Evaluating the Color Fidelity of ITMOs and HDR Color Appearance Models

MEKIDES ASSEFA ABEBE, Technicolor Research & Innovation and Université de Poitiers
TANIA POULI and JONATHAN KERVEC, Technicolor

With the increasing availability of high-dynamic-range (HDR) displays comes the need to remaster existing content in a way that takes advantage of the extended range of luminance and contrast that such displays offer. At the same time, it is crucial that the creative intent of the director is preserved through such changes as much as possible. In this article, we compare several approaches for dynamic range extension to assess their ability to correctly reproduce the color appearance of standard dynamic range (SDR) images on HDR displays. A number of state-of-the-art inverse tone mapping operators (ITMOs) combined with a standard chromatic adaptation transform (CAT) as well as some HDR color appearance models have been evaluated through a psychophysical study, making use of an HDR display as well as HDR ground-truth data. We found that global ITMOs lead to the most reliable performance when combined with a standard CAT, while more complex methods were found to be more scene dependent, and often less preferred than the unprocessed SDR image. HDR color appearance models, albeit being the most complete solutions for accurate color reproduction, were found to not be well suited to the problem of dynamic range expansion, suggesting that further research may be necessary to provide accurate color management in the context of inverse tone mapping.

CCS Concepts: • **Computing methodologies → Model verification and validation**; *Image processing*; • **Hardware → Emerging tools and methodologies;**

Additional Key Words and Phrases: Color appearance modeling, ITMO, high-dynamic-range imaging, subjective evaluation

ACM Reference Format:

Mekides Assefa Abebe, Tania Pouli, and Jonathan Kervec. 2015. Evaluating the color fidelity of ITMOs and HDR color appearance models. ACM Trans. Appl. Percept. 12, 4, Article 14 (September 2015), 16 pages.
DOI: http://dx.doi.org/10.1145/2808232

1. INTRODUCTION

High-dynamic-range (HDR) imaging has attracted a lot of attention over more than a decade, both within the research community and, more recently, in the industrial world. Briefly, HDR encompasses tools and techniques for capturing, processing, storing, and displaying content with a wider range of luminance [Reinhard et al. 2012]. A key problem in this area has been that of remapping the dynamic range of images or video content to displays of different capabilities. Techniques for compressing the dynamic range of HDR content for standard-dynamic-range (SDR) displays are known as tone mapping operators (TMOs), while solutions addressing the inverse problem are historically referred to as reverse or inverse tone mapping operators (ITMOs).

Authors' addresses: M. A. Abebe; email: mekides123@gmail.com; T. Pouli; email: Tania.Pouli@technicolor.com; J. Kervec; email: jonathan.kervec@technicolor.com.

DOI: http://dx.doi.org/10.1145/2808232

The latter case, namely, the extension of the luminance and contrast range of content, is likely to significantly increase in importance in the coming years. Although HDR displays are starting to become available to consumers, content is not as readily available. Most current commercial efforts to offer HDR content in fact rely on manual remastering by professional artists—a process that is neither simple nor cheap. Given the vast quantities of content currently available, however, the need for automated solutions for HDR content remastering is becoming apparent. A number of ITMO algorithms have been proposed in the literature, which will be discussed in more detail in the following section. The focus of existing techniques has been on exploring different approaches for maximizing dynamic range without creating artifacts. However, in digital color image reproduction, if the viewing conditions change, both luminance and chromatic information may need to be modified to preserve the appearance of the image. Therefore, just expanding or compressing the luminance range and the contrast of SDR content may not be sufficient for accurate HDR color image reproduction or, in a content creation context, for preserving the director's intent.

In fact, there are several color appearance phenomena that are affected by a change in the luminance of the stimuli, the background, and the viewing environment [Reinhard et al. 2008; Akyüz and Reinhard 2006]. The Stevens effect describes the increase in local perceptual brightness contrast due to the increase in luminance level, while the Hunts effect states that an increase in luminance level results in an increase in perceived colorfulness. Helmholtz-Kohlrausch and Bezold-Bruckes effects show how an increase in luminance level also affects the brightness and causes a hue shift [Fairchild 2013]. Therefore, to accurately model such appearance effects when modifying the luminance and contrast of content, a more complete model of human vision and color perception may be needed.

Traditional color and image appearance models are mainly optimized according to low-luminance-level datasets and are designed for perceptual appearance of reflective surfaces and backlit advertisement materials [CIETC:801 2000]. Therefore, they cannot be readily applied to HDR image reproduction. A small number of recent models [Kuang et al. 2007a; Kim et al. 2009; Reinhard et al. 2012] have been designed with HDR luminance levels in mind; however, to our knowledge, none of them have been explicitly evaluated in the context of dynamic range extension. At the same time, although several studies exist comparing ITMOs subjectively [Akyüz et al. 2007; Banterle et al. 2009; Masia et al. 2009], the ability of existing ITMO solutions to preserve the color appearance of content has not been formally assessed.

To evaluate the appearance reproduction capabilities of existing ITMOs and color appearance models (CAMs) and to better understand where their limitations lie, we perform a psychophysical study based on the ITU-recommended SAMVIQ protocol [ITU-R 2012], where we compare several color appearance models and ITMOs in their ability to match the appearance of native HDR content when extending the dynamic range of SDR images of the same scenes. We explore the combination of ITMO algorithms with standard chromatic adaptation transforms [Sharma 2002], akin to the work of Akyuz et al. [2004]. In addition, to better quantify the effect of each method, we analyze appearance attributes in both HDR and expanded SDR images, processed by different methods.

2. RELATED WORK

2.1 Inverse Tone Mapping Operators

Solutions that aim at expanding the luminance and contrast range of SDR images in order to prepare them for viewing on HDR displays are known as ITMOs. Some approaches focus on increasing the effective bit depth of images in order to reduce quantization artifacts [Daly and Feng 2003, 2004]. However, most ITMOs assume that part of the image will be more or less well exposed and therefore does not require much change, while highlights and bright areas should be expanded.

In an effort to enhance clipped highlights in non-HDR environment maps, Landis [2002] proposed a power function for expanding the luminance of SDR images. Akyüz et al. [2007] explored several different gamma-like global curves through a psychophysical experiment to assess their effectiveness for expanding the dynamic range of images, finding that in fact a parameterizable linear expansion was preferred. In a more extensive study, Masia et al. [2009] and Masia and Gutierrez [2011] come to similar conclusions and propose an image-dependent gamma function for expanding images with large overexposed areas.

More sophisticated expansion strategies have also been proposed. Several methods aim at constructing a map of highlights that allows them to not only expand bright areas more but also reshape the otherwise flat appearance of these areas. Banterle et al. [2006, 2007] achieve that through a weighted combination of an inverse sigmoidal expansion and the original image depending on such a map, an approach further extended by Huo et al. [2014], who add a local *dodging and burning* step to further expand luminance in both the bright and dark direction. A similar approach is followed by Kovaleski and Oliveira [2009], who use a set of bilateral filtering steps to obtain the expansion map. Meylan et al. [2006, 2007] opt for linear expansion but expand diffuse and highlight areas differently, while Rempel et al. [2007] combine linear expansion with a Gaussian luminance reshaping step for highlight areas.

Highlight reconstruction is managed more explicitly in some methods. Didyk et al. [2008] use several features to classify bright areas into diffuse surfaces, reflections, and light sources and propose different treatments for each case, while Wang et al. [2007] employ user-guided inpainting combined with luminance reshaping to reconstruct the texture and detail in highlight areas.

In all these cases, the emphasis of proposed methods is on expanding the dynamic range of luminance and contrast information in images. However, the appearance of images depends not only on luminance but also on chromatic information. As such, to accurately recover the appearance of the original HDR scene when expanding the dynamic range of content, it is necessary to also consider the interaction between chromatic and luminance information. For this, more elaborate models are needed, which take into consideration perceptual phenomena regulating the appearance of color.

2.2 HDR Appearance Models

A CAM can be defined as "...any model with some form of a chromatic adaptation transform and includes predictors of at least the relative color-appearance attributes of lightness, chroma, and hue" [Sharma 2002]. Effectively, CAMs try to predict the appearance of a given stimulus under different viewing conditions, given parameters about the stimulus itself, the original scene, and the target viewing environment [Fairchild 2013]. To model appearance changes, these models are traditionally optimized based on psychophysical data that describe appearance changes for different patches and environments as described by viewers. As these models are based on simple stimuli, however, they cannot be readily generalized to complete images. For such complex stimuli, spatially localized adaptation and other spatial phenomena should be considered.

To model the appearance of complex, spatially varying stimuli, image appearance models, such as s-CIELAB [Zhang and Silverstein 1997] and iCAM [Fairchild and Johnson 2002, 2004], have been developed. These models consider spatial aspects in addition to colorimetric ones and try to model adaptation as a local process. Traditionally, however, these models have been optimized using SDR stimuli and as such they cannot accurately predict appearance changes in HDR. To that end, the iCAM model was extended into iCAM06 [Kuang et al. 2007a], by incorporating a dynamic range compression step. To extend the sets of available color appearance data to HDR luminance levels, Kim et al. [2009] conducted a series of psychophysical experiments and subsequently constructed a model based on the resulting data. More recently, Reinhard et al. [2012] proposed a simpler model that also takes into account display characteristics, which was validated based on an extensive set of both SDR and HDR data.

Although the more recent appearance models consider HDR luminance levels and offer sophisticated tone compression capabilities, to our knowledge, none of the available models explicitly consider the challenges relating the opposite problem, namely, going from low to much higher luminance levels. In the case of iCAM06 [Kuang et al. 2007a], such a scenario is precluded by the formulation of the model itself, which is always compressive. However, the models of Kim et al. [2009] and Reinhard et al. [2012] are not explicitly limited and therefore could be used in a dynamic range extension context. Because of that, we consider both these models in our experiment.

To take advantage of the best of both tone and color reproduction worlds, Akyuz et al. [2004] and Akyüz and Reinhard [2006] proposed a combination of tone mapping and color appearance modeling, whereby the existing color appearance models were used as an independent preprocessing step for the state-of-the-art tone mapping solutions. In a similar vain, we try to improve color reproduction of the ITMOs we test by combining them with a simple chromatic adaptation transform. This will be discussed in more detail in Section 3.

2.3 Evaluations of ITMOs

Since several tone mapping operators, inverse tone mapping operators, and color appearance models have been proposed in past years, the question of which one is best has been a recurring one, leading to several subjective and objective studies being performed.

As tone compression has attracted more attention in the literature, several studies comparing TMOs exist. Tone-mapped images were evaluated with respect to the reference based on their perceptual similarity in Yoshida et al. [2005] and Ledda et al. [2004] or they were compared among each other based on the observers preferences in Drago et al. [2002], Kuang et al. [2007b], Seetzen et al. [2004], and Ramanarayanan et al. [2007]. Čadík et al. [2008], on the other hand, evaluated tone-mapping operators both with and without reference, while more recently, Wanat et al. [2012] studied the common artifacts and color rendition problems of tone-mapping operators for HDR videos.

At the same time, there are very few experimental studies related to ITMOs. Banterle et al. [2009] conducted a psychophysical experiment where they evaluated several inverse tone-mapping operators for HDR display visualization and image-based lightning. Their investigation mainly focused on determining if more complex models are required to solve the SDR to HDR problem and if there is a correlation between image content and inverse tone mapping quality. Their results show that more complicated nonlinear algorithms perform better than simpler ones for almost all types of image contents they have used.

The opposite conclusion is given in a set of psychovisual experiments conducted by Akyüz et al. [2007] and Masia et al. [2009], who found that simpler, global methods are preferred. Akyüz et al. [2007] evaluated ITMOs based on subjective preferences and visual attributes of the experimental images, such as naturalness, visibility, visual appeal, and spaciousness. It should be noted, however, that in this experiment, only a simple linear scaling, nonlinear scaling with $\gamma < 1$, and nonlinear scaling with $\gamma > 1$ curves were evaluated. Their overall results suggested that linearly, stretching the luminance of the SDR image to the HDR display range was able to give an equal or even better HDR experience than a true HDR image. Masia et al. [2009], on the other hand, evaluated ITMOs' performance for underexposed and overexposed SDR images specifically. Their results showed that if the SDR image is not properly exposed, spatial artifacts created by more sophisticated ITMOs will make the result less preferred than a simple gamma or linear ITMO results.

The previous studies focused on comparing ITMOs based on the overall quality of their results and their capability to recover details in the over- and underexposed regions. To our knowledge, no studies exist that are specifically intended for evaluating color appearance reproduction quality of ITMOs and HDR color appearance models. Given the increasing need for HDR content, methods that can accurately remaster existing content to HDR will likely become necessary. Therefore, to determine whether existing methods are fitting for this purpose, we conduct a psychophysical evaluation of state-of-the-art ITMOs and HDR appearance models, focusing specifically on appearance reproduction criteria.

3. EXPERIMENT OVERVIEW

Although several studies exist that evaluate the comparative performance of ITMOs [Akyüz et al. 2007; Masia et al. 2009; Banterle et al. 2009], as discussed in the previous section, the issue of color reproduction in the context of dynamic range extension has not yet been studied. To that end, the goal of our study is to assess:

—the performance of current HDR color appearance models for an inverse tone-mapping application,
—the ability of existing ITMOs to reproduce the appearance of the original HDR scene, and
—the relative performance of all the previous methods in terms of color reproduction when expanding the dynamic range of an image.

Participants were shown images of the same scene processed in six different ways, as well as the absolute reference HDR image and a corresponding SDR input image of the same scene, and were asked to rate how similar the color appearance of each image was to the reference. The appearance attributes of hue, saturation, colorfulness, and brightness were specifically evaluated. The images shown included the input SDR image; the ground-truth HDR image, which was the same as the reference; the results of four selected ITMOs; and the results from two recent HDR color appearance models.

Specifically, the four ITMOs evaluated in this experiment are the methods by Akyüz et al. [2007], Meylan et al. [2007], Banterle et al. [2006], and Rempel et al. [2007]. These methods were preferred in previous studies and were found to lead to the most robust performances in our experiments with several images. Among the proposed HDR color appearance models, we have chosen the models proposed by Kim et al. [2009] and Reinhard et al. [2012]. These two models, as well as iCAM06 [Kuang et al. 2007a], are designed for HDR image reproduction in the context of dynamic range compression. However, we were not able to use iCAM06 as an inverse tone-mapping operator due to its explicit tone compression step, whereas the Kim et al. and Reinhard et al. models were able to expand the dynamic range of the SDR image given appropriate viewing condition parameters.

Since the ITMOs tested do not have any explicit color processing step, we combined their results with a simple chromatic adaptation transform to match the white point of the target display. Without this step, the differences due to white point mismatches would be too large to assess the behavior of each algorithm. We chose the CIECAM02 chromatic adaptation transform [Fairchild 2013] since it has been used in several current color appearance models, including the model by Kim et al. [2009], which was also evaluated in our study.

4. EXPERIMENTAL STIMULI

As we were interested in studying the quality of the appearance reproduction of different algorithms, the ground-truth images used in our experiment needed to come with corresponding radiometric data. Few databases exist that provide such data for HDR content. Among them, the RIT HDR Photographic Survey [Fairchild 2007] and the sky-centered database of Gryaditskaya et al. [2014] provide calibrated HDR image data with their respective absolute luminance multipliers. The HDR images used as a reference in the experiment are taken from these two databases.

Fig. 1. Experimental SDR images generated from the absolute radiometric HDR images. Images 8 and 9 are taken from the dataset of Gryaditskaya et al. [2014]; the remaining images are from the HDR Photographic Survey [Fairchild 2007].

For this experiment, we have chosen 10 images, representing different types of scenes. As can be seen in Figure 1, our collection of images includes scenes from indoor and outdoor, nighttime and daytime, and closeup views. The images also include dark details, diffuse surfaces, specular highlights, and light sources.

We have transformed the floating-point RGB linear images into CIE 1931 XYZ tristimulus values according to Equation (1) using the standard [IEC 1999], the sRGB characterization matrix M_{srgb}. The resulting XYZ images are then converted to their respective absolute radiometric values by multiplying their luminance channel with the provided luminance multipliers L_m.

Prior to performing the full experiment, we performed a pilot study in which we visualized our experimental images on the SIM2 HDR display both by using their absolute radiometric values directly and by scaling them to different luminance levels. Given the scalings provided in their respective databases, we observed that most of our experimental ground-truth images have an absolute radiometric maximum luminance of greater than 4,000 cd/m^2, whereas for a few of them (Figure 1, numbers 2 and 8), the absolute scaling led to most of the scene being within scotopic luminance ranges, making it impossible to provide any assessments about the color reproduction within the majority of the scene. Therefore, in order to avoid overly dark scenes, we opted for scaling any images that had a peak luminance less than 4,000 cd/m^2 to this value.

Finally, the absolute XYZ radiometric values are converted back to RGB using the inverse of Equation (1).

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = M_{\text{srgb}} \begin{pmatrix} R \\ G \\ B \end{pmatrix}, \tag{1}$$

where

$$M_{\text{srgb}} = \begin{pmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2127 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9504 \end{pmatrix}. \tag{2}$$

As in this experiment we are assessing appearance reproduction, it is necessary to compare processed images against a ground truth. In our case, HDR images serve as ground truth, as they represent the light in real scenes, while SDR images extracted from each HDR serve as input to each algorithm. Our experimental SDR images, shown in Figure 1, are generated from their absolute radiometric HDR images by simulating an exposure bracketing process, using the implementation within the Matlab HDR Toolbox [Banterle et al. 2011], which simulates the typical algorithm implemented in common digital cameras.

Then, the generated SDR images are processed by the selected HDR CAMs and ITMOs. From our preliminary tests, we found that the color appearance difference between the ITMO results and the ground truth was very large in the chromatic direction, as the ITMO results were not adapted to the white point of the display and viewing environment. The two CAMs apply their own dark, light, and chromatic adaptation steps to preserve the same appearance in given different viewing conditions, whereas the ITMOs only expand the dynamic range in spite of the change in viewing conditions. Therefore, conducting the experiment in such conditions would only show the importance of chromatic adaptation transforms, without allowing us to measure contrast and dynamic range reproduction capabilities. For this reason, we have adapted all the stimuli to the white point of the same target display.

To achieve that, the linear RGB results of the ITMOs, the linear RGB SDR image, and the linear RGB HDR ground-truth image are first converted to the XYZ color space, again, based on Equation (1), and the M_{srgb} matrix. Then, the CAT02 chromatic adaptation transform is applied using the target HDR display white point. This transform is also used for the chromatic adaptation step within the Kim et al. [2009] model. Finally, the HDR CAM of Reinhard et al. [2012] models the pupil size, bleaching, and photoreceptor responses in order to simulate the human visual system dark, light, and chromatic adaptations during changes in the viewing environments and conditions.

Finally, all chromatically adapted XYZ results of the ITMOs and the HDR CAMs, together with the adapted XYZ values of the SDR and HDR ground-truth images, are converted to the linear RGB values of the SIM2 HDR47 display, which was characterized as described in Section 4.2. The XYZ to RGB conversion is done according to Equation (1) using the SIM2 HDR47 display characterization matrix M_{sim2} (Equation (3)). Finally, the resulting RGB image values are transformed to the display linear RGBs through the three one-dimensional LUT transformations, which accounts for the nonlinear electro-optical transfer function (EOTF) of the display.

4.1 Algorithm Parameters

To process images using the different ITMOs, we selected parameter values recommended by their respective authors. For the HDR CAMs of Kim et al. and Reinhard et al., the scene parameters were predicted from the input image using the parameter estimation module from the model of Kim et al. [2009]. We chose to use the same prediction module to determine parameters for both CAMs to reduce the number of covarying variables. Since for both models, several assumptions are made for estimating scene parameters from the input image, we found this solution most reliable for our purposes.

The scene white point is computed as the XYZ value of the pixel with the highest luminance value in the input SDR image, and the respective adapting luminance is computed as 20% of the luminance of the predicted white point. For the destination white point, we have measured and used the XYZ values of the SIM2 HDR display white point. Similar to the scene adapting luminance, the destination viewing adapting luminance is computed as 20% of the luminance of the measured display white point. For both HDR CAMs, the surround factor is set to 1, assuming a complete adaptation. The complete list of parameters used for each method is given in Table I and example results for one scene are shown in Figure 7.

4.2 Display Calibration

For our experiment, it was crucial that the behavior of the display used was taken into account in the preparation of the stimuli to ensure accurate color reproduction. To that end, several properties of the display needed to be measured.

Although the SIM2 displays very high peak luminance in theory (more than 4,000cd/m^2), in practice, the highest luminance it can achieve depends on the size of the patch displayed. As such, before

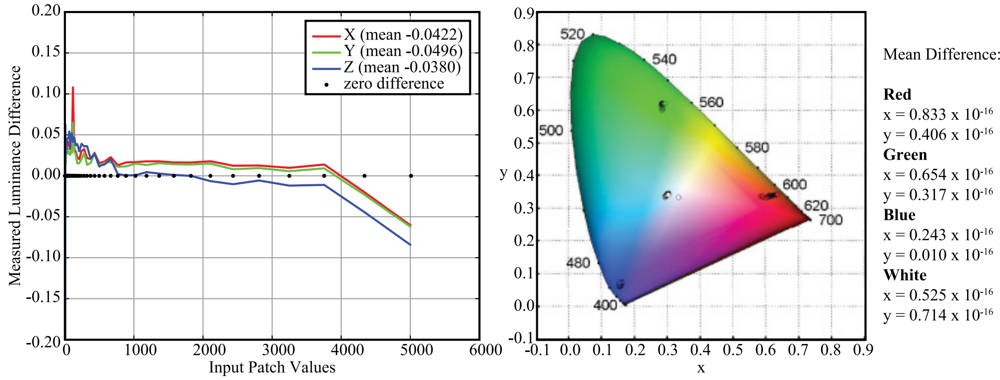
Table I. Used Parameters for the Algorithms Evaluated in the Experiment

Algorithm	Parameter name	Values
Kim et al. [2009] and Reinhard et al. [2012] CAMs	Scene maximum white point	Predicted from input image
	Scene adapting luminance	20% of luminance of white
	Scene surrounding factor F	1
	Viewing maximum white	XYZ = 3531.7 4000 4235.7
	Viewing adapting luminance	800cd/m^2
	Viewing surrounding factor F	1
Meylan et al. [2007] ITMO	Maximum luminance	4000cd/m^2
	λ	0.67
Akyüz et al. [2007] ITMO	Maximum luminance	4000cd/m^2
	Akyuz γ	1
Rempel et al. [2007] ITMO	Maximum luminance	4000cd/m^2
	Noise removal	yes
Banterle et al. [2006] ITMO	Maximum luminance	400cd/m^2
	Expansion operator	Inverse of photographic tone-mapping operator
	Stretching factor	0.68
	Light source clamping threshold	Automatically computed
	Expand map	Computed for each color channel

performing colorimetric characterization, it was necessary to select a patch size that would allow us to measure the full response of this display. To measure the relation between patch size and achievable peak luminance, we prepared and measured the luminance Y, in cd/m^2, of several white patches of different sizes. As can be seen in Figure 3, the display starts to have nonlinear behavior for patches with size of 1,000 $\times$ 1,000 and greater. In contrast, the maximum patch size with which the display has approximately linear behavior is 800 $\times$ 800. This size was used in further characterization that we performed. Since all our experimental images do not contain wide bright regions, such characterization will be sufficiently accurate. Also, the highest luminance Y that still yields an acceptable level of spatial dependence for such patches is found to be 4,000cd/m^2.

To characterize the color reproduction of the display, we have evaluated the channel independence of the display [Neuhaus 2010], that is, how the value of one channel may affect those of other channels for a single pixel. To do that, we computed the percentage difference between the XYZ values of measured gray patches and the sum of XYZ values of measured RGB patches. Then we checked for chromaticity consistency of the display by computing the average mean difference within the chromaticity points of each RGB primary patch and the average RGB patch. Since there was a very minor difference, as shown in Figure 2, it is acceptable to assume that the display channels are independent of each other and that the display is chromatically consistent. Additionally, as demonstrated in Figure 3 for patches of 800 $\times$ 800 pixels, the relationship between the input and output luminance values of the display is indeed linear.

Given these measurements, we can then directly define and use the normalized reference primary (NPM) matrix M_{sim2} (Equation (3)) for defining the relationship between the normalized RGB signals and CIE tristimulus values XYZ. This matrix was obtained following the SMPTE recommendation [SMPTE 2002]. Finally, we have generated one-dimensional lookup tables (LUTs) for each RGB channel, defining the transformation that accounts for the nonlinear EOTF of the display using the



a. Channel Independence

b. Chromaticity Consistency

Fig. 2. Channel independence and chromaticity constancy of SIM2 HDR display. (a) The average differences between the additive combination of RGB primary measurements and the white patch measurements. (b) The chromaticity points of the measured RGB primaries of the SIM2 HDR display within CIE 1931 chromaticity diagram and the average disparity from the mean RGB chromaticity point.

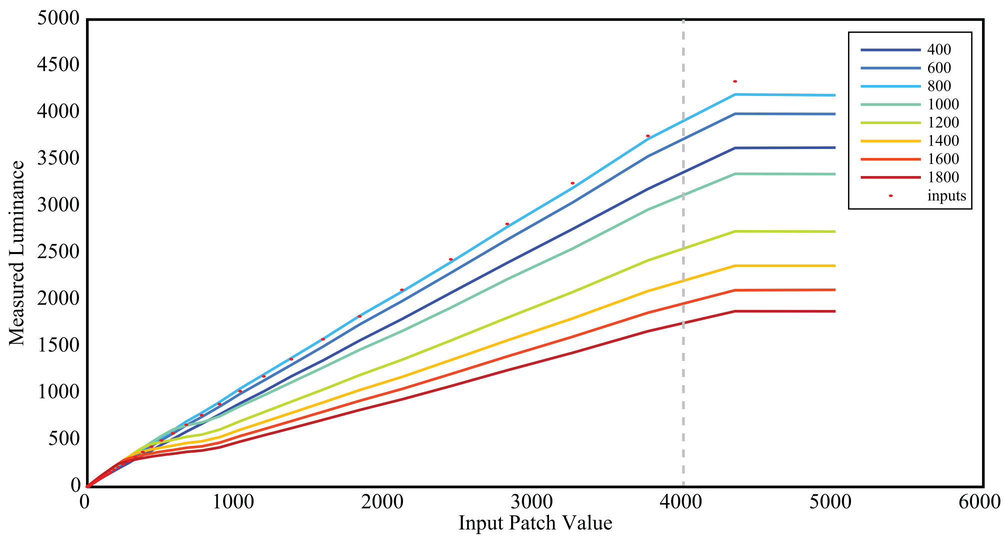


Fig. 3. Testing the power limitation of the display. The different lines in this figure correspond to the display response using patches of different resolutions (detailed in the legend). For each patch size, we have generated several patches with logarithmically increasing intensity values from 0 to 5,011. Finally, the measured luminance values of the patches, for each respective patch size, are plotted.

resulting NPM matrix M_{sim2} (Equation (3)) and spline-based interpolation:

$$M_{\text{sim2}} = \begin{pmatrix} 3.3704 & -1.3948 & -0.5491 \\ -1.0111 & 1.8606 & 0.0304 \\ -0.0410 & 0.0160 & 0.9633 \end{pmatrix}. \tag{3}$$

a. Example Stimulus Screen   b. Experiment GUI

Fig. 4. Experimental setup: (a) example stimulus displayed during the experiment on the SIM2 display; (b) the graphical user interface used for the execution of the experiment, designed according to the SAMVIQ protocol.

5. EXPERIMENTAL SETUP

The experiment was performed in a black room and stimuli were shown on a characterized SIM2 HDR display, discuessed in Section 4.2. The display has a 42' diagonal size, HD resolution, and peak luminance of 4,000cd/m^2. We used a 64-bit Dell Latitude E6530 laptop computer, with 8GB RAM and 2.6GHz processor speed, for displaying the user interface of the experiment, shown in Figure 4(b).

The viewing distance for the observers was set to 1.5m and participants were free to adjust their chair and move their head so that they were approximately centered on the display. The laptop and mouse for navigating through the experiment GUI were placed such that they would not block the view of the displayed stimuli or disturb the observers' focus.

The resolution of the stimuli was limited to a size of 1,286 $\times$ 724 pixels, so that only the central portion of the display was used, to minimize directional viewing issues. The rest of the display was set to gray with 20% of the maximum luminance of the display. An example of the stimulus viewing screen is shown in Figure 4(a).

5.1 Procedure

A total of 21 participants took part in the experiment (7 female and 14 male), with a normal or corrected color vision. The total age range of our participants was 23 to 53.

The experiment followed the subjective evaluation methodology recommended by ITU, known as SAMVIQ (Subjective Assessment Methodology for Video Quality) [ITU-R 2012]. It was supported by an interactive graphical user interface, shown in Figure 4. Using the interface, the participants were able to choose between the eight processed images and give their scores using the provided slider.

For a given scene, we asked the observers to rate each processed image based on their similarity in color appearance with the ground-truth HDR image. We instructed them to particularly focus their assessment on aspects of hue, colorfulness, saturation, and brightness.

Each processed image was scored on a scale from 0 (bad) to 100 (excellent). Participants were allowed to view the reference image whenever they desired and could alter the scores for each alternative within a series, as described by SAMVIQ. Once all scores were entered for a given scene, participants could save their decisions and proceed to the next scene.

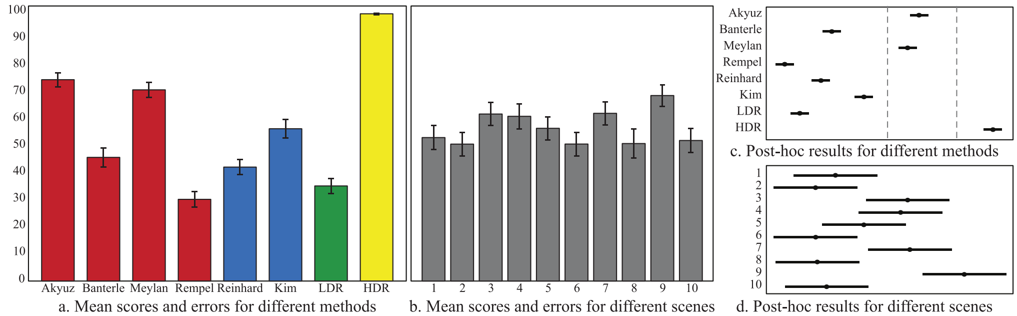


Fig. 5. Mean scores and corresponding errors (95% confidence interval) averaged over all methods (a) and scenes (b). The corresponding results from the post hoc analysis are shown in (c) and (d).

6. RESULTS AND DISCUSSION

In our experiment, we evaluated four ITMO solutions that were combined with a standard CAT to ensure accurate white point mapping, two HDR color appearance models, and the unprocessed SDR image against the HDR ground truth. A set of 10 images was used, leading to a total of 80 trials per participant.

Figure 5(a) shows the mean scores for each method as well as the SDR input and HDR ground truth, over all images. The error bars represent a 95% confidence interval. As would be expected, the HDR was given an almost perfect score. The highest-scoring methods were found to be the approach of Akyüz et al. [2007] and Meylan et al. [2007], with a mean score of 74.97% and 71.20%, respectively.

We further analyzed our results for significance using ANOVA, finding that differences in scores were significant ($F(7,1672) = 274.772$, $p < 0.0001$). Post hoc analysis using the Tukey-Kramer test also supported this conclusion, indicating also that the two global methods of Akyüz et al. [2007] and Meylan et al. [2007] were statistically indistinguishable from each other. Figure 5(c) shows the results from this analysis for visualization.

Interestingly, the two best-performing methods follow a global approach: Akyüz et al. employ a parametric linear expansion, while Meylan et al. opt for a piecewise linear approach that manages diffuse and highlight areas differently. Their performance in our experiment is in accordance with previous studies that found simpler solutions to be preferred [Masia et al. 2009]. The more complex ITMO solutions of Banterle et al. [2006] and Rempel et al. [2007], on the other hand, led to significantly lower scores, which in the case of Rempel et al. were statistically not distinguishable from the SDR input, indicating that more complex solutions may not always be capable of improving the appearance and color reproduction of the input when expanding to a higher dynamic range.

The mean scores of the two HDR appearance models tested were 56.69% for the model of Kim et al. [2009] and 42.43% for Reinhard et al. [2012]. Although these models account for many aspects of the scene and viewing environments and can handle extended dynamic ranges, they are optimized for going from a higher scene luminance to lower viewing luminance levels. As such, they are mostly suited for tone compression, which explains the relatively low scores they obtained. In practice, we found that the HDR appearance models led to accurate color reproduction relative to the HDR ground-truth scene but with loss of local contrast on some occasions.

In our analysis, we also found that the performance of different algorithms is dependent on the scene content. As discussed in the previous section, we selected images representing a varied set of scenes, including indoor, outdoor, and closeup views. Figure 5(b) shows the average scores and corresponding

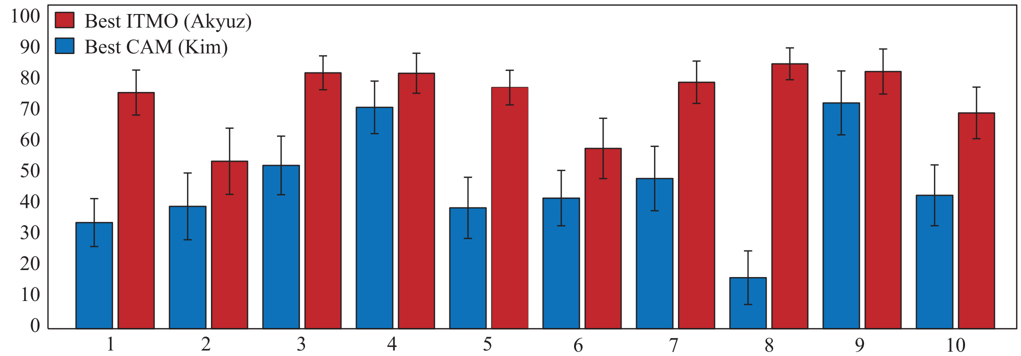


Fig. 6. Mean scores and errors for each image for the best-scoring ITMO and CAM: the best ITMO method based on the experiment scores was the global method of Akyüz et al. [2007], and the best CAM was found to be the model by Kim et al. [2009].

errors for each scene. Differences between different scenes were found to be significant (F(9,1670) = 7.796, $p < 0.0001$).

The SDR versions of each scene are shown in Figure 1 and the results of post hoc analysis on scenes are shown in Figure 5(d). Overall, we found that all methods performed better on outdoor daylight scenes, relative to other scene types, perhaps due the less extreme dynamic range present in these scenes. Additionally, the illuminant white point in these scenes is likely not far from a D65 white point ($X = 95.047, Y = 100.00, Z = 108.883$), and therefore also not far from the SIM2 display white point, simplifying the task of both the CAT and the color appearance models tested.

Overall, although simpler methods and simpler scenes may achieve more accurate reproduction when compared with the HDR ground truth, our results suggest that more work is necessary to ensure accurate management of color and contrast reproduction in the context of dynamic range expansion. To explore this issue further, we look at the best-performing ITMO method [Akyüz et al. 2007] and HDR CAM [Kim et al. 2009] separately for each image. Mean scores and errors for each are shown in Figure 6.

Although in both cases the highest scores are obtained for outdoor daylight images, the Kim et al. model performs best in cases where global contrast and dynamic range are less extreme. Additionally, the ITMO in this case outperforms the HDR CAM on all images tested, indicating that in dynamic range expansion scenarios, accurate reproduction of the contrast and luminance range is much more critical relative to color appearance phenomena.

7. ANALYSIS OF COLOR APPEARANCE ATTRIBUTES

In this section, we briefly evaluate the behaviors of some of the experimental methods when increasing luminance. The basic color appearance attributes of lightness, chroma, and saturation of the two ITMOs that received the higher scores in our subjective test, as well as the two HDR CAMS, were investigated in the IPT color space [Ebner and Fairchild 1998].

The IPT is an opponent color space akin to CIELab with three coordinates, I, P, and T, representing the lightness, red-green, and yellow-blue opponent dimensions, respectively. The model first transforms the input XYZ values, which are normalized to D65, to the LMS cone response space and then compresses each LMS signal by a power of 0.43. This color space is also used as a simple color appearance model due to its ability to predict lightness, chroma, and hue appearance attributes. Although

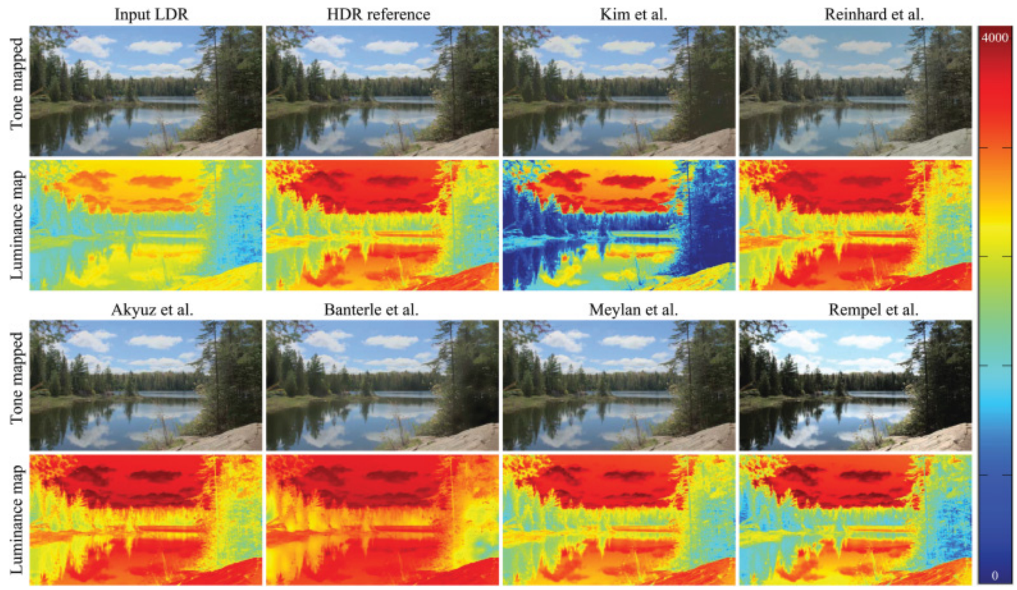


Fig. 7. Example results from each of the tested methods and color maps visualizing the luminance range encoded in each image. Note that results in this figure are tone mapped as their dynamic range is much higher than what could be reproduced on paper.

the more commonly used CIELab space offers similar properties, we opted for the IPT for its accurate and constant prediction of hue without disrupting the prediction of other appearance attributes.

For this comparison, we have generated a 500 × 500 pixel patch with its chromaticity set to $xy = (0.6232, 0.3383)$. The luminance of all the pixels was set to 100cd/m^2. We then processed the patch nine times using the four methods [Akyüz et al. 2007; Meylan et al. 2007; Kim et al. 2009; Reinhard et al. 2012] for nine different maximum viewing luminances (250, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, and 4,000cd/m^2). Finally, the results are transformed into the IPT color space and the appearance attributes of lightness (L), chroma (C), and saturation (S) are computed according to Equations (4), (5), and (6), respectively:

$$L = I, \tag{4}$$

$$C = \sqrt{P^2 + T^2}, \tag{5}$$

$$S = \frac{C}{L}. \tag{6}$$

As can be seen in Figure 8, in all methods, both lightness and chroma increase with the increase in luminance of the viewing condition. These two attributes increase more for the two ITMOs than the HDR CAMs. At the same time, the saturation remains constant in all cases.

Our results also suggest an interesting potential relation between subjective scores (Section 6) and two of the appearance correlates in processed results, namely, lightness and chroma, as can be seen from Figures 8(a) and 8(b) specifically. In agreement with the study of Akyüz et al. [2007], brighter and more colorful images are usually preferred by viewers.

For dynamic range compression, preserving the colorfulness and the saturation of the original HDR scene as much as possible will increase the accuracy of color reproduction [Mantiuk et al. 2009; Pouli et al. 2013]. However, conventional SDR digital cameras do not preserve the color appearance of the

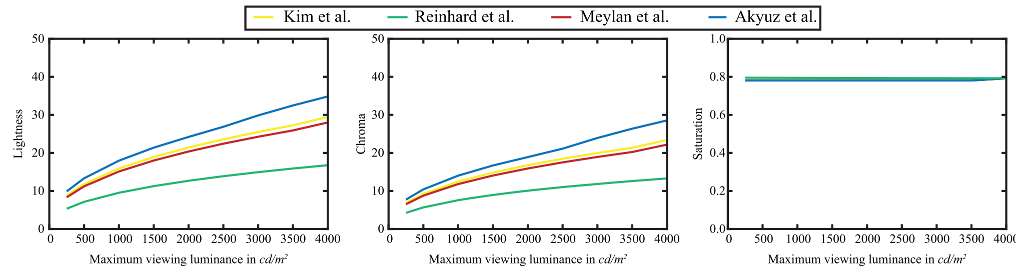


Fig. 8. Basic appearance attributes, computed in IPT color space.

real-world scene, leading to potential impairments in color information. Therefore, when performing dynamic range expansion with the goal of preserving the appearance of the original real-world scene, we should not necessarily retain the colorfulness and saturation of the SDR images. Rather, as shown in our overall results, expanding the dynamic range with concurrent increases in colorfulness and saturation may lead to more preferable results, deemed closer to the ground-truth scene.

8. CONCLUSIONS

As HDR displays are starting to become available to the consumer market, the need for solutions that can reliably expand the dynamic range of existing content while preserving the director's intent is becoming apparent. Although many ITMOs have been proposed in the literature for dynamic range expansion, their ability to accurately reproduce the appearance of processed images has not been thoroughly studied. At the same time, HDR color appearance models have been evaluated mostly in the context of tone compression. As such, it is not clear whether existing methods are sufficient for this purpose.

To answer this question, we evaluated several ITMOs and HDR appearance models assessing their ability to reproduce the appearance of an original HDR scene when expanding the luminance and contrast range of an SDR image of the same scene. In contrast to previous studies, our focus is on appearance reproduction rather than preference. To evaluate this aspect, we followed the ITU-recommended protocol and compared each processed image against the HDR ground truth on a calibrated HDR monitor.

Our results indicate that contrast and luminance reproduction play a more important role than color reproduction. Additionally, we find that simpler, global methods lead to better results when compared to the HDR ground truth. At the same time, we find that the performance of all methods is dependent on image content, suggesting that more work is necessary to allow for accurate appearance reproduction in the context of dynamic range expansion.

REFERENCES

Ahmet Oğuz Akyüz, Roland Fleming, Bernhard E. Riecke, Erik Reinhard, and Heinrich H. Bülthoff. 2007. Do HDR displays support LDR content?: A psychophysical evaluation. *ACM Transactions on Graphics* 26, 3, Article 38 (July 2007). DOI:http://dx.doi.org/10.1145/1276377.1276425

Ahmet Oğuz Akyüz and Erik Reinhard. 2006. Color appearance in high-dynamic-range imaging. *SPIE Journal of Electronic Imaging* 15, 3 (2006), 033001–1–033001–12.

Ahmet Oguz Akyuz, Erik Reinhard, and Sumanta Pattanaik. 2004. Color appearance models and dynamic range reduction. In *1st ACM Symposium on Applied Perception in Graphics and Visualization (APGV'04)*. 166.

Francesco Banterle, Alessandro Artusi, Kurt Debattista, and Alan Chalmers. 2011. *Advanced High Dynamic Range Imaging: Theory and Practice*. AK Peters (CRC Press), Natick, MA.

Francesco Banterle, Patrick Ledda, Kurt Debattista, Alessandro Artusi, Marina Bloj, and Alan Chalmers. 2009. A psychophysical evaluation of inverse tone mapping techniques. *Computer Graphics Forum* 28, 1 (2009), 13–25.

Francesco Banterle, Patrick Ledda, Kurt Debattista, and Alan Chalmers. 2006. Inverse tone mapping. In *Proceedings of the 4th International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia (GRAPHITE'06)*. 349–356.

Francesco Banterle, Patrick Ledda, Kurt Debattista, Alan Chalmers, and Marina Bloj. 2007. A framework for inverse tone mapping. *Visual Computer* 23, 7 (2007), 467–478. DOI:http://dx.doi.org/10.1007/s00371-007-0124-9

Martin Čadík, Michael Wimmer, Laszlo Neumann, and Alessandro Artusi. 2008. Evaluation of HDR tone mapping methods using essential perceptual attributes. *Computers & Graphics* 32, 3 (2008), 330–349.

CIETC:801. 2000. Specification of Colour Appearance for Reflective Media and Self-Luminous Display Comparisons. Retrieved from http://www.cie.co.at/index.php?i_ca_id=806.

Scott J. Daly and Xiaofan Feng. 2003. Bit-depth extension using spatiotemporal microdither based on models of the equivalent input noise of the visual system. In *Electronic Imaging 2003*. International Society for Optics and Photonics. 455–466.

Scott J. Daly and Xiaofan Feng. 2004. Decontouring: Prevention and removal of false contour artifacts. In *Proceedings of SPIE*, Vol. 5292. 130–149.

Piotr Didyk, Rafał Mantiuk, Matthias Hein, and Hans-Peter Seidel. 2008. Enhancement of bright video features for HDR displays. *Computer Graphics Forum (Proceedings of Eurographics Symposium on Rendering 2008, Sarajevo, Bosnia, and Herzegovina)* 27, 4 (2008), 1265–1274.

Frédéric Drago, William L. Martens, Karol Myszkowski, and Hans-Peter Seidel. 2002. *Perceptual Evaluation of Tone Mapping Operators with Regard to Similarity and Preference*. Technical Report MPI-I-2002-4-002. Max Plank Institut für Informatik.

Fritz Ebner and Mark D. Fairchild. 1998. Development and testing of a color space (IPT) with improved hue uniformity. In *Proceedings of the 6th Color Imaging Conference: Color Science, Systems and Applications*. 8–13.

M. D. Fairchild. 2007. The HDR photographic survey. In *Proceedings of the 15th Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, Vol. 15. Society for Imaging Science and Technology, 233–238.

Mark D. Fairchild. 2013. *Color Appearance Models*. John Wiley & Sons.

Mark D. Fairchild and G. M. Johnson. 2002. Meet iCAM: An image color appearance model. In *Proceedings of the 10th IS&T/SID Color Imaging Conference*. 33–38.

M. D. Fairchild and G. M. Johnson. 2004. The iCAM framework for image appearance, image differences, and image quality. *Journal of Electronic Imaging* 13 (January 2004), 126–138. DOI:http://dx.doi.org/10.1117/1.1635368

Yulia Gryaditskaya, Tania Pouli, Erik Reinhard, and Hans-Peter Seidel. 2014. Sky based light metering for high dynamic range images. *Computer Graphics Forum (Proceedings of Pacific Graphics)* 33, 7 (2014), 61–69. DOI:http://dx.doi.org/10.1111/cgf.12474

Yongqing Huo, Fan Yang, and Vincent Brost. 2014. Inverse tone mapping based upon retina response. *Scientific World Journal* 2014, 168564 (2014), 5. DOI:http://dx.doi.org/10.1155/2014/168564

IEC. 1999. IEC 61966-2-1:1999: Multimedia systems and equipment - Colour measurement and management - Part 2-1: Colour management - Default RGB colour space - sRGB. Retrieved from https://webstore.iec.ch/publication/6169.

ITU-R. 2012. Recommendation ITU-R BT.500-13: Methodology for the subjective assessment of the quality of television pictures. Retrieved from http://www.itu.int/rec/R-REC-BT.500-13-201201-I.

Mark D. Fairchild and Jiangtao Kuang. 2007. iCAM06, HDR, and Image Appearance. Retrieve from http://www.rit-mcsl.org/StudentResearch/CIC15_iCAM06.pdf.

Min H. Kim, Tim Weyrich, and Jan Kautz. 2009. Modeling human color perception under extended luminance levels. *ACM Transactions on Graphics (Proceedings SIGGRAPH'09)* 28, 3 (2009), 27:1–9.

Rafael Pacheco Kovaleski and Manuel M. Oliveira. 2009. High-quality brightness enhancement functions for real-time reverse tone mapping. *Visual Computer* 25, 5–7 (2009), 539–547.

Jiangtao Kuang, Garrett M. Johnson, and Mark D. Fairchild. 2007a. iCAM06: A refined image appearance model for HDR image rendering. *Journal of Visual Communication and Image Representation* 18, 5 (2007), 406–414. Special issue on High Dynamic Range Imaging.

Jiangtao Kuang, Hiroshi Yamaguchi, Changmeng Liu, Garrett M. Johnson, and Mark D. Fairchild. 2007b. Evaluating HDR rendering algorithms. *ACM Transactions on Applied Perception* 4, 2 (2007).

Hayden Landis. 2002. Production-ready global illumination. In *Siggraph Course Notes 16*.

Patrick Ledda, Alan Chalmers, and Helge Seetzen. 2004. A psychological validation of tonemapping operators using a high dynamic range display. In *Proceedings of the 1st ACM Symposium on Applied Perception in Graphics and Visualization (APGV'04)*. 159.

Radosław Mantiuk, Rafał Mantiuk, Anna Tomaszewska, and Wolfgang Heidrich. 2009. Color correction for tone mapping. *Computer Graphics Forum (Proceedings of EUROGRAPHICS'09)* 28, 2 (2009), 193–202.

Belen Masia, Sandra Agustin, Roland W. Fleming, Olga Sorkine, and Diego Gutierrez. 2009. Evaluation of reverse tone mapping through varying exposure conditions. *ACM Transactions on Graphics* 28, 5 (2009), 1–8. DOI:http://dx.doi.org/10.1145/1618452.1618506

Belen Masia and Diego Gutierrez. 2011. *Multilinear Regression for Gamma Expansion of Overexposed Content*. Technical Report. Universidad de Zaragoza (Technical Report RR-03-11).

Laurence Meylan, Scott Daly, and Sabine Süsstrunk. 2007. Tone mapping for high dynamic range displays. In *Proceedings of IS&T/SPIE Electronic Imaging: Human Vision and Electronic Imaging XII*, Vol. 6492.

L. Meylan and S. Susstrunk. 2006. High dynamic range image rendering with a retinex-based adaptive filter. *IEEE Transactions on Image Processing* 15, 9 (2006), 2820–2830.

Chantal Neuhaus. 2010. *Colorimetric Characterization of a High-Dynamic Range Display*. Ph.D. Dissertation.

Tania Pouli, Alessandro Artusi, Francesco Banterle, Ahmet Oğuz Akyüz, Hans-Peter Seidel, and Erik Reinhard. 2013. Color correction for tone reproduction. In *Proceedings of the Color and Imaging Conference*, Vol. 2013. Society for Imaging Science and Technology, 215–220.

G. Ramanarayanan, J. A. Ferwerda, B. Walter, and K. Bala. 2007. Visual equivalence: Towards a new standard for image fidelity. *ACM Transactions on Graphics* 26, 3 (2007), 76.

Erik Reinhard, Erum Arif Khan, Ahmet Oğuz Akyüz, and Garrett M. Johnson. 2008. *Color Imaging: Fundamentals and Applications*. AK Peters, Wellesley.

E. Reinhard, T. Pouli, T. Kunkel, B. Long, A. Ballestad, and G. Damberg. 2012. Calibrated image appearance reproduction. *ACM Transactions on Graphics* 31, 6 (2012), 201.

Allan G. Rempel, Matthew Trentacoste, Helge Seetzen, H. David Young, Wolfgang Heidrich, Lorne Whitehead, and Greg Ward. 2007. Ldr2Hdr: On-the-fly reverse tone mapping of legacy video and photographs. *ACM Transactions on Graphics* 26, 3 (2007), 39. DOI:http://dx.doi.org/10.1145/1276377.1276426

Helge Seetzen, Wolfgang Heidrich, Wolfgang Stuerzlinger, Greg Ward, Lorne Whitehead, Matthew Trentacoste, Abhijeet Ghosh, and Andrejs Vorozcovs. 2004. High dynamic range display systems. *ACM Transactions on Graphics* 23, 3 (August 2004), 760–768. DOI:http://dx.doi.org/10.1145/1015706.1015797

Gaurav Sharma. 2002. *Digital Color Imaging Handbook*. CRC Press, Boca Raton, FL, 171 pages.

SMPTE. 2002. RP 177-1993: Derivation of Basic Television Color Equations. (2002).

Robert Wanat, Josselin Petit, and Rafal Mantiuk. 2012. Physical and perceptual limitations of a projector-based high dynamic range display. In *TPCG*. 9–16.

Lvdi Wang, Li-Yi Wei, Kun Zhou, Baining Guo, and Heung-Yeung Shum. 2007. High dynamic range image hallucination. In *EGSR*. Eurographics Association, 321–326.

Akiko Yoshida, Volker Blanz, Karol Myszkowski, and Hans Peter Seidel. 2005. Perceptual evaluation of tone mapping operators with real-world scenes. In *Human Vision & Electronic Imaging X*. SPIE, 192–203.

Xuemei Zhang, B. A. Wandell, D. A. Silverstein, and J. E. Farrell. 1997. Color image quality metric S-CIELAB and its application on halftone texture visibility. In *IEEE COMPCON Symposium Digest*. 44–48.

Received July 2015; revised July 2015; accepted July 2015