HIGH DYNAMIC RANGE IMAGE ACQUISITION USING FLASH IMAGE

Ryo Matsuoka, Tatsuya Baba, Masahiro Okuda*  
Univ. of Kitakyushu,  
Faculty of Environmental Engineering, JAPAN

Keiichiro Shirai  
Shinshu University  
Faculty of Engineering, JAPAN

ABSTRACT
We propose a denoising technique using multiple exposure image integration. When acquiring a dark scene, the detail of the dark area is often deteriorated by sensor noise. For a high dynamic range image acquisition, denoising in dark areas is a critical issue, since the dark area is, in general, enhanced by a tone-mapping and the noise is made more visible when displaying it on an output devise. In our method, a flash image is utilized as well as a no-flash multiple exposure images to further reduce the noise. Multiple exposure integration is performed in a wavelet domain, where noise removal is achieved by the wavelet-shrinkage for multiple exposures. Our method works well especially for noise in shadows. We show the validity of the proposed algorithm by simulating the method with some actual noisy images.

Index Terms— Image denoising, Wavelet transforms, weight function, HDRI, color-line

1. INTRODUCTION
The human visual system can capture a wide dynamic range of irradiance, while the dynamic range of CCD or CMOS sensors does not cover the perceptual range of real scenes. It is important in many applications to catch a wide range of irradiance of a natural scene and preserve the irradiance value in each pixel. In the application of CG, a high dynamic range image (HDRI) is widely used for high quality rendering with image based lighting [1]. In addition, recently it has been applied to the surveillance camera, the in-vehicle camera and the photographic development with high resolution, and so on.

In general, the HDRI is generated by combining some photographs taken with multiple exposure settings [2]-[5]. To get a high dynamic range, we should have several photographs with short to long exposures. As usual, the dynamic range is defined by the ratio of darkest and brightest pixel intensities of an image, where the darkest point is usually defined as the lowest value in a range that is not dominated by noise. Thus the image denoising is often required to acquire the high dynamic range. Although the process of multiple exposure integration inherently has a property to reduce the noise [6], it is often insufficient and noises appear especially in a dark area. When taking photos with a hand held camera in a dark lighting condition, a high ISO setting is required to restore the dark area without blurring artifacts, which yields noisy images and then brings down the dynamic range.

In the last decade, to improve the noise due to a high ISO setting, many techniques [7]-[8] have been proposed based on a flash image.

2. PROPOSED METHOD

2.1. Outline
Fig. 1 shows the procedure of our restoration algorithm. Our method mainly consists of two steps: (I) Restoration of high exposure Image by flash image, and (II) image integration in the wavelet domain. Unlike the conventional multiple exposure integration, we use a flash image as well as no-flash multiple exposure images. The flash image is utilized to restore dark (low irradiance) regions in a high exposure image, which is described in Sec.2.2 in detail. In the second step, we integrate the multiple exposure images that includes the restored high exposure image. To integrate the multiple exposures, the images are converted to radiance domain by compensating the non-linearity of a sensor to linearize the response. A weighting function, which is designed to reduce the error, is multiplied to the images. Then, denoising procedure is applied to them. In our method, two types of the wavelet shrinkage is performed, which is explained in Sec.2.4.

2.2. Restoration of High Exposure Image by Flash Image
When acquiring high dynamic range images, one often suffers from sensor noise that appears especially in a dark scene. Since a dark area is often enhanced by a tone-mapping operator, the noise in the shadows is made more visible. Thus the denoising in shadows is a critical problem. In our framework, the dark area of the scene
is mainly restored by the high exposure image. Our aim here is to
denoise the high exposure image by using the flash image. The pro-
cEDURE is shown in the upper part of Fig. 1.

For denoising, our method utilizes the property of the local color
linearity [8],[9], that is, the RGB distribution of a local window can
be approximated as a single line in the RGB space. Figure 2 shows
a local region of a flash and no-flash image (a), (b) and their color
distributions (c). We can see from the figure that the two color dis-
tributions of the no-flash image can be approximated by the affine
transform of the flash image. Thus the aim of the restoration here is

\[ p_i = A_i f_i + b_i, \]

where \( A \) and \( b \) are \( 3 \times 3 \) transform matrix and \( 3 \times 1 \) shift vector,
\( f \) and \( p \) are \( 3 \times 1 \) RGB vectors of the flash and high exposure (i.e.
no-flash) images, respectively. The subscript \( i \) is a pixel index.

The RGB color of the flash image is transformed by the matrix
\( A \). The affine transform \( A \) and \( b \) are determined by minimizing the
cost function:

\[
E = \sum_i \sum_{j \in N(i)} \left\{ \omega_{i,j} \cdot \rho \left( A_i f_j + b_i - p_j \right) + \lambda \| b_i \|^2 + \epsilon \| A_i \|^2 \right\},
\]

where \( \omega_{i,j} \) is a bilateral weighting function [8], and \( \rho \) is a robust
function. We found that the luminance difference between the flash
and no-flash images can be compensated by scaling rather than shift-
ing. Therefore, in our case, a large shift vector sometimes adversely
affects the results. Thus we added the penalty term \( \lambda \| b_i \|^2 \) to the
cost.

If one chooses \( \rho(x) = \sqrt{x^2 + \epsilon} \), \( \omega_{i,j} = 1 \) and \( \lambda = 0 \), Eq.(1)
coincides with the formulation of the guided filter [10]. In our case,
we use \( \rho(x) = (-\sigma_0^2/2) \exp(-|x|^2/\sigma_0^2) \) to reduce the effect of outliers. In practice, we minimize (1) by using the IRLS (Iterative
Reweighted Least Squares) method (for detail, see [8]). Once the
optimal transform, which consists of \( A \) and \( b \), is found, a restored
image is calculated by

\[
\hat{p}_i = C_i W_i, \quad \left\{ \begin{array}{c} C_i = \sum_j \omega_{i,j} A_j f_j + b_j \\ W_i = \sum_j \omega_{i,j} \end{array} \right. \]

We call this procedure the LCDP (Local Color Distribution Projec-
tion).

Fig. 2. Linearity of color distributions in corresponding local
regions of no-flash and flash images.

2.3. Alpha map

For the region where the flash fully reaches, the proposed restoration
method in the previous section is used, while we apply the conven-
tional bilateral smoothing [16] for the other region. To discriminate
the two regions, we generate an alpha map, which has large values
in the region where the flash sufficiently reaches, and has small val-
ues in the no-flash regions. Based on the alpha map, we merge the
images. The step is described as follows.

We use the flash image \( f \) and the high exposure image \( p \). To
generate the flash map, we estimate the irradiance of the images:

\[
R_p = (a_p g^{-1}(p)) / t_p, \]

where \( a_p \) and \( t_p \) are the gain of the ISO sensitivity and the shutter
speed of \( p \). \( g^{-1} \) is the camera response curve. \( R_p \) is the estimated
irradiance. The irradiance of the flash image \( R_f \) is also calculated.
Then, we find the alpha map \( \Omega \):

\[
\Omega = M_p (R_f - R_p),
\]

where \( M_p \) is a masking function that has 0 in saturated pixels either
of the flash or high exposure images. The pixel values of \( \Omega \) are
normalized to the range of \([0, 1]\). Additionally we apply smoothing
filter and gamma correction to Ω in order to make it robust to noises. Based on Ω, we merge the images:

\[ p' = Ω p - (1 - Ω) \hat{p}. \]  

(5)

where \( \hat{p} \) is the image smoothed by the bilateral filter.

### 2.4. Denoising by Shrinkage

After the high exposure image is restored, we integrate the multiple exposure images. Before the integration, difference (noise) between the images is relieved in an inter-image shrinkage as follows. First we divide the images to \( L_m \) and \( H_m \) using (6) and (7).

\[ L_m(n) = \frac{M_m(n)p_m(n) + M_{m+1}(n+1)p_m(n+1)}{2}, \]  

(6)

\[ H_m(n) = \frac{M_m(n)p_m(n) - M_{m+1}(n+1)p_m(n+1)}{2}, \]  

(7)

where \( n = 1, \cdots, N - 1 \), and \( N \) is the number of multiple exposures. We assume that lower value of \( n \) indicate lower exposure. \( p_m(n) \) is the \( n \)-channel of the \( n \)-th exposure, and \( M_m(n) \) is a masking function that has 0 if the original image is 0 (under exposure) or 1 (over exposure), and 1 otherwise. After the decomposition, we apply hard thresholding to \( H_m \) in (7).

\[ H'_m(n) = \begin{cases} 0, & \text{if } H_m < \text{threshold} \\ H_m(n), & \text{otherwise} \end{cases} \]  

(8)

This determination of thresholding is performed only for the green channel, and if \( H_m < \text{threshold} \) at a pixel, the coefficients of the three channels are set to zero, since if the thresholding is performed for the three channels independently, color balance is often destroyed. After the thresholding, we reconstruct the images by

\[ \hat{p}'_m(n) = L_m(n) + H'_m(n). \]  

(9)

In the second step, we integrate the images in the wavelet domain. In general, a low exposure yields a dark image and its low pixel values with noises are enhanced by the camera response curve and the tone-mapping operator. Summing up images may remove the noise to some extent [6]. However the simple integration does not always remove noises adequately. In this section we try to remove the noise by shrinkage for multiple exposure image.

Before applying the wavelet transform, we weight the image by

\[ \hat{p}'_m(n) = w_m(n) \cdot p_m(n), \]  

(10)

where we use the weight function of [18], which is expressed by

\[ w(i) = m_j(i), \]  

(11)

where \( g' \) is the derivative of \( g \), and \( m_j(i) \) is a masking function that has 0 in the saturation pixel. Using this weight function, one can control the suppression effect of the sensor noise and quantization error by varying the parameters. Here \( b_1 = 0.001, b_2 = 0.01, b_3 = 0.001 \) are used and fixed for all images. This paper aims at removal of the dark area’s noise, and so the weight function by the above parameter setup reduces a sensor noise.

Next, we employ the wavelet shrinkage to our integration problem to remove the noise. The weighted images, \( \hat{p}'_m \) in (10) is converted by the Haar-based shift invariant wavelet transform (non-subsampling). In the wavelet conversion, a low pass/high pass filter pair is performed to it in the horizontal and vertical direction respectively, and four subbands (\( LL, LH, HL, \) and \( HH \)) are produced, and then repeatedly the \( LL \) band is transformed to four bands.

Here we introduce the wavelet shrinkage for the multiple exposure integration. We modify the shrinkage [11], [12] to a multiple exposure integration. The problem is defined to minimize the cost function:

\[ \min_h E(h) = |h|^0 + \lambda \sum_{n=1}^{N} (h - h(n))^2, \]  

(12)

where \( h(n) \) is an input wavelet coefficient of the weighted \( n \)-th exposure image and \( h \) is an output wavelet coefficient formed by the high dynamic range. Additionally, \( \lambda \) is the parameter to control noise removal. By differentiating \( E(h) \) with respect to \( h \) and setting it to 0, we derive the optimal wavelet coefficients:

\[ h^* = \begin{cases} 0, & \text{if } 1 - \lambda \left( \frac{1}{N} \sum_n h(n) \right)^2 < 0 \\ \frac{1}{N} \sum_n h(n), & \text{otherwise} \end{cases} \]  

(13)

The lowest band is simply merged by the weighted mean. Note that roles of (13) are not only the shrinkage based denoising but also multiple exposure fusion in the wavelet domain in which it is different from the conventional shrinkage.

### 3. EXPERIMENTAL RESULTS

To evaluate the method quantitatively, we artificially created ground truth as follows. First we prepare three images with different exposures, each of which is obtained by averaging 15 photographs, and we select randomly one of them for each as input images. Then the three images are integrated by conventional method and our method. In our method, we have a flash image as an input as well.

First we compare the results of high exposure image restoration described in Sec.2.2 to confirm the validity of the LCDP. Fig.3 shows four images of high exposure image: the ground truth, noisy image, our result, and the result of the conventional method [7]. Our method and [7] use the flash image as a second input. One can see from the figure that denoising capability of the two methods are almost same, but our method preserves more detail than [7].

Next we show the results of the multiple exposure integration. The results of our method, the conventional integration (simple weighted sum of the images), BM3D [15] and Bilateral filter [16] are shown in Fig.4. The image at the top is the tone-mapped version of a ground truth HDR image. Table 1 indicates the qualitative results (PSNR) of three images (a)-(c). We enhance the detail of the images by the three methods: Reinhard’s method [1], Jinno et al.’s method [17], and tone-map function (tonemap.m) in MATLAB image processing toolbox, and calculate the PSNR of the tone-mapped images. The results show that our method outperforms the others.

### 4. CONCLUSION

We proposed a method for the HDR acquisition with detail restoration, especially the extremely dark scene. The proposed method can significantly restore the detail of dark area compared with the conventional integration method and some denoising technique.
Fig. 3. Result: (upper left) Full frame of ground truth, (upper right) Input, (lower left) Our method, (lower right) Petschnigg et al. [7]

Table 1. Comparison in PSNR for images (a)-(c)

<table>
<thead>
<tr>
<th>Tone-mapping</th>
<th>[1]</th>
<th>[17]</th>
<th>MAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>conventional method</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td>28.31</td>
<td>21.24</td>
<td>22.30</td>
</tr>
<tr>
<td>(b)</td>
<td>27.87</td>
<td>20.21</td>
<td>22.37</td>
</tr>
<tr>
<td>(c)</td>
<td>30.17</td>
<td>19.56</td>
<td>23.38</td>
</tr>
<tr>
<td>our method</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td>30.14</td>
<td>24.34</td>
<td>27.97</td>
</tr>
<tr>
<td>(b)</td>
<td>29.93</td>
<td>24.02</td>
<td>25.70</td>
</tr>
<tr>
<td>(c)</td>
<td>32.15</td>
<td>22.62</td>
<td>23.60</td>
</tr>
<tr>
<td>BM3D</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td>28.46</td>
<td>21.81</td>
<td>22.33</td>
</tr>
<tr>
<td>(b)</td>
<td>27.96</td>
<td>20.80</td>
<td>22.08</td>
</tr>
<tr>
<td>(c)</td>
<td>28.23</td>
<td>17.71</td>
<td>23.54</td>
</tr>
<tr>
<td>Bilateral Filter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td>28.21</td>
<td>21.70</td>
<td>22.38</td>
</tr>
<tr>
<td>(b)</td>
<td>27.34</td>
<td>19.94</td>
<td>22.01</td>
</tr>
<tr>
<td>(c)</td>
<td>27.78</td>
<td>17.87</td>
<td>23.22</td>
</tr>
</tbody>
</table>

Fig. 4. Result: (top left) Flash image, (top middle) Ground truth, (top right) Noisy image, (lower left) our method, (lower middle) BM3D, (lower right) BRF
5. REFERENCES


