Tone Mapping Operators on Small Screen Devices: An Evaluation Study

C. Urbano¹, L. Magalhães²³, J. Moura²⁴, M. Bessa²⁴, A. Marcos⁵ and A. Chalmers⁶

¹Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria, Leiria, Portugal
carlos.urbano@ipleiria.pt
²Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal
{lmgalha, jpmoura, maxbessa}@utad.pt
³Institute for Systems and Computer Engineering of Porto (INESC Porto), Porto, Portugal
⁴Knowledge Engineering and Decision Support Research Center (GECAD), Porto, Portugal
⁵Universidade Aberta/DCT, Lisboa, Portugal
marcos@univ-ab.pt
⁶The Digital Laboratory, WMG, University of Warwick, UK
alan.chalmers@warwick.ac.uk

Abstract

In the last decade, an increasing number of techniques have been developed to reproduce high dynamic range imagery on traditional displays. These techniques, known as Tone Mapping Operators (TMOs), have been compared and ranked in different ways according to several image characteristics. However, none of these algorithms has been developed specifically for small screen devices (SSD). In this paper, we present an evaluation of currently used TMOs to show that SSDs with limited size, resolution and colour depth require specific research to find or create an appropriate solution. The research described in this paper is based on psychophysical experiments; using three different types of displays (CRT, LCD and SSD). The obtained results show that rankings obtained are similar for the LCD and CRT but are significantly different for the SSD. Furthermore, these rankings show additionally that some characteristics of TMOs need to be emphasized to obtain better high-fidelity mapped images for SSDs.

Keywords: high dynamic range, human visual system, image reproduction, psychophysics, small screen device, tone mapping, visual perception


1. Introduction

Nowadays, it is easy to obtain high dynamic range (HDR) images either by capturing them from the real world with a typical semi-pro digital camera or by rendering them on computer. The vast ranges of intensities that are encoded within these images are similar to what Human Visual System (HVS) can perceive.

Although high dynamic range imagery is easily obtainable problems arise regarding the reproduction of such rich images using common modern display devices that have low dynamic range (LDR). Therefore, it is necessary to use techniques that scale-down the dynamic range to fit the range of the display while preserving the appearance of the HDR image. To solve or minimize that problem, a large number of high-quality tone mapping operators (TMOs) have been developed.

1.1. Tone mapping operators

These mapping algorithms can be classified as global or local [Devlin’02]. The former are simple and fast as they map each
pixel based on its intensity and global image characteristics, regardless of the pixel’s spatial location. The later take into account the pixel’s surroundings to be mapped. This means that a pixel of a given intensity will be mapped to a different value depending on whether it is located in a dark or bright area.

Another aspect concerning tone mapping reproduction that was not taken into account in the above classification is time. If a TMO was designed for handling animated sequences of images considering, the HVS adaptation over time it can be designated as time dependent. Usually these TMOs are global. On the other hand if a TMO was designed specifically for isolated images, it is designated as time-independent. Recently, some algorithms were specially created to achieve interactive rates when running on modern graphics hardware.

Another important issue regarding TMOs lies on image reproduction. There are three possible approaches: perceptual, cognitive and aesthetical [Cadik’06]. In the classical perceptual approach, the tone mapping operator tries to simulate the human vision process. For example, due to scotopic vision a scene viewed at night would be represented as blurred and nearly monochromatic. The cognitive approach is used when it is important to understand the fine details or the structure of the visible lines in the result. Finally, if the goal is merely a pleasant appearance of the image, then it is called an aesthetical approach. In this paper, we concentrate only on the classical perceptual approach with the aim to characterize the image quality in a perceptual sense.

Despite this large number of TMOs and premises, none of them were specifically conceived for small screen devices (SSDs) with very low dynamic range (VLDR). However, some studies were also trying to improve image visualization on SSDs such as the refinement of the CIE colour appearance model [Park’07] and recently the development of an adaptive tone mapping that minimize visible contrast distortions for a range of output devices, ranging from e-paper to HDR displays [Mantiuk’08]. This technique can adjust image for optimum contrast visibility taking into account ambient illumination and display characteristics.

SSDs, also known as small form-factor devices (SFF), are characterized by limited size, resolution and colour depth. Recent mobile devices are equipped with 24-bit displays with VGA resolution (640 × 480) and can go up to 5” in size [Capin’08]. However, typically they have LCDs screens with about 2.5”–3.5”, a resolution of 320 × 240 (QVGA), and 16-bit colour-depth.

1.2. Objectives

The goal of this study is to verify whether or not the development of TMOs specifically for SSD with its known limitations in size, resolution and colour depth needs a special and different approach comparing with the existing ones. To avoid misunderstandings it is important to emphasize that the obtained rankings resulting from psychophysical experiments are not supposed to be a formal comparison of TMOs because the TMO’s default parameters were used and also because this work does not include some new important tone mapping algorithms.

2. Previous Work

In recent years, some TMOs comparisons have been done using psychophysical experiments where subjects compared images mapped with different approaches. These studies differ in some aspects from each other. Some use a comparison reference such as the real scene or the correspondent HDR image showed on an HDR display device; some make pairwise comparison; others ask subjects to rate images. To better analyse, them a brief description of those most often referred in the literature are described next.

In 2003, Drago et al. had as the objective the definition of a comparison methodology for existing TMOs to better understand their strengths and weaknesses [Drago’03a].

Yoshida et al. conducted a psychophysical experiment based on a direct comparison between the appearance of real-world scenes and HDR images of these scenes displayed on an LDR monitor [Yoshida’05]. The main goal of this experiment was to assess the differences in how tone mapped images are perceived and to find out which attributes of image appearance account for these differences when tone mapped images are compared directly with their corresponding real-world scenes rather than with each other.

Ledda et al. made a series of experiments using for the first time an HDR display device for TMOs evaluation [Ledda’05]. The main purpose of this investigation was not simply to determine which is the best algorithm compared to the HDR reference, but more generally, to propose an experimental methodology to validate such operators.

In 2006, Cadik et al. presented an overview of image quality attributes of different TMOs and proposed a schema of relationship between these attributes leading to the definition of an overall image quality measure [Cadik’06]. The authors performed subjective psychophysical tests to prove the proposed relationship scheme and also to evaluate existing tone mapping methods with regard to these attributes. In 2008, Cadik et al. extended their previous experiments with two new scenes (an outdoor scene and a night scene) using an identical design conception [Cadik’08].

The work done by Ashikhmin and Goyal in 2006 showed that it is important to use real environments and scenes to meaningfully judge and compare relative performance of tone-mapping techniques [Ashikhmin’06].

Yoshida et al. in 2006 had as major outcome a better understanding of how users adjust TMO parameters to achieve...
Table 1: Resume of previous TMO comparison studies.

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th># TMOs</th>
<th>Display Type</th>
<th>Comparison Method</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dragoi et al. (2003)</td>
<td>11</td>
<td>6</td>
<td>CRT</td>
<td>pairwise</td>
<td>detail, naturalness, contrast</td>
</tr>
<tr>
<td>Yoshida et al. (2005)</td>
<td>14</td>
<td>6</td>
<td>LCD</td>
<td>all + real</td>
<td>brightness, detail</td>
</tr>
<tr>
<td>Ledda et al. (2006)</td>
<td>28</td>
<td>6</td>
<td>LCD</td>
<td>pairwise + real (HDR)</td>
<td>similarity</td>
</tr>
<tr>
<td>Cadik et al. (2006)</td>
<td>10</td>
<td>6</td>
<td>CRT</td>
<td>printouts</td>
<td>ranking, detail, contrast, artifacts</td>
</tr>
<tr>
<td>Ashikmin and Goyal (2006)</td>
<td>15</td>
<td>5</td>
<td>1024x768</td>
<td>all + real</td>
<td>similarity</td>
</tr>
<tr>
<td>Yoshida et al. (2006)</td>
<td>25</td>
<td>6</td>
<td>HDR</td>
<td>image (parameters adjustment)</td>
<td>preference</td>
</tr>
<tr>
<td>Cadik et al. (2006)</td>
<td>10</td>
<td>6</td>
<td>HDR</td>
<td>image + real (rating-scale)</td>
<td>performance, naturalness, contrast, sharpness, highlights, contrast, shadow color</td>
</tr>
<tr>
<td>Kuang et al. (2007)</td>
<td>19</td>
<td>7</td>
<td>LCD</td>
<td>pairwise</td>
<td>natural clutter, overall accuracy</td>
</tr>
<tr>
<td>Kuang et al. (2007)</td>
<td>19</td>
<td>7</td>
<td>LCD</td>
<td>pairwise</td>
<td>image + real (rating-scale)</td>
</tr>
</tbody>
</table>

This study about prior experiments for TMO comparison was used to conceive our own experimental design. The advantages and disadvantages of each experiment were very helpful to avoid an ad hoc and erroneous experimental design. As an example of some important item we take into account were the use of the real scene as reference so that each participant decision was not made by self preference or likeness but by similarity with the real scene; the use of a considerable and statistically accepted number of participants in each experiment; paired comparison; maximum randomization; constant environment illumination conditions, etc.

It is important to remark that the aim of this work is not comparison of the results of our work with the ones above because the TMOs, attributes, conditions and mainly the goals are completely different.

Our experimental design is described in Section 3.1.

3. Research Methodology

Great care must be taken when designing psychophysical experiments to ensure that the experimental framework is valid and robust and ensure the use of randomization and avoiding biased variables.

3.1. Experimental design

An initial experimental design was defined and presented at IASK2007 [Urbano’07]. This has been subsequently modified to improve its suitability for the goals of this work. A description of the experimental design is described later.

Two sets of experiments were conducted each one using a different scene. Each set of experiments had three parts, one using two traditional CRT displays, a second using two traditional LCD displays and a third using two SSD displays (PDAs). In this study, the term traditional display refers to CRTs or LCDs with a screen size not less than 14 inch. The PDAs have been chosen because they present characteristics that are representative of a typical SSD. In the experiment’s description, the term PDA is used throughout instead of SSD.

In all cases within the experiments, subjects were asked to make paired comparison (also known as forced choice comparison or two alternative forced choice) and choose one of the two tone mapped images shown on the two calibrated display devices that looked most similar to the real scene in front of them (Figure 1). In each comparison, the subjects made the choice based on four (separately) specific image properties being tested. The properties evaluated were colour, detail, contrast and naturalness. These were chosen because they are known to well characterize an image [Cadik’06] and also because they are simple to understand by participants that are not experts in computer graphics. Examples of different images varying the four image properties were shown to each participant for better and homogeneous understanding.
Although the use of paired comparison technique implies a large number of trials which becomes time consuming, such a technique makes it easier to evaluate and compare the performance of each subject and is considered to be more accurate and precise than rating [Kuang’07].

3.1.1. Research hypothesis

We predicted that the most appropriate TMO is different for PDAs compared to a traditional display. The experimental hypothesis was thus that there exists a significant difference between the obtained TMO ranking for a PDA and the ones obtained for traditional displays. This hypothesis was tested against the null hypothesis ($H_0$), which maintains that there are no differences between rankings.

$$H_0 : \text{ranking}_{\text{SSD}} = \text{ranking}_{\text{CRT}} = \text{ranking}_{\text{LCD}}.$$  

3.1.2. Participants

Six different groups of 19 people each participated in the experiments (two scenes $\times$ three display types), making a total of 114 subjects. Although these included several graduate students having diverse backgrounds, most of them were not experts in tone mapping or human perception. All subjects had normal or corrected to normal vision, with ages ranging from 18 to 40 years old. Enough time was given to each observer to adapt to the light level at the experiment location before starting the experiment (from 2 to 3 min.). No time limit was imposed on the selection process.

3.1.3. Conditions

A scene with a high range of intensities was carefully created for each experimental set. They are indoor scenes under controlled illumination environment comprising a light source and many specifically chosen objects. All scenes included Macbeth charts and two different kind of Snellen chart, see Figures 2 and 3 because these objects enabled the subjects to better evaluate the colour and detail attributes when comparing each pair of tone mapped images with the real scene in front of them.

The HDR images were acquired for each scene using a Canon EOS 350D digital camera. Three photos with varying exposition time were taken from a tripod in the same position and viewing angle the participants would have. The final HDR images were generated using Photomatix Pro [Photomatixpro] (Figure 4). The other scene is illustrated in Figure 5 and is characterized by having objects with high level of specularity.

Seven tone mapped images were then created using Qtpfsgui [Qtpfsgui] and the tone mapping operators used were: Gradient Domain [Fattal’02], Spatially Varying [Ashikhmin’02], Bilateral Filtering [Durand’02], Logarithmic Mapping [Drago’03b], Local Photoreceptor Based [Pattanaik’00], Photographic Tone Reproduction [Reinhard’02] and Photoreceptor Model [Reinhard’05].
These TMOs were chosen for the experiment because we already have their implementations and are believed to represent the high-quality techniques for still images.

The default parameters of each TMO were used since that was not relevant for the purpose of the experiments.

Each participant was asked to observe all possible pair combinations of the tone mapped images, for a total of 21 different pairs of images. On every occasion each participant observed two images plus the real scene. The real scene was located at the centre and should be used as reference for the comparisons. At the left and right, tone mapped versions of the reference were shown on calibrated display devices (CRT, LCD or PDA). To avoid confusions each subject was informed that they have to evaluate perceptual similarity. For example, it was bad if an image has less detail than the real scene. It is also bad if it has more detail. So, the closer to the real scene the better it should be. As result of all experiments we got a total of 2 scenes × 19 participants × 3 display types × 4 attributes × 21 pairs = 9576 values of observation.

Three different types of display devices were used for each scene. First, the tone mapped images were displayed on two 17 inch CRT Dell, then on two 17 inch TFT LCD Dell and at last on two HTC P3300 with 2.8 inch display size (Figure 6).

For each experiment, the devices were carefully s-RGB calibrated. Great care with devices calibration was taken. Display devices with same mark and model do not assure similar visualization. That happened even with PDA where some device rejections were made. The PDA device can only enable brightness calibration making very hard to find two similar devices. In the traditional monitors, the gamma value was carefully calibrated to match the gamma used by algorithms because most TMOs require this issue for best performance. The images shown on the 17 inch CRT and LCD devices had resolutions of 1024 × 682, whereas the ones shown on the PDAs had a resolution of 240 × 320 (QVGA).

To achieve these resolutions some resizing had to be done since the photos captured by the digital camera had resolution of 3456 × 2304. The resize could be done in one of three possible states of the mapped image generation process (Figure 7).

After performing some simple visual tests, we observed that all solutions lead to different results. However, comparing the three resulting images, the worst seems to be the one achieved at point 2. Because usually the HDR images will not be specially generated for PDA, we assume that the resize should occur at points 2 or 3. In our experiments, we...
decided to make the resize at point 3 because, as said before, it generally leads to visually better results. Anyway, the resize was performed the same way for all TMO, making this issue not relevant for the experiments purposes. Resizes were done using ACDSeePro image software from ACD Systems [Acddseepro].

There were thus two independent variables (the displays devices and the TMOs) and four dependent variables (the image attributes).

3.1.4. Procedure

A pilot test was performed with a few experienced participants to refine the experimental conditions and procedures.

Each subject made his/her choice based on a specific property being tested. All pairs of images were randomly shown and the participant had to choose the image which is perceivably similar to the real scene in front of him/her based on the specified property. The subject identified his/her choice by clicking on that image first, then clicking on the other one. Special software was developed to manage when and where each image was to be shown and to store the results.

Each participant took about 20–30 min to complete the 21 pairs × 4 attributes ≈ 84 comparisons.

4. Results and Discussion

As already mentioned, although the use of paired comparison technique implies a large number of time-consuming trials, such an approach makes it easier to evaluate and compare the performance of each subject.

The advantage of paired comparison is not only simplicity, because subjects only have to make straightforward judgments, it also allows an evaluation of the transitivity, that is, the within-subject consistency of the data, as well as the between-subject consistency. Part of our analysis procedure is very similar to that used by Ledda et al. [Ledda’05] and is described later.

For statistical analysis purposes, for each scene, 12 preference matrices were created (three device types and four image attributes). For example, Table 2 shows the preference matrix for the colour attribute evaluated in PDAs where, for each row, one can observe how many times a specific TMO was preferred against the others, for example Drago’s TMO (Dr) was preferred 12 times when compared with Fattal’s TMO (Fa). Note that we decided to use simplified designation for each TMO where Fattal’s TMO is Fa, Durand’s TMO is Du, Ashikhmin’s TMO is As, Pattanaik’s TMO is Pu, Reinhard2002’s TMO is R2 and Reinhard2004’s TMO is R4.

If a participant preferred TMO A rather B and B rather C then logically he/she would prefer A rather C. If this happens for all judgments the full consistency/transitivity will be achieved. The coefficient of consistency, ζ [Kendall’40], allows an evaluation of the transitivity for each participant. The coefficient of consistency is defined as

\[ ζ = 1 - \frac{24}{t(2t^2 - 1)} \sum_{i=0}^{t-1} \left( \frac{p_{ij}}{t(t-1)} \right)^2 \]

where \( t \) is the number of TMOs to compare and \( p_{ij} \) is the number of preferences scored by TMOi \( (i = 1, 2, ..., t) \). For example, Table 2 shows that Fattal’s TMO scored 55 in the first scene for the colour attribute in PDA.

Values of ζ close to 1 indicate that there were good consistency within-subject. For example, the scene 1 participants’ coefficient of consistency for the contrast in the CRT was 0.8276.

Another important measure is the between-subject consistency or agreement. This can be done using the Kendall Coefficient of Agreement [Kendall’40] defined as

\[ u = \frac{2 \sum_{i>j} \left( \frac{p_{ij}}{t(t-1)} \right)}{s(s-1)t(t-1)} - 1, \]

where \( p_{ij} \) is the number of times TMOi is preferred to TMOj and \( s \) is the number of subjects.

Because the number of subjects is odd (19), the Kendall’s u ranges from \(-1/19\) (when agreement is minimum) and 1 (when agreement is maximum). To measure the significance of the Coefficient of Agreement, we may test the null hypothesis \( H_0 \) against the alternative hypothesis \( H_1 \), where

\[ H_0: \text{no agreement between subjects} \]

\[ H_1: \text{degree of agreement greater than if the evaluation of the comparison had been done randomly} \]

We may use the chi-squared test statistics \( (χ^2) \) to determine the significance of \( u \) [Siegel’698],

\[ χ^2 = \frac{t(t-1)(1 + u(s-1))}{2} \]
Table 3: Tables showing the Kendall coefficient of agreement ($\tau$), coefficient of consistency ($\zeta$) and significance $p$ according to chi-square tests statistics ($\chi^2$) for each device type and attribute. The corresponding TMO rankings are also indicated.

Table 4: Overall score of each TMO. In each column the best value(s) is(are)bolded and painted with a different colour.

$\chi^2$ is asymptotically distributed with $(t(t-1))/2$ degrees of freedom. The statistically significance of the obtained values were easily determined from tables of probability.

Table 3 shows that the agreement among observers is statistically significant and that each observer was very consistent in his/her choices. This means that we have a valid and good base of work.

Yet from Table 3 it is possible to observe all the partial TMO rankings for each type of display device and comparison criterion.

Table 4 shows the overall score for each TMO highlighting the differences. The total absolute scores, the sum of the four partial attribute scores, show that the results for the CRTs and LCDs are similar, but both are very different than PDAs.
Table 5: Final TMO rankings where the arrows show the main changes in the PDA ranking when compared with CRT and LCD rankings.

<table>
<thead>
<tr>
<th>Scene 1</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
<th>6th</th>
<th>7th</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRT</td>
<td>R2</td>
<td>Du</td>
<td>Dr</td>
<td>R4</td>
<td>R4</td>
<td>Pa</td>
<td>Fa</td>
</tr>
<tr>
<td>LCD</td>
<td>R2</td>
<td>Du</td>
<td>Dr</td>
<td>Fa</td>
<td>As</td>
<td>Pa</td>
<td>Pa</td>
</tr>
<tr>
<td>PDA</td>
<td>Dr</td>
<td>Dr</td>
<td>Pa</td>
<td>Pa</td>
<td>Pa</td>
<td>Pa</td>
<td>Pa</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scene 2</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
<th>6th</th>
<th>7th</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRT</td>
<td>R2</td>
<td>R2</td>
<td>Du</td>
<td>Du</td>
<td>As</td>
<td>Pa</td>
<td>Fa</td>
</tr>
<tr>
<td>LCD</td>
<td>R2</td>
<td>R2</td>
<td>Dr</td>
<td>Dr</td>
<td>Pa</td>
<td>As</td>
<td>As</td>
</tr>
<tr>
<td>PDA</td>
<td>R4</td>
<td>R4</td>
<td>R4</td>
<td>Pa</td>
<td>Fa</td>
<td>Fa</td>
<td>Fa</td>
</tr>
</tbody>
</table>

Table 6: Pearson chi-square tests [Spss].

<table>
<thead>
<tr>
<th>Chi-square tests - Scene 1</th>
<th>value</th>
<th>degrees freedom</th>
<th>Asymp. Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRT vs. LCD vs. PDA</td>
<td>153.7</td>
<td>6</td>
<td>0.000</td>
</tr>
<tr>
<td>CRT vs. LCD</td>
<td>0.306</td>
<td>6</td>
<td>0.999</td>
</tr>
<tr>
<td>CRT vs. PDA</td>
<td>110.76</td>
<td>12</td>
<td>0.000</td>
</tr>
<tr>
<td>LCD vs. PDA</td>
<td>107.55</td>
<td>12</td>
<td>0.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chi-square tests - Scene 2</th>
<th>value</th>
<th>degrees freedom</th>
<th>Asymp. Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRT vs. LCD vs. PDA</td>
<td>168.618</td>
<td>12</td>
<td>0.000</td>
</tr>
<tr>
<td>CRT vs. LCD</td>
<td>3.741</td>
<td>6</td>
<td>0.712</td>
</tr>
<tr>
<td>CRT vs. PDA</td>
<td>108.205</td>
<td>12</td>
<td>0.000</td>
</tr>
<tr>
<td>LCD vs. PDA</td>
<td>129.601</td>
<td>12</td>
<td>0.000</td>
</tr>
</tbody>
</table>

and this happens in both scenes. This difference can also be clearly seen in the final ranking for each TMO, Table 5. Curiously, besides we use the default TMO’s parameters, the resulting TMO ranking for LCD and CRT are quite similar do the previous related work. In those studies R2 was usually one of the betters and Dr also had a good performance; Fa was considered not very natural.

To measure the significance, the identified difference, Pearson’s chi-square tests were performed. As can be seen in the first row correspondent to each scene in Table 6, there are significant differences between CRT, LCD and PDA rankings. That is, we may reject the null hypothesis $H_0$ defined in Section 3.1.1. The other three possible cases were then analysed: CRT versus LCD, CRT versus PDA and LCD versus PDA. The corresponding null hypotheses for each one are:

- $H_0$: $\text{ranking}_{\text{CRT}} = \text{ranking}_{\text{LCD}}$ (CRT vs. LCD)
- $H_0$: $\text{ranking}_{\text{CRT}} = \text{ranking}_{\text{PDA}}$ (CRT vs. PDA)
- $H_0$: $\text{ranking}_{\text{LCD}} = \text{ranking}_{\text{PDA}}$ (LCD vs. PDA)

The results are shown in Table 6.

These tests show that we cannot reject the null hypothesis for CRT versus LCD. This means that the rankings obtained for the CRT and LCD are not significantly different. In other hand, we can reject the null hypothesis for CRT versus PDA and LCD versus PDA because $p < 0.005$, resulting that CRT and LCD rankings are both significantly different from the PDA ranking.

These results show that people see the same tone mapped image differently in PDA than they see in CRT or LCD.

The experiments performed not only show that, in fact, there is some differences, but also indicate guidelines for the development or adjusting of TMOs for small screen devices such as PDAs. Table 5 shows that the TMOs from Fattal et al. (Fa) have raised in the ranking compared with the CRT/LCD counterparts. In minor scale, the same happened with the TMO from Drago et al. (Dr).

With this observation in mind we can positively ask what special characteristics these TMOs have? According to the results, for traditional displays Fa’s TMO tends to exaggerate detail reproduction but this characteristic seems to be important to maintain perceptual similarity in PDAs. Dr’s TMO, on the other hand, is characterized by having visually more saturated colours and the mapped image has a higher overall brightness. These characteristics are noted as disadvantage when displayed in a CRT or LCD, but seems to be important factors in PDAs. This fact is indeed in accordance with deRidder’s work, where he mentioned that higher colour saturation is needed to compensate the reduced brightness of a display to achieve a more natural image perception [Deridder’96].

Although R2’s TMO came out on top of the CRT and LCD rankings, it loses strength in the PDA ranking (Table 4). A possible solution to get a better TMO for PDA could be to develop a hybrid approach using Dr’s TMO as base and improve the details with the Fa’s TMO.

All images used in the experiments and also the collected data can be accessed at Carlos Urbano’s personal web site http://www.estg.ipleiria.pt/~carlos.urbano/.

5. Conclusions

The results of the experiments show that the limited size, resolution and colour depth of SSDs require a different approach when tone mapping HDR images for display on such a device.

The results also show that, in fact, some image characteristics need to be emphasized by the TMO to obtain perceptually better images for SSDs. These characteristics are stronger detail reproduction, more saturated colours and overall brighter image appearance.
The knowledge gained from these results will be used to develop a new tone mapping operator for small screen devices, providing the best scene preservation of all attributes including detail, contrast and colour. It will be very interesting to compare and discuss our new TMO with the display adaptive developed by Mantiuk et al. [Mantiuk’08].

Acknowledgement
The authors thank anonymous reviewers for suggestions, which objectively improved this paper.

References


© 2010 The Authors

Computer Graphics Forum © 2010 The Eurographics Association and Blackwell Publishing Ltd.


