Multimedia Tools and Applications
https://doi.org/10.1007/s11042-021-11182-7

An experimental evaluation of visual similarity for HDR images

Merve Aydinlilar[1] · **Ahmet Oguz Akyuz**[1] · **Sibel Tari**[1]

Received: 20 August 2020 / Revised: 9 January 2021 / Accepted: 24 June 2021 /
Published online: 31 July 2021

Abstract
In this paper, we investigate visual similarity for high dynamic range (HDR) images. We collect crowdsourcing data through a web-based experimental interface, in which the participants are asked to choose one of the two candidate images as being more similar to the query image. Triplets forming the query-and-candidates sets are obtained by random sampling from existing HDR data sets. Experimental control factors include choice of tone mapping operator (TMO), choice of distance metric, and choice of image feature. The image features that we experiment with are chosen from the features that are commonly used in the usual low dynamic range setting including features learned via Convolutional Neural Networks. The set of image features also includes combined features where the combination coefficients are estimated using logistic regression. We compute correlations between human judgments and quantitative features to understand how much each feature contributes to visual similarity. Combined features yield nearly 84% agreement with human judgments when applied on tone mapped images. Though we observed that using common features directly on raw or linearly scaled HDR images yield subpar correlation estimates compared to using them on tone mapped HDR images, we did not observe significant effect due to the choice of TMO on the estimates. As an application, we propose an improvement to style-based tone mapping for more correctly imparting desired styles to HDR images with different characteristics.

Keywords HDR imaging · Visual similarity · Tone mapping

✉ Ahmet Oguz Akyuz
akyuz@ceng.metu.edu.tr

Merve Aydinlilar
merve@ceng.metu.edu.tr

Sibel Tari
sibel@ceng.metu.edu.tr

[1] Middle East Technical University, Department of Computer Engineering, Ankara, Turkey

1 Introduction

Assessing visual similarity of images is an important task for various applications including image retrieval and indexing [36], classification and clustering [30], image editing and style transfer [53]. Due to its importance, significant amount of research is dedicated to measuring image similarity. Because visual similarity is a perceptual phenomenon without ground truth, several human experiments are conducted as well. All these works, however, assume that the image is given in standard low dynamic range format where the brightness information is suitably quantized to match the dynamic range of traditional image display devices. Yet, growing number of applications utilize High Dynamic Range (HDR) images with unbounded brightness values. It may be argued that HDR images pose no extra challenges and approaches designed for LDR images may directly be used to assess visual similarity of HDR images as well. There are several counter-arguments, however. First, HDR images contain potentially uncalibrated floating point data and two images that have vastly different pixel values may actually be very similar to each other. Second, the richness of information in an HDR image, despite causing difficulties, may aid in similarity assessment. For example, pixel values corresponding to a bright light source can be much higher than that of a white reflecting surface in an HDR image, while the two objects are likely to map to similar intensities in an LDR image. Third using a standard similarity measure for an HDR image requires tone mapping, a problem for which a multitude of algorithms, each with a number of parameters, exist [76].

Hence, there is a need for investigating visual similarity for HDR images. This need is the motivation of the present work where we experimentally investigate assessing visual similarity between two images. To this end, we collect subjective human judgments using crowdsourcing and evaluate features by comparing them to human judgments. Our data collection via crowdsourcing is performed in two stages. In the first stage, 100 HDR images from several HDR image databases are used in a pairwise assessment task. Due to a large number of combinations, this phase primarily serves as an unbiased exploration of a large search space. In the second stage, we focus on the already tested images from the first phase to collect multiple responses for each test case.

Our experimental control factors include choice of tone mapping operator, choice of distance metric, and choice of image feature. We use commonly used low-level image features such as color, luminance, and texture histograms; advanced features such as GIST [50]; deeply learned features [66]; and combined features estimated using logistic regression. To our knowledge our work serves as the first rigorous attempt to evaluate how visual similarity can be assessed between HDR images. Using our findings, we propose a tone mapping methodology where tone mapping parameters are automatically computed to impart a certain user-defined style to a given HDR image using the similarity between this image and several calibration images that are used to create this style.

In the following, we first review the related work on image similarity and HDR imaging (Section 2). We then introduce our experiments (Section 3) followed by a description of the relevant features and metrics (Section 4). We then share our results (Section 5) and describe an application that benefits from our findings (Section 6). Finally, we conclude our paper by reiterating the key findings, drawing out its limitations, and outlining several future research directions (Section 7).

2 Related work

2.1 Image similarity

Traditionally, image similarity is measured by measuring the distance between hand crafted features extracted from each image. These hand crafted features include simple descriptors such as color/luminance histograms, or improved ideas, including histogram of oriented gradients [8], GIST [50], SIFT [37], SURF [3]. These features are compared using several types of distance metrics. Recently, deep convolutional neural networks (DCNNs) became the state of art for image classification. Starting with AlexNet [32] and followed by deeper networks such as VGG [66], GoogLeNet[68], and ResNet [27], DCNNs started to perform near human level success for image classification. Their success lead to use feature vectors that have been obtained from DCNNs for image retrieval [21, 73]. Unlike previous approaches that are based on hand-crafted features, DCNNs learn the feature vector itself directly from the image. One major drawback of using DCNNs is the need for using very large labeled datasets for training, which is difficult to obtain or not available at all for most problem domains. Transfer learning [77] aims to solve this problem by using pretrained networks on large scale datasets such as ImageNet [60]. The basic method is to give the images to the pre-trained network and use the output of the last fully connected layers as feature vectors [10, 73] – an approach that we also adopt in our work.

Visual similarity is a perceptual phenomenon without ground-truth data. This makes collecting data using crowdsourcing experiments valuable. Indeed, there are several crowdsourcing-based works [30, 38, 62] that address shape or style similarity problems and conduct user experiments to either derive or validate models. Of most related to our work are two similarity studies that also employ subjective experiments [48, 58].

In the first study, human participants are asked to judge image similarity using two different experiments: one involving printouts of images (called table scaling) and the other using a computer based comparison (called computer scaling) [58]. These results are compared with computational similarity approaches [18] and simple CIELAB histograms. It was found that both table and computer scaling yield similar results and color is a major factor influencing similarity for human observers. In the second study, user experiments are conducted to evaluate the relationship between an image-indexing system and perceived similarity in an LDR setting [48]. The tested image indexing system is based on basic properties of early stages of human vision – chromaticity, luminance, and texture. Two-alternative forced-choice (2AFC) method is used for all experiments. Three images are shown to the observer, the query image and two test images. Of these two images one image is called the target and the other the distractor. These images are selected based on the rankings obtained from the image-indexing system. Then the correlation between the users' preference and index rank is investigated. First, each index, chromaticity, luminance, and texture are calculated separately. From these indexes chromaticity is found to give the best results. Then for the second experiment, combinations of the indexes are evaluated. The combination of chromaticity and texture indices are found to give better results than chromaticity alone and the combination of all indices are found to give the best result.

As reviewed in this section, although visual image similarity is an extensively studied subject [36], to our knowledge there is no study that directly addresses this problem for HDR images.

2.2 HDR imaging

The need for HDR imaging was realized for the first time in computer graphics to deal with the requirements of physically accurate lighting simulation systems [20]. Such systems produced numerically unbounded pixel values, necessitating their storage in HDR file formats [35]. HDR images have typically been termed as "scene-referred" as opposed to "display-referred" – a term used for LDR images [56]. However, as display devices have traditionally been low dynamic range, displaying these images on LDR devices required an operation known as tone mapping [71, 74]. Numerous tone mapping operators (TMOs) have been developed in literature ranging from simple contrast adjustments to complex algorithms modeling the human visual system [17] and the properties of display devices [41]. Many methods have also been produced to create photographic HDR images of real-world scenes [9], including dynamic scenes [29, 64]. Besides computer graphics, HDR imaging has many application areas including studying of fossils [69], cultural heritage and archaeology [25], structural engineering [23], architecture [6], medical imaging [26, 57], forensics [5], and automotive industry [75].

While HDR imaging has long been an active field of research, recent developments in HDR imaging [2, 7, 56], in particular those pertaining to HDR image and video capture [19, 70], display systems [63], and HDR video streaming standards [61] are likely to make HDR content more ubiquitous in the near future. However, despite the practical improvements in the field, there is also a need for fundamental and experimental research that explores various aspects related to HDR imaging and dynamic range. Hanhart et al. investigated the performance of various objective metrics in quantifying visual distortions of HDR images commensurate with subjective opinions [24]. The authors found HDR-VDP-2 [43] and HDR-VQM [46] to be the best predictors of visual quality. In another study, Grimaldi et al. investigated how image statistics change as a function of dynamic range and found that there are indeed differences between HDR and LDR images [22]. The authors, also found, however, that the majority of these differences are accounted for by the early visual processing that takes place in the human visual system. However, these works do not consider the HDR image similarity problem.

Given the lack of visual similarity studies on HDR images, understanding the nature of image similarity for HDR images and developing an objective similarity measure is the primary goal of this paper. Secondly, we show how such a metric could be leveraged to solve an important tone mapping problem, which is how to tone map different HDR images such that they consistently follow a user-defined style.

3 Experiments

3.1 Experimental design

To measure perceptual similarity between HDR images, we conducted a 2AFC experiment. The experiment is publicly available[1]. As we needed a large number of responses, we designed a web-based interface to collect crowdsourcing data. We used the HDRHTML technique [42] for visualizing HDR images on web browsers. This technique uses a windowing approach to select a desired exposure range from the HDR image, which is indicated

[1] https://user.ceng.metu.edu.tr/~merve/userstudy/

by a slider set by the user. By dynamically adjusting the position of the slider, the user can efficiently view the entire exposure range contained within the HDR image. These sliders are normally overlaid with the image histogram. We removed this overlay to prevent the image histogram from affecting the observers' decisions. Figure 1 shows a sample trial from the experiment. An HDR reference image was shown at the top and two HDR test images were shown at the bottom. The sliders, which were mandatory to be adjusted, allowed all images to be inspected at different exposure levels.

In each experimental session, 33 such image triplets were displayed to the observers. Thus, an experimental session consisted of 33 trials. In each trial, the observers were asked to choose which of the two test images was visually more similar to the reference image. Here it is important to note that we did not ask users to decide for a specific type of similarity such as object, color, etc. By intentionally leaving the definition of visual similarity vague, we hoped to achieve a range of responses, which in overall, would converge to a common sense understanding for similarity. All trials, except for the verification ones, were generated randomly from the dataset during the runtime of the experiment. Three of these triplets were used for verification. They contained an obviously similar reference and test image pair to evaluate the reliability of an observer. If an observer failed to provide the correct answer even for one of these trials, his or her data was discarded as being unreliable. These trials were distributed evenly across the experiment to ensure that observers were attentive throughout. Before the experiment began, observers were informed about their task and the expected duration of the experiment, which was at most 20 minutes at a normal pace. During

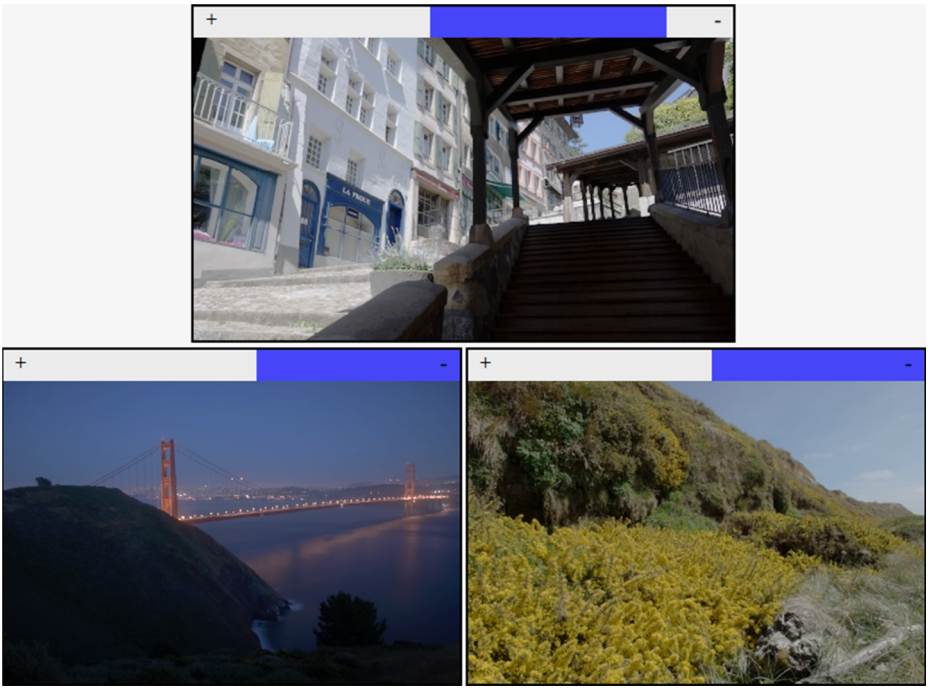


Fig. 1 A sample trial from the experiment. The observers were asked to choose the most similar image to the reference image (top) from the test images (bottom). All images could be examined at different exposure levels by adjusting their sliders

the experiment, observers were required to use the exposure sliders for each image before they made a selection. Image selection was done by clicking on one of the test images. The selection was indicated using a green border around the selected image. Observers could change their selection until they pressed the "Next" button. The progress of an observer was indicated using a small progress bar at the bottom center of the screen. At the end of the experiment, observers were informed with a final page confirming the conclusion of the experiment and were presented with unique session ids. They were required to enter this id to the crowdsourcing platform to verify that they have finished the experiment.

3.2 Dataset

The set of images used in visual similarity experiments should be sufficiently diverse. Although such datasets exists for LDR images, and several HDR image *quality* datasets exist for HDR images [33, 34], there is no specific visual similarity dataset for HDR images. We therefore decided to select 100 HDR images from various databases to present observers with a diverse set of images[2]. The used datasets were: Fairchild's HDR Photographic Survey [14], HDR-Eye [47], DEIMOS [31], Empa HDR Image Database [13], and pfstools HDR Image Gallery [40]. Thumbnails for the used images are shown in Fig. 2.

3.3 Crowdsourcing

In order to reach as many people as possible, the experiment was published at Microworkers crowdsourcing platform[3]. The number of paid users that participated in the experiment through this platform was 801. For each completed experiment 0.3$ were paid. Among these, 165 sessions were discarded due to incorrect responses given to the verification trials. Age, gender, and familiarity with computer graphics/image processing distribution of the participants are shown in Fig. 3.

After collecting the experimental results, it was found that 18747 unique image triplets were judged by the observers. This amounts to approximately 11.6% of the total possible triplets that can be obtained from 100 images, $C(100, 3)$. Experiment sessions were independent and random for each participant, but it was guaranteed that a single session consisted of only unique triplets.

This design resulted in a single response for the majority of the triplets. Some triplets received two responses and only a few received three or more. As such we considered this first phase of the experiment as a random exploration of all possible comparisons. However, as judging similarity based on a single response could be too subjective, we extended the experiment as discussed below to collect multiple responses for each triplet.

3.4 Extension to the experiment

The first phase of the experiment was extended to obtain three evaluations per triplet. Unlike the first phase where triplets were generated randomly, the second phase solely used the triplets that had been evaluated before. To achieve this, we sorted the triplets from the first phase in descending order. If a triplet had more than three responses, we randomly

[2]We unfortunately discovered after the experiments were conducted that one image was duplicated under different names. See the images in 2nd row-4th column and 9th row-3rd column in Fig. 2. In our analysis, we discarded the few trials in which this image was duplicated.

[3]www.microworkers.com

Fig. 2 HDR images used in the visual similarity experiments

selected three of them. Triplets with exactly three responses were used as is. These two cases occurred very rarely. Next, triplets with two responses, and then a single response were presented randomly to obtain a total of 4890 triplets that had been evaluated three times.

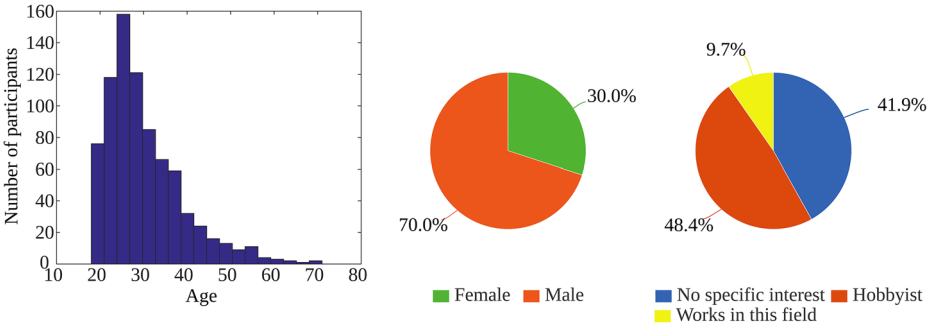


Fig. 3 Age, gender, and computer graphics/image processing familiarity distribution of the participants

Table 1 HDR Image features and distances

Feature	Representation	Distance Metric
Color	2D chromaticity histogram	Earth Mover's Distance (EMD)
Luminance	1D (relative) luminance histogram	EMD
Texture	Histogram of gradients	EMD
GIST	Feature vector	Cosine distance
VGG16/VGG19 - fc6	Fused fc6 layer	Cosine distance
VGG16/VGG19 - fc7	Fused fc7 layer	Cosine distance

Among these thrice evaluated triplets, 2170 triplets were judged consistently by all three observers. The remaining 2820 triplets generated two-to-one responses. Similar to the first part of the experiment, the second part also contained the same validity checks to eliminate the responses of inattentive observers.

4 Features & metrics

4.1 Features

In this study, five kinds of features are used to model HDR images: color, luminance, texture, GIST, and DCNN features. Table 1 lists these features together with their representations and the distance metric used for each feature. The following sections outline the details of these features and the corresponding distance metrics.

4.1.1 Color

Since the early days of the image similarity research, color has been used as one of the most discriminative cues [48]. In this study, we used the a and b channels of the CIELAB color space [28] to represent chromaticity information. This is an opponent color space, where the a channel represents red/green opponent colors and the b channel yellow/blue opponent colors. We used a 2D chromaticity histogram to represent the distribution of colors in a given image. Each dimension contained 15 bins for a total of 225 bins. Figure 4 shows this histogram for the Mason Lake image from the dataset [14].

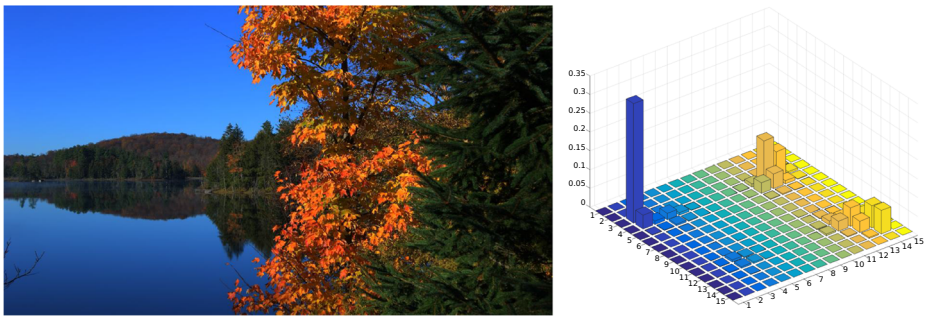


Fig. 4 Sample image (left) and the corresponding 2D ab histogram (right)

4.1.2 Texture

Texture is the second most used feature for content based image retrieval systems after chromatic features. This feature is especially helpful for discriminating images that have similar color but different spatial characteristics such as blue sky and sea or sand and buildings. To represent the texture information we used histogram of gradient magnitudes [65].

4.1.3 Luminance

The main difference between an HDR and an LDR image is the much wider range of luminance distribution for the former. A single HDR image may contain very low luminances corresponding to highly shadowed regions as well as very high luminances corresponding to bright highlights. Therefore, we hypothesized that the luminance distribution of an HDR image may be an important cue for visual similarity. The luminance distribution is modeled using a 1D (relative) luminance histogram with 50 bins.

4.1.4 GIST descriptor

The GIST descriptor [50] aims to represent the dominant spatial structure of a scene by using low level multi-scale representations. This descriptor defines the scene as a whole rather than focusing on individual objects or regions. Discriminative properties of a scene are listed as naturalness, openness, roughness, expansion, and ruggedness. The class of a scene, e.g., man-made, natural, indoor, outdoor, etc., is determined by these properties.

The procedure for extracting GIST descriptors consists of applying Gabor filters that are scaled and oriented differently to the input image, dividing the filter response map into a grid in order to have spatial information, averaging the filter response in each grid, and concatenating the results to obtain the final feature vector, i.e. the GIST descriptor.

4.1.5 Deep learning features

Recently, DCNNs have started to dominate object recognition and image classification tasks, achieving near human success rates [32, 66, 79]. These models are trained with large prelabeled datasets and develop a hierarchical model that becomes more aware of the content of the image rather than the underlying pixel values. To our knowledge currently there is no DCNN model that is trained on HDR images for the purpose of image indexing, scene classification, or visual similarity tasks. Furthermore, there is no prelabeled large HDR image dataset to use for training a DCNN model from scratch. Therefore in this study, we used transfer learning method to employ pretrained DCNNs for our perceptual similarity problem.

For feature extraction, pretrained AlexNet [32] and two variants of VGG networks, VGG16 and VGG19, are used [66]. All networks are trained on the ImageNet [60] dataset, but we also evaluated their performance when trained using different datasets. For transfer learning, the last fully connected layer, which contains classification outputs, is removed and the remaining 4096 dimensional two fully connected layers, **fc6** and **fc7**, are used as feature vectors. As suggested by Simonyan and Zisserman [67], the results obtained from VGG16 and VGG19 are fused (by taking an average) and it is observed that the fused version performs better than both VGG16 and VGG19. The distance between the feature vectors are calculated using cosine distance, which is a commonly used distance metric for deep learning features.

4.1.6 Computational analysis

As attentive readers may inquire about the computational complexity of these features, a brief analysis is provided in this section. The color and luminance features are merely histograms and they can be computed in $O(N)$ time, where N represents the number of image pixels. The GIST descriptor is based on computing the convolution of 32 Gabor filters at 4 scales and 8 orientations to compute feature maps and then averaging the features to obtain a 16 element vector per feature map. These vectors are then concatenated to find the final feature vector. Therefore, the GIST feature also has a linear time computational complexity. The texture feature represented by HOG is similar to GIST in terms of its operation, albeit being somewhat simpler, and also has a linear complexity. Finally, the deeply learned features are computed by a single forward application of the VGG network, which also involves several convolutions at multiple scales as well as application of activation functions. As the convolutional kernel sizes are negligible compared to the image size, the computation of VGG features also has a linear time computational complexity. In summary, all of the features can be computed at real-time rates in modern hardware, especially if GPUs are utilized.

4.2 Dissimilarity measures

The use of a proper distance metric is as important as the features themselves. Each feature representation may require a different distance metric. In this section, we briefly describe the definitions and properties of the dissimilarity measures that we used for different types of features.

4.2.1 Euclidean distance

The Euclidean distance between two histograms p and q is calculated as:

$$d_{euc}(p,q) = \sqrt{\sum_i (p_i - q_i)^2}, \tag{1}$$

where i is the bin index. In general, dissimilarity obtained by Euclidean distance for histograms is not satisfactory as it does not take bin proximity into account.

4.2.2 Bhattacharyya distance

Bhattacharyya distance [4] measures the overlap between two distributions. If p and q are two histograms, it can be calculated as:

$$d_{bhat}(p,q) = -\ln\left(\sum_i \sqrt{p_i.q_i}\right). \tag{2}$$

For our HDR similarity problem Bhattacharyya distance gives slightly better results than Euclidean distance. However, it also suffers from the same problem that the proximity of the bins is not taken into account.

4.2.3 Earth mover's distance

Earth Mover's Distance (EMD) is a dissimilarity metric commonly used for image the retrieval problems [59]. EMD aims to capture the perceptual similarity between two distributions by calculating the minimal cost of transforming one distribution to the other. Unlike the other dissimilarity metrics, EMD can be calculated for varying-size partitions of the data, called signatures. Signatures consist of dominant clusters of the data, represented as $s_i = (m_i, w_i)$ pairs where m_i is the cluster center and w_i is the size of the cluster. EMD does not require the signatures to have the same number of clusters – ground distances between cluster centers are sufficient. Histograms are signatures with bin centers corresponding to cluster centers, m_i, and normalized bin values to weights, w_i.

The total amount of work to transform distribution p to q with flow f is:

$$WORK(P, Q, F) = \sum_{i}^{m}\sum_{j}^{n} d_{ij} f_{ij}, \tag{3}$$

where d_{ij} is the ground distance between cluster centers i and j. The optimal flow f that results with the minimum work, can be found by any linear optimization algorithm. When f is calculated, the EMD between p and q is defined as:

$$d_{\text{EMD}}(p, q) = \frac{\sum\sum d_{ij} f_{ij}}{\sum\sum f_{ij}}. \tag{4}$$

In our problem, bin centers correspond to color values (ab values in the CIELAB space) and ground distances are calculated as Euclidean because of the perceptual uniformity of the CIELAB color space.

Figure 5 compares the effect of these three distance metrics for a sample image from the dataset. The image on the first column is the query image, and in each row, the most similar five images from the dataset are shown. The distance metric used in first row is Euclidean, the second row is Bhattacharyya, and the last row is the EMD. It can be argued that more similar images are found using the EMD metric.

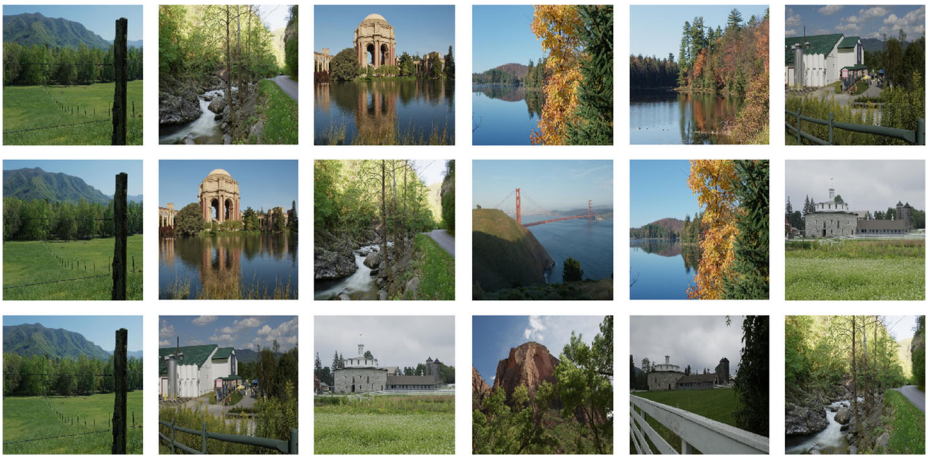


Fig. 5 A comparison of dissimilarity metrics for histogram-based features. The leftmost image is the query image, the most similar five images from the dataset are shown in each row: Euclidean distance (first row), Bhattacharyya distance (second row), Earth Mover's distance (third row)

4.2.4 Cosine distance

Cosine distance between two feature vectors p and q is calculated as:

$$d_{cos}(p,q) = 1 - \frac{\sum_{i=1}^{n} p_i q_i}{\sqrt{\sum_{i=1}^{n} p_i^2}\sqrt{\sum_{i=1}^{n} q_i^2}}. \tag{5}$$

Cosine distance is a widely used distance metric for deep representations. In this study, we used cosine distance for calculating the distances between DCNN feature vectors and GIST features.

5 Analysis & results

Having discussed the details of our crowdsourcing study along with experimented image features, we now explain how we relate features to human judgments. We first present our analysis method for assessing the correlation between each feature type and the experiment results. We then discuss two possible ways to combine the features for developing a more effective similarity model. In our evaluations, we used HDR images directly, as well as by linear scaling, and applying several tone mapping operators. For this purpose, we used the pfstmo software library [40], which provides a reliable implementation of several commonly used TMOs.

5.1 Raw feature correlations

Assume that $t_i = R_i - A_i - B_i$ represents the i^{th} triplet (i.e. trial) with R_i being the reference image, A_i the left test image, and B_i the right test image. This triplet could have been evaluated one or more times by different observers. Let $n(A_i)$ and $n(B_i)$ represent the number of times that each image was found more similar to R_i than the other. From this information, we created a binary vector to encode the participants' responses:

$$P = (x_1, \ldots, x_N), \tag{6}$$

where each element is defined as:

$$x_i = \begin{cases} 1, & \text{if } n(A_i) > n(B_i) \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

For each feature type f, we also computed the feature representations of each image as $f(R_i)$, $f(A_i)$, $f(B_i)$ and computed their similarity to each other to obtain the following binary vector:

$$F = (y_1, \ldots, y_N), \tag{8}$$

where

$$y_i = \begin{cases} 1, & \text{if } d(f(R_i), f(A_i)) < d(f(R_i), f(B_i)) \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

In this equation d represents the distance metric that was chosen to be used for feature f.

This encoding gave rise to two binary vectors, P and F, with the former computed from user responses and the latter from feature similarities. There are many approaches to compute the correlation between two such vectors. We used the Sokal-Michener correlation,

which is a simple, intuitive, and effective way to correlate two binary vectors [78]. This correlation is defined as

$$s = \frac{S_{11}(P, F) + S_{00}(P, F)}{N}, \tag{10}$$

with S_{11} and S_{00} representing the total count of matching ones and zeros respectively:

$$S_{11}(P, F) = P \cdot F, \tag{11}$$

$$S_{00}(P, F) = \neg P \cdot \neg F, \tag{12}$$

Note that the correlation coefficient s can take a value in range [0, 1]. In the following, we multiply this coefficient by 100 to represent the correlations as percentages.

The raw feature correlations with the first (Section 3.3) and the extended experiment (Section 3.4) are reported in Tables 2 and 3, respectively. In these tables, the leftmost column indicates the processing type applied to the images before the computation of features. "HDR-original" represents the unaltered HDR image whereas "HDR-linear" represents its linearly scaled version. The other processing types all include the application of a certain tone mapping operator. For all processing types, except the original, the images were gamma-corrected and scaled to [0, 255] range. However, it should be noted that the computation of features were performed in spaces that are suitable for feature types. For example, the luminance feature was computed in a linear color space (we applied degamma operation to the images). The color feature was computed in the CIELAB space, which is assumed to be perceptually uniform. VGG features were computed on gamma corrected images, which is typical as these models are trained on non-linear inputs.

We also conducted ANOVA to determine which features are statistically different from each other across different tone mapping operators (HDR-original and HDR-linear were excluded as they generally yielded lower correlations). The results, computed for the

Table 2 Individual feature correlations with the first part of the experiment

Processing Type	VGG16	VGG19	Color	Luminance	Texture	GIST
HDR-original	56.79	58.09	55.10	53.14	52.39	56.82
HDR-linear	63.54	63.31	55.69	54.07	54.36	58.18
Drago et al. [11]	65.88	65.74	56.73	57.45	51.17	58.23
Mai et al. [39]	65.28	65.13	56.01	56.77	51.90	57.57
Reinhard et al. (local) [55]	65.82	65.63	56.58	54.77	51.43	57.89
Reinhard et al. (global) [55]	65.75	65.52	56.59	54.68	51.39	57.92
Durand & Dorsey [12]	66.17	65.43	55.77	55.12	51.79	57.85
Mantiuk et al. [44]	65.42	65.33	56.29	55.38	52.08	58.03
Reinhard & Devlin [54]	65.28	65.20	57.15	55.89	54.85	58.33
Fattal et al. [15]	65.90	65.72	56.39	57.46	51.92	58.19
Mantiuk et al. [41]	65.71	65.74	55.98	56.99	51.84	57.86
Ferradans et al. [16]	66.02	65.90	55.18	56.51	51.99	58.33
Pattanaik et al. [52]	64.46	64.38	53.04	54.61	53.06	57.84

The numbers indicate the Sokal-Michener correlation scaled by 100 to represent percentages

Table 3 Individual feature correlations with the second part of the experiment

Processing Type	VGG16	VGG19	Color	Luminance	Texture	GIST
HDR-original	64.88	67.14	60.23	58.39	54.42	63.50
HDR-linear	75.58	76.13	60.78	57.79	57.97	65.71
Drago et al. [11]	80.88	81.80	62.58	62.72	53.87	65.39
Mai et al. [39]	80.00	79.95	61.11	61.66	53.46	64.06
Reinhard et al. (local) [55]	80.92	81.61	62.21	58.16	53.87	64.88
Reinhard et al. (global) [55]	80.92	81.57	62.21	57.97	54.75	64.75
Durand & Dorsey [12]	81.75	81.34	62.07	59.22	53.00	64.19
Mantiuk et al. [44]	80.41	80.65	61.15	59.59	52.49	64.47
Reinhard & Devlin [54]	80.37	80.41	64.15	61.43	60.55	65.44
Fattal et al. [15]	80.51	80.92	62.30	64.24	52.90	65.02
Mantiuk et al. [41]	80.00	80.78	62.12	61.71	54.56	64.19
Ferradans et al. [16]	81.38	82.21	58.39	61.61	55.02	65.25
Pattanaik et al. [52]	78.66	78.11	57.33	58.66	55.71	64.52
Mean	80.53	80.85	61.42	60.63	54.56	64.74

The numbers indicate the Sokal-Michener correlation scaled by 100 to represent percentages. The means are computed only for the tone mapping operators.

extended experiment, indicated significant differences between the features: $F(5, 60) = 533.1$, $p < 0.001$. As a post-hoc test, we conducted Tukey's honestly significant differences analysis that includes corrections for multiple comparisons. According to this test, we found VGG16 and VGG19 in the same significance group. Similarly color and luminance features were found in the same group as well.

Finally, we conducted Fisher's exact test [72] to determine which features have significantly different correlations with user choices. This test computes a statistical significance probability from a 2×2 contingency matrix which encodes how many responses match and mismatch with two selected feature types. p values larger than 0.05 are assumed to indicate that the two features are statistically similar. According to this analysis, only the VGG features were found to correlate similarly with user responses. The p value for the color and luminance features were 0.03 indicating a statistically significant difference. All other feature pairs were found to be statistically different with much lower p values. We used Ferradans et al.'s TMO [16] for this analysis.

5.2 Feature combination

Given the individual correlations reported in the previous tables, a natural question that follows is if we can combine them to develop a single objective metric that better correlates with human's assessment of similarity for HDR images. To this end, we performed two types of linear regression analysis yielding two related but different models.

5.2.1 Model one

In our first analysis, we aimed to develop a model that predicts which of the two test images is more similar to the reference image using the pairwise distances between the test

and reference images. Assuming that j is a feature index, one can compute these pairwise differences as follows:

$$a_j = d_j(f_j(R), f_j(A)), \tag{13}$$

$$b_j = d_j(f_j(R), f_j(B)). \tag{14}$$

Here d_j represents the distance metric chosen for the j^{th} feature. The model takes as input these differences for all features (i.e. $j \in \{1, 2, 3, 4, 5, 6\}$) and computes their weighted average as its response:

$$\begin{aligned} r = c_0 + \; & c_1(a_1 - b_1) + c_2(a_2 - b_2) + c_3(a_3 - b_3) + \\ & c_4(a_4 - b_4) + c_5(a_5 - b_5) + c_6(a_6 - b_6) \end{aligned} \tag{15}$$

To compute the unknown coefficients we used logistic regression as our dependent data (i.e. user responses) were binary: given one reference and two test images, the user selects either the left image or the right one, encoded as 1 and 0.

The regression was performed between the two vectors, namely the P vector from (6), and the model response R comprised of the following elements:

$$R = (r_1, \ldots, r_N), \tag{16}$$

where

$$r_i = [a_{i1} - b_{i1} \;\; \ldots \;\; a_{i6} - b_{i6}]. \tag{17}$$

The logistic regression models the logarithm of the odds as the response of the model:

$$ln\left(\frac{Pr(x = 1)}{1 - Pr(x = 1)}\right) = r. \tag{18}$$

From this equation, it can be derived that the probability of a user responding 1 (i.e. selecting the left image) is equal to

$$Pr(x = 1) = \frac{1}{1 + e^{-r}}. \tag{19}$$

If we find $Pr(x = 1) > 0.5$, we assume that the model has selected the left image. Otherwise, the model's response was taken as the right image.

To measure the effectiveness of this model we used 10-fold cross validation. In each fold, 90% of the trials were selected for training and the remaining 10% for testing. This process was repeated 10 times while ensuring that each test fold is mutually exclusive from each other. Similar to the analysis of individual features, we assessed the success of this model against both the original (V1) and the extended experiment (V2). The results are shown in Table 4. It can be seen that the feature combination, on average, improves the success of each presentation type by about 3% to 4%. The best three results are obtained by Ferradans et al.'s [16], Drago et al.'s [11], and Reinhard et al.'s [55] TMO algorithms. The reported coefficients are computed by using the entire dataset from the second part of the experiment (V2) due to its higher correlation with the combined features.

5.2.2 Model two

Despite the first regression model yielding high correlations exceeding 80% for most algorithms, it has an important drawback. It requires a triplet of images, one reference and two test, as input to the model. While this matches the presentation type in our experiment, a more desirable model should be able to take only two images (e.g., a query image and a test image) and produce a relative similarity score between them. This may allow, for instance,

Table 4 The correlations of the first regression model with the user responses

Processing Type	V1	V2	c0	c1	c2	c3	c4	c5	c6
HDR-original	60.67	70.76	0.0573	0.0768	-3.3241	-0.0028	-0.0124	-0.2921	-10.7505
HDR-linear	64.81	78.83	0.0005	-5.4801	-5.9902	-0.0074	-0.0635	-0.3289	-10.1782
Drago et al. [11]	67.36	**83.49**	-0.0423	-7.8751	-7.6339	-0.0506	-0.0958	0.0043	-7.3615
Mai et al. [39]	66.70	81.78	0.0085	-5.2932	-7.9526	-0.0601	-0.1078	-0.0358	-4.9275
Reinhard et al. (local) [55]	67.19	**83.21**	-0.0154	-7.3838	-8.7207	-0.0688	-0.0853	0.0145	-7.8380
Reinhard et al. (global) [55]	67.34	83.16	-0.0230	-7.0932	-8.8856	-0.0687	-0.0783	0.0101	-7.4470
Durand & Dorsey [12]	66.92	83.03	-0.0604	-8.1694	-7.3044	-0.0977	-0.0147	0.0082	-6.8549
Mantiuk et al. [44]	66.64	81.74	0.0220	-6.1999	-8.0462	-0.1081	-0.0286	-0.0102	-10.0494
Reinhard & Devlin [54]	66.72	82.75	-0.0332	-5.6555	-8.8871	-0.1284	-0.0144	-0.0254	-7.9970
Fattal et al. [15]	67.25	82.56	-0.0025	-6.2320	-8.3176	-0.1120	-0.0272	-0.0143	-7.9175
Mantiuk et al. [41]	66.91	82.15	-0.0005	-5.7555	-8.4433	-0.0777	-0.0548	-0.0041	-6.8189
Ferradans et al. [16]	67.21	**83.53**	0.0226	-5.7782	-9.8899	-0.0801	-0.0432	-0.0060	-7.4090
Pattanaik et al. [52]	65.02	79.89	-0.0365	-7.5194	-5.6052	0.0132	-0.0565	0.0211	-6.1389

V1 and V2 represent the first and extended experiments respectively. The coefficients are reported for the extended experiment only due to its higher correlation with the user data

ranking the similarity of multiple images with a query image as in image-based search applications.

In order to allow for this possibility, our second regression model was designed in the following manner. For each trial, $t_i = R_i - A_i - B_i, i \in \{1, \ldots, N\}$, we inserted two elements to our user response vector:

$$x_{2i-1} = \begin{cases} 1, & \text{if } n(A_i) > n(B_i) \\ 0, & \text{otherwise,} \end{cases} \tag{20}$$

$$x_{2i} = \neg x_{2i-1}, \tag{21}$$

yielding a vector of size $2N$:

$$P = (x_1, x_2, \ldots, x_{2N}). \tag{22}$$

As for the model's inputs each element of the feature vector was computed as

$$y_{2i-1} = [a_1 \ \ldots \ a_6], \tag{23}$$

$$y_{2i} = [b_1 \ \ldots \ b_6], \tag{24}$$

yielding

$$F = (y_1, y_2, \ldots, y_{2N}). \tag{25}$$

In summary, the elements of the feature vector always followed the A, B order, whereas the corresponding elements in the user vector were 1 for the selected image and 0 for the other image. This second regression model learns to produce the following response given the feature differences between a reference and test image:

$$r_a = c_0 + c_1 a_1 + c_2 a_2 + c_3 a_3 + c_4 a_4 + c_5 a_5 + c_6 a_6 \tag{26}$$

By converting this response to probability values as in (19), one can compute a relative degree of similarity between the two images. To validate this model, we computed the model response twice by using $R_i - A_i$ and $R_i - B_i$ image pairs:

$$Pr(x = \text{left}) = \frac{1}{1+e^{-r_a}} \tag{27}$$

$$Pr(x = \text{right}) = \frac{1}{1+e^{-r_b}} \tag{28}$$

Given a triplet, if we found $Pr(x = \text{left}) > Pr(x = \text{right})$ we assumed the model to have selected the left image. Otherwise, it was assumed that the model selects the right one. The correlation of this model with the user responses was calculated as in the previous model yielding the results in Table 5. The best result of the second model was found for Drago et al.'s [11] TMO in the extended experiment. The model achieved a correlation of 83.81% with the user responses.

6 Application: style-based tone mapping

The existence of numerous tone mapping operators that are available paved the way for many studies that are conducted for selecting the best one [51]. However, tone mapping can be conducted for different purposes, and rendering the resulting images to follow a consistent style can be one of them. For example, in a movie production process, making all frames consistently tone mapped, regardless of the content of the frames, can be a desired operation to impart a certain look and feel to the viewers. In this section, we show how an earlier work by Akyüz et al. [49] that pursues this goal can be improved with the findings of the current study.

In Akyüz et al., a small set of calibration images are first tone mapped by an artist. The artist uses a modified version of the generic TMO [45], which allows the artist to control the overall brightness, contrast, saturation, and detail present in the final image. When a new

Table 5 The correlations of the second regression model with the user responses

Processing Type	V1	V2	c0	c1	c2	c3	c4	c5	c6
HDR-original	60.75	70.80	2.9323	-0.1531	-2.8191	-0.0024	-0.0063	-0.2494	-7.1569
HDR-linear	64.65	78.50	5.5623	-3.9224	-3.7111	-0.0149	-0.0048	-0.2164	-5.5490
Drago et al. [11]	67.52	**83.81**	8.3967	-5.5248	-4.0845	-0.0280	-0.0587	-0.0054	-3.3575
Mai et al. [39]	66.72	81.73	7.4594	-4.0822	-5.0859	-0.0196	-0.0532	-0.0326	-0.9064
Reinhard et al. (local) [55]	67.35	83.53	8.2123	-5.6104	-4.3743	-0.0249	-0.0290	0.0063	-3.6705
Reinhard et al. (global) [55]	67.20	83.16	8.2162	-5.3915	-4.5828	-0.0259	-0.0264	0.0049	-3.6673
Durand & Dorsey [12]	66.81	82.61	8.2396	-5.9298	-4.0539	-0.0560	-0.0081	0.0077	-3.1001
Mantiuk et al. [44]	66.50	82.10	7.6833	-4.3626	-4.8232	-0.0658	-0.0212	-0.0082	-3.4889
Reinhard & Devlin [54]	66.56	82.61	8.5676	-4.6656	-4.9324	-0.0936	-0.0075	-0.0262	-2.8843
Fattal et al. [15]	67.07	82.79	8.0671	-4.4119	-4.8215	-0.0716	-0.0191	-0.0141	-2.8938
Mantiuk et al. [41]	66.57	82.01	7.7805	-3.9872	-5.3899	-0.0228	-0.0319	-0.0084	-2.6625
Ferradans et al. [16]	67.33	83.16	8.5911	-5.0432	-4.9965	-0.0541	-0.0258	-0.0089	-2.3825
Pattanaik et al. [52]	64.94	79.84	6.6735	-5.6657	-3.4601	0.0109	-0.0295	0.0241	-1.8922

V1 and V2 represent the first and extended experiments respectively. The coefficients are reported for the extended experiment only due to its higher correlation with the user data

image needs to be tone mapped using a style created in the calibration phase, its similarity to the calibration images are found and the tone mapping parameters are interpolated:

$$\mathbf{t} = \frac{\sum_{i=1}^{N} \frac{1}{d(\mathbf{f},\mathbf{f_i})} \mathbf{t_i}}{\sum_{i=1}^{N} \frac{1}{d(\mathbf{f},\mathbf{f_i})}} \tag{29}$$

Here, $\mathbf{t_i}$ are the tone mapping parameters and $\mathbf{f_i}$ the feature vector for the calibration image i. $\mathbf{f}$ is the feature vector of the current input image and $\mathbf{t}$ its computed tone mapping parameters. Finally d is a distance function that measures the similarity between the two features vectors. In the original work of Akyüz et al. [49], the feature vector is represented by a 60 dimensional HSV and gradient histogram. The distance metric is the Euclidean distance. Here we show two modifications of this method that are made possible by the experimental findings of the current study.

6.1 Version I

In the first version, features given in Table 1 are extracted from the selected HDR image and calibration images. Then, distances between these features are calculated using the corresponding distance metrics given in the same table. The weighted average of these feature distances are calculated using the coefficients obtained from the logistic regression model (26), with the idea that less important features should also contribute less to the distance. This operation can be summarized with the following equation:

$$d_i = \sum_{j=1}^{6} c_j d_j(\mathbf{f_j}, \mathbf{f_{ij}}), \tag{30}$$

where c_j is the coefficient of the j^{th} feature, $\mathbf{f_j}$ is the j^{th} feature of the input image, $\mathbf{f_{ij}}$ is the same for the i^{th} calibration image, and finally d_j is the distance metric for the j^{th} feature. The result d_i represents the combined distance between the input image and the corresponding calibration image. These combined distances are calculated between the selected HDR image and all calibration images. The tone mapping parameters for the selected HDR image are then interpolated using inverse distance transform as in (29).

6.2 Version II

While the previous approach calculates a single distance value between images and use this value to interpolate *all* tone mapping parameters, Version II relates model features with tone mapping parameters and interpolates individual tone mapping parameters with different weights. To achieve this, we use the relationships defined in Table 6.

For example, the brightness parameter t_b is computed by interpolating the t_{b_i} parameters of the calibration images by using the similarity of the luminance features:

$$t_b = \frac{\sum_{i=1}^{N} \frac{1}{d_{\text{lum}}(\text{lum},\text{lum}_i)} t_{b_i}}{\sum_{i=1}^{N} \frac{1}{d_{\text{lum}}(\text{lum},\text{lum}_i)}} \tag{31}$$

Other parameters are interpolated analogously. Because GIST and deep learning features are not directly linked to a specific appearance phenomenon but are measures of overall similarity, we did not directly link them to specific tone mapping parameters. Instead we experimented with merging them using the individually interpolated parameters as follows:

$$\mathbf{t} = w_0\mathbf{t_0} + w_1\mathbf{t_1} + w_2\mathbf{t_2}, \tag{32}$$

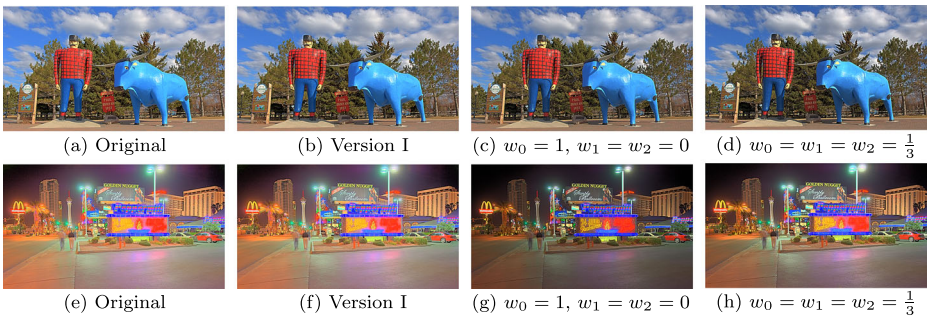
Table 6 Model features used for interpolation of tone mapping parameters used in Version II

Tone mapping parameter	Model feature
Brightness (t_b)	Luminance
Contrast (t_c)	Luminance
Black point (t_{bp})	Luminance
White point (t_{wp})	Luminance
Color saturation (t_s)	Chromaticity
Small detail strength (t_{λ_s})	Texture
Medium detail strength (t_{λ_m})	Texture
Large detail strength (t_{λ_l})	Texture

where $\mathbf{t_0}$ represents individually interpolated TMO parameters, $\mathbf{t_1}$ TMO parameters interpolated as a whole using GIST similarity only, and $\mathbf{t_2}$ TMO parameters interpolated as a whole using solely deep learning feature similarity. The weights control the influence of TMO parameters that are computed by using these different approaches.

6.3 Results

In Figure 6, we compare several results obtained by using the original style based tone mapping method as well as with the modifications proposed above. In the first row, we show the results of the "Paul Bunyan" scene from the HDR Photographic Survey [14]. This scene depicts a bright outdoors environment with colorful foreground objects. It may be noted that all results are similar but the individual parameter interpolation with equally weighted GIST and deep learning features (d) has slightly higher contrast (please refer to supplementary full resolution images for better comparison). The overall colorful style is preserved in all images. In the second row, we show the "Peppermill" night scene from the same dataset. For this scene the difference of Version II is more clear as images in (c) and (d) exhibit a darker rendering, which is more suitable for a night scene. The reason for this darkening effect is that the t_b parameter for tone mapping becomes more similar to the t_b parameter of the night image in the calibration set due to the similarity of the *luminance* features between these



(a) Original (b) Version I (c) $w_0 = 1,\ w_1 = w_2 = 0$ (d) $w_0 = w_1 = w_2 = \frac{1}{3}$

(e) Original (f) Version I (g) $w_0 = 1,\ w_1 = w_2 = 0$ (h) $w_0 = w_1 = w_2 = \frac{1}{3}$

Fig. 6 Application of our findings for the style-based tone mapping problem. Original results are shown in the first column, followed by Version I in the second column and two variants of Version II in the last two columns

images. The addition of GIST and deep learning features in (d) yields a slightly brighter image compared to (c). We encourage the readers to refer to the electronic supplementary materials for more clear observation of the differences.

7 Summary and discussion

We performed two user experiments followed by statistical analyses to get an in-depth understanding of image similarity for HDR images. We first collected a large number of human similarity responses via crowdsourcing, and then evaluated several image features with respect to the collected data. Evaluation is performed both on individual features and on their combination. When combining features, two models both of which using logistic regression are considered. Although both models are found to perform comparably, the second one permits direct one-to-one comparison, making it more suitable for practical applications. We show one such application, namely style-based tone mapping that benefits from the experimental findings. Key observations obtained in our work are the following:

1. When properly tone mapped images are used as compared to using either original or linearly scaled HDR images, higher correlations with human responses are obtained.
2. Most tone mapping operators (TMOs) yield comparable performance.
3. Deeply learned features, in comparison to hand-crafted features, correlate better with the human responses.
4. Among hand-crafted features, GIST yields the highest correlation, followed by color, luminance, and texture.
5. All of the estimated correlations for the second experiment are higher in comparison to those for the first experiment.

The first observation highlights the importance of using tone mapped data for HDR image similarity. While tone mapping is a lossy process, it brings the data to a more meaningful range for the computation of most features. However, some features are less dependent on tone mapping. For instance the texture feature represented by the histogram of oriented gradients is found to produce about the same correlation regardless of whether HDR or tone mapped data is used. This is followed by the color feature represented by 2D chromaticity histogram. Among the hand-crafted features the largest difference is observed for luminance when tone mapped data is used. This can be expected as non-linear luminance compression often eliminates large gaps in luminance histogram where little useful information is present.

Perhaps unexpectedly, the second observation suggests that TMOs perform comparably. Although there exists a large number of TMO evaluation studies, we are not aware of any work that compares TMOs for the task of HDR image similarity. The lowest performing operator is found to be Pattanaik et al.'s [52] algorithm. It is, however, known that this algorithm highly depends on calibrated input data and viewing conditions as it tries to accurately model the human visual system.

As for the third observation, it is not surprising to find that features obtained from a DCNN [66] trained over a large image dataset [60] outperform simple hand-crafted features. Similar findings are reported by image retrieval studies conducted for low dynamic images [21, 73]. For HDR images, our findings indicate that deep features are mostly useful if the images are tone mapped to the 8-bit per color channel domain first. This is also expected as the training data of DCNNs are comprised of such images.

The fourth observation indicates that the GIST descriptor surpasses the texture, luminance, and color features for HDR image similarity. In addition to outperforming them, in fact, it performs surprisingly consistently across different processing types. Despite having a smaller correlation with the user data than the deep features, it exhibits less variability overall. This may be a desirable property as it appears to be minimally affected by how an HDR image is processed.

Pertaining to our last observation, it can be argued that seeking multiple consistent responses by the participants are important; not only for developing a more reliable model but also for assessing the correlation of different features with user responses. For instance, inspection of Tables 2 and 3 reveals that while deep features correlate better with the user responses, this difference is clearly magnified for the second experiment. In other words, as the experimental findings become more reliable the merits and drawbacks of different features become more noticeable.

Our work also has certain limitations and drawbacks. Firstly, we relied on crowdsourcing, which was necessary to reach a wider audience but made it impossible to control the viewing conditions of the participants. Different results could have been obtained if the experiments were done in a laboratory environment with controlled display and lighting conditions. Secondly, the participants compared the HDR images on standard monitors and used sliders to visualize different image regions. Again, different results could have been obtained if participants viewed the images on an HDR display. Finally, we intentionally did not define the meaning of similarity and left this to the interpretation of the participants. To reduce this uncertainty, future studies may explicitly define what is meant by similarity such as object similarity, color similarity, indoor-outdoor similarity, time-of-day similarity, etc.

We believe that our work simply scratches the surface of the HDR image similarity problem. The proposed models can be extended with different types of features. Further experiments which consider ranking and rating tasks as well as pairwise comparisons can be conducted. Evaluations may include DCNNs that are either fine-tuned or trained with HDR data from the ground up. Given the large number of image quality datasets and subjective evaluations in the form of mean opinions scores (MOS), whether image quality and similarity correlate with each other in the context of HDR imaging can be investigated. Image saliency can also be taken into account for similarity judgments as it was found to improve performance in some other domains [1]. Perhaps most importantly, the effect of *calibrated* HDR images for image similarity and retrieval tasks can be studied. As objects are represented with their true luminances in calibrated data, this may simplify similarity assessment between the images. Finally, with emerging standards for HDR video streaming such as HDR10+ and Dolby Vision, we envision the HDR video similarity problem to gain importance in near future.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11042-021-11182-7.

Declarations

Conflict of Interests All authors declare that they have no conflict of interest.

References

1. Amirkhani D, Bastanfard A (2019) Inpainted image quality evaluation based on saliency map features. In: 2019 5th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), pp 1–6
2. Banterle F, Artusi A, Debattista K, Chalmers A (2011) Advanced high dynamic range imaging: Theory and practice, First. CRC Press (AK Peters), Natick, MA
3. Bay H, Tuytelaars T, Van Gool L (2006) Surf: Speeded up robust features. In: European conference on computer vision, pp 404–417. Springer
4. Bhattacharyya A (1946) On a measure of divergence between two multinomial populations. Sankhyō: The Indian Journal of Statistics, pp 401–406
5. Brown KC, Bryant T, Watkins MD (2010) The forensic application of high dynamic range photography. J Forensic Identification 60(4):449–459
6. Cai H (2013) High dynamic range photogrammetry for synchronous luminance and geometry measurement. Light Res Technol 45(2):230–257
7. Chalmers A, Campisi P, Shirley P, Olaizola IG (2016) High dynamic range video: concepts, technologies and applications. Academic Press
8. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: international Conference on computer vision & Pattern Recognition (CVPR'05), vol 1, pp 886–893. IEEE Computer Society
9. Debevec PE, Malik J (1997) Recovering high dynamic range radiance maps from photographs. In: SIGGRAPH 97 Conference Proceedings, pp 369–378
10. Donahue J, Jia Y, Vinyals O, Hoffman J, Zhang N, Tzeng E, Darrell T (2014) Decaf: A deep convolutional activation feature for generic visual recognition. In: International conference on machine learning, pp 647–655
11. Drago F, Myszkowski K, Annen T, Chiba N (2003) Adaptive logarithmic mapping for displaying high contrast scenes. In: Computer Graphics Forum, vol 22, pp 419–426. Wiley Online Library
12. Durand F, Dorsey J (2002) Fast bilateral filtering for the display of high-dynamic-range images. ACM Trans Graph 21(3):257–266
13. Empa hdr image database. http://www.empamedia.ethz.ch/hdrdatabase/ Accessed: 2017-08-26
14. Fairchild MD (2007) The hdr photographic survey. In: Color and Imaging Conference, pp 233–238. Society for Imaging Science and Technology
15. Fattal R, Lischinski D, Werman M (2002) Gradient domain high dynamic range compression. ACM Trans Graph 21(3):249–256
16. Ferradans S, Bertalmio M, Provenzi E, Caselles V (2011) An analysis of visual adaptation and contrast perception for tone mapping. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(10):2002–2012
17. Ferwerda JA, Pattanaik S, Shirley P, Greenberg DP (1996) A model of visual adaptation for realistic image synthesis. In: SIGGRAPH 96 Conference Proceedings, pp 249–258
18. Frese T, Bouman CA, Allebach JP (1997) Methodology for designing image similarity metrics based on human visual system models. In: Human Vision and Electronic Imaging II, vol 3016, pp 472–483. International Society for Optics and Photonics
19. Froehlich J, Grandinetti S, Eberhardt B, Walter S, Schilling A, Brendel H (2014) Creating cinematic wide gamut hdr-video for the evaluation of tone mapping operators and hdr-displays. In: Digital Photography X, vol 9023, p 90230X. International Society for Optics and Photonics
20. Glassner AS (1995) Principles of digital image synthesis: Vol. 1, Elsevier
21. Gordo A, Almazán J, Revaud J, Larlus D (2016) Deep image retrieval: Learning global representations for image search. In: European conference on computer vision, pp 241–257. Springer
22. Grimaldi A, Kane D, Bertalmío M (2019) Statistics of natural images as a function of dynamic range. J Vis 19(2):13–13. https://doi.org/10.1167/19.2.13
23. Grinzato E, Cadelano G, Bison P, Petracca A (2009) Seismic risk evaluation aided by ir thermography. In: SPIE Defense, Security, and Sensing, pp 72990C–72990C. International Society for Optics and Photonics
24. Hanhart P, Bernardo MV, Pereira M, Pinheiro AMG, Ebrahimi T (2015) Benchmarking of objective quality metrics for hdr image quality assessment. EURASIP Journal on Image and Video Processing 2015(1):1–18
25. Happa J, Artusi A, Czanner S, Chalmers A (2010) High dynamic range video for cultural heritage documentation and experimental archaeology. In: Proceedings of the 11th International conference on Virtual Reality, Archaeology and Cultural Heritage, pp 17–24. Eurographics Association

26. Harifi S, Bastanfard A (2015) Efficient iris segmentation based on converting iris images to high dynamic range images. In: 2015 Second International Conference on Computing Technology and Information Management (ICCTIM), pp 115–119. IEEE
27. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
28. ISO EN (2011) 11664-4 colorimetry—part 4: Cie 1976 l* a* b* colour space. CEN (European Committee for Standardization): Brussels, Belgium
29. Kalantari NK, Ramamoorthi R (2017) Deep high dynamic range imaging of dynamic scenes. ACM Trans. Graph 36(4):144
30. Kleiman Y, Goldberg G, Amsterdamer Y, Cohen-Or D (2016) Toward semantic image similarity from crowdsourced clustering. Vis Comput 32(6-8):1045–1055
31. Klíma M, Fliegel K, Pata P, Vitek S, Blažek M, Dostal P, Krasula L, Kratochvíl T, Rícnỳ V, Slanina M et al (2011) Deimos–an open source image database. Radioengineering, vol 20 (4)
32. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
33. Kundu D, Ghadiyaram D, Bovik AC, Evans BL (2016) Espl-live hdr image quality database. Online: http://signal.ece.utexas.edu/debarati/HDRDatabase.zip,[Mar, 2017]
34. Kundu D, Ghadiyaram D, Bovik AC, Evans BL (2017) Large-scale crowdsourced study for high dynamic range images. IEEE Trans Image Process 26(10):4725–4740
35. Larson GW, Shakespeare RA (1998) Rendering with radiance. Morgan Kaufmann Publishers
36. Liu Y, Zhang D, Lu G, Ma W-Y (2007) A survey of content-based image retrieval with high-level semantics. Pattern recognition 40(1):262–282
37. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. International journal of computer vision 60(2):91–110
38. Lun Z, Kalogerakis E, Sheffer A (2015) Elements of style: learning perceptual shape style similarity. ACM Transactions on Graphics (TOG) 34(4):84
39. Mai Z, Mansour H, Mantiuk R, Nasiopoulos P, Ward R, Heidrich W (2011) Optimizing a tone curve for backward-compatible high dynamic range image and video compression. IEEE Trans Image Process 20(6):1558–1571. https://doi.org/10.1109/TIP.2010.2095866
40. Mantiuk R (2007) High dynamic range imaging: towards the limits of the human visual perception. Forsch. Wiss. Rechnen 72:11–27
41. Mantiuk R, Daly S, Kerofsky L (2008) Display adaptive tone mapping. ACM Trans. Graph. 27:68:1–68:10. https://doi.org/10.1145/1360612.1360667
42. Mantiuk R, Heidrich W (2009) Visualizing high dynamic range images in a web browser. J Graphics, GPU, and Game Tools 14(1):43–53
43. Mantiuk R, Kim KJ, Rempel AG, Heidrich W (2011) Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. ACM Trans. Graph. 30(4):40:1–40:14. https://doi.org/10.1145/2010324.1964935
44. Mantiuk R, Myszkowski K, Seidel H-P (2006) A perceptual framework for contrast processing of high dynamic range images. ACM Transactions on Applied Perception (TAP) 3(3):286–308
45. Mantiuk R, Seidel H-P (2008) Modeling a generic tone-mapping operator. Computer Graphics Forum 27(2):699–708
46. Narwaria M, Da Silva MP, Le Callet P (2015) Hdr-vqm: An objective quality measure for high dynamic range video. Signal Process Image Commun 35:46–60
47. Nemoto H, Korshunov P, Hanhart P, Ebrahimi T (2015) Visual attention in ldr and hdr images. In: 9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)
48. Neumann D, Gegenfurtner KR (2006) Image retrieval and perceptual similarity. ACM Transactions on Applied Perception (TAP) 3(1):31–47
49. Oğuz Akyüz A, Bloch MAC, Hadimli K (2013) Style-based tone mapping for hdr images. In: SIGGRAPH Asia 2013 Technical Briefs. ACM. No. 39
50. Oliva A, Torralba A (2001) Modeling the shape of the scene: A holistic representation of the spatial envelope. International journal of computer vision 42(3):145–175
51. Parraga CA, Otazu X et al (2018) Which tone-mapping operator is the best? a comparative study of perceptual quality. JOSA A 35(4):626–638
52. Pattanaik SN, Tumblin J, Yee H, Greenberg DP (2000) Time-dependent visual adaptation for fast realistic image display. In: Proceedings of the 27th annual conference on Computer graphics and interactive techniques, pp 47–54. ACM Press/Addison-Wesley Publishing Co.
53. Rawat S, Gairola S, Shah R, Narayanan PJ (2018) Find me a sky: A data-driven method for color-consistent sky search and replacement. In: International Conference on Multimedia Modeling, pp 216–228. Springer

54. Reinhard E, Devlin K (2005) Dynamic range reduction inspired by photoreceptor physiology. IEEE Trans Vis Comput Graph 11(1):13–24
55. Reinhard E, Stark M, Shirley P, Ferwerda J (2002) Photographic tone reproduction for digital images. ACM Trans Graph 21(3):267–276
56. Reinhard E, Ward G, Pattanaik S, Debevec P (2010) High dynamic range imaging: Acquisition, display and image-based lighting, Second. Morgan Kaufmann, San Francisco
57. Rizzi A, Barricelli BR, Bonanomi C, Albani L, Gianini G (2018) Visual glare limits of hdr displays in medical imaging. IET Comput Vis 12(7):976–988
58. Rogowitz BE, Frese T, Smith JR, Bouman CA, Kalin EB (1998) Perceptual image similarity experiments. In: Photonics West'98 Electronic Imaging, pp 576–590
59. Rubner Y, Tomasi C, Guibas LJ (2000) The earth mover's distance as a metric for image retrieval. International journal of computer vision 40(2):99–121
60. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M et al (2015) Imagenet large scale visual recognition challenge. International journal of computer vision 115(3):211–252
61. STANDARD SMPTE (2016) Dynamic metadata for color volume transform–core components
62. Saleh B, Dontcheva M, Hertzmann A, Liu Z (2015) Learning style similarity for searching infographics. In: Proceedings of the 41st graphics interface conference, pp 59–64. Canadian Information Processing Society
63. Seetzen H, Heidrich W, Stuerzlinger W, Ward G, Whitehead L, Trentacoste M, Ghosh A, Vorozcovs A (2004) High dynamic range display systems. ACM Trans Graph 23(3):760–768
64. Sen P, Kalantari NK, Yaesoubi M, Darabi S, Goldman DB, Shechtman E (2012) Robust patch-based hdr reconstruction of dynamic scenes. ACM Trans. Graph. 31(6):203
65. Sharma M, Ghosh H (2015) Histogram of gradient magnitudes: a rotation invariant texture-descriptor. In: 2015 IEEE International Conference on Image Processing (ICIP), pp 4614–4618. IEEE
66. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
67. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
68. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1–9
69. Theodor JM, Furr RS (2009) High dynamic range imaging as applied to paleontological specimen photography. Palaeontol Electron, 12(1)
70. Tocci MD, Kiser C, Tocci N, Sen P (2011) A versatile HDR video production system. In: ACM Transactions on Graphics (TOG), 30, p 41. ACM
71. Tumblin J, Rushmeier H (1993) Tone reproduction for computer generated images. IEEE Comput Graph Appl 13(6):42–48
72. Upton GJG (1992) Fisher's exact test. J Royal Statistical Society: Series A (Statistics in Society) 155(3):395–402
73. Wan J, Wang D, Hoi SCH, Wu P, Zhu J, Zhang Y, Li J (2014) Deep learning for content-based image retrieval: A comprehensive study. In: Proceedings of the 22nd ACM international conference on Multimedia, pp 157–166
74. Ward G, Rushmeier H, Piatko C (1997) A visibility matching tone reproduction operator for high dynamic range scenes. IEEE Trans. on Visualization and Comp. Graphics, 3(4)
75. Wu H-HP, Lee Y-P, Chang S-H (2012) Fast measurement of automotive headlamps based on high dynamic range imaging. Applied optics 51(28):6870–6880
76. Yeganeh H, Wang Z (2012) Objective quality assessment of tone-mapped images. IEEE Transactions on Image processing 22(2):657–667
77. Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? In: Advances in neural information processing systems, pp 3320–3328
78. Zhang B, Srihari SN (2003) Properties of binary vector dissimilarity measures. In: Proc. JCIS Int'l Conf. Computer Vision, Pattern Recognition, and Image Processing, 1. Citeseer
79. Zhou B, Zhao H, Puig X, Fidler S, Barriuso A, Torralba A (2017) Scene parsing through ade20k dataset. In: Proc. CVPR