HIGH DYNAMIC RANGE IMAGE WITH TWO EXPOSURES

[1] *Chih-Chun Lee* (*李知駿*), [2] *Chiou-Shann Fuh* (*傅楸善*)

[1] Dept. of Computer Science and Information Engineering,
National Taiwan University, Taipei, Taiwan
E-mail: r99922115@ntu.edu.tw
[2] Dept. of Computer Science and Information Engineering,
National Taiwan University, Taipei, Taiwan
E-mail: fuh@csie.ntu.edu.tw

ABSTRACT

In related work, most algorithms generate high dynamic range image by fusing several differently exposed images. In this paper, we propose an image fusion pipeline that generates a high dynamic range image by fusing long- and short-exposure raw images. Some algorithms generating high dynamic range image with two exposure images may incur artifacts. Our method uses the property of raw image to overcome the artifact. By performing several quality measures such as brightness, contrast, and saturation, we estimate a weighting map, and then fuse the two raw images by multi-resolution blending technique to generate seamless high dynamic range image. We also combine the image alignment and ghost removal algorithm to deal with vibration problem and moving objects and make the fusion result clearer and better.

Keywords *High Dynamic Range; Image Fusion;*

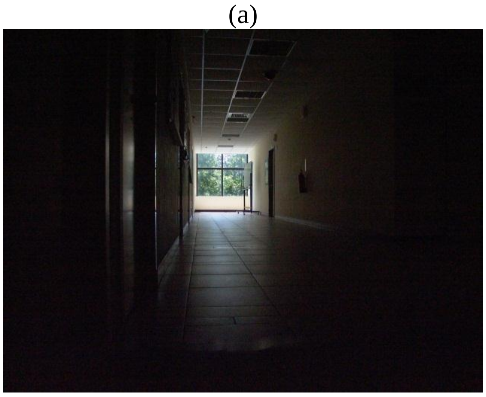
1. INTRODUCTION

The dynamic range is a ratio between the most brightness area and the most darkness area in a scene. When taking picture with outdoor scenes, we can often see some place in the image is over-exposed or under-exposed because the dynamic range in real world is much higher than the range of most image sensors, but human eyes have higher dynamic range than image sensors. For this reason, we need high dynamic range imaging to increase the dynamic range of images and make the final images look like the scenes viewed from the human eyes.

High dynamic range imaging has been extensively studied in the past years, due to the widely used consumer digital cameras and smart phones using this technology to improve the quality of images. However, increasing the dynamic range by improving the hardware costs much, so many researchers address this problem by software solutions. The most common method is fusing a stack of different exposure images to generate a high dynamic range image, because different exposure images capture various parts of dynamic range in a scene.

In this paper, we propose an image fusion pipeline that generates a high dynamic range image by fusing two exposure images. This pipeline chooses two appropriate exposure images from exposure sequence and applies alignment and ghost removal algorithms to deal with vibration problem and moving objects, and blends them by performing simple quality measures.

(a)



(b)

Fig. 1: (a) Indoor is clear, and outdoor is over-exposure. (b) Outdoor is clear, and indoor is under-exposure.

2. RELATED WORK

High dynamic range (HDR) imaging technique has been studied in the past years. Some researchers tried to extend the dynamic range by developing new HDR camera prototypes [5]. However, due to the high price on hardware, most researchers solve this problem with software solutions that achieve high dynamic range imaging by fusing a stack of different exposure images. Generally speaking, the existing algorithms can be classified into static high dynamic range and dynamic high dynamic range [15].

Static high dynamic range algorithms require the target scene completely still. Because of the process of these algorithms cannot erase the moving objects in the target scene. If we use these algorithms in dynamic scene, the final image will suffer from ghosting artifacts. The standard high dynamic range process consists of two steps: determining the camera response curve [8] and recovering the radiance map [6]. The radiance map is a high dynamic range image. We will describe in Section 2.1. The second step is performing tone mapping algorithm to the radiance map. Tone mapping is used to compress the high dynamic range into low dynamic range of normal monitors. In addition to recovering the radiance map, some other algorithms try to generate a tone-mapped-like high dynamic range image by fusing exposure images directly instead of determining a camera response curve [13]. These algorithms do not need to apply tone mapping algorithms so they are more efficient than recovering the radiance map. We will give an example in Section 2.2.

Dynamic high dynamic range algorithms usually estimate motion objects first and then generate a ghost map. By removing the contributions of these regions in the ghost map, we can produce a ghost-free high dynamic range result. Many different methods are used to detect the motion objects, such as variance measurement [3] and entropy calculation [4]. These algorithms generate radiance map after calculating the ghost map. Some other methods such as gradient estimation [15] erase ghost artifacts and generate high dynamic range result by fusing exposure images directly.

2.1. Recovering the Radiance Map

E. Debevec and J. Malik [8] present a method of recovering high dynamic range radiance maps from photographs taken with camera. Their algorithm uses differently exposed photographs to recover the response function of the imaging process. With the known response function, the algorithm can fuse the multiple photographs into a single, high dynamic range radiance map whose pixel values are proportional to the true radiance values in the real world.

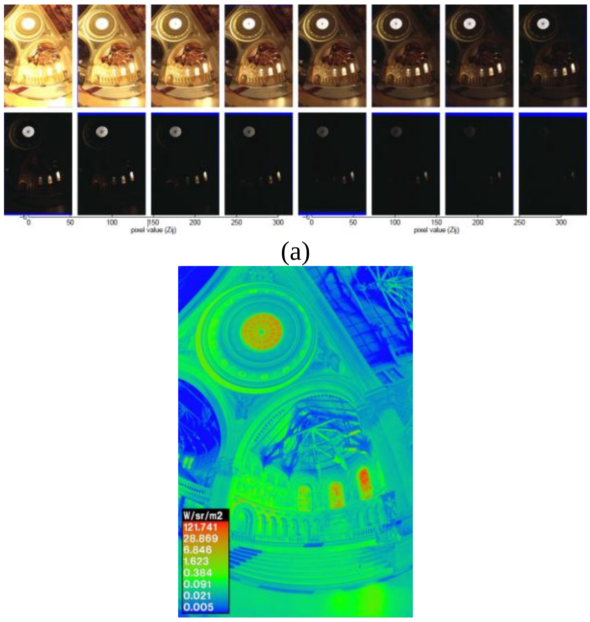


Fig. 2: (a) Exposure images [8]. (b) Radiance map [8].

2.2. Exposure Fusion

T. Mertens, J. Kautz, and F. Van Reeth proposed a technique for fusing a stack of exposure sequence into a high quality image, without converting to high dynamic range image first [13]. Their idea keeps only the best parts in multi-exposure image sequence. By performing several perceptual quality measures, we give the appropriate pixels higher weighting. Thus those best parts in multi-exposure image sequence can be preserved in the final image. Besides, they use multi-solution blending to avoid brightness variation.

The goal is to generate weight maps for each input image. A high weight means that a pixel should appear in the final image. They use three quality measures to estimate weight, contrast, saturation, and well-exposedness. The contrast measure wants to detect the high frequency details. Since they think that the details will decrease when the exposure is longer or lower, they try to keep edges and texture in each input image. By applying a Laplacian filter to each input image and take the absolute value of the filter response, it can give high weight to edges and texture. We use indicator C to represent contrast. The saturation measure wants to find the colorful regions. They want final brilliant image, so they assign high weight to high saturation pixels. The saturation is computed as the standard deviation within R, G, and B channels. We use indicator S to represent saturation. The well-exposedness measure wants to locate pixels which are exposed appropriately. They want to keep pixel values which are not underexposure or overexposure. They use a Gaussian curve to assign weight. We use indicator E to represent exposedness. After several quality measures, we integrate these results into a weight map.

Fig. 3: Weight maps for each input image [13].

They use a multi-resolution blending method [9]. This method can blend two images seamlessly by using alpha mask and implementing at multiple resolutions. These input images are decomposed into Laplacian pyramid and fused in each level separately. They improve the blending method in this case where we blend *N* input images and regard *N* weight maps as alpha mask.

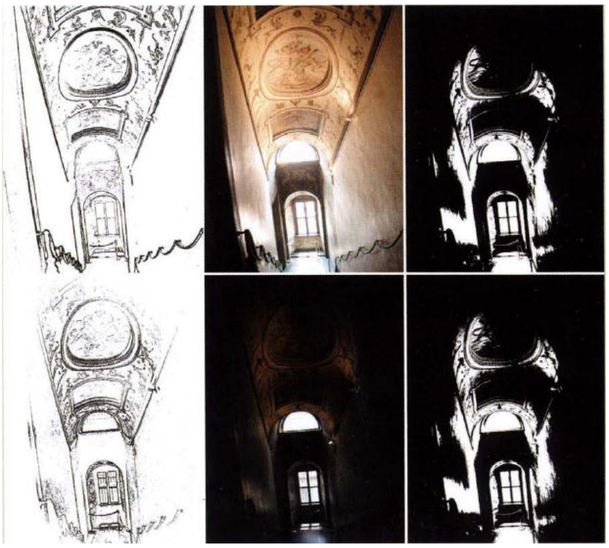
3. BACKGROUND

3.1. Image Alignment

Generally speaking, we capture all exposure images with the aid of a tripod to make sure the pixels of several differently exposed images are aligned. However, even if we use the tripod, the pixels will still have some shift. All high dynamic range imaging algorithms introduced in Section 2 needs the input exposure sequence totally aligned. Otherwise, details of result image will become blurry. Thus, the image alignment algorithm is needed to compensate the offset.

Fig. 4: Left image fused from unaligned exposures and right image fused from aligned exposures. We can see that the details of right are clearer than those of left [3].

However, aligning different exposure images cannot use conventional alignment algorithms such as detecting and matching edges. Since edges decrease when the pixel value is near overexposure or underexposure, we use Median Threshold Bitmap (MTB) algorithm to align the exposure images [3].



(a) (b) (c)

Fig. 5: (a) Corresponding edge maps of (b). (b) Original unaligned images. (c) MTB results of (b) [3].

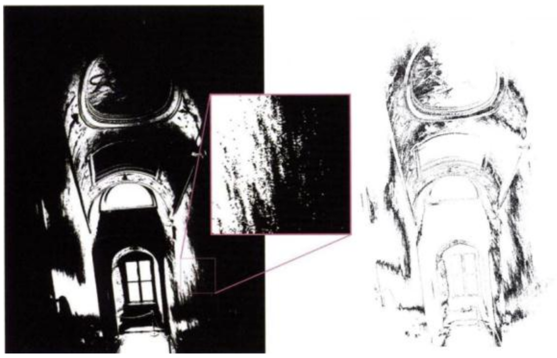
MTB algorithm is insensitive to image exposure. It can align *N* grayscale images fast by transforming them into bitmaps. The computation focuses on integer pixel offsets, because they can be used to quickly recombine the exposures without resampling. It is inappropriate for arbitrary camera movements such as zooming and tilting, but empirical evidence suggests that handheld sequences do not require rotational alignment in about 90% of cases [3].

In order to apply MTB algorithm to *N* exposure images, we select one of the *N* images as the reference image, and the output of the algorithm is a series of *N*-1 (*x*,*y*) integer offsets for each of the remaining images relative to this reference. Then we determine the median value by computing the histogram of grayscale image. We regard the median value as a threshold and use the threshold to create a bitmap. If the pixel value is less than or equal to threshold, the pixel value is set to 0, otherwise is set to 1. In contrast to the edge maps, the results of MTB are nearly identical for the two exposures. We can apply exclusive-or (XOR) operator to bitmaps to show where the two images are misaligned. All we have to do is minimizing the XOR difference. The brute force method is to test every offset within the allowed range, computing the XOR difference at each offset and taking the coordinate pair corresponding to the minimum difference. Another more efficient method is based on an image pyramid.

First we down-sample the MTB results with log(max_offset) levels past the base resolution. To compute the offset, we start with the lowest-resolution pair and compute the minimum difference offset within a range of ± 1 pixel. At the next resolution level, we multiply this offset by 2 and compute the minimum difference offset within a ± 1 pixel range of this previous offset [3]. This continues to the original level, and we get the final offset result.

Fig. 6: Pyramid of MTB results is used to align exposures one bit at a time [3].

In order to make it more resistant to noise, we discard pixels that are close to the threshold. We generate an exclusion map where all bits are set to 0 if pixel value is within ±4 of the median value. Every time we compute the minimum difference offset, we combine the exclusion map with XOR result by AND operator to ignore the noise.



(a) (b)

Fig. 7: (a) The HDR result without ghost artifact. (b) Uncertainty image [4].

3.2. Ghost Removal

After image alignment, we still have some problems to solve. In our early description, ghost artifacts are caused by moving object such as people passing through when taking exposure sequence. Ghost removal usually estimates motion objects first and then generates a ghost map.

(a)

(b)

Fig. 8: (a) An exposure sequence with moving object. (b) Result HDR image with ghost artifact [4].

K. Jacobs, C. Loscos, and G. Ward introduce a method to generate ghost map by estimating entropy in exposure sequence [4]. Entropy is a statistical measure that defines the uncertainty that remains about a system [4]. The entropy of an image gives a measure of the uncertainty of the pixels in the image. If all intensity values are equal, the entropy is zero and if all intensity values are different, the entropy is high. They use local entropy to determine whether moving objects pass through or not. The local entropy is high in areas with many details, such as edges. These areas do not change between the images in the exposure sequence, except when corrupted by moving object or saturation. Their method creates an uncertainty image *UI*. Pixels in *UI* with high values indicate moving objects.

(a)

(b)

Fig. 9: (a) The HDR result without ghost artifact. (b) Uncertainty image [4].

4. OUR METHOD

We propose an image fusion pipeline that can generate an HDR image by fusing two differently exposed raw images. First, we have to determine the input images from the exposure sequence, and then we align them by Median Threshold Bitmap (MTB) which is introduced in Section 3.1. Finally, we fuse two raw images by a weighting map which is generated by moving object detection, brightness measurement, contrast measurement, and saturation measurement.

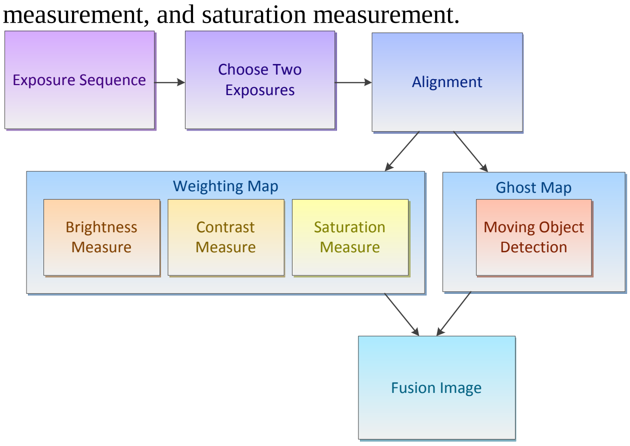


Fig. 10: The flow of high dynamic range image pipeline.

4.1. Choose Exposure Images

How to choose exposure images is an important part in high dynamic range image pipeline. The long exposure should contain details of dark area and the short exposure should contain details of bright area so that the fusion image can preserve all details in the scene. If we do not choose appropriate exposures, the effect of this method will degrade. Thus we design a method to choose exposure images.

The idea is very simple, we want the long exposure to have details of dark area and the short exposure to have details of bright area. First, compute the average intensity of each exposure image of exposure sequence. We take the medium intensity as a reference image. And then apply Otsu's Method [7] to reference image to find a suitable threshold to split the reference image into bright area and dark area.

Otsu's Method can be used to create binary images, which uses variance to estimate threshold. Variance can be used to measure homogeneity of a group. Otsu's method continuously separates the histogram of reference image into two groups and minimizes within-group variance.

After applying Otsu's method, the reference image can be split into bright area and dark area. In order to find long-exposure image, we compute the average intensity of bright area in each image of exposure sequence, and choose the image with average intensity nearest medium intensity as the short-exposure image. The way to find long- exposure image is similar, the dissimilitude is that we compute the average intensity in dark area.

4.2. Image Fusion

After choosing two exposure images, we use Medium Threshold Bitmap (MTB) to align these images to ensure the following measures will be accurate. Our method shares the same spirit with Exposure Fusion in Section 2.2 but it cannot deal with dynamic scene. We apply several quality measures to long- and short-exposure images to estimate weighting map and ghost map.

In order to generate ghost map, we use entropy measurement in Section 3.2. After moving object detection, we get a ghost map that shows the regions of moving objects in the long- and short-exposure images. These regions are determined as motion pixels, otherwise static pixels. When we fuse the two exposure images, we consider whether or not the pixels are determined as motion pixels. If they are motion pixels, the fusion image is computed using only the short-exposure image data. If they are static pixels, we fuse the long- and short-exposure images by weighting map. We use indicator G to represent ghost map.

The weighting map is composed of brightness measure, contrast measure, and saturation measure. In these measures, the brightness is the principal feature in weighting map. About brightness measure, the details of bright area should totally come from short-exposure image and the details of dark area should totally come from long-exposure. We apply Otsu's method to long-exposure image to split it into bright area and short area, and then we set the weighting of dark area 1, otherwise 0. We use indicator B to represent brightness. Next we apply a Laplacian filter to the long- and short-exposure images to estimate contrast. We take the absolute value of the filter output and normalize it as weighting. We use indicator C to represent contrast. The saturation measure applies to the long- and short-exposure images and normalizes the response as weighting. It assigns high weighting to the high saturation pixels. We use indicator S to represent saturation.

For each pixel, we combine the results from the different measures:

$$W_L(x,y) = B_L(x,y) \times (1 + C_L(x,y)) \times (1 + S_L(x,y)) \quad (4.1)$$

$$W_S(x,y) = (1 - B_L(x,y)) \times (1 + C_S(x,y)) \times (1 + S_S(x,y)) \quad (4.2)$$

the subscripts L and S mean long- or short-exposure images, respectively. Next we normalize the value of long- and short-weighting maps such that they sum to one at each pixel.

$$\widehat{W}_L(x,y) = \frac{W_L(x,y)}{W_L(x,y) + W_S(x,y)} \times G(x,y) \quad (4.3)$$

$$\widehat{W}_S(x,y) = \frac{W_S(x,y)}{W_L(x,y) + W_S(x,y)} \times G(x,y) \quad (4.4)$$

where $\widehat{W}(x,y)$ and $G(x,y)$ denote the fusion weighting result of ghost map. Then we use the following equation to fuse long- and short-exposure images.

$$F(x,y) = \widehat{W}_L(x,y) \times I_L(x,y) + \widehat{W}_S(x,y) \times I_S(x,y) \times C \quad (4.5)$$

where $F(x, y)$ and $I(x, y)$ denote the fusion image and intensity. The constant C is the ratio of the long- and short-exposure time. This constant increases the intensity of short-exposure to compensate the intensity offset between long- and short-exposure images. Finally, the high dynamic range image is produced by long- and short-exposure images seamlessly based on a multi-resolution blending similar to Exposure Fusion in Section 2.2.

5. EXPERIMENTAL RESULTS

(a)

(b)

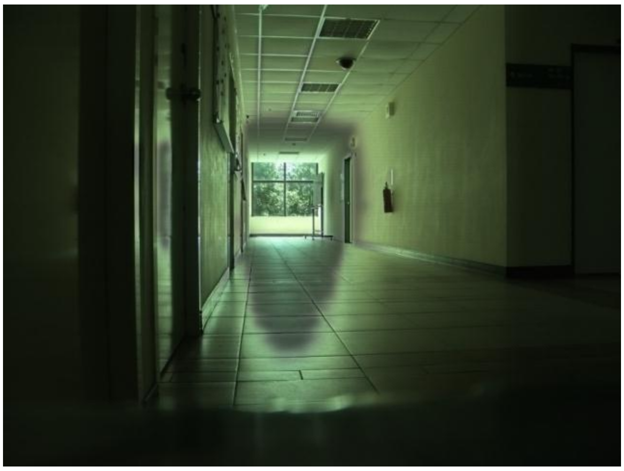
(c) (d)

Fig. 11: (a) Photoshop. (b) Our method. (c) Moving objects in (a). (d) Moving objects in (b)

(a)

(b)

Fig. 12: (a) Photoshop. (b) Our method.



(a)

(b)

Fig. 13: (a) Exposure Fusion. (b) Our method.

(a)

(b)

Fig. 14: (a) Exposure Fusion. (b) Our method.

(a)

(b)

Fig. 15: (a) Exposure Fusion. (b) Our method.

6. CONCLUSION AND FUTURE WORK

In this paper, we propose an image fusion pipeline to generate a high dynamic range image with two differently exposed images. Our method effective eliminates the brightness artifacts which occur when fusing high dynamic range image with two inappropriate exposure images. This method can also preserve the details of the long- and short-exposure and extend the dynamic range of the final image. However, seam problem may occur if the exposure ratio between long- and short-exposure is too high. Because the relationship between the exposure and the intensity of the raw images is not linear, there exists a little offset. When we compensate the bright area, the offset will also be amplified. This makes the boundary between bright area and dark area more distinct. If we limit the exposure ratio between long- and short-exposure, this problem can be solved.

REFERENCES

[1] Exposureguide.com, "Focusing Basics," http://www.exposureguide.com/focusing-basics.htm, 2012.

[2] Film Alley, "What Is Single CCD and 3-CCD System," http://www.filmalley.com/articles/1ccd%20vs%203ccd/, 2004.

[3] G. Ward. "Fast, robust image registration for compositing high dynamic range photographs from handheld exposures," Journal of Graphics Tools, vol. 8, no. 2, pp. 17-30, 2004.

[4] K. Jacobs, C. Loscos, and G.Ward. "Automatic high-dynamic range image generation for dynamic scenes," IEEE Computer Graphics and Applications, vol. 28, no. 2, pp. 84–93, 2008.

[5] M. Aggarwal and N. Ahuja. "Split Aperture Imaging for High Dynamic Range," International Journal of Computer Vision, vol. 58, pp. 7-17, 2004.

[6] M. D. Grossberg and S. K. Nayer. "Determining the Camera Response from Image: What Is Knowable?," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 11, pp. 1455-1467, 2003.

[7] N. Ostu. "A Threshold Selection Method from Gray-Level Histograms," IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, pp. 62-66, Jan. 1979.

[8] P. E. Debevec and J. Malik, "Recovering High Dynamic Range Radiance Maps from Photographs," Proceeding of ACM SIGGRAPH, Los Angeles, pp. 369-378, 1997.

[9] P. Burt and T. Adelson, "The Laplacian Pyramid as a Compact Image Code," IEEE Transaction on Communication, vol. 31, no. 4, pp. 532–540, Apr. 1983.

[10] S. Paris and F. Durand. "A Fast Approximation of The Bilateral Filter Using a Signal Processing Approach," ECCV, 2006.

[11] S. P. Tang, "Image Sensors (CCD and CMOS)," http://kcs.kcjh.ptc.edu.tw/~spt/computer/digital-image/CCD-CMOS.htm, 2012.

[12] Shree K. Nayar, "Computer Vision Laboratory," http://www1.cs.columbia.edu/CAVE/projects/rad_cal/images/dorfDiagram.jpg

[13] T. Mertens, J. Kautz, and F. V. Reeth, "Exposure Fusion," Conputer Graphics Forum, vol. 28, no. 1, pp. 161-171, 2009.

[14] W. C. Kao, C. C. Hsu, L. Y. Chen, C. C. Kao, and S. H. Chen, "Integrating Image Fusion and Motion Stabilization for Capturing Still Images in High Dynamic Range Scenes," IEEE Transaction Consumer Electronics, vol. 52, no. 3, pp. 735-741, 2006.

[15] W. Zhang and W. K, Cham, "Gradient-Directed Composition of Multi-Exposure Images," Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, pp. 530-536, 2010.

[16] Wikipedia, "Dynamic Range," http://en.wikipedia.org/wiki/Dynamic_range, 2012.

[17] Wikipedia, "Color Filter Array," http://en.wikipedia.org/wiki/Color_filter_array, 2012.

[18] Wikipedia, "Exposure (Photography)," http://en.wikipedia.org/wiki/Exposure_(photography), 2012.

[19] Wikipedia, "Aperture," http://en.wikipedia.org/wiki/Aperture, 2012.