TONE-MAPPING FOR AN HDR SURVEILLANCE SYSTEM USING SIFT FEATURES

Takao Jinno, Shigeru Kuriyama

Toyohashi University of Technology
Department of Computer Science and Engineering

Masahiro Okuda

The University of Kitakyushu
Faculty of Environmental Engineering

ABSTRACT

This paper proposes a tone-mapping method for High Dynamic Range (HDR) surveillance systems, which can improve the accuracy of SIFT object tracking in the systems. HDR surveillance systems can store minute details for the whole radiance range without over- or under-exposure. Although the minute details can be able to increase the accuracy of object detection and tracking, we have some issues of using them to existing SIFT. HDR contents should be tone-mapped before calculating SIFT feature since SIFT is designed for common Low Dynamic Range (LDR) contents. Most of conventional tone-mapping methods lose the minute details through the tone-mapping process, or degrade the accuracy of tracking since they also enhances the noise and invalid features. We propose the tone-mapping method for HDR surveillance system using SIFT features, which can enhance the minute details without invalid features.

***Index Terms*—** HDR surveillance system, object tracking, SIFT feature, tone-mapping

1. INTRODUCTION

Robust and accurate object tracking in high contrast scenes is one of the most important characteristics of surveillance system. SIFT feature [1] is well-known feature and it can be used as robust key features for object tracking. This approach is employed by many object tracking methods since it is robust to scaling, rotation, and changes in the lighting. HDR images can store minute details for the whole radiance range even if in high contrast scenes, where the minute details is features consist of very weak edges. However, it cannot directly apply to conventional object tracking using SIFT feature. HDR images have wide dynamic range, high bit-depth, and linear pixel values to scene radiance. Since SIFT is designed for common LDR images that have nonlinear pixel values, it is difficult to direct use SIFT features of HDR images for object tracking. Therefore, to achieve robust and accurate SIFT object tracking, HDR images should be tone-mapped adequately before input them into SIFT. Using tone-mapped LDR images instead of HDR images is also able to make a transmission quantity and a computational cost low.

The present paper proposes a tone-mapping method to improve accuracy of existing SIFT object tracking in HDR surveillance system. Many global and local tone-mapping methods have been proposed [2]-[6], and our previous study proposed high contrast tone-mapping using multi-scale contrast enhancement to enhance (MSCE) the minute details of HDR images [7]. Although the minute details preserved in HDR images are important features in untrackable case that moving objects have only some thin features, they are often lost through many global tone-mapping methods [4]. Many conventional local tone-mapping methods [5], [6] can enhance and preserve the minute details, but they do not always improve the accuracy of existing SIFT object tracking since most of them aim to improve visual quality or reconstruct human appearance model. If we push aside their aims, enhancing high frequency details that may contain some noise and preserving global contrast caused by scene lighting are not necessary for SIFT object tracking. This paper proposes a tone-mapping based on [7] for SIFT object tracking, and it preferentially enhances valid thin features in multi-scale processing, and moderates scene lighting effects. Our method can track the untrackable objects.

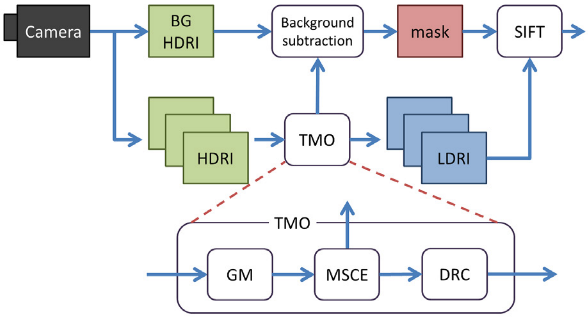
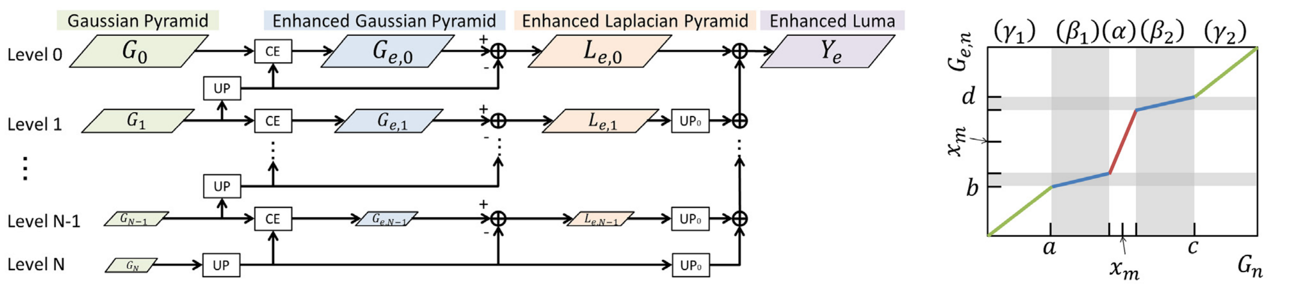


Fig. 1. Our assumed HDR camera surveillance system: (BG HDRI) BackGround HDR Image, (TMO) Tone-Mapping Operator, (GM) Gamma Mapping, (MSCE) Multi-Scale Contrast Enhancement, (DRC) Dynamic Range Compression.

2. PROPOSED METHOD

Figure 1 shows overview of our assumed HDR surveillance system. Our tone-mapping method enhances the minute details by using high contrast tone-mapping (Section 2.1). Enhancing details increases SIFT feature points, and then its computational cost is huge. Proposed method estimates an object mask by using background subtraction in Laplacian



(a) The flow of our multi-scale processing (MSCE) : (UP)upsampling, (UP_0)0-th level up-sample operator, (CE)contrast enhancement

(b) The mapping function of our contrast enhancement : $x_m = UP\,(G_{n+1})$

Fig. 2. Multi-scale contrast enhancement

field (Section 2.2). The gradients caused by scene lighting on the moving objects bring the SIFT object tracking to ruin. Our method moderates their scene lighting effects (Section 2.3), and then obtains our tone-mapped LDR images. Using them as input data, existing SIFT object tracking methods can track untrackable objects.

2.1. High contrast tone-mapping

This section explains our tone-mapping method and TMO in Figure 1 shows its flow. This method enhances the minute details before dynamic range compression to avoid losing them. HDR images are linear with respect to the scene radiance, whereas the Human Visual System (HVS) has a nonlinear response. The nonlinearity yields a difference in the contrast magnitude, i.e., the contrast magnitude is larger in the high luminance range than in the low luminance range of HDR images. Our method applies gamma mapping to the original HDR image to mimic the nonlinearity of the HVS. This mapping can remove the nonlinearity in an approximate manner. Multi-scale contrast enhancement (MSCE in Figure 1 and Figure 2(a)) and dynamic range compression (DRC in Figure 1) are performed in the luminance field. Our MSCE builds an N-level Gaussian pyramid $G_0, G_1, ..., G_N$, and then enhances each level of the pyramid by using a mapping function (CE in Figure 2(a) and Figure 2(b)).

The mapping functions at an n-th level varies depending on the up-sampled $(n+1)$-th level of the Gaussian pyramid $UP(G_{n+1})$. It consists of five lines as shown in Figure 2(b). Six parameters h_α, h_β, h_γ, v_α, v_β and v_γ control these five lines.

$$CE(x) = \begin{cases} g_\gamma \cdot (x-a)+b & : \text{if } (\gamma_1) \\ g_\beta \cdot (x-a)+b & : \text{else if } (\beta_1) \\ g_\alpha \cdot (x-x_m)+x_m & : \text{else if } (\alpha) \\ g_\beta \cdot (x-c)+d & : \text{else if } (\beta_2) \\ g_\gamma \cdot (x-c)+d & : \text{else (i.e.}(\gamma_2)) \end{cases}, \quad (1)$$

where $g_\alpha = v_\alpha/h_\alpha$, $g_\beta = v_\beta/h_\beta$, $g_\gamma = v_\gamma/h_\gamma$, $v_{\alpha,\beta} = v_\alpha + v_\beta$, $h_{\alpha,\beta} = h_\alpha + h_\beta$, $a = x_m - h_{\alpha,\beta}$, $b = x_m - v_{\alpha,\beta}$, $c = x_m + h_{\alpha,\beta}$, and $d = x_m + v_{\alpha,\beta}$.

$(\gamma_1)\ :\ x \le x_m - h_{\alpha,\beta}$

$(\beta_1)\ :\ x_m - h_{\alpha,\beta} < x \le x_m - h_\alpha$

$(\alpha)\ :\ x_m - h_\alpha < x \le x_m + h_\alpha$

$(\beta_2)\ :\ x_m + h_\alpha < x \le x_m + h_{\alpha,\beta}$

$(\gamma_2)\ :\ x_m + h_{\alpha,\beta} < x$

Additionally, $x = G_n$ and $x_m = UP(G_{n+1})$ in this paper. This mapping function can enhance the minute details, however it also enhances the noise and uncertain weak features. Their invalid edges tend to exist in only high frequency band. Our high contrast tone-mapping method applies weaker contrast enhancement to small edges in high frequency band than in low frequency band. Our method uses $h_\alpha \cdot (N/n^{0.3})$ instead of h_α in the mapping function, and then moderates effects of enhancing invalid edges.

The next step calculates an enhanced Laplacian pyramid $L_{e,0}, L_{e,1}, ..., L_{e,N-1}$ from an enhanced Gaussian pyramid $G_{e,0}, G_{e,1}, ..., G_{e,N-1}$, before deriving an enhanced luminance map, as follows.

$$Y_e^{(n)} = UP(Y_e^{(n+1)}) + L_{e,n}, \quad (2)$$

where $n = N-2, N-1, ..., 0$, $UP()$ is the upsampling operator, $Y_e^{(N-1)} = G_{e,N-1}$ and $Y_e = Y_e^{(0)}$. The enhanced luminance map Y_e still has a high dynamic range, so its dynamic range needs to be compressed. Our method compresses the dynamic range using the following equation.

$$Y_{LDR} = \frac{Y_e}{1 + UP(G_1)}. \quad (3)$$

Finally, our method colorizes Y_{LDR} based on the original HDR image, before acquiring a tone-mapped image.

2.2. Moving object detection

Preserving and enhancing the minute details increases the number of points at which SIFT calculates features, and then their computational cost becomes huge. This paper refines their points by using an object mask, where our method uses simple background subtraction to detect moving objects. Background subtraction techniques often fail when the moving objects share a similar color or luminance with the

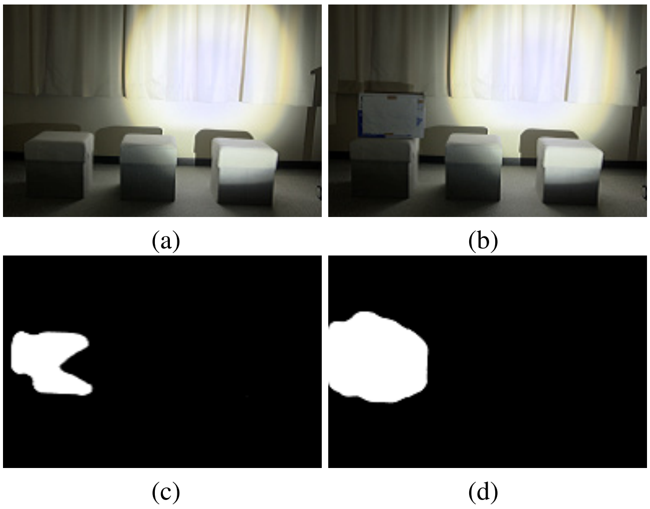


Fig. 3. Estimation results for a moving object: (a) background frame, (b) target frame, (c) conventional method with adapted threshold, (d) our method.

background. Their untracable objects also have some weak features e.g. outline, texture or smooth gradient, however their features are just too weak. Our high contrast tone-mapping (Section 2.1) can enhance them. This paper applies a simple background subtraction method by using enhanced Laplacian pyramid to make the object mask. This approach subtracts Laplacian pyramid of background ($L_{bg,e,n}$) from target one ($L_{e,n}$) at each scale levels, and then sums them, where Laplacian pyramid of target image is acquired in our tone-mapping process.

$$E = \sum_{n=0}^{N-1} \left\{ w_n \cdot UP_0(|L_{e,n} - L_{bg,e,n}|)^{0.1} \right\}, \tag{4}$$

where n is scale level, N is rank of Laplacian pyramid, a multiplier factor 0.1 compresses the dynamic range of each subtraction, and a weight $w_n = (1 + \frac{n}{N-1})$. Subtractions at the higher resolution scales are often affected by noise, thus using the weight w_n to sum the multi-scale subtractions can moderate the effects of noise. This background subtraction approach can abstract object outlines and differences of texture from the subtractions of lower resolution scale and higher resolution scale, respectively. Our method obtains the object mask by thresholding E. In our experiments, we set mean value $\bar{E}$ as the threshold. This mask can reduce the unnecessary feature points.

Figures 3(c) and (d) shows the results with conventional moving object detection method and with ours, respectively. The conventional method [4] uses for tone-mapping and it detects the object area using an existing simple background subtraction technique. [4] is one of the most well-known tone-mapping methods. The conventional method with adapted threshold can detect the boundary of the moving object, but it fails inside a moving object that does not have large edges(Figure 3(c)). Our method can detect the whole area of the moving object (Figure 3(d)).

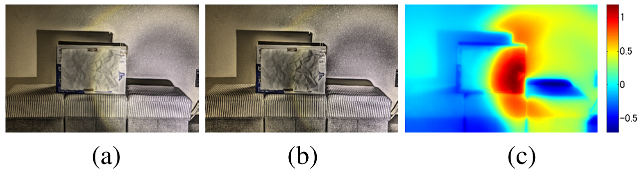


Fig. 4. Lighting effect removal: (a) non-removal, (b) removed, (c) estimated lighting effect Y_c.

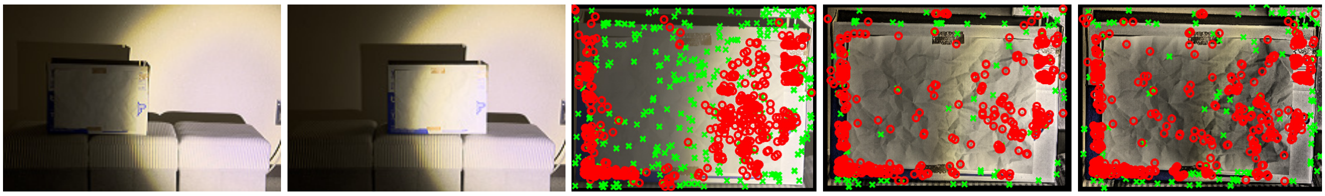
2.3. Lighting effect removal

SIFT object tracking often fails when the lighting conditions on the moving objects change, e.g., Figure 4(a) and (b). In this case, the global scene lighting does not vary whereas the lighting conditions on the objects change. Changing the lighting conditions affects the feature of the object. Thus, it is necessary to remove the lighting effects before calculating the SIFT features. This specific lighting effect is the illumination component, which is calculated by removing the texture and details (i.e., the reflectance component) from the luminance map. It is difficult to calculate the accurate illumination component, which also includes the necessary features composed of small shadows such as textures caused by 3D shapes on surfaces. Therefore, the lighting effect referred to in this paper is the overall or global illumination (i.e., rough illumination) and our method only needs to remove this effect. Our method estimates the lighting effect roughly using the Laplacian pyramid and moderates it before calculating the SIFT features. The high resolution scale of the Laplacian pyramid denotes local contrasts whereas the low resolution scale denotes global contrasts. Our method calculates the lighting effect as follows:

$$Y_c = \sum_{n=0}^{N-1} C_n \cdot UP_0(L_n), \tag{5}$$

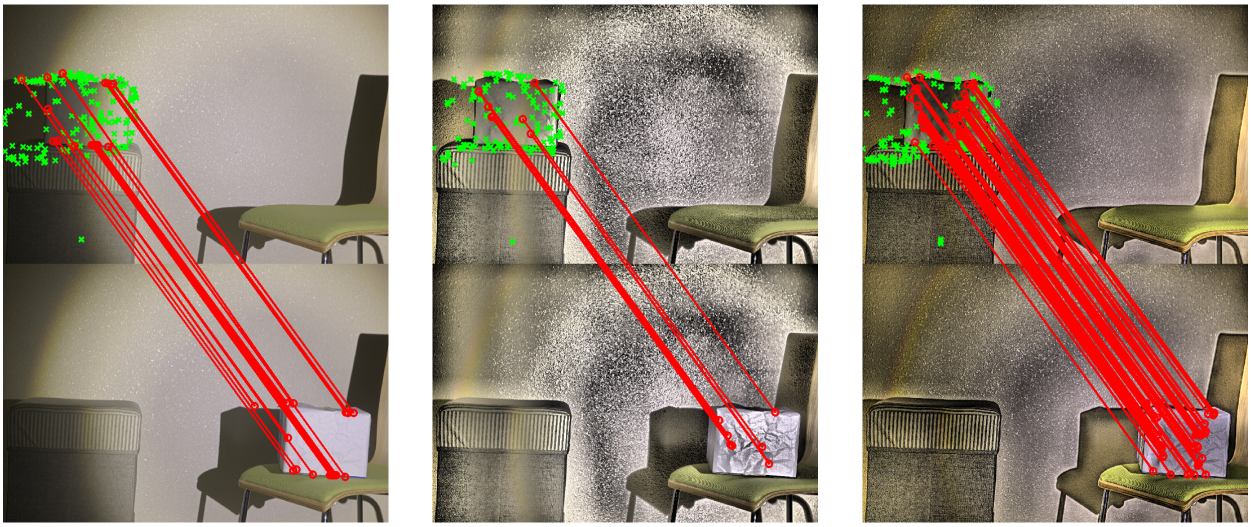
where N is the rank of the Laplacian pyramid and C_n is a weight. This weight was set as $C_n = 0.8/N^2 \cdot n^3$ based on experimental results and it approaches zero at the high resolution scale. This weight avoids adding the local details of the high resolution scale to the lighting effect Y_c. Subtracting Y_c from the enhanced luminance Y_e can moderate the lighting effect. Figure 4 shows the estimated lighting effect and the results of its removal.

Figure 5 shows validity of our lighting effect removal method for SIFT object tracking. Our method without lighting effect removal (Fig.5(d)) can track more SIFT feature points in an untrackable area compared with the conventional method (Fig.5(c)). Our tone-mapping method suppresses large contrasts, which moderates the strong lighting effect. Our method with the lighting effect removal (Fig.5(e)) can also track many more SIFT feature points in the untrackable area.



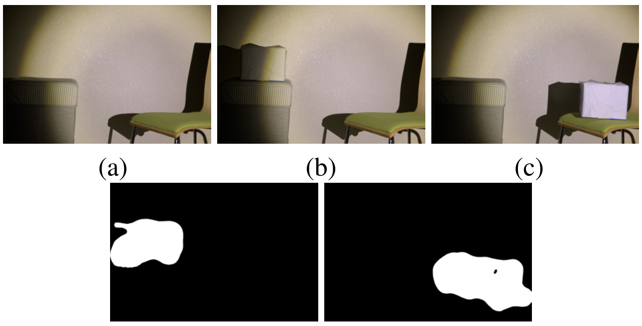
(a) (b) (c) (d) (e)

Fig. 5. Results of lighting effect removal: (a) reference frame, (b) target frame, (c) conventional method [4], (d) non-removal, (e) our method. The trackable points are red circles, while the failed points are green x-marks.



(a) (b) (c)

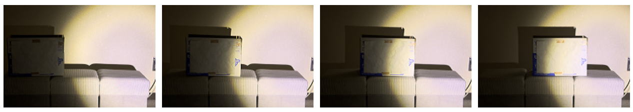
Fig. 6. Comparison of the results: (a) Reinhard [4], (b) Reinhard[4] + CLAHE, (c) Our method. The red circles denote valid feature points and green x-marks indicate failed feature points.



(a) (b) (c)

(d) (e)

Fig. 7. Input images and masks of the experiment (Figure 6): (a) background image, (b) reference image, (c) target image, (d) mask of current image, (e) mask of target image.



"box1" "box2" "box3" "box4"

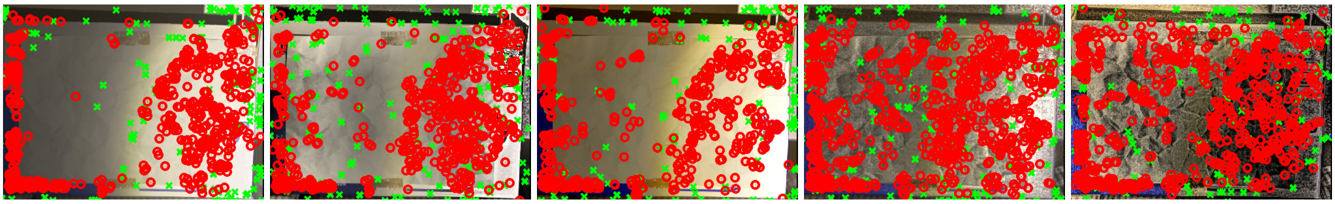
Fig. 8. Scene used in the experiment (Table.1).

3. RESULTS

This paper compares the accuracy improvement with existing SIFT object tracking using our method and the conventional methods. This section first compares conventional natural tone-mapping method [4] with our method. The results are shown in Figure 6, and this experiment uses the input images and masks shown in Figure 7. Figure 6(a) is the object tracking result using [4]. No points could be tracked using the SIFT features, with the exception of the boundary of the object. Since [4] does not allow high contrast enhancement, most of the minute details are lost or too weak. We improved [4] by adding contrast enhancement after tone-mapping. Figure 6(b) shows the result using enhanced [4], where we used contrast limited adaptive histogram equalization (CLAHE) for contrast enhancement. In this case, the points were tracked in the region that only had the minute features whereas the points on the boundary often failed. The boundaries are large edges whereas the minute features are small edges. Thus, general local contrast enhancement has a variable enhancement effect depending on the surrounding luminance range. The conventional method does not remove the lighting effect, so it is difficult to achieve robust object tracking using SIFT features. Our method can enhance the details

Table 1. Number of valid SIFT feature points and the accuracy of SIFT object tracking.

	Reinhard[4]		Reinhard[4] + CLAHE		Subband[5]		Local Laplacian[6]		Our method	
Frame name	Total points	Success rate	Total points	Success rate	Total points	Success rate	Total points	Success rate	Total points	Success rate
box 1-2	515	65.0%	707	69.3%	725	66.1%	822	75.8%	762	76.6%
box 2-3	564	75.7%	639	70.9%	797	80.8%	1185	88.3%	963	86.7%
box 3-4	685	80.0%	751	78.3%	709	79.4%	959	83.3%	791	85.5%



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

Fig. 9. Results (reference:"box3", target:"box4"): (a) Reinhard[4], (b) Reinhard + CLAHE, (c) Subband[5], (d) Local Laplacian[6], (e) Our method. The trackable points are red circles, while the failed points are green x-marks.

of the whole luminance range uniformly using our lighting effect removal method (Figure 6(c)).

This section secondly compares the conventional local tone-mapping methods with ours. This experiment uses the sequence consists of still HDR images shown in Figure 8. In this sequence, a box moves from the dark area to the bright area. To make the sequence, this experiment fused three multiple exposure images captured using a Canon 40D. The images shown in Figure 8 were tone-mapped using [4] and their resolution is 1944×1296. Table 1 shows the results of SIFT object tracking, where "box 1-2" denotes the object tracking result using Figure 8 "box1" and "box2". The conventional method performed better when the object was in a bright area rather than a dark area. This was because the minute features were larger in the bright area. Our method delivered higher object tracking with greater accuracy than the conventional methods. Figure 9 shows the success points and the failed points when reference image is "box3" and target one is "box4". [4] has high success rate but Figure 9(a) explains success points concentrated on the object outline or bright area. The success rate of the enhanced [4] is lower than [4], but Figure 9(b) shows contrast enhancement (CLAHE) increases the success points in dark area. The results of [5] are similar to the results of the enhanced [4]. The results of [6] are comparable to ours, but our algorithm is more adapted for SIFT features and our computational cost is far lower than it.

4. CONCLUSION

Present paper proposed a tone-mapping for the HDR surveillance system. Our method improves the accuracy of existing SIFT object tracking methods used in HDR surveillance systems. In addition, our method enhances minute details and removes the lighting effect, which can increase the success rate when tracking previously untrackable objects using SIFT.

In future, we will reduce the computational demand so this approach is suitable for real-time processing. We will also make further accuracy improvements when tracking untrackable object using SIFT features. The tone-mapping algorithm [7] on which the proposed method based can be used to high efficiency two-layer HDR coding. We will apply the proposed method to two-layer HDR coding.

5. REFERENCES

[1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, 60, 2 (2004), pp. 91–110.

[2] E. Reinhard, S. Pattanaik, G. Ward, and P. Debevec, "High Dynamic Range Imaging: Acquisition, Display, and Image- Based Lighting," *Morgan Kaufmann Publisher*, 2005.

[3] K. Devlin, A. Chalmers, A. Wilkie and W. Purgathofer, "Tone Reproduction and Physically Based Spectral Rendering," *Eurographics*, 2002.

[4] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic Tone Reproduction for Digital Images," *ACM Trans on Graphics*. 21, 3, 267–276, 2002.

[5] Y. Li, L. Sharan and E. H. Adelson, "Compressing and companding high dynamic range images with subband," *Proceedings of SIGGRAPH '05*, Vol. 24, Issue. 3, pp. 836-844, July, 2005.

[6] S. Paris, S. W. Hasinoff, and J. Kautz, "Local Laplacian Filters: Edge-aware Image Processing with a Laplacian Pyramid," *SIGGRAPH*, 2011.

[7] T. jinno, H. Watanabe and M. Okuda, "High contrast tone-mapping and its application for two-layer High Dynamic Range coding," *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Dec. 2012.