

# User Rating Based Comparison of Tone Mapping Operators for HDR Images

## Multimedia Communications AS 2015

Taivo Pungas  
15-928-336

January 29, 2016

## 1 Introduction

### 1.1 Motivation

Most classified ads – ranging from real estate to used guitars – contain images. In most cases these images are taken by non-experts, and even when a professional photographer is used, photos might not reflect the advertised object accurately; one possible cause is the low dynamic range (LDR) of the images.

Assuming the look and feel of images play an important part in buying decisions, we naturally get to the questions: do scenes or objects depicted in high dynamic range (HDR) images look better compared to LDR images? Which tone mapping operator (TMO) gives the best result?

### 1.2 Approach

I will:

1. Create HDR images by combining multiple images taken at different exposures and tone-map the images using a Matlab toolbox<sup>1</sup> from [1].
2. In a survey, ask users how pleasant they find the images.
3. Compare the results of LDR images and different TMOs for HDR images.

### 1.3 Expected Outcome

I will rank different TMOs and compare them with LDR images. I will compare the results for different lighting conditions.

---

<sup>1</sup>[https://github.com/banterle/HDR\\_Toolbox](https://github.com/banterle/HDR_Toolbox)

## 2 Theoretical Background

### 2.1 HDR images

The luminance of a region of an image describes the intensity of light originating from that region and captured in the camera sensor.

A scene can be said to have high dynamic range if the amount of light is very high in some regions and very low in others. Standing in a dark room during the day and looking towards a window is an example of a high dynamic range scene: a lot of light is coming from the window but very little from the surrounding walls.

The dynamic range of a camera [monitor] is a measure of the range of luminance values that can be captured [displayed] using the device. The dynamic range of typical commercial cameras is low, which means it is hard to capture details in both light and dark areas of the image. Furthermore, even if an HDR image can be captured (e.g. by merging images taken at different exposure values), most monitors are unable to display a high dynamic range of values (i.e. HDR images can not be reproduced as-is on these monitors).

### 2.2 Tone mapping operators

We can solve to the problem of reducing a large range of luminance values (from the HDR image) to a small range (the capability of monitors) by mapping the large range to the small range using a *tone mapping operator* (TMO). "This reduction of the range attempts to keep some characteristics of the original content such as local and global contrast, details, etc. Furthermore, the perception of the tone mapped image should match the perception of the real-world scene." [1]

Having multiple criteria implies multiple optimisation goals and so we should expect that there is no single optimal solution; rather, choosing a TMO involves some amount of artistic liberty. The goal of this work is to find the TMO that is considered aesthetically most pleasing by users.

## 3 Implementation

### 3.1 Pipeline

The end-to-end pipeline of selecting images for surveys was as follows:

- Capture 40 scenes. For each scene, three images were taken: one with automatic settings chosen by the camera (Olympus E-400) and two others with exposure values (EV) -1 and +1 using automatic exposure bracketing (AEB). The images were taken in Tallinn, Estonia, during dusk of December 19, 2015. All images were taken using a camera stand to avoid shift in the scenes.
- Choose 4 scenes out of the 40 so that the scenes would be dissimilar in composition, without ghosting, and without motion blur (due to camera movement).

- Resize images from 3648x2736 to 1067x800 to speed up HDR image generation and tone-mapping.
- Create HDR image for each scene using the three AEB images.
- Run all 27 available TMOs in the toolbox [1] with their default parameters (which are set to values recommended by the TMOs' original authors).
- Apply gamma correction to tone-mapped images where necessary.
- Pick top 4 TMOs to compare in the user survey such that the TMOs produce images that are visually different enough and don't have gross aesthetic failures (e.g. extremely low brightness). The chosen TMOs were Drago [2], Kuang [3], Mertens [4] and WardHistAdj [5].

This process yielded 16 images: one for each TMO-scene pair.

## 3.2 Resources

All of the resources described below are available at the GitHub repository of this project<sup>2</sup>.

**Code** The implementation can be found in the folder `code`:

- `pipeline.m` runs the pipeline from generating HDR images to mapping them to LDR images using TMOs.
- `build_hdrs.m` creates HDR images from given sets of LDR images taken at different exposures.
- `build_comparison.m` creates a single tiled image of all tone images compared in this work for easy visual comparison.

**Images** All images can be found in the folder `images`.

- `combined.jpg` is a tiled image of all tone-mapped images compared in this work.
- `combined_all.jpg` is a tiled image of all tone-mapped images generated, including ones not shown to users in the survey.
- `best_worst.jpg` is a comparison of the best and worst tone-mapped images according to the user ratings.
- Folders `kalamaja2`, `niguliste`, `ptln1`, and `toompea4` each contain the files for one scene:
  - `PC*.JPG`, the three images taken using AEB.

---

<sup>2</sup><https://github.com/taivop/eth-mm-c-project>

- `hdr_image.pfm`, the HDR image generated from this scene.
- `tmo__Original.JPG`, the image taken using automatic settings chosen by the camera.
- `tmo_<name>.jpg`, the tone-mapped images, where `<name>` is the name of the TMO used.

**Data analysis** Survey result data and the visualisation script can be found in the folder `analysis`.

- `results.csv` contains results of the user survey. The columns are coded as two characters `XY` where `X` is the first letter of the scene (e.g. `T` stands for Toompea) and `Y` is the first letter of the TMO (e.g. `M` stands for Mertens).
- `plotting.R` contains the R scripts used for analysing and visualising data.

## 4 Experiments and Results

### 4.1 Survey method

In the user survey, each user was presented with five images for each scene: four tone-mapped images from the HDR-TMO pipeline, plus the original image captured with automatic camera settings. The images were presented in a fixed order that was the same for all respondents. This yielded a total of 20 images for each user to rate; I chose such a low number of TMOs and scenes since I expected a sample of less than 50 users. Figure 5 shows all the images included in the survey.

Each image was shown on a separate page and the user was asked to evaluate each image for individual aesthetic merit, not compare it with previous images. The user was asked "How much do you like the following image?" and the answers were given on a Likert scale ranging 1 to 7 with integer steps. A rating of 1 corresponded to "Do not like at all", 4 to "Neutral" and 7 to "Like very much".

The survey was implemented using Google Forms and conducted in Estonian as the primary audience were Estonians. I acquired 126 participants by individually inviting (roughly 20 respondents) and sharing the questionnaire on my Facebook page (roughly 100 respondents). Each of the 126 participants gave a total of 20 ratings yielding a total of 2520 ratings.

### 4.2 Results

From Figure 1, we can see that there is no TMO that always performs better than others, however Mertens and Drago perform worse in all scenes than Kuang and WardHistAdj. The same is visible from the rating distributions on Figure 2 and on Figure 4: Drago and Mertens are rated consistently lower. In interpreting the data, it is important to note that Toompea4 was a much darker scene than others – this shows in the data as clearly the performance of all TMOs is low on that scene.

When we compare the TMOs' ratings with those of the original images, we see a very uneven performance: The original images rank first on two scenes (significantly so in the scene Niguliste), fourth on one scene and second/third on one scene.

For ease of reference, the best and worst images for each scene are shown in Figure 3.

## 5 Discussion

Given that Kuang is better than all other TMOs in all images but one and delivers consistent performance across all images, I would rank it above WardHistAdj. Mertens also beat Drago in all images but one and was more consistent in aesthetic quality, so from the results of this survey, I would rank the TMOs as:

$$\text{Kuang} > \text{WardHistAdj} > \text{Mertens} > \text{Drago}$$

Given that the aesthetic performance of the original images is highly uneven but beats all TMOs in some cases, I recommend considering the original image as an option if the

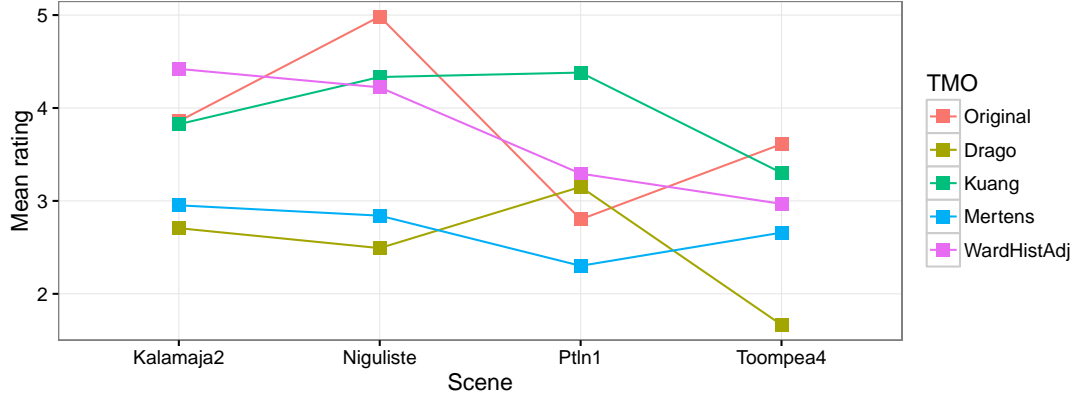


Figure 1: Mean user ratings by scene and TMO.

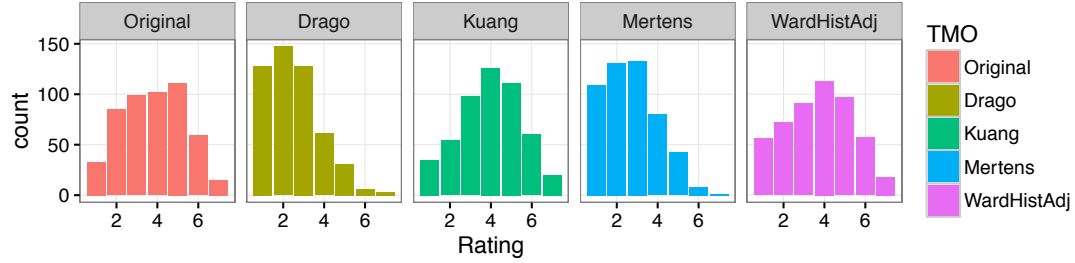


Figure 2: Distribution of ratings by TMO

goal is to select the most visually pleasing image. It is also interesting that the original image performs relatively better whenever the scene is dark (Niguliste and Toompea4) – this can be taken into account in an automated image aesthetics improvement algorithm like the one proposed in [6].

**Room for improvement** There are several features of this study that could be improved upon. First, increasing the user sample size would allow making stronger statements about more scenes and more TMOs, which would allow us to make more general statements about the relative aesthetic performance of TMOs.

Second, image presentation order in the user survey should be randomised to remove undesired effects caused by giving all users a baseline (the image that was first shown).

Third, users may be better at comparing pairs of images rather than rating images in isolation. One way to achieve this would be to use the Elo rating system (as a colleague pointed out to me in an informal chat).

Fourth, the TMO parameters should also be varied. In the current study, the performance of some TMOs may have been hindered by a suboptimal choice of parameters. This, however, would need a much larger sample size or automated image aesthetics analysis.



Figure 3: The best and worst tone-mapped images according to user ratings.

## 6 Summary and conclusion

In this study, I compared how different TMOs are rated by users. I proposed a user-rating-based ranking of TMOs and found that the original (i.e. LDR) image may significantly outperform tone-mapped HDR images, especially in low-light conditions where all TMOs perform worse than in better-lit scenes.



## References

- [1] Francesco Banterle, Alessandro Artusi, Kurt Debattista, and Alan Chalmers. *Advanced High Dynamic Range Imaging: Theory and Practice*. AK Peters (CRC Press), Natick, MA, USA, 2011.
- [2] Frédéric Drago, Karol Myszkowski, Thomas Annen, and Norishige Chiba. Adaptive logarithmic mapping for displaying high contrast scenes. In *Computer Graphics Forum*, volume 22, pages 419–426. Wiley Online Library, 2003.
- [3] Jiangtao Kuang, Garrett M Johnson, and Mark D Fairchild. icam06: A refined image appearance model for hdr image rendering. *Journal of Visual Communication and Image Representation*, 18(5):406–414, 2007.
- [4] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion. In *Computer Graphics and Applications, 2007. PG'07. 15th Pacific Conference on*, pages 382–390. IEEE, 2007.
- [5] Gregory Ward Larson, Holly Rushmeier, and Christine Piatko. A visibility matching tone reproduction operator for high dynamic range scenes. *Visualization and Computer Graphics, IEEE Transactions on*, 3(4):291–306, 1997.
- [6] Tunc Ozan Aydin, Aljoscha Smolic, and Markus Gross. Automated aesthetic analysis of photographic images. *Visualization and Computer Graphics, IEEE Transactions on*, 21(1):31–42, 2015.

## Appendix A: distribution of ratings by TMO and scene.

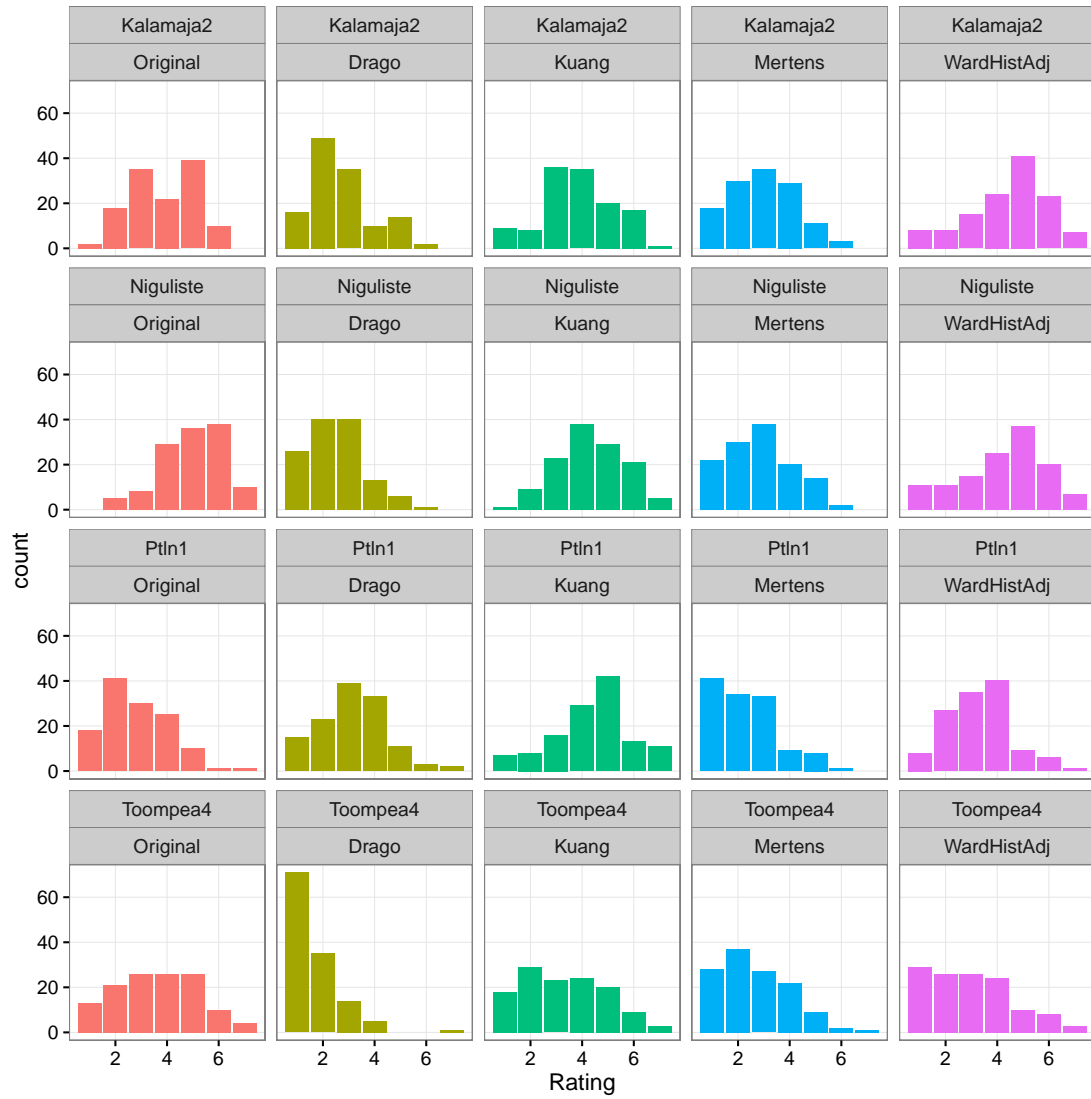


Figure 4: Distribution of ratings by TMO and scene.

## Appendix B: all images shown to users.



Figure 5: Tone-mapped images (top 4 rows) and the original image taken with automatic settings (bottom row).