Recursive HDR Image Generation from Differently Exposed Images

Andrey Vavilin, Kang-Hyun Jo
Graduate School of Electrical Engineering
University of Ulsan, Ulsan, Korea
andrey.korea@gmail.com, jkh2008@ulsan.islab.ac.kr

Abstract

Dynamic range limitation of CCD-cameras may cause distortions and data loses in images. Such limitations are strongly effect to the further image processing. In this paper simple tone reproduction algorithm is presented. The method is based on analyzing the level of useful information in image subregions and choosing best exposure for each subregion. Local region analysis is based on such properties like average contrast, intensity density, entropy etc. This technique allows recovering details in overexposed and underexposed parts of image. Resulted HDR image simplifies further image processing in applications which deals with object detection and recognition.

Keywords: *HDR, image enhancement, local feature analysis.*

1. INTRODUCTION

High dynamic range (HDR) imaging techniques are designed to produce images faithfully depict the full visual dynamics of real world scenes. Most of algorithms are based on reconstructing scene luminance using several differently exposed images. All algorithms could be classified into two main categories: spatially uniform (non-local) and spatially varying (local).

Spatially varying approach [1,3,4,6] is based on the fact that humans are capable of viewing high contrasts scenes thanks to the local control sensitivity in the retina. This suggests that a position-dependent scale factor might reduce scene contrasts acceptably and allow displaying them on a low dynamic range device. This approach converts the original scene or real-world intensities to the displayed image intensities, using a position-dependent multiplying term

Non-local approaches [2,5] do not imitate local adaptation. Instead, almost all image synthesis, recording, and display processes use an implicit normalizing step to map the original scene intensities to the target display intensities without disturbing any scene contrasts that fall within the range of the display device. This normalizing consists of a single constant multiplier. Image normalizing has two important properties: it preserves all reproducible scene contrasts and it discards the intensities of the original scene or image. Contrast, the ratio of any two intensities, is not changed if the same multiplier scales both intensities. Normalizing implicitly assumes that scaling does not change the appearance, as if all the perceptually important information were carried by the contrasts alone, but scaling display intensities can strongly affect a viewer's estimates of scene contrasts and intensities. Normalizing fails to capture dramatic appearance changes at the extremes of lighting, such as gradual loss of color vision, changes in acuity, and changes in contrast sensitivity. Proposed algorithm is based on combining parts of differently exposed images in order to select best exposure for each region instead of recovering the real physical luminance for each pixel.

2. ALGORITHM DESCRIPTION

The input of algorithm is a set of differently exposed images $I^{(1)},...,I^{(N)}$. As the result the program generates single image which contains information from each of input images. Main scheme of proposed method is shown at Figure.1.
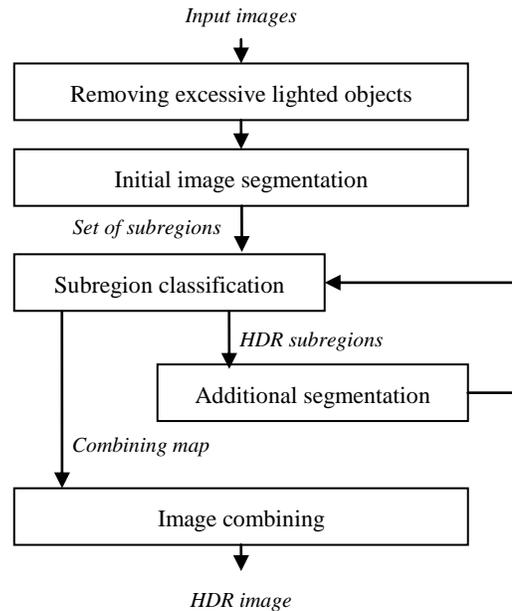

Figure 1: Main algorithm scheme.

2.1 Local properties estimation

The key point of proposed algorithm is region classification. In this paper we consider two types of regions, namely LDR and HDR. LDR regions are regions with dynamic range lower than limited by image properties. We consider dynamic range to be limited by 8-bit jpeg format as the most widely used in consumer cameras. In other words, this region could be represented by one exposure. HDR regions are regions in which some information was lost due to limitations of camera. In this paper we tried to decompose these regions into subregions and apply algorithm recursively. It allows more precise detail recovering shadows and highlights. For the purpose of region classification we compute its local such as mean intensity (1), intensity deviation (2), entropy (3) and the level of details (4). Suppose k-th region has width kw, height kh and top left corner has coordinates (kx, ky). For the i-th exposure, local properties of k-th region can be expressed by equations (1)~(4).

$$\mu_k^{(i)} = \frac{1}{A_k} \sum_{x=kx}^{kx+kw-1} \sum_{y=ky}^{ky+kh-1} I^{(i)}(x, y) \tag{1}$$

where $I^{(i)}(x,y)$ is pixel intensity and $A_k = kw \cdot kh$ is the area of k-th region.

$$\sigma_k^{(i)} = \sqrt{\frac{1}{A_k} \sum_{x=kx}^{kx+kw-1} \sum_{y=ky}^{ky+kh-1} \left(I^{(i)}(x,y) - \mu_k^{(i)}\right)^2} \qquad (2)$$

$$H_k^{(i)} = \sum_{j=0}^{255} p[j] \log_2(p[j]) \qquad (3)$$

where $p[j] = \dfrac{1}{A_k} \sum_{I_k^{(i)}=j} 1$

$$D_k^{(i)} = \frac{1}{255} \sum_{x=kx}^{kx+kw-1} \sum_{y=ky}^{ky+kh-1} \max\left(\Delta I_x^{(i)}(x,y), \Delta I_y^{(i)}(x,y)\right) \qquad (4)$$

where $\Delta I_x^{(i)}$ is defined as $\Delta I_x^{(i)} = \left|I^{(i)}(x+1,y) - I^{(i)}(x,y)\right|$ and $\Delta I_y^{(i)} = \left|I^{(i)}(x,y+1) - I^{(i)}(x,y)\right|$.

Using these properties we can formalize the "usefulness" of exposure i for k-th region as:

$$U(i,k) = \overline{H}_k^{(i)} + \overline{D}_k^{(i)} + \overline{\sigma}_k^{(i)} \qquad (5)$$

where $\overline{H}_k^{(i)}, \overline{D}_k^{(i)}, \overline{\sigma}_k^{(i)}$ are normalized entropy, level of details and intensity deviation.

$$\overline{H}_k^{(i)} = H_k^{(i)} \Big/ \max_{j=1..N} H_k^{(j)} \qquad (6)$$

$$\overline{D}_k^{(i)} = D_k^{(i)} \Big/ \max_{j=1..N} D_k^{(j)} \qquad (7)$$

$$\overline{\sigma}_k^{(i)} = \sigma_k^{(i)} \Big/ \max_{j=1..N} \sigma_k^{(j)} \qquad (8)$$

The best exposure for k-th region is

$$R(k) = \arg\left(\max_{j=1..N} U(j,k)\right) \qquad (9)$$

To classify k-th region as HDR or LDR we used 2 assumptions. Region is HDR if:

1. it contains overexposed or underexposed regions;

2. from the set of differently exposed regions $I_k^{(1)}, ..., I_k^{(N)}$ more than one will have high value of detail evaluation function (5)

Overexposed and underexposed regions are regions in which pixel intensities are close to 255 and 0 respectively. Hence we can say what region is HDR if

$$\begin{cases} \mu_k^{(i)} + \sigma_k^{(i)} \geq 200 \ or \ \mu_k^{(i)} - \sigma_k^{(i)} \leq 55 \\ \left|\left\{ j \mid U(j,k) > 0.8 \max_p U(p,k) \right\}\right| > 1 \end{cases} \qquad (10)$$

Equations (9) and (10) are sensitive to very dark and very bright objects in scene. In other words, if there will be an object which color is close black or white in all exposures it may cause equations (9) and (10) produce wrong result. In real scene this situation may happen if the real color of the object is black or white or in if we do not have exposure which can reflect the real object color. To overcome this problem we remove all such pixels

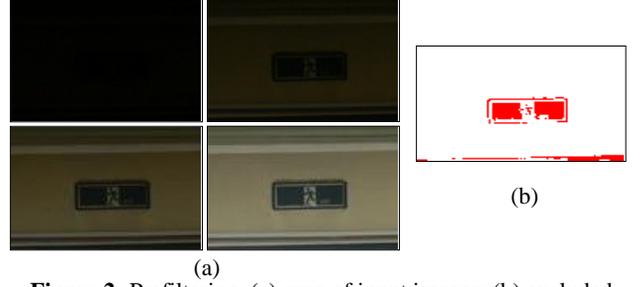from computational process. Example of that filtering is shown in Figure 2.

(a)

(b)

Figure 2: Prefiltering. (a) crop of input images; (b) excluded pixels.

The filtering process is based on following definitions:

Pixel (x,y) is black if $I^{(i)}(x,y) < 30 \ \forall i \in [1,N]$ (11)

Pixel (x,y) is white if $I^{(i)}(x,y) > 225 \ \forall i \in [1,N]$ (12)

Pixels which satisfy condition (11) or (12) are excluded before segmentation process in order to make region classification works correctly. However, these pixels will be used during image combining process.

2.2 Image segmentation

Initially image is segmented into 4 rectangular regions (2 regions in both vertical and horizontal directions) of the same size and apply classification algorithm which was described in the previous section. After this stage all regions classified as HDR are recursively segmented into 4 subregions so that width and height of new subregions two times smaller than original. Process stops if there are no HDR regions in image or if the size of region is smaller than 50 pixels in any dimension. This technique allows more precise reconstruction of image without local contrast decreasing. Example of segmentation for four differently exposed images is shown at Figure 3.
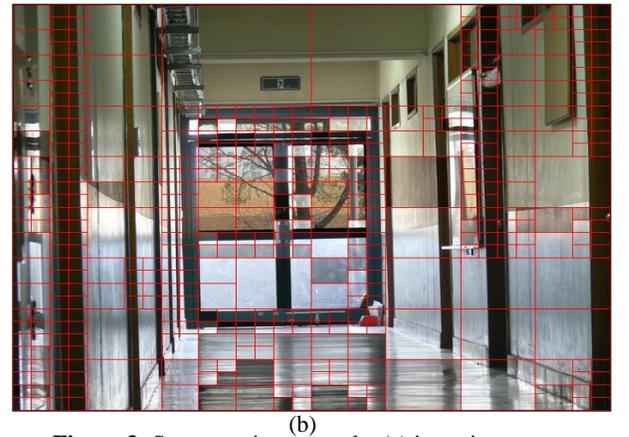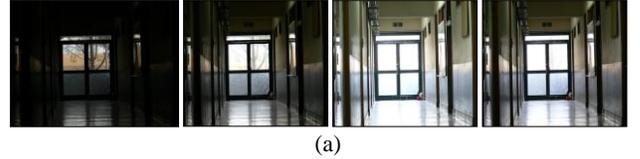

(a)

(b)

Figure 3: Segmentation example. (a) input images; (b) segmentation results.

2.3 Combining HDR image

The result of image segmentation is a map that shows regions and corresponding exposures. Direct combination based on this map will give us a mosaic image like the one shown in Figure 3(b). This picture indicates very sharp transitions which arise at the borders of regions. The sharpness depends on the difference between the exposures selected for connected regions. To eliminate this effect we applied Gaussian blending function. Value of each pixel of output image is calculated as a superposition of exposures selected for this region and exposures selected for the neighboring regions. Weights of exposures are selected according to the blending function. The weight of exposure selected for k-th region at point (x, y) is defined as:

$$B^k(x, y) = G^k(x, y) \bigg/ \sum_p G^p(x, y) \qquad (13)$$

where $G^k(x, y)$ is a Gaussian function with maximum at the center of k-th region

$$G^k(x, y) = \exp\left\{ -\left(\frac{(x - kc)^2}{2\sigma_x^2} + \frac{+(y - kc)^2}{2\sigma_y^2} \right) \right\} \qquad (14)$$

where (kc, kc) is the center of the k-th region, and σ_x, σ_y stands for the standard deviation of 2D Gaussian function in each direction respectively. In proposed framework we selected these values depending from the size of corresponding region.

$$\begin{aligned} \sigma_x &= \sigma_x(k) = 1.5kw \\ \sigma_y &= \sigma_y(k) = 1.5ky \end{aligned} \qquad (15)$$

So, the influence of exposure selected for the k-th region for the (x, y) pixel depends on the distance between the center (kx, ky) of the region and the adjacent pixel and from the region's size. This idea is presented in Figure 4.
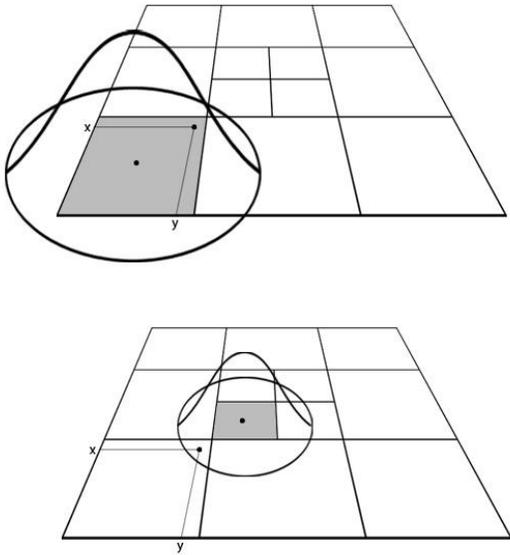

Figure 4 Illustration of regions convolved by Gaussian kernel functions in different scale.

Applying function defined by Equation (13) we formalize the output of algorithm as

$$I_{out}(x, y) = \sum_k B^k(x, y) I_{R(k)}(x, y) \qquad (16)$$

2.4 Optimizations

To make the proposed algorithm work faster several optimizations were done. First, we reduce the amount of data for computing local image properties for (9) by using each second row and column of image in computations. Another point of optimization is calculation of Gaussian for the blending function (Equation (14)). Instead of permanent recalculation lookup tables could be used. As long as Gaussian is separable function its 2D response could be computed as a multiplication of corresponding 1D functions. Hence, two lookup tables are needed for each region size. Gaussian function is a symmetrical function; hence size of each table can be reduced by using absolute values. Suppose T^w and T^h are lookup tables for the region with size $w \times h$. In this sense we could reformulate equation (14) as

$$G^k(x, y) = L(k, x, y) \cdot T^{kw}[|kx - x|] \cdot T^{kh}[|ky - y|] \qquad (17)$$

where $L(k, x)$ is a range limitation function defined as

$$L(k, x) = \begin{cases} 1, & \text{if } |kx - x| < sizeof\left(T^{kw}\right) \text{ and} \\ & |ky - y| < sizeof\left(T^{kh}\right) \\ 0, & \text{otherwise} \end{cases} \qquad (18)$$

Total number of tables depends on maximum depth of segmentation d which could be found using the limitation for the minimum region size.

$$d = \min\left(\left\lfloor \log_2\left(\frac{w_{image}}{w_{\min}} \right) \right\rfloor, \left\lfloor \log_2\left(\frac{h_{image}}{h_{\min}} \right) \right\rfloor \right) \qquad (19)$$

where w_{image}, h_{image} are width and height of the image and $w_{\min}$, $h_{\min}$ are minimal width and height of the segments respectively. The number of tables is $NT = 2d$.

Number of elements in table for segment with width w can be estimated as

$$sizeof\left(T^w\right) = 3\sigma \overset{(15)}{=} 4.5w \qquad (20)$$

Applying this optimization strategy is a tradeoff between computation speed and memory usage. Lookup tables can reduce computational time but it requires a lot of memory for storing data structures. Experiments show that in common case initial segmentation into 4 regions is ineffective especially for the big images. The more effective approach is based on defining maximum segment size and start with segmenting into

$$d_{init} = \min\left(\left\lceil \log_2\left(\frac{w_{image}}{w_{\max}} \right), \log_2\left(\frac{h_{image}}{h_{\max}} \right) \right\rceil \right) \qquad (21)$$

This will reduce the number of lookup tables and their sizes.

3. EXPERIMENTAL RESULTS

All experiments were done on Pentium-IV with 1 GB RAM under Borland Builder environment.

For our experiments we used images taken by Nikon D50 under different lighting conditions. All images were taken in 8bit jpeg quality (instead of 12bit raw format). Resolution of pictures varies from 800x600 up to 2000x3000. To obtain images with different exposure the camera was fixed on a tripod and its aperture was set up for a constant value while the shutter speed was changing. The number of different exposures from one scene to another varies from 3 to 9. All images could be separated into 3 groups depending on dynamic range of the scene. First group of images contains scenes which require 5 or more exposures for correct reconstruction. Second group contains scenes which required 3 exposures. Scenes included in last group can be represented by only 1 exposure. Experimental results are shown in Table1. Values in the table show each ration of correctly reconstructed and the number of images. By correct image reconstruction we mean suitable selection of exposures for all region and combining the image without color inconsistency.

Table 1: Experimental results.

# of exposures	Dynamic range		
	Low	**Medium**	**High**
3	15/15	10/10	-
4-5	10/10	10/10	9/9
6-9	-	5	11/11

The effectiveness of algorithm is illustrated by Figures 5.
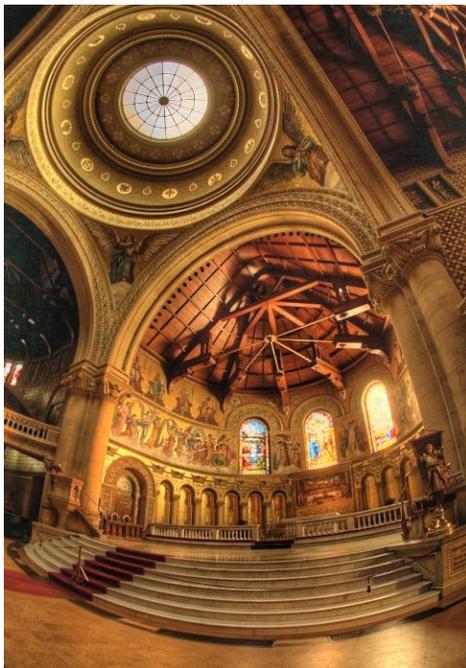

Figure 5 Image generated from sixteen photographs of a church taken at 1-stop increments

Figure 6: Image processing example. (Input images for this scene are shown in Figure 3(a)).

4. CONCLUSIONS

Proposed method allows correct representation of details in overexposed and underexposed areas of image which could be lost due to the high dynamic range of the scene. Image is reconstructed without decreasing of local contrast which is a specific property of global-based techniques. Furthermore, no additional noise is produced. Proposed method is based on several local properties such as entropy, deviation and local contrast which make it less sensitive to the input image noise. Using regions of different size allows more precise detail recovering on the borders of regions with high contrast changing. Proposed method avoids camera response curve calibration and needs no information about exposure for obtaining HDR image. This could be useful as long as we do not have such information in the common case. Average computational time for five input images with resolution 1600x1200 on Pentium-IV is around 15 seconds.

5. REFERENCES

[1] Chiu K., Herf M., Shirley P., Swamy S., Wang C., Zimmerman K. "Spatially Nonunifrom Scaling Functions for High Contrast Images". Proceedings of Graphics Interface 93, pp.245-254, 1993.

[2] Ferwerda J., Pattanaik S.N., Shirley P. and Greenberg D.P. "A Model of Visual Adaptation for Realistic Image Synthesis", Proceedings of SIGGRAPH 1996, pp. 249-258.

[3] Guoping Qiu, Jian Guan, Jian Duan and Min Chen, "Tone Mapping for HDR Image using Optimization – A New Closed Form Solution", Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06).

[4] Pattanaik S.N., Ferwerda J, Fairchild M.D., and Greenberg D.P. "A Multiscale Model of Adaptation and Spatial Vision for Realistic Image Display", Proceedings of SIGGRAPH 1998, pp. 287-298.

[5] Pattanaik S.N., Tumblin J.E.,Yee H. and Greenberg D.P. "Time-Dependent Visual Adaptation for Realistic Real-Time Image Display",Proceedings of SIGGRAPH 2000, pp.47-54.

[6] Rahman Z., D. J. Jobson, and G. A. Woodell. "Multi-scale retinex for colour image enhancement", Proceedings of International Conference on Image Processing, vol.3, pp. 1003-006, Lausanne, Switzerland 16-19 September 1996.