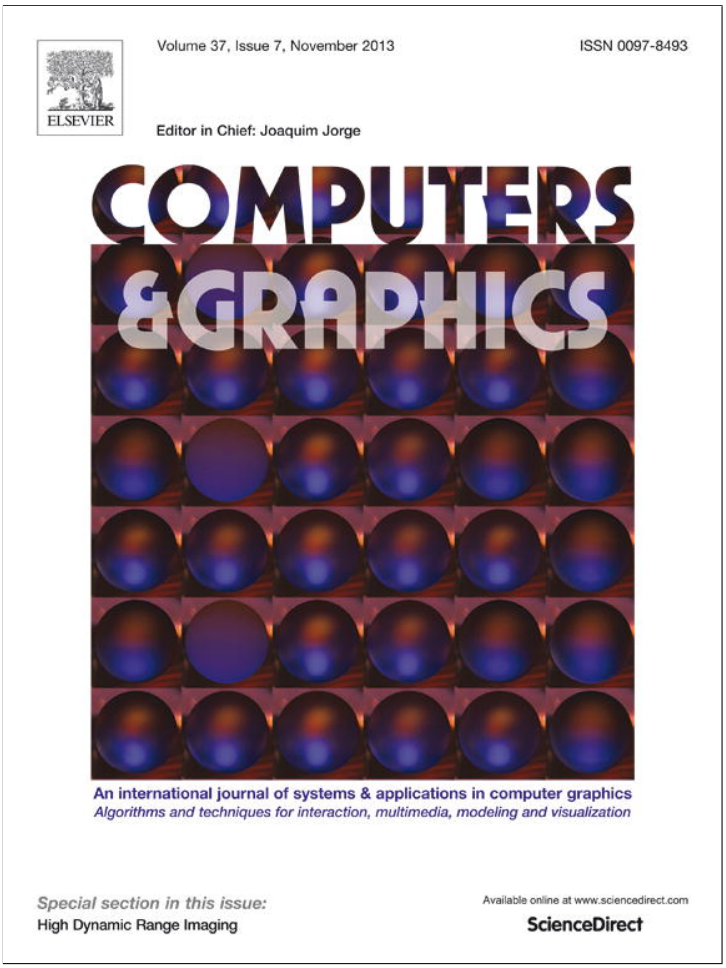

Computers & Graphics 37 (2013) 885–895

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on HDR Imaging

An evaluation of image reproduction algorithms for high contrast scenes on large and small screen display devices

Ahmet Oğuz Akyüz *, M. Levent Eksert, M. Selin Aydin

Department of Computer Engineering, Middle East Technical University, Universiteler Mahallesi, Dumlupinar Bulvari No: 1, Cankaya, 06800 Ankara, Turkey

ARTICLE INFO

Article history:
Received 1 March 2013
Received in revised form
10 July 2013
Accepted 15 July 2013
Available online 31 July 2013

Keywords:
HDR imaging
Tone mapping
Exposure fusion

ABSTRACT

Rendering high contrast scenes on display devices with limited dynamic range is a challenging task. Two groups of algorithms have emerged to take up this challenge: tone mapping operators (TMOs) and more recently exposure fusion (EF) techniques. While several formal evaluation studies comparing TMOs exist, no formal evaluation has yet been performed that compares EF techniques with each other or compares them against TMOs. Moreover, with the advancements in hand-held devices and programmable digital cameras it became possible to directly capture and view high dynamic range (HDR) content on these devices which are characterized by their small screens. However, currently very little is known about how to best visualize a high contrast scene on a small screen. Thus the primary goal of this paper is to provide answers to both of these questions by conducting a series of rigorous psychophysical experiments. Our results suggest that the best tone mapping algorithms are generally superior to EF algorithms except for the reproduction of colors. Furthermore, contrary to some previous work, we find that the differences between algorithms are barely perceptible on small screens and therefore one can opt for a simpler solution than a more complex and accurate one.

1. Introduction

Reproducing high contrast scenes on mediums with low dynamic range (LDR) has been a difficult problem. Limitations of digital cameras as well as display devices pose serious challenges. However, to some extent, the techniques of HDR imaging that were developed over the last two decades made it possible to overcome these limitations [1,2].

Among these, two specific types of algorithms are of interest to this paper. These are tone mapping operators and exposure fusion algorithms (EFAs). Tone mapping (TM) refers to the notion of controllably reducing the dynamic range of an HDR image to reproduce it as accurately as possible on a medium of lower dynamic range [3]. The input to a TMO is an HDR image and the output is an LDR one. Exposure fusion, on the other hand, takes a sequence of LDR images as input where each image has a different exposure. By extracting and merging the information from such a sequence, it attempts to create a detail-rich single LDR image. For both types of algorithms multiple solutions have been proposed (Sections 2.1 and 2.2).

The question of which algorithm provides the best solution is an important one, and several studies were conducted in an attempt to answer it (Section 2.3). However a clear picture has not yet emerged and more work is needed to reach reliable conclusions. This is also necessitated by the fact that new algorithms are being proposed on a continuous basis which are not covered by earlier experiments. Furthermore, most earlier work focused on validating TMOs; rigorous evaluations of EFAs and studies that compare EFAs with TMOs are lacking. Thus, the first goal of this paper is to provide a deeper insight into these issues.

Secondly, until recently, creating HDR images has been a tedious process where the individual exposures captured by a digital camera through manual- or auto-bracketing were combined into an HDR image offline on the computer [4]. However, we can assume that this laborious workflow is going to be replaced by an automatic approach where HDR image generation takes place exclusively inside capture devices. Such devices will have to provide instant feedback to the user by depicting the captured scene as accurately as possible. In fact, in-camera HDR feature has already been integrated into several commodity products such as Pentax K-7, as well as iPhone and Android smart phones. A recent patent filed by Canon Inc. [5] also describes a technology that allows in-camera generation of HDR images through individually varying pixel exposures [6].

However, we hypothesize that while digital cameras may allow direct capture of HDR imagery soon, small camera screens are less likely to support direct display of HDR content in near future. Therefore, the second goal of this study is to answer an important

* Corresponding author. Tel.: +90 3122105565.
E-mail address: akyuz@ceng.metu.edu.tr (A. Oğuz Akyüz).

0097-8493/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cag.2013.07.004

question that will arise during this transition: how to best display HDR content on a small screen which is characterized by smaller angular pixel size and inferior display characteristics?

2. Related work

In this section, we briefly review the previous work on tone mapping and exposure fusion by putting more emphasis on the methods that were evaluated in this study. We then discuss the previous evaluation studies and highlight the differences of our work.

2.1. Tone mapping

Tone mapping is the process that is commonly employed to visualize HDR content on LDR displays. Many TMOs have so far been proposed with different properties. TMOs are generally classified as global and local, intensity domain and frequency domain, based on human visual system, and those that are not based on any perceptual phenomena [1].

Among the most commonly used operators is the photographic TMO [7]. This operator belongs to the class of sigmoidal tone reproduction operators. It comes in two flavors, namely the global and local operator. In the current study, we evaluated the global version of this operator. Another commonly used operator is the multi-scale decomposition based algorithm of Li et al. [8]. This operator follows an analysis-synthesis framework where the input image is separated into multiple subbands which are compressed independently using different gain factors. The compressed subbands are recombined in the synthesis phase.

A more recent operator is Mantiuk et al.'s display adaptive operator [9]. By modeling the human visual system and taking into characteristics of a target display, this operator attempts to minimize the visibility of contrast distortions. The most recent operator that we included in this study is the linear windowed TMO [10]. It operates by first determining local constraints in small overlapping windows, and then integrating them to solve a global optimization problem. While the method is local, it is found to be resistant against common artifacts such as contrast reversals.

There exist many other notable tone mapping operators, although a comprehensive review is beyond our scope (see [1,2] for excellent reviews). Our motivation for focusing on these four operators can be explained as follows. The photographic TMO [7] has performed very well in earlier validation studies [11–16]. Thus determining its performance against EFAs and in multiple display conditions is valuable. Display adaptive TMO [9], on the other hand, was selected because of its ability to adjust itself for different display conditions. Because we had two different display conditions, including this operator was also essential. Finally, Li et al.'s [8] and Shan et al.'s [10] methods were selected due to the overall high quality of their results and their implementations being publicly available by the original authors.

2.2. Exposure fusion

Exposure fusion is an alternative technique to tone mapping (assuming that one has access to individual exposures). It involves directly creating a contrast-rich LDR image from a set of bracketed exposures. It is typically formalized as

$$H(x,y) = \sum_{i=1}^{N} W_i(x,y) I_i(x,y) \quad (1)$$

where I_i are the input images, W_i are the weight maps, and H is the fused image. Different EFAs are distinguished by their computation of weight maps and blending strategies.

For instance, Goshtasby partitions the image into tiles of uniform size and computes entropy for each tile [17]. The tiles with the maximum entropy are selected for the fusion process. Blending is performed via rational Gaussians [18]. Mertens et al. [19], on the other hand, operate on individual pixels and weight pixels based on their saturation, contrast, and well-exposedness. Blending is performed via Laplacian pyramids [20].

In a more recent algorithm, Zhang and Cham [21] propose to use gradient magnitudes as an indicator of the exposure quality. The rationale is that in detail-rich regions gradient magnitude will be high and it will tend toward zero for under- and over-exposed regions. The weight maps calculated based on the gradient magnitudes are refined using cross-bilateral filtering [22] and blending is performed by using multiresolution splines [23].

In this study, we included Goshtasby's [17], Mertens et al.'s [19], and Zhang and Cham's [21] EFAs. Goshtasby's and Mertens et al.'s methods were included as being two of the pioneering works in exposure fusion. Zhang and Cham's method was included due to the high quality of its results despite being a simple algorithm.

2.3. Subjective quality evaluation

Evaluating image quality is a challenging task in computer graphics. To date, the most reliable method remains to be subjective evaluation which is based on collecting responses from human observers [1]. However, subjective evaluation is tedious and time consuming, and require meticulously designed experiments. Many subjective evaluation studies for comparing tone mapping operators have hitherto been conducted. These include pairwise comparison without a reference [24], with HDR display reference [14], and with real-world scene reference [25]. Direct ranking or rating experiments of several algorithms with and without reference were also performed [13,26,15,27] as well as experiments using non-photographic stimuli [28].

In another subjective evaluation, the observers were asked to modulate several TMO parameters such as contrast, brightness, and saturation [29]. The results of this study were later used for developing generic [30] and display adaptive tone mapping operators [9]. More recently, Urbano et al. conducted an experiment using small-screen devices and found that humans' preference for TMOs is different for these devices than for desktop monitors [16].

Our study is partially inspired by the work of Urbano et al., although there are important differences. In that study, only TMOs were compared whereas in ours we compare TMOs and EFAs. Urbano et al.'s study did not include recent notable tone mapping operators such as [8–10]. Finally, their evaluation was conducted on a low resolution PDA device, whereas we used a Canon EOS550D dSLR camera to simulate a more realistic HDR image capture scenario.

3. Experimental design and analysis

Various experimental designs have been used in previous tone mapping evaluation studies. These include rating [11,13], ranking [26], and pairwise comparisons with or without reference [14,28,16]. Each method have their inherent advantages and disadvantages. For instance, rating experiments require the observers to make accurate judgements on a continuous or discrete scale. Making such judgements is difficult for naïve observers especially if there is no reference. Ranking studies, on the other hand, require the observer to order the stimuli according to a given criteria. Their difficulty arises from the fact that the observer needs to compare multiple stimuli at the same time and put an order on items which may not be ordered due to circular

preferences known as *triads*. The paired comparison design circumvents these drawbacks as only two items are shown at a time, simplifying the observer's task, and allowing for the possibility of detecting triads. However, it takes the longest time as each stimulus needs to be compared with each other. More importantly, comparison designs cannot assess how close the compared images are to their real world counterparts.

In this study, we opted for the method of paired comparisons in our experimental framework mainly because it produces the smallest measurement variance and produces the most accurate results [31]. This method involves presenting pairs of images to observers and recording their judgement for each pair. If there are t algorithms to compare the total number of judgements amounts to $T=\binom{t}{2}$. Each observer's judgements are recorded in a preference matrix. In this matrix, a value of 1 in cell (i,j) indicates that the method in row i is preferred over the method in column j.

3.1. Kendall's ζ

The method of paired comparisons has certain benefits over a rating or ranking experiment, albeit it takes longer to execute. First, the task is easier for observers as they only need to judge the items two at a time instead of making a decision for all of them at once. Secondly, it allows for *triads*, i.e. circular preferences, to occur. For instance, an observer's preference for three items may be $A \rightarrow B$, $B \rightarrow C$, and $C \rightarrow A$. While this might suggest that the observer is not capable of making a reliable judgement, i.e. the observer is inconsistent, it may also mean that the evaluated items cannot be linearly ordered. In general, one can deduce that if all observers are inconsistent the items are either too similar or impossible to judge in a linear order. However, if only a few subjects are inconsistent, they can be deemed to be poor judges and their data can be discarded. The consistency can be computed by using the following formula [32]:

$$\zeta = \begin{cases} 1-\dfrac{24c}{t^3-t} & \text{if } t \text{ is odd,} \\ 1-\dfrac{24c}{t^3-4t} & \text{if } t \text{ is even,} \end{cases} \tag{2}$$

where c is the number of triads.

3.2. Differences among algorithm scores

As in analysis of variance (ANOVA), one first needs to determine whether there are significant differences between the scores of the algorithms as a whole before partitioning them into similarity subgroups. To this we used the least significant difference method [33]. If we denote the total score of each algorithm by a_i where $i=1,2,\ldots,t$ and the desired significance level by p, this method involves comparing

$$D = 4\left[\sum_{i=1}^{t} a_i^2 - \frac{1}{4}tn^2(t-1)^2\right]/(nt) \tag{3}$$

with the upper $100p\%$ point of the χ^2 distribution with $(t-1)$ degrees of freedom where n is the number of subjects. The null hypothesis that all algorithms are equal can be rejected if D value is greater than the critical χ^2 value. In our experiments we set $p=0.05$ following the common practice.

3.3. Test of equality of two pre-assigned treatments

Determining a statistically significant D value indicates that there are indeed differences between the evaluated methods that are not likely due to chance. The question then becomes how to determine which methods are significantly different from each other. This can be answered by using the test of equality of two pre-assigned treatments [33]:

$$m_c = \lceil 1.96(0.5nt)^{0.5} + 0.5 \rceil \tag{4}$$

Any score difference not less than m_c can then be considered as statistically significant.

4. Experiments

Having described the details of our experimental design and analysis, we now focus on the specifics of our study. We included the following algorithms in our evaluation. The labels in parenthesis serve as their identifiers in the rest of the text. The motivation for selecting them as well as their main advantages and disadvantages were discussed in Section 2.

- Tile-based fusion [17] (A).
- Subband TMO [8] (B).
- Display adaptive TMO [9] (C).
- Exposure fusion [19] (D).
- Photographic TMO [7] (E).
- Linear windowed TMO [10] (F).
- Gradient-directed fusion [21] (G).

4.1. Stimuli

We used images of four different scenes captured with a Canon EOS 550D camera using exposure bracketing. For all scenes, we used 9 exposures captured in the camera RAW format which is known to be quantimetrically linear with respect to sensor irradiance [34,35]. We validated the linearity of our own camera by capturing a color checker at different exposure levels and manually comparing the pixel values of various patches. All scenes together with their false color visualizations and log relative luminance histograms are shown in Fig. 1. Note that the images were not calibrated up to an absolute scale and therefore these visualizations are only meaningful for each image separately (i.e. the same color may map to different luminance for each image). The dynamic ranges of each image are reported in Table 1. As can be seen from this table, most images had similar dynamic ranges except for the *toys* image which had a lower dynamic range. The *lamp* image had the highest dynamic range especially after excluding 1% of the darkest and lightest pixels.

The reason for using the RAW format as opposed to JPEG output can be explained as follows. Exposure fusion algorithms do not require the knowledge of the camera response function (CRF). They simply fuse the input images in the non-linear color space in which they were originally recorded. Tone mapping algorithms, on the other hand, take as input HDR images which need to be created from linearized exposures using the inverse CRF. As several CRF recovery methods exist, with each producing potentially different response curves [36], using one of these algorithms would make our results liable to the accuracy of that algorithm. Therefore, we started with RAW images and converted them to JPEG format using the non-linear mapping defined in the sRGB standard [37]. These images were directly used as input for exposure fusion algorithms.

As for tone mapping, we first created the HDR images using [1]

$$I_j = \sum_{i=1}^{N} \frac{f^{-1}(p_{ij})w(p_{ij})}{t_i} \Big/ \sum_{i=1}^{N} w\left(p_{ij}\right), \tag{5}$$

where N is the number of exposures, p_{ij} is the value of pixel j in image i, f is the camera response function (inverse sRGB gamma, in this case), w is a weighting function used to attenuate the

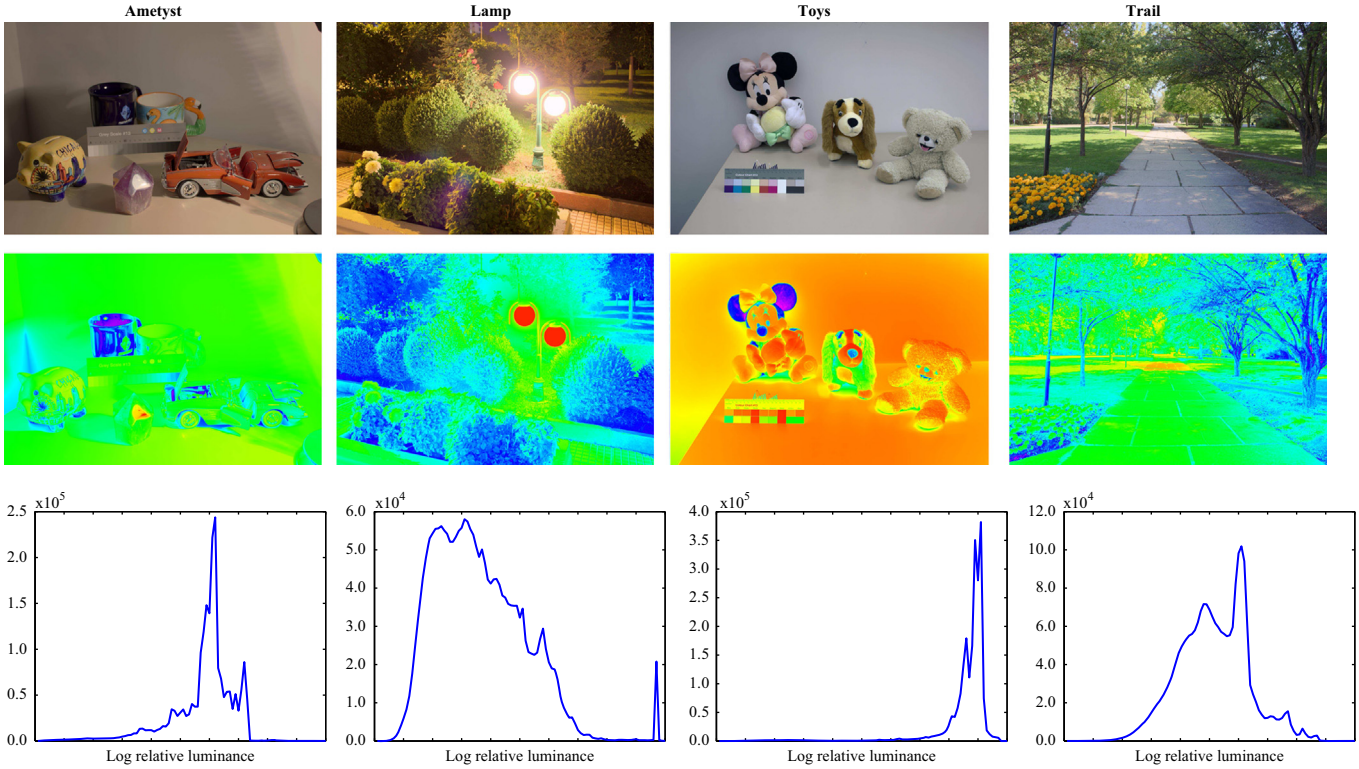

Fig. 1. The images that we used in our experiments together with their false color representations and histograms.

Table 1
The dynamic ranges of the images we used in our experiments. DR represents the order of magnitude difference between the maximum and minimum luminances. DR^1 represents the dynamic range after excluding 1% of outliers from both ends.

Dynamic range	Ametyst	Lamp	Toys	Trail
DR	4.27	4.51	2.99	4.33
DR^1	2.40	3.75	2.08	2.46

Table 2
Parameters of the algorithms used in our experiments. For all TMOs we used a gamma value of 0.45 and did not perform post tone mapping saturation adjustment.

Alg.	Parameters
A	$d=160$, $\sigma=160$, $\Delta=32$
B	$\alpha=0.2$, $\gamma=0.6$
C	$E_{amb}=10$, $L_{min}=0.5$, $k=0.01$, $L_{max}=80$
D	$\omega_C=1$, $\omega_E=1$, $\omega_S=1$
E	$\alpha=0.18$, $L_{white}=1e20$
F	$\epsilon=0.1$, $\kappa=0.05$, *window size* $=3$, $\beta_1=0.6$, $\beta_2=0.2$, $\beta_3=0.1$, $s=1$
G	$l=9$, $\tau=0.9$

contribution of poorly exposed pixels [38], and t_i is the exposure time. We then tone mapped the HDR images using the evaluated algorithms with their default parameters (see Table 2). The tone mapped images were output in the sRGB color space as well. As such, both fused and tone mapped images were stored in the same color space to make them comparable with each other.

As for producing the final stimuli, we used bicubic interpolation to match the original resolution of the images (5184×3456) to that of the target display devices (1920×1200 for the desktop display and 720×480 for the camera display). The aspect ratio was preserved during resizing. We opted for bicubic interpolation as it produces fewer artifacts than bilinear and nearest neighbor interpolation. However, we did not conduct further experiments to investigate the effect of the interpolation strategy on the final results.

4.2. Experiment one: color, contrast, and detail

Our first experiment was aimed at evaluating naturalness of colors, reproduction of contrast, and visibility of details. To evaluate algorithms with respect to these three criteria we used images of 3 real world scenes. These included an outdoor day scene (*trail*), an outdoor night scene (*lamp*), and an indoor scene of closeby objects (*ametyst*). The algorithms' outputs for all three scenes are shown in Fig. 2 (the fourth scene, *toys*, is used in the second experiment).

The experimental procedure was the method of paired comparisons. The participants were presented with two images in random order from the stimuli. They could switch between the images to analyze their differences. Once they indicated their decision, a neutral gray screen was shown for a brief period followed by a new pair of random images. The observers first judged all pairs for the attribute of color, then contrast, and finally detail. Going through the entire sequence once for each attribute instead of interleaving the attributes appeared to be an easier task as the observers did not have to constantly change their criteria of decision.

All stimuli were shown on a NEC SpectraView Reference 241W monitor calibrated to the sRGB profile using an X-Rite i1Display Pro colorimeter. The peak display luminance was set to 80 cd/m^2 for full sRGB compliance. The black level was measured as 0.5 cd/m^2. The participants viewed the display in a dark room approximately from 70 cm, although no head mounting was used to avoid discomfort. At this distance, the angular size of a center pixel was approximately 0.022° in both dimensions.

A gender balanced set of 15 naïve observers between the ages 20 and 35 with normal or corrected to normal color vision took part in this part of the experiment. The duration of the experiment

Fig. 2. The output of all algorithms for all images.

did not exceed 30 min for each observer. Participants did not appear to have difficulty in understanding the meaning of the evaluated attributes. Therefore, no pre-experiment training was conducted. The participants' instructions for experiment one are given in Appendix A for reference.

4.3. *Experiment two: similarity*

The second experiment was similar to the first one except that we measured similarity of the resulting images to the real world scenes. In addition to the NEC monitor, we introduced a second

display device namely the LCD screen of a Canon EOS 550D camera. The maximum luminance and the black level of the camera's LCD screen were measured as 125 cd/m^2 and 0.5 cd/m^2 respectively. The screen was 3 in. diagonally and had a resolution of 720×480. The viewers held the camera at approximately 30 cm from their eye point which resulted in an angular pixel size of 0.017° in both dimensions. About 15 observers participated in the second experiment. Only one observer participated in both experiments.

As the display conditions changed for the second experiment, we adjusted the parameters of the display adaptive TMO accordingly. To this end, we set $L_{max} = 125$ for the camera screen and $E_{amb} = 250$ when the room lights were turned on (only for the *toys* scene).

Due to difficulties in programmatically controlling the camera's LCD screen, a manual approach was adopted for recording the observers' answers. To this end, all of the randomly selected image pairs from the stimuli were uploaded to the camera's memory card. Between each pair a neutral gray image was inserted to mark the boundary between different stimuli. The participants could switch between the images within a pair using the camera's control buttons. The participants indicated their answers verbally to the experimenter by specifying the overlaid file name (e.g. IMG_1001, IMG_1002, and so on). These answers were decoded after the experiment to create the preference matrices.

Only two scenes were used in the second experiment. The first one was the *ametyst* image from the first experiment and the second one was a set of toys under standard fluorescent office lighting (Fig. 3). The instructions for the second experiment are given in Appendix B for reference.

Fig. 3. Our experimental setup for the similarity experiments.

5. Results and analysis

The accumulated preference matrices for both experiments are reported in Tables 3 and 4 respectively. In these tables, each block represents the accumulated preferences of all observers for a given scene and attribute condition. The bottom blocks (labeled as Total) demonstrate the aggregated preferences across all scenes. Sim.-d and Sim.-c denote the similarity results for desktop and camera viewing conditions.

Before creating these matrices, we measured the consistency of all observers. As can be seen from Table 5, most observers attained a high degree of consistency. As none of the observers appeared to be making random decisions (i.e. inconsistencies were not by chance), we did not discard any observer's data in the following analysis. An interesting observation was that the color attribute on average received a lower consistency score than contrast and detail suggesting that it was harder to make judgements for color than the other two.

Next, we performed the significance analysis. As can be seen from Table 6, there were statistically significant differences for each scene and attribute combinations. Therefore we proceeded with separating the algorithms into statistical similarity groups.

At this point, it is necessary to note that this experiment produced a large amount of data which is impossible to discuss in entirety in this paper. Therefore, we conducted most of the following analysis based on the total scores. The full set of results are provided in the supplementary materials.

As can be seen from Fig. 4, EFAs were judged to produce more natural colors compared to TMOs. Based on the score rankings, three of the first four operators were found to be EFAs. Zhang and Cham's method [21] received the highest score; however, it was not statistically better than Mertens et al.'s algorithm [19]. The highest scoring TMO was Mantiuk et al.'s display adaptive operator [9], which performed significantly better than the other TMOs.

As for contrast and detail, Li et al.'s operator [8] outperformed all other methods. It stood out as statistically better than the other methods for detail, but was found to be in the same significance group with Mertens et al. [19] for contrast. Based on these, we can argue that if Li et al.'s TMO is augmented with a post tone mapping color correction step it can prove to be a very effective TMO. However, it is also worth noting that this operator is not free of halo artifacts as shown in Fig. 5. This appeared to be the main reason for why this operator received low scores in the similarity experiment discussed below.

Let us make a few more observations before discussing the second experiment. It appeared that Shan et al.'s linear windowed TMO received low scores for all three attributes. We believe that this could be due to: (1) the algorithm in general produced darker images, and (2) it gave rise to lower color saturation. It is interesting to note that while Li et al.'s method was not preferred due to production of over-saturated images, Shan et al.'s method was not preferred for the opposite reason. This, again, underlines the importance of correct color management in tone mapping. Secondly, Mertens et al.'s EFA performed consistently well in all three attributes. As such, it stood out as a balanced and effective algorithm.

As for similarity, Mantiuk et al.'s operator [9] outperformed all others for both desktop and camera conditions. We attribute this to two factors: (1) as this operator includes a model of the human visual system it could be expected that it computes images that are more similar to human perception, and (2) as this operator takes the display conditions into account such as peak luminance, black level, and ambient illumination it can adapt to changing display conditions while other operators cannot without parameter tuning.

Finally, five of the seven operators attained very similar preference scores for the camera display condition. This suggests that the limited form factor of the camera display hides any significant differences between different algorithms. As such, it may be acceptable to opt for a simpler approach than seeking out complicated and more accurate algorithms when it comes to tone mapping for the small screen.

Another interesting finding for this case was that Zhang and Cham's gradient directed EFA [21] received very low similarity scores. Deeper analysis reveals that this is mainly caused by its very dark reproduction of colors for the toys scene. As this scene is characterized by large uniform regions with small gradients, it may have posed a worst case scenario for this operator.

6. Discussion

The overall findings of this study are three-fold: (1) Li et al.'s TMO [8] is better in reproducing contrast, detail, and Mantiuk

Table 3
The accumulated preference matrices of all observers for the first experiment.

Algorithm	Color								Contrast								Detail							
	A	B	C	D	E	F	G	**T**	A	B	C	D	E	F	G	**T**	A	B	C	D	E	F	G	**T**
Ametyst																								
A	0	12	3	2	5	7	2	**31**	0	3	11	8	14	14	9	**59**	0	0	9	11	9	6	13	**48**
B	3	0	2	5	3	6	5	**24**	12	0	12	8	15	15	14	**76**	15	0	15	13	14	15	15	**87**
C	12	13	0	10	10	9	11	**65**	4	3	0	5	13	15	9	**49**	6	0	0	10	5	5	13	**39**
D	13	10	5	0	6	8	10	**52**	7	7	10	0	14	14	13	**65**	4	2	5	0	7	4	15	**37**
E	10	12	5	9	0	9	11	**56**	1	0	2	1	0	5	2	**11**	6	1	10	8	0	8	12	**45**
F	8	9	6	7	6	0	5	**41**	1	0	0	1	10	0	3	**15**	9	0	10	11	7	0	13	**50**
G	13	10	4	5	4	10	0	**46**	6	1	6	2	13	12	0	**40**	2	0	2	0	3	2	0	**9**
Lamp																								
A	0	7	9	5	11	11	3	**46**	0	10	5	4	5	13	8	**45**	0	0	8	3	8	15	10	**44**
B	8	0	9	6	13	8	4	**48**	5	0	6	4	7	14	8	**44**	15	0	14	14	14	15	14	**86**
C	6	6	0	4	12	6	3	**37**	10	9	0	12	9	13	10	**63**	7	1	0	4	5	15	13	**45**
D	10	9	11	0	14	10	5	**59**	11	11	3	0	6	13	11	**55**	12	1	11	0	10	15	15	**64**
E	4	2	3	1	0	4	3	**17**	10	8	6	9	0	13	11	**57**	7	1	10	5	0	15	11	**49**
F	4	7	9	5	11	0	3	**39**	2	1	2	2	2	0	2	**11**	0	0	0	0	0	0	1	**1**
G	12	11	12	10	12	12	0	**69**	7	7	5	4	4	13	0	**40**	5	1	2	0	4	14	0	**26**
Trail																								
A	0	12	8	9	10	10	5	**54**	0	3	2	0	10	10	2	**27**	0	1	10	0	7	10	4	**32**
B	3	0	3	0	3	5	2	**16**	12	0	10	12	14	14	13	**75**	14	0	15	14	15	14	15	**87**
C	7	12	0	7	8	9	4	**47**	13	5	0	5	10	13	11	**57**	5	0	0	0	3	10	0	**18**
D	6	15	8	0	10	10	9	**58**	15	3	10	0	14	13	11	**66**	15	1	15	0	12	14	13	**70**
E	5	12	7	5	0	11	4	**44**	5	1	5	1	0	9	3	**24**	8	0	12	3	0	12	3	**38**
F	5	10	6	5	4	0	4	**34**	5	1	2	2	6	0	2	**18**	5	1	5	1	3	0	1	**16**
G	10	13	11	6	11	11	0	**62**	13	2	4	4	12	13	0	**48**	11	0	15	2	12	14	0	**54**
Total																								
A	0	31	20	16	26	28	10	**131**	0	16	18	12	29	37	19	**131**	0	1	27	14	24	31	27	**124**
B	14	0	14	11	19	19	11	**88**	29	0	28	24	36	43	35	**195**	44	0	44	41	43	44	44	**260**
C	25	31	0	21	30	24	18	**149**	27	17	0	22	32	41	30	**169**	18	1	0	14	13	30	26	**102**
D	29	34	24	0	30	28	24	**169**	33	21	23	0	34	40	35	**186**	31	4	31	0	29	33	43	**171**
E	19	26	15	15	0	24	18	**117**	16	9	13	11	0	27	16	**92**	21	2	32	16	0	35	26	**132**
F	17	26	21	17	21	0	12	**114**	8	2	4	5	18	0	7	**44**	14	1	15	12	10	0	15	**67**
G	35	34	27	21	27	33	0	**177**	26	10	15	10	29	38	0	**128**	18	1	19	2	19	30	0	**89**

Table 4
The accumulated preference matrices of all observers for the second experiment. Sim.-d and Sim.-c denote the similarities for desktop and camera display conditions.

Algorithm	Sim.-d								Sim.-c							
	A	B	C	D	E	F	G	**T**	A	B	C	D	E	F	G	**T**
Ametyst																
A	0	9	6	6	7	8	7	**43**	0	4	12	7	7	10	9	**49**
B	6	0	4	7	8	9	11	**45**	11	0	13	10	12	12	12	**70**
C	9	11	0	4	9	10	8	**51**	3	2	0	3	7	11	5	**31**
D	9	8	11	0	12	10	10	**60**	8	5	12	0	10	12	9	**56**
E	8	7	6	3	0	8	6	**38**	8	3	8	5	0	10	7	**41**
F	7	6	5	5	7	0	8	**38**	5	3	4	3	5	0	6	**26**
G	8	4	7	5	9	7	0	**40**	6	3	10	6	8	9	0	**42**
Toys																
A	0	15	2	11	10	5	14	**57**	0	14	3	10	7	4	13	**51**
B	0	0	0	2	2	2	4	**10**	1	0	1	2	3	0	6	**13**
C	13	15	0	12	13	9	15	**77**	12	14	0	14	12	7	14	**73**
D	4	13	3	0	6	2	14	**42**	5	13	1	0	4	2	15	**40**
E	5	13	2	9	0	1	14	**44**	8	12	3	11	0	3	14	**51**
F	10	13	6	13	14	0	13	**69**	11	15	8	13	12	0	13	**72**
G	1	11	0	1	1	2	0	**16**	2	9	1	0	1	2	0	**15**
Total																
A	0	24	8	17	17	13	21	**100**	0	18	15	17	14	14	22	**100**
B	6	0	4	9	10	11	15	**55**	12	0	14	12	15	12	18	**83**
C	22	26	0	16	22	19	23	**128**	15	16	0	17	19	18	19	**104**
D	13	21	14	0	18	12	24	**102**	13	18	13	0	14	14	24	**96**
E	13	20	8	12	0	9	20	**82**	16	15	11	16	0	13	21	**92**
F	17	19	11	18	21	0	21	**107**	16	18	12	16	17	0	19	**98**
G	9	15	7	6	10	9	0	**56**	8	12	11	6	9	11	0	**57**

Table 5
Consistency values of all observers. Note that the similarity experiment was performed by a different set of observers except observer 15 who participated in both experiments.

Obsv.	Ametyst					Lamp			Trail			Toys	
	Color	Contrast	Detail	Sim.-d	Sim.-c	Color	Contrast	Detail	Color	Contrast	Detail	Sim.-d	Sim.-c
1	0.786	1.000	1.000	0.643	0.857	0.929	1.000	0.929	0.929	1.000	1.000	0.643	0.786
2	0.429	0.857	0.643	0.571	1.000	1.000	1.000	1.000	0.929	0.929	0.929	0.643	0.929
3	0.571	0.786	0.714	0.643	1.000	0.929	0.929	0.929	0.714	1.000	1.000	0.286	1.000
4	0.714	0.929	0.714	0.571	0.714	1.000	1.000	1.000	0.714	0.929	1.000	1.000	1.000
5	0.714	0.857	0.929	0.643	0.929	0.643	0.857	0.714	0.500	0.571	0.571	1.000	1.000
6	1.000	1.000	1.000	0.357	0.929	0.758	0.786	0.929	0.714	0.714	0.571	0.786	0.786
7	0.571	0.786	0.929	0.357	0.857	0.643	1.000	0.929	0.714	0.929	1.000	0.714	1.000
8	0.286	1.000	1.000	0.714	1.000	0.929	1.000	0.857	1.000	1.000	1.000	0.929	0.929
9	0.571	0.786	0.929	0.286	0.714	0.758	0.857	1.000	0.929	0.643	0.857	1.000	0.214
10	0.714	0.857	0.643	0.857	0.929	0.643	0.857	1.000	0.714	0.929	0.929	0.857	0.929
11	0.286	0.929	0.929	0.500	0.786	0.429	0.857	0.768	0.571	1.000	0.857	0.929	1.000
12	1.000	0.500	1.000	0.429	0.929	0.571	1.000	1.000	0.857	0.929	1.000	0.929	1.000
13	0.571	1.000	0.929	1.000	0.929	0.571	0.786	0.857	0.786	1.000	1.000	0.929	0.714
14	0.786	0.929	0.714	0.929	1.000	0.571	0.929	0.857	0.857	1.000	0.929	0.929	0.857
15	0.571	0.857	0.714	1.000	1.000	0.857	0.857	1.000	1.000	0.857	1.000	1.000	0.857
Avg.	0.638	0.872	0.852	0.633	0.905	0.749	0.914	0.918	0.795	0.895	0.910	0.838	0.867

Table 6
D values for the first experiment (Section 3.2). At significance level of $p=0.05$ and 6 degrees of freedom, values greater than 12.59 are statistically significant.

Scene	Color	Contrast	Detail	Sim.-d	Sim.-c
Ametyst	46.63	139.20	121.68	14.78	51.20
Lamp	63.47	66.67	165.94	–	–
Trail	57.37	113.83	162.21	–	–
Toys	–	–	–	145.52	134.63
Total	77.54	222.88	315.94	82.23	29.93

et al.'s TMO [9] is better in preserving similarity to the actual scenes when the images are presented on a desktop display; (2) exposure fusion algorithms tend to be better in color reproduction; (3) the differences between the operators are largely subdued when the images are viewed on a camera screen. In this section, we will attempt to provide possible explanations for each of these findings.

Before interpreting these findings it may be worthwhile to consider whether these two types of algorithms (EFAs and TMOs) can be considered as part of a common framework. First, one should note that the input to an EFA is a set of bracketed exposures that are stored in the non-linear color space of a digital camera. TMOs, on the other hand, take as input an HDR image that is ideally stored in a linear color space. At this point, two extra pieces of information have already been used in the tone mapping pipeline: (1) the camera response function for linearizing the input exposures and (2) the exposure times used for normalizing the exposures. Thus, one can argue that a TMO has a more complete view of the scene that is being reproduced. Theoretically, as part of its processing pipeline, a TMO can reproduce the individual exposures from the given HDR image and perform a task similar to exposure fusion. An EFA, on the other hand, is designed to work without the knowledge of the CRF nor the exposure times. Thus, compared to TMOs, EFAs have a more limited view of the scene being processed. From this point of view, TMOs appear to have an advantage over EFAs.

However, an alternative interpretation is also possible. EFAs directly combine the input exposures that are stored in the non-linear color space of a digital camera. This color space is often optimized to look appealing to most human observers. Because the processing is done directly in this space, the fused image exhibits its characteristics as well. In theory, a TMO can also output images in a carefully designed color space that is appealing to humans. However, in practice most TMOs focus on luminance compression rather than optimizing the tone curves.

What makes some TMOs' contrast and detail reproduction better? We offer the following explanation. EFAs work by computing a weight map for each exposure based on various criteria (e.g. well-exposedness, saturation, contrast, etc.). This weight map guides which pixel in which proportion will contribute to the fused result. However, in order to reduce the seams between pixels coming from different images, the computed weight maps are blurred prior to fusion. This results in blending between pixels that are not necessarily well-exposed which may cause loss of small details in high frequency image regions. As for contrast, EFAs take the best exposed regions from multiple photographs. While this minimizes the presence of under- and over-exposed regions it can also reduce the amount of contrast between dark and light regions. Both of these effects are demonstrated using an artificial bipartite luminance distribution in Fig. 6.

As for color reproduction, it is well-known that tone mapping algorithms distort colors because of the non-linear luminance compression that they employ. In general, they compress high luminances more than medium and low luminances leaving the chromaticities intact. However, according to the Hunt effect, the colorfulness of a stimulus depends on its luminance [39]. Therefore changing only the luminance and leaving the chromaticities intact do not preserve color appearance. A few ad hoc solutions exist to tackle this problem. Among these, the most commonly used one is to adjust the saturation by

$$\mathbf{C}_o = \left(\frac{\mathbf{C}_i}{L_i}\right)^s L_o, \tag{6}$$

where $\mathbf{C}_i = (R_i, G_i, B_i)$ represents the input color, $\mathbf{C}_o$ represents the output color, and L_i and L_o represent the input and output luminances respectively. The problem of this approach is that there is no principled way of setting the correct value for the saturation parameter, s.

More recently, Mantiuk et al. [40] proposed a color correction method for TMOs based on psychophysical experiments. Their method allows one to determine the value of the saturation parameter based on the degree of the luminance compression. However, the proposed method is only applicable if the dynamic range compression function can be represented as a power curve. As local operators do not comply with this requirement this correction method cannot be applied for them.

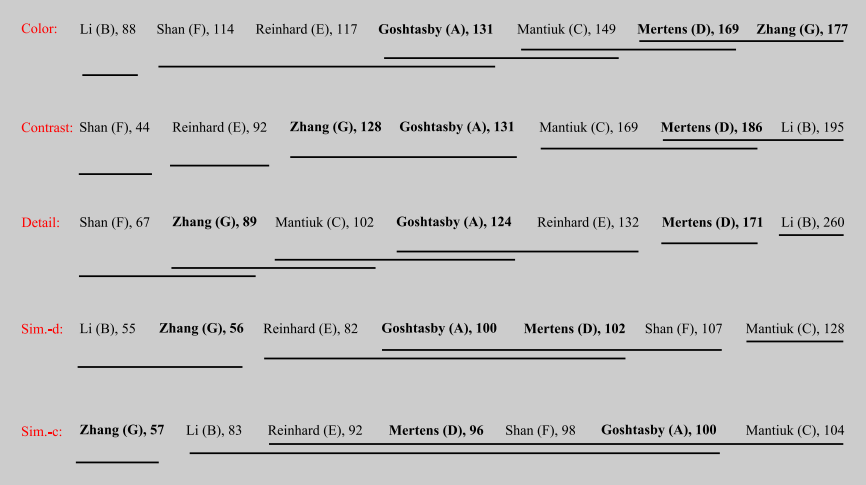

Fig. 4. Significance groups based on the total scores for each attribute. The algorithms underlined by the same line segment are statistically similar. EFAs are shown in bold.

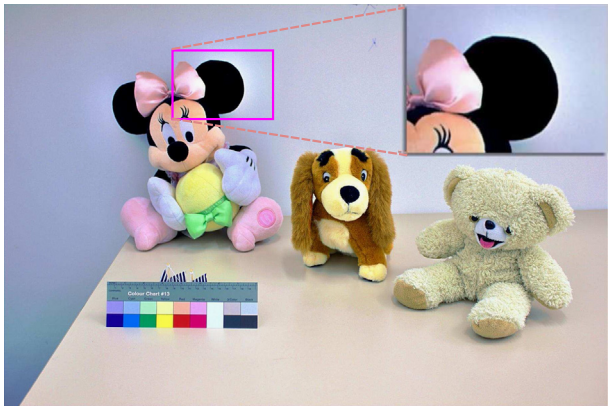

Fig. 5. Halo artifacts produced by Li et al.'s subband TMO [8].

Due to this immature state of color management in tone mapping, in this study we opted to avoid any ad hoc or limited solutions. Our results indicate that without color correction, TMOs are surpassed by EFAs in color reproduction.

Finally, we discuss the subdued differences between the algorithms in the camera display condition. As reported in Section 4, we measured the angular size of the central pixel for the desktop display and the camera screen as 0.022° and 0.017° respectively. This corresponds to approximately 45 and 59 cycles per degree respectively. It is well-known that the human visual system has an inverted U shaped contrast sensitivity with respect to spatial frequency. The contrast sensitivity function (CSF) proposed by Mannos and Sakrison is defined as [41]

$$A(f) = 2.6(0.0192 + 0.114f)\exp(-0.114f^{1.1}), \quad (7)$$

where f represents the spatial frequency. As we can see from Fig. 7, the CSF makes a peak around $f = 8$ cycles per degree and decreases for higher frequencies. In particular, it can be seen that the sensitivity at the spatial frequency of the camera screen is lower than that of the desktop monitor. Therefore, it may be argued that some of the high frequency content that influences the observers' decisions are not visible on the camera display which in turn results in subdued differences between the algorithms.

As in most psychophysical experiments, our study also has several caveats. First, in our experiments, we did not attempt to match the pixel angular size between the two display conditions. Our goal was to evaluate the performance of the algorithms under typical viewing distances. However, it would be interesting to conduct further research to study the impact of this parameter. Secondly, we emphasize that all of the tested algorithms have several parameters and our results are only valid for those parameter values that we used in our experiments (Table 2). As in all TMO validation studies, the rankings of the operators could have been different for different sets of parameters. Finally, in our experiments, we used a total number of four images. While this number is not atypical for TMO validation studies, one should be careful about generalizing the results to a diverse set of real-world environments.

7. Relationship to earlier TMO validation studies

In order to reach reliable conclusions about how various algorithms actually perform, it is important to compare and contrast results of different validation studies. Unfortunately, this is a difficult task as neither all studies evaluate the same algorithms nor they ask the same research questions. However, we will attempt to provide a brief analysis to elucidate our findings in light of the earlier work.

In previous studies, there appeared to be a general agreement that the photographic TMO [7] performed among the best operators [11–16]. It was sometimes accompanied by other operators such as bilateral filtering [42], histogram adjustment [43], iCAM [25], and gradient domain compression [44] although they were less consistent.

Among these notable algorithms we only included the photographic TMO in our experiments as a leading operator. Our findings, however, suggest that the newer TMOs such as the subband method [8] and the display adaptive TMO [9] generally outperform the photographic TMO. Shan et al.'s linear windowed operator [10], on the other hand, could surpass the photographic TMO only for the similarity task but not in the color, contrast, or detail tasks. Also most EFAs appeared to perform better than the photographic TMO.

8. Conclusions

Our study reveals several interesting findings. First, we found that the best TMO has outperformed the best EFA for preserving contrast, detail, and similarity. However, all TMOs fall short when it comes to accurate reproduction of colors. This emphasizes

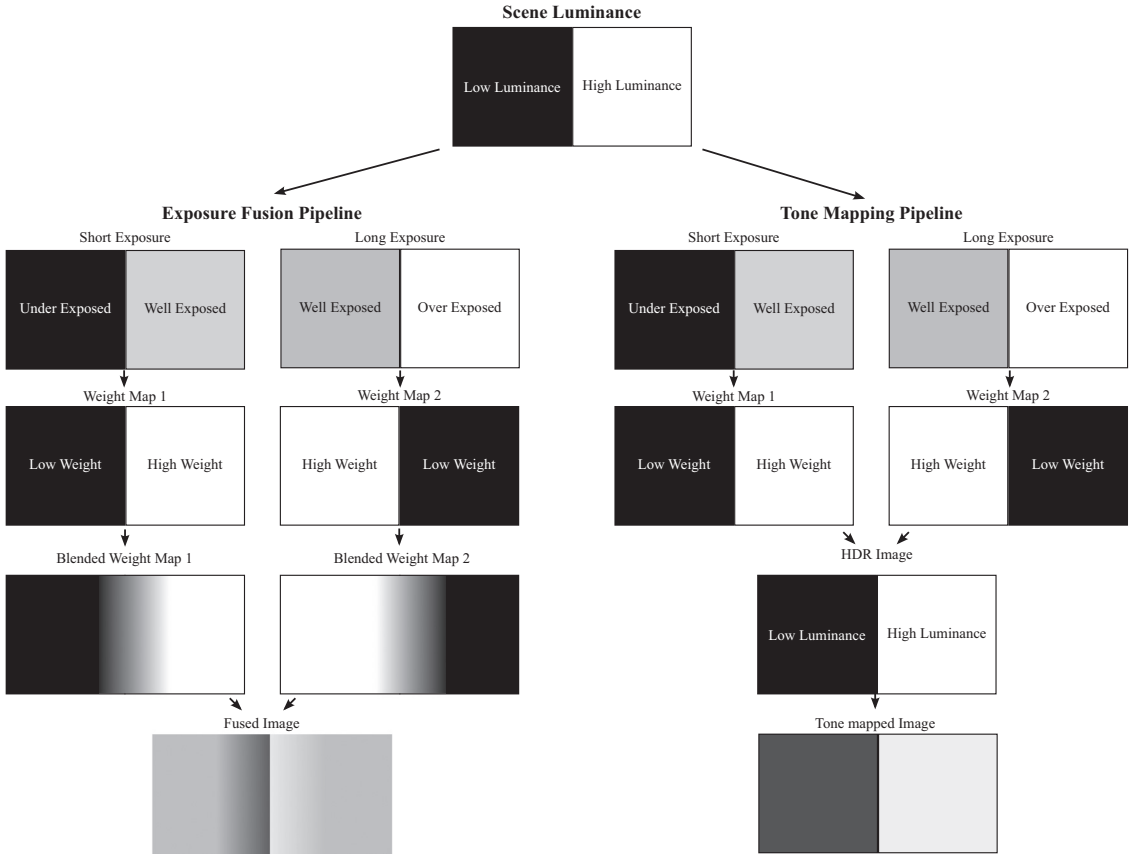

Fig. 6. The comparison of the two imaging pipelines. A high contrast scene is captured using two exposures (short and long) where one half of the scene is well-exposed in each exposure. In both pipelines weight maps are generated for each exposure. However, while the weight maps are blurred in the EF pipeline to avoid seams during blending of exposures, no blurring occurs in the TM pipeline (see Eq. (5)). Due to this blending, the small scale details may be lost in the EF pipeline. Also because the best exposed regions are taken from the input exposures, the contrast between the light and dark regions may be subdued in the EF pipeline.

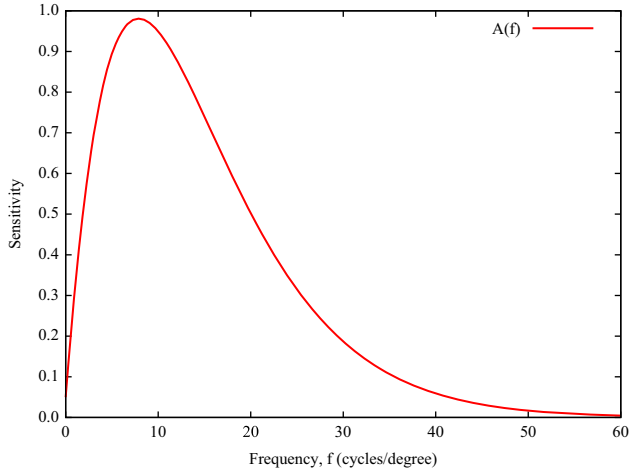

Fig. 7. The contrast sensitivity function modeled by Mannos and Sakrison [41].

the importance of addressing color issues for tone mapping. Our second finding was that rankings of operators indeed change between small screen and large screen devices confirming the results of earlier experiments [16]. However, we also found that most operators have comparable performance when their results are viewed on a small screen which may be attributed to the reduced contrast sensitivity of the human eye at increased spatial frequencies. This suggests that when it comes to tone mapping for the small screen one may get away with simpler and more efficient algorithms.

Appendix A. Instructions for experiment one

In this experiment, you will be presented with multiple image pairs. For each pair, you will be asked to choose the image that you prefer according to three different criteria, namely:

1. Naturalness of colors.
2. Sensation of contrast.
3. Visibility of details.

For (1), choose the image whose colors appear more natural to you. For (2), choose the image that appears to have more contrast. For (3), choose the image where details are more visible.

The experiment should take about 30–40 min assuming that you spend 10–15 s on each image pair. To avoid fatigue, do not spend too much on each stimulus. Note that you can terminate the experiment if you feel any discomfort.

Thank you for your participation. You can ask the experimenter should you have any questions.

Appendix B. Instructions for experiment two

In this experiment, you will be presented with multiple image pairs in two different display devices. For each pair, you will be asked to choose the image that you think is more similar to its real world version.

For the desktop monitor, switch between the images using the keys “1” and “2” on the numpad. When you make your decision, press the “enter” key to continue with the next pair.

For the camera LCD monitor, switch between the images using the camera's "left" and "right" buttons. When you make your decision, tell the name of that image to the experimenter. You can then move on to the next pair (each pair is separated by a gray image).

The experiment should take about 15–20 min assuming that you spend 10–15 s on each image pair. To avoid fatigue, do not spend too much on each stimulus. Note that you can terminate the experiment if you feel any discomfort.

Thank you for your participation. You can ask the experimenter should you have any questions.

Appendix C. Supplementary data

Supplementary data associated with this article can be found in the online version at http://dx.doi.org.10.1016/j.cag.2013.07.004.

References

[1] Reinhard E, Ward G, Pattanaik S, Debevec P. High dynamic range imaging: acquisition, display and image-based lighting. second edition. San Francisco: Morgan Kaufmann; 2010.

[2] Banterle F, Artusi A, Debattista K, Chalmers A. Advanced high dynamic range imaging: theory and practice. first edition. Natick, MA: CRC Press (AK Peters); 2011.

[3] Tumblin J, Rushmeier H. Tone reproduction for computer generated images. IEEE Computer Graphics and Applications 1993;13(6):42–8.

[4] Debevec PE, Malik J. Recovering high dynamic range radiance maps from photographs. In: SIGGRAPH 97 conference proceedings; 1997. p. 369–78.

[5] Canon Inc. Image capturing apparatus and image capturing method. US 2010/0141792 A1; 2010.

[6] Nayar S, Branzoi V. Adaptive dynamic range imaging: optical control of pixel exposures over space and time. In: Proceedings of the 9th IEEE international conference on computer vision, vol. 2; 2003. p. 1168–75. http://dx.doi.org/10.1109/ICCV.2003.1238624.

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

[8] Y. Li, L. Sharan, E.H. Adelson, Compressing and companding high dynamic range images with subband architectures. In: SIGGRAPH '05: ACM SIGGRAPH 2005 papers. New York, NY, USA: ACM Press; 2005, p. 836–44.

[9] Mantiuk R, Daly S, Kerofsky L. Display adaptive tone mapping. ACM Transactions on Graphics 2008;27(3) 68:1–10.

[10] Shan Q, Jia J, Brown M. Globally optimized linear windowed tone mapping. IEEE Transactions on Visualization and Computer Graphics 2010;16(4):663–75, http://dx.doi.org/10.1109/TVCG.2009.92.

[11] Drago F, Martens WL, Myszkowski K, Seidel HP. Perceptual evaluation of tone mapping operators with regard to similarity and preference. Technical Report MPI-I-2002-4-002. Max Plank Institut für Informatik; 2002.

[12] Kuang J, Yamaguchi H, Johnson GM, Fairchild MD. Testing HDR image rendering algorithms. In: Proceedings of IS&T/SID 12th color image conference; 2004. p. 315–20.

[13] Yoshida A, Blanz V, Myszkowski K, Seidel H. Perceptual evaluation of tone mapping operators with real-world scenes. In: Woods Andrew J, Bolas Mark T, Merritt John O, McDowall Ian E, editors. Stereoscopic displays and virtual reality systems XII. Proceedings of the SPIE, vol. 5666; 2005, p. 192–203.

[14] Ledda P, Chalmers A, Troscianko T, Seetzen H. Evaluation of tone mapping operators using a high dynamic range display. ACM Transactions on Graphics 2005;24(3):640–8.

[15] Čadík M, Wimmer M, Neumann L, Artusi A. Image attributes and quality for evaluation of tone mapping operators. In: Proceedings of the 14th Pacific Conference on Computer Graphics and Applications. Taipei, Taiwan; 2006, p. 35–44.

[16] Urbano C, Magalhães L, Moura J, Bessa M, Marcos A, Chalmers A. Tone mapping operators on small screen devices: an evaluation study. Computer Graphics Forum 2010;29(8):2469–78, http://dx.doi.org/10.1111/j.1467-8659.2010.01758.x.

[17] Goshtasby AA. Fusion of multi-exposure images. Image and Vision Computing 2005;23(6):611–8, http://dx.doi.org/10.1016/j.imavis.2005.02.004.

[18] Goshtasby A. Design and recovery of 2-d and 3-d shapes using rational Gaussian curves and surfaces. International Journal of Computer Vision 1993;10(3):233–56.

[19] Mertens T, Kautz J, Van Reeth F. Exposure fusion. In: 15th Pacific conference on computer graphics and applications, PG'07; 2007. p. 382–90. http://dx.doi.org/10.1109/PG.2007.17.

[20] Burt P, Adelson E. The Laplacian pyramid as a compact image code. IEEE Transactions on Communications 1983;31(4):532–40, http://dx.doi.org/10.1109/TCOM.1983.1095851.

[21] Zhang W, Cham WK. Gradient-directed composition of multi-exposure images. In: IEEE conference on computer vision and pattern recognition (CVPR); 2010. p. 530–6. http://dx.doi.org/10.1109/CVPR.2010.5540168.

[22] Paris S, Durand F. A fast approximation of the bilateral filter using a signal processing approach. In: Leonardis A, Bischof H, Pinz A, editors. Computer vision—ECCV 2006. Lecture notes in computer science, vol. 3954. Berlin/Heidelberg: Springer; 2006. p. 568–80 ISBN 978-3-540-33838-3.

[23] Burt PJ, Adelson EH. A multiresolution spline with application to image mosaics. ACM Transactions on Graphics 1983;2(4):217–36.

[24] Drago F, Martens WL, Myszkowski K, Chiba N. Design of a tone mapping operator for high dynamic range images based upon psychophysical evaluation and preference mapping. In: IS&T SPIE electronic imaging 2003. The human vision and electronic imaging VIII conference; 2003. p. 321–31.

[25] Kuang J, Johnson GM, Fairchild MD. icam06: a refined image appearance model for HDR image rendering. Journal of Visual Communication and Image Representation 2007;18(5):406–14, http://dx.doi.org/10.1016/j.jvcir.2007.06.003 Special issue on High Dynamic Range Imaging, URL: ⟨http://www.sciencedirect.com/science/article/pii/S1047320307000533⟩.

[26] Ashikhmin M, Goyal J. A reality check for tone-mapping operators. ACM Transactions on Applied Perception 2006;3(4):399–411.

[27] Čadík M, Wimmer M, Neumann L, Artusi A. Evaluation of HDR tone mapping methods using essential perceptual attributes. Computers & Graphics 2008;32(3):330–49.

[28] Akyüz AO, Reinhard E. Perceptual evaluation of tone reproduction operators using the Cornsweet–Craik–O'brien illusion. ACM Transactions on Applied Perception 2008;4(4):1–29.

[29] Yoshida A, Mantiuk R, Myszkowski K, Seidel HP. Analysis of reproducing real-world appearance on displays of varying dynamic range. Computer Graphics Forum 2006;25(3):415–26.

[30] Mantiuk R, Seidel HP. Modeling a generic tone-mapping operator. Computer Graphics Forum 2008;27(2):699–708.

[31] Mantiuk RK, Tomaszewska A, Mantiuk R. Comparison of four subjective methods for image quality assessment. Computer Graphics Forum 2012;31(8):2478–91, http://dx.doi.org/10.1111/j.1467-8659.2012.03188.x.

[32] Kendall MG, Smith BB. On the method of paired comparisons. Biometrika 1940;31(3/4):324–45.

[33] Starks TH, David HA. Significance tests for paired-comparison experiments. Biometrika 1961;48(1/2):95–108.

[34] Manders C, Mann S. Determining camera response functions from comparagrams of images with their raw datafile counterparts. In: Proceedings of international symposium on intelligent multimedia, video and speech processing; 2004. p. 418–21. http://dx.doi.org/10.1109/ISIMP.2004.1434089.

[35] Chakrabarti A, Scharstein D, Zickler T. An empirical camera model for internet color vision. In: Proceedings of the British machine vision conference; 2009.

[36] Gençtav A, Akyüz AO. Evaluation of radiometric camera response recovery methods. In: SIGGRAPH Asia 2011 posters, SA '11. New York, NY, USA: ACM. ISBN 978-1-4503-1137-3; 2011, p. 15.

[37] ITU. ITU-R Recommendation BT.709-5. Parameter values for the HDTV standards for production and for international programme exchange. Geneva: ITU (International Telecommunication Union); 2002.

[38] Mitsunaga T, Nayar SK. Radiometric self calibration. In: Proceedings of CVPR, vol. 2; 1999. p. 374–80.

[39] Fairchild MD. Color appearance models. second editionJohn Wiley & Sons; 2005.

[40] Mantiuk R, Mantiuk R, Tomaszweska A, Heidrich W. Color correction for tone mapping. Computer Graphics Forum 2009;28(2):193–202.

[41] Mannos J, Sakrison D. The effects of a visual fidelity criterion of the encoding of images. IEEE Transactions on Information Theory 1974;20(4):525–36 http://dx.doi.org/10.1109/TIT.1974.1055250.

[42] Durand F, Dorsey J. Fast bilateral filtering for the display of high-dynamic-range images. ACM Transactions on Graphics 2002;21(3):257–66.

[43] Ward G, Rushmeier H, Piatko C. A visibility matching tone reproduction operator for high dynamic range scenes. IEEE Transactions on Visualization and Computer Graphics 1997;3(4).

[44] Fattal R, Lischinski D, Werman M. Gradient domain high dynamic range compression. ACM Transactions on Graphics 2002;21(3):249–56.