# CONTENT-ADAPTIVE INVERSE TONE MAPPING

*Pin-Hung Kuo, Chi-Sun Tang, and Shao-Yi Chien*

Graduate Institute of Electronics Engineering, National Taiwan University, Taiwan

## ABSTRACT

Tone mapping is an important technique used for displaying high dynamic range (HDR) content on low dynamic range (LDR) devices. On the other hand, inverse tone mapping enables LDR content to appear with an HDR effect on HDR displays. The existing inverse tone mapping algorithms usually focus on enhancing the luminance in over-exposed regions with less (or even no) effort on the process of the well-exposed regions. In this paper, we propose an algorithm with not only enhancement in the over-exposed regions but also in the remaining well-exposed regions. This paper provides an "histogram-based" method for inverse tone mapping. The proposed algorithm contains a content-adaptive inverse tone mapping operator, which has different responses with different scene characteristics. Scene classification is included in this algorithm to select the environment parameters. Lastly, enhancement of the over-exposed regions, which reconstructs the truncated information, is performed.

***Index Terms*—** Inverse Tone Mapping, Scene Recognition, Support Vector Machine

## 1. INTRODUCTION

The popularity of HDR images has been increasing in recent years. More and more devices have shown this trend where they have built-in HDR capture capability, mobile phones and consumer digital cameras for example. However, the problem lies in how to exhibit such a large luminance range on traditional displays. In the past decade, a large number of tone mapping algorithms were proposed to solve this problem. Although there are usually big differences between these algorithms, we can roughly sort them into two classes: global tone mapping and local tone mapping. In global tone mapping, every pixel is mapped to an LDR value following a same response curve. In other words, global tone mapping produces LDR images with few artifacts and a perception similar to that of HDR images. However, to maintain the contrast of every part in the image, it has to inevitably compress or discard some luminance information. On the other hand, local tone mapping aims to reproduce the image to preserve local details in every part of the image despite luminance variation. As a result, halo artifacts and perceptual distortions may be produced in the boundary of bright and dark regions.

Recently, the dynamic range of displays is getting higher. The available HDR displays in consumer market is a predictable future. However, the content that have been captured and stored are all still in LDR formats. To make HDR displays compatible with LDR contents, inverse tone mapping is necessary for the HDR display. As the literal meaning, inverse tone mapping performs an inverse process of traditional tone mapping. In video or image capture, cameras have unique responses to compress the world luminance to displayable luminance and truncates the extremely high luminance in over-exposed regions. Therefore, the inverse tone mapping algorithm can be mainly divided into two parts: inverse tone mapping operator and over-exposure enhancement.

An inverse tone mapping operator is used to estimate the response of LDR to HDR, which conjugates to tone mapping. However, deciding the parameters of the operator is difficult, which dominates the quality of inverse tone mapping. As scene varies, choosing the appropriate parameter and maintaining a visually consistent inverse tone-mapped output sequence is difficult. In this paper, we propose an algorithm for reconstructing the response curve and determining the parameters in a content-aware manner.

The second part of the over-exposure enhancement aims to recover the truncated luminance during capture. This is difficult because no information exists on these over-exposed regions in the LDR images. A perfect luminance reconstruction is impossible, but we can still enhance these regions to have a perception similar to the real scene. In this paper, a robust method is proposed to determine the over-exposed regions and enhance their luminance.

In the remaining parts of this paper, Section 2 introduces the previous works. In Section 3, the proposed algorithm of the inverse tone mapping, support vector machine (SVM) scene classifier, and over-exposure enhancement are described in detail. The Section 4 shows the experiment results. Finally, Section 5 concludes this paper.

## 2. RELATED WORKS

To acquire as much information as possible of the real world scene, devices have to compress luminance during image/video capture. In fact, even original tone mapping executes a similar compression, for both camera and tone mapping algorithm aim to show a vast luminance range on devices

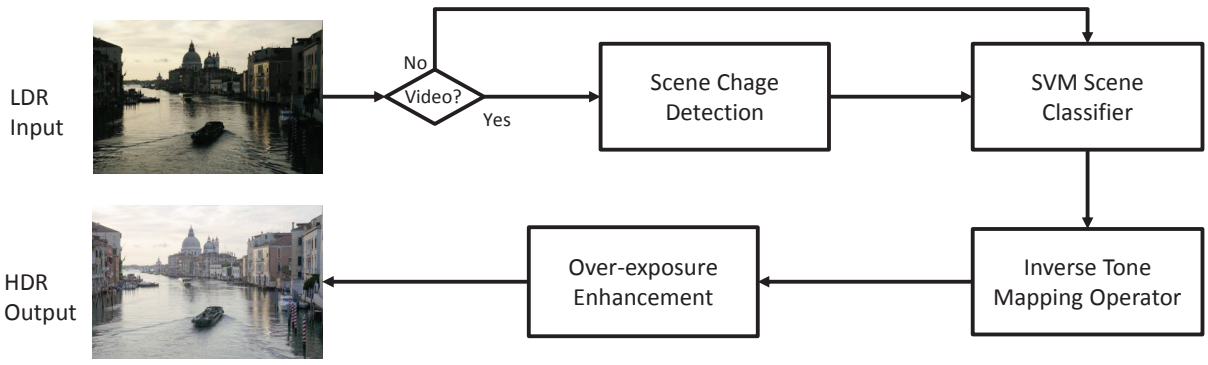


**Fig. 1**. The flow chart of the system.

with a relatively narrow luminance range capability. On the contrary, the first step of inverse tone mapping is to reverse these tone mapping processes to restore the compressed dynamic range to the real world dynamic range. In [1], Banterle et al. derived the inverse tone mapping operator from the work of Reinhard et al. [2]. Rempel et al. [3] chose to inverse the gamma, which has been commonly adopted in standard video and television formats. Although these algorithms perform well with the correct parameters, the parameters are usually decided by the user rather than by a robust mechanism. Therefore, we propose an inverse tone mapping operator whose parameters can be decided based on the content.

After the inverse tone mapping operator, the truncated luminance information in the over-exposed region should be reconstructed. The main problem is determining whether a region is over-exposed or not. Meylan et al. [4] proposed a simple but useful algorithm to approach the response curve with two linear lines. It divides the input image into a diffuse region and specular region. The specular region, which is the over-exposed or brightest region of the entire image, will be mapped to the HDR domain with a gain higher than the diffuse region. In [1], a median cut algorithm for light probe sampling [5] was adopted to determine the light sources. For lower computing complexity and easy to be accumulated by FPGA or GPU, Rempel et al. [3] made efforts to estimate a region of over-exposure rather than separate bright points. An iterative process will be performed starting from high value points. After each iteration, these bright points spread out and form bigger bright regions. This process will continue till the boundary of the bright region reaches edges with large gradients. We can determine over-exposed regions with a fade out effect that matches our perception to bright light sources by this algorithm. However, the hard-decided starting points are usually erroneous, and hence, the well-exposed regions are treated as over-exposed regions. Moreover, bad parameters result in the divergence of the iteration. So we propose an algorithm with a more robust mechanism.

All the above inverse tone mapping algorithms focus on processing of still images. Although most of them can be extended to video, in fact, the process is just performed frame by frame. As a result, the algorithm proposed by this paper aims to process video taking the content into account.

## 3. PROPOSED ALGORITHM

The design of our inverse tone mapping algorithm is inspired by the work of Schlick [6]. Schlick proposed the algorithm by taking human perceptual characteristics into consideration. With some environment-dependent parameters, the tone mapping response varies with different scene luminance. However, these manually decided parameters may affect the result drastically. There is no guarantee that every user can estimate the scene luminance or the parameters correctly. Moreover, deciding the parameters frame by frame is not practical in video processing. With regard to over-exposure enhancement, the previous works [3] and [4] both provided solutions. [3] can enhance the over-exposed region as a light source in it, but the parameters have to be adjusted for different cases. On the other hand, [4] has a robust performance in the specular point search but a coarser over-exposed region compared to the work of Remple et al. [3].

The algorithm proposed by this paper mainly consists of two components: scene-adaptive inverse tone mapping and over-exposure enhancement. The scene-adaptive inverse tone mapping algorithm includes a scene classifier and the inverse tone mapping operator. The proposed inverse tone mapping operator is based on the work of Schlick [6]. This inverse tone mapping operator contains environment-dependant parameters, which is selected by a user in the original operator. To decide such parameters, a support vector machine (SVM)is adopted. An SVM scene classifier was trained to recognize the scenes so that the parameters of the inverse tone mapping operator can be decided automatically. For the over-exposure enhancement step, our method is based on the works of Rempel et al. [3] and Meylan et al. [4] to make up for the deficiency in each work.

Both components and their details will be described in the following section.

### 3.1. Scene-adaptive Inverse Tone Mapping

As far as the response for a particular image is concerned, the existing previous works have already provided satisfying performances. In fact, they always take the image characteristics into account, such as the maximum luminance [1], the ability of the target display [3], and entire contrast [4] in the image. Nevertheless, most of them focus on single image processing, which may lead to flicker or an unreal perception as the video plays. In view of this, we propose an inverse tone mapping operator as well as a scene classifier to decide the parameter of this operator.

#### *3.1.1. Inverse Tone Mapping Operator*

As described in the above section, this work is based on that of Schlick [6]. The work of Schlick is

$$L_d = \frac{pL_w(x,y)}{(p-1)L_w(x,y) + L_{max}}, \text{where } p \in [1,\infty) \quad (1)$$

$L_d$ and $L_w$ refers to the display luminance and world luminance, respectively. The parameter $p$ can be approximated by

$$p = \frac{\delta L_0}{N}\frac{L_{max}}{L_{min}} \quad (2)$$

The $L_0$ term in (2) is the just-noticeable difference (JND). In other words, $p$ may be seem as the smallest value that is not black after tone mapping. As described in [6], the $L_0$ can be decided with a simple experiment which show the patches with different luminance at random positions on a black background. The minimum luminance can be observed different from black is $L_0$. In our experiment, the $L_0$ is about $7-10$. The $L_{max}$ and $L_{min}$ represent to the maximum luminance and minimum luminance in world scenes. Although these two values vary with scenes, $L_{max}/L_{min}$ can be seemed as the contrast ratio of one scene. As described in [7], the temporal dynamic range of human eye is about 3 orders, so we chose $L_{max}/L_{min}$ as $10^3$. With $N = 256$, the value of $p$ is 27.34375. For simplicity, we set $p$ as 30 in this paper.

The reason for choosing this operator is because of perception concerning and also its simplicity. With (1), the proposed inverse tone mapping operator can be directly derived from (1) as follows:

$$L_w = \frac{L_d L_{max}}{p(1-L_d) + L_d} \quad (3)$$

With the experiment-decided $p$, the only parameter to be decided in (3) is $L_{max}$. The next subsection will introduce the method how to determine this value.

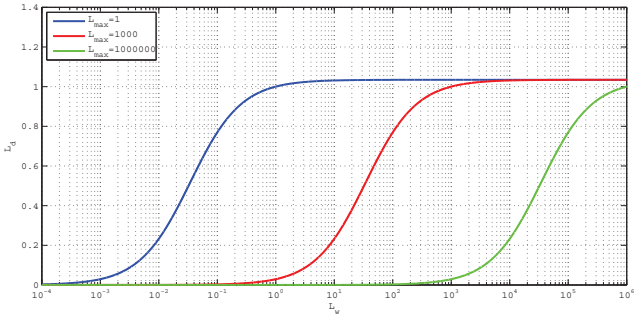


**Fig. 2**. The tone mapping operator response of Schlick ($p = 30$).

#### *3.1.2. Scene Classifier*

The SVM is a machine learning method that has similar roots to neural networks. Nowadays, it is widely adopted in scene classification. We use the libsvm tool provided by Chih-Chung Chang and Chih-Jen Lin [8] to train our classifier. In general, users of the SVM should extract the features from the training data and tag every training data with a class label. After training, the output is a model that can be used for classifying data. The data set of SUN [9] and images on the internet are used as the training data in this paper. One of the features extracted from the data is a histogram in 512 dimensions, where each RGB channel is quantized from 256 bins to 8 bins. The last feature in the vector is the ratio of specular points to the entire image pixels, which will be introduced in Section.3.2. This classifier aids us in determining the parameters of the operator, taking scene information into consideration.

There are three classes in our SVM classifier. The three classes are "bright", "midtone," and "dim". These three classes represent the scenes of high, middle, and low luminance levels. For example, a scene in moon light will be classified in the "dim" class, "midtone" usually corresponds to the indoor scenes and "bright" class contains the most bright scenes such as under the sunlight. In other words, the class of an image depends on the luminance level and histogram, rather than the objects in the image. With three luminance levels and dynamic ranges, the content can be mapped to the parameters matching to the scenes. According to [7], the ambient luminance in sunlight condition is about $10^5 cd/m^2$, $10^2 cd/m^2$ for indoor lighting condition and $10^{-1} cd/m^2$ for moonlight condition. We assume that the $L_{max}$'s of these three conditions are 10 times to the ambient luminance, where is $10^6 cd/m^2$, $10^3 cd/m^2$ and $1 cd/m^2$ respectively. Fig.2 shows the responses of these 3 $L_{max}$'s.

In our experiments, the SVM usually classifies images correctly. However, mistaken detection happens at boundary of scene changes. To compensate this effect, we apply a shot detection improvement. We assume that the frames of one

shot should be of the same scene. So the minor mis-detected frames will be forced to be of the same scene of the remaining frames.

In general cases, it does improve the performance of the classification. In spite of this, this improvement has to be adopted for the reason that it prevents the inverse tone mapped HDR video from flicker, as a result of mapping the frames in one shot with different responses. The table 1 shows the results of this improvement. In some case this improvement cannot work for the mis-detected frames are as many as the correctly detected frames. In such cases, we add a gradual mechanism to eliminate the flickers. The $L_{max}$'s of two adjacent frames with different SVM scenes will increase or decrease gradually. For example, a "dim" scene after a "midtone" scene has a $L_{max}$ about $100cd/m^2$ lower than the $1000cd/m^2$ rather than jumping to $1cd/m^2$.

With this classifier, we can decide the three $L_{max}$'s correlating to the three scenes and choose the parameters according to the results of the classification.

### 3.2. Over-Exposure Enhancement

Although we can reconstruct from LDR images to HDR by the method mentioned in the previous section, the truncated information in over-exposed regions is still lost. The perfect recovery of such information is impossible, but we can synthesize such information to make it as real as possible. As mentioned in the above section, we combine the advantages of the previous works to achieve this.

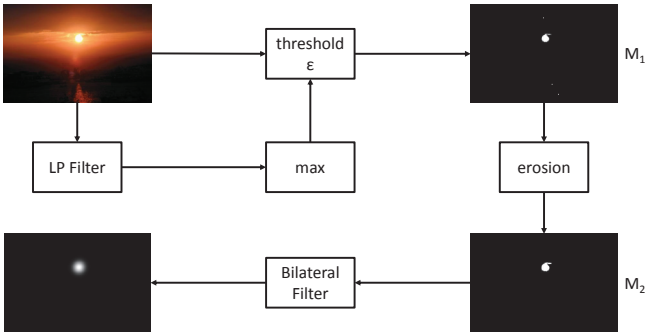


**Fig. 3**. The over-exposure enhancement flow.

Fig.3 shows the over-exposure enhancement flow. At first, the input image is passed through one low-pass filter. The kernel size of the filter is about $0.1max(width, height)$, and the filter we adopted in this paper is an average filter as the selection of Meylan et al. [4]. After passing through this filter, the biggest values in the images are chosen to be the threshold $\epsilon$. Then, we have the first binary map $M_1$ with threshold $\epsilon$

$$M_1(p) = \begin{cases} 1 & \text{if } I(p) > \epsilon \\ 0 & \text{otherwise} \end{cases}$$

With morphological erosion to eliminate single 1's resulted from noise, we can acquire binary map $M_2$:

$$M_2(p) = \begin{cases} 0 & \text{if } M_1(p) = 0 \text{ or } dil(M_1(p)) = 0 \\ 1 & \text{otherwise} \end{cases}$$

1's in $M_2$ can be regarded as the bright points or the specular points which is mentioned in 3.1.2. Intuitively, a bright point lights up the space around it and become dimmer as the distance increases. To figure out the over-exposed region with such visual effects, we pass $M_2$ and the input image through a joint bilateral filter with a large kernel size, leading to the "fade out" effect. Besides, the joint bilateral filter keeps the light from spilling over edges with large gradient. As described in [3], this flood-fill algorithm should be stopped at the pixels with strong edges, as the light will be shaded by some objects. This proposed method has the effect very close to [3] while this method is more robust than the previous works. First, the binary map $M_2$ is computed according to the characteristics of the image, rather than a threshold for every image. For example, a lossy compressed image and a uncompressed image need different thresholds, and an image of brighter scene and that of dimmer scene need different thresholds too. The method proposed by this paper provides a more robust approach to find these bright points. Second, the flood-fill step in [3] requires another gradient threshold to decide where the flood has to be stopped, this threshold has to be adjusted with different images too. Moreover, the algorithm in [3] is iterative, which means that the iteration may be divergency. On the other hand, the bilateral filter with large $\sigma_r$ of range term provides stable results. In our experiments, bilateral filter with kernel size as the average filter and a $\sigma_r$ higher than 100 performs well for most images. The results of over-exposure enhancement are shown in the most right column of Fig.5. In addition, these specular points, i.e., 1's in $M_2$, are one of the SVM features in scene classifier.

## 4. EXPERIMENTAL RESULTS

In this section, we discuss the results in Table 1 at first. As described in 3.1.2, the shot detection improvement works for most cases in our experiments. However, a scene with several moving objects may be less accurate in the prediction. Under these circumstances, the accuracy of this shot will be 0 after shot detection improvement. For example, in the video Bravia, we have seven shots improved by this method, but the accuracies of three shots become 0 after this step. Although this method cannot always improve the accuracy of prediction, it is necessary in this algorithm because it makes the inverse tone mapping response curves uniform in one shot, which is mentioned in 3.1.2.

To tell whether the method proposed in this paper works or not, we adopt the HDR-visual difference prediction (HDR-VDP) [10]. The HDR-VDP is a perceptual criterion between HDR and HDR or HDR and LDR. Fig.5 shows the results. In the figure, we compare the results of this work with that of [1]. Table 2 shows the values of HDR-VDP in detail. Fig. 5 shows that our algorithm preserves the perceptual characteris-

| Video | Shots | Correct | Correct* | Frames | Correct | Accuracy | Correct* | Accuracy* |
|---|---|---|---|---|---|---|---|---|
| Totoro | 20 | 18 | 20 | 550 | 536 | 97% | 550 | 100% |
| Night at the Museum | 20 | 14 | 14 | 476 | 375 | 79% | 401 | 84% |
| Pirates of the Caribbean | 8 | 7 | 8 | 313 | 286 | 91% | 313 | 100% |
| Bravia | 38 | 18 | 22 | 833 | 495 | 59% | 507 | 61% |

**Table 1**. Experimental results of the SVM classifier. The * represents the results with shot detection improvement.

| | Works of Banterle et al. [1] | | Our Algorithm | |
|---|---|---|---|---|
| Image | $p > 0.75$ | $p > 0.95$ | $p > 0.75$ | $p > 0.95$ |
| Bridge | 44.223% | 37.323% | 0.041% | 0.019% |
| Sunset | 11.298% | 8.115% | 0.021% | 0.014% |
| Winter | 17.409% | 11.391% | 0.004% | 0.003% |

**Table 2**. The probability of HDR-VDP corresponds to Fig.5.

tics of the input content. Although the results of our work and [1] are all produced with the same $L_max$, our work produces overwhelming results.

The Fig. 4 shows the results of the comparison between our inverse tone mapping HDR and the synthesized HDR image from three LDR images. Although this result is not as good as our expecting, it exhibits the preservation of the perceptual characteristics in the brighter areas. The darker areas in the image are tagged with probable visual differences, and the synthesized HDR contains more information in these areas; such information is not available in the normal exposure LDR image.

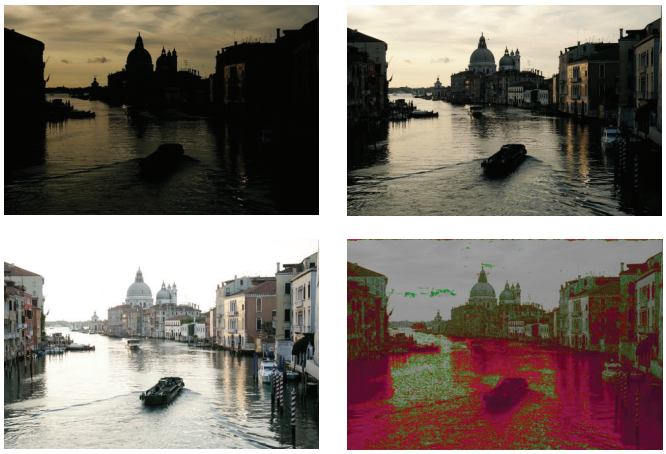

**Fig. 4**. HDR-VDP comparison between HDR and HDR. The images in the top row are under exposure and normal exposure images, and the left image in the bottom row is an over-exposure image. The right image in the bottom row is the HDR-VDP comparison between the HDR image synthesized from the three LDR images and the HDR inverse-tone mapped from the normal exposure LDR image. The region with red color is the region where the possibility of difference to be noticed higher than 0.95, and the green means possibility than 0.75.

## 5. CONCLUSION

A content-adaptive inverse tone mapping algorithm is proposed in this paper, which is currently the most complete work of inverse tone mapping. The algorithm not only enhances the over-exposure region but also takes the different scene responses into account. Besides, this work is a "histogram-based" method: the SVM classifier and scene change detection are both based on the 512-bin histograms.

To achieve the automatic parameter decision, the SVM classifier is adopted in this work. From the experimental results, SVM is adequate for scene recognition. Contents with similar characteristics will be classified into a same class, and the different mapping responses can be applied to these contents.

In this paper, we focus on inverse tone mapping for videos, which take temporal information into consideration, rather than just tone mapping the video frame by frame. The proposed algorithm also provides auto-adjusted tone mapping responses for each kind of images.

## 6. REFERENCES


[1] Francesco Banterle, Patrick Ledda, Kurt Debattista, and Alan Chalmers, "Inverse tone mapping," in *Proceedings of the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia*, New York, NY, USA, 2006, GRAPHITE '06, pp. 349–356, ACM.

[2] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda, "Photographic tone reproduction for digital images," in *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 2002, SIGGRAPH '02, pp. 267–276, ACM.

[3] Allan G. Rempel, Matthew Trentacoste, Helge Seetzen, H. David Young, Wolfgang Heidrich, Lorne Whitehead, and Greg Ward, "Ldr2hdr: on-the-fly reverse tone map-

**Fig. 5**. Results of HDR-VDP and over-exposure enhancement. The columns from left to right are the original LDR images, the HDR-VDP results of Banterle et al. [1], the HDR-VDP results of our method and the results of over-exposure enhancement respectively. The green and red regions in the second and third columns represent the same possibilities as in Fig.4


ping of legacy video and photographs," in *ACM SIGGRAPH 2007 papers*, New York, NY, USA, 2007, SIGGRAPH '07, ACM.

[4] Laurence Meylan, Scott Daly, and Sabine Susstrunk, "Tone mapping for high dynamic range displays," in *Proc. IS&T/SPIE Electronic Imaging: Human Vision and Electronic Imaging XII*, 2007, vol. 6492.

[5] Paul Debevec, "A median cut algorithm for light probe sampling," in *ACM SIGGRAPH 2005 Posters*, New York, NY, USA, 2005, SIGGRAPH '05, ACM.

[6] Christophe Schlick, "Quantization techniques for visualization of high dynamic range pictures," in *Photorealistic Rendering Techniques*. 1994, pp. 7–20, Springer-Verlag.

[7] B A Wandell, *Foundations of Vision*, vol. 21, Sinauer Associates, 1995.

[8] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.

[9] Jianxiong Xiao, J. Hays, K.A. Ehinger, A. Oliva, and A. Torralba, "Sun database: Large-scale scene recognition from abbey to zoo," in *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, june 2010, pp. 3485 –3492.

[10] Rafał Mantiuk, Karol Myszkowski, and Hans-Peter Seidel, "Visible difference predicator for high dynamic range images," in *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, 2004, pp. 2763–2769.