Improved Sectional Tone Mapping for High Dynamic Range Imagery

Jaclyn Pytlarz  
M.S. Candidate, Computational and Mathematical Engineering  
Stanford University  
jpytlarz@stanford.edu

Andrew Misiolek  
M.S. Candidate, Electrical Engineering  
Stanford University  
misiolek@stanford.edu

Jack O’Reilly  
M.S. Candidate, Electrical Engineering  
Stanford University  
oreilly@stanford.edu

Abstract—High dynamic range (HDR) imaging has become common in consumer entertainment. As tools for capture, content creation, encoding, and transmission evolve to support HDR formats, the variety of display capabilities is increasing. Hence, there is value in determining solutions for remapping native HDR imagery for display on lower quality HDR and legacy standard dynamic range (SDR) displays. In this paper we expand upon a sectional tone mapping approach by Lenzen [9] attempting to improve temporal stability while maintaining picture quality. We give results from both an objective metric on flicker and a subjective user study on preference and flicker. Our results indicate success in reducing halos and improving temporal stability through edge aware filtering, without significant loss to image quality.

Index Terms—HDR, SDR, high dynamic range, tone mapping, edge detection, Drago operator, sectional tone mapping

I. INTRODUCTION

Most high dynamic range content today is graded on displays that have a luminance range between 0.005 – 1000 cd/m². However consumer display peak luminance levels can easily range between 400 – 3000 cd/m². Global tone mapping, wherein a single transfer curve is applied to the full HDR image, is a common procedure used today for this task of displaying high dynamic range content on these displays which support lower dynamic range. This global tone mapping operator may be static or dynamic (changing frame by frame) as described by Kiser, Reinhard, Tocci, and Tocci [8], but its nature as a single operator requires a compromise between local contrast and brightness. This and other drawbacks identified in global tone mapping have led to proposals for more sophisticated tone mapping methods. More recently, research has sparked in the realm of local tone mapping. Here, local tone mapping refers to small area contrast enhancements, similar to sharpening filters, which help recover lost detail during the global tone mapping process as described in [5]. These methods may lead to image artifacts such as halos and additionally can be computationally expensive, especially as we move towards 8K 120fps video content.

An alternative approach utilizes sectional tone mapping as proposed by Lenzen [9]. This method defines 16x9 regions which are tone mapped independently. The resulting image contrast and details appear much improved. However, we believe there is room for improvement. Lenzen [9] accounts for temporal stability by blending 25 percent with a global mapping operator. While this improves stability, it also diminishes the benefit of the sectional operator. Additionally, despite this filtering, the sectional tone mapping approach still suffers from temporal artifacts.

In this paper, we propose a mapping method that improves upon the sectional tone mapping approach by defining edge aware regions in place of the 16x9 blocks. This paper first provides background context to this approach, including a detailed overview of the main components of Lenzen’s work [9]. We then discuss the addition of edge aware processing to this method, with consideration of algorithm efficiency, seeking to maintain or improve upon the speed of the sectional tone mapping implementation. We will conclude by introducing qualitative and quantitative metrics for evaluating the value of our improvement to the sectional tone mapping approach.

II. HISTORY

The search for desirable tone mapping operators has existed since near the advent of photography. In the era before digital photography, solutions were principally oriented around film stock and print development system design to provide desired curves. As digital workflows began to see wider use, it became possible to consider a wider array of approaches.

One of the first approaches for digital tone mapping came in the form of a simple global tone mapping operator proposed by Stockham [10] (1972):

\[ L_d = \frac{\log(L_w + 1)}{\log(L_{\text{max}} + 1)} \]

The values \( L_w \), \( L_{\text{max}} \), and \( L_d \) correspond to world luminance, maximum scene luminance, and the luminance display factor. Note that the logarithm base doesn’t matter. This simple operator takes inspiration from experimental observations that the brightness \( B \) perceived by the human visual system can be modeled by a power function \( B = kL^\gamma \). Additional advantages of this approach include statelessness, temporal stability, and relatively low computational complexity. However, as in the case of many global operators, this approach can result in SDR output with regions suffering from overexposure or lost detail.
Drago et al. [2] (2003) sought to improve on this approach by proposing a new global tone mapping operator that incorporated further characteristics of the human visual system, including the eye’s ability to adapt to the overall changing luminance of a scene. Their work produced the following operator:

\[
L_d = \frac{L_{d_{\text{max}}} \cdot 0.01}{\log_{10}(L_{w_{\text{max}}} + 1)} \cdot \frac{\log(L_w + 1)}{\log(2 + \text{bias}_b(L_w/ L_{w_{\text{max}}})* 8}
\]

Here, \(\text{bias}_b(t)\) is a power function defined over the unit interval which produces values also within the unit interval. Curvature above or below the identity function may be chosen by the parameter \(b\), and it is defined thus:

\[
\text{bias}_b(t) = t^{\frac{\log(b)}{\log(0.5)}}
\]

\(b\) is an intuitive parameter chosen by empirical results to be 0.85, a value which Lenzen later used in his modified version of the Drago operator. \(L_w\) and \(L_{w_{\text{max}}}\) are defined as before except that they are divided by the quantity \(L_{wa}\), called the world adaptation luminance by Tumblin and Rushmeier [12]. They proposed computing this value in the following manner:

\[
\log_{10}(L_{wa}) = E[\log_{10}(L_w)] + 0.84
\]

\(E[\log_{10}(L_w)]\) is estimated by averaging the log luminance over the pixels in an image. The constant 0.84 corresponds to the 8.4\(dB\) offset commonly found to be the adaptation state above the diffuse white in a scene.

In a following section, we will discuss how Lenzen adapted Drago’s technique in order to propose a new spatially varying tone mapping technique.

### III. Tone Mapping Algorithms

In this section we will describe our design for a local tone mapping operator. We will begin by describing the sectional tone mapping approach utilizing a virtual aperture algorithm which was the motivation for our work. Then we will describe our implementation which uses a spatially blurred edge preserved mid-map as a guide to a global tone mapping operator.

#### A. The Virtual Aperture Algorithm

As part of his work that inspired our project, Lenzen [9] presented a method for combining a luminance transfer curve inspired by Drago’s [2] approach as well as histogram-based methods such as those initially proposed by Ward et. al [13]. The implementation of this algorithm is characterized by a lengthy array of algorithm parameters, which are listed in table I. A very significant portion of our project involved effectively reverse engineering portions of this algorithm in order to determine a parameterization which would produce desirable results. When contacted, Lenzen replied that a reference implementation was not publicly available.

The virtual aperture approach is split into a number of steps but the goal is to calculate a 9x16 grid of local max’s that can later be used to guide the Drago tone mapping operator. This is done by splitting the image into 9x16 blocks and addressing each block individually. The histogram of one such block is show in Fig 1.

Next we define a bounding box where the width of the box depends on the target display luminance and the height depends on the number of pixels within the region. A Gaussian is fit on top of the bounding box as seen in Fig 2. This Gaussian also has a number of associated tuning parameters to define the standard deviation and height. The \(\mu\) and \(\sigma\) from the Gaussian definition are the two terms that are filtered over time to attempt to improve temporal stability. Our implementation included a crude scene change detector using deviations in \(\mu\) and \(\sigma\).

Next, the histogram and the Gaussian are multiplied together to both accentuate and normalize the input signal. The cut-off threshold is then defined as an additional tuning parameter. Indicated in Fig 3 by the dashed red line, the location of this cut-off becomes the local maximum for a single block. This
is repeated for all of the blocks in the image.

1) Sectional Tone Mapping Implementation: After repeating this process of using the virtual aperture for each of the blocks, we now have a matrix of max luminance values. These need to be resized to the original image size. We chose to use bilinear interpolation to ensure no over/under shoots as these would cause the tone mapping operator to attempt to divide by zero.

For improved temporal stability, it was recommended to blend the virtual aperture max values (\( z = 0.25 \)) with max values found by processing the image as a whole (the global \( L_{uw_{\text{max}}} \)). The last step is to use the modified Drago [2] global tone mapping operator to map the image. In the Drago operator we still have our input \( L_w \), but now instead of our \( L_{uw_{\text{max}}} \) being a single number, it is a different value per pixel in the image. It will be lower where the original image is lower luminance and higher where the original pixels are higher luminance.

Fig. 3: Choosing \( Y_{\text{max}} \)

B. Edge-Aware Luminance Mapping

Our motivation for developing a new local mapping method was our observation that, while the sectional tone mapping approach offered pleasing results for still images, the decisions to segment in a 9x16 grid resulted in temporal flickering despite attempts to minimize it. Additionally the 9x16 regions seemed unconnected to any characteristic of the human visual system. We propose a method that computes an edge-aware transformation of the input luminances. This transform blurs the image heavily while attempting to maintain sharp edges at high gradients. This should improve temporal stability while maintaining perceptual quality. In the results section we show that it allows for greater detail at areas of interest within the image while also improving temporal stability compared to the sectional tone mapping approach.

Our algorithm engineering process included multiple possible filtering approaches. Our experimentation included many filter types including the Interpolated Convolution (IC), Normalized Convolution (NC), Bilateral Filter, Local Laplacian Filters, and MATLAB’s "imguidedfilter" function based on the Guided filter [7]. The Bilateral Filter [3] and Local Laplacian filters [11] are referenced in multiple tone-mapping situations.

Fig. 4: Guided filtered image from bistro_01 [6]
However, based on our initial temporal flicker analysis, as described in our evaluation methods, some of the most promising filters were the more recent IC or NC Filters [4]. Flicker analysis is shown in Fig 5. These filters produced sharp edges, but would require approximately 10 seconds to process per frame. Due to the computational complexity, we moved onto other filter types. The most promising filter that is both fast to compute and produced decently blurred edge-preserved results is the guided filter (imguidedfilter.m in Matlab). The result of this filter is shown in Fig 4. We implemented the guided filter with a 20x20 filter size and a threshold of 0.5.

We apply the guided filter to the luminance of the source image, then use the result as the guide in the Drago tone mapper in place of the $L_{\text{max}}$ that was used in the sectional approach. Similar to the world adapting luminance method, we compute $L_{\text{max}}$ by:

$$L_{\text{max}} = 10^{\left(L_{\text{umG}} + 0.84\right)}$$

As was done in the sectional method, we also blend (z = 0.25) with the global $L_{\text{max}}$.

IV. DATASET

Froelich et al. [6] created a large dataset of scenes captured in the form of series of wide gamut HDR images. Like Lenzen [9], we chose to evaluate our algorithm on the clips color graded in ITU-R BT.2100 [1] (PQ) color space with input luminance ranging from 0.005 – 4000 cd/m^2. These clips are meant to include a wide variety of scene content, with particular emphasis on changing brightness, specular highlights, high contrast skin tones, and more.

In particular, we chose the scenes referred to as bistro_01, showgirl_01, and fishing_longshot to evaluate our algorithm. These sequences were chosen specifically to evaluate temporal stability, detail resolution, and perceptual color preservation in the face of camera motion (fishing_longshot), specular highlights (bistro_01), and human skin tones (showgirl_01).

V. EVALUATION METHODS

From early in the project it was of central concern to determine methods for evaluating the success of our proposed tone mapping algorithm. To evaluate the performance of our tone mapping algorithm, we compared it with the following algorithms. We have done this both objectively and subjectively.

- Global Tone Mapping - Stockham
- Global Tone Mapping with Virtual Aperture - Lenzen
- Sectional Tone Mapping - Lenzen

A. Flicker Probability

The approach described by Lenzen [9] revealed an opportunity to improve the temporal stability of the tone-mapping operation. To accomplish this we devised an objective metric for basic flicker analysis between a given output and corresponding input image sequence. This is done by comparing the ratio between the source changes frame to frame against the test image changes from frame to frame. When the test images have a ratio that varies more than the source, we know we are introducing some amount of flicker. We blur the result before taking the mean to eliminate small pixel differences.

$$\text{ImDiff}_S = |\text{SourceA} - \text{SourceB}|$$

$$\text{ImDiff}_T = |\text{TestA} - \text{TestB}|$$

$$\text{ratio} = \frac{\text{ImDiff}_S + 0.001}{\text{ImDiff}_T + 0.001}$$

$$\text{ratioE}(i) = E[\text{blur}(\text{ratio})]$$

This objective metric is motivated by the idea that the per-pixel ratio of luminances between successive images in the output should roughly match the corresponding ratios for the input image sequence. In a “perfect” case of a global tone mapping operator, the relationship between input and output luminance values should be smooth.

B. Subjective Methods

Our primary metric for evaluating algorithm performance was a subjective experiment. Our study was structured in two parts. The first part, which was meant to evaluate overall suitability of the tone mapping operator with regard to desirable SDR output, involved showing the viewer a sequence of still images separated with black images in between. For one still from each of the three sequences, four mapping alternatives were shown (‘Stockham’, ‘global’, ‘sectional’, and ‘edge-aware’). Observers were asked to rank the images by order of preference while considering realism, contrast, brightness, and color consistency. The second part of our experiment was meant to evaluate temporal stability of each algorithm.
Observers were first introduced to the concept of ‘flicker’. Thereafter, each observer was shown a succession of the four mapping alternatives for each sequence, corresponding to the algorithm implementations, and was asked to rank the renditions in descending order of stability. Our study includes responses from 23 individuals with a near even distribution of gender, ages ranging from 23-70 and includes expert and non-expert observers.

VI. RESULTS

Our findings for the algorithm efficacy are tabulated and presented in this section.

A. Flicker Probability Calculations

We have analyzed our edge-aware results against the sectional approach as seen in Fig 7. Although we see both methods varying over time, the sectional method (seen in black) has much higher local variance which appears like noise in the signal. It’s this noise that presents as flicker in the resultant image. Our objective metric is corroborated by the subjective results given in the next section.

B. Subjective Results

The results of our subjective user study are given in Fig 8 and Fig 9. Starting with user preference (Fig 8) we notice that that Stockham global operator performed extremely poorly. Most of these images were more washed out and bright, so this was anticipated. We see that our edge-aware method was most preferred for the Bistro scene, but was less preferred for the other two scenes. Feedback from observers showed that the “flattness” of the edge-aware method was less desired for the other two scenes so there is room for improvement. It is
surprising that the global operator performed as well as it did as this is a much simpler approach. It’s possible that the black interstitial made the task too difficult.

In Fig 9 we explore results for temporal stability. We see, as expected, that the sectional approach performs very poorly. In contrast, our edge-aware method performs very well for temporal stability. In fact, it is nearly outperforming the global operator. This may have to do with the filter on the global operator being too quick.

C. Example Output Images

Examples of the rendered output for each algorithm are shown in figure 10 for reference.

VII. Conclusions

We have described our design for an improved local tone mapping operator and compared it against three other tone mapping methods for both subjective preference as well as temporal stability. We have shown that our edge-aware method has decent performance as far as image quality and preference. We have also shown that we have improved temporal stability dramatically over the sectional tone mapping approach, even outperforming global mapping in most cases. The main area for future improvement is increasing the local contrast of the edge mapping. This can likely be accomplished by finding an improved edge aware filter to allow for more blurring.

ACKNOWLEDGMENT

We would like to thank the Professor Girod and the EE368 course staff for continued support throughout the quarter.

REFERENCES


APPENDIX: WHO DID WHAT

Our project was broken down roughly in the following manner by effort:

- **Each team member** proposed specific a project idea
- **Jaclyn’s** project idea was determined to be the most interesting, feasible, and fully-formed
- **Andrew** performed primary investigation of edge-aware filtering techniques for our proposed algorithm enhancement
- **Jack** provided the initial implementation of Lenzen’s virtual aperture algorithm
- **Jaclyn** designed and implemented the algorithm for computing an objective flicker probability metric
- **Andrew** performed temporal stability analysis on edge-aware filters
- **Jaclyn** found tuning parameters for and fixed implementation issues with the virtual aperture algorithm
- **Jack** added the simple Stockham mapping operator as a baseline for comparison
- **All team members** solicited in-person feedback for the subjective algorithm performance study
- **All team members** contributed to the final project report