Perceptual Evaluation of Tone Reproduction Operators using the Cornsweet-Craik-O'Brien Illusion

AHMET OĞUZ AKYÜZ
University of Central Florida
Max Planck Institute for Biological Cybernetics

and
ERIK REINHARD
University of Bristol
University of Central Florida

High dynamic range images cannot be directly displayed on conventional display devices, but have to be tone-mapped first. For this purpose, a large set of tone reproduction operators is currently available. However, it is unclear which operator is most suitable for any given task. In addition, different tasks may place different requirements upon each operator. In this paper we evaluate several tone reproduction operators using a paradigm that does not require the construction of a real high dynamic range scene, nor does it require the availability of a high dynamic range display device. The user study involves a task that relates to the evaluation of contrast, which is an important attribute that needs to be preserved under tone reproduction.

Categories and Subject Descriptors: I.3 [**Computer Graphics**]: Miscellaneous; J.4 [**Social and Behavioral Sciences**]: Psychology

General Terms: Tone Reproduction, Experimentation, High dynamic range imaging
Additional Key Words and Phrases: Tone mapping algorithms, validation studies, dynamic range compression

1. INTRODUCTION

With high dynamic range (HDR) imaging entering the main-stream, there is an increasing need for reducing the dynamic range of HDR images for display on conventional display devices. A large set of tone reproduction operators is currently available to accomplish this (for a review see [Reinhard et al. 2005]). The ultimate goal of tone reproduction operators is to reduce the dynamic range while preserving the visual experience an HDR image offers. Given the limited dynamic range of most display devices, this goal is usually not achievable and therefore second-best solutions are sought. These include preserving one or more key attributes of an image such as its brightness, contrast, or visibility.

Although the main purpose of tone-mapping is usually to produce visually pleasing representations, this purpose may very well depend on the application at hand. For some application areas, such as medical imaging and visualization, visibility preserving tone-mapping could be preferred as it may ease diagnosis and help identification. For entertainment or special effects, contrast and brightness preservation could be given more emphasis in order to increase the visual appeal of the reproduced image. In color reproduction, preserving brightness or color appearance could be desired to convey the appearance of the

original scene as accurately as possible.

It is therefore essential to know which operators preserve which of these key attributes. Such knowledge may lead us to make educated choices when choosing an operator to be used in a given application.

We may evaluate tone-mapping operators based on people's preference, the criteria being realism or visual appeal. Although this may reveal people's preference in general, it does not help discover the underlying reasons for these preferences. Another approach would be to rank each operator for each desired attribute, such as contrast, brightness, visibility, visual appearance and color appearance. A collection of such studies may help explain people's preferences derived from the former type of validation studies. In our work, we adopt the latter approach and evaluate tone reproduction operators based on preservation of *contrast*. In that, the human eye is very sensitive to differences rather than absolute values, contrast is an important attribute that needs to be preserved under tone-mapping.

To be able to isolate and evaluate a single attribute such as contrast we use a relatively simple stimulus. Our stimulus is derived from the well-known Cornsweet-Craik-O'Brien visual illusion [O'Brien 1958; Craik 1966; Cornsweet 1970]. The advantage of using this illusion is two fold. First, it yields a very simple brightness profile which could be evaluated with a simple experimental design. Second, an image containing this illusion can be created at any dynamic range; both high and low. Therefore we can use the low dynamic range version as a reference, and compare the tone-mapped images against it. As such, we can compare tone mapping operators against a reference without using an HDR display device.

Specifically, the input to our experiment is a single high dynamic range image, where each scanline contains the Cornsweet-Craik-O'Brien illusion. This image is tone-mapped with several commonly used operators and scanlines of interest are chosen for psychophysical evaluation. In the psychophysical evaluation, we determine the perceived contrasts in these scanlines with a 2-AFC procedure and compare them against a reference. As a result, we categorize operators in terms of how well they preserve the Cornsweet illusion.

In the following section, we review previous validation studies of tone-mapping operators. We motivate our design choices in Section 3. Background information pertaining to the tone reproduction operators evaluated in our experiment is given in Section 4. The experimental design is detailed in Section 5, and results are presented in Section 6 followed by their discussion in Section 7. The relation of our results to the previous validation studies is given in Section 8 and conclusions are drawn in Section 9.

2. PREVIOUS WORK

Validation studies of tone-mapping operators may be divided into two categories. In the first category, images tone-mapped by various operators are compared against each other, the criterion being visual appeal or realism. Since the original HDR images are not used in the comparison, these studies only reveal the relative rankings of tone-mapping algorithms. In the second category, tone-mapped images are compared against either their HDR originals displayed on an HDR monitor or real world scenes. Although the latter type of studies may arguably yield more realistic rankings, they also have important limitations. For instance, studies involving real world scenes are limited to indoor scenes for reproducibility purposes. On the other hand, current HDR displays are still rare and costly at the same time notoriously difficult to calibrate. Furthermore, the peak luminance of current HDR

displays are about $3000\,cd/m^2$, which is still much lower than luminances found in many real world scenes [Seetzen et al. 2004].

Drago et al. [2002] conducted a psychophysical experiment in order to evaluate tone-mapping operators with regard to similarity and preference. In their experiment, participants were shown tone-mapped images in pairs and were asked to rate them in terms of similarity to each other as well as naturalness and participants' preference in a global sense. The resulting ranking therefore is most meaningful if the general application area is entertainment, where the operator producing the most preferred images is arguably the most desirable one.

Another pairwise comparison-based validation study was performed by Kuang et al. [2004] (also summarized by Fairchild et al.[2004]). In this study, participants were asked to indicate their preference on both color and grayscale (luminance channel only) images produced by several different operators. This study shows that the ranking of operators for grayscale and color images are strongly correlated. Thus, the dominant factor determining participants' preference is the quality of the luminance remapping, rather than color fidelity.

One of the advantages of the pairwise comparison paradigm is that no real scene is required for the assessment. Any HDR image may be tone-mapped with any number of operators, and their comparative preference may be recorded in a relatively straightforward user study. The results may depend on the content of the images, an effect which may be counter-balanced by incorporating a large- and varied-enough set of images. At the same time, this presents a disadvantage, because comparison-based evaluations are not able to assess how perceptually close the tone-mapped image is to its original, unless the original is also used in the comparison.

The advent of HDR display devices made it possible to carry out comparison based experiments by also using the original HDR images in the comparison [Seetzen et al. 2004]. Ledda et al. conducted such a psychophysical experiment to determine which tone reproduction operators render good representations of real world scenes [Ledda et al. 2004; Ledda et al. 2005]. By using an HDR display device, the authors test the closeness of the tone mapped images to linearly scaled images on the HDR monitor. Images that are most closely related to a direct rendition on an HDR monitor could be termed most accurate — a distinctly desirable property. This work is therefore complementary to the above preference-based user studies.

Yoshida et al. [2005] performed a user study by comparing tone-mapped images with actual scenes. The participants were asked to rate naturalness, contrast, brightness, and detail reproduction in images tone-mapped with different operators with respect to the corresponding real world scenes. Although this could be the ultimate test to determine how accurate the actual scene is reproduced on a display device, the necessity to set-up a controlled scene would limit the imagery that may be evaluated.

In a more recent study, Ashikhmin and Goyal [2006] investigate whether purely low dynamic range comparison based studies are sufficient to determine operator performance. The authors compare results of their two experiments, where only one of them involves real world scenes. The study indicates that results may differ significantly when original real world scenes are used in the comparison.

An evaluation study with somewhat contradictory results to Ashikhmin and Goyal's study is performed by Čadík et al. [2006]. The authors use image attributes such as bright-

ness, contrast, detail reproduction, and color appearance to define an image quality metric, and evaluate tone-mapping operators by this metric. Here, one group of subjects rank tone-mapped images without seeing the original scene, while a second group performs the same task while they are allowed to see the actual scene. The results of these two experiments are highly-correlated, a finding contrary to Ashikhmin and Goyal's results.

Based on these previous validation studies, an overall picture has not yet emerged. Sometimes we find the results of a validation study in contradiction with a previous one. For instance, gradient domain tone-mapping [Fattal et al. 2002] which performed worst in Kuang et al. [2004] performs best in Ashikhmin and Goyal [2006]. This suggests that differences between images in the test set and the experimental methodologies used may be larger than the differences between tone reproduction operators. Given this scenario, the field of tone-mapping needs a more systematic approach to validating tone-mapping operators. One possibility is to individually study how well different image attributes are preserved under tone-mapping. In our study we pursue this possibility by focusing on a single attribute, namely contrast, which needs to be preserved under tone-mapping.

3. MOTIVATION

Tone reproduction operators prepare HDR images for display on low dynamic range display devices. Most operators aim to preserve one or more of the key attributes of HDR images, such as brightness, contrast, visibility or appearance, during dynamic range compression.

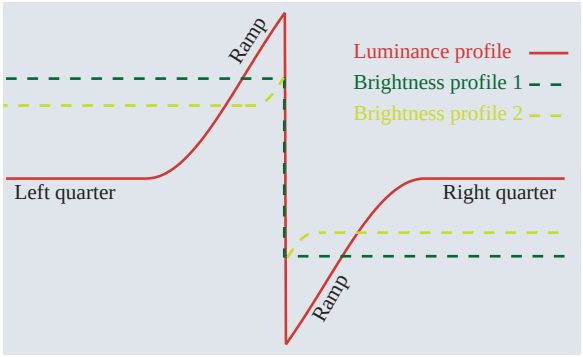
Since the human eye is very sensitive to differences rather than absolute values, it is mainly contrast that makes objects and colors in an image discernible from each other. As such, preservation of contrast is an important task for tone-mapping operators. In this study we therefore evaluate how well tone reproduction operators preserve contrast.

Our evaluation relies on the existence of a 2-D functional form, which may be reproduced at any dynamic range—both high and low—and may therefore be directly displayed in its low dynamic range form. The high dynamic range version may be tone-mapped with any number of tone reproduction operators, and then displayed on the same display device. The low dynamic range image then can be used to compare the tone-mapped images against it; i.e. it acts as a reference.

The functional form from which we derive our HDR image should exhibit the following three properties. First, it should allow parameterization to yield both high and low dynamic range variants. Second, it should be simple enough to allow measurement of a specific attribute — in our case contrast. Third, it should contain a large and continuous set of luminance values, so that the operators are evaluated for a large range of luminances rather than a few discrete luminance levels. An HDR image satisfying all three requirements can be created using the Cornsweet-Craik-O'Brien visual illusion [O'Brien 1958; Craik 1966; Cornsweet 1970].

This illusion consists of two identical gray patches separated by two non-linear ramps of opposite polarity as shown in Figure 1(a). A sharp gradient is present between the two ramps. This luminance profile induces the perception of a step function. Hence, its brightness profile is as shown in Figure 1(b). Note that although the luminance is identical on the left and right quarters, the entire patch looks like a bi-partite field due to a smooth variation followed by a sharp gradient in the middle. In fact, if the sharp gradient and the ramps are occluded by a small piece of paper, the left and right quarters are correctly

(a) The Cornsweet illusion



(b) The luminance profile and two possible brightness interpretations of the Cornsweet illusion

Fig. 1. *The top figure demonstrates the illusion in a rectangular patch, and the bottom figures shows its luminance and two possible brightness profiles.*

interpreted as being of the same luminance. For brevity, we will refer to this illusion as the Cornsweet illusion and its luminance profile as the Cornsweet profile in the remainder of this paper.

The Cornsweet illusion is characterized by its *ramp width* and *physical contrast* (Figure 2). The physical contrast C associated with the Cornsweet illusion is defined by [Dooley and Greenfield 1977; Kingdom and Moulden 1988]:

$$C = \frac{L_{max} - L_{min}}{L_{max} + L_{min}} \tag{1}$$

Both contrast and ramp width affect the strength of the illusion. As a general rule, it is found that low contrast ($< 20\%$) and wide ramps yield an apparent contrast equal to a real step having the same physical contrast [Dooley and Greenfield 1977]. As the contrast is increased (around $30\% - 40\%$), the illusion reduces relative to a real step. Also, the strength of the illusion remains constant if the ramp width and the total width is increased together such that their ratio remains the same [Wachtler and Wehrhahn 1997]. Figure 1(b) demonstrates two possible brightness interpretations of the Cornsweet profile. Depending on its functional form and parameters, either of the profiles could be perceived.

The Cornsweet ramp can be created with different functional forms such as linear [Wachtler and Wehrhahn 1997], parabolic [Dooley and Greenfield 1977], sinusoidal and exponential [Kingdom and Moulden 1988] ramps. The C1 continuity between the ramps and the constant sections ensures that the ramps cannot be differentiated from the constant-luminance sections. It may be argued that low level center-surround processes, such as those found in the retina [Kuffler 1953; Marr 1982; Dowling 1987; Palmer 1999], are the cause for this failure to detect the ramps. These processes are tuned to sharp discontinu-

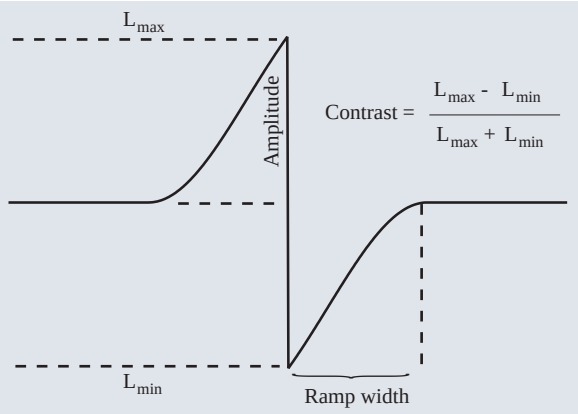


Fig. 2. *The parameters of the Cornsweet illusion.*

ities rather than slow gradients. The sharp discontinuity in the middle is therefore the only feature that is detected by low level visual processing. Center-surround processing of an actual step function would produce a similar signal. Hence, the Cornsweet illusion and a step function are perceived similarly.

It is believed that later visual processing receives edge information from the retina and then reconstructs the original signal by filling in [Walls 1954; Krauskopf 1963; Land and McCann 1971; Grossberg and Mingolla 1985; Pessoa et al. 1998; Elder 1999]. One possible reason for this encoding and decoding procedure is to overcome the communication bottleneck presented by the optic nerve. However, filling-in problems are ill-posed and therefore need to be guided by making assumptions on the scene being viewed.

One of the assumptions made by the filling-in procedure may be that the areas between edges are of more or less constant luminance. Although the luminance profile shown in Figure 1(b) violates this assumption, the human visual system only detects the sharp gradient in the middle, and therefore reconstructs a step function with the left half being brighter than the right half. That is, it propagates the sharp discontinuity to both left and right halves. Since there is a mismatch between the actual luminance profile and the perceived luminance profile (i.e. brightness profile), a visual illusion results.

We use the Cornsweet illusion in the following way. An HDR image is created where each scanline is a rendition of a low dynamic range Cornsweet profile. The luminance of each scanline is increased from top to bottom to give the image high dynamic range content[1]. This image is tone-mapped with several operators, and scanlines of particular interest are chosen for subsequent evaluation. We hypothesize that tone reproduction operators will alter the shape of the luminance profile and therefore alter the strength of the Cornsweet illusion. Thus, by measuring the strength of this illusion in a 2AFC design, we are able to determine how well contrast is preserved by each tone reproduction operator.

In addition, a low dynamic range image is created where every scanline follows the same Cornsweet profile as the corresponding scanlines in the HDR image. In other words,

[1]For clarity, we define the dynamic range of our input image as the quotient of its maximum and minimum luminances. Therefore, the dynamic range may take values between one and (theoretically) infinity. On the other hand, (horizontal) contrast is defined by Equation 1 and it only takes values between zero and one.

the parameters discussed above — ramp width and physical contrast — are identical for corresponding scanlines of each image, although the absolute luminances are different. As such, this image represents an ideal (but hypothetical) tone-mapping, which can serve as a reference for the task at hand.

In the evaluation, we ask subjects to compare the tone-mapped scanlines against real steps to assess the contrast after tone-mapping. Similarly, we determine the perceived contrasts induced by the scanlines coming from the reference image. Finally, we compare the contrasts induced by tone-mapped profiles against reference profiles, in order to estimate how well the Cornsweet illusion is preserved.

4. TONE REPRODUCTION OPERATORS

Our selection of operators to include in the evaluation is based on two criteria. First, the selection should cover a representative range of different approaches to tone reproduction, including both local and global operators. Second, the operators should either be parameter-free or provide a sensible set of default parameters to enable the evaluation to depend less on the choice of parameter settings.

The seven tone-mapping operators evaluated in this paper are the histogram adjustment technique [Ward et al. 1997], the revised Tumblin and Rushmeier operator [Tumblin et al. 1999], bilateral filtering [Durand and Dorsey 2002], gradient domain compression [Fattal et al. 2002], photographic tone reproduction (both global and local versions) [Reinhard et al. 2002], and iCAM [Fairchild and Johnson 2002; 2004]. Each of these operators is discussed in detail by Reinhard et al. [2005]. Therefore only a brief discussion is given below. In addition to these operators, we include a linearly scaled version of the input image as well as a reference image (see Section 5.2) which is created directly in low dynamic range.

The histogram adjustment operator uses a modified cumulative histogram to provide a global mapping between input and output luminances. The technique is similar to histogram equalization, but the cumulative histogram is modified such that local contrasts cannot be exaggerated. Thus, for each image a curve is derived based on the luminance histogram which is identically applied to all pixels. It may therefore be seen as a global operator.

Tumblin and Rushmeier's operator is a global operator designed to preserve brightness. An estimate of the brightness is computed from the HDR image using a power law which is derived from Stevens' human contrast sensitivity. Then a mapping is applied such that the tone-mapped image elicits the same brightness as the original HDR image under display conditions.

The global photographic tone reproduction operator is an example of sigmoidal compression. Therefore, it is representative of several other operators which are functionally similar [Schlick 1994; Pattanaik et al. 1998; Tumblin et al. 1999; Pattanaik et al. 2000; Reinhard and Devlin 2005]. The local photographic tone reproduction employs a local scale selection mechanism based on a Gaussian pyramid to compute a local adaptation value for each pixel, which is then applied in a sigmoidal compression function. Ashikhmin's operator is functionally similar to the local photographic operator [Ashikhmin 2002].

The iCAM model is a spatially variant descendant of a sequence of color appearance models [Fairchild 2005], which include CIECAM97 and CIECAM02. It is aimed at both compressing luminances as well as matching color appearance. It is the first operator which

Operator	Parameter	Value
Histogram adjustment technique	Mimic human vision	off
The revised Tumblin and Rushmeier operator	Display adaptation luminance	$25\,cd/m^2$
	Maximum display contrast	50
Fast bilateral filtering	Down sampling factor	1
	Base contrast	2
Gradient domain compression	Gradient magnitude threshold (α)	0.80
	Gradient magnitude exponent (β)	0.85
Photographic tone reproduction (global)	Key	0.55
	Luminance of white (L_{white})	$3340.80\,cd/m^2$
Photographic tone reproduction (local)	Key	0.55
	Luminance of white (L_{white})	$3340.80\,cd/m^2$
iCAM	Filter size	0.25 image size
	The degree of image white to D65 transformation (D_1)	0.1
	The degree of D65 to monitor white transformation (D_2)	0.0

Table I. *The input parameters of each tone-mapping operator evaluated in the experiment.*

can be viewed as both a tone reproduction operator as well as a color appearance model.

Bilateral tone reproduction is a local operator which splits the input into base and detail layers [Durand and Dorsey 2002]. The base layer is constructed by applying a bilateral filter [Tomasi and Manduchi 1998], which smoothes over small details while preserving sharp contrasts. It is therefore an edge-preserving smoothing operator. The detail layer is constructed by dividing the input image by the base layer. The base layer is then compressed in the log domain, before it is recombined with the detail layer to yield the final displayable image.

Gradient domain compression computes a gradient field of the luminance image and attenuates high gradients more than low gradients. The resulting gradient field is integrated to form a displayable image. The integration is accomplished by solving a Poisson equation using the Full Multigrid Method [Fattal et al. 2002].

Linear scaling is a naïve approach to tone-mapping, although it performs well in images with moderate dynamic range [Ward et al. 1997]. It is included in the experiment to observe where it stands among other operators in preserving the Cornsweet illusion.

The reference image has the same width and height as the original HDR image, and is directly created in low dynamic range (i.e. no tone-mapping is performed to create the reference image). In that, it has the same Cornsweet illusion parameters (i.e. ramp width and physical contrast) it may serve as a reference (see Section 5.2).

We used default parameters of the selected operators in our experiments. For the photographic operator, we applied its associated automatic parameter estimation algorithm [Reinhard 2003]. The full list of input parameters and their settings for each tone-mapping operator is given in Table I.

5. EXPERIMENTAL DESIGN

Our experimental design is based on an HDR input image which contains the Cornsweet illusion in every scanline. That is, each scanline itself is low dynamic range and follows a Cornsweet profile, but the image has a high dynamic range as a result of luminance variation from the top to the bottom of the image. This image is tone-mapped with the tested operators and scanlines of particular interest are chosen for the subsequent evaluation. In the evaluation, a stimulus is derived from each selected scanline and compared against

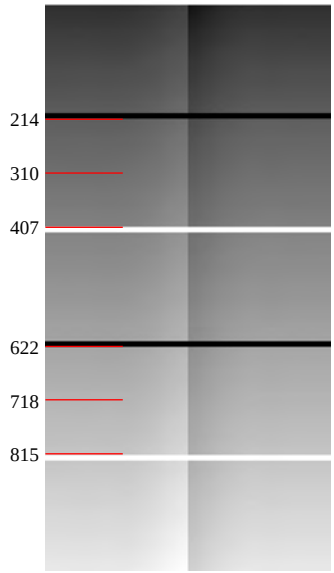


Fig. 3. *The high dynamic range input image (linearly scaled for display purposes). The red lines are the scanlines selected for the psychophysical evaluation; the red color is not part of the stimulus. The scanlines* 310 *and* 718 *are from smooth regions whereas the other scanlines are just next to black and white stripes (high-contrast regions). The resolution of the image is* 512×1024 *and each stripe is* 10 *pixels high.*

an actual step stimulus in a 2AFC design to quantify the strength of the illusion for that scanline. The details of each part are explained in the following sections.

5.1 The HDR input image

We have created a 512 by 1024 HDR image in which each scanline follows a low dynamic range Cornsweet profile, but neighboring scanlines have increasingly larger values, so that the whole image has a high dynamic range. This difference increases exponentially to create a perceptually linear stimulus from top to bottom as shown in Figure 3 (red lines are not part of the image and indicate the scanlines selected for psychophysical evaluation).

In addition, we have added dark and light horizontal bands to the image, to enable us to assess the behavior of tone reproduction operators near sharp contrasts.

The luminances of the black and white stripes are set to 10^{-1} and $10^{6}\, cd/m^2$, which are the minimum and the maximum values in the image[2]. The Cornsweet scanlines range from approximately 10^{3} to $10^{6}\, cd/m^2$. The purpose of this mapping is to create an HDR image which has a moderate dynamic range (i.e. $10^{3} : 1$) in most regions, but has a high dynamic range (i.e. $10^{7} : 1$) in total. In fact, HDR images of many scenes possess a similar distribution, in that they also have a moderate dynamic range in most parts of the image, but their dynamic range is high due to the concomitant presence of highlights and darks shadows in smaller regions.

We followed Kingdom and Moulden [1988], and based our Cornsweet profile on an

[2]These values are roughly equivalent to the reflection of white paper under moonlight to the reflection of white paper under sunlight.

Scanline	Minimum	Maximum	Average
214	89,435	187,389	138,412
310	136,044	253,752	194,898
407	187,793	327,435	257,614
622	321,566	517,904	419,735
718	390,990	616,753	503,871
815	468,070	726,502	597,286

Table II. *The minimum, maximum, and average luminance values of the selected scanlines. All values are in* cd/m^2.

exponential form. We set the physical contrast, C, to 20% and the ramp width to 25% of the entire profile. For these parameter settings, the Cornsweet illusion is found to be indistinguishable from a real step of equal physical contrast [Dooley and Greenfield 1977].

In addition, we have found that increasing the amplitude of the discontinuity in the middle in lock-step with intensity (from top to bottom of the image), produces a similar strength Cornsweet illusion for all scanlines. Thus, lighter scanlines will also have a larger amplitude. This induces a Cornsweet illusion that is of approximately equal strength for all luminance values.

As such, the functional form of our HDR input is defined on a unit square image as follows:

$$
\begin{aligned}
G(y) &= 0.70\frac{1-e^{y}}{1-e} \\
H(x) &= 0.176\frac{1-e^{-x^2}}{1-e^{-1}} \\
R(x,y) &= \begin{cases} H(4\,(x-0.25))\,(0.5+G(y)) & x \in [0.25, 0.50) \\ -H(4\,(0.75-x))\,(0.5+G(y)) & x \in [0.50, 0.75) \\ 0 & \text{otherwise} \end{cases} \\
L(x,y) &= 10^{-1} + 10^{6}\,(G(y)+R(x,y)+0.414)
\end{aligned}
\tag{2}
$$

The constants in this equation ensure that the stimulus remains within the desired dynamic range while still allowing headroom to let the ramp on each scanline drop below and rise above the base level. They were chosen for practical purposes and could possibly be defined differently without harming the experiment.

The red lines in this figure indicate which scanlines were selected for the subsequent psychophysical evaluation. These were scanlines $214, 310, 407, 622, 718,$ and 815, comprising both light and dark scanlines in smooth regions, dark scanlines near dark and light stripes, and light scanlines near dark and light stripes. The minimum, maximum, and average luminance values for these scanlines are listed in Table II.

5.2 The LDR reference image

The reference image is created by replacing the multiplier 10^6 with 256 in Equation 2, yielding:

$$L'(x,y) = 10^{-1} + 256\,(G(y)+R(x,y)+0.414) \tag{3}$$

Note that only the global scale is altered, and the two parameters (i.e. the physical contrast and the ramp width) that affect the strength of the Cornsweet illusion are kept intact.

5.3 Tone-mapped images

The HDR image is then tone-mapped with the results shown in Figure 4. From each of the tone-mapped images and as well as the reference image, we select 6 scanlines and evaluate the strength of the Cornsweet illusion. The individual profiles of the selected scanlines are shown in Figure 5.

For most scanlines the typical shape of the Cornsweet profile is preserved, although the amplitude of the discontinuity in the middle is changed. However, some of the scanlines tone-mapped by bilateral filtering and gradient domain compression have unusual profiles as shown in the figure. These correspond to the scanlines adjacent to the black stripes. We explain the reasons for these distortions in Section 7.

5.4 Stimuli

The stimuli are created from the test scanlines of the tone-mapped images. Since one single scanline is too thin to be used as a stimulus, we replicate each scanline to form a stimulus of 204 pixels high. The full width of the scanlines is used, which is equal to 512 pixels. As there are 9 images and 6 stimuli from each image, we obtained a total of 54 stimuli.

A step stimulus of equal dimensions is created for each Cornsweet stimulus. The purpose of this step stimulus is to quantify the strength of the Cornsweet illusion. The initial step size of each stimulus is set by the authors such that it produces a contrast similar to its corresponding Cornsweet stimulus. During the experiment, the magnitude of this step was varied in response to the participants' input. When a step stimulus becomes indistinguishable from its corresponding Cornsweet stimulus, the pixel difference between the left and right halves (i.e. step size) represents the strength of the Cornsweet illusion.

We normalize each test scanline to have an average pixel value of 100. Normalization is performed to control for the overall luminance level. Otherwise operators that produce brighter images would produce a stronger illusion. Normalization also makes it possible to compare the scanlines coming from dark and light regions of the input image.

Consider the example shown in Figure 6, where the top scanline with luminance profile shown with a solid red line is tone-mapped by two different operators. Although, the only difference between the two tone-mapped profiles is a global scaling factor c, the scanline output of the first operator has higher strength of illusion than the second one (the step size shown with a dashed green line shows the strength of the illusion). However, this should not be a distinguishing factor between tone-mapping operators since the output of any operator can be scaled with any arbitrary constant to change the overall luminance.

5.5 Experimental set-up

The stimuli were presented on a Sony Trinitron GDM-500PS monitor with a resolution of 1024×768 and driven by an S3 video adapter. The monitor was calibrated to a correlated color temperature of $6500K$ and $50\ cd/m^2$ maximum luminance using a GretagMacbeth Eye-One Display 2 calibration device. The stimuli were displayed on a background emitting $0.89\,cd/m^2$ (the display has a black level of $0.27\,cd/m^2$). The stimuli were viewed at a distance of approximately $150\,cm$ to the monitor. At this distance, the width and the height of the stimuli spanned 7.5 and 2.3 visual degrees respectively. The environment was illuminated by a GretagMacbeth D_{65} daylight simulator aimed at a white surface behind the monitor. There were no other sources of light in the room which was an office environment. A photograph of the experimental set-up is shown in Figure 7.

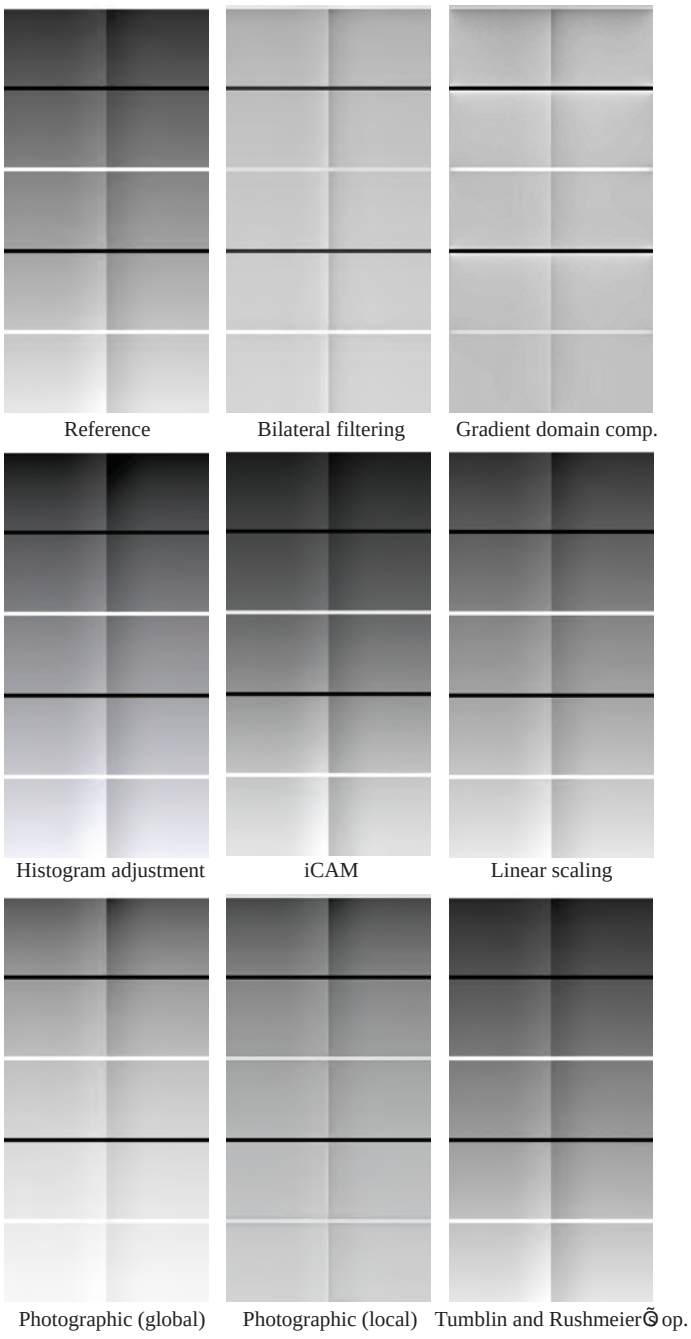


Fig. 4. *The tone-mapped images with each operator and the reference image.*

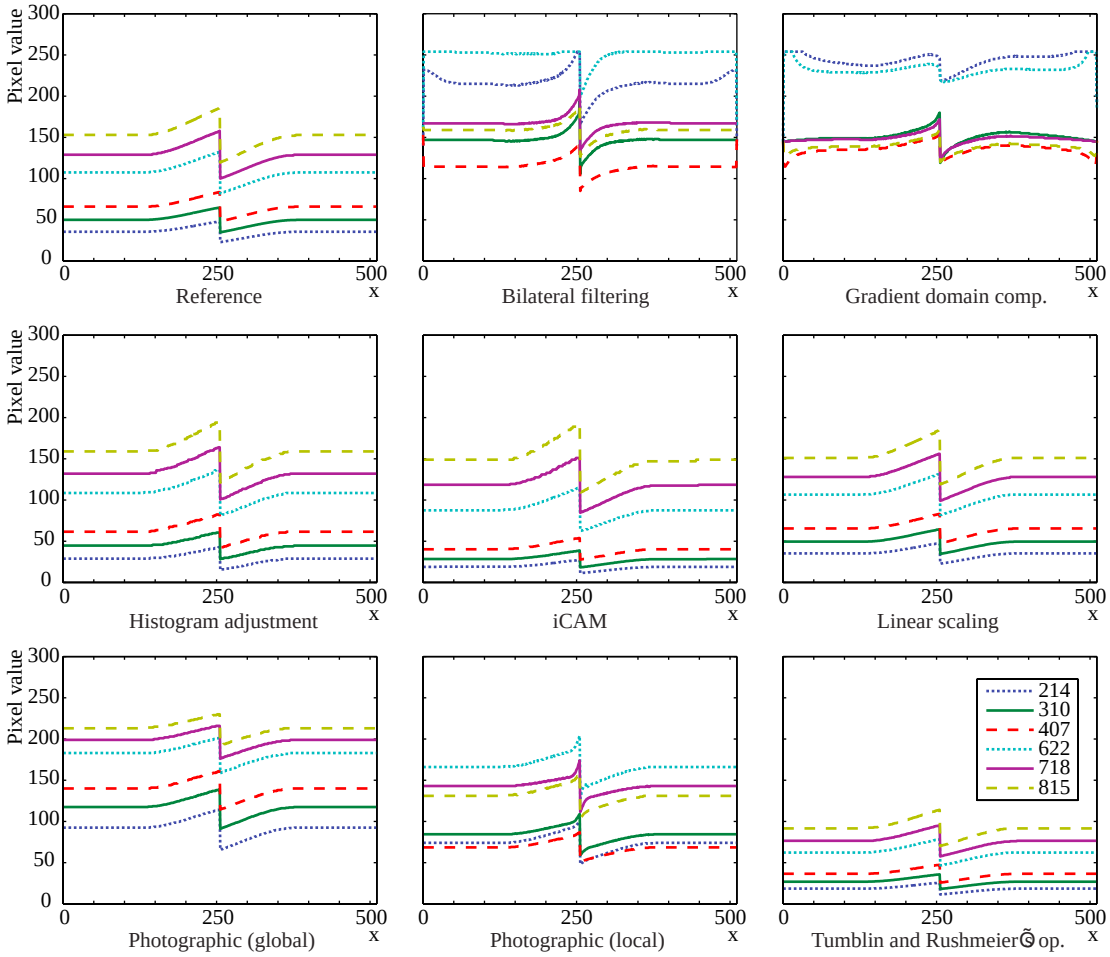


Fig. 5. *The profiles of the test scanlines after tone-mapping. Dotted, solid, and dashed lines represent scanlines next to black stripes, coming from smooth regions, and next to white stripes respectively.*

5.6 Procedure

The experiment was conducted using a temporally interleaved 2AFC procedure. Each trial consisted of a brief display of a tone-mapped (and replicated) scanline followed by a step stimulus (or vice-versa). The order of presentation was randomized. Trials for each of the 54 scanlines were interleaved in random order.

The start of each trial was indicated with a short beep, the duration of which was adjusted to render the beep audible but unobtrusive. The stimulus exposure duration was 0.5 seconds with an inter-stimulus interval of 0.1 seconds. The interval time between trials was 1.0 second during which the participant was given the opportunity to select one of the stimuli by hitting either the 1 or the 2 key on the keyboard.

The observers' task was to indicate which of the two stimuli elicited a bigger contrast (see the Appendix for a verbatim copy of the instructions given to the participants). The

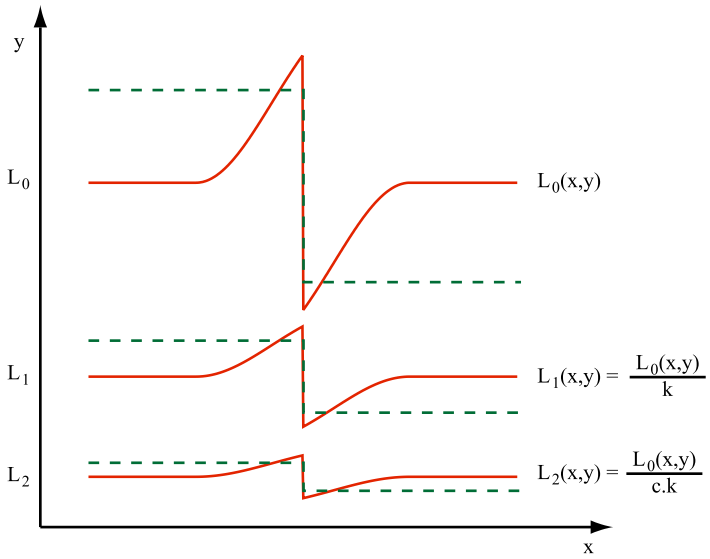


Fig. 6. *The original luminance profile (red solid line) shown at the top is tone-mapped by two operator to two values that only differ by a global scaling factor. This affects the resulting strength of the illusion (shown by dashed green lines), which we avoid by normalization.*

same pair was repeatedly shown (in randomized order) until the observer indicated a response. Participants were encouraged to give a response after every trial without letting a pair of stimuli repeat. The reason to repeat a trial if no response was given, was to allow the experiment to be interrupted if a participant needed a break. In addition, every 150 trials the participant was forced to pause for a few minutes.

If the observer indicates for a particular trial that the step stimulus has a higher contrast its contrast is reduced; otherwise its contrast is increased. The schematic representation of this procedure is illustrated in Figure 8.

A typical set of responses is shown in Figure 9. In this figure, each local minimum and maximum is called a reversal, and the experiment stops after 14 reversals. All trials prior to the first 3 reversals are considered as warm-ups, and are excluded from the subsequent analysis. With this configuration all the participants finished the experiment between 45 to 60 minutes.

The amount by which a step size is incremented or decremented after each trial is an important factor that affects both the accuracy and the rate of convergence. After a pilot study which involved both authors, we settled upon the following relation between the step size S_k for trial k and the new step size S_{k+1} for trial $k+1$:

$$S_{k+1} = S_k \pm S_k \, (0.45 + \zeta_1 \, 0.1) + \zeta_2 \, 0.01 \tag{4}$$

Here, ζ_1 and ζ_2 are uniformly distributed random variables between 0 and 1. Note that a change in the step size depends on the magnitude of the step itself. This follows from Weber's law which suggests that the amount of a noticeable change in a signal is directly proportional to the value of the signal.

Fig. 7. *A photograph of the experimental set-up, taken from the view of the participants. The fluorescent lights of the room were off during the experiment, and the only lighting was provided by the GretagMacbeth D_{65} daylight simulator, which illuminated the white surface behind the monitor, as shown.*

6. RESULTS

We collected data from 13 participants, 12 of which were naïve to the purpose of the experiment. There were 5 females and 8 males in the group, who were between 21 and 39 years old. All participants had normal or corrected to normal vision.

The data were analyzed with a repeated measures analysis of variance (ANOVA) procedure. The two within-subjects variables were the tone mapping algorithm and the scanline. We did find statistically significant differences between the average step sizes of different tone-mapping operators ($F(1.89, 22.62) = 53.50\ p < 0.001$). We show average step sizes together with their 95% confidence intervals in Figure 10.

To determine which operators are significantly different from others we conducted pairwise comparisons with Bonferonni correction against type-I errors[3]. All analyses were carried out with an alpha value of 0.05. The similarity groups are depicted in Figure 11. In this figure, statistically similar operators are shown in the same set. Note that sets are not mutually exclusive, that is an operator can belong to more than one set.

As Figure 11 shows, the operators that are in the same group as the reference image are: the local photographic operator, histogram adjustment, linear scaling, Tumblin and Rushmeier, and iCAM. The other three operators: gradient domain compression, bilateral

[3]Here, type-I error is the probability of finding a difference among operators where there is not. A type-II error would be not discovering a difference despite there actually being a difference.

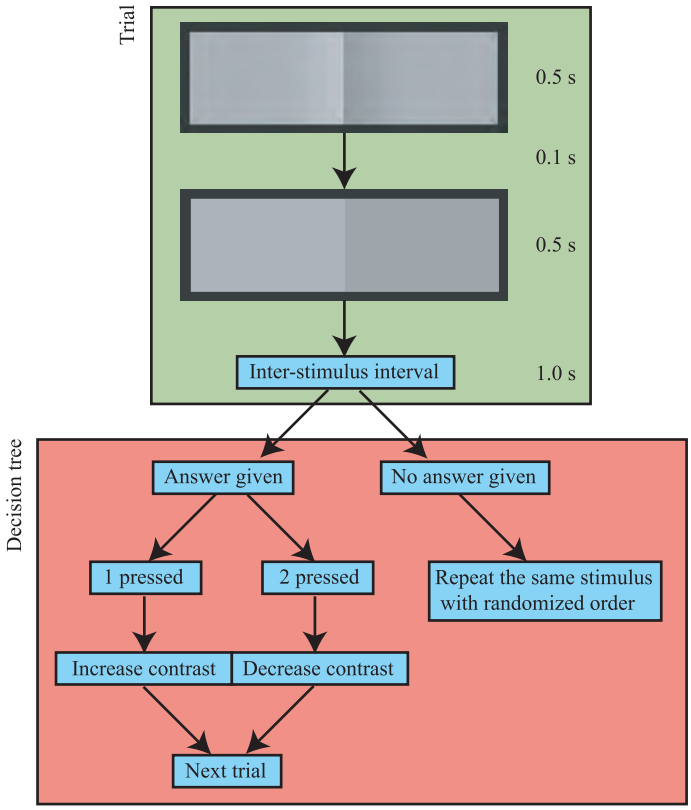


Fig. 8. *The schematic representation of the experiment.*

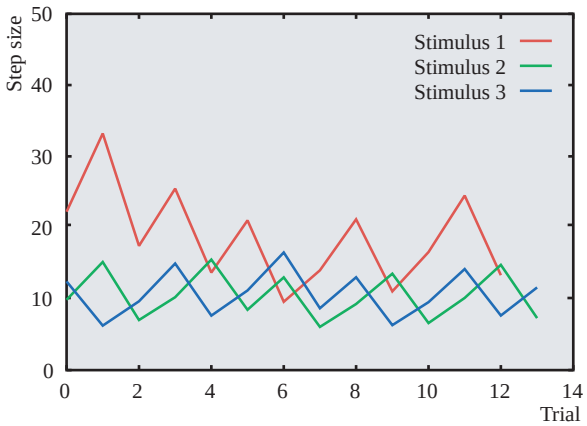


Fig. 9. *The plot of measured step size for three arbitrary stimuli that were used in the experiment.*

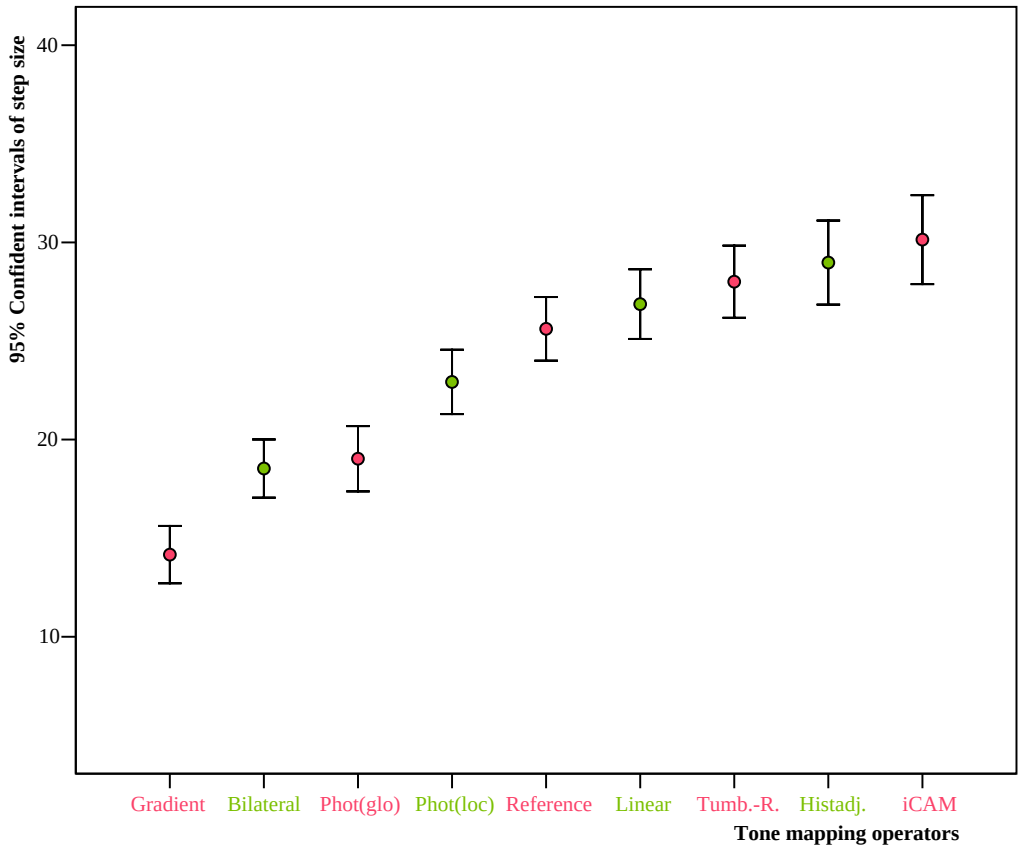


Fig. 10. *Average step sizes of the tone reproduction operators. The error bars represent* 95% *confidence intervals.*

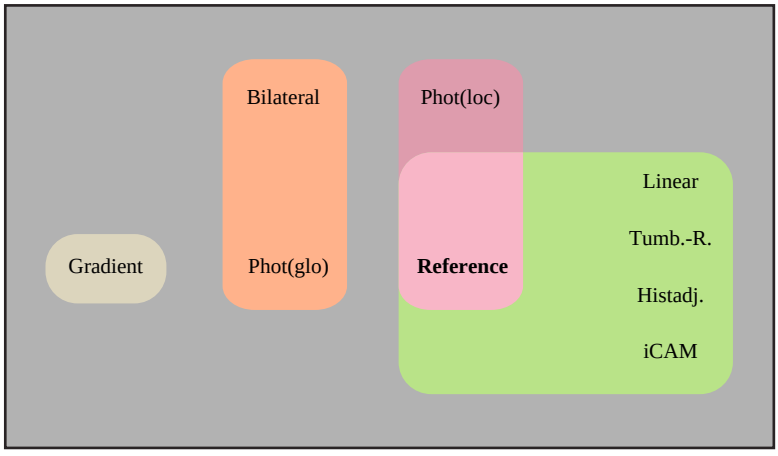


Fig. 11. *The statistical similarity groups of the tested operators.*

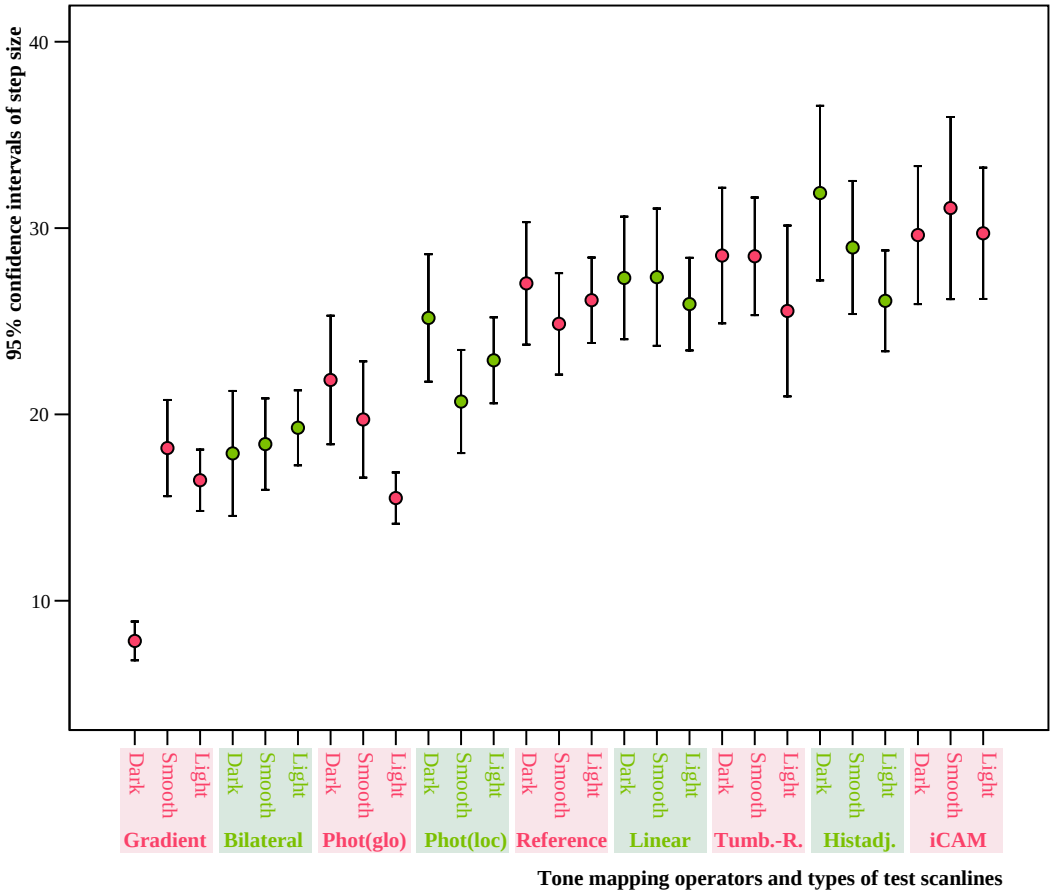


Fig. 12. *The 95% confidence intervals of the step size for three different types of scanlines: scanlines from dark and light high-contrast regions, and from smooth regions. Each line represents the average of two scanlines in each type.*

filtering, and the global photographic operator are outside this group.

This classification is based on the aggregate result of all of the 6 test scanlines. However, bilateral filtering and gradient domain compression operators suffer from haloing artifacts for scanlines next to the black stripes. As such, it is not surprising for these two operators to yield the lowest contrast when aggregate results are considered.

Deeper insight can be gained by separately analyzing the three types of scanlines. Figure 12 demonstrates the average step size elicited by each operator for scanlines next to the dark and light stripes as well the scanlines coming from the smooth regions. This figure shows that operators behave differently in different regions of the input image.

Bilateral filtering, iCAM, linear scaling, the reference, and Tumblin and Rushmeier's operator elicit a consistent step size for all three types of scanlines. However, the other operators induce significantly different step sizes for different scanlines as shown in Figure 13. These results are enumerated in Table III. In the remainder of this section we analyze each type of scanline separately.

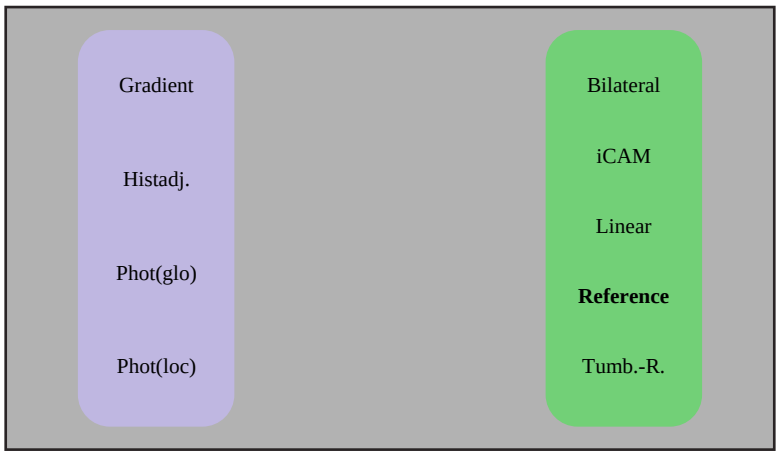


Fig. 13. *The operators that produced consistent step size for each type of scanline is shown in the green set (right) whereas the inconsistent operators is shown in the blue set (left).*

Operator	**F-number**	**p-value**
iCAM	$F(2, 50) = 0.876$	0.423
Histogram adjustment	$F(2, 50) = 12.420$	< 0.001
Tumblin-Rushmeier	$F(2, 50) = 0.916$	0.407
Linear scaling	$F(2, 50) = 0.868$	0.426
Reference	$F(2, 50) = 1.887$	0.162
Local photographic operator	$F(2, 50) = 7.680$	0.001
Global photographic operator	$F(2, 50) = 17.130$	< 0.001
Bilateral filter	$F(2, 50) = 0.868$	0.426
Gradient domain compression	$F(2, 50) = 49.993$	< 0.001

Table III. *The results for the tone-mapping consistency of the operators across different regions of the input HDR image.*

6.1 Scanlines next to black stripes

Figure 14 shows the 95% confidence intervals of step size for scanlines next to the black stripes. Significant differences are found between operators ($F(8, 96) = 46.227$ $p < 0.001$). As we can see from Figure 15, the similarity groups are almost identical to those of the aggregate results with the exception that both photographic operators fall into the same group. This is a surprising result as these scanlines are from high-contrast regions. We would expect the local operator to create a significantly higher contrast. In fact, by examining Figure 5, we can see that the physical contrast is exaggerated more by the local operator (especially for the cyan line). However, we know that as the physical contrast of the Cornsweet profile is increased the apparent contrast stops increasing beyond a certain level [Dooley and Greenfield 1977]. This could be the reason for having a similar apparent contrast for both operators.

Figure 14 reveals a trend, which is that the step size of the brighter scanline (622) is smaller than that of the darker scanline (214) for all operators. This shows most operators tend to compress high luminance regions in an HDR image more significantly, resulting in less detail for those regions. This behavior is expected from operators which use a

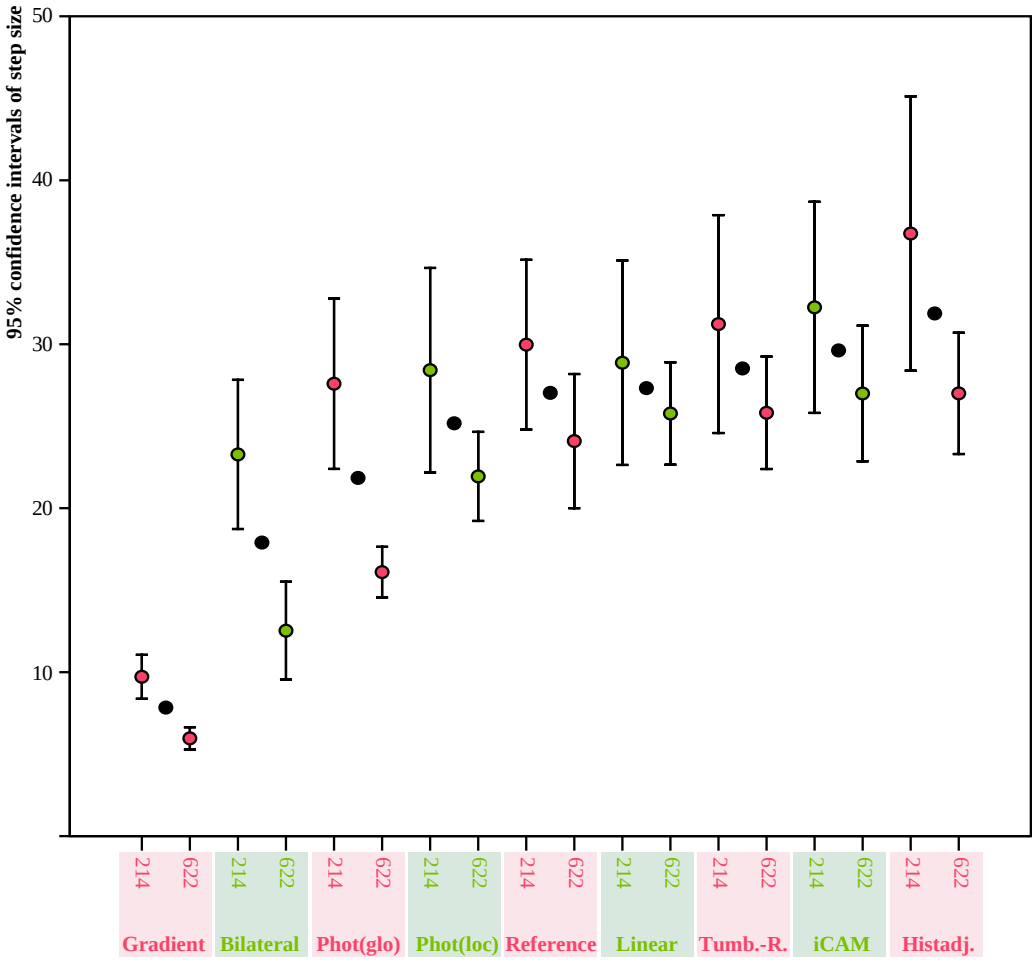


Fig. 14. *The 95% confidence intervals of the step size for scanlines next to the black stripes. Black dots show the mean step size for the two scanlines.*

sigmoidal compression curve.

For this type of scanline the gradient domain compression produces a very low contrast. Figure 5 shows that these scanlines are mapped to a very high value, and the Cornsweet profile is severely distorted. Another operator which severely distorts the Cornsweet profile is bilateral filtering (see Figure 5).

When the Cornsweet profile is significantly distorted it may cease to elicit a step profile. However since the sharp discontinuity in the middle is preserved, the left and right halves of the stimulus

6.2 Scanlines next to white stripes

Significant differences exist between operators for this type of scanline ($F(2.665, 31.981) = 32.101$ $p < 0.001$). The confidence intervals and the statistical similarity groups are shown in Figures 16 and 17. The global photographic operator produces the lowest step size, but it

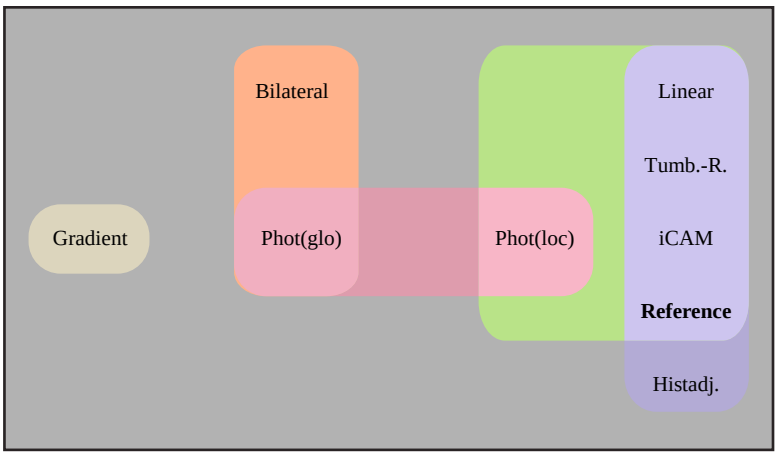


Fig. 15. *The statistical similarity groups for the scanlines next to the black stripes.*

is not statistically different from gradient domain compression. Also bilateral filtering is in the same group as the gradient domain compression. These three operators are statistically different from the other operators, all of which are in the same set as the reference.

These scanlines give insight into the behavior of the local photographic operator. Inspection of Figure 5 shows that scanlines that are next to the white stripes are more severely compressed than the other types of scanlines. On the other hand, scanlines next to dark stripes are least compressed. Thus, the contrast is increased for both of these regions. This behavior shows itself in real images with better preservation of details around highlights and in dark regions.

6.3 Scanlines from smooth regions

As Figure 18 shows, the step size is more uniformly distributed although significant differences exist between operators ($F(2.892, 34.710) = 21.011\ p < 0.001$). All the operators are in the same group as the reference with the exception of the bilateral filter and the global photographic operator as shown in Figure 19.

In this figure the presence of iCAM and the global photographic operator in the same group seems surprising as the difference between their average step sizes is high. However, what makes them indistinguishable is their high variance. Also note that for smooth scanlines the operators behave more similarly to each other as opposed a more varied behavior near high-contrast regions.

Gradient domain compression and bilateral filtering operators yield higher contrast in smooth regions than in contrast-rich regions. This naturally follows from the fact that these scanlines do not have haloing artifacts. However, these two operators still yield the lowest contrast among all operators with the exception of the global photographic operator for scanline 718.

7. DISCUSSION

Some of the results given in the previous section can be related to the functional form of each operator, while others have less clear explanations. For instance gradient domain compression attenuates high gradients more than low gradients. For the test HDR image,

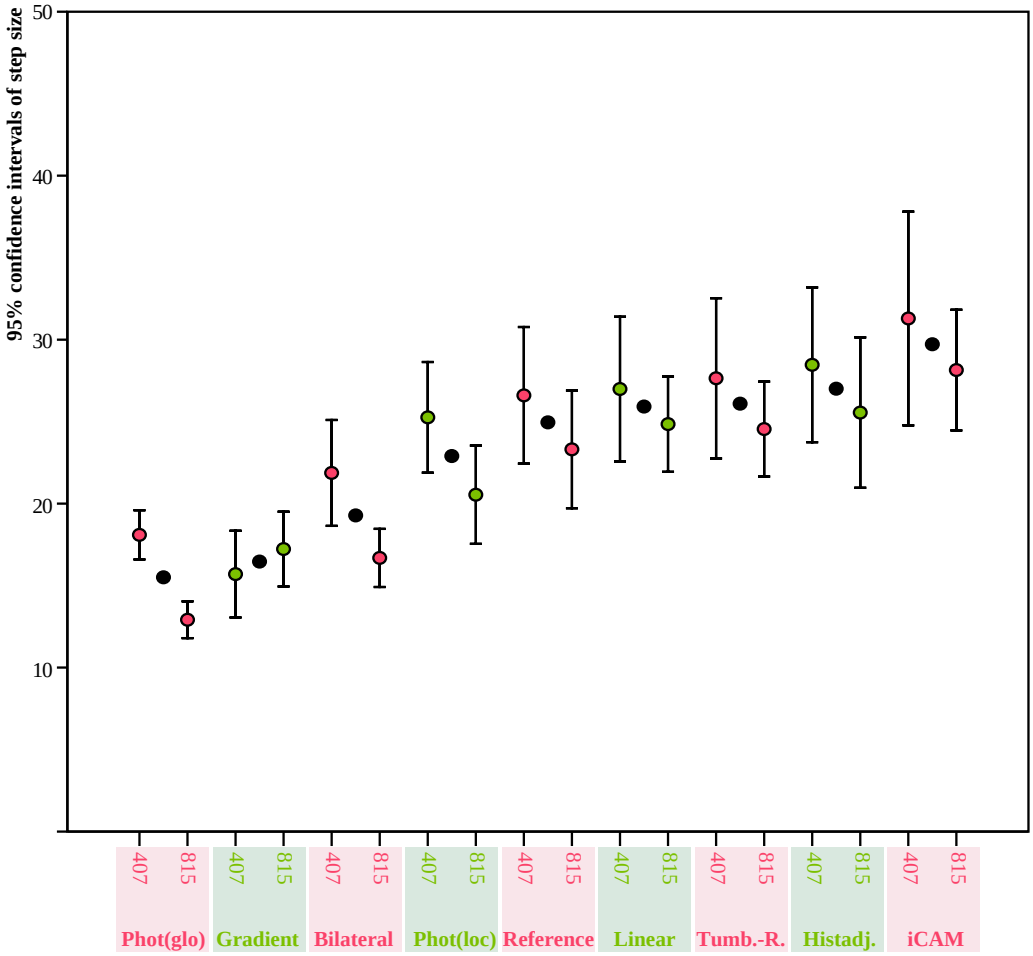


Fig. 16. *The 95% confidence intervals of the step size for scanlines next to the white stripes. Black dots show the mean step size for the two scanlines.*

the highest gradient along a scanline is the sharp transition point in the middle, and therefore it is attenuated the most. We view this as the main cause of low contrast produced by this operator.

Another cause of low-contrast for gradient domain compression is the occurrence of halo artifacts for scanlines adjacent to the black bands. This operator recovers the tone-mapped image by integrating a compressed gradient field. Integration is performed numerically using the Full Multigrid Method [Press et al. 1992]. This operation essentially involves searching the space of all 2D potential functions whose gradient is closest to the attenuated gradients in the least squared sense. In that a solution minimizing the *global* least square error is searched, some pixel values may be misassigned causing gradient reversals. The HDR image we use could be a more challenging input for this operator, since its gradient field is fairly uniform, except significant peaks next to the black and white stripes.

In our experiment bilateral filtering based operator also yields low contrast. For some

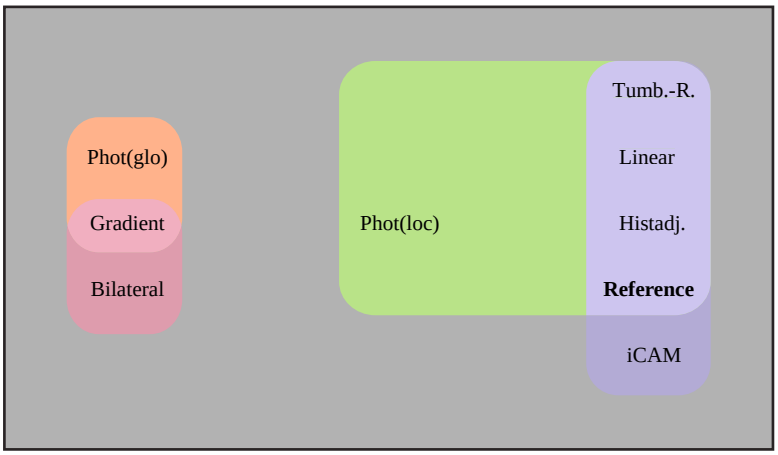


Fig. 17. *The statistical similarity groups for the scanlines next to the white stripes.*

scanlines this may be attributed to halo artifacts. While edge-preserving contrast reduction methods significantly reduce halos, they do not completely eliminate them [Tumblin and Turk 1999; Tumblin 1999]. Durand and Dorsey [2002] propose an explanation in terms of signal to noise ratio and argue that halos correspond to pixels where there is not enough information to decouple large scale and small scale features. These pixels, called uncertain pixels, are typically found in high-contrast boundaries where fewer pixels are averaged due to influence of an edge stopping function. For those pixels a statistical estimator is computed with very little data, potentially causing the halo artifacts.

To circumvent this problem, Durand and Dorsey [2002] suggest to apply an extra step of smoothing for uncertain pixels using a 2 by 2 Gaussian kernel. The final values of these pixels are obtained as a linear combination of the Gaussian smoothed image and the output of the bilateral filter. The higher the uncertainty of a pixel the more weight it gets from the Gaussian smoothed image. This causes some blurring around sharp edges, which results in halos in the tone-mapped image. Due to the small Gaussian kernel, these artifacts may be masked in real images.

The second reason of the low contrast by bilateral filtering is that it smoothes slowly varying gradients. Such gradients exist in the ramps of the Cornsweet scanlines, and smoothing them results in shorter ramp widths. However, shorter ramps are known to induce smaller contrasts [Dooley and Greenfield 1977].

In our experiment, the iCAM produces highest overall contrast, although it is not statistically different from the reference image. While this algorithm uses a sigmoidal compression, the main cause of high contrast is most likely clipping performed in the last stage of the algorithm [Johnson and Fairchild 2003]. During clipping a small percentage of pixels are allowed to burn out leaving headroom for higher dynamic range for darker pixels.

For the last two scanlines (i.e. 718 and 815), the global photographic tone-mapping operator produces the smallest contrast. This may be attributed to the sigmoidal compression employed in the algorithm. Although a true sigmoidal compression compresses both high and low luminances, this operator uses a compression scheme which compresses mainly the high luminances [Reinhard et al. 2002].

For the two scanlines that are next to the light intensity bands, the local photographic op-

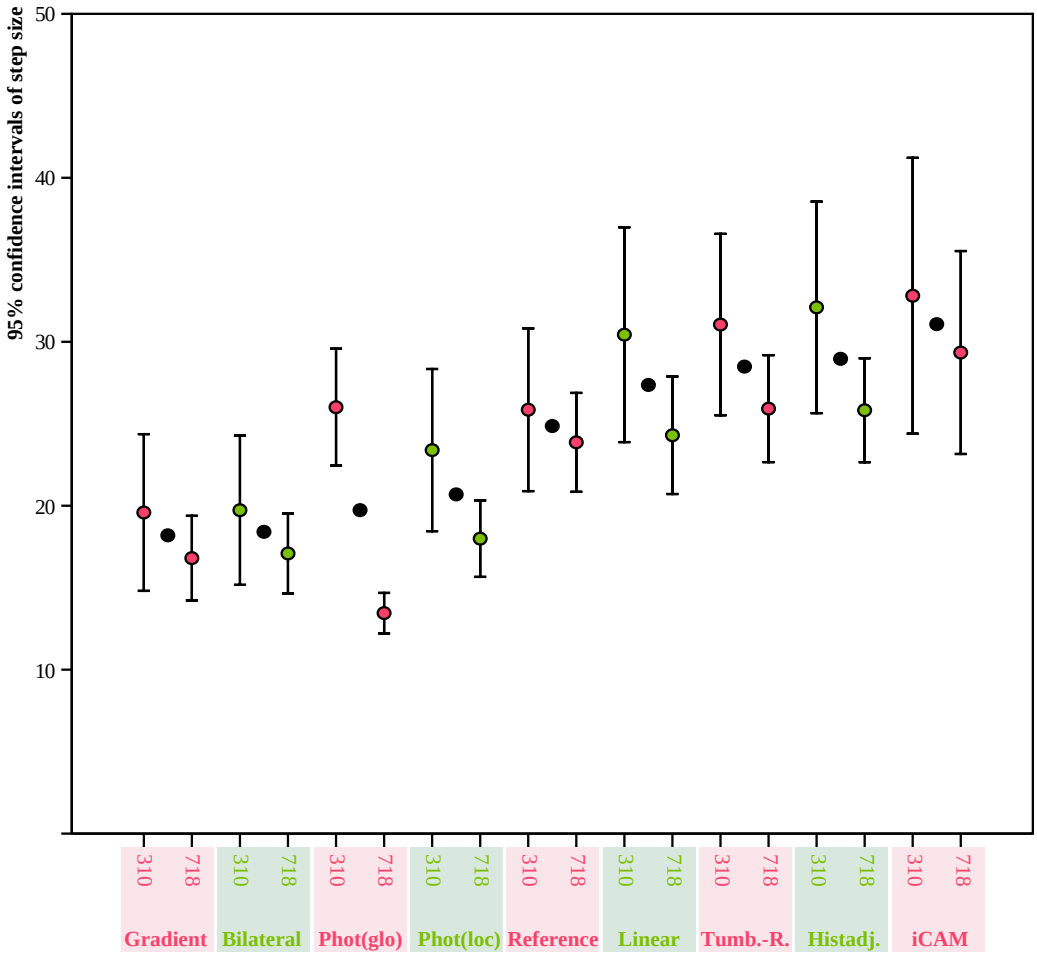


Fig. 18. *The 95% confidence intervals of the step size for scanlines coming from the smooth regions. Black dots show the mean step size for the two scanlines.*

erator produces significantly higher contrast than the global operator. This is an expected result since the local operator is designed to preserve contrast by altering the sigmoidal compression curve according to the neighborhood of the pixel of interest. The local operator produces higher contrast also for the scanlines next to the dark bands, however for those scanlines the contrast difference with the global operator is not significant.

Tumblin and Rushmeier's operator is a global tone-mapping operator which aims to match the brightness of the original scene with the brightness of the displayed image. In doing so, input luminances are mapped to output luminances via a nonlinear function akin to a gamma curve. Due to its nonlinearity, pixels with different luminances are compressed differently. Thus, we do not expect to find significant differences between scanlines selected from high contrast and smooth regions. This is indeed the case for this operator, where there is a monotonic decrease in contrast as we move from dark to light regions in the image.

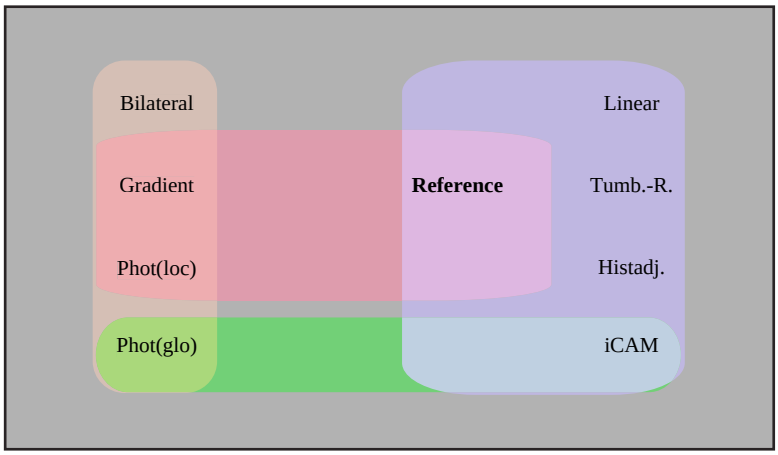


Fig. 19. *The statistical similarity groups for the scanlines coming from the smooth regions.*

The histogram adjustment operator is found to be similar to the reference image in all types of scanlines. This result is despite the fact the input HDR image has a very sparse histogram, with two peaks corresponding to black and white stripes. Such an histogram constitutes a more difficult case for this operator because its cumulative histogram contains a large region of luminance values which will be mapped to the same output. As such, this operator could potentially perform better for HDR images with more highly populated histograms.

When a tone-mapping algorithm alters the shape of the Cornsweet profile significantly, such a profile may cease to induce a perfect step edge. In this case, luminance variations may be detected in either half of the profile. In our study, this occurs only for the two scanlines next to the black bands for bilateral filtering and gradient domain compression algorithms. As shown in Figure 5, the luminance profiles of these two scanlines are significantly distorted, and as such they do not induce perfect steps. However, due to the persistence of the sharp discontinuity in the middle an overall contrast between the left and right halves is still perceived. In these cases, we expect the subjects to base their decisions on the overall contrast between the left and right halves of a stimulus.

The LDR reference approach we use in our study has several benefits as well as potential limitations. Its chief benefit is that it allows to compare tone-mapping operators against a reference without using an HDR display device or a real scene. In addition to being rare and expensive, current HDR display devices are not well suited for psychophysical studies that use simple stimuli such as the one in this experiment[4]. On the other hand, real scenes contain a multitude of factors which need to be meticulously accounted for. The comparison of a 3-D world against 2-D images is itself a challenging task for participants. The LDR reference approach utilizes a well understood and calibrated CRT monitor, and therefore bypasses the problems associated with both approaches.

One challenge of the LDR reference approach is that the reference image should repre-

[4]Since the current HDR displays are composed of a low resolution LED layer behind a high resolution LCD panel, displaying dots, lines, or simple stimuli such as the one we use in this experiment cause halo artifacts. These artifacts are due to light leakage from LCD pixels and may go unnoticed when natural images are displayed.

sent an ideal tone-mapping of its HDR original. Given the difficulties in determining an ideal tone-mapping for a natural HDR image, this may limit us to the use of only abstract stimuli generated from functions whose parameters can be adjusted as in our experiment.

8. RELATION TO PREVIOUS VALIDATION STUDIES

Using pairwise comparisons, Drago et al. [2002] evaluated operators' performance with regard to naturalness and subjects' preference. The results show that the photographic tone-mapping operator [Reinhard et al. 2002] produces the most natural looking images, and also it is also preferred the most by the subjects. In our study, we found this operator to preserve contrast very well and this may be one of the reasons behind subjects' preference in Drago et al.'s experiment. Drago et al. also found that the histogram adjustment technique [Ward et al. 1997] enhances contrast, but perhaps for this reason, its resulting images are not considered very natural. We have also found histogram adjustment to produce high contrast. However, in our experiment this operator does not enhance contrast but accurately reproduce it. Therefore, we may argue that the histogram adjustment technique may be used where high-contrast is desired.

In another pairwise comparison study, Kuang et al. [2004] found bilateral filtering [Durand and Dorsey 2002] and photographic tone-mapping [Reinhard et al. 2002] the most preferred, and gradient domain compression [Fattal et al. 2002] the least preferred operators. This result is only in partial agreement with our study. In our study, the gradient domain operator produces the lowest contrast, and this may explain why it is not preferred by observers in Kuang et al.'s study. However, we have also found bilateral filtering to yield low contrast although this operator is preferred the most in Kuang et al.'s experiment.

Yoshida et al. evaluated tone-mapping operators using real world scenes [2005], with results similar to ours. The authors found global operators to generate more *overall* contrast than local operators. This trend is also seen in Figure 4 of our study, where global operators show a larger variation of luminance in the vertical direction. One exception is iCAM, which although being a local operator yields high overall contrast. This is most likely due to its large kernel size, i.e. one quarter of the image, which makes iCAM behave similar to the global operators. The authors also found histogram adjustment [Ward et al. 1997] and photographic tone-mapping [Reinhard et al. 2002] to produce the most natural looking images. This result is partially contradicted by Drago et al.'s study where the histogram adjustment operator was not considered natural. However, the good performance of the photographic tone-mapping operator is also supported by this study.

Ledda et al. [2005] performed an evaluation study where tone-mapped images are compared against linearly scaled images displayed on an HDR monitor. The authors found the iCAM [Fairchild and Johnson 2002] and the photographic tone reproduction operator [Reinhard et al. 2002] to be the most similar to the reference images displayed on the HDR monitor. Conversely, bilateral filtering is found to be the least similar to the reference. These results are largely in parallel to ours where we also found that iCAM and the photographic operator are in the same statistical group as the reference. Bilateral filtering is not in this group.

In a recent study, Ashikhmin and Goyal investigate whether subjects' awareness of real world scenes play a role in their interpretation of naturalness [2006]. The authors conclude that if subjects can see the original scenes while evaluating tone-mapped images, the final rankings may change significantly compared to participants not knowing the original

scenes. When subjects are allowed to see the actual scenes, the gradient domain compression is systematically chosen as the most natural, although there is no clear winner in the absence of actual scene information. This result is in contradiction with our study where this operator poorly reproduces contrast.

In another evaluation study, Cădík et al. [2006] also asked subjects to express the overall image quality with and without seeing the real world scene[5]. The results show that the knowledge of the original scene plays only a minor role in subjects' preference, a finding in conflict with Ashikhmin and Goyal's evaluation. In this study, the photographic tone-mapping and histogram adjustment operators produced the best overall quality, while gradient domain compression and bilateral filtering based operators were low in ranking. These results are in agreement with our contrast preservation findings.

9. CONCLUSIONS

In this paper we present an evaluation study of tone reproduction operators using the Cornsweet illusion. We focus on a single task, related to contrast perception, in an artificially created high dynamic range image. This provides straightforward control over many factors, and therefore affords a robust experimental design.

The human visual system is very sensitive to contrast, and we therefore argue that tone reproduction operators should preserve contrast as one of their primary goals. Our experimental set-up is chosen such that we are able to assess contrast reproduction after tone-mapping. The use of the Cornsweet illusion allows us to evaluate the operators for a continuous range of luminances while still affording a simple task. This would have been very difficult to achieve if either natural images or pure step stimuli were used. Natural images, in general, are too complex to single out and evaluate a specific attribute. On the other hand, pure step stimuli would be too simple to hold meaning, since any anomalous tone-mapping of luminance values between steps would go unnoticed.

Our study has two immediate conclusions. First, tone-mapping operators do not preserve the Cornsweet illusion found in an high dynamic range image equally well. Some of them accentuate the illusion, while others make it less pronounced. Second, most tone-mapping operators alter the strength of the Cornsweet illusion differently for different regions of the input HDR image. This is partially due to gradient reversals that may occur in high-contrast regions, particularly for local operators. For global operators, the shape of the compression curve they use seem to have a profound effect on their consistency for different regions of the input image.

The HDR image we use is low dynamic range in every scanline, but covers a wide range of luminances from top to bottom. Therefore, we argue that an ideal operator should preserve the horizontal contrast and only compress the vertical contrast. Such a hypothetical tone-mapping is represented by our reference image. As such, we regard operators that induce statistically similar contrasts to the reference image as optimal in terms of preserving the Cornsweet illusion. The possibility to compare tone-mapping operators against a reference without requiring the use of an HDR display device is a novel idea that we introduce in this paper.

Tone-mapping operators are designed to work on real images, rather than artificial images such as the one used in this paper. However, there are two benefits to using artificial

[5]Only one indoor scene is used in this study, although the authors evaluate 14 different operators, the highest number of operators evaluated in a single study so far.

stimuli to evaluate tone reproduction operators. First, the luminance distribution of the image can be steered to represent various types of scenes. For instance, the image we use has a histogram that resembles a predominantly bright scene with some very bright and very dark regions. An outdoor scene with dark shadows and highlights would have a similar histogram. Second, artificial images may help discover artifacts and anomalies that could be introduced by some operators. Many artifacts would go unnoticed in natural images as a result of masking effects. For example, we noticed that bilateral filtering based tone-mapping may cause halos of 1 pixel wide around sharp edges due to its uncertainty computation. Gradient domain compression may also cause halos, as the numerical solution it employs does not guarantee to preserve gradient directions.

We believe validation studies involving natural images and abstract patterns are both worthwhile and important. While studies based on natural images reveal people's preference for some operators over others, studies involving abstract patterns reveal the underlying causes of these preferences. While the former type of studies come up with subjective rankings meaningful only for humans, the latter type should explain the objective grounds behind those rankings.

To a large extent, our results are in agreement with those of the previous validation studies. Therefore, we may speculate that there exist a correlation between the reproduction of contrast and naturalness as well as people's subjective preference. Despite similarities, some differences also exist among previous validation studies including ours. The most notable as we see is the fluctuating performances of the bilateral filtering and gradient domain based tone-mapping operators. While they perform very good in some studies, their performance drops in others. This suggest that subtle implementation details and parameter settings of these operators maybe playing a major role in the quality of their outputs.

In this study we focused on contrast preservation. Nevertheless, contrast is not the only factor that determines suitability for different tasks. As such, other validation studies would be required to fully characterize tone reproduction operators. Such validation studies may for instance focus on operators' ability to preserve visibility or brightness.

Appendix: Participants' instructions

The following instructions were given to the participants prior to the start of the experiment:

> "You will hear a beep and then you will see two displays in short succession. The left and right half of each display will have a different grey value. Please indicate if the first or the second display showed the biggest difference in grey value by hitting the '1' or the '2' key. Guess if you can't decide, but always give an answer (1 or 2). A new trial starts when you hear the next beep. If you feel you have made a mistake, do not worry, but just continue with the next trial.
> You may take a break at any time if you want; this will not harm the experiment. In addition, there will be regular breaks scheduled during the experiment. If for whatever reason you wish to terminate the experiment you may do so as well without consequence to you."

REFERENCES

ASHIKHMIN, M. 2002. A tone mapping algorithm for high contrast images. In *Proceedings of 13th Eurographics Workshop on Rendering*. 145–155.

ASHIKHMIN, M. AND GOYAL, J. 2006. A reality check for tone mapping operators. to appear in ACM Transactions on Applied Perception 3(4).

CORNSWEET, T. 1970. *Visual perception*. Academic Press, New York.

CRAIK, K. J. W. 1966. The nature of psychology: a selection of papers, essays and other writings. S. L. Sherwood, Ed. Cambridge University Press, Cambridge, 94–97.

CĂDÍK, M., WIMMER, M., NEUMANN, L., AND ARTUSI, A. 2006. Image attributes and quality for evaluation of tone mapping operators. to appear in Proceedings of Pacific Graphics.

DOOLEY, R. P. AND GREENFIELD, M. I. 1977. Measurement of edge-induced visual contrast and a spatial-frequency interaction of the cornsweet illusion. *J. Opt. Soc. Am. 67*, 6 (June), 761–765.

DOWLING, J. E. 1987. *The Retina: An Approachable Part of the Brain*. Harvard University Press.

DRAGO, F., MARTENS, W. L., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2002. Perceptual evaluation of tone mapping operators with regard to similarity and preference. Tech. Rep. MPI-I-2002-4-002, Max Plank Institut für Informatik.

DURAND, F. AND DORSEY, J. 2002. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics 21*, 3, 257–266.

ELDER, J. H. 1999. Are edges incomplete? *International Journal of Computer Vision 34*, 2/3, 97–122.

FAIRCHILD, M. D. 2005. *Color appearance models*. John Wiley & Sons. Second edition.

FAIRCHILD, M. D. AND JOHNSON, G. M. 2002. Meet iCAM: an image color appearance model. In *IS&T/SID* 10^{th} *Color Imaging Conference*. Scottsdale, 33–38.

FAIRCHILD, M. D. AND JOHNSON, G. M. 2004. The iCAM framework for image appearance, image differences, and image quality. *Journal of Electronic Imaging*.

FAIRCHILD, M. D., JOHNSON, G. M., KUANG, J., AND YAMAGUCHI, H. 2004. Image appearance modelling and high-dynamic-range image rendering. In *First ACM Symposium on Applied Perception in Graphics and Visualization (APGV)*. 159.

FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. *ACM Transactions on Graphics 21*, 3, 249–256.

GROSSBERG, S. AND MINGOLLA, E. 1985. Neural dynamics of form perception: boundary adaptation, illusory figures, and neon color spreading. *Psychological Review 92*, 173–211.

JOHNSON, G. M. AND FAIRCHILD, M. D. 2003. Rendering HDR images. In *IS&T/SID* 11^{th} *Color Imaging Conference*. Scottsdale, 36–41.

KINGDOM, F. AND MOULDEN, B. 1988. Border effects on brightness: a review of findings, models and issues. *Spatial Vision 3*, 4, 225–262.

KRAUSKOPF, J. 1963. Effect of retinal stabilization on the appearance of heterochromatic targets. *J Opt Soc Am 53*, 741–744.

KUANG, J., YAMAGUCHI, H., JOHNSON, G. M., AND FAIRCHILD, M. D. 2004. Testing hdr image rendering algorithms. In *in Proceedings of IS&T/SID 12th Color Imaging Conference*. 315–320.

KUFFLER, S. W. 1953. Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology 16*, 37–68.

LAND, E. H. AND MCCANN, J. J. 1971. Lightness and retinex theory. *J. Opt. Soc. Am. 63*, 1, 1–11.

LEDDA, P., CHALMERS, A., AND SEETZEN, H. 2004. A psychophysical validation of tone mapping operators using a high dynamic range display. In *APGV '04: Proceedings of the 1st Symposium on Applied perception in graphics and visualization*. ACM Press, 159.

LEDDA, P., CHALMERS, A., TROSCIANKO, T., AND SEETZEN, H. 2005. Evaluation of tone mapping operators using a high dynamic range display. *ACM Transactions on Graphics 24*, 3, 640–648.

MARR, D. 1982. *Vision, a computational investigation into the human representation and processing of visual information*. W H Freeman and Company, San Fransisco.

O'BRIEN, V. 1958. Contour perception, illusion and reality. *Journal of the Optical Society of America 49*, 112–119.

PALMER, S. E. 1999. *Vision Science: Photons to Phenomenology*. The MIT Press, Cambridge, Massachusetts.

PATTANAIK, S. N., FERWERDA, J. A., FAIRCHILD, M. D., AND GREENBERG, D. P. 1998. A multiscale model of adaptation and spatial vision for realistic image display. In *SIGGRAPH 98 Conference Proceedings*. 287–298.

PATTANAIK, S. N., TUMBLIN, J., YEE, H., AND GREENBERG, D. P. 2000. Time-dependent visual adaptation for fast realistic display. In *SIGGRAPH 2000 Conference Proceedings*. 47–54.

PESSOA, L., THOMPSON, E., AND NOË, A. 1998. Finding out about filling-in: A guide to perceptual completion for visual science and the philosophy of perception. *Behavioral and Brain Sciences 32*, 723 – 802.

PRESS, W. H., TEUKOLSKY, S. A., WETTERLING, W. T., AND FLANNERY, B. P. 1992. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press.

REINHARD, E. 2003. Parameter estimation for photographic tone reproduction. *Journal of Graphics Tools 7*, 1, 45–51.

REINHARD, E. AND DEVLIN, K. 2005. Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization and Computer Graphics 11*, 1 (January/February), 13–24.

REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic tone reproduction for digital images. *ACM Transactions on Graphics 21*, 3, 267–276.

REINHARD, E., WARD, G., DEBEVEC, P., AND PATTANAIK, S. 2005. *High Dynamic Range Imaging: Acquisition, Display, and Image Based Lighting*. Morgan Kaufmann, San Francisco, CA.

SCHLICK, C. 1994. Quantization techniques for the visualization of high dynamic range pictures. In *Photorealistic Rendering Techniques*, P. Shirley, G. Sakas, and S. Müller, Eds. Springer-Verlag Berlin Heidelberg New York, 7–20.

SEETZEN, H., HEIDRICH, W., STUERZLINGER, W., WARD, G., WHITEHEAD, L., TRENTACOSTE, M., GHOSH, A., AND VOROZCOVS, A. 2004. High dynamic range display systems. In *ACM Trans. on Graphics*. Vol. 23. 760–768.

TOMASI, C. AND MANDUCHI, R. 1998. Bilateral filtering for gray and color images. In *Proc. of the 1998 IEEE International Conference on Computer Vision*. IEEE, 839–846.

TUMBLIN, J. 1999. Three methods of detail-preserving contrast reduction for displayed images. Ph.D. thesis, College of Computing Georgia Inst. of Technology.

TUMBLIN, J., HODGINS, J. K., AND GUENTER, B. K. 1999. Two methods for display of high contrast images. *ACM Transactions on Graphics 18 (1)*, 56–94.

TUMBLIN, J. AND TURK, G. 1999. LCIS: A boundary hierarchy for detail-preserving contrast reduction. In *Siggraph 1999, Computer Graphics Proceedings*, A. Rockwood, Ed. Annual Conference Series. Addison Wesley Longman, Los Angeles, 83–90.

WACHTLER, T. AND WEHRHAHN, C. 1997. The craik - o'brien - cornsweet illusion in colour: Quantitative characterisation and comparison with luminance. *Perception 26*, 1423 – 1430.

WALLS, G. 1954. The filling-in process. *American Journal of Optometry 31*, 329–340.

WARD, G., RUSHMEIER, H., AND PIATKO, C. 1997. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics 3*, 4.

YOSHIDA, A., BLANZ, V., MYSZKOWSKI, K., AND SEIDEL, H. 2005. Perceptual evaluation of tone mapping operators with real-world scenes. In *Stereoscopic Displays and Virtual Reality Systems XII. Edited by Woods, Andrew J.; Bolas, Mark T.; Merritt, John O.; McDowall, Ian E. Proceedings of the SPIE, Volume 5666, pp. 192-203 (2005).* 192–203.