Virtual Reality (2013) 17:253–262
DOI 10.1007/s10055-012-0215-4

SI: EVALUATING VIRTUAL WORLDS

Evaluation of tone mapping operators in night-time virtual worlds

Josselin Petit · Roland Brémond · Ariane Tom

Received: 10 August 2011 / Accepted: 19 September 2012 / Published online: 30 September 2012
© Springer-Verlag London 2012

Abstract The subjective quality of a virtual world depends on the quality of displayed images. In the present paper, we address a technical aspect of image quality in virtual environments. Due to the recent development of high dynamic range (HDR) imaging in computer graphics applications, tone mapping operators (TMO) are needed in the graphic pipeline, and their impact on the final image quality needs to be tested. Previous evaluations of such operators have emphasized the fact that the specific merit of a given operator may depend on both the scene and the application. The dynamic behavior of tone mapping operators was not tested before, and we have designed two psychophysical experiments in order to assess the relevance of various TMO for a specific class of virtual worlds, outdoor scenes at night and an interactive application, to explore an outdoor virtual world at night. In a first experiment, 5 HDR video clips were tone-mapped using 8 operators from the literature, resulting in 40 videos. These 40 videos were presented to 14 subjects, which were asked to rate their realism. However, the subject's evaluation was not a direct comparison with the HDR videos. In a second experiment, 9 HDR photographs of urban scenes at night were tone-mapped with the same 8 operators. The resulting 72 photographs were presented to 13 subjects, at the location where the photographs were taken. The subjects were asked to rate the realism of each tone-mapped image, displayed on a laptop, with respect to the physical scene they experienced. The first experiment emphasized the importance of modeling the temporal visual adaptation for a night-time application.

Keywords Virtual environments · Image rendering · Tone mapping · Presence · Subjective evaluation

1 Introduction

Realistic rendering of virtual worlds is a key issue in many interactive applications, such as video games, but also in the movie industry, where computed images are displayed at 25 Hz. In a previous paper, we have addressed the realism of rendering techniques in virtual environments in two steps: a high dynamic range (HDR) physical simulation of lighting is followed by a tone mapping (Petit and Brémond 2010), see Fig. 1. HDR refers to the luminance range in the virtual world, which is in most cases impossible to render on usual display devices (low dynamic range, LDR displays). The first step focuses on physical realism (photometry and colorimetry), while the second step focuses on perceptual realism (Ferwerda 2003).

The subjective quality of a virtual world is a multi-factor issue. Due to the recent development of HDR imaging in computer graphics applications, tone mapping operators (TMO) have been developed in order to map HDR computed images into the low dynamic range (LDR) of usual displays. These TMO operators are more and more needed in the graphic pipeline. TMO were introduced by Tumblin and Rushmeier in 1993 to the field of computer graphics (Tumblin and Rushmeier 1993). In a first decade, many operators were proposed to address the issue of displaying HDR images onto low dynamic range display devices (Reinhard et al. 2005). However, no "killer" operator emerged, with better performance whatever the images and whatever the display devices. Thus, tone mapping evaluation became an important field of research (Drago et al. 2003; Kuang et al. 2004, 2005; Ledda et al. 2005; Yoshida

---

J. Petit · R. Brémond (✉) · A. Tom
Université Paris Est, IFSTTAR, IM, LEPSiS, Paris, France
e-mail: roland.bremond@ifsttar.fr

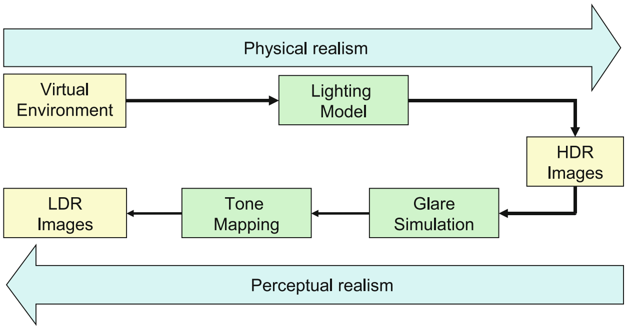


Fig. 1 Proposed approach for a realistic rendering in virtual environments: A physically based HDR rendering is followed by a perceptual tone mapping

et al. 2005; McNamara 2006; Ashikhmin and Goyal 2006; Kuang et al. 2007; Čadík et al. 2008; Grave and Brémond 2008). Whereas previous studies were restricted to still images, we considered in the present paper TMO performance in dynamic situations, relevant for movies and interactive applications. Thus, we assess tone mapping operators in terms of subjective realism of the rendered images in virtual environments, which contributes to the subject's immersion feeling.

2 Previous work

2.1 Tone mapping operators

Many TMO have been proposed since 1993, most of them being reviewed in Reinhard et al.'s (2005) book. They have been classified as global/local: global algorithms map the entire image with the same function [e.g., Ward's operator (Ward 1994)], while local operators consider local adaptation and may convert two pixels with the same intensity values in the HDR image to different intensity values in the displayed image, depending on the pixel neighbor [e.g., Pattanaik et al. (1998)]. These operators may also be classified as static versus dynamic, although very few model the temporal visual adaptation (Pattanaik et al. 2000; Durand and Dorsey 2000; Ledda et al. 2004; Irawan and Marschner 2005). They may also be classified depending on their performance (global operators tend to be faster than local ones), or depending on their field of inspiration. For instance, Reinhard proposed a now classical operator inspired by photographic art (Reinhard et al. 2002) and later another one which mimic the photobiology of retinal sensors (Reinhard et al. 2005).

2.2 Tone mapping evaluation

After a decade when dozens of TMO were published, these operators have been assessed with psychophysical methodologies. Two main methodologies are used. The first one compared, without visual HDR reference, various tone-mapped images computed from the same HDR image, with rating, scaling, or pair comparison techniques. The second approach directly compares the tone-mapped images to HDR stimuli. These HDR stimuli may be physical scenes, providing that the HDR images were built from photographs of these scenes (Debevec and Malik 1997), or the HDR images displayed on HDR display devices (Seetzen et al. 2004).

2.2.1 TMO evaluation without reference

Drago et al (2003) were the first to perform a study in which tone mapping operators were evaluated in terms of visual quality. Then, Kuang conducted in-deep researches on TMO evaluation (Kuang et al. 2005, 2006, 2004, 2007). In one experiment, a paired comparison experiment allowed to compare 8 operators (Kuang et al. 2004). Eight HDR images were tone-mapped and displayed to 30 subjects, focusing on color and luminance issues. The tested operators led to the same results when applied to grayscale versus color images. Durand and Dorsey's operator (2002) and Reinhard's operator (2002) were rated best. In another experiment, the authors compared the overall image preference with preferences for six image attributes (Kuang et al. 2005). They showed that shadow details, overall contrast, sharpness, and colorfulness have high correlations with the overall preference.

2.2.2 TMO evaluation with an HDR displayed reference

The development of HDR display devices, using LED backlights behind a liquid crystal display (LCD) pannel (Seetzen et al. 2004), opened a new field of research for tone mapping evaluation. Ledda et al. (2005). ran an evaluation of 6 operators, comparing tone-mapped images displayed on a LCD to HDR reference images displayed on an HDR display device. Subjects had to choose, from two tone-mapped images, the one they felt the closest to the HDR reference. The relative merit of these operators could thus be assessed: a good operator should produce an LDR image the subject finds close to the reference HDR image.

Grave and Brémond (2008) have compared 5 tone mapping operators, using a digital-light processing (DLP) video-projection as an HDR reference, compared to a LCD display as a LDR display device. Unlike previous evaluations, they focused on visual performance rather than subjective judgment. Three HDR images were displayed in LDR and HDR to 30 subjects, who were asked to detect a short (100 ms) gap in a Landolt ring displayed at the center of the images. Two hundred Landolt ring tests were presented for each displayed image, in order to accurately compute the subject's detection performance. The TMO proposed in the paper was rated best for this task, together with Pattanaik's operator, which mimics the multi-scale behavior of the human visual system (Pattanaik et al. 1998).

Yoshida et al. (2006) analyzed the reproduction of HDR images on displays of varying luminance range, without actually rating the respective merit of each operator. Akyuz et al. (2007) investigated how LDR images are best displayed on HDR monitors. They showed that HDR displays outperform LDR ones in terms of subjective preference of the observer, which was expected. More surprisingly, tone-mapped HDR images were not better than the best single LDR exposure.

2.2.3 TMO evaluation with an HDR physical reference

Yoshida et al. (2005) were the first to use real-world scenes as HDR reference. Their experimental protocol used HDR photographs of 2 indoor environments (Robertson et al. 1999) instead of computer graphics images. Fourteen observers were asked to compare the physical scenes to tone-mapped images displayed on a LCD display device. The comparison was made in terms of appearance (overall brightness, contrast, detail reproduction in dark and bright regions) and realism. Finally, the scores were normalized over each attribute and each subject. Brightness and contrast were found to be the two main components of visual appearance, and surprisingly, the best operator with respect to brightness was the linear one. Drago's algorithm (2003) led to good results, Ward's operator too, except for the reproduction of details.

Kuang et al. (2006) used 3 indoor scenes and 19 subjects to evaluate 7 operators. Using paired comparisons, the authors evaluated image contrast, colorfulness, and overall accuracy. The results showed that an operator based on bilateral filtering (Durand and Dorsey 2002) outperformed the other algorithms.

Čadík et al. (2008) performed two psychovisual experiments, with and without HDR physical references. For the "reality check" evaluation, three environments were captured with an HDR imaging technique (Debevec and Malik 1997) (indoor, outdoor, and outdoor at night). This evaluation is the more exhaustive to date, as 14 TMO were tested. Ten subjects were asked to rate the similarity between the physical scene and the tone-mapped images, as a whole or focusing on specific attributes (brightness, contrast, reproduction of color, reproduction of details, etc.). The best operators were those from Ward 1994; Tumblin and Turk 1999 and Reinhard et al. 2002, and the best linear clip, which outperformed all of them, which should be compared to Yoshida's results (Yoshida et al. 2005).

Ashikhmin and Goyal (2006) compared 5 TMO using physical references. Their point was to compare TMO evaluation with and without physical reference. Their results showed no difference between the "preference" and "fidelity" criteria when no reference scene is experienced, while a significant difference was found when the subjects could compare the tone-mapped images with the reference indoor scenes. As a result, one may expect that assessing the realism of tone-mapped images without reference would in fact assess the subject's preference.

Several important results emerge from the above-listed TMO evaluations. First, no dynamic evaluation was published to date on the dynamic aspects of tone mapping. Such a dynamic evaluation would deserve interactive applications, such as the use of TMO in virtual reality. Second, various operators (including a linear clip) may perform well, depending on the criteria (most of the time, the overall appearance), on the experimental protocol, and on the environment. Kuang's results emphasized the fact that an operator's performance may strongly depend on the input HDR image (Kuang et al. 2007). Considering a specific application, namely, to explore an outdoor virtual world at night, the input images belong to a specific subclass of HDR images, with a low average luminance, while light sources continuously enter and leave the field of view. Thus, temporal visual adaptation is expected to play a significant role.

3 TMO Evaluation in dynamic night-time environments

In the following, we consider night-time virtual environments and compare the merit of several well-performing TMO from the literature. Our goal is to have a better understanding on the impact of this rendering step, in the graphic pipeline, on the subject's presence. The rationale for focusing on these environments is twofold:

- For a specific application such as moving in a night-time virtual environment, we felt it more robust to select an operator which is optimal for this very application, even at the cost of being sub-optimal for other ones. This application-based approach already gave good results in Grave and Brémond (2008).
- Night-time environments raise specific rendering problems: very dark areas (much darker than what a LCD displays as "black") use to appear near bright light sources. This specificity is much emphasized in dynamic conditions, given that glare and temporal adaptation is of great importance at night. Moreover, most TMO compute global parameters, such as light adaptation, from the input images. Very fast changes may occur in the mean and maximum luminance values in urban scenes at night (e.g., when a light source enters the field of view in a dark scene), resulting in a disturbing effect on the screen.

An experimental protocol was designed in order to compare the operator's performance in a night-time outdoor virtual environment. The proposed TMO evaluation framework includes two steps. The first one uses tone-mapped videos as visual stimuli and includes no reference (HDR) display. The tone-mapped videos are computed from HDR videos, which in turn were computed from two virtual environments at night (Petit and Brémond 2010). Participants were asked to rate, after each video, its overall "realism" with respect to their experience of driving at night. This main merit of this experiment was to assess the dynamic aspects of tone mapping, in conditions close to the target situation: exploring a virtual world.

Then, although a direct comparison with the HDR reference is not the main issue in virtual reality applications, it was necessary to control whether operators which perform well in the first (dynamic) experiment are still valuable in a "reality check" evaluation. The best protocol, there, would have been to intoduce the same HDR videos in the evaluation, as was done in (Ashikhmin and Goyal 2006) for photographs. Unfortunately, no HDR display device was available in the lab at the time of the experiment. Moreover, it may be argued that the first experiment not only rated TMO but also the quality of the virtual databases. Thus, a second experiment was designed, in order to rate the merit of the same TMO as in the first experiment, when applied to HDR photographs of urban environments at night. The tone-mapped images were compared to real-world scenes, as in (Yoshida et al. 2005; Ashikhmin and Goyal 2006; Kuang et al. 2007; Čadík et al. 2008).

4 Experiments

4.1 Tone mapping operators

In order to compare state of the art tone mapping operators for night-time virtual environments, three HDR videos were rendered and stored. Then, they were tone-mapped by 8 classical tone mapping operators (Ward 1994; Rahman et al. 1996; Ward et al. 1997; Reinhard et al. 2002; Durand and Dorsey 2002; Choudhury and Tumblin 2003; Reinhard et al. 2005; Irawan and Marschner 2005). The C implementation of these operators was used for each of these operators available in Reinhard's book (Reinhard et al. 2005), except the last one, which was implemented by the first author. The main features of these algorithms are as follow (see a more complete description in (Reinhard et al. 2005), and of course in the corresponding papers):

- Ward's 1994 operator is the first to be grounded on a visual performance criterion (contrast detection) (Ward 1994). He was rated best (together with 5 other operators) in Čadík's study (2008).
- Rahman's operator (1996) is based on the Retinex Theory, which intend to mimic the visual sensitivity to lightness (Land and McCann 1971). He was rated best in Drago et al. (2003)'s study, together with Reinhard's operator (Reinhard et al. 2002).
- Ward et al. (1997) proposed a histogram adjustment method, including a visual performance adjustment in order not to emphasize local contrasts, with respect to human vision. He was rated best or among the best operators in several studies (Drago et al. 2003; Ledda et al. 2005; Kuang et al. 2007).
- Reinhard's operator is inspired by photographic art (dodging and burning) (Reinhard et al. 2002). He was rated best or among the best operators in many studies (Drago et al. 2003; Kuang et al. 2004; Ledda et al. 2005; Kuang et al. 2007; Čadík et al. 2008).
- Durand and Dorsey (2002) used a bilateral filter in order to split HDR images into a base layer (low spatial frequencies) and a contrast layer. Only the base layer is tone-mapped. This operator was rated best in two studies (Kuang et al. 2004, 2007).
- Choudhury proposed an improvement of Durand operator (2002), using a tri-lateral filter instead of the bilateral filter, thus improving the reproduction of details (Choudhury and Tumblin 2003). It was not tested in previous TMO evaluations.
- Another algorithm proposed by Reinhard mimics the behavior of the photo-sensors inside the retina (Reinhard et al. 2005). It was not tested in previous TMO evaluations.
- Irawan's operator (2005) is close to Ward's histogram adjustment technique as far as static images are concerned (Ward et al. 1997) and includes a temporal adaptation model from Pattanaik et al. 2000. Neither (Pattanaik et al. 2000) nor (Irawan and Marschner 2005) was tested in previous TMO evaluations.

4.2 Experiment 1: Computer graphics HDR videos

The first experiment took place in a dark room (no windows, walls painted in black) in order to avoid light reflections to produce a veil luminance on the display. Two virtual environments were used. The first one (*Rivoli* in the following) was designed in order to reproduce as faithfully as possible a small quartier in Paris (France) (Giannopulu et al. 2008).The second one (*Tunnel* in the following) was designed in order to reproduce a tunnel in the neighbor of Paris [courtesy of LIVIC Lab. (Gruyer et al. 2010)]. Both environments were rendered in physical units, using a coarse real-time physical lighting simulation (Petit and

Brémond 2010). HDR videos were produced while one of the authors made a virtual exploration of the VR environments. Then, these HDR videos were tone-mapped by each of the 8 operators listed above, leading to 8 tone-mapped videos for each virtual promenade.

Fourteen subjects (9 men, 5 women) participated to the first experiment. Although some of them worked in the field of digital image, they were naive to the purpose of the experiment. After 8 training videos, they were asked to rate each of a series of 40 short video clips (5 HDR videos × 8 TMO) according to their visual experience of urban environments at night. This rating was done with a numerical rating scale, from 1 (poorly realistic) to 10 (very realistic). This subjective rating was taken as the dependent variable. Participant were displayed 5 series (5 virtual promenades) of 8 videos (8 operators). The promenades were displayed in a random order, and for each promenade, the 8 videos were also displayed in a random order.

4.3 Experiment 2: HDR photographs

The second experience focused on the same operators applied to still images. HDR photographs of urban environments at night were built with a multi-exposure technique: several photographs of the same scene are taken at various exposures; then, the HDR image is built off-line (Debevec and Malik 1997). From these HDR photographs, it was then possible to apply the same 8 tone mapping operators to produce LDR versions of the photographs.

Thirteen subjects participated in this experiment (9 men, 4 women). The subjects followed the experimenter in the streets, on a way around the lab with 10 stops. All experimental data were collected approximately at the same hour of the night as when the photographs were shot. At each stop, the subjects were situated where the HDR photograph had been shot. They were displayed tone-mapped images on a laptop in random order and were asked to rate the "realism" of the displayed images with respect to the real scene they experienced. They used the same numerical rating scale as in experience 1, from 1 (poorly realistic) to 10 (very realistic), with reference to what they actually see. Data from the first site were not considered in the statistical analysis, because we considered it as a training situation Fig. 2.

5 Results

A statistical analysis (ANOVA) was performed for both experiments (Sects. 5.1 and 5.2). The significance criterion was set to $p = 0.05$ Fig. 3.

5.1 Experiment 1

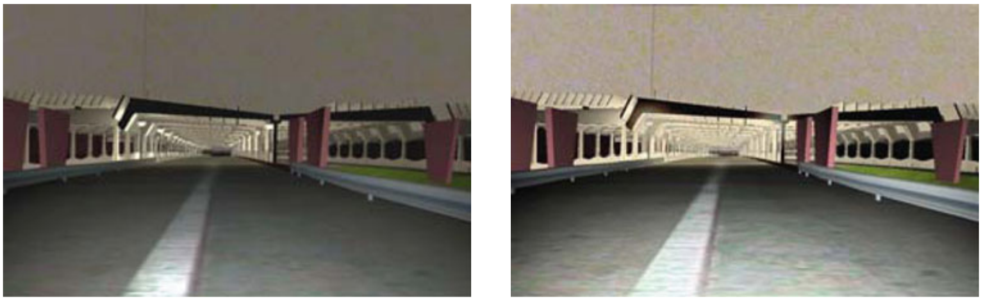
In the first experiment, where tone-mapped VR videos were displayed in laboratory conditions, main effects were found both for the *video* factor ($F(4,56) = 3.19$; $p < .05$) and for the *TMO* factor ($F(7,98) = 32.74$; $p < .05$). Post hoc analysis for all pairs of algorithms (HSD Tukey tests) showed that Irawan and Marschner (2005) significantly outperforms all other operators. When comparing it with the second best, Reinhard's biologically inspired operator (Reinhard et al. 2005), a significant difference was found Table 1.

An analysis was made on these post hoc tests, looking for clusters of operators which lead to the same performance, in a statistical sense. In Table 2, the post hoc tests are presented with the operators in rank order. A statistical difference is denoted by a star (*), while operators with the same performance (in a statistical sense) are denoted NS (Not Significant). Clusters of "NS" may then appear as boxes in Table 2, meaning that the operators in a given cluster are identical in a statistical sense (2 by 2 post hoc tests all are NS). Note that a given operator may belong to several consecutive clusters; for instance, Rahman's operator (1996) belongs to the second and third clusters.

Irawan et al.'s TMO are ranked alone in the first (best performance) cluster, while the second cluster gathers four operators (Reinhard et al. 2005; Ward et al. 1997; Rahman et al. 1996; Choudhury and Tumblin 2003). The third one includes Rahman, Choudhury, and Reinhard photographic operator (Rahman et al. 1996; Choudhury and Tumblin 2003; Reinhard et al. 2002). Ward and Reinhard (Reinhard et al. 2002; Ward 1994) are together in the fourth cluster, while the last one gathers Ward (1994) and Durand and Dorsey (2002). The cluster ranking is summarized in Table 3. Reinhard et al. (2005) and Ward et al. (1997) cannot be distinguished (they are both included in Cluster 2, and in no other cluster). The same happens with Rahman et al. (1996) and Choudhury and Tumblin (2003), which are included in clusters 2 and 3.

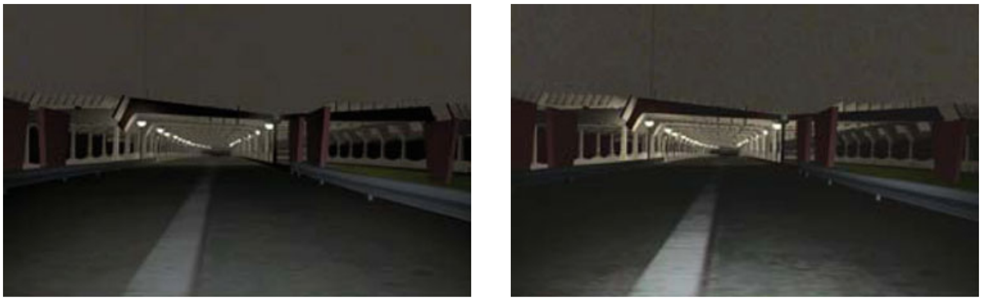
Verbal data collected after the experiment tend to show that a "flicker" effect was experienced with Durand's operator (Durand and Dorsey 2002), due to the high temporal variations in the computation of internal parameters), which resulted in a significant negative effect on the subject's judgment. In addition, most static operators had a drawback when the virtual camera entered/leaved a region with light sources in the field of view, such as leaving the tunnel (fast change in the max/mean luminance value). This may explain the fine results of Irawan's operator (2005), which is the only tested operator to include a temporal adaptation model.

Ward (1994) Rahman et al. (1996)



Ward et al. (1997) Reinhard et al. (2002)

Durand and Dorsey (2002) Choudhury et al. (2003)



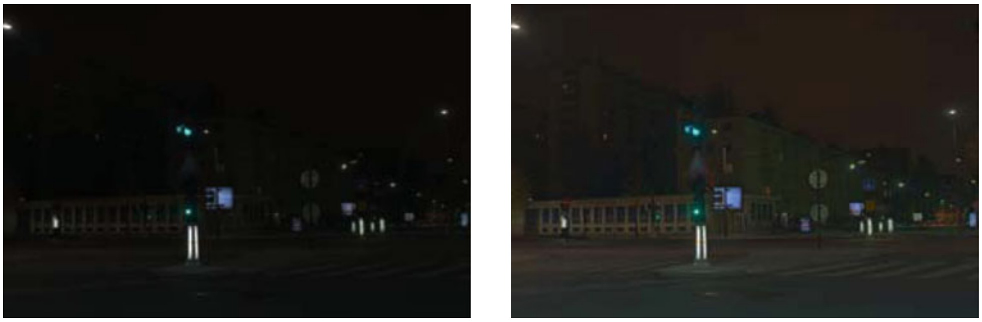
Reinhard et al. (2005) Irawan et al. (2005)

Fig. 2 Sample images from the tone-mapped HDR videos displayed to the participants (*Tunnel* environment). Please note that printed images are not representative of the visual experience of a display on a LCD device with $\gamma = 2$

5.2 Experiment 2

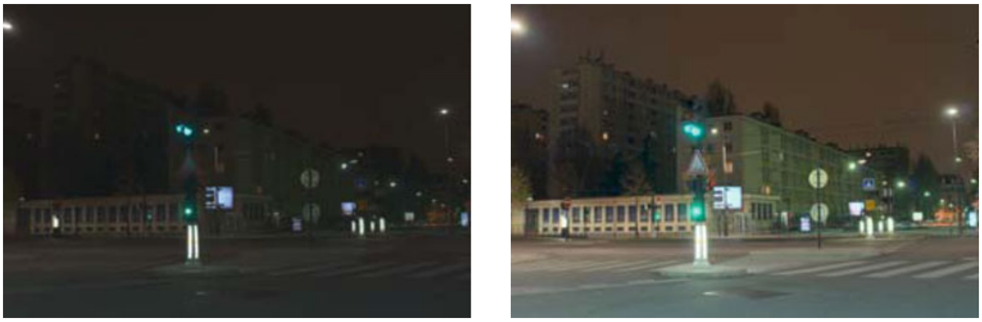
The same 8 TMO were also compared in a static situation, with a "reality check" method close to previous studies (Yoshida et al. 2005; Ashikhmin and Goyal 2006; Kuang et al. 2007; Čadík et al. 2008). Thus, the advantage given to Irawan's operator (2005) by the temporal adaptation model vanished, and the operators could be compared on the basis of their "static" merit.

In this second experiment, no effect was found for the *photograph* factor ($F(8,40) = 1.50$; $p = 0.19$), while a strong effect was found for the *TMO* factor ($F(7,35) = 8.53$; $p < .05$). However, when comparing 2 by 2 the algorithms with post hoc tests (HSD Tukey), most comparisons were not significant: the only significant difference was for Ward's TMO (Ward 1994), which was found worst than all others, except Reinhard's biologically inspired operator (Reinhard et al. 2005).



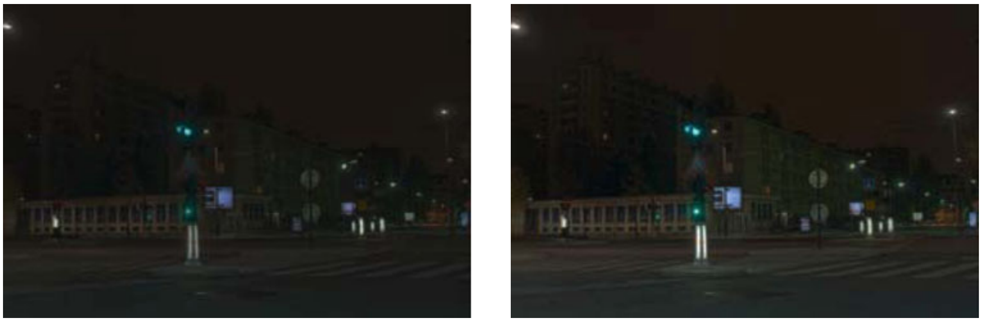
Ward (1994)

Rahman et al. (1996)



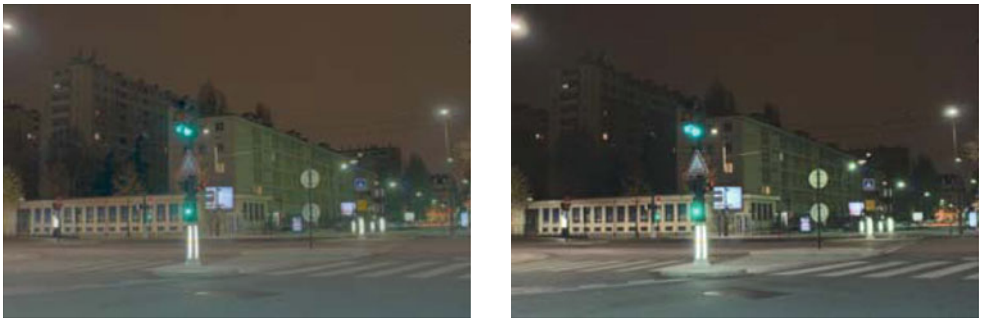
Ward et al. (1997)

Reinhard et al. (2002)



Durand and Dorsey (2002)

Choudhury et al. (2003)



Reinhard et al. (2005)

Irawan et al. (2005)

Fig. 3 Sample images displayed to the participants in the second experiment. The 8 images are tone-mapped from the same HDR photograph (an urban landscape in Paris, France)

Given that the *photograph* factor was not significant, it was possible to compute a new ANOVA based on the mean performance value over the 9 photographs. Of course, the *TMO* factor was still significant ($F(7,84) = 20.27$; $p < .05$). Interestingly, post hoc tests now showed significant differences between the operators (see Table 4).

Looking for clusters of operators as in Sect. 5.1, 5 clusters were found (see the boxes in Table 4). The 3 first operators (Choudhury and Tumblin 2003; Rahman et al. 1996; Irawan and Marschner 2005) could not be distinguished on a statistical basis and are included in the first cluster. Cluster 2 includes (Rahman et al. 1996; Irawan and Marschner 2005;

Table 1 Mean score of the 8 tested TMO for the VR video and the photograph experiments

TMO	videos	rank	photographs	rank
Ward (1994)	3.27	7	2.65	8
Rahman et al. (1996)	5.06	4	5.92	2
Ward et al. (1997)	5.58	3	4.29	5
Reinhard et al. (2002)	4.29	6	3.96	6
Durand and Dorsey (2002)	2.67	8	4.87	4
Choudhury and Tumblin (2003)	4.91	5	6.10	1
Reinhard et al. (2005)	5.62	2	3.90	7
Irawan and Marschner (2005)	7.47	1	5.43	3

Table 2 HSD Tukey post hoc tests, comparing 2 by 2 the 8 algorithms in the video experiment

	I05	R05	W97	R96	C03	R02	W94	D02
I05	–	*	*	*	*	*	*	*
R05	*	–	NS	NS	NS	*	*	*
W97	*	NS	–	NS	NS	*	*	*
R96	*	NS	NS	–	NS	NS	*	*
C03	*	NS	NS	NS	–	NS	*	*
R02	*	*	*	NS	NS	–	NS	*
W94	*	*	*	*	*	NS	–	NS
D02	*	*	*	*	*	*	NS	–

Asterisks indicate that there is a significant difference between the two algorithms, with $p < 0.05$

NS: $p \geq 0.05$

Table 3 Cluster ranking of the 8 tested TMO, based on Tukey's post hoc tests on the "VR video" experiment

TMO	I05	R05	W97	R96	C03	R02	W94	D02
cluster #	1	2	2	2 and 3	2 and 3	3 and 4	4 and 5	5

Table 4 HSD Tukey post hoc tests, comparing 2 by 2 the 8 tested TMO in the "photograph" experiment

	C03	R96	I05	D02	W97	R02	R05	W94
C03	–	NS	NS	*	*	*	*	*
R96	NS	–	NS	NS	*	*	*	*
I05	NS	NS	–	NS	NS	*	*	*
D02	*	NS	NS	–	NS	NS	NS	*
W97	*	*	NS	NS	–	NS	NS	*
R02	*	*	*	NS	NS	–	NS	*
R05	*	*	*	NS	NS	NS	–	NS
W94	*	*	*	*	*	*	NS	–

Asterisks indicate that there is a significant difference between the two algorithms, with $p < 0.05$

NS: $p \geq 0.05$

and Durand and Dorsey 2002) and Cluster 3 includes (Irawan and Marschner 2005; Durand and Dorsey 2002; and Ward et al. 1997). The fourth cluster includes 4 operators (Durand and Dorsey 2002; Ward et al. 1997; Reinhard et al. 2002, 2005), and Ward (1994) is alone in the last one. The cluster ranking is summarized in Table 5.

5.3 Overall results

The two cluster analysis are summarized in Table 6. As we used the default parameters values for each operator, it may happen that a better tuning would have lead to better results for some of them. Anyhow, based on the tested TMO,

Table 5 Cluster ranking of the 8 tested operators, based on Tukey's post hoc tests on the "photograph" experiment

TMO	C03	R96	I05	D02	W97	R02	R05	W94
cluster #	1	1 and 2	1, 2 and 3	2, 3 and 4	3 and 4	4	4 and 5	5

Table 6 Cluster classification of the 8 tested TMO in the video and photograph experiments (cluster #)

TMO	videos	photographs
Irawan and Marschner (2005)	1	1 and 2 and 3
Ward et al. (1997)	2	3 and 4
Reinhard et al. (2005)	2	4 and 5
Choudhury and Tumblin (2003)	2 and 3	1
Rahman et al. (1996)	2 and 3	1 and 2
Reinhard et al. (2002)	3 and 4	4
Ward (1994)	4 and 5	5
Durand and Dorsey (2002)	5	2 and 3 and 4

Irawan (Irawan and Marschner 2005) is the only operator to be rated in the best cluster in both experiments, showing that his performance is not only due to Pattanaik's temporal adaptation model (Pattanaik et al. 2000). Interestingly, Ward et al. 1997, which is the core of Irawan et al.'s static model, leads to poorer results even in the static experiment, which emphasizes Irawan's improvements of Ward's histogram adjustment method.

Given the objective of this study, one cannot say that Table 6 compares the relative merit of the 8 tested operators. The hypothesis under study was limited to the selection of a TMO when exploring a virtual world, in night-time outdoor conditions. The hierarchy between operators for a given task/environment may dramatically change for another task/environment.

6 Conclusion

We have compared the performance of several good-performing TMO (as rated by previous TMO evaluation experiments), together with some operators which were not tested before (Choudhury and Tumblin 2003; Reinhard et al. 2005; Irawan and Marschner 2005), for a specific application: the rendering of outdoor environments at night in virtual worlds. Two experiments were designed, in order to rate the dynamic and static aspects of the operators' performance. Based on post hoc tests, clusters of operators with equal merit were built for each experiment.

Based on our data, Irawan et al.'s operator may be selected for interactive applications in outdoor environments at night, as it is rated best in dynamic conditions and is equivalent with the two operators rated in the first cluster in the static condition. Indeed, a real-time implementation of this operator in virtual environments is available (Petit and Brémond 2010).

These results also give insights for the development of new dynamic operators, if one could answer the following question: is it possible to build a dynamic TMO out of any static TMO? Previous dynamic operators model the time course of adaptation (Pattanaik et al. 2000; Durand and Dorsey 2000; Ledda et al. 2004; Irawan and Marschner 2005), biasing a static tone mapping operator to take into account the fact that the visual system is not fully adapted. Thus, such operators are made of two parts, a static vision model (which needs the adaptation luminance as an input) and a dynamic adaptation model. For instance, Pattanaik et al. (2000) derived a human vision model from Hunt's book (1995), to build a static TMO on it, and Durand and Dorsey (2000) started from Ward's operator [see above, (Ward 1994)], which was grounded on psychophysical data.

It seems sensible, then, to "upgrade" any TMO inspired by vision science into the same framework: an upgraded model from Reinhard operator (Reinhard et al. 2005) would simply include a temporal version of the adaptation computation. But, this upgrade cannot apply so easily when the static operator does not refer to adaptation luminance: this is what happens, for instance, with Rahman's et al. (1996) and Choudhury's operators (2003), which were rated in the best cluster in our static experiment. This opens a new challenge: how to "upgrade" a static TMO to a dynamic TMO, when the adaptation luminance is not an input of the static operator?

On the other hand, future work is needed to assess the impact of image rendering on the subject's presence, exploring more aspects of the immersion feeling. For instance, people are not expected to be sensitive to the same components of the visual environment (Wallach et al. 2010); and subjective preference is only one among other methods to assess the presence feeling in virtual worlds (Bormann 2006; Lepecq et al. 2009).

Acknowledgements The authors would like to express appreciation to Sophie Lemonnier and Jean-Philippe tarel for their helpful advices. The authors are grateful for subject's participation to the psychophysical experiments.

References

Akyuz AO, Fleming R, Riecke BE, Reinhard E, Bulthoff HH (2007) Do HDR displays support LDR content? A psychophysical evaluation. ACM Trans Graph 26(3):Art. 38

Ashikhmin M, Goyal J (2006) A reality check for tone mapping operators. ACM Trans Appl Percept 3(4):399–411

Bormann K (2006) Subjective performance. Virtual Real 9:226–233

Čadík M, Wimmer M, Neumann L, Artusi A (2008) Evaluation of HDR tone mapping methods using essential perceptual attributes. Comput Graph 32(3):330–349

Choudhury P, Tumblin J (2003) The trilateral filter for high contrast images and meshes. In: Proceedings of the eurographics symposium on rendering, pp 186–196

Debevec P, Malik J (1997) Recovering high dynamic range radiance maps from photographs. In: Proceedings of SIGGRAPH. ACM, Los Angeles, CA, pp 369–378

Drago F, Myszkowski K, Annen T, Chiba N (2003) Adaptive logarithmic mapping for displaying high contrast scenes. In: Proceedings of eurographics. Granada (Spain) pp 419–426

Drago F, Martens WL, Myszkowski K, Seidel HP (2003) Perceptual evaluation of tone mapping operators. In: ACM SIGGRAPH conference abstracts and applications

Durand F, Dorsey J (2000) Interactive tone mapping. In: Proceedings of eurographics symposium on rendering. Brno, Czech Rep., pp 219–230

Durand F, Dorsey J (2002) Fast bilateral filtering of HDR images. In: Proceedings of SIGGRAPH. ACM, New Orleans, Lousiana, pp 257–265

Ferwerda JA (2003) Three varieties of realism in computer graphics. In: Proceedings of SPIE human vision and electronic imaging. San Jose, CA, pp 290–297

Giannopulu I, Bertin R, Brémond R, Espié S, Kapoula Z (2008) Visuomotor strategies using driving simulators in virtual and pre-recorded environment. Adv Transp Stud 14:49–56

Grave J., Brémond R (2008) A tone mapping algorithm for road visibility experiments. ACM Trans Appl Percept 5(2): article 12

Gruyer D, Glaser S, Monnier B (2010) SiVIC, a virtual platform for ADAS and PADAS prototyping, test and evaluation. In: Proceedings of FISITA world automotive congress. Budapest (Hungary)

Hunt RWG (1995) The reproduction of colour. 5th edn. Fountain Press Ltd, London

Irawan PAJ, Marschner SR (2005) Perceptually based tone mapping of high dynamic range image streams. In: Proceedings of eurographics symposium on rendering. Konstanz (Germany), pp 231–242

Kuang J, Johnson GM, Fairchild MD (2005) Image preference scaling for HDR image rendering. In: Proceedings of IS&T/SID 13th color imaging conference. Scotsdale, Arizona

Kuang J, Liu C, Johnson G, Fairchild M (2006) Evaluation of HDR image rendering algorithms using real-world scenes. In: Proceedings of international congress of imaging science

Kuang J, Yamaguchi H, Johnson GM, Fairchild MD (2004) Testing HDR image rendering algorithms. In: Proceedings of IS&T/SID 12th color imaging conference. Scotsdale, Arizona

Kuang J, Yamaguchi H, Liu C, Johnson GM, Fairchild MD (2007) Evaluating HDR rendering algorithms. ACM Trans Appl Percept 3(3):286–308

Land E, McCann J (1971) Lightness and retinex theory. J Opt Soc Am 61(1):1–11

Ledda P, Chalmers A, Troscianko T, Seetzen H (2005) Evaluation of tone mapping operators using a high dynamic range display. In: Proceedings of SIGGRAPH. ACM, Los Angeles, CA, pp 640–648

Ledda P, Santos LP, Chalmers A (2004) A local model of eye adaptation for high dynamic range images. In: Proceedings of AFRIGRAPH. ACM, Stellenbosch (South Africa), pp 151–160

Lepecq JC, Bringoux L, Pergandi JM, Coyle T, Mestre D (2009) Afforded actions as a behavioral assessment of physical presence in virtual environemnts. Virtual Real 13:141–151

McNamara A (2006) Exploring visual and automatic measures of perceptual fidelity in real and simulated imagery. ACM Trans Appl Percept 3(3):217–238

Pattanaik SN, Ferwerda JA, Fairchild MD, Greenberg DP (1998) A multiscale model of adaptation and spatial vision for realistic image display. Comput Graph 32(Annual Conference Series): 287–298

Pattanaik SN, Tumblin J, Yee H, Greenberg DP (2000) Time-dependent visual adaptation for fast realistic image display. In: Proceedings of SIGGRAPH. ACM, New Orleans, Lousiana, pp 47–54

Petit J, Brémond R (2010) A high dynamic range rendering pipeline for interactive applications: in search for perceptual realism. The Visual Computer 28(6–8):533–542. Special issue CGI 2010

Rahman ZU, Jobson DJ, Woodell GA (1996) A multiscale retinex for color rendition and dynamic range compression. In: SPIE Proceedings: Applications of digital image processing XIX, vol. 2847. Denver, CO

Reinhard E, Stark M, Shirley P, Ferwerda J (2002) Photographic tone reproduction for digital images. In: Proceedings of SIGGRAPH. ACM, San Antonio, Texas, pp 267–276

Reinhard E, Ward G, Pattanaik SN, Debevec P (2005) High dynamic range imaging: acquisition, display, and image-based lighting. Morgan Kaufmann, USA

Robertson MA, Borman S, Stevenson RL (1999) Dynamic range improvement through multiple exposures. In: Proceedings of IEEE international conference on image processing. Kobe (Japan)

Seetzen H, Heidrich W, Stuerzlinger W, Ward G, Whitehead L, Trentacoste M, Ghosh A, Vorozcovs A (2004) High dynamic range display systems. ACM Trans Graph 23(3):760–768

Tumblin J, Rushmeier H (1993) Tone reproduction for realistic images. IEEE Comput Graph Appl 13(6):42–48

Tumblin J, Turk G (1999) Low curvature image simplifiers (LCIS). In: Proceedings of SIGGRAPH. Addison Wesley, Los Angeles, CA, pp 83–90

Wallach HS, Safir MP, Samana R (2010) Personality variables and presence. Virtual Real 14:3–13

Ward G (1994) A contrast-based scale factor for luminance display. Graph Gems IV, Academic Press Professional, Inc., San Diego, CA, pp 415-421

Ward G, Rushmeier H, Piatko C (1997) A visibility matching tone reproduction operator for high dynamic range scenes. IEEE Trans Vis Comput Graph 3(4):291–306

Yoshida A, Blanz V, Myszkowski K, Seidel HP (2005) Perceptual evaluation of tone mapping operators with real world scenes. In: Proceedings of the SPIE, vol. 5666. San Jose, CA, pp 192–203

Yoshida A, Mantiuk R, Myszkowski K, Seidel HP (2006) Analysis of reproducing real-world appearance on displays of varying dynamic range. Comput Graph Forum 25(3):415–426