Spatio-Temporal Tone Mapping Operator based on a Retina model

Alexandre Benoit, David Alleysson, Jeanny Hérault, Patrick Le Callet

▶ **To cite this version:**

Alexandre Benoit, David Alleysson, Jeanny Hérault, Patrick Le Callet. Spatio-Temporal Tone Mapping Operator based on a Retina model. CCIW 2009 - Computational Color Imaging Workshop (CCIW'09), Mar 2009, Saint Etienne, France. pp.12-22. hal-00366126

HAL Id: hal-00366126
https://hal.science/hal-00366126

Submitted on 28 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Spatio-Temporal Tone Mapping Operator based on a Retina model

Alexandre Benoit[1*], David Alleysson[2], Jeanny Herault[3], Patrick Le Callet[4]

[1]LISTIC Annecy le Vieux, [2]LPNC Grenoble, [3]Gipsa Lab Grenoble, [4]IRCCyN Nantes, France
*alexandre.benoit@univ-savoie.fr

Abstract

From moonlight to bright sun shine, real world visual scenes contain a very wide range of luminance; they are said to be High Dynamic Range (HDR). Our visual system is well adapted to explore and analyze such a variable visual content. It is now possible to acquire such HDR contents with digital cameras; however it is not possible to render them all on standard displays, which have only Low Dynamic Range (LDR) capabilities. This rendering usually generates bad exposure or loss of information. It is necessary to develop locally adaptive Tone Mapping Operators (TMO) to compress a HDR content to a LDR one and keep as much information as possible. The human retina is known to perform such a task to overcome the limited range of values which can be coded by neurons. The purpose of this paper is to present a TMO inspired from the retina properties. The presented biological model allows reliable dynamic range compression with natural color constancy properties. Moreover, its non-separable spatio-temporal filter enhances HDR video content processing with an added temporal constancy.

Key works: High Dynamic Range compression, tone mapping, retina model, color constancy, video sequence tone mapping.

1 Introduction

In this paper, we propose a method to compress High Dynamic Range images in order to make visual data perceptible on display media with lower dynamic range capabilities. HDR images are our real life visual world; our eyes perceive everyday a wide variety of visual scenes with really different luminance values. Our visual system is able to cope with such a wide variety of input signals and extract salient information. However, we can notice that, as discussed in [1], neurons cannot code a wide range of input values. Thus, at the retina level a compression process occurs in order to preserve all relevant information in the coding process, including color.
A similar idea has been recently introduced with digital High Dynamic Range imaging. It is now possible to create HDR images even with standard digital cameras [2] or light simulation [3], nevertheless the development of HDR displays, which

would be able to render all the acquired data is still under development [4]. Current displays are Low Dynamic Range and direct HDR image visualization would hide a large part of the information. An alternative is the development of Tone Mapping Operators [5, 6] which allow HDR images to be rendered on standard LDR displays and preserve most of the information to be seen. Nevertheless, as discussed in [7] a HDR image cannot be perceived similarly to its LDR version. Human factors related to this problem are not already known but some artifacts created by the tone mapping conversion can already be measured [8]. They are related to halo effects and color distortions and lead to naturalness corruption. Nevertheless, studies on the quality perception [5, 6, 9] are the only current assessment solutions.
The current challenge is to design a TMO able to limit artifacts and preserve the general ambiance of the original HDR visual scene. Several operators have already been proposed and compared [6, 9] and lead to a wide variety of approaches. From computer vision methods to the one inspired by the visual system, each TMO presents a different approach and requires specific parameters set for each processed image. In addition, a new challenge is related to video sequence processing and has not yet been explored. The aim is to generate successive tone mapped images which allow a natural perception sensation without temporal instabilities created by frame-by-frame image optimization.
In this paper, we propose a new TMO based on a retina model. The approach models retina local adaptation properties, as described in Meylan *et al*. work [10] and is completed by specific spatio-temporal filters of the retina. The added contribution involves retina processes which enable spatio-temporal noise removal, temporal stability introduction and spectral whitening. The paper is presented as follows: section 2 describes the proposed retina model and its properties for HDR compression. Section 3 illustrates the effect of such a filter in the case of static and dynamic content processing.

2 Retina biological model

The human retina architecture is based on cellular layers which process the visual information from the photoreceptors visual data entry point to the ganglion cells output. The input signals are locally processed step by step so that details, motion and color information are enhanced and conditioned for high level analysis at the visual cortex level. Here, we focus on the known parts of the human retina, which are suitable for a Tone Mapping Operator design. The aim is to show that tone mapping is already performed in low level vision so that higher level visual tasks are facilitated. Furthermore, modeling these early vision properties leads to a fast and efficient TMO. We choose to work with the model described in [1], which takes into account the different low level processes occurring in the retina. We particularly focus on the foveal vision area and its output called Parvocellular channel, which brings the details and color information to the central nervous system. The aspects taken into account are detailed in the following:

- Local luminance and local contrast adaptation at the photoreceptor and ganglion cells levels. This biological property is directly linked to our dynamic range compression topic.
- The spatio-temporal filtering occurring at the Outer Plexiform Layer level (OPL). This filtering allows input image frequency spectrum to be whiten and enhances image details. Moreover, its temporal properties allow noise reduction and temporal stability.
- Color sampling: the input image is spatially sampled by sensors with different color sensitivities. Our TMO allows gray scale and color images to be processed in the same way and introduces color constancy properties.

The architecture of the proposed model follows the biological model architecture and is depicted on figure 1.

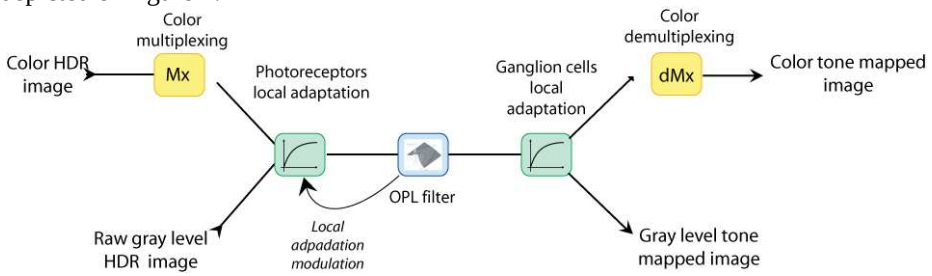

Fig. 1: simplified model of the proposed retina model. Color processing is an additional processing, which requires preliminary input image multiplexing and output picture demultiplexing.

The input image can be either a raw gray image or a multiplexed color one. Then, color processing appears as an additional processing, which consists of preliminarily multiplexed color information followed by a filtering stage, then a demultiplexing stage. The key-point of the model is actually the two local adaptation steps, corresponding respectively to photoreceptors and ganglion cells, and the OPL filter placed between them. As discussed in [1], the photoreceptor's local adaptation is modulated by the OPL filter.

2.1 Local adaptation

Photoreceptors are able to adjust their sensitivity with respect to the luminance of their spatio-temporal neighborhood. This is modeled by the Michaelis-Menten [1] relation, which is normalized for a luminance range of [0, *Vmax*] (Eq. 1). *Vmax* represents the maximum pixel value in the image-255 in the case of standard 8 bits images. It may vary greatly in case of High Dynamic Range images.

$$C(p) = \frac{R(p)}{R(p) + R_0(p)} \left(V_{\max} + R_0(p)\right) \quad (1)$$

$$R_0(p) = V_0 . L(p) + V_{\max}\left(1 - V_0\right) \quad (2)$$

In this relation, the response *C(p)* of photoreceptor *p* depends on the current excitation *R(p)* and on a compression parameter $R_0(p)$ which is linearly linked to the local luminance *L(p)* (cf. Eq.2) of the neighborhood of the photoreceptor *p*. This local luminance *L(p)* is computed by applying a spatial low pass filter to the input image. This low pass filtering is actually achieved by the horizontal cells network as presented later on. Moreover, in order to increase flexibility and make the system more accurate, we add in $R_0(p)$ the contribution of a static compression parameter V_0 of value range [0;1]. Compression effect is reinforced when V_0 tends to *1* or is attenuated when reaching *0*. Photoreceptors V_0 value is set to 0.7 as a generic experimental value.
Figure 2 shows the evolution of sensitivity with respect to $R_0(p)$ and illustrates the effect of such a compression on a back-lit picture. Sensitivity is reinforced for low values of $R_0(p)$ and is kept linear for high values. As a result, this model enhances contrast visibility in dark areas while maintaining it in bright areas.

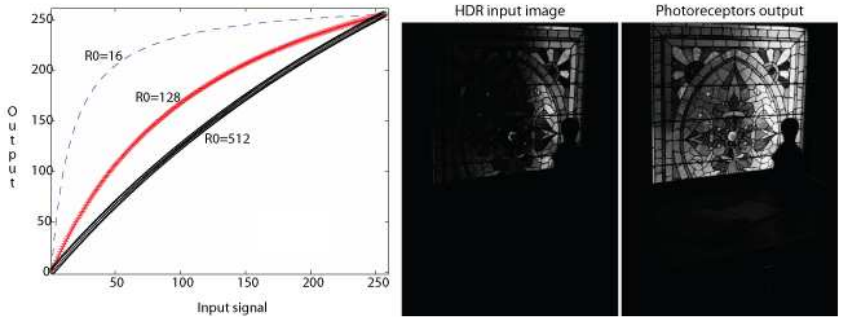

Fig. 2: photoreceptors local luminance adaptation. Left: output response with regard to the local R0 (p) value. Right: effect on a HDR image (from: www.openexr.org).

2.2 OPL: Spatio-temporal filtering and contour enhancement

The cellular interactions of the OPL layer can be modeled with a non-separable spatio-temporal filter [1] whose transfer function for 1D signal is defined in eq. (3) where *fs* and *ft* denote respectively spatial and temporal frequency. Its transfer function is drawn on figure 3.a. This filter can be considered as a difference between two low-pass spatio-temporal filters which model the photoreceptor network *ph* and the horizontal cell network *h* of the retina. As discussed in [10, 1], the output of the horizontal cells network (F_h) is very low spatial frequency limited and can be interpreted as the local luminance *L(p)* required by the photoreceptors local adaptation step (eq. 2). Moreover, F_h filter's temporal low pass effect allows local luminance computation to be temporally smoothed. Finally, as a general rule, global F_{OPL} filter has a spatio-temporal high-pass effect in low frequencies which results in a spectral whitening of the input. Its high-frequency low-pass effect enables the removal of the structural noise.

$$F_{OPL}(f_s, f_t) = F_{ph}(f_s, f_t) \cdot \left(1 - F_h(f_s, f_t)\right)$$
$$\text{with} \quad F_i(f_s, f_t) = \frac{1}{1 + \beta_i + 2\alpha_i\left(1 - \cos(2\pi f_s)\right) + j2\pi\tau_i f_t} \qquad (3)$$

with subscript *i=ph* or *i=h*.

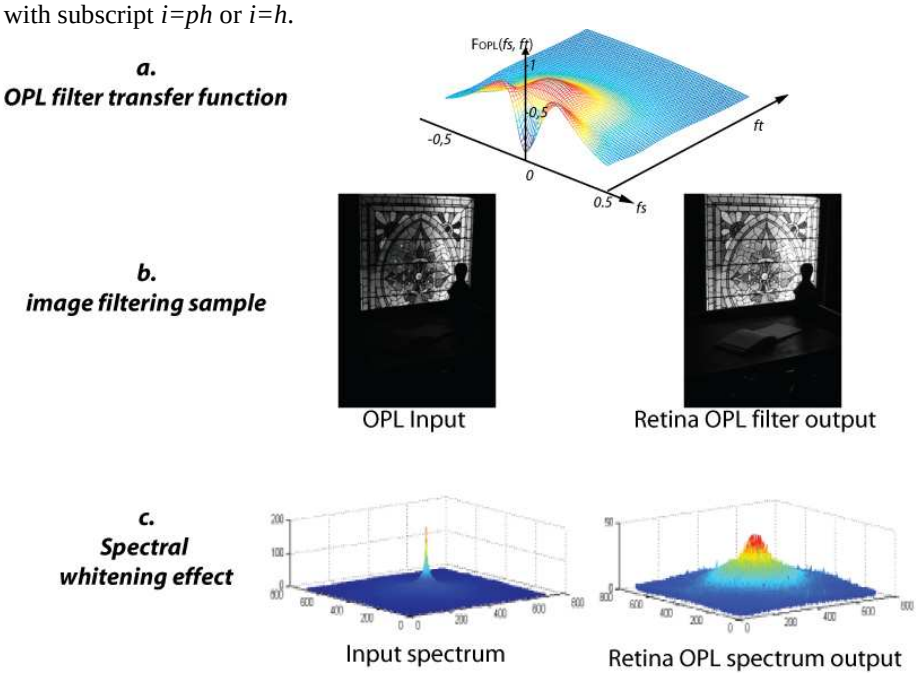

Fig. 3: OPL filter transfer function and illustration of its effect on an image and its spectrum (input image is the output of the photoreceptors previous stage).

β_{ph} is the gain of filter F_{ph}. Setting β_{ph} to 0 cancels luminance information and a higher value allows luminance to be partially processed; β_{ph} is typically set to 0.7 which allows a good general effect. Parameters α_i and τ_i stand respectively for the space and time constants of the filters. Temporal noise is minimized when τ_{ph} is increased. Besides, the higher τ_h the slower temporal adaptation. α_{ph} spatial filtering constant should be remained low (close to 0) in order to preserve high spatial frequencies. α_h allows local luminance to be extracted; its generic value is set to a space constant of 7 pixels.

One of the most relevant effects is spectral whitening, which compensates for the 1/f spectrum of natural images as shown in figure 3.b-c. It happens on a limited portion of the spatial frequency spectrum.

2.3 Retina ganglion cells final dynamic compression step

The ganglion cells of the Parvocellular channel receive the information coming from the OPL. They act as local contrast enhancers and are modeled by the Michaelis-Menten law similar to the photoreceptors as discussed in [11] but with different parameters. Indeed, the local luminance value is related to a smaller area around each cell since receptor fields are smaller. Also, as a general rule, the compression effect is more powerful at the ganglion cells level. The parameter V_0 is then higher than the one of photoreceptors and is typically set to 0.9. Figure 4 shows results of this local

adaptation on the OPL's output stage. This last filtering allows the final luminance compression of originally very dark areas.

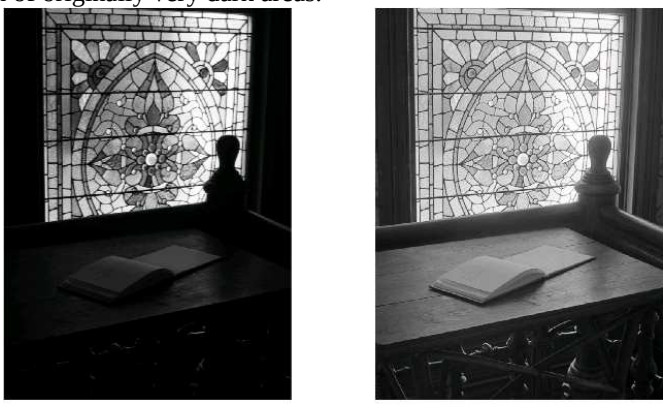

Fig.: Effect of the ganglion cells local adaptation (input image is the output of the OPL previous stage)

2.4 Color processing

To deal with color, it is possible to take advantage of the photoreceptors color sampling properties. As described in [10], since in the fovea photoreceptors (mainly cones) sample the visual scene with three different sensitivities (Long, Medium and Short wavelengths: with L, M and S cones), the spectrum of the color sampled image presents special properties. Luminance spectrum is centered on the low frequencies while color information spectrum is located in higher frequencies. Modeling such properties consists in multiplexing the color image before processing it with the previously presented retina model and demultiplexing it afterwards, as presented in figure 1. As a consequence, designing a TMO with such an approach while preserving high spatial frequencies allows luminance map to be toned and color information to be kept.

Fig. 5: color multiplexing and demultiplexing on the TMO “boundaries”.

Moreover, from a computational cost point of view, processing color only requires an additional color multiplexing/demultiplexing stage to the gray level processing core. Towards this end, we propose to use the color demultiplexing algorithm presented in [12] which supports different color sampling methods (Bayer, Diagonal, Random). Following the TMO scheme depicted on figure 1, figure 5 shows the color

management steps of the algorithm. Using Bayer color sampling, color spectrum is translated in high spatial frequencies. Thus, since OPL filter with a low value of α_{ph} keeps those frequencies, color information remains unchanged.
To sum up, the human retina naturally achieves color tone mapping and its simplified model allows digital HDR images to be processed in the same way. Notice that this current application is suitable to deal with high dynamic range in terms of luminance; issues related to gamut limitation are not considered.

3 Design of a retina based TMO

As shown in the previous section, the retina model includes many properties required for low level image tone mapping. Indeed, its preliminary goal is to condition visual signals in order to make them compliant with the subsequent image analysis "operators". In addition to its image processing properties, this model has the ability to present a low computational cost: the gray level processing part requires only 16 operations per pixel. Color management depends on the chosen demultiplexing method -here it requires less than 200 linear operation per pixels.
Consequently, we follow the biological model and use all the tools described on figure 1. Actually, the proposed TMO is rather similar to the one proposed by Meylan & al. [10] but it adds the contribution of the OPL filter which achieves a more complete retina model, taking into account temporal aspects. The main advantage is that the OPL intermediate filter allows spectral whitening and temporal information management. For simplification, we first describe its properties for static visual scenes and then deal with video sequences processing.

3.1 Static content

Even if human factors related to HDR imaging are not well known, some critical points have been reported in literature. A brief overview of the main issues and the response introduced by our method are presented in the following. Refer to figure 6 for image samples.

- The first point is related to artifacts generation in the tone mapping process: since luminance is compressed, algorithms have a tendency to generate halos and contours around the boundaries of high luminance gradient of visual scenes [5, 13]; e.g, sun and lights boundaries. Our algorithm limits such artifacts since low spatial frequency energy is lowered and logarithmic compression of photoreceptors is very local (typically 7-pixel wide).
- Color constancy is the second challenge. It is often impaired when color components are manipulated independently [2]. Due to multiplexing stage, our method naturally remains color information in high spatial frequencies and do not change them. Furthermore, objects color information is preserved, independently of the illuminant color [1].
- Natural rendering is the general issue for image tone mapping. Even though details extraction in bright and dark areas of a visual scene is the final goal, it can change the image initial balance and reduce its dynamic [2, 14, 6]. Global ambiance luminance should be preserved in order to maintain the initial « goal of

the image » and the areas at which the photographer wants the observers to look at. Concerning that specific issue, our algorithm allows luminance ambiance to be less compressed by lowering the value of photoreceptors V_0 parameter. It is a matter of trade-off: either luminance ambiance survival or all details extraction.

In addition, it is generally difficult to identify a unique TMO parameters setting which would allow many different images to be tone mapped and look "natural". Each TMO algorithm generally has its own parameters which should be optimized for each image, resulting in a supervised tone mapping process. However, some algorithms already present good parameters stability and generate natural-looking images with a large set of pictures, e.g. the Meylan *et al.* [10], Mantiuk's [15] algorithm and our contribution.

Figure 6 gives examples of tone mapping results with the proposed algorithm and other results coming from [10], another human visual system based TMO [15], which includes visual cortex models but less retina properties, and Reinhard's TMO [2], which is more focused on local adaptation retina properties. Note that generic default values were used as unique parameter set for each method. On these examples, color constancy is achieved for both [10] and our method since they manage color in a similar way. Meylan's method compresses luminance to a larger extent, which allows details to be more visible. In parallel, the proposed method keeps better global luminance ambiance and provide minimal halo effects. As a compromise, dark details may remain hidden because the variance of $R_0(p)$ is less due to an additional temporal filtering. While comparing all methods, we can see how different results can be from one algorithm to the other. Actually, apart from constancy problems, naturalness and artifacts limitation versus details extraction trade off is the most difficult to reach. Also, our method, [10] and [15] appear in these samples as optimal TMOs with a constant parameters set. However, only future visual quality tests will allow rigorous comparisons to be made.

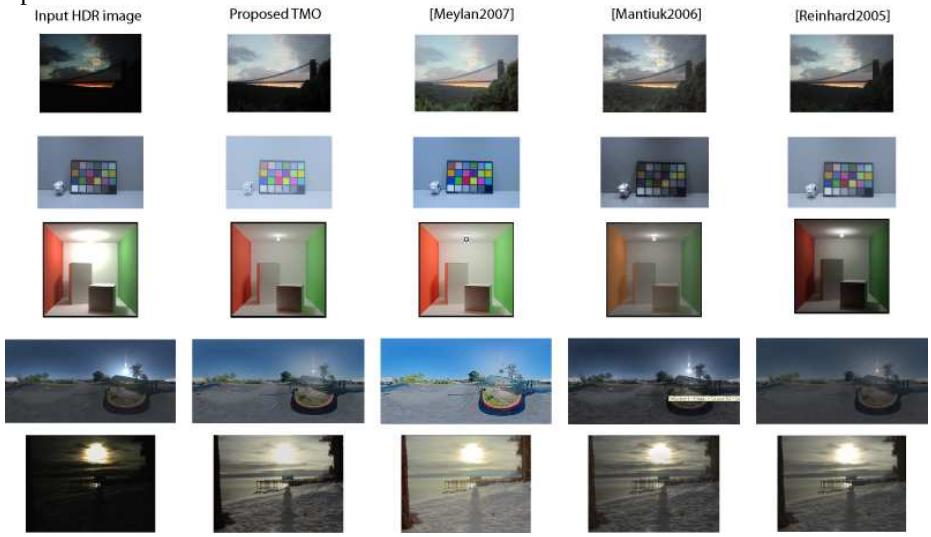

Fig. 6 : tone mapping results comparison between our approach and other existing methods (refer to [16] for higher resolution results). Our algorithm does not generate halos (refer to the two last examples), nevertheless, dark areas' details can be hidden (refer to the first two

examples). The other compared algorithms can better extract details in dark areas but halos or color distortion can appear. Also, locked parameters and accurate tone mapping are not supported by all the algorithms presented in this sample.

3.2 Video content

As mentioned in [17], TMO encounter critical problems when visual content changes with time. Indeed, from frame-to-frame critical light changes can appear and since current TMO are independently optimized for each frame, the reconstructed video sequence can be degraded by successive uncorrelated tone mapping operations. This problem is even more difficult to overcome when dealing with global image TMO, which generate a tone mapped image depending on the global image luminance. This problem is more limited in the case of local operators such as [10, 13] and our method. However, video content tone mapping is a very recent challenge and it is not already possible to compare results due to the lack of real HDR video contents. Here, we propose in figure 7 some samples of a HDR video sequence shoot to illustrate some properties of our method. Sudden light changes do not impact global luminance of the tone mapped images.

Fig.7: HDR video sequence tone mapping results samples (refer to [16] for higher resolution results). In this test sequence, the sun generates optic halos on the camera, the most important changes are related to sudden sun disappearing and appearing depending on camera position and objects on the visual scene. Our method allows details to be constantly extracted whatever luminance change there are. Moreover, colors and image dynamic remain stable.

Our TMO presents temporal constants, which acts on local luminance processing adaptation. Particularly, the spatio-temporal horizontal cells low pass filter of the OPL filter computes the local luminance with a memory effect. This results in smooth transitions when luminance changes suddenly. Also, the photoreceptors stage allows spatio-temporal noise removal at the beginning of the tone mapping process.

Conclusion

This paper proposes a Tone Mapping Operator (TMO) based on a human retina model. It imitates some parts of the foveal retina functionalities including its luminance compression properties and adds the contribution of temporal information processing. This “biologic” algorithm is able to process a wide variety of images with natural rendering and has a good potential for video sequences processing. Moreover, the method presents good color constancy properties and its color management flexibility enables easier practical implementation.

Future work could aim at identifying human factors, which would allow to identify the main clues of the most appropriate TMO and adjust its parameters in that sense, as

well as visual perception tests in order to further assess the subjective quality of the Tone Mapping operators.

Acknowledgement

This work was supported by FuturIm@ge project within the “Media and Network” French Cluster. Also, we thank Grzegorz Krawczyk from MPI Informatik who gave us access to these rare HDR video sequences.

References

[1] Hérault J., Durette B.:Modeling Visual Perception for Image Processing, *F. Sandoval et al. (Eds.): IWANN 2007*, LNCS 4507, 662–675, Springer-Verlag Berlin Heidelberg. (2007).
[2] Reinhard E., Ward G., Debevec P., Pattanaik S.: *High Dynamic Range Imaging: Acquisition, Display, and Image Based Lighting*, Morgan Kaufmann, San Francisco, (2005).
[3] Ward, G.: 'The RADIANCE Lighting Simulation and Rendering System''Computer Graphics Proceedings, Annual Conference Series', (1994).
[4] Hoefflinger, B.: *High-Dynamic-Range (HDR) VisionMicroelectronics, Image Processing, Computer Graphics*, Springer, (2007).
[5] Ahmet, O. A. ; Erik, R.: Perceptual Evaluation of Tone Reproduction Operators Using the Cornsweet-Craik-O'Brien Illusion, *ACM Transactions on Applied Perception* **4**(4), 1—29, (2008).
[6] Jiangtao, K., Hiroshi, Y., Changmeng, L., Garrett, M. J., Mark, Fairchild .D. : 'Evaluating HDR rendering algorithms', *ACM Trans. Appl. Percept.* **4**(2), 9, (2007).
[7] Johnson, G.: Cares and concerns of CIE TC8-08 : Spatial appearance modeling and HDR rendering, *in* 'SPIE proceedings series, Image quality and system performance', pp. 148-156, (2005).
[8] Tunç O. Aydin, Mantiuk R., Myszkowski K., Seidel H-P.: Dynamic Range Independent Image Quality Assessment. To appear in: ACM Transactions on Graphics (Proc. of SIGGRAPH'08) 27 (3), (2008)
[9] Kuang J., Yamaguchi H., Liu C., Johnson G.M., Fairchild M.D.: Evaluating HDR rendering algorithms," *ACM Transactions on Applied Perception, 4* Article 9 (2007).
[10] Meylan, L.; Alleysson, D., Susstrunk, S.: A Model of Retinal Local Adaptation for the Tone Mapping of Color Filter Array Images', *Journal of the Optical Society of America A (JOSA A)* 24(9), 2807—2816, (2007).
[11] Smirnakis S.M., Berry M.J., Warland D.K., Bialek W., Meister M.: Adaptation of Retinal Processing to Image Contrast and Spatial Scale *Nature vol. 386, 69-73*, (1997).
[12] Chaix de Lavarène B., Alleysson D., Hérault J. : Practical Implementation of LMMSE Demosaicing Using Luminance and Chrominance Spaces. Computer Vision and Image Understanding, Volume 107, Issues 1, Pages 3-13 (July 2007).
[13] Mantiuk R., Daly S., Kerofsky L.: Display Adaptive Tone Mapping. To appear in: ACM Transactions on Graphics (Proc. of SIGGRAPH'08) 27 (3), (2008).
[14] Yoshida, A.. Blanz, V.. Myszkowski, K.. Seidel, H.: Perceptual Evaluation of Tone Mapping Operators with Real-World Sceness, *in* Bernice E. Rogowitz; Thrasyvoulos N. Pappas & Scott J. Daly, ed.,'Human Vision and Electronic Imaging X, IS&T/SPIE's 17th Annual Symposium on Electronic Imaging (2005)', SPIE, San Jose, USA, pp. 192—203, (2005).
[15] Mantiuk R., Myszkowski K., Seidel H.-P.: A Perceptual Framework for Contrast Processing of High Dynamic Range Images (revised and extended version). In ACM Transactions on Applied Perception 3, 3, pp. 286-308, (2006).
[16] proposed TMO results at *http://benoit.alexandre.vision.googlepages.com/HDR*
[17] Didyk P., Mantiuk R., Hein M., Seidel H.-P.: Enhancement of Bright Video Features for HDR Displays. To appear in: Computer Graphics Forum (Proc. of EGSR'08) (2008).