbf{r} - \mathbb{1} \cdot \frac{\mathbb{1}^T (\mathbf{I}^T \mathbf{I})^{-1} \mathbf{I}^T \mathbf{r} - 1}{\mathbb{1}^T (\mathbf{I}^T \mathbf{I})^{-1} \mathbb{1}} \right] \tag{4}$$

where $\mathbf{I} = (\mathbf{i}^1, \mathbf{i}^1, \cdots, \mathbf{i}^N)$, $\mathbf{i}^k = \mathrm{vec}(\mathbf{I}^k)$. The variable $\mathbf{r} = \mathrm{vec}(\mathbf{R})$ and $\mathbb{1}$ is the vector whose all entries are 1. However, the answer can not be found when the inverse matrix of $\mathbf{I}^T\mathbf{I}$ does not exist. To solve (3) in any cases and more efficiently, gram matrix has been involved, let $\mathbf{G}_{jk}$ be the (j, k) entry of gram matrix:

$$\mathbf{G}_{jk} = \left(\langle \mathbf{i}^j - \mathbf{r}, \mathbf{i}^k - \mathbf{r} \rangle\right)_{j,k \in [1,N]} \tag{5}$$

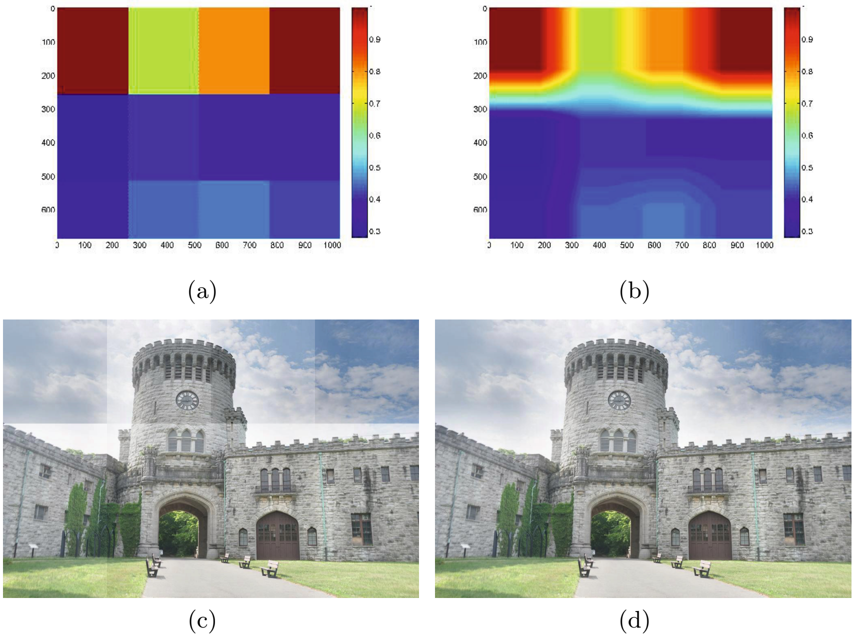
Then we compute the equation:

$$\mathbf{W} = (\mathbf{G} + \delta\mathbf{E})^{-1} \cdot \mathbb{1} \tag{6}$$

δ is a small number and $\mathbf{E}$ is the identity matrix. $\mathbf{G} + \delta\mathbf{E}$ is to make the matrix a positive-definite matrix and invertible. After that, the answer $\hat{\mathbf{W}}$ is solved by normalizing $\mathbf{W}$.

$$\hat{\mathbf{W}} = \frac{\mathbf{W}}{\sum_{k=1}^{N} W^k} \tag{7}$$

The result is shown in Fig. 3(b). The halo does not exist and the visual quality is improved. Fig. 3(b) contains more information than any of the images in Fig. 1 but the information is less than that in Fig. 3(a).



(a) (b) (c) (d)

Fig. 4. (a) shows the un-smoothed weight map. (b) shows the weight map that has been smoothed. (c) is the result fused by un-smoothed weight map. (d) is the result fused by smoothed weight map.

To achieve more information, one weight per image is not satisfying. We do the fusion on $M \times M$ sized windows, which is much larger than $m \times m$ sized blocks. By this way, we can keep more local information. Then we use the mean filter to smooth the edges of weight map between windows. The 2-dimensional mean filter can be decomposed into two 1-dimensional mean filter and is able to run parallel. Besides, only pixels around the window edges need to be smoothed, the filter process can be quite efficient. According to the analysis of [12], fusion in groups of two results in lower computation complexity. In our fusion, we also take advantage of this concept to compute faster and get better result than fusing all images together at a time. For example, we first fuse the images in Fig. 1(a) and Fig. 1(b) to get the "image12". Secondly, we fuse "image12" and Fig. 1(c) to achieve the result. Fig. 4 shows the second step.

3 Experiments

Our experiments are done on the JPG-encoded photographs. We use 16×16 sized blocks and 256×256 sized windows for all the examples. In our exposure fusion, we do not need any information about the camera, except in Fig. 6. In Fig. 6, we need the exposure time of every image to create the HDR format image to test the tone mapping algorithm.

In Fig. 5, we compare our result to Tom Mertens et al.'s approach [7], our result is a little bit brighter than theirs but we both contains all the details and information in the exposure sequence. The halo does not exist in the experiment.

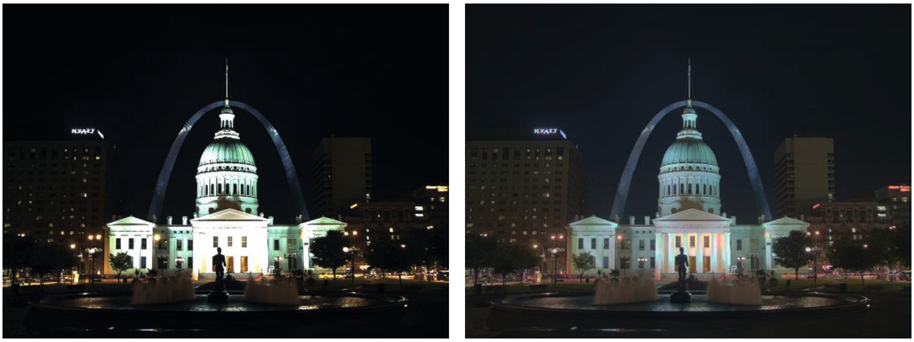
(a) Mertens et al. [7]　　(b) Our algorithm

Fig. 5. Comparison with other exposure fusion. [7] is shown in (a). Our algorithm is shown in (b).

Fig. 6 is the multi-exposure sequence. We generate a HDR format image with [1] from this sequence. Then we use the HDR format image to run Durand et al. [3], Reinhard et al. [9] and Drago et al. [2] tone mapping methods. Three tone mapping results and our exposure fusion result are shown in Fig. 7. Our algorithm result looks better than these tone mapping results but in the area of neon, the word is not that clear.

(a) 1/30s (b) 1/4s (c) 2/1s (d) 15/1s

Fig. 6. Images with different exposure time for HDR recovering



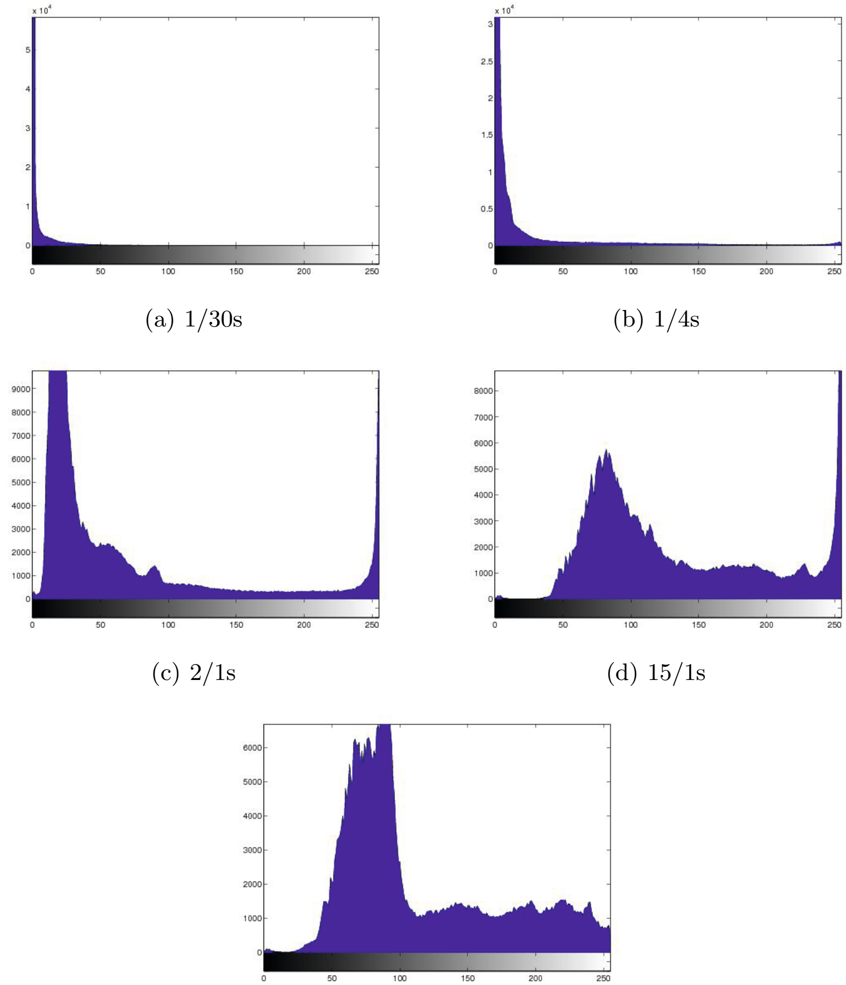
(a) Durand et al. [3] (b) Reinhard et al. [9]

(c) Drago et al. [2] (d) Our algorithm

Fig. 7. Comparison with tone mapping operators. [3] is shown in (a). [9] is shown in (b). [2] is shown in (c). Our algorithm is shown in (d).

In Fig. 8, we also compare the histogram of the images in the sequence to our result. Histogram of Fig. 6 are shown from Fig. 8(a) to Fig. 8(d) and the histogram of our result is shown in Fig. 8(e). We can see that our algorithm can center the pixel values and under- or over-exposure pixels decrease a lot.

From these comparisons, the proposed algorithm successfully extracts well-exposure parts from each image and performs well in visual quality.



(a) 1/30s

(b) 1/4s

(c) 2/1s

(d) 15/1s

(e) Our result

Fig. 8. Comparison in histograms. (a)-(d) are the histograms of images in Fig. 6. (e) is the histogram of our result.

4 Conclusion

In this paper, we exploit the locality properties in the multi-exposure sequence to extract information from different exposure images. Also, we borrow the concept from LLE to model our fusion and achieve good results. In our exposure fusion, no camera information has to be known and no HDR format images will be generated.

As our algorithm can run parallel in block-level, it is more suitable to take advantage of this parallelism and implement our algorithm on GPU. Finally, as HDR video cameras are developing quickly, e.g. RED-EPIC camera, we are trying to apply this technique on multi-exposure video sequences for the next step.

Acknowledgements. This work is partially supported by grants from the Chinese National Natural Science Foundation under contract No.61171139 and National High Technology Research and Development Program of China (863 Program) under contract No.2012AA011703.

References

1. Debevec, P.E., Malik, J.: Recovering high dynamic range radiance maps from photographs. In: ACM SIGGRAPH 2008 Classes, p. 31. ACM (2008)
2. Drago, F., Myszkowski, K., Annen, T., Chiba, N.: Adaptive logarithmic mapping for displaying high contrast scenes. Computer Graphics Forum 22, 419–426 (2003)
3. Durand, F., Dorsey, J.: Fast bilateral filtering for the display of high-dynamic-range images. ACM Transactions on Graphics (TOG) 21, 257–266 (2002)
4. Goshtasby, A.A.: Fusion of multi-exposure images. Image and Vision Computing 23, 611–618 (2005)
5. Grosch, T.: Fast and robust high dynamic range image generation with camera and object movement. In: Vision, Modeling and Visualization, pp. 277–284. RWTH Aachen (2006)
6. Kotwal, K., Chaudhuri, S.: An optimization-based approach to fusion of multi-exposure, low dynamic range images. In: 2011 IEEE International Conference on Information Fusion (FUSION), Chicago, IL, USA, July 5-8, p. 7 (2011)
7. Mertens, T., Kautz, J., Van Reeth, F.: Exposure fusion. In: 15th Pacific Conference on Computer Graphics and Applications, PG 2007, pp. 382–390. IEEE (2007)
8. Reinhard, E., Heidrich, W., Debevec, P., Pattanaik, S., Ward, G., Myszkowski, K.: High dynamic range imaging: acquisition, display, and image-based lighting. Morgan Kaufmann (2010)
9. Reinhard, E., Stark, M.M., Shirley, P., Ferwerda, J.A.: Photographic tone reproduction for digital images. In: SIGGRAPH, pp. 267–276 (2002)
10. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500), 2323–2326 (2000)
11. Vanmali, A.V., Deshmukh, S.S., Gadre, V.M.: Low complexity detail preserving multi-exposure image fusion for images with balanced exposure. In: 2013 National Conference on Communications (NCC), pp. 1–5. IEEE (2013)
12. Wang, J., Xu, D., Li, B.: Exposure fusion based on steerable pyramid for displaying high dynamic range scenes. Optical Engineering 48(11), 117003 (2009)
13. Ward, G.: Fast, robust image registration for compositing high dynamic range photographs from handheld exposures. Journal of Graphics Tools 8, 17–30 (2003)