Technical Section

Tone mapping high dynamic 3D scenes with global lightness coherency ☆

Ific Goudé [a,*], Jérémy Lacoche [b], Rémi Cozot [c]

[a] *Univ Rennes, CNRS, IRISA, France*
[b] *Orange Labs, France*
[c] *Littoral Opal Coast University, France*

ARTICLE INFO

ABSTRACT

We propose a new approach for real-time Tone Mapping Operator dedicated to High Dynamic Range rendering of interactive 3D scenes. The proposed method considers the whole scene lighting in order to preserve the global coherency. This is the major contribution of our method. Indeed, most of existing Tone Mapping Operators only consider the image rendered by the camera at the current frame and simulate the Human Visual System accommodation to bright and dark luminance. Consequently, after an adaptation time, the lighting design of the 3D scene is lost. For example, two rooms with a high contrast (one dark and one bright) can be perceived with the same luminance level after adaptation. To cope with this coherency issue, we adapt an existing Tone Mapping Operator that combines (1) a global Tone Mapping Operator which takes into account the High Dynamic Range of the whole scene and (2) a viewport frame based Tone Mapping Operator that enhances the image contrast. Thus, we preserve the global lighting coherency and enhance the contrast for each rendered image. Furthermore, we present a subjective evaluation that shows that our method provides a better user experience than previous methods in the case of visualization on Head-Mounted Display.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

In computer graphics, High Dynamic Range (HDR) images of 3D scenes are rendered with all perceptible dark and bright details that befall in the real world. However, displaying such HDR images on standard screens or Head-Mounted Displays (HMDs) is impossible because of their limited dynamic range. To overcome this limitation and to be able to watch HDR contents on Standard Dynamic Range (SDR) displays, the high dynamic range of these contents is compressed to fit the limited dynamic range of the display. This operation is ensured by a Tone Mapping Operator (TMO) that seeks to preserve as best as possible global and local contrasts.

In this paper, we tackle real-time tone mapping of HDR rendering of 3D scenes taking into account both spatial and temporal coherency. Such applications like video games involve a camera moved by a user through a 3D scene. At each frame, the part of the scene that stands in front of the camera is rendered in HDR. The HDR image is then tone mapped and displayed onto a SDR display (or a HMD). As described later in the paper, processing each image independently without accounting for spatial or temporal coherency may produce visual artifacts [1] which degrades the image quality. We propose a tone mapping method which accounts for the whole scene to ensure a spatial and temporal coherency. That way, our TMO allows content consumers to better feel the lighting atmosphere of the scene and avoids visual artifacts such as flickering effect. Furthermore, in order to demonstrate the efficiency of our TMO, we conducted a user study to evaluate the fidelity of tone mapped content with respect to the HDR reference using different TMOs. Compared to other methods, we found that our TMO better preserves fidelity and is more appreciated. While our TMO works for interactive rendering of 3D scenes displayed on 2D screens, we focused on Virtual Reality (VR) contents because of the growth of this technology and the challenge of the existing tone mapping operators to target the limited dynamic range of today's HMDs. Moreover, in the VR use case, the user, and consequently the camera, is constantly moving, which increases artifacts occurrence and degrades the immersion feeling. Finally, we propose a new subjective study protocol to compare HDR content visualized on a 2D display with its tone-mapped version visualized on a HMD.

The paper is organized as follows. Section 2 presents related work regarding real-time tone mapping. In Section 3, we detail an existing TMO dedicated to the visualization of 360° HDR images on HMDs [2] and we propose different changes to adapt this TMO to the problem of walking through 3D scenes rendered in HDR. The implementation of our TMO is detailed in Section 4 while some results are presented in Section 5. Finally, we describe the subjective study we conducted to evaluate our solution over previous TMOs in Section 6. Limits and future work are discussed in Section 7 to conclude the paper.

2. Related work

The research in the field of TMOs started many years ago and has given rise to many different methods. The main purpose is to visualize HDR images, while preserving all details and contrasts on a display with a lower dynamic range. It amounts to a compression problem with loss, that seeks to preserve global and local contrasts as better as possible. We can distinguish two main groups of TMOs: global ones process all pixels of an image with the same operator, while local ones process each pixel according to its neighborhood [3,4]. Many TMOs are based on the Human Visual System (HVS) sensitivity and on how the HVS accommodates to bright and dark luminances to produce a pleasant and coherent tone mapped image. HVS lightness perception has long been studied and can be simulated using mathematical models. The logarithm of the luminance seems to fit well with the perceived lightness [5,6]. Tone mapping an HDR image using the logarithm of the luminance compresses its dynamic range by coherently preserving the perceived contrast. A common way to tone map HDR images is to find a luminance of adaptation to transform the input image luminance to an output luminance on the display [7,8]. More specifically, few TMOs simulate the visual adaptation over time to improve the feeling of contrast when visualizing a sequence of tone mapped images [9,10]. Another way to improve the contrast for scenes with very high luminances is to simulate the maladaptation phenomenon [11–13]. This effect occurs when a visible area is bleached after having looked at something very bright (the sun for example). During few seconds, colors and local contrasts are lost in this view area.

More recently, many works have tackled TMOs for HDR videos [14–16]. Video tone mapping operators have to cope with spatio-temporal coherency issues [1]. A complete survey on video tone mapping has been proposed by Eilertsen et al. [17], it especially details methods that take temporal coherency into account. For example, Zonal Brightness Coherency TMO proposed by Boitard et al. [14] processes the tone mapping of a video by considering spatio-temporal segments to preserve the temporal coherency and avoid incoherent luminance changes. However, this solution cannot be used for real-time tone mapping of HDR images rendered from a 3D scene as the whole video sequence has to be known. Eilertsen et al. [16] proposed a real-time noise-aware tone mapping that answers the main issues of temporal consistency described by Boitard et al. [1]. Nevertheless, this TMO requires too much processing power to be processed in real-time and does not care about object brightness consistency.

Another TMO proposed by Goudé et al. [2] addresses the issue of spatial coherency in the case of still HDR panoramas visualized on HMDs. In that case, the sequence of successive viewport images is unknown and the tone mapping has to be applied in real-time. The authors propose to apply two different TMOs, a first one to the whole panorama that preserves the global coherency and a second one to the viewport image only to enhance the local contrast. The viewport TMO is smoothed over time to mimic the eye accommodation to the luminance level and consequently avoid the flickering effect produced by sudden changes in the image. The final tone mapped image is a combination of both TMOs in the logarithmic domain.

Durand and Dorsey [18] proposed an Interactive tone mapping that simulates the visual adaptation. This TMO is dedicated to interactive walk in a 3D scene rendered in HDR and tone mapped on the fly. They added a post processing pass that accounts for chromatic adaptation and adds flares around light sources to increase the brightness effect. Due to visual adaptation, this TMO ensures temporal coherency but does not take into account the whole scene to tackle the spatial coherency.

Recall that our objective is to walk through a 3D scene, render it in real-time from the current camera, tone map the resulting HDR viewport image (which corresponds to the current position of the camera) and display it on an SDR screen, while preserving spatial and temporal coherency. Spatial coherency is obtained by accounting for the luminance of the whole scene. Temporal coherency is ensured by efficiently smoothing transitions between successive views. In this work, the image sequence is unknown because the camera is controlled by the user and the lighting conditions of the scene may change over time (turning on and off light sources, moving objects, etc.) Our solution relies on a combination of two TMOs as described by Goudé et al. in [2].

The first one is a global TMO that takes into account the luminance distribution of the whole scene to preserve the spatial and temporal coherency, while a second TMO, that only considers the image rendered on the viewport, enhances the local contrast and the temporal coherency. A more detailed description of our method is given in the following section.

3. Dynamic range of a 3D scene

3.1. Problem statement

A previous work about tone mapping 360° HDR images raises the issues of spatial and temporal coherency when these images are visualized on a HMD [2]. At each time step, only a part viewport of the 360° image is displayed on the HMD, corresponding to where the user is looking at. To visualize these limited parts of the 360° image one after the other, a naive approach is to tone map each viewport image individually. This way, each current image in the viewport is well tone mapped but the sequence visualized as a whole can suffer from temporal incoherence. Indeed, the used TMO is based on a global operator proposed by Reinhard et al. [19] that computes luminance values of the HDR image: the log-luminance average and the maximum luminance value. If sudden changes in luminance happen from a view to another, the operation of tone mapping can produce flickering artifacts that breaks the temporal coherency. The solution initially proposed by Yu [20] is to smooth the tone mapping values over time in order to mimic the way the human eyes accommodate to bright and dark luminances. In this way, flickering artifacts are avoided by the smooth transitions and the temporal coherency is then preserved. Nevertheless, processing each image in the viewport irrespective of the entire 360° image tends to produce tone mapped images with the same overall luminance average. Dark areas of the 360° image will be brightened while bright areas will be darkened. Thus, the spatial coherency of the 360° image is lost because all areas will be rendered with the same average luminance.

Goudé et al. proposed to solve the problem of spatial coherency by applying a TMO to the entire 360° image. Tone mapping the entire 360° image in one go with the same global operator [19] preserved the spatial coherency. Dark areas are still dark on the tone mapped image. Moreover, there is no more problem of temporal coherency (like flickering effect) as the tone mapping is applied once. Tone mapping the entire 360° image only once could solve the incoherence issues but it is to the detriment of reducing the
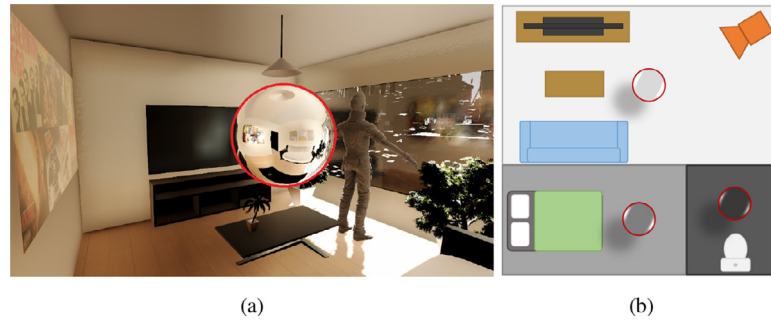
(a)   (b)

Fig. 1. (a) The environment map of the living room rendered by a 360° camera is represented by a sphere. (b) Scene example with three 360° cameras (circled in red) placed in different rooms. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

used dynamic range of the tone mapped image in the viewport. Effectively, the dynamic range of a 360° image can be potentially much higher than a limited area of the same image in the viewport. Reducing the used dynamic range leads to a reduced contrast in the tone mapped image that degrades the visual quality. The authors suggest to keep advantages of those two TMOs by combining them together.

Finally, a smoothed TMO applied to the viewport enhances the contrast and is temporally coherent while a TMO applied to the 360° image preserves the spatial coherency. Combining these two TMOs provides a balanced result. The global TMO can be processed once at the start and the viewport TMO is processed at each time step. To obtain a coherent tone mapped image, the combination is performed in the logarithmic domain (as explained in Section 4). This approach (called HMD-TMO from now on) can be reused in the case of interactive HDR rendering of 3D scenes. Indeed, the viewport TMO can be used as is for each frame, while the global one has to be modified because the luminances of a 3D scene differs from the luminance of a 360° image. For this reason, HMD-TMO cannot be adapted to 3D scenes, it is then not evaluated in our study. The adaptation of HMD-TMO to 3D scenes is the objective and the main contribution of the paper; it is detailed in the following section.

3.2. Approximating the luminance distribution

While the dynamic range of a still image has a clear definition (minimum and maximum luminance values), the dynamic range of an entire scene is a more complex notion. Looking at the luminance of all pixels of an image gives the dynamic range of this image.

On the other hand, a straightforward approach to compute the dynamic range of a 3D scene is to look at the luminance of every space point of the scene. Indeed, for each 3D position (x, y, z) in the scene, if we compute the luminance coming from every direction (θ, ϕ), we can obtain the light field of the scene defined by the five-dimensional plenoptic function $L(x, y, z, \theta, \phi)$ [21]. Proceeding this way would give the most precise estimation of the luminance of the scene. In comparison, while the luminance of a pixel for an image is given by $L(x, y)$, the luminance at a position of a 3D scene that comes from a direction is given by $L(x, y, z, \theta, \phi)$. Nevertheless, computing such a luminance in a 3D scene lead to place a large number of 360° cameras at many points of the scene [22] and render the scene for all of these cameras, which is a complex, memory and time consuming process. We assume that the luminance of the scene can be approximated by placing only few 360° cameras at key positions as illustrated in Fig. 1.

The rendered images are 360° images of the surrounding environment [23] called environment maps (EMs) from now on. Of course, the placement of the 360° cameras depends on the scene topology, and it is not excluded to put several cameras in the same

room if this one is huge or with a particular topology. At the end, we get a set of HDR EMs rendered by all 360° cameras placed in the scene. Looking at the luminance of all pixels of those EMs gives an approximation of the dynamic range of the scene. But the positioning of the 360° cameras may affect the luminance of the pixels, and then the result of the tone mapping.

3.3. Importance of the 360° cameras positioning

The positioning of the 360° cameras could affect the actual dynamic range of the scene. If the EMs are not placed correctly in the scene (for example not close to sources of high luminance such as light sources, specular and glossy surfaces, etc.) then the dynamic range of the scene is not well captured, which affects the TMO. We will precisely detail later how the TMO processes the rendered HDR image.

In the example of the Fig. 2, when the 360° camera is placed close to a window (first row), the rendered EM is mainly bright (left image), the tone mapped image is dark (right image). On the other hand, when the 360° camera is far from the window (second row), the EM is darker (left image) and the tone mapped image is bright (right image). In both tone mapped images (right side), the position of the 360° camera is represented by a sphere circled in red. The placement of the 360° cameras in the scene has a real impact on the tone mapped image and it is left to the content creator to place the 360° cameras depending on the expected result. Besides the influence of 360° cameras positioning, the rendering method used by these cameras to generate the HDR EMs can also impact the resulting tone mapping.

3.4. Comparison of rendering methods

In order to obtain the most representative luminance distribution of the scene, a realistic simulation of the light interaction with objects surfaces should be done. Ray tracing based methods produce more realistic images but are still too slow to render multiple EMs of complex 3D scenes in real-time. Nevertheless, approaching a realistic rendering in real-time can be achieved using some simplifications. We decided to render direct lighting and add an ambient light and an ambiant occlusion terms that coarsely approach indirect lighting.

The rendering performed by these two methods (ray tracing and ambient based rendering) produces slightly different results that are illustrated in Fig. 3. Moreover, as detailed later, only the log-luminance average value ($\bar{L}_w$ in Eq. (2)) of the EMs is needed to compute our global TMO.

We compare the log-luminance average of the environment map rendered using ray tracing (Fig. 3a) and when approximating it with ambient and occlusion terms (Fig. 3b). The difference between these two computed values ($\bar{L}_w$) does not impact the resulting TMO. Thus, all EMs are rendered in real-time by computing
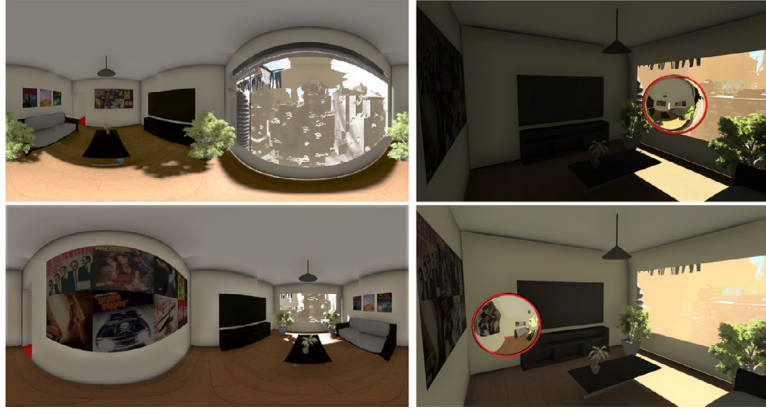
Fig. 2. First row: the 360° camera is close to the window, the environment map (left) is mainly bright and the TMO darken the image (right). Second row: the 360° camera is far from the window, the environment map (left) is mainly dark, the TMO brighten the image (right). (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)
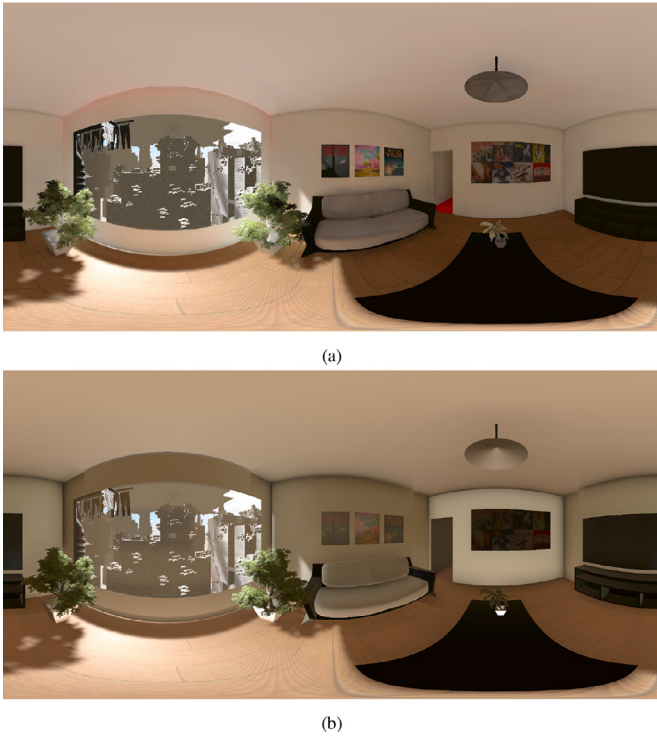

(a)

(b)

Fig. 3. (a) Ray tracing rendering: $\bar{L}_w = 0.012$ (b) Ambient based rendering: $\bar{L}_w = 0.01$. The log-luminance average does not differ enough to significantly impact the result of the TMO.

direct lighting added to an ambient term and an ambient occlusion. Our global TMO is then performed using all the rendered EMs as described in Section 4. To finish, as the process has to be real-time, we propose three optimizations that can increase significantly the rendering time.

3.5. Optimizations

First, we propose to render the EMs with a low resolution as only the log-luminance average value is needed by the global TMO. This value is pretty equivalent for a high or a reasonably low resolution.

Secondly, the number of EMs to render impacts the processing time. As the scene can be huge, dozen of 360° cameras can be useful to obtain the global TMO. We propose to limit the number of

EMs to render by considering only the nearest cameras from the user. The number of cameras to take into account is a user-defined parameter. The impact of number of EMs to render and their resolutions is shown in Fig. 4.

Moreover, the scene is dynamic: lights can be turned on and off, doors can be opened and closed, etc. To obtain a correct global TMO, EMs should be rendered at each frame, which is time-consuming. We finally propose to render the EMs only at the first frame, and when the scene changes. A change is detected only for objects that can significantly affect the luminance of the scene (lights and doors for example, moving little objects does not impact the luminance too much). If any change is detected on a object close to a 360° camera, the EM is rendered again to take into account this change in real-time.

4. TMO for HDR rendering of 3D scenes

4.1. Global TMO of a 3D scene

Before describing our TMO that considers the luminance of the whole scene, we will explain the basics of HMD-TMO [2]. The global TMO performed by HMD-TMO is applied to the 360° HDR image that is visualised on the HMD. During all the process, the luminance is calculated following the Rec. 709 standard:

$$L_w = 0.2126 \times R + 0.7152 \times G + 0.0722 \times B. \tag{1}$$

First, the log-luminance average $\bar{L}_w$ of the 360° image is computed as:

$$\bar{L}_w = \frac{1}{N} exp\left(\sum_{x,y} log\big(\delta + L_w(x,y)\big)\right), \tag{2}$$

where L_w is the world luminance of a pixel (x, y). A small value δ is used to avoid singularities and N is the number of pixels of the 360° image. Then, the global TMO processes each pixel (x_d, y_d) of the viewport using the Photographic Tone Reproduction operator proposed by Reinhard et al. [19]:

$$L(x_d, y_d) = \frac{a}{\bar{L}_w} L_w(x_d, y_d), \tag{3}$$

$$G(x_d, y_d) = \frac{L(x_d, y_d)}{1 + L(x_d, y_d)}, \tag{4}$$

where a is a user defined value (default value is equal to 0.18) and G is the global tone mapped luminance.

In our case, we also want to tone map the HDR image rendered by the user camera using Eqs. (3) and (4). However, the log-luminance average $\bar{L}_w$ (defined in Eq. (2)) cannot be computed in
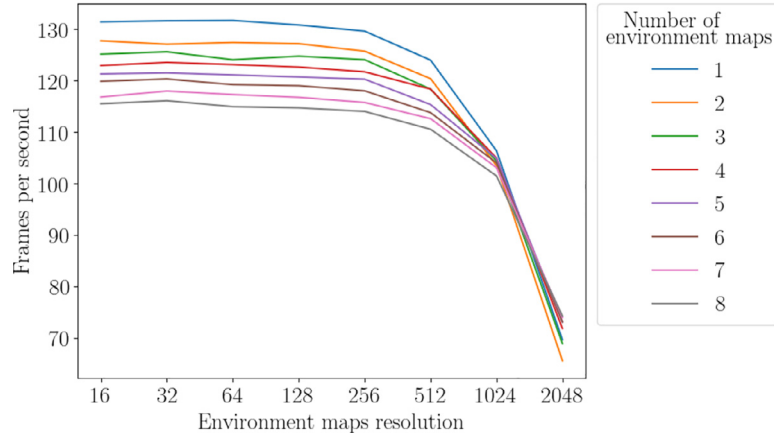
Fig. 4. Impact of the number of EMs to render and their resolution for a complex 3D scene (CPU Intel core i5 6600 3.30 GHz, 16GO RAM, NVidia GTX 980Ti).

the same way as the scene is not represented by a unique 360° HDR image. We have to take into account the set of HDR EMs obtained before to compute the log-luminance average of the whole scene. First, we compute the log-luminance average $\bar{L}_w$ of each HDR EM_i rendered by the 360° cameras:

$$\bar{L}_w(EM_i) = \frac{1}{N_i} exp\Big(\sum_{x,y} log\big(\delta + L_w(EM_i, x_t, y_t)\big)\Big), \quad (5)$$

where L_w is the world luminance of a pixel (x_t, y_t) of the EM_i at a time t. A small value δ is still used to avoid singularities and N_i is the number of pixels for the EM_i. The global log-luminance average of the whole scene is then approximated by the probability for the user camera to be in a particular room at a moment. As we do not have any a priori on the position of the user in the scene, the probability of being in a room is the same everywhere. That conducts in an arithmetic mean of the log-luminance averages of all EMs:

$$\bar{L}_w = \frac{1}{M}\Big(\sum_i \bar{L}_w(EM_i)\Big), \quad (6)$$

where M denotes the number of 360° cameras in the scene. Our newly defined $\bar{L}_w$ (Eq. (6)) replaces the one given in Eq. (3).

At this point, the spatial coherency is preserved as the TMO is computed considering the log-luminance average of the entire scene. However, temporal changes are not considered because luminance variations in the scene can happen (lights turn on and off, doors can be opened and closed, etc). To take into account these variations, the computation of the log-luminance average of the whole scene (Eq. (4)) is performed at each time step. Indeed, if an area of the scene suddenly becomes very bright, the scene luminance globally increases and the tone mapped image is darkened. In order to avoid a sudden change in the tone mapped image, we decided to temporally smooth the value of the log-luminance average of the whole scene in the same way Goudé et al. smoothed the viewport TMO in HMD-TMO:

$$\bar{L}'_w(t) = \alpha_g \bar{L}_w(t) + (1 - \alpha_g)\bar{L}'_w(t-1), \quad (7)$$

where $\bar{L}'_w(t)$ is the smoothed log-luminance average value of the entire scene between two successive time steps and α_g is a time dependent interpolation variable which defines how fast is the smoothing. We fixed $\alpha_g = 0.1$ in our implementation, which means the transition takes about 10 seconds. The smoothed log-luminance average value $\bar{L}'_w(t)$ replaces $\bar{L}_w$ in Eq. (3). This allows us to adapt the tone mapping to temporal changes of lighting in the scene with smooth transitions.

4.2. Viewport TMO of the user camera

As detailed in HMD-TMO, a TMO only applied to the global scene results in a very limited use of the dynamic range of the display. To use a higher dynamic range and then enhance the contrast in the tone mapped image, a second TMO has to be applied to the viewport. Goudé et al. proposed to compute a second tone mapping of the image rendered in the viewport and combine it with the global TMO. The used operation is the same, but the log-luminance average is that of the HDR image of the viewport.

$$\bar{L}_w = \frac{1}{N} exp\Big(\sum_{x_d, y_d} log\big(\delta + L_w(x_d, y_d)\big)\Big), \quad (8)$$

$$\bar{L}'_w(t) = \alpha_v \bar{L}_w(t) + (1 - \alpha_v)\bar{L}'_w(t-1), \quad (9)$$

$$L(x_d, y_d) = \frac{a}{\bar{L}_w} L_w(x_d, y_d), \quad (10)$$

$$V(x_d, y_d) = \frac{L(x_d, y_d)}{1 + L(x_d, y_d)}, \quad (11)$$

where L_w is the world luminance of the pixel (x_d, y_d) in the viewport image, $\bar{L}_w$ is the viewport log-luminance average, and a is still the user defined value. $\bar{L}'_w(t)$ is the smoothed value between two successive views and α_v is a time dependent interpolation variable ($\alpha_v = 1.0$ as defined in [2]). Finally, V is the viewport tone-mapped luminance. There, the temporal coherency is also preserved by smoothing the $\bar{L}_w$ value through time (Eq. (9)). Besides, the eye temporal adaptation is implicitly involved in these smooth transitions.

4.3. Resulting TMO: combination of global and viewport TMOs

Finally, the combination of both, global and viewport TMOs is done in the logarithmic domain as proposed by Goudé et al.:

$$L_d = exp\Big(\big(\beta \times ln(G)\big) + \big((1 - \beta) \times ln(V)\big)\Big), \quad (12)$$

$$L_d = G^\beta \times V^{1-\beta}, \quad (13)$$

where G and V are respectively the resulting luminance of the global and the viewport TMO. β is a variable in the range (0,1) that gives more emphasize on the global or the viewport result and L_d is the final displayed luminance. Obviously, L_d, G and V are evaluated for each pixel (x_d, y_d).

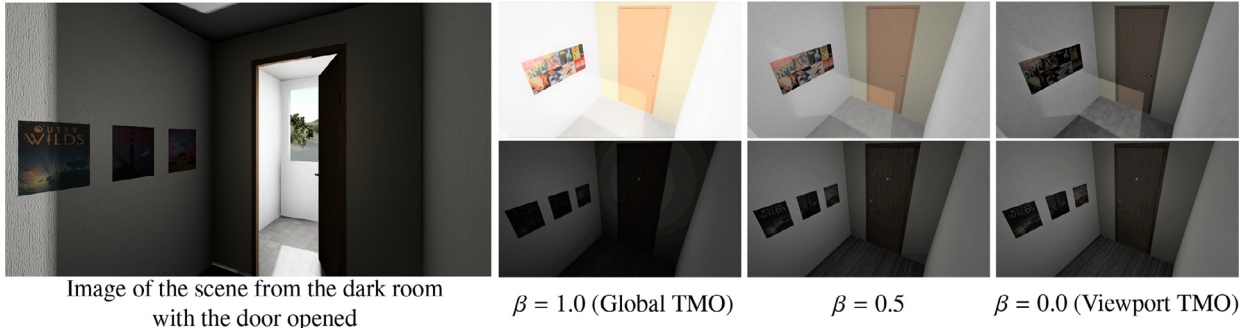| Image of the scene from the dark room with the door opened | $\beta = 1.0$ (Global TMO) | $\beta = 0.5$ | $\beta = 0.0$ (Viewport TMO) |

Fig. 5. The β variable of the combination (Eq. (13)) gives more emphasize on the global TMO or on the viewport TMO. First row and second row are respectively the tone map rendering of the bright room and the dark room in the same scene showed by the left image.
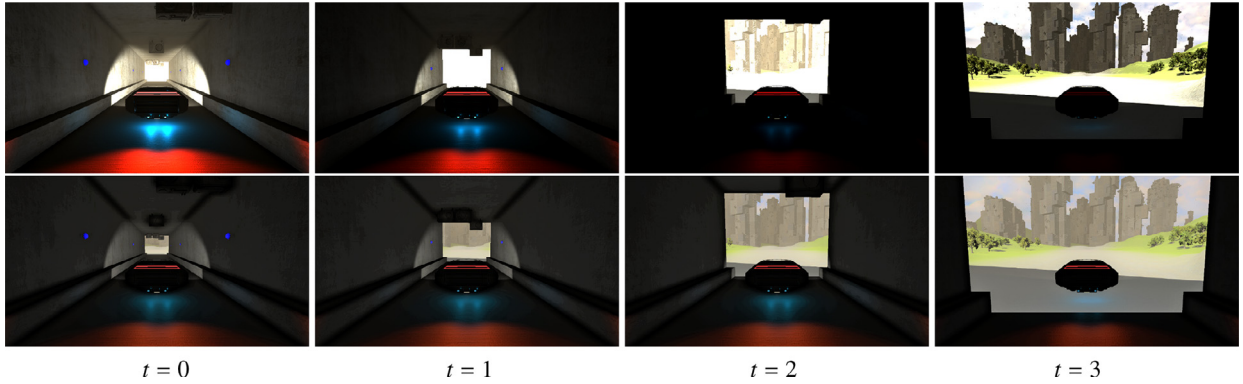

| $t = 0$ | $t = 1$ | $t = 2$ | $t = 3$ |

Fig. 6. Comparison between eye-adaptation (first row) and our TMO (second row) on a image sequence. The eye-adaptation TMO only applied to the viewport does not preserve the temporal coherency. The tunnel is less and less visible while the outside becomes visible when the car gets out. Our method preserves the temporal coherency, both inside and outside of the tunnel are always visible. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

4.4. Colorization

During the entire process, only the luminance is managed. The color is added at the end using the Schlick's approach [24]:

$$C' = \left(\frac{C}{L_w} \right)^s L_d, \tag{14}$$

where C and C' are respectively the input and output trichromatic values (RGB), L_w the world luminance and L_d the tone mapped luminance. The saturation parameter s is set to 0.7 for all of our results. We used a basic method to manage the color, but several approaches of color correction exist for tone mapping [25,26] and can be easily applied to this TMO. The efficiency of our TMO is illustrated in the following section.

5. Results

In this section, we evaluate our TMO for scenes rendered in HDR. We compare tone mapped images either rendered from two different viewpoints of the scene or from different time with a camera that moves through the scene.

We start by giving a detailed example of the effect of combining both TMOs illustrated by Fig. 5. In this example, we compare two rooms of a same scene. One is very bright (first row) and the second is very dark (second row). When considering the global TMO only ($\beta = 1.0$), the global contrast between the two rooms is well preserved. However, details in both images are lost. The light reflection on the wall in the bright room is clipped, the floor in the dark room is totally black. Regarding the result when considering the viewport TMO only ($\beta = 0.0$), the contrast in each image is enhanced, all details are visible. However, the global coherency is completely lost, both rooms seem equally lit, it is impossible to say

if one room is brighter than the other. A geometric mean between the two TMOs ($\beta = 0.5$) gives a good compromise for both images. The images of both bright and dark rooms are well detailed and the global contrast is preserved.

We also compared our method with a TMO applied to the viewport only with smooth transitions between views that simulate the visual adaptation to the light[1]. We are interesting in the temporal coherency as illustrated in Fig. 6. This example corresponds to a scene where a car is driving in a tunnel [9,27]. With the naive TMO only applied to the viewport (first row), the temporal coherency is lost. At start ($t = 0$), the car is inside the tunnel, only the spotlights are illuminating the area. We see in the background that the end of the tunnel is completely clipped. At the end ($t = 3$), the car is outside the tunnel, the image is mainly bright. We can not see anymore the red and blue lights from the back of the car. Over time, the HDR image became brighter and brighter, which leads to a darker and darker tone mapped image. With our TMO (second row), the inside and the outside of the tunnel are visible at each time in the tone mapped image because we take into account the luminance of the whole scene. Even if the contrast is reduced in each image independently, the temporal coherency is globally preserved. See the accompanying video for more details.

6. Subjective evaluation

In this section we detail our evaluation protocol regarding the efficiency of our TMO compared with two others. We suppose that our method is more faithful with respect to the HDR content and that users appreciate more our TMO. To confirm that, we conducted a subjective evaluation where participants had to judge an

---

[1] Eye-adaptation TMO implemented in Unity3D (https://unity.com/).

HDR content tone mapped with three different TMOs compared with the same HDR content visualized on an HDR display (the ground truth). Our panel consists of 18 participants (9 females and 9 males), between 23 and 44 years old. 9 were expert in VR with a daily use while the others had only a few VR experience before the experiment.

6.1. Experiment protocol

The objective is to allow a user to compare scenes visualized on the HDR display (the ground truth) with the tone mapped result visualized on a HMD.

6.1.1. Evaluating TMOs

As explained before, HMD-TMO cannot be evaluated in our study. Even if HMD-TMO ensures visual coherency while displaying 360° HDR images, this coherency is limited to head rotation movement while in our case we tackle walk through 3D scenes. That is why we chose to evaluate our TMO by comparing it with two others that can be computed in real-time. The first one is a naive method denoted Logarithm that is based on the Adaptive logarithmic mapping for displaying high contrast scenes proposed by Drago et al. [28] with smoothed values over time to avoid flickering artifacts. In the implementation of this first TMO, the b factor defined in the related paper is set to 0.8 as suggested by the authors. The second one is the eye-adaptation TMO implemented in Unity3D. This TMO is based on the Photographic tone reproduction for digital images [19] with smoothed values too. In addition to ensuring temporal coherency, this implementation tries to preserve the global coherency by introducing a scale factor. This parameter applies an offset to the display luminance depending on the world log-luminance average of the current image. In other words, if the HDR image is globally dark, this method applies a negative offset to darken the displayed image. It works similarly for bright images. This method is effective for a user defined luminance range. For our purpose we fixed the low and high limits corresponding to the minimum and maximum luminances that can be encoded by the used HDR display (ie. the dynamic range of the SIM2 display goes from 10^{-4} to $10^4 cd/m^2$).

To sum up, the three TMOs used in the experiment are:

1. Our TMO described in Section 4
2. Eye-adaptation: The default eye-adaptation TMO implemented in Unity3D
3. Logarithm: Adaptive logarithmic mapping [28] with smoothed values to avoid flickering.

6.1.2. Evaluation criteria

We asked participants to evaluate the three TMOs regarding two aspects. First, they had to grade the visual fidelity of the tone mapped content visualized on the HMD compared with the HDR content visualized on the HDR screen. Then, we ask them to rank their global appreciation regarding the visual quality of the tone mapped content regardless of the HDR reference. Finally, they had to rank the three TMOs by order of preference.

6.1.3. Evaluating 3D scenes

In order to evaluate our TMO, we developed two interactive scenes in Unity3D. These scenes are interactive as the camera can move and a door can be opened or closed. Both scenes have very contrasted areas (very dark and very bright) to confront evaluated TMOs with difficult conditions. The first one takes place in a car and the scene consists in driving in a tunnel. Outside the tunnel, the sun illuminates the scene while inside the tunnel, the car's headlights are the only light sources as showed in Fig. 6. The second scene takes place in a little apartment. One room has a huge

window and is illuminated by the sun. The second room is totally closed and is illuminated by a weak light. The two rooms are separated by a door as illustrated in Fig. 5.

6.1.4. Evaluation procedure

The procedure can be described as follows. We start by explaining to participants the test protocol and ask them to fill in the consent form. First, we ask them to interact with the HDR scene during few minutes in order to view the different areas of the scene. The interaction with the scene is very simple: participants use a mouse to manipulate the position and the orientation of the camera in the scene. The content visualized on the HDR display is the ground truth.

Once the HDR scene has been explored, we present to participants exactly the same scene on the HMD. In that case, participants move physically in rotation and/or translation depending on the tested scene. The scene of the apartment had the same dimensions that the physical experiment room, so participants are entirely free of movement. For the scene of the tunnel in the car, participants are sat at the driver position and we give them a mouse to move the car forward and backward (by pressing left and right buttons). A constant speed was applied to the car in order to reduce the effect of motion sickness experienced by users [29,30]. We observed this effect thanks to a Simulator Sickness Questionnaire (SSQ) [31]. The image on the HMD is tone mapped using one of the three evaluated TMOs.

Finally, when the HDR content and the tone mapped content have been seen, we ask participants to give their grades regarding the two criteria: (1) The fidelity of the tone mapped content compared with the HDR reference is graded using a linear scale between 0 (not faithful at all) and 10 (very faithful). (2) The global appreciation regarding the visual quality of the tone mapped content is graded using a linear scale between 0 (not appreciated at all) and 10 (very appreciated). To sum up, a protocol loop consists of three steps:

1. Interact with the HDR content on the SIM2 HDR screen
2. Interact with the tone mapped content on the HMD
3. Grade the visual fidelity and the global appreciation

Once the three TMOs for a scene have been tried, we also ask participants to rank the three TMOs from the best to the worst in order of preference. For the whole study, we evaluate three TMOs for two scenes, which means the participants follow the protocol six times in total. We also randomized the order the scenes and the TMOs are tested in order to avoid any bias.

6.1.5. Software and hardware material

Hardware used were a SIM2 HDR display for the ground truth and an HTC Vive pro as a HMD. In order to render HDR images, an image effect shader compliant with the SIM2 display has been implemented. We implemented our solution in Unity3D as it is a friendly tool for VR and HDR developments. Tone mapping values are computed in real-time with compute shaders. Then we apply our TMO as an image effect shader on the rendered HDR image. All the process has to be real-time and run at least at 90 frames per second to ensure a pleasant experience in case of VR. We will now present the results of the experiment.

6.2. Experiment results

As our experiment involved physical navigation as well as virtual navigation using a controller, users fulfilled the SSQ before and after the test. Indeed, cybersickness could have an impact on our results. We followed the approach proposed by Kennedy [31] in order to compute the SSQ score of each user before and after the experiment. The mean (M) and standard deviation (SD) scores of
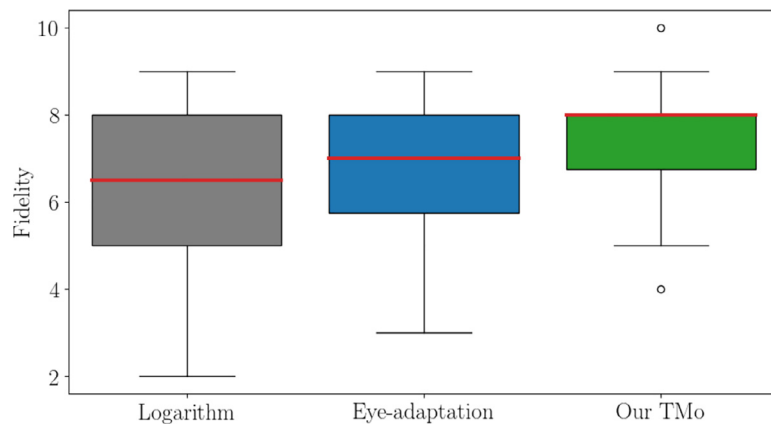
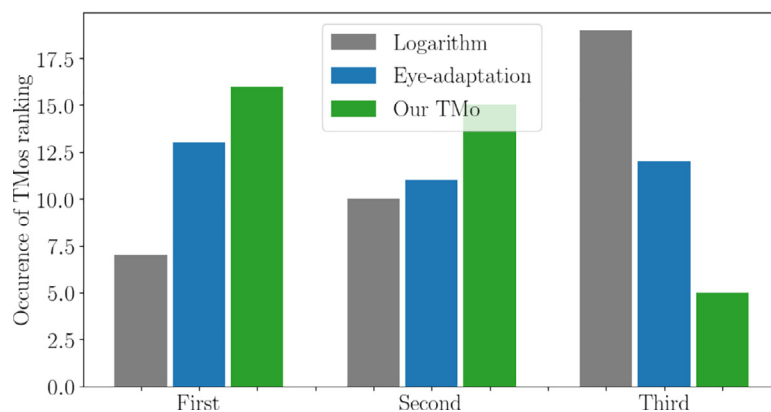**Fig. 7.** Box plot for the fidelity scores overall.

Fig. 8. Occurrence of TMOs ranked first, second and third overall. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

Table 1
Fidelity means and standard deviations for the different TMOs.

		Fidelity		
		Logarithm	Eye-adaptation	Our TMO
Scene 1	M	6.39	6.56	**7.50**
	SD	1.82	1.89	**1.15**
Scene 2	M	6.17	7.00	**7.39**
	SD	1.69	1.64	**1.58**

Table 2
Global appreciation means and standard deviations for the different TMOs.

		Global appreciation		
		Logarithm	Eye-adaptation	Our TMO
Scene 1	M	6.83	6.94	**7.44**
	SD	2.20	1.66	**1.38**
Scene 2	M	6.28	7.22	**7.61**
	SD	1.81	1.86	**1.65**

the SSQ were slightly greater at the end of the experiment (start: $M = 11.01$, $SD = 15.04$; end: $M = 14.13$, $SD = 21.75$) mainly due to the disorientation symptoms. However, these scores can be considered relatively low. As these scores do not follow a normal distribution, we used a Wilcoxon Matched Pair Signed Rank Test to determine if this difference was significant. The result suggests that our experiment did not have a significant impact on cyberskickness ($p > 0.05$).

Regarding the different scores given for fidelity and appreciation, we first analyzed both criteria for each scene (scores presented in Tables 1 and 2).

Globally, our TMO obtained the best scores while Logarithm obtained the worst for both fidelity and appreciation criteria. However, we performed an analysis of variance (ANOVA) on these results but we did not find any significant difference.

Results were also analyzed independently from the scene, by adding the scores obtained for both scenes for each criteria: fidelity and appreciation.

We also performed an ANOVA on these results and we found a significant difference for the fidelity criteria ($p < 0.05$) as represented in Fig. 7. Then, for comparing each pair of conditions, we performed a post hoc Turkey test and we only found a significant difference between Logarithm and our TMO conditions ($p < 0.05$). No significant difference was found for global appreciation.

Regarding the ranking of the three TMOs given at the end of the experiment, results are shown in Fig. 8. Whatever the scene, our TMO (green) is the one which is the most ranked first and second and the least ranked third. The Logarithm TMO (grey) is often ranked third while the Eye-adaptation TMO (blue) is equally ranked first, second and third. Indeed, our TMO tends to be the one which satisfies most users.

To sum up, the non-significant results do not allow us to give a clear conclusion. However, they suggest that our method is globally more faithful and more appreciated than the others, especially compared with the Logarithm method.

7. Conclusion

Rendering in computer graphics produces more and more HDR images. Preserving the artistic intent and lighting atmosphere of the original content is essential during the tone mapping process. Previous methods do not consider the whole scene to tone map rendered images, which causes a loss of global coherency. To overcome this limitation, we proposed to consider the whole scene to compute a global tone mapping combined with a viewport tone mapping to render our final image. The entire process is performed in real-time and can be applied to more complex scenes. We showed that our method is slightly preferred when compared with other ones.

Nevertheless, this TMO suffers from inconsistency when a sudden change appear in a distant room: the global luminance changes and causes an undesired evolution of the tone mapped image. This issue is managed thanks to smooth transitions, but could be tackled differently considering visual adaptation or maladaptation. Our future work heads toward fixing this incoherent result. Moreover, an analysis of the scene topology to determine an efficient placement of the 360° cameras could lead to a semi or fully automated process. Finally, more subjective evaluation should be conducted for manifold 3D scenes with various topology, lighting design and more interactivity. It could be very interesting to evaluate our TMO on SDR 2D screens to confirm its efficiency compared with authors TMOs.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Ific Goudé: Conceptualization, Methodology, Software, Validation, Writing - original draft, Visualization. **Jérémy Lacoche:** Formal analysis, Investigation, Data curation, Writing - original draft. **Rémi Cozot:** Writing - review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgments

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cag.2020.08.004.

References

[1] Boitard R, Cozot R, Thoreau D, Bouatouch K. Survey of temporal brightness artifacts in video tone mapping. In: Proceedings of HDRi2014-second international conference and SME workshop on HDR imaging, 9.

[2] Goudé I, Cozot R, Banterle F. Hmd-tmo: a tone mapping operator for 360 hdr images visualization for head mounted displays. In: Proceedings of computer graphics international conference. Springer; 2019. p. 216–27.

[3] Reinhard E, Heidrich W, Debevec P, Pattanaik S, Ward G, Myszkowski K. High dynamic range imaging: acquisition, display, and image-based lighting. Morgan Kaufmann; 2010.

[4] Banterle F, Artusi A, Debattista K, Chalmers A. Advanced high dynamic range imaging. AK Peters/CRC Press; 2017.

[5] Fechner GT, Howes DH, Boring EG. Elements of psychophysics, 1. Holt, Rinehart and Winston New York; 1966.

[6] Fairchild MD. Color appearance models. John Wiley & Sons; 2013.

[7] Ferwerda JA, Pattanaik SN, Shirley P, Greenberg DP. A model of visual adaptation for realistic image synthesis. In: Proceedings of the 23rd annual conference on computer graphics and interactive techniques. ACM; 1996. p. 249–58.

[8] Pattanaik SN, Ferwerda JA, Fairchild MD, Greenberg DP. A multiscale model of adaptation and spatial vision for realistic image display. In: Proceedings of the 25th annual conference on computer graphics and interactive techniques; 1998. p. 287–98.

[9] Pattanaik SN, Tumblin J, Yee H, Greenberg DP. Time-dependent visual adaptation for fast realistic image display. In: Proceedings of the 27th annual conference on computer graphics and interactive techniques; 2000. p. 47–54.

[10] Ledda P, Santos LP, Chalmers A. A local model of eye adaptation for high dynamic range images. In: Proceedings of the 3rd international conference on computer graphics, virtual reality, visualisation and interaction in Africa. ACM; 2004. p. 151–60.

[11] Pajak D, Cadík M, Aydin TO, Myszkowski K, Seidel H-P. Visual maladaptation in contrast domain. In: Human vision and electronic imaging XV, 7527. International Society for Optics and Photonics; 2010. p. 752710.

[12] Ritschel T, Eisemann E. A computational model of afterimages. In: Computer graphics forum, 31. Wiley Online Library; 2012. p. 529–34.

[13] E Jacobs D, Gallo O, A Cooper E, Pulli K, Levoy M. Simulating the visual experience of very bright and very dark scenes. ACM Trans Graph 2015;34(3):1–15.

[14] Boitard R, Cozot R, Thoreau D, Bouatouch K. Zonal brightness coherency for video tone mapping. Signal Process: Image Commun 2014b;29(2):229–46.

[15] Aydin TO, Stefanoski N, Croci S, Gross M, Smolic A. Temporally coherent local tone mapping of hdr video. ACM Trans Graph 2014;33(6):1–13.

[16] Eilertsen G, Mantiuk RK, Unger J. Real-time noise-aware tone mapping. ACM Trans Graph 2015;34:198.

[17] Eilertsen G, Mantiuk RK, Unger J. A comparative review of tone-mapping algorithms for high dynamic range video. In: Computer graphics forum, 36. Wiley Online Library; 2017. p. 565–92.

[18] Durand F, Dorsey J. Interactive tone mapping. In: Rendering techniques 2000. Springer; 2000. p. 219–30.

[19] Reinhard E, Stark M, Shirley P, Ferwerda J. Photographic tone reproduction for digital images. ACM Trans Graph 2002;21(3):267–76.

[20] Yu M. Dynamic tone mapping with head-mounted displays. Standford University Report2015; 5.

[21] Adelson EH, Bergen JR, et al. The plenoptic function and the elements of early vision, 2. Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology; 1991.

[22] McMillan L, Bishop G. Plenoptic modeling: An image-based rendering system. In: Proceedings of the 22nd annual conference on computer graphics and interactive techniques; 1995. p. 39–46.

[23] Debevec P. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In: Proceedings of ACM SIGGRAPH 2008 classes. ACM; 2008. p. 32.

[24] Schlick C. Quantization techniques for visualization of high dynamic range pictures. In: Photorealistic rendering techniques. Springer; 1995. p. 7–20.

[25] Mantiuk R, Mantiuk R, Tomaszewska A, Heidrich W. Color correction for tone mapping. In: Computer graphics forum, 28. Wiley Online Library; 2009. p. 193–202.

[26] Artusi A, Pouli T, Banterle F, Akyüz AO. Automatic saturation correction for dynamic range management algorithms. Signal Process: Image Commun 2018;63:100–12.

[27] Irawan P, Ferwerda JA, Marschner SR. Perceptually based tone mapping of high dynamic range image streams. In: Rendering techniques; 2005. p. 231–42.

[28] Drago F, Myszkowski K, Annen T, Chiba N. Adaptive logarithmic mapping for displaying high contrast scenes. In: Computer graphics forum, 22. Wiley Online Library; 2003. p. 419–26.

[29] LaValle S. Virtual reality2016;.

[30] LaViola Jr JJ. A discussion of cybersickness in virtual environments. ACM Sigchi Bull 2000;32(1):47–56.

[31] Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. IntJ Aviat Psychol 1993;3(3):203–20.