Dynamic Range Expansion of Video Sequences: a Subjective Quality Assessment Study

Francesca De Simone, Giuseppe Valenzise, Paul Lauga, Frederic Dufaux, Francesco Banterle

HAL Id: hal-01082628
https://hal.archives-ouvertes.fr/hal-01082628

Submitted on 13 Nov 2014

Dynamic Range Expansion of Video Sequences: a Subjective Quality Assessment Study

Francesca De Simone, Giuseppe Valenzise,
Paul Lauga, Frederic Dufaux
Multimedia Signal Processing Group
Institut Mines-Télécom, Télécom ParisTech, CNRS LTCI
Paris, France

Francesco Banterle
Visual Computing Lab
ISTI-CNR
Pisa, Italy

Abstract—The goal of this paper is to assess whether existing solutions to expand low dynamic range (LDR) still images to high dynamic range (HDR) displays can apply to moving pictures. To this end, we conduct a subjective quality assessment test on four state-of-the-art expansion operators, using high-quality input LDR content. Our results show that, perhaps surprisingly, no temporal artifacts, such as flickering due to global illumination changes, are introduced applying frame-based algorithms. In addition, they confirm previous studies in that operators as simple as linear expansion can yield acceptable (or even excellent) quality of the HDR video.

Index Terms—High dynamic range, expansion operators, HDR display, subjective quality assessment

I. Introduction

Next generation high dynamic range (HDR) displays [1] can reproduce far higher luminance and contrast levels than their LCD or CRT predecessors, and they are expected to gain a large television market share in the coming years. However, most of the existing image and video content is low dynamic range (LDR) and needs to be converted to HDR to be viewed on HDR displays. Recently, this has motivated research on LDR-to-HDR conversion [2], also known as *dynamic range expansion* or *inverse/reverse tone mapping*. Such expansion is in general an ill-posed problem, since physical radiance values are lost when the image is acquired and stored in LDR format. Thus, it requires a set of strong a priori assumptions on the original scene in order to be solved. In a more constrained scenario, where the goal is to render LDR content on an HDR screen, the problem is more tractable, though its solution is not unique. As a result, various *expansion operators* (EOs) have been proposed [2–10]. Evaluating these techniques plays a key role in the design of new EOs and helps to improve the characteristics of HDR displays.

Few works in the literature have compared the performance of EOs [11–13]. Assessing the quality of expanded HDR content is not simply a matter of *fidelity* to the original LDR, but entails complex and challenging *aesthetic* considerations, difficult to model in an objective quality metric. As a consequence, these studies are mainly based on subjective visual quality assessment experiments. However, to our knowledge, there is no formal and detailed subjective study yet that

has explicitly analyzed how existing EOs perform when the temporal dimension is added to the problem.

In this paper, we focus on the expansion of LDR video, with the aim to answer the following research questions: *Are existing EOs adequate to support video? If yes, which is the best algorithm?* A subjective test based on pairwise comparisons has been designed, to rank four EOs of different algorithmic complexity, representative of the main approaches proposed so far. We target the problem of real-time displaying of high-quality professional content, which is representative of a typical broadcasting scenario. Thus, EOs that cannot be run in real-time, due to heavy processing or time-consuming offline training, have been excluded from the comparison. Our results show that existing EOs yield an acceptable (in some case, even excellent) visual quality, and even if frame-based, they do not produce temporal artifacts, such as flickering due to global illumination changes.

The rest of the paper is organized as follows: in Section II we review the main approaches in the literature to LDR expansion and previous subjective studies on EOs; we describe the experimental setup in Section III and the proposed test methodology in Section IV. Results are discussed in Section V. Finally, Section VI concludes the paper.

II. Background and related work

A. Expansion Operators (EOs)

Expansion operators can be broadly classified according to how the LDR pixel values are expanded to the luminance range of the HDR display. In the simplest algorithms, the same *global expansion* function is applied to all image pixels. Akyuz et al. [11] found that a linear expansion of the gamma-corrected LDR values to the available display luminance range can yield a good HDR experience. However, the provided content needs to be correctly exposed. In the case of over-exposed LDR content (i.e., large saturated areas), a non-linear expansion (e.g., adaptive gamma) needs to be employed to avoid loss of visual contrast [13].

The class of *segmentation* EOs attempts to reproduce the levels of contrast associated to highlights, such as specular reflections or light sources, which result in saturated pixels in the LDR image. Meylan et al. [4, 5] proposed to classify pixels in two classes: diffuse and specular regions. Diffuse

(a) "vespa"   (b) "firedance"   (c) "sunset"

(d) "tennis"   (e) "stairs"   (f) "naples"

Fig. 1. Sample frames from the six test sequences used for the test.

regions are expanded linearly up to a fraction ρ of the available HDR display luminance. The remaining $1 - \rho$ fraction of the display range is then linearly allocated to the specular regions. Didyk et al. [7] employed a multi-class approach with learning that requires manual intervention from the user. The method in [6] requires heavy manual intervention, too. A more general approach is the *local* class, in which the expansion is localized by employing a pixel-wise *expand map*, which describes the amount of expansion to be applied to each pixel. This approach was proposed in different works [2, 3, 8].

The last category of EOs collects algorithms that expand LDR content based on the *local contrast* perceived by the human visual system. Among these, the EO of Huo et al. [9] expands LDR pixels by inverting the local retina response without the need of segmentation or expand maps. Instead, the EO by Kuo et al. [10] inverts a class of parametric tone mapping operators (TMOs), i.e. functions used to display HDR content on LDR screens, by selecting parameters based on scene classification. However, this approach focuses on reconstructing the original photometric radiance values of the scene, rather than displaying LDR content on HDR displays, which is the scenario considered in this paper.

B. Subjective quality assessment of EOs

The subjective visual quality of expanded HDR pictures was firstly considered by Akyuz et al. [11], who designed an experiment to analyze the reasons why HDR visual experience is superior to the LDR one. The original HDR picture was compared with the LDR image, expanded via global linear and non-linear expansion, targeting different luminance and contrast levels. The LDR image was chosen among those used to create the HDR image, as the subjective best exposure. Results showed that, overall, average luminance is more important than contrast for quality perception, and even simple linear scaling of a correctly exposed LDR picture can successfully yield a good HDR experience, as long as the viewer judges *visual appeal* rather than *fidelity to the real scene*.

The performance of five EOs [2–4, 6, 11] were compared in [12], to assess which expanded content resembles more the original HDR picture. LDR images used as input to the EOs were produced from the source HDR content, by automatically selecting the best exposure. The results of the test indicate that, when viewers are asked to judge *fidelity to the HDR source*

rather than visual appeal, non-linear contrast enhancement [2, 3, 6] is preferable to simple linear expansion [4, 11].

Masia et al. [13] considered the effect of the exposure of the LDR input on the produced HDR picture. The tested EOs include [2, 3] and linear expansion [11]. The LDR version of each image was shown to viewers, who were asked to rate the quality of each of the four renditions according to the *fidelity with respect to the real world appearance of the scene*. No HDR ground truth was presented. The reported results suggest that the tested EOs can in general perform well on originally dark content (independently from the exposure). However, for brighter inputs, over-exposed pictures lead to lower quality HDR renditions.

III. Experimental setup

A. Source LDR content and considered EOs

We analyzed several LDR video sequences, downloaded from YouTube, as potential test material to be included in our experiment. All the considered sequences had 4K spatial resolution, but we decimated them to HD (1920×1080 pixels) to meet our display's resolution. We focused on high quality video footage: no visible compression artifacts are present, the content is well exposed, and has been filmed and post-processed professionally (as in a typical broadcast scenario). Differently from previous studies [11, 12], no HDR ground truth was available. For each content, we computed the spatial and temporal indexes (SI and TI, respectively) [14], which describe the spatial and temporal complexity of the content, as well as two luminance and contrast-related features: the image key [15], which gives a measure of the brightness of the image; the Michelson contrast [16], computed locally using a 32×32 pixels sliding window. Based on the semantic interest of each content and on the diversity of the considered characteristics, we selected the six video sequences shown in Fig. 1. All videos have a duration between $4.5\ s$ and $10\ s$, and except "naples" and "tennis", which have two scene changes, the rest of the sequences feature the same visual scene. We excluded longer sequences with several scene changes for two reasons. First, based on a pilot test with expert viewers, we observed that scene changes did not influence the performance of the considered EOs. Second, the performance of EOs change with the characteristics of the scene (e.g., key, contrast, etc.), as discussed in Section V. As a result, a video segment with multiple scenes would have time-varying quality, which complicates the subjective quality assessment task. Fig. 2 reports the boxplot of content characteristics for the selected material.

In order to produce the HDR videos, we considered four EOs, representative of the different approaches proposed in the literature. The algorithms include (letters denote the abbreviation used in the rest of the paper to refer to each algorithm): (A) simple linear scaling[1] of LDR values to meet the HDR

---

[1]In order to be robust to obvious artifacts due to frame-by-frame oscillations of the minimum and maximum input LDR values, we clipped 1% of the darkest and brightest pixels before the expansion.

display luminance range [11]; (M) the highlight enhancement method by Meylan et al. [5] (with $\rho = 0.67$); (R) the method based on expand map by Rempel et al. [2]; and (H) the perceptual algorithm by Huo et al. [9].

B. Display, test environment and viewing conditions

Expanded HDR videos were displayed on the SIM2 HDR47 display, which has HD1080 resolution with declared contrast ratio higher than $4 \cdot 10^6$ [17]. Using a light probe, we verified the linear response of the monitor and we measured a peak luminance of $4250\ cd/m^2$ when 60% of the screen surface is white. The value $L_{\max} = 4250\ cd/m^2$ was taken as reference maximum display luminance for the four EOs. Display-referred HDR values for our screen are obtained by dividing physical cd/m^2 units by the luminance efficacy at equal energy white, i.e., by the constant factor 179. We set up a test space with mid gray non-reflective background, isolated from external sources of lights, as recommended in [18, 19]. A lamp at $6500\ K$ color temperature was placed behind the HDR screen to ensure ambient illumination while avoiding the presence of any direct light source (apart from the HDR display) in the field of view of the user. The resulting ambient light measured in front of the screen, when this is off, was of approximatively $10\ cd/m^2$. One viewer at time participated to our viewing sessions, sitting at a distance of 1.6 times the diagonal of the display (i.e., 2 m in our case), which corresponds to a $30°$ viewing angle.

IV. TEST METHOD

A mixed, quantitative and qualitative, test approach was used in our experiment. The viewer was presented twice with a pair of video sequences, *a* and *b*, played sequentially, with a two second mid-grey screen showing the letter *a* or *b* before the rendering of the corresponding video; *a* and *b* depict the same content, expanded using two different EOs. After the presentation, the viewer was asked to answer the following questions: 1) *"Did you prefer a, b or did you have no preference?"*, i.e. three forced-choice pair comparison (PC) method [20]; 2) *"If you had a preference, why did you prefer a or b?"*, i.e., qualitative evaluation to motivate the PC choice; 3) *"Would you say the quality of the preferred stimulus (or both if you did not have a preference) is acceptable if you would get this video on your TV?"*, i.e. acceptability rating; 4) *"How would you rate the overall quality of the preferred stimulus (or both if you did not have a preference)?"*, i.e., overall quality Absolute Categorical Rating (ACR) using 5-levels discrete quality scale [14]. Each viewer was left free to take as much time as needed to answer these questions, by directly handling the interface.

A complete block design PC method was employed and each pair of stimuli was presented in both orders: thus, being n the number of compared EOs, for each content, $n \times (n - 1) = 12$ pairs were considered, and in total 72 stimuli (12×6 contents) were presented to each viewer. To avoid viewing sessions longer than 20 minutes, the entire test was split in three sessions: each viewer performed a session

in a different day. The list of test stimuli included in each session and their order were randomized for each viewer, with the constraint that no consecutive presentations of the same content would occur. Two dummy presentations were included at the beginning of each session to stabilize viewers' scores. After completing each session, the viewer was also asked to rate the perceived visual fatigue, as in [21].

Before the first session, each viewer was asked to fill a form to characterize his/her familiarity with HDR technology and multimedia user's habits. Each viewer was also screened for correct visual and color acuity. A training session, including three video scenes (different from the test material) and the same EOs as those considered in the test, was performed before the first session, to familiarize the viewer with the interface as well as with the paper rating sheet.

V. RESULTS AND DISCUSSION

A panel of fifteen ($M = 15$) people, one third women, two thirds men, with average age of 28.8 years old (min = 22, max = 38), took part in our test. Most of them were not familiar with HDR imaging and declared not to have a HDTV at home.

In order to detect and remove possible outliers from the panel, we leveraged the transitivity property of the ordering induced by PC tests, as proposed in [22]. First, we counted for each subject how often the transitivity property was violated, i.e., for each triad of stimuli i, j, k, we checked whether: $i > j \wedge j > k \wedge (k > i \vee k \equiv i)$, or $i > j \wedge j \equiv k \wedge k > i$, or $i \equiv j \wedge j > k \wedge k > i$, or $i < j \wedge j < k \wedge (k < i \vee k \equiv i)$, or $i < j \wedge j \equiv k \wedge k < i$, or $i \equiv j \wedge j < k \wedge k < i$, where $>$ denotes the preference relation and $\equiv$ denotes a tie. Then, we computed, across the entire dataset, the transitivity satisfaction rate $\xi = 1 - C/T$, where C is the number of transitivity violations and T is the number of possible triads ($T = 24$). If ξ was less than a threshold, the subject was considered as an outlier. In our results, only one subject had transitivity satisfaction rate below 0.9. Thus, he was considered as an outlier and his results were discarded.

Table I(a) shows the normalized winning frequencies (w_{ij}), averaged across the six video contents, of having algorithm i (on the rows) preferred to algorithm j (on the columns), where $i, j = 1, 2, ..., n$ and n is the number of different EOs. These are computed as $w_{ij} = p_{ij} + t_{ij}/2$, where p_{ij} is the normalized frequency of stimulus i being preferred to stimulus j (*"I prefer a (b)"*) and t_{ij} is the normalized frequency of the tie (*"I have no preference"*). The resulting ranking of the EOs, global and per content, are shown in Table I(b).

In terms of overall ranking and winning frequencies, it can be observed that, in most cases, the simple linear expansion (A) ranks first, while R ranks last. Considering the results of the qualitative evaluation, this can be explained by the fact that in general the expanded HDR video using R was judged as "too dark" and "not colorful", giving the impression to be "unnatural" to viewers.

In terms of acceptability, the expanded videos were almost always judged acceptable in terms of visual quality (only 8% of the video stimuli was judged unacceptable by at least one
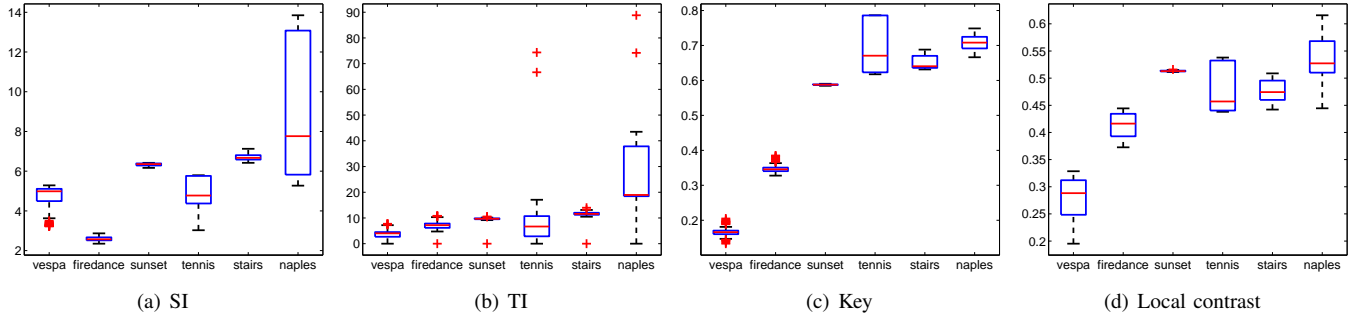
Fig. 2. Boxplot of video characteristics for the selected video contents. Both image key and local contrast have a value range of $[0, 1]$.
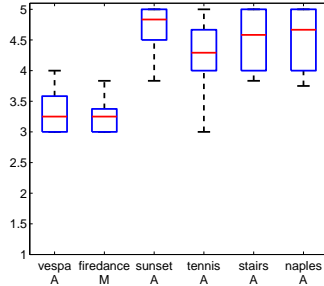

Fig. 3. Boxplot of the ACR results (1 = "Bad", 2 = "Poor", 3 = "Fair", 4 = "Good", 5 = "Excellent") across the panel of viewers for best ranking algorithm for each content (indicated by tick labels on x axis).

third of the panel). Analyzing the results of the ACR quality assessment (Fig. 3), it can be noticed that, in most of the cases, the best performing algorithms obtain good or even excellent results. The most challenging scenes are "vespa" and "firedance", which have low key and local contrast values (Fig. 2(c) and 2(d)), being mostly dark: in these cases, the ACR scores are between fair and good, and users' qualitative feedbacks report that expanded videos are "noisy". Viewers' comments also reveal that the preferred stimulus usually is the one which is deemed as "brighter" and "with more details". This somehow confirms the findings in [11], which showed that viewers prefer the brightest stimulus, as far as it displays a good contrast. This explains why the simple linear expansion algorithm (A) ranked first in the test: since the input LDR material was high quality and well exposed, the linear expansion enables to increase the dynamic range while enhancing details and giving the same overall impression of contrast as in the original LDR. Nevertheless, if the input LDR is too dark, linear expansion amplifies noise, which is more visible on brighter displays [23]. Interestingly, even in the cases when the quality of the stimulus was considered not acceptable, viewers did not signal any specific temporal artifacts such as flickering due to global illumination changes: "dark" and "noisy" are the most frequent adjectives used to describe the worst stimuli.

Finally, concerning the visual fatigue results in Table II, differently from previous studies [21], symptoms such as "eye burning/pricking", "pain around the eyes" and "headache" were slightly perceived by some users.

VI. CONCLUSION AND FUTURE WORK

The research question considered in this paper was: *what is the visual quality yielded by existing EOs when applied to video?* Our subjective tests with four representative EOs have somewhat confirmed some of the results previously found in the literature for the expansion of still images, i.e., simple algorithms such as linear range expansion can provide visually acceptable and, in most cases, very good visual quality when visual appeal is judged, rather than fidelity to the original HDR content or to reality [11]. Most interestingly, we found that none of the examined frame-based EOs introduced visible temporal artifacts, such as flickering due to global illumination changes, as it could be argued due to the frame-by-frame nature of these algorithms. Our conclusions apply to the case of high-quality, professionally shot/edited LDR content. This is a realistic scenario in the case of broadcasting. Future work will extend this scenario, considering the effects of distortions such as compression artifacts on the input LDR video.

TABLE I
RESULTS OF PC TESTS.

(a) Overall winning frequencies (w_{ij}), averaged over the six videos.

	A	H	M	R
A	0	0.76	0.65	0.88
H	0.24	0	0.45	0.81
M	0.35	0.55	0	0.74
R	0.12	0.19	0.26	0

(b) Algorithms ranking across test videos.

	1st	2nd	3rd	4th
vespa	A	M	H	R
firedance	M	A	H	R
sunset	A	H	R	M
tennis	A	H	M	R
stairs	A	M	H	R
naples	A	H	M	R
Avg	A	M	H	R

TABLE II
RESULTS OF VISUAL FATIGUE QUESTIONNAIRE (0 = "I did not perceive this symtom" - 10 = "I perceived this symtom a lot").

	Mean	Std
Double vision	0.52	1.04
Problems in focusing	1.59	1.92
Burning/pricking sensation in the eyes	2.02	2.37
Blurred vision	1.17	1.42
Tearing/watery eyes	0.26	0.48
Pain around the eyes	1.88	2.61
Headache	2.12	2.35
Image floating	0.43	0.96
Color change	0.24	0.66

REFERENCES

[1] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, "High dynamic range display systems," in *ACM SIGGRAPH 2004 Papers*, New York, NY, USA, 2004, SIGGRAPH '04, pp. 760–768, ACM.

[2] A. G. Rempel, M. Trentacoste, H. Seetzen, H. D. Young, W. Heidrich, L. Whitehead, and G. Ward, "LDR2HDR: On-the-fly reverse tone mapping of legacy video and photographs," in *ACM SIGGRAPH 2007 Papers*, New York, NY, USA, 2007, SIGGRAPH '07, ACM.

[3] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers, "Inverse tone mapping," in *Proc. of the 4th International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia*, New York, NY, USA, 2006, GRAPHITE '06, pp. 349–356, ACM.

[4] L. Meylan, S. Daly, and S. Susstrunk, "The reproduction of specular highlights on high dynamic range displays," in *Proc. of the 14th Color Imagining Conference*, 2006.

[5] L. Meylan, S. Daly, and S. Susstrunk, "Tone Mapping For High Dynamic Range Displays," in *Proc. IS&T/SPIE Electronic Imaging: Human Vision and Electronic Imaging XII*, 2007, vol. 6492.

[6] L. Wang, L. Wei, K. Zhou, B. Guo, and H. Shum, "High dynamic range image hallucination," in *Proc. 18th Eurographics Conference on Rendering Techniques*. Eurographics Association, 2007, pp. 321–326.

[7] P. Didyk, R. Mantiuk, M. Hein, and H. Seidel, "Enhancement of bright video features for HDR displays," *Computer Graphics Forum*, vol. 27, no. 4, pp. 1265–1274, 2008.

[8] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers, "Expanding low dynamic range videos for high dynamic range applications," in *Proc. of the 24th Spring Conference on Computer Graphics*. ACM, 2008, pp. 33–41.

[9] Y. Huo, F. Yang, L. Dong, and V. Brost, "Physiological inverse tone mapping based on retina response," *The Visual Computer, Springer*, vol. 30, no. 5, pp. 507–517, 2014.

[10] P. Kuo, C. Tang, and S. Chien, "Content-adaptive inverse tone mapping," in *Proc. IEEE Visual Communications and Image Processing*, November 2012, pp. 1–6.

[11] A. O. Akyüz, R. Fleming, B. E. Riecke, E. Reinhard, and H. H. Bülthoff, "Do HDR displays support LDR content?: A psychophysical evaluation," in *ACM SIGGRAPH 2007 Papers*, New York, NY, USA, 2007, SIGGRAPH '07, ACM.

[12] F. Banterle, P. Ledda, K. Debattista, M. Bloj, A. Artusi, and A. Chalmers, "A psychophysical evaluation of inverse tone mapping techniques," *Computer Graphics Forum*, vol. 28, no. 1, pp. 13–25, March 2009.

[13] B. Masia, S. Agustin, R. W. Fleming, O. Sorkine, and D. Gutierrez, "Evaluation of reverse tone mapping through varying exposure conditions.," *ACM Transactions on Graphics*, vol. 28, no. 5, 2009.

[14] ITU-T, "Subjective Video Quality Assessment Methods for Multimedia Applications," ITU-T Recommendation P.910, April 2008.

[15] A. O. Akyüz and E. Reinhard, "Color appearance in high-dynamic-range imaging," *Journal of Electronic Imaging*, vol. 15, no. 3, pp. 033001, 2006.

[16] A. Michelson, *Studies in optics*, Chicago Press, 1927.

[17] SIM2, "http://www.sim2.com/hdr/," June 2014.

[18] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," ITU-R Recommendation BT. 500, January 2012.

[19] ITU-R, "General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays," ITU-R Recommendation BT. 2022, August 2012.

[20] H. A. David, *The Method of Paired Comparisons*, Charles Griffin, 1969.

[21] A. G. Rempel, W. Heidrich, H. Li, and R. Mantiuk, "Video viewing preferences for HDR displays under varying ambient illumination," in *Proc. of the 6th Symposium on Applied Perception in Graphics and Visualization*, New York, NY, USA, 2009, APGV '09, pp. 45–52, ACM.

[22] J. Lee, F. De Simone, and T. Ebrahimi, "Subjective quality evaluation via paired comparison: Application to scalable video coding," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 882–893, 2011.

[23] T. O. Aydın, R. Mantiuk, and H. Seidel, "Extending quality metrics to full dynamic range images," in *Proc. of SPIE Electronic Imaging: Human Vision and Electronic Imaging XIII*, San Jose, USA, January 2008, pp. 6806–10.