SUBJECTIVE EVALUATION OF TONE-MAPPING METHODS ON 3D IMAGES

Zicong Mai, Colin Doutre, Panos Nasiopoulos, and Rabab K. Ward

Department of Electrical and Computer Engineering
The University of British Columbia

ABSTRACT

High dynamic range (HDR) imaging provides superior picture quality to traditional 8 bit, low dynamic range (LDR), image representations. Capturing images/videos in HDR format can avoid problems with over and under exposures. Tone-mapping is a process that converts from HDR to LDR, so that HDR content can be shown on existing displays. Tone mapping has been extensively studied in the context of 2D images/video but not for 3D content. This paper addresses the problem of presenting 3D HDR content on stereoscopic LDR displays and presents a subjective psychophysical experiment that evaluates existing tone-mapping operators on 3D HDR images. The results show that 3D content derived using tone-mapping is much preferred to that is captured directly with a pair of LDR cameras. Global tone-mapping methods (which better preserve global contrast) are found to produce images with better 3D effect than local tone-mapping operators (which produce images with high amounts of detail/texture). Also, the brightness of the tone-mapped images is found to be highly collated with perceived 3D quality.

Index Terms— 3D, stereoscopic, high dynamic range, HDR, tone-mapping, subjective tests

1. INTRODUCTION

High dynamic range (HDR) images/videos provide superior picture quality by supporting a very large luminance range, which is comparable to what human vision is able to cover. On the other hand, existing content is considered to be in low dynamic range (LDR), which allows only a limited range of contrast and is far below the capability of human eyes. In order to carry the immense amount of information, HDR signals are encoded with at least 10 bits per color channel [1], as opposed to LDR signals which are represented by only eight bits. Although the majority of today's displays can support only LDR content, they can all provide much better picture quality if the content is first captured in HDR and then converted to the LDR format. The process that converts HDR to LDR signals is called tone-mapping. A large variety of tone-mapping operators (TMOs) have been developed for different purposes, ranging from photographic tone reproduction to adaption to different displays [2]-[3], [8]-[12]. Combining HDR capturing and tone mapping produces higher quality images with less over and under saturated areas compared to LDR capturing even when an LDR display is used.

3D displays provide another way for significantly enhancing users' visual experience by providing a sense of depth. Factors such as brightness [13] and scene detail [14] have been noted to affect the visual comfort and quality of 3D content. HDR imaging and tone mapping can affect these factors.

Ideally, content could be captured in a 3D HDR representation and viewed on a 3D HDR display, to achieve a more lifelike picture quality. However, existing 3D displays can support only 8-bit LDR content. In order for HDR content to be displayed on existing imaging systems, we need to generate LDR signals for each view of the 3D HDR pair. That is, tone-mapping needs to be applied to each view.

Tone mapping of 2D images and videos has been extensively studied and several subjective evaluations have been performed. Drago et al. [4] performed subjective tests on six TMOs with 11 observers and found that the most salient attributes of tone-mapping are naturalness and detail. Yoshida et al. [5] evaluated seven TMOs using two real-world scenes in the comparison and concluded that global (spatially invariant) TMOs are preferred for the reproduction of overall contrast and brightness while local (spatially variant) TMOs perform better for reproducing details in highlights. In 2007, Kuang et al. [6] evaluated eight tone-mapping methods having subjects rate four specific attributes of the images; namely, global contrast, colorfulness, shadow details and sharpness.

Inspired by the fact that the impact of tone mapping on 3D content has not been studied, in this work we conduct subjective tests that evaluate tone-mapping methods on 3D displays. The objective is not to rank existing TMOs, but to understand which attributes of tone-mapping methods will contribute to good 3D representation and how they differ from the 2D scenario. This paper is organized as follows: An overview of tone-mapping and a short description of the seven TMOs used in our experiments are presented in

This work was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC).

Section 2. Section 3 describes the framework of our experiments. Results and discussions can be found in Section 4. Conclusions are presented in Section 5.

2. TONE-MAPPING OPERATORS

Tone-mapping operators are used for reducing the dynamic range in such a way that the resulting image or video may be displayed on LDR devices. A review of many popular operators can be found in [7].

TMOs can be categorized into two broad classes: global and local operators. Global TMOs use a single (nonlinear) mapping curve for every pixel while in the case of local operators the mapping of each pixel is a spatially variant function of neighboring pixels. In general, global TMOs are more computationally efficient and better retain the sensation of the brightness of the original HDR signal. On the other hand, local operators usually preserve more details and provide higher local contrast.

Since dozens of tone-mapping operators are available today, it is impractical to evaluate all of them. Therefore, in our experiments, seven TMOs, four global and three local, are selected that cover most design principles of tone-mapping. Below are short descriptions of these seven tone-mapping operators. The italic text appearing in parenthesis immediately after each name is used to represent the corresponding TMO in the rest of this paper.

Photographic tone-mapping (*Photo*) [2]: It simulates the dodging and burning techniques in photographic tone reproduction. This scheme supports both global and local versions. The local version computes a local brightness adaptation level for each pixel by constructing a Guassian pyramid and identifying a largest possible neighborhood that contains relatively low contrast. Then such adaption level is used together with a sigmoidal function to compress the dynamic range. In this paper we tested only the local operator since the behaviors of the two versions are similar. Most previous studies also considered only the local version [4], [5], [15].

Adaptive logarithmic tone-mapping (*Log*) [8]: This global TMO takes into account the human vision system and is formulated as a logarithmic compressive function. The base of the logarithm is varied according to the input intensity.

Fast bilateral filtering (*Bi*) [9]: This is a local TMO. It first generates a base layer using a bilateral filter, which preserves the edges and smoothes the rest of the HDR image. Then, an enhancement layer is produced by dividing the HDR image by the base layer. The LDR image is derived by adding the enhancement layer and a compressed dynamic range version of the base layer.

Tone-mapping for backward-compatible HDR compression (*Comp*) [10]: It is a global tone-mapping method that was inspired by the observation that most tone-mapped content, especially videos, will undergo lossy image/video encoding such as H.264/AVC compression. This TMO thus aims at minimizing the join information loss due to the tone-mapping and the image/video encoding processes.

Gradient domain tone-mapping (*Grad*) [11]: It achieves dynamic range reduction in the gradient domain. This tone-mapping method attenuates large gradients by manipulating gradients of the image in different spatial scales. The LDR image is generated by solving a Poison equation in an adjusted gradient field.

Display adaptive tone-mapping (*Disp*) [3]: This tone-mapping operator is formulated using a global piece-wise function which is uniquely determined by minimizing the perceptual contrast distortion between the original HDR signal and the resulting tone-mapped content on various

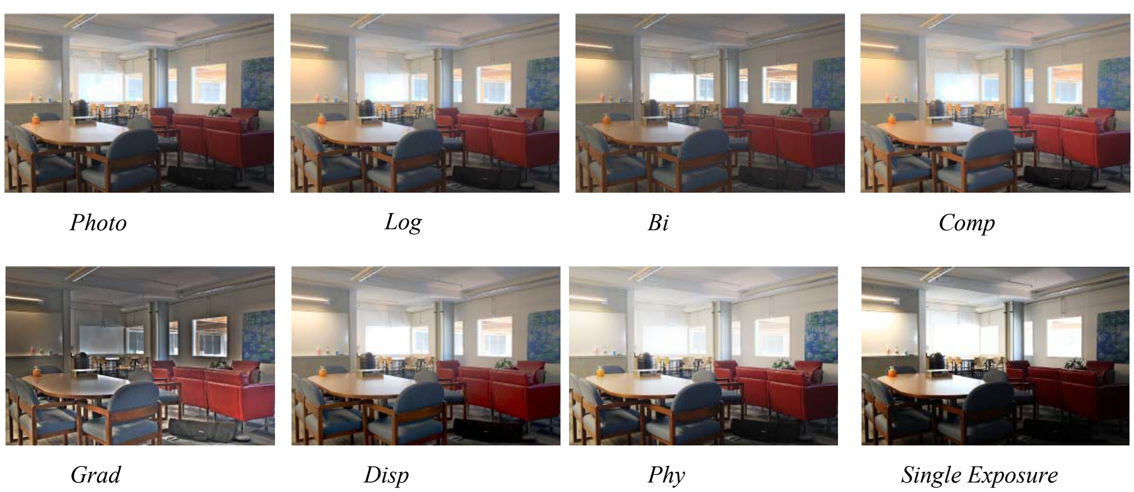

Fig. 1 LDR images generated from the same HDR scene by tone-mapping operators evaluated in this paper. The image captured using one single LDR exposure is also included in this figure.

displays. The display models vary from e-paper to high contrast output devices.

Tone-mapping by photoreceptor physiology *(Phy)* [12]**:** Based on the finds in physiology that vision adaption occurs in the photoreceptor, this global TMO achieves tone reproduction by modeling the behavior of photoreceptors. An advantage of this operator is that it can be easily controlled by three intuitive user parameters: brightness, contrast and chromatic adaptation.

In order to give the reader an idea of the visual properties of each TMO, in Fig. 1 we show the LDR images generated by the operators tested in this paper. The same scene, captured with a single LDR exposure, is also shown. Note that the tone-mapped versions of the image avoid the clipping that is present in the LDR capture. It can also be seen that the local TMOs (particularly *Bi* and *Grad*) produce images with more details/texture than the global TMOs.

3. EXPERIMENTAL FRAMEWORK

Several of the selected tone-mapping operators require their user parameters to be specified. Therefore, prior to the main subjective test, we conducted a pilot study that asked subjects to choose the best possible parameters for each of the TMOs for each scene respectively. We found that the fine-tuned parameters by different subjects were very similar for the same TMO and the same scene, so average values were used to generate tone-mapped images for each of the scenes.

Sixteen subjects participated in our main experiment. They had diverse cultural backgrounds and differed in age (24 – 38 years old). All of them had normal or corrected vision and little experience with 3D and HDR content. The display device used in our test was a *Samsung UN55C7000* 3D TV, 55 inch, 1920 * 1080p at 240Hz, which uses active shutter glasses.

In our main experiment, subjects needed to evaluate stereoscopic LDR images on a 3D display. Particularly, subjects were asked to evaluate eight scenes: four indoors and four outdoors (see Fig. 2). Eight different LDR images were derived from each scene: seven of them are tone-mapped using the selected TMOs described in the previous session while one of them is a well-exposed image captured in LDR. The display order of the different tone mapped versions of each scene was randomized. In total, 64 images were tested. In order to help the subjects better adapt their visions to a new image, we inserted a plain grey slide between tested images to cushion the transition.

For each of the 64 3D images in the test, subjects were asked to evaluate two aspects: i) the 3D effect and ii) the overall quality. The 3D effect score rates the impression of depth and how comfortable the 3D scene is perceived. On the other hand, the overall quality is a measure of how much the user 'likes' each image. It reflects a combination of image attributes including contrast, naturalness, sharpness, detail reproduction as well as 3D effect.

A scoring bar ranging from 0 – 10 is associated with each aspect to be evaluated; a higher score means better quality. It is a continuous grading scheme, which means rating a non-integer number such as 4.5, is permitted. For each scene, subjects are asked to rate the eight LDR images consecutively at first. Then, they are allowed to go back and perform paired comparison on images where subjects are not confident with the ratings. The combination of the two evaluation techniques, rating scale and paired comparison, will keep the experiment time-efficient and, at the same time, guarantee high accuracy of the collected data. The pace of the entire test could be fully controlled by the subject; it took about 30 minutes on average to finish one user study.

Fig. 2 Scenes used in the subjective test. Since HDR content cannot be shown on the paper or most monitors, all the images displayed above are tone-mapped using the photographic TMO.

4. RESULTS AND DISCUSSION

As described previously, our experiment consists of 64 images – 8 (TMOs) x 8 (scenes). Two aspects are evaluated for each image: 3D effect and overall quality. Below we conduct a statistical study on the data collected. To account for the fact that different people would have different standards regarding the quality scaling, prior to the main process and analysis and for each aspect evaluated, we normalize the scores as $(S_{i,j} - \mu_i) / \sigma_i$, where $S_{i,j}$ is a score from the *i*th subject and the *j*th image. The variables μ_i and σ_i denote the average value and the standard deviation over all the images for the *i*th subject, respectively.

First of all, we perform the statistical significance test. Its objective is to determine if there are significant differences between the tone-mapping methods for the two aspects, 3D effects and overall 3D quality. Hogg's one-way analysis of variance (ANOVA) [13] is applied for each aspect. We state the null hypothesis as: there is no significant distinction between the collected values for the tested tone-mapping methods. The smaller the returned *p* value, the more significant is the difference. A typical threshold of the *p* value that rejects the null hypothesis is 0.01 or 0.05. The *p* values for 3D effect and overall 3D quality are 0 and 2.4799e-5 respectively, which are far below the threshold. Therefore, statistically the tone-mapping methods evaluated in our experiment are found to have significant differences for each of the two rated aspects.

The results for the 3D effect and the overall 3D quality (with 95% confidences intervals) are shown in Fig. 3 and Fig. 4 respectively. In both figures, the results for each of the eight scenes are provided on the left while the average performance is given on the right. The x axis denotes the tone-mapping schemes, and the y axis shows the normalized quality standings.

The average results of the 3D effect and the overall 3D quality have a similar trend except for the gradient TMO. The gradient TMO produces images with such high amount of detail that the images tone-mapped by it may be perceptually unreal (see Fig. 5). This explains why it has a much lower rating in terms of overall quality, which takes into account the naturalness of an image. The other seven schemes generate more natural looking LDR content, and their performance indicates that the overall perceived 3D quality has high correlation with the 3D effect.

Fig. 5 Demonstration of unnaturalness by the gradient TMO.

All TMOs had significantly higher scores for 3D effect compared to the single exposure LDR (Fig. 4). The LDR images often have clipped regions (over- or under-exposed regions), which interfere with the 3D effect. The lack of information in these regions will make the 3D perception uncomfortable, since the viewer's brain has little evidence on how to fuse the images. This impairs the perceived 3D effect, and also lowers the overall 3D quality. This drawback of LDR capturing shows that there is significant benefit to capturing in HDR format and then performing tone-mapping, even though an LDR display is used to view the images in the end.

Another interesting observation from Fig. 3 is that the global TMOs (*Log, Comp* and *Disp*) have generally better performance in terms of a 3D effect than local TMOs (*Bi* and *Grad*) which are designed to produce images with more details. One exception is the local operator *Photo* which is rated high on average, and this is because the design of the local version of *photo* has a lot of influence from its global counterpart, and the behaviors of the two versions are quite similar. Relatively poor results can also be observed for the global operator *Phy*. This may be due to its unsatisfying behavior (which produces noticeable under- or over-exposed areas) for scenes containing ultra high dynamic range, such as Bulletin and LabWindow. Our finding that global TMOs usually outperform local TMOs for 3D effect is a bit unexpected. One might think that the greater amount of details present in the images from the local operators may help 3D fusion. The results suggest that the amount of details provided by a well designed global TMO are sufficient for the human visual system to comfortably fuse a stereo image, so long as they do not perform much clipping in bright/dark regions. Another factor that may account for the better performance of global TMOs is that, according to previous studies, Global TMOs produce more natural looking images [5]. Images with more natural appearance may strengthen monoscopic depth cues and therefore give better 3D effect. Furthermore, the heavy details from local TMOs may not perfectly match in the 3D pair since the tone mapping is applied to the left and the right views independently in our experiment. Local TMOs often amplify noise, resulting in details that do not match between the left and right, which could inhibit the perceived 3D effect. An interesting topic for future work could be designing a local tone-mapping method that jointly considers the information in the two views in order to generate images high in detail that are consistent between the views.

3D filmmakers have noted that brightness is an important factor in producing pleasant looking 3D movies [13]. Since different TMOs will produce images with different brightness levels, we examined how the average brightness of the images from a TMO affects the subjective 3D effect. Fig. 6 shows the rating of the 3D effect against the average brightness level of each TMO. Each circle in the

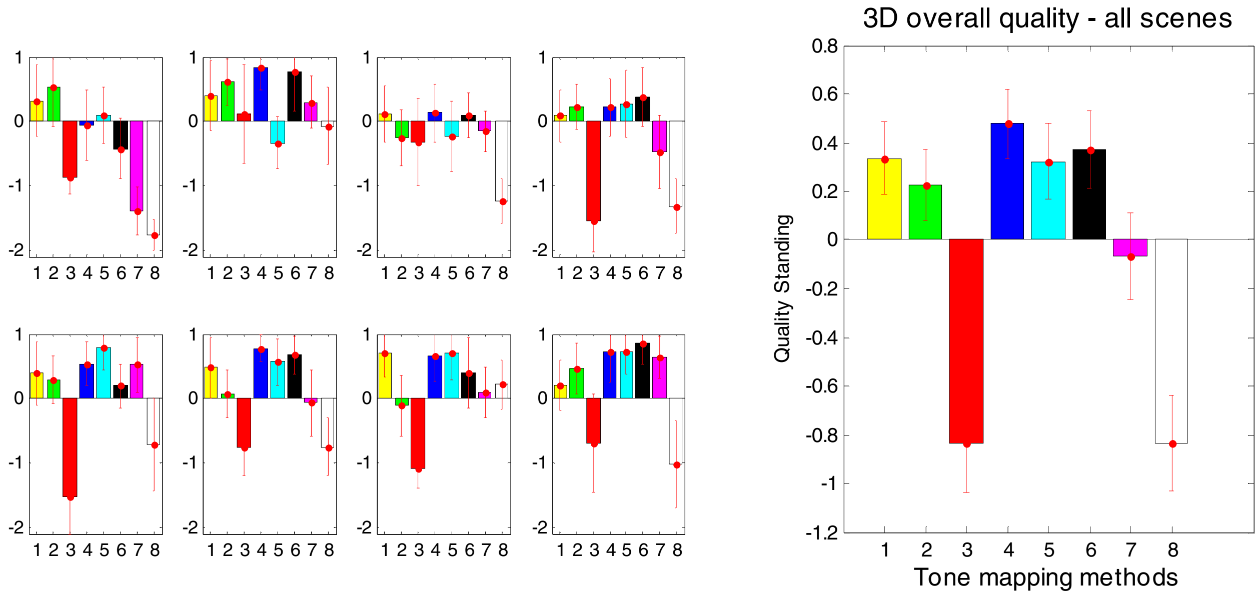

Fig. 3 Results (with 95% confidence interval) for overall 3D quality. The eight smaller figures on the left corresponds to ratings for individual scenes. The top row (from left to right): "Bulletin", "ChemEngEntrance", "ICICSBuilding" and "LabWindow"; The bottom row (from left to right): "Library", "SauderSchool", "MeetingTable" and "Stairs". The larger figure on the right shows the average results over the eight scenes. In each of the plot, the x axis denotes the quality standing: the higher the better. The y axis represents the tone-mapping schemes: (1) Adaptive logarithmic tone-mapping (*Log*), (2) Fast bilateral filtering (*Bi*), (3) Gradient domain tone-mapping (*Grad*), (4) Tone-mapping for backward-compatible HDR compression (*Comp*), (5) Display adaptive tone-mapping (*Disp*), (6) Photographic tone-mapping (*Photo*), (7) Tone-mapping by photoreceptor physiology (Phy), and (8) Captured with normal LDR cameras with a single exposure.

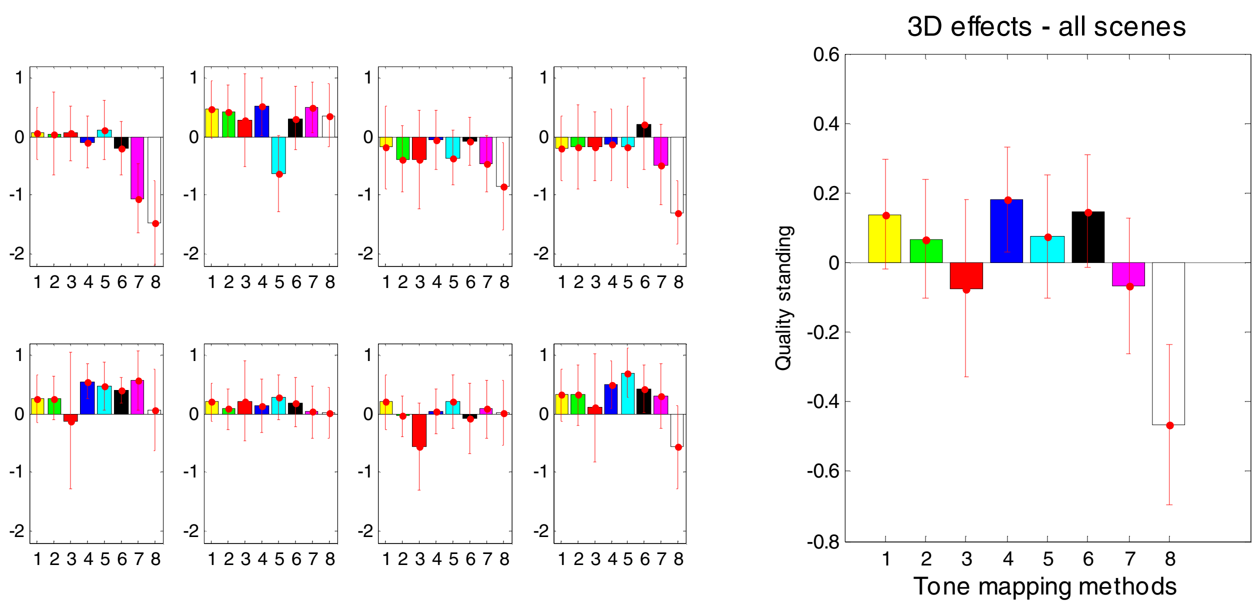

Fig. 4 Results (with 95% confidence interval) for 3D effect. Notations are the same as Fig.3.

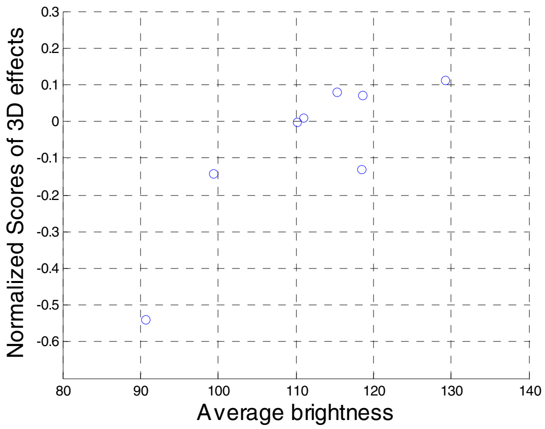

Fig. 6 Relationship between the 3D effect versus average brightness value. Each circle in the figure corresponds to a tone-mapping method, and its brightness value (x coordinate) and the score (y coordinate) are averaged over the eight LDR images produced using that method.

figure corresponds to one TMO; the brightness value (x coordinate) and the 3D effect score (y coordinate) are averaged over the eight LDR images. There is a strong correlation between the average brightness and 3D effect score, with brighter images giving better 3D effect. This suggests that quality of a 3D image can be improved by using a tone mapping that gives a brighter image. Future work will examine this issue more closely.

5. CONCLUSIONS

This paper presents a subjective evaluation of tone mapping operators on 3D images. We evaluated seven tone-mapping schemes, as well as LDR capturing, in two aspects: 3D effect and overall 3D quality. The results indicate that global TMOs are generally preferred to local TMOs, despite the fact that they produce images with fewer details. All of the tested TMOs had significantly higher scores for 3D effect than LDR capturing, which shows that capturing in HDR and then applying tone mapping produces better 3D images than direct LDR capturing. We also found that the average brightness of the resulting tone-mapped image plays an important role in providing a good 3D effect and hence a good overall 3D quality.

6. REFERENCES

[1] R. Mantiuk, G. Krawczyk, K. Myszkowski, and H.-P. Seidel, "Perception-motivated high dynamic range video encoding," *ACM Transactions on Graphics*, vol. 23, no. 3, pp.730–738, 2004.

[2] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 21, no. 3, pp. 267–276, 2002.

[3] R. Mantiuk, S. Daly, and L. Kerofsky, "Display Adaptive Tone Mapping," *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 27, no. 3, pp. 68–68, 2008.

[4] F. Drago, W. L. Martens, K. Myszkowski, and H.-P. Seidel, "Perceptual Evaluation of Tone Mapping Operators," In *ACM Proc. SIGGRAPH 2003 Conference on Sketches and Applications*, pp. 1, 2003.

[5] A. Yoshida, V. Blanz, K. Myszkowski, and H.-P. Seidel, "Perceptual Evaluation of Tone Mapping Operators with Real-World Scenes," *SPIE: Human Vision & Electronic Image X*, San Jose, CA, USA, pp. 192-203, 2005.

[6] J. Kuang, H. Yamaguchi, C. Liu, G. M. Johnson, and M. D. Fairchild, "Evaluating HDR Rendering Algorithms," *ACM Transactions on Applied Perception*, vol. 4, no. 2, article 9, 2007.

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

[8] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive Logarithmic Mapping for Displaying High Contrast Scenes," *Computer Graphics Forum (Proc. of Eurographics)*, vol. 22, no. 3, pp. 419–426, 2003.

[9] F. Durand, and J. Dorsey, "Fast Bilateral Filtering for the Display of High-Dynamic-Range Images," *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 21, no. 3, pp.257–266, 2002.

[10] Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, R. Ward, W. Heidrich, " Optimizing a Tone Curve for Backward-Compatible High Dynamic Range Image/Video Compression ", *IEEE Transactions on Image Processing (TIP)*, accepted, 2010.

[11] R. Fattal, D. Lischinski, and M. Werman, "Gradient Domain High Dynamic Range Compression," *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 21, no. 3, pp.249–256, 2002.

[12] E. Reinhard, and K. Devlin, "Dynamic Range Reduction Inspired by Photoreceptor Physiology," *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, no. 1, pp. 13-24, 2005.

[13] B. Mendiburu, "3D Movie Making – Stereoscopic Digital Cinema from Script to Screen." Elsevier, 2008.

[14] G. Jones, D. Lee, N. Holliman, and D. Ezra, "Controlling Perceived Depth in Stereoscopic Images," Proc. SPIE, Stereoscopic Displays and Virtual Reality Systems vol. 4297, pp. 42–53, 2001.

[15] M. Cadik, M. Wimmer, L. Neumann, and A. Artusi, "Evaluation of HDR tone-mapping methods using essential perceptual attributes," Jounral of Computer and Graphic, vol.32, pp. 330-349.

[16] R. V. Hogg, and J. Ledolter, "Engineering Statistics," New York: MacMillan, 1987.