



# Wide Color Gamut Image Content Characterization: Method, Evaluation, and Applications

Junghyuk Lee, Toinon Vigier, Patrick Le Callet, Jong-Seok Lee

## ► To cite this version:

Junghyuk Lee, Toinon Vigier, Patrick Le Callet, Jong-Seok Lee. Wide Color Gamut Image Content Characterization: Method, Evaluation, and Applications. IEEE Transactions on Multimedia, inPress, 10.1109/TMM.2020.3032026 . hal-03205745

**HAL Id: hal-03205745**

**<https://hal.science/hal-03205745>**

Submitted on 22 Apr 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Wide Color Gamut Image Content Characterization: Method, Evaluation, and Applications

Junghyuk Lee, Toinon Vigier, Patrick Le Callet, *Fellow, IEEE*, and Jong-Seok Lee, *Senior Member, IEEE*

**Abstract**—In this paper, we propose a novel framework to characterize a wide color gamut image content based on perceived quality due to the processes that change color gamut, and demonstrate two practical use cases where the framework can be applied. We first introduce the main framework and implementation details. Then, we provide analysis for understanding of existing wide color gamut datasets with quantitative characterization criteria on their characteristics, where four criteria, i.e., coverage, total coverage, uniformity, and total uniformity, are proposed. Finally, the framework is applied to content selection in a gamut mapping evaluation scenario in order to enhance reliability and robustness of the evaluation results. As a result, the framework fulfils content characterization for studies where quality of experience of wide color gamut stimuli is involved.

**Index Terms**—Wide color gamut, color gamut mapping, content characterization, content selection, quality of experience.

## I. INTRODUCTION

IN order to provide more realistic and higher visual quality of experience (QoE) of multimedia contents to viewers, technologies related to wide color gamut (WCG) have emerged. Since the HDTV standard ITU-R Rec.709 [2], several WCGs have been proposed. International Telecommunication Union (ITU) approved Rec.2020 [3] as the standard color gamut for UHDTV, which covers the widest area of the CIE 1931 space [4] (see Fig. 1). Recently, many devices including mobile devices support WCGs as a process of transition to Rec.2020 [5]. Considering various environments of multimedia content consumption, gamut mapping is often inevitable in order to match the original color to displaying devices.

In this situation, several gamut mapping algorithms (GMAs) have been proposed as well as the standard algorithms in the CIE guideline [6]. Among them, gamut reduction aims to reproduce details and color quality of WCG images in smaller gamuts, and maps colors from a large source gamut

to a smaller target gamut. For instance, a gamut reduction algorithm proposed in [7] iteratively modifies the color of each pixel based on adaptive local contrast according to the Retinex theory [8]. In developing and evaluating methods related to color representation of visual contents including WCG and GMA, it is important to assess how the result will be perceived by human observers. In order to assess perceived QoE, subjective and/or objective studies are usually conducted [9], [10], [11], [12], [13].

When subjective or objective QoE evaluation is conducted, one of the primary steps is to select a representative and compact set of source contents, whose processed versions are assessed. This step, equipped with a proper content characterization method, is important not only to conduct an experiment efficiently with limited resources (especially for subjective evaluation) but also to draw reliable and reproducible conclusions. If the contents for an experiment are biased and not representative in their characteristics, the results may be biased and not be generalizable for other types of contents. Thus, it is important to select representative contents according to the purpose of a specific experiment. Towards this, it is necessary to objectively measure the representativeness and suitability of a set of contents.

In this paper, we propose a novel framework to characterize WCG contents and its applications.<sup>1</sup> We note that WCG contents are frequently exposed to the gamut mapping processes targeting diverse displaying environments. Therefore, it is important to consider the perceptual difference caused by gamut reduction for the WCG contents. Thus, our main idea is to measure perceptual difference due to successive gamut reduction in order to characterize a WCG content. We also validate the framework by applying it to two applications involving content characterization and selection in practical WCG-related studies.

Our main contributions are summarized as follows:

- 1) We propose an objective framework for WCG content characterization based on perceptual properties related to differences due to color gamut change. We obtain the perceptual difference due to gamut reduction by predicting the subjective score with an objective metric.
- 2) In order to demonstrate its effectiveness, we apply the framework to practical applications related to WCG. As one of the applications, we propose multiple criteria characterizing WCG datasets quantitatively based on the

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the “ICT Consilience Creative Program” (IITP-2019-2017-0-01015) supervised by the IITP (Institute for Information & communications Technology Promotion), the International Research & Development Program of the National Research Foundation of Korea (NRF) funded by the Korea government (MSIT) (NRF-2016K1A3A1A21005710), and the Science and Technology Amicable Research (STAR) Program funded by the Partenariats Hubert Curien (PHC) and the Campus France (PHC-STAR 36664WK).

A preliminary version of this work was presented at the International Conference on Image Processing (ICIP) in 2018 [1].

© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

<sup>1</sup>Our code is publicly available at <https://github.com/junghyuk-lee/WCG-content-characterization>

proposed perceptual difference. Using them, we conduct analysis of existing WCG datasets.

- 3) In addition, we apply the framework in a scenario of benchmarking GMAs. We demonstrate that the reliability of the benchmarking is maximized by content selection using the proposed framework.

Note that this paper has distinguished contributions compared to our preliminary work [1] in various respects. While the preliminary work only introduces the basic idea of the proposed framework, this paper provides its detailed description and further analysis along with the shared source code. In addition, we present the two practical applications involving WCG contents and demonstrate the effectiveness of our framework for content characterization.

The rest of this paper is organized as follows. In Section II, we briefly survey the related works. In Section III, we present the proposed framework for WCG content characterization and provides its implementation details. In Section IV, we describe how the proposed framework is applied to quantify characteristics of WCG datasets and provide analysis of existing WCG datasets. In Section V, we describe another use case of the proposed framework for comparison of GMAs. Finally, Section VI provides concluding remarks.

## II. RELATED WORKS

### A. Gamut Mapping

In order to reproduce the original color of contents in devices having smaller color gamuts, several GMAs have been proposed. They can be categorized into global and local strategies. The former changes all colors of out-of-gamut pixels towards the inside of the target gamut by gamut compression or clipping [14], [15], [16], [17], [18]. QoE of the gamut-reduced image often decreases since the color may become blurred around the pixels where the color changes. The latter considers spatial relationship between pixels at the expense of increased computational complexity in order to enhance perceived quality of the gamut-reduced images [7], [19], [20], [21], [22], [23], [24], [25].

### B. QoE Assessment of Gamut Mapping

QoE of gamut-mapped contents is usually assessed by conducting subjective or objective studies. In [26], a psychophysical experiment is conducted to evaluate four GMAs, where the subjective quality of the gamut-reduced images is assessed. In [27], [28], [29], [30], [31], various color image difference metrics are proposed to measure objective quality of gamut-mapped images. In [32], subjective scores of gamut-reduced images using different GMAs are obtained by a psychophysical experiment, and are used to evaluate four objective metrics.

However, in [33], it is concluded that the color difference measured by objective metrics and the perceived image difference between original and gamut-reduced images do not correlated well. There are attempts to improve objective metrics by employing spatial filtering that simulates the human visual system [34] and by extracting features based on perceptually important distortion [35]. On the other hand, studies

that consider measuring QoE of WCG contents are rare. In [36], a physiological experiment is conducted to measure electroencephalography during watching WCG video contents.

### C. Content Characterization

Winkler [37] quantifies the characteristics of the contents in existing image and video datasets, including spatial information and colorfulness for color images, and motion vectors for video contents, based on which the representativeness of a set of contents can be evaluated [38], [39]. In [40], it is suggested to consider attributes of the test material such as brightness, colorfulness, amount of motion, scene cuts, types of the content, etc. for subjective video quality assessment. In [41], contrast, colorfulness, and naturalness are considered to characterize tone-mapped images for HDR contents. In [42], a content selection procedure for light field images is proposed using high-level features consisting of depth properties, disparity range of pixels, refocusing features, etc. as well as general image quality features.

In [43], however, it is argued that those simple characteristics do not sufficiently cover the perceptual aspects of visual contents when processing steps (i.e. tone-mapping) are involved. Therefore, an approach is proposed to characterize HDR contents in the viewpoint of whether an HDR content is challenging for tone mapping operators. It focuses on the perceptual change due to the dynamic range reduction that is frequently applied to HDR contents. Using this characterization method, a framework to build a representative HDR dataset is proposed in [44]. In a similar spirit, we propose a novel characterization framework for WCG contents.

## III. PROPOSED FRAMEWORK

### A. General Algorithm

We propose a framework for WCG content characterization based on the perceptual change caused by gamut mapping. We define WCG content characteristics as degrees of the perceptual differences due to successive gamut reduction. The overall procedure of the proposed method is summarized in Algorithm 1.

The framework in Algorithm 1 produces an  $N$ -dimensional feature vector of perceptual difference for each WCG source content. First, we obtain  $N$  gamut-reduced images by applying a gamut reduction operator that converts the color gamut of the reference image  $G_0$  into a target gamut  $G_n$  ( $n = 1, \dots, N$ ). For each gamut-reduced image  $I_n$ , we apply an objective metric that measures the perceptual difference from the reference image  $I_0$ . Finally, we obtain a feature vector  $D$  describing the behavior of the WCG content in terms of perceptual difference due to gamut reduction. We can utilize this feature in various applications such as WCG dataset analysis, content clustering, and selection, which will be presented in Sections IV and V.

### B. Obtaining Ground Truth of Perceptual Difference

Hereafter, we provide implementation details of the proposed framework. In Algorithm 1, we use an objective metric  $PD$  to measure the perceptual difference due to gamut

---

**Algorithm 1** General framework for WCG content characterization
 

---

# **Input**  $I_0$ : WCG source image  
 # **Output**  $D$ : vector of perceptual differences of  $I_0$   
 #  $N$ : number of target gamut spaces for gamut reduction  
 #  $G_0$ : reference gamut space that covers all colors of  $I_0$   
 #  $G_n$ :  $n$ -th gamut space smaller than  $G_{n-1}$  ( $n = 1, \dots, N$ )  
 #  $f_{GR}(I, G)$ : function that generates a gamut-reduced image with all colors in gamut  $G$  from image  $I$   
 #  $PD(I, I')$ : function that measures the perceptual difference between images  $I$  and  $I'$

**for**  $n = 1 : N$  **do**

    Generate  $I_n = f_{GR}(I_{n-1}, G_n)$

    Calculate  $d_n = PD(I_n, I_0)$

**end for**

Obtain  $D = [d_1, d_2, \dots, d_N]^T$

---

reduction. Although various image quality metrics have been proposed in literature, metrics specifically designed to measure perceptual difference of images exposed to color gamut change do not exist. Therefore, we conduct a subjective test in which the mean opinion score (MOS) of the perceptual difference between gamut-mapped images is measured. MOS is then used to benchmark existing color metrics and optimize the best one via nonlinear transformation.

1) *Data*: We collect 54 images consisting of scenes from HdM-HDR-2014 [45] and Arri Alexa sample footage<sup>2</sup>, in short, HdM and Arri, respectively. HdM contains videos filmed in a professional cinematography environment with dynamic ranges up to 18 stops and a color gamut close to Rec.2020. Especially, it focuses on the WCG by containing videos with highly saturated color and lights. The perceptual difference of the videos is large when the gamut is reduced. Arri is a video sample footage provided by the ARRI company. Contents of the dataset are in various natural topics with up to the Rec.2020 color gamut. Compared to HdM, color differences are not large when the gamut is not much reduced. The collected image set is divided into training and validation sets of 30 and 24 images, respectively. The HDR images from HdM are converted to the standard dynamic range with a fixed value of exposure.

We use DCI-P3 as a reference WCG, which originates from the cinema industry. And, we use two target gamuts for gamut reduction (i.e.,  $G_1$  and  $G_2$ ): Rec.709 and Toy. With widespread displays abiding by the HDTV standard, gamut reduction from P3 to Rec.709 frequently happens to WCG contents. In addition, to cover a high degree of gamut reduction, we employ an artificially created gamut, called Toy, which has been used in the state-of-the-art WCG studies [7], [46]. It is smaller than Rec.709 and produces large perceptual difference when the gamut of a WCG image is reduced to it. The choice of these two gamuts is based on our preliminary experiments, where for the gamuts between P3 and Rec.709, the gamut-reduced images are not visually distinguishable from those in P3 nor Rec.709; in addition, gamuts smaller than Toy give rise

TABLE I  
RGB PRIMARY COLORS IN THE CIE 1931 COLOR SPACE

Gamuts	Red Primaries		Green Primaries		Blue Primaries	
	x	y	x	y	x	y
DCI-P3	0.680	0.320	0.265	0.690	0.150	0.060
Rec.709	0.640	0.330	0.300	0.600	0.150	0.060
Toy	0.570	0.320	0.300	0.530	0.190	0.130

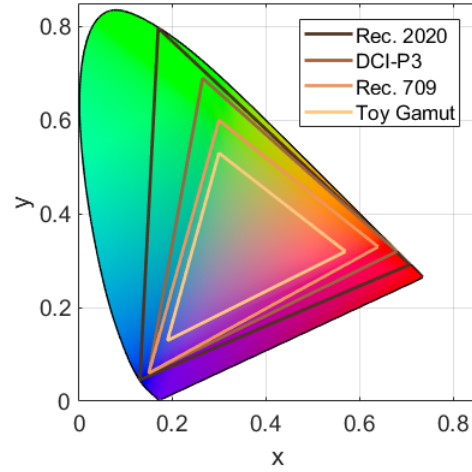


Fig. 1. Color spaces of color gamuts considered in this work on CIE 1931 chromaticity diagram.

to too much color distortion in the gamut-reduced images and thus are not practically meaningful. For gamut reduction, we consider a simple gamut mapping algorithm because complex and time-consuming algorithms are not preferred in the content characterization process. Hence, we use the gamut clipping method that maps colors outside the target gamut at the nearest boundary of the target gamut.

2) *Subjective Test*: We adopt the paired comparison test methodology [47] for the subjective test, because the difference due to gamut reduction is mostly subtle perceptual difference rather than large quality distortion. The reference image in the P3 gamut and one of the gamut-reduced images produced in Section III-B1 are shown in a side-by-side manner. The images are compared in terms of color difference on a three-point scale: no difference (0), slight difference (1), and clear difference (2).

The test is conducted under the standardized test room condition complying with the laboratory condition described in ITU-R BT.500 such as luminance of the monitor, room illumination, observers, etc. [48]. We use an EZIO ColorEdge monitor that can display up to the P3 color gamut. We heuristically crop each image in half-width ( $960 \times 1080$  pixels) to show both images side-by-side on a single monitor. Participants are 51 healthy non-expert volunteer subjects consisting of 26 males and 25 females, who are screened by a color and vision test. We obtain the MOS for each of the 60 images (30 source images  $\times$  two target gamuts) by taking the average value of the ratings over the subjects.

<sup>2</sup><https://www.arri.com/en/learn-help/learn-help-camera-system/camera-sample-footage>

The test consists of an exercise and a test sessions. During the exercise session, the test methodology is described to the subjects with five exercise stimuli that are different from the test stimuli. The test session proceeds sequentially for each pair of images as follows. First, a reference image and one of its gamut-reduced versions are displayed on the monitor up to five seconds. Then, the monitor turns into a gray screen. At anytime during these steps, the subjects can enter their rating using a keyboard. Finally, the monitor turns into (or stays) gray for one second for a break and then the next pair is shown. The viewing order of the stimuli is set random for each subject. The arrangement of the reference image or the gamut-reduced image (i.e., left or right side) is also randomized for each pair. At the beginning of the test session, three dummy pairs are shown for stabilization, which are also different from the test stimuli.

### C. Fitting Objective Metric

In order to approximate the subjective score of the color difference due to gamut reduction in an objective manner, we employ the color extension of the structural similarity index (*cssim*) [49], [50], which can effectively measure perceptually significant structural differences due to gamut reduction between two color images. The preliminary study [1] shows that it performs best with the highest accuracy among eight commonly-used objective color difference metrics [51], [52], [53], [54], [55], [56], [57].

For each pair of the reference and gamut-reduced image, we measure the *cssim* score. The score is further fitted to the MOS by a monotonic nonlinear function as described in [58]:

$$f(x) = \frac{\alpha}{1 + 10^{\beta(\gamma-x)}}, \quad (1)$$

where the fitted values of the parameters are  $\alpha = 2$ ,  $\beta = -3.5$ , and  $\gamma = 1.9$ . The result of fitting for the training dataset is shown in Fig. 2. In order to evaluate the prediction performance, we obtain MOS for the validation dataset from 20 subjects by following the same procedure described in Section III-B2. The Pearson correlation coefficients (PCCs) between the ground truth MOS and predicted MOS using the fitting function are 0.92 and 0.80 for the training and validation sets, respectively. Therefore, we calculate the perceptual difference in Algorithm 1 as

$$d_n = PD(I_n, I_0) = f(cssim(I_n, I_0)). \quad (2)$$

### D. Validation

We validate the framework by applying it to a simple content selection task. As mentioned in Section I, using representative contents is crucial to draw reliable conclusion in studies on QoE of WCG images. In the task, the main objective is to select representative images that have diverse behaviors in terms of the perceptual difference due to successive gamut reduction. We use the framework to obtain predicted perceptual differences due to gamut reduction to the two target gamuts (Rec.709 and Toy) as two-dimensional features characterizing the 24 candidate images in the validation

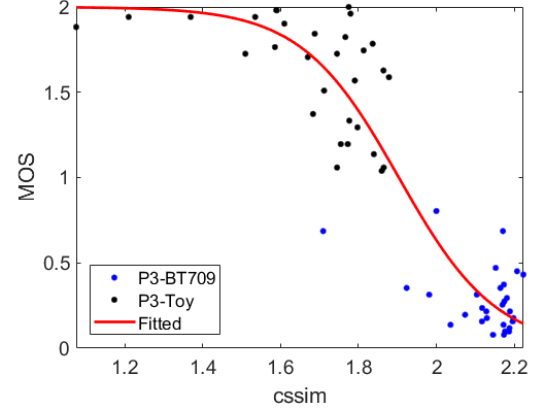


Fig. 2. Fitted sigmoid function to predict MOS of perceptual difference using *cssim*. The ground truth and predicted MOS are shown as dots and a red line, respectively.

dataset. Then, the  $k$ -means clustering algorithm is applied to the predicted perceptual differences. The value of  $k$  determines the number of representative clusters for content selection, which should be chosen by the user according to the purpose of content selection. In this experiment, we set the value of  $k$  to five based on the distribution of the images in terms of the predicted perceptual differences. One image for each cluster is randomly selected to construct a representative image set, which maximizes the coverage of the feature space. For comparison, we also apply a random selection method where five images are selected randomly in the same dataset.

The result for each selection method is shown in Fig. 3. It can be seen that the selected images by our framework in Fig. 3a are more spread than the randomly selected images. In Fig. 3b, however, the selected images are biased to the upper-side of the feature space. In this case, images having small perceptual difference by severe gamut reduction are not considered, and the obtained image set cannot be said to be representative. Fig. 4 shows two example images (marked in Fig. 3a) in different gamuts. In Fig. 4a, as predicted, large perceptual differences are observed for both gamut-reduced images compared to the reference P3 image, i.e., the overall color of the scene and the green laser lights at the top area. On the contrary, Fig. 4b hardly shows any difference between gamut-reduced ones, which is also predicted in Fig. 3a. By selecting images with diverse characteristics, a representative dataset can be constructed by our framework.

We also evaluate robustness of content selection with our framework. For each of the two methods (random selection and our framework), the selection task is repeated two times to obtain two sets of selected images, and the PCC between the MOSs of the two sets is measured. We consider that a high value of PCC by a selection method represents a high level of robustness of the method, because it means that the characteristics of the selected images are consistent regardless of repetition or random effects. We repeat the procedure 100 times. Much higher PCC values are obtained by our framework than random selection (0.83 vs. 0.15 on average), which is found to be statistically significant via a t-test,  $t(137.1) =$

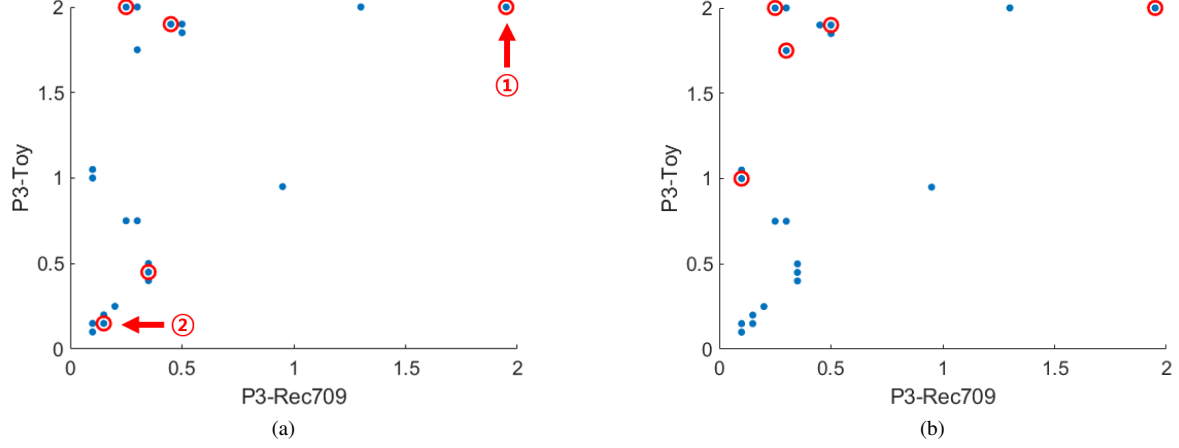


Fig. 3. Example of content selection with (a) our framework and (b) random selection. The x- and y-axis are predicted perceptual differences of images from the P3 to Rec.709 and Toy gamuts, respectively. Among the data points shown as blue dots, the selected images are marked with red circles. Note that the lower-right area of each figure is empty because as the gamut is reduced more, the perceptual difference becomes larger, thus the value of the y-axis would be always bigger than that of the x-axis.

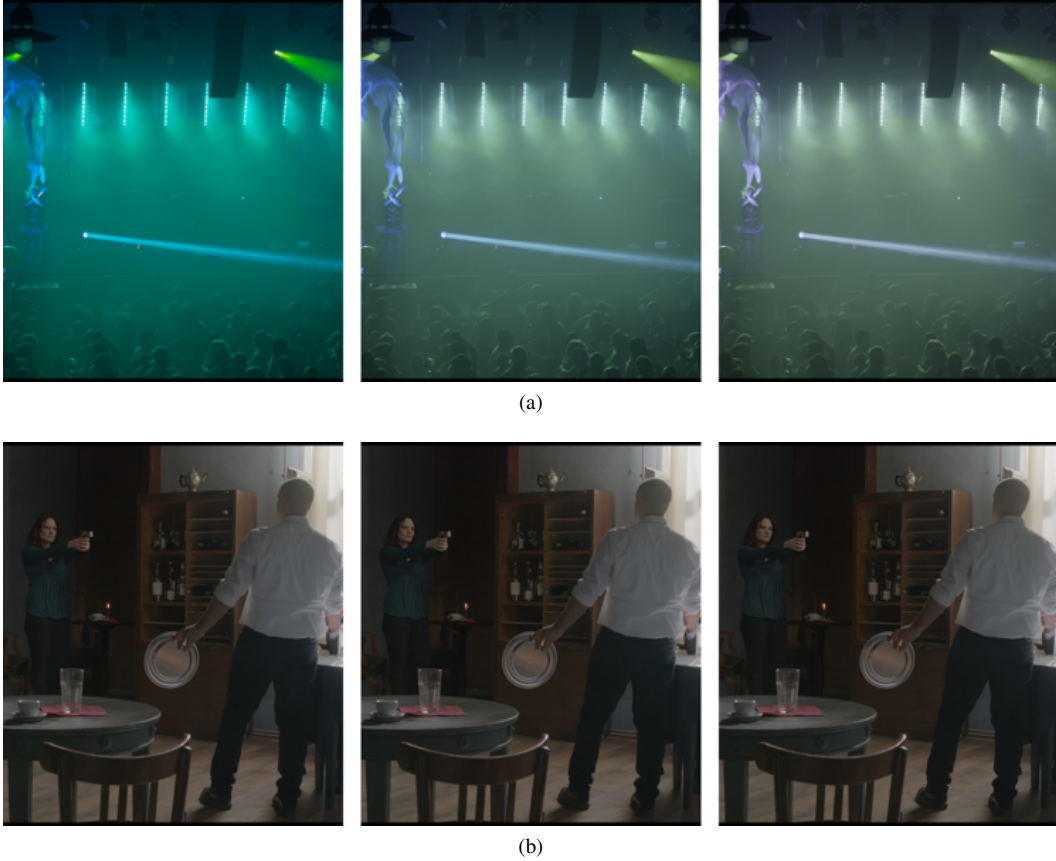


Fig. 4. Selected images corresponding (a) ① and (b) ② of Fig. 3a in the P3, Rec.709, and Toy gamuts (left, middle, and right panels, respectively).

21.1,  $p < 0.001^3$ .

#### IV. APPLICATION TO WCG DATASET CHARACTERIZATION

In this section, we apply the proposed framework to characterization of WCG image datasets. We describe dataset

characterization criteria and analyze existing WCG datasets based on them. Characterizing datasets helps an experimenter to determine or construct a suitable dataset for studies related to QoE of WCG contents.

##### A. Dataset Characterization Criteria

By extending the dataset characterization criteria presented in [37], we propose to measure four statistics of perceptual

<sup>3</sup>The statistical significance of higher PCC values by our framework is obtained in all cases with  $k$  from 2 to 10.



difference measured by the framework as follows. In [37], three statistics of various characteristics extracted from images or videos in the dataset are proposed. They are two criteria measuring the coverage and uniformity in each dimension, and a multidimensional coverage criterion. In addition to these, we also consider the multidimensional uniformity. Note that we normalize the perceptual differences in each dimension to scale the span of the criteria within  $[0, 1]$ , i.e.,  $\tilde{d}_i = d_i/s_i$ , where  $s_i$  is a normalization factor that is equal to the maximum possible value of MOS ( $s_i = 2$  in our case) because the minimum value of MOS is zero.

1) *Coverage*: To quantify how wide the range of the perceptual differences covered by the images of a dataset, we measure the difference between the smallest and largest perceptual difference values of the images. Specifically, coverage  $C_i$  for gamut space  $i$  is calculated as

$$C_i = \max(z_i) - \min(z_i), \quad (3)$$

where  $z_i$  is a set of the normalized perceptual differences  $\tilde{d}_i$  of all images in the dataset when the gamut is reduced to target gamut  $i$  from the reference gamut ( $i = 1, \dots, N$ ). The maximum value of  $C_i$  is obtained when the dataset contains images corresponding to both no-difference (MOS = 0) and clear difference (MOS = 2) for the  $i$ th target gamut. In other words, one image has less or no colors outside the  $i$ th gamut space so that it does not cause perceptual difference by gamut reduction, but the other image contains lots of colors outside the space and thus its perceptual difference can be clearly observed by gamut reduction.

2) *Total Coverage*: This is the relative area occupied by the data points in the space of perceptual differences. It is similar to  $C_i$ , but considers the interaction of different dimensions in  $Z = \{z_1, z_2, \dots, z_N\}$ . It is calculated as follows:

$$C_{total} = \sqrt[N]{\int \text{convex}(Z)}, \quad (4)$$

where  $\text{convex}(Z)$  returns the convex hull for  $N$ -dimensional vectors in  $Z$ .  $C_{total}$  becomes the largest when the dataset consists of images having the maximum coverage of perceptual difference for all target gamuts. Using a dataset having a large value of  $C_{total}$  in an experiment implies that images having extreme perceptual characteristics (i.e., both severe and little perceptual differences) under gamut change are employed.

3) *Uniformity*: While the above coverage measures consider the range of perceptual differences observed in the images, uniformity measures how evenly the perceptual differences are distributed within the range. For this, we use the information entropy, which is popularly used to measure the uniformity of a distribution. In other words, we construct the histogram of  $z_i$ , and then compute its entropy as follows in order to quantify the uniformity of the distribution of the perceptual differences.

$$U_i = - \sum_{k=1}^B p_{i,k} \log_B p_{i,k}, \quad (5)$$

where  $B$  is the number of bins of the histogram and  $p_{i,k}$  is the ratio of the images of which perceptual differences are in the

range of the  $k$ th bin. The uniformity has the largest value of 1 when the perceptual differences of the dataset are uniformly distributed. It becomes low when the dataset contains images having similar perceptual differences, and reaches 0 when the perceptual differences are the same for all images.

4) *Total Uniformity*: This measures the uniformity of perceptual differences over the whole dimensions of reduced target gamuts. In this case, we compute the  $N$ -dimensional histogram of  $Z$  and its entropy, i.e.,

$$U_{total} = - \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^B q_{i,k} \log_B q_{i,k}, \quad (6)$$

where  $B$  is the number of bins for each dimension of the histogram and  $q_{i,k}$  is the normalized count in the  $k$ th bin (normalized over the whole dimension). It becomes the largest value (i.e., 1) when a dataset contains diverse images in terms of perceptual differences and the perceptual differences are uniformly distributed over all target gamuts. On the other hand, it has the lowest value of 0 when the dataset contains images that show the same amount of perceptual difference for all target gamuts. A dataset having a large value of  $U_{total}$  is beneficial to conduct experiments with images having diverse perceptual characteristics under gamut change.

## B. Analysis of Existing Datasets

We analyze the two existing WCG datasets, HdM and Arri, in terms of the four criteria described above<sup>4</sup>. In this experiment, we collect 38 and 11 images from each dataset, respectively. We use the perceptual difference for the 49 images due to successive gamut reduction from the reference P3 gamut to the Rec.709 and Toy gamuts as in Section III-C. We then measure the four criteria of the two WCG datasets. For (total) uniformity, we use 10 bins for each dimension of the histograms (i.e.,  $B = 10$ ). The measured criteria are summarized in TABLE II. In addition, the distributions of the perceptual difference for the two datasets are shown in Fig. 5.

First, the coverages of the two datasets have different behaviors depending on the target gamuts. The perceptual differences of the images in the HdM dataset cover over about a half of the scale for both target gamuts as shown in Fig. 5a. For the case of gamut reduction to Toy, the perceptual difference is biased to large values because most images of HdM contain many pixels with highly saturated colors, which produces large perceptual difference when the gamut is reduced. On the contrary, pixels with highly saturated color are few in the images of the Arri dataset, so the coverage criterion for Rec.709 is low while that for Toy is high as shown in Fig. 5b.

Similarly to the results of the dimension-wise coverage criterion, the HdM dataset has a medium level of total coverage of perceptual differences, showing the convex hull covering almost the upper-half area in Fig. 5a. On the other hand, although the coverage value for the Toy gamut is large as shown in Fig. 5b, the total coverage of the Arri dataset is small due to the extremely low coverage for Rec.709. Note that  $z_{Toy}$

<sup>4</sup>These are the only publicly available datasets that support Rec.2020.

TABLE II  
RESULTS OF WCG DATASET CHARACTERIZATION

Criteria	HdM			Arri		
	Toy	Rec.709	Total	Toy	Rec.709	Total
Coverage	0.512	0.647	0.412	0.933	0.017	0.093
Uniformity	0.707	0.509	0.550	0.713	0.000	0.357

would be always higher than  $z_{709}$  for the same image because the details of color are more distorted in Toy, so the practical maximum possible value of total coverage is 0.707 ( $=\sqrt{0.5}$ ).

In terms of uniformity, the perceptual differences caused by the large gamut difference (i.e., the case of Toy) are quite uniformly distributed for both datasets. For the small gamut reduction (to Rec.709), the perceptual differences of the HdM dataset are slightly biased to low values. The perceptual differences of the Arri dataset are extremely biased, so all data points are allocated in a single bin and the uniformity is zero.

In the case of total uniformity, there exist differences between the two datasets. The perceptual differences of the HdM dataset are quite uniformly distributed on the two-dimensional space in Fig. 5a, although the data points are slightly biased to the upper region (where large perceptual differences occur due to large gamut reduction). For the Arri dataset, the perceptual differences are biased to the left-side in Fig. 5b, so the total uniformity becomes low.

Overall, each of the two datasets has its own strengths and limitations in a complementary manner. HdM has a relatively small coverage of  $z_{Toy}$ , while Arri has limited characteristics in the Rec.709 gamut. For example, if the Arri dataset is used for an experiment involving gamut changes, the experiment would draw biased conclusion for small gamut difference. Based on this understanding, one can choose either of the two datasets for particular research problems; for instance, the Arri dataset could be more effective for the experiments that focus on large gamut difference. Furthermore, one can obtain an enhanced dataset by supplementing one of the two datasets with particular contents having characteristics desired for the given objective.

## V. APPLICATION TO EVALUATION OF GAMUT MAPPING ALGORITHMS

In this section, we present another practical application of the proposed framework, which is the problem of evaluation of GMAs. In this scenario, the proposed framework plays a role to select image contents used for performance comparison of different GMAs. We demonstrate the reliability of the framework for selection of representative contents for fair comparison.

### A. Scenario

The main goal of the scenario is to benchmark performance of GMAs. Each GMA is applied to a set of source image contents having wide gamuts, and its performance is measured by an objective quality metric in terms of perceptual color information loss in the gamut-reduced images in comparison

to the original ones. Here, which image dataset is used is an important issue. For instance, if images that do not have color profiles challenging enough to reveal distinguished gamut mapping performance, the GMAs may be evaluated to perform similarly, which may not be the case if challenging images are included. Therefore, careful selection of the images is required to obtain unbiased benchmarking results, for which the proposed framework can be used. Therefore, our objective is to evaluate the reliability of the benchmarking results between different source content selection methods.

We limit the number of GMAs for comparison to two in order to validate the effectiveness of the proposed framework clearly rather than to present extensive benchmarking of many GMAs. One is the state-of-the-art gamut reduction algorithm [7] that adaptively modifies local contrast of pixels residing outside of the target gamut based on the Retinex theory [8]. For the other one, we use the gamut compression algorithm [6] that maps the entire color of the source image inside the target gamut in the CIE 1931 space.

To evaluate the performance of gamut mapping, we use the color image difference (CID) [35] that predicts perceptual color difference between the reference and gamut-reduced image, which is used to evaluate performance of the gamut reduction algorithm in [7]. As the main objective of conducting the scenario, we focus on the reliability and robustness of test results with representative contents selected by our framework. First, the selected dataset should sufficiently cover diverse gamut characteristics so that it is representative. Second, in terms of robustness, experiments with content selection followed by the same procedure should produce consistent results and conclusions regardless of repetition.

### B. Content Selection

The pool of candidate source images consists of half-HD ( $960 \times 1080$  pixels) WCG images from both the HdM and Arri datasets. After excluding images containing no or too few pixels in WCG (outside the Rec.709 gamut) from the data used in Section IV-B, 35 candidate images are used. The reference gamut is Rec.2020, and we use three target gamuts for gamut mapping: P3, Rec.709, and Toy.

The proposed framework is applied to select representative images from the pool. As described in Section III-C, each candidate image is represented by a two-dimensional perceptual feature vector. Then, the  $k$ -means clustering algorithm with  $k = 3$  is used to group them into three clusters, from each of which three images are randomly selected. For comparison, content selection using an existing content feature, colorfulness [53], is also conducted. It measures the variety and intensity of colors in an image. The colorfulness features computed for the candidate images are also clustered into three groups and three images are randomly chosen from each group. These content selection procedures are repeated 100 times with different random seeds.



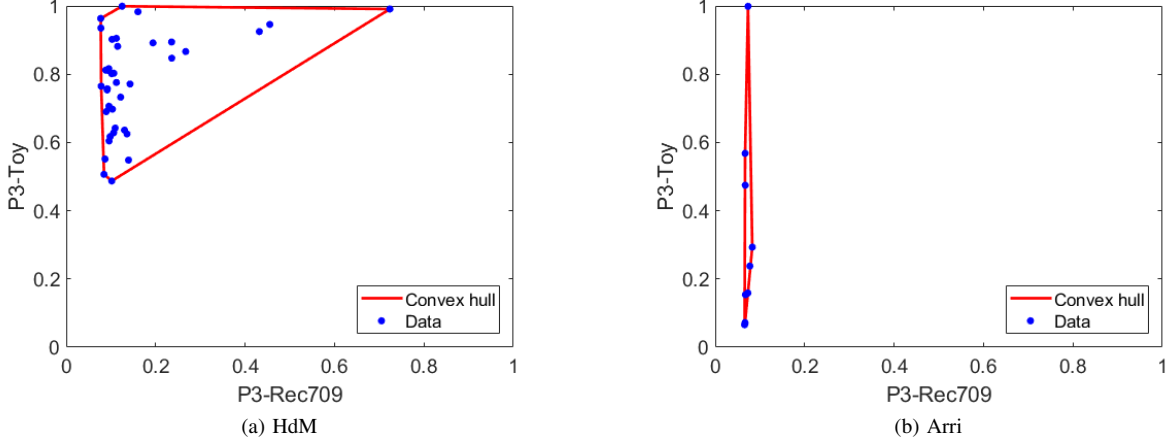


Fig. 5. Measured perceptual differences and corresponding convex hulls for the (a) HdM and (b) Arri datasets.

### C. Evaluation

In order to compare the two GMAs, we define CID gain  $g_t$  for target gamut  $t$  and for a source image as

$$g_t = \text{CID}(I_0, \text{GC}(I_0, t)) - \text{CID}(I_0, \text{GR}(I_0, t)), \quad (7)$$

where  $I_0$  is the reference image, and  $\text{GC}(I_0, t)$  and  $\text{GR}(I_0, t)$  are the gamut-compressed and gamut-reduced versions of  $I_0$ , respectively.  $g_t$  becomes positive when the gamut reduction algorithm performs better than the gamut compression algorithm, and its absolute value indicates the degree of the performance difference.

Using the CID gains for 100 repetitions, the two content selection methods are compared with respect to two aspects: robustness and representativeness. First, a content selection method is considered to be robust when the CID gains remain consistent, i.e., the averages and standard deviations of the CID gains over the selected images are similar across the repetitions. Second, a dataset of images chosen by a content selection method is regarded as being representative if the images have diverse color characteristics. Thus, the CID gains lie in a wide range, resulting in a large average and standard deviation over the images.

### D. Results

Fig. 6 shows the average and standard deviation of CID gains for the selected images with respect to the target gamut and selection method. In all cases, the average CID gains are positive, which indicates that the gamut reduction algorithm produces gamut-reduced images with smaller difference from the reference ones compared to the gamut compression algorithm. When the three target gamuts are compared, a smaller gamut yields larger CID gains because more color distortion is introduced by the gamut compression algorithm than the gamut reduction algorithm as the gamut difference becomes larger.

The two selection methods show clearly distinct results. First, the average and standard deviation of the CID gains appear more similar across 100 trials when the proposed

framework is used, particularly when the target gamut is small. In order to statistically assess this, we conduct one-sided F-tests under the null hypothesis that the two populations (one for the proposed framework and the other for the method using colorfulness) of the average (or standard deviation) values of the CID gains have the same variance. The results are shown in TABLE III, which confirms that the cases involving large gamut changes show statistically significant difference (i.e., Rec.709 and Toy for the average and Toy for the standard deviation). Note that for P3, the gamut difference from Rec.2020 is small, so the average and standard deviation of the CID gains are also small. These results demonstrate that the selection method has an impact on the results of GMA comparison, where content selection using the proposed framework provides improved robustness.

Second, on average, the average and standard deviation values are larger for the case using the proposed framework than for the case using colorfulness. Since many images in the pool are not challenging for GMAs as shown in Section IV-B, for which the CID gain is small, a larger average or standard deviation value indicates a more representative dataset. We perform one-sided t-tests under the null hypothesis that the two populations of the average (or standard deviation) values of the CID gains have the same mean. As shown in TABLE III, the null hypothesis is rejected in all cases, indicating that the average and standard deviation values are significantly larger for our method. This confirms representativeness of the dataset obtained using our method and, consequently, reliability of the results of the benchmarking.

For comparison, we provide further results using selection features other than colorfulness. We use two no-reference color quality metrics: contrast enhancement based contrast-changed image quality measure (CEIQ) [59] and accelerated screen image quality evaluator (ASIQE) [60]. The former is a metric based on a learned support vector machine using multiple features estimating contrast distortion, while the latter assesses image quality considering four types of quality features consisting of picture complexity, screen content statistics, global brightness quality, and sharpness of details. We conduct

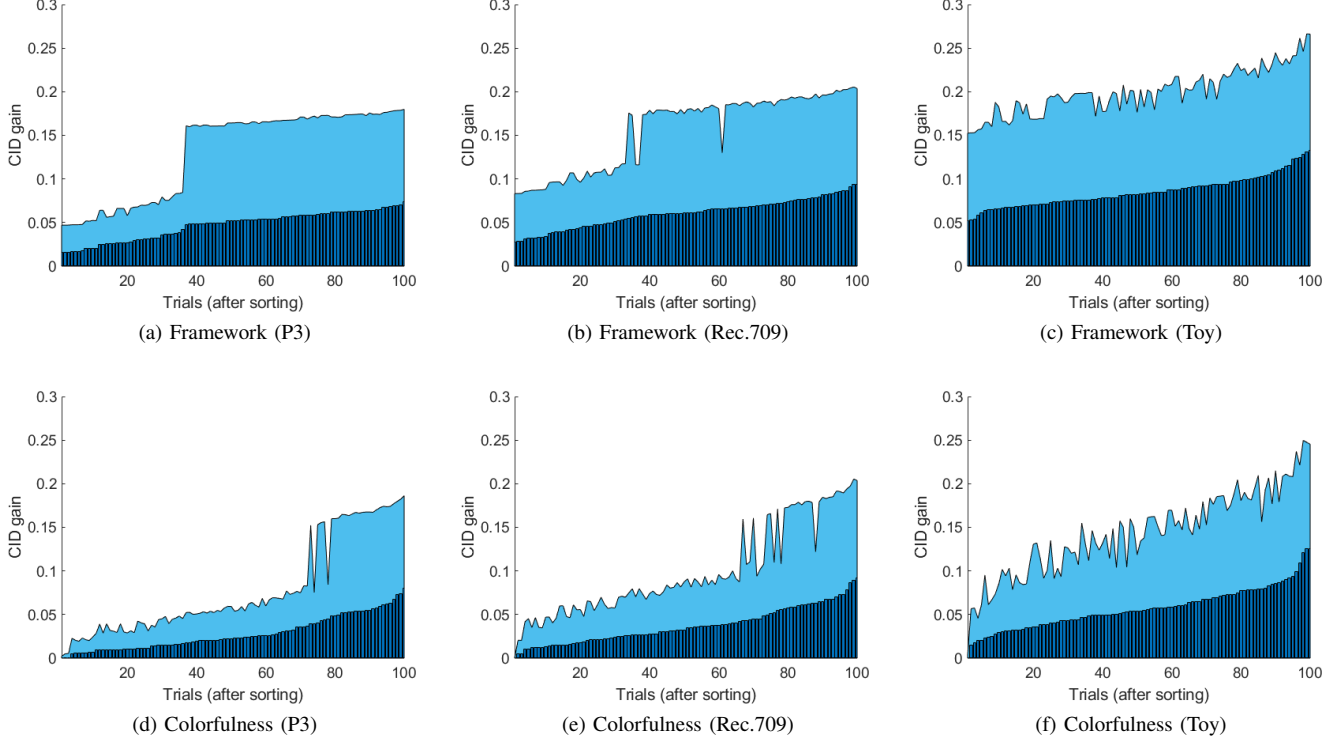


Fig. 6. CID gains for the images selected by the method using the proposed framework or the one using colorfulness. The average and standard deviation values of the CID gains over the selected images are represented by the bars with dark color and the shaded area, respectively.

TABLE III

RESULT OF THE STATISTICAL TESTS COMPARING THE SELECTION METHOD USING THE PROPOSED FRAMEWORK AND THE ONE USING COLORFULNESS. THE DEGREES OF FREEDOM OF T-TESTS ON AVERAGE AND STANDARD DEVIATION USING THE WELCH-SATTERTHWAITE EQUATION ARE 181.8 AND 134.8, RESPECTIVELY. STATISTICAL SIGNIFICANCE (BONFERRONI-CORRECTED FOR MULTIPLE COMPARISON) IS MARKED IN BOLD.

Target gamut	Average		Standard deviation	
	Statistics	<i>p</i> -value	Statistics	<i>p</i> -value
<b>F-tests</b>				
P3	$F = 0.72$	0.050	$F = 0.94$	0.385
Rec.709	$F = 0.63$	<b>0.012</b>	$F = 0.72$	0.053
Toy	$F = 0.54$	<b>0.001</b>	$F = 0.19$	<b>&lt; 0.001</b>
<b>t-tests</b>				
P3	$t = 7.40$	<b>&lt; 0.001</b>	$t = 6.43$	<b>&lt; 0.001</b>
Rec.709	$t = 8.79$	<b>&lt; 0.001</b>	$t = 7.50$	<b>&lt; 0.001</b>
Toy	$t = 9.65$	<b>&lt; 0.001</b>	$t = 9.43$	<b>&lt; 0.001</b>

statistical tests comparing the CID gains obtained by our framework and the method using either CEIQ or ASIQE. The results are shown in TABLE IV. Similar to the results in TABLE III using colorfulness, statistical significance is observed for F-tests in the cases of the large gamut change (i.e. between Rec.709 and Toy) and for t-tests in all cases. Thus, our framework can effectively select representative contents reliably compared to the methods using these image quality metrics.

TABLE IV

RESULT OF THE STATISTICAL TESTS COMPARING THE SELECTION METHOD USING THE PROPOSED FRAMEWORK AND THE ONES USING CEIQ AND ASIQE. STATISTICAL SIGNIFICANCE IS MARKED IN BOLD.

Target gamut	Average		Standard deviation	
	Statistics	<i>p</i> -value	Statistics	<i>p</i> -value
<b>F-tests (CEIQ)</b>				
P3	$F = 0.78$	0.103	$F = 0.83$	0.170
Rec.709	$F = 0.66$	<b>0.020</b>	$F = 0.57$	<b>&lt; 0.001</b>
Toy	$F = 0.49$	<b>&lt; 0.001</b>	$F = 0.14$	<b>&lt; 0.001</b>
<b>t-tests</b>				
P3	$t = 10.73$	<b>&lt; 0.001</b>	$t = 10.44$	<b>&lt; 0.001</b>
Rec.709	$t = 12.04$	<b>&lt; 0.001</b>	$t = 11.80$	<b>&lt; 0.001</b>
Toy	$t = 12.23$	<b>&lt; 0.001</b>	$t = 12.78$	<b>&lt; 0.001</b>
<b>F-tests (ASIQE)</b>				
P3	$F = 0.76$	0.082	$F = 0.92$	0.340
Rec.709	$F = 0.57$	<b>0.002</b>	$F = 0.62$	<b>0.009</b>
Toy	$F = 0.44$	<b>&lt; 0.001</b>	$F = 0.11$	<b>&lt; 0.001</b>
<b>t-tests</b>				
P3	$t = 9.77$	<b>&lt; 0.001</b>	$t = 9.27$	<b>&lt; 0.001</b>
Rec.709	$t = 10.81$	<b>&lt; 0.001</b>	$t = 11.00$	<b>&lt; 0.001</b>
Toy	$t = 10.90$	<b>&lt; 0.001</b>	$t = 13.24$	<b>&lt; 0.001</b>

## VI. CONCLUSION

We proposed a content characterization method for a WCG image content and evaluated it in practical applications. The main idea was to obtain perceptual color differences due to successive gamut reduction as content characteristics for

the WCG content. As one of the practical use cases of the framework, we analyzed existing datasets by measuring dataset characterization criteria on the WCG characteristics. Four criteria consisting of coverage, total coverage, uniformity, and total uniformity effectively characterized WCG datasets. In addition, we validated WCG content characteristics as a content selection feature in a GMA benchmarking scenario. Using the framework, we were able to select representative WCG contents, and draw robust and reliable benchmarking results.

In the future, the proposed framework can be improved in several ways. First, we employed *cssim* for objective quality assessment due to its superiority. If metrics that perform better than *cssim* are developed in the future, e.g., deep learning-based methods, our framework could benefit from employing such improved metrics. Second, the scope of the framework could be extended to video contents by considering the temporal dimension of color perception.

## REFERENCES

- [1] J. Lee, T. Vigier, P. Le Callet, and J.-S. Lee, "A perception-based framework for wide color gamut content selection," in *Proceedings of IEEE International Conference on Image Processing*, Oct. 2018, pp. 709–713.
- [2] ITU, "ITU-R BT.709-6. Parameter values for the HDTV standards for production and international programme exchange," Tech. Rep., 2015.
- [3] —, "ITU-R BT.2020-2. Parameter values for ultra-high definition television systems for production and international programme exchange," Tech. Rep., 2015.
- [4] T. Smith and J. Guild, "The C.I.E. colorimetric standards and their use," *Transactions of the Optical Society*, vol. 33, no. 3, pp. 73–134, 1931.
- [5] C. Chinnock, "The status of wide color gamut UHD-TVs," Tech. Rep., 2016.
- [6] CIE, "Guidelines for the evaluation of gamut mapping algorithms," Tech. Rep., 2004.
- [7] S. W. Zamir, J. Vazquez-Corral, and M. Bertalmio, "Gamut mapping in cinematography through perceptually-based contrast modification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 3, pp. 490–503, 2014.
- [8] E. H. Land and J. J. McCann, "Lightness and retinex theory," *Journal of the Optical Society of America*, vol. 61, no. 1, pp. 1–11, 1971.
- [9] K. Gu, S. Wang, H. Yang, W. Lin, G. Zhai, X. Yang, and W. Zhang, "Saliency-guided quality assessment of screen content images," *IEEE Transactions on Multimedia*, vol. 18, no. 6, pp. 1098–1110, Jun. 2016.
- [10] K. Gu, S. Wang, G. Zhai, S. Ma, X. Yang, W. Lin, W. Zhang, and W. Gao, "Blind quality assessment of tone-mapped images via analysis of information, naturalness, and structure," *IEEE Transactions on Multimedia*, vol. 18, no. 3, pp. 432–443, Mar. 2016.
- [11] Z. Fan, T. Jiang, and T. Huang, "Active sampling exploiting reliable informativeness for subjective image quality assessment based on pairwise comparison," *IEEE Transactions on Multimedia*, vol. 19, no. 12, pp. 2720–2735, Dec. 2017.
- [12] X. Min, K. Gu, G. Zhai, J. Liu, X. Yang, and C. W. Chen, "Blind quality assessment based on pseudo-reference image," *IEEE Transactions on Multimedia*, vol. 20, no. 8, pp. 2049–2062, Aug. 2018.
- [13] P. G. Freitas, W. Y. L. Akamine, and M. C. Q. Farias, "No-reference image quality assessment using orthogonal color planes patterns," *IEEE Transactions on Multimedia*, vol. 20, no. 12, pp. 3353–3360, Dec. 2018.
- [14] M. C. Stone, W. B. Cowan, and J. C. Beatty, "Color gamut mapping and the printing of digital color images," *ACM Transactions on Graphics*, vol. 7, no. 4, pp. 249–292, Oct. 1988.
- [15] G. M. Murch and J. M. Taylor, "Color in computer graphics: Manipulating and matching color," in *Advances in Computer Graphics V*. Springer, 1989, pp. 19–47.
- [16] F. Ebner and M. D. Fairchild, "Gamut mapping from below: Finding minimum perceptual distances for colors outside the gamut volume," *Color Research & Application*, vol. 22, no. 6, pp. 402–413, 1997.
- [17] J. Morovic and M. R. Luo, "Gamut mapping algorithms based on psychophysical experiment," in *Proceedings of the Color and Imaging Conference*, vol. 1997, no. 1, 1997, pp. 44–49.
- [18] S. O. Naoya Katoh, Masahiko Ito, "Three-dimensional gamut mapping using various color difference formulae and color spaces," *Journal of Electronic Imaging*, vol. 8, no. 4, pp. 365–379, 1999.
- [19] J. Morovic and Y. Wang, "A multi-resolution, full-colour spatial gamut mapping algorithm," in *Proceedings of the Color and Imaging Conference*, vol. 2003, no. 1, 2003, pp. 282–287.
- [20] P. Zolliker and K. Simon, "Adding local contrast to global gamut mapping algorithms," in *Proceedings of the Conference on Colour in Graphics, Imaging, and Vision*, vol. 2006, no. 1, 2006, pp. 257–261.
- [21] I. Farup, C. Gatta, and A. Rizzi, "A multiscale framework for spatial gamut mapping," *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2423–2435, 2007.
- [22] P. Zolliker and K. Simon, "Retaining local image information in gamut mapping algorithms," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 664–672, Mar. 2007.
- [23] Ø. Kolås and I. Farup, "Efficient hue-preserving and edge-preserving spatial color gamut mapping," in *Proceedings of the Color and Imaging Conference*, 2007, pp. 207–212.
- [24] A. Alsam and I. Farup, "Spatial colour gamut mapping by orthogonal projection of gradients onto constant hue lines," in *Advances in Visual Computing*, 2012, pp. 556–565.
- [25] C. Gatta and I. Farup, "Gamut mapping in RGB colour spaces with the iterative ratios diffusion algorithm," in *Proceedings of the IS&T International Symposium on Electronic Imaging*, 2017, pp. 12–20.
- [26] F. Dugay, I. Farup, and J. Y. Hardeberg, "Perceptual evaluation of color gamut mapping algorithms," *Color Research & Application*, vol. 33, no. 6, pp. 470–476, 2008.
- [27] CIE, "Recommendations on uniform color spaces, color-difference equations, psychometric color terms," Tech. Rep., 1978.
- [28] X. Zhang and B. A. Wandell, "A spatial extension of CIELAB for digital color image reproduction," in *SID International Symposium Digest of Technical Papers*, vol. 27, 1996, pp. 731–734.
- [29] M. D. Fairchild and G. M. Johnson, "The iCAM framework for image appearance, image differences, and image quality," *Journal of Electronic Imaging*, vol. 13, pp. 126–138, Jan. 2004.
- [30] G. Hong and M. R. Luo, "Perceptually-based color difference for complex images," in *Proceedings of the Congress of the International Colour Association*, vol. 4421, 2002, pp. 618–622.
- [31] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
- [32] N. Bonnier, F. Schmitt, H. Brettel, and S. Berche, "Evaluation of spatial gamut mapping algorithms," in *Proceedings of Color and Imaging Conference*, 2006, pp. 56–61.
- [33] J. Y. Hardeberg, E. Bando, and M. Pedersen, "Evaluating colour image difference metrics for gamut-mapped images," *Coloration Technology*, vol. 124, no. 4, pp. 243–253, 2008.
- [34] M. Pedersen and J. Y. Hardeberg, "A new spatial hue angle metric for perceptual image difference," in *In Proceedings of the Computational Color Imaging Workshop*, 2009, pp. 81–90.
- [35] I. Lissner, J. Preiss, P. Urban, M. S. Lichtenauer, and P. Zolliker, "Image-difference prediction: From grayscale to color," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 435–446, 2013.
- [36] D. Darcy, E. Gitterman, A. Brandmeyer, S. Daly, and P. Crum, "Physiological capture of augmented viewing states: objective measures of high-dynamic-range and wide-color-gamut viewing experiences," in *Proceedings of the IS&T International Symposium on Human Vision and Electronic Imaging*, 2016, pp. HVEI126:1–9.
- [37] S. Winkler, "Analysis of public image and video databases for quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 616–625, 2012.
- [38] F. Zhang, F. M. Moss, R. Baddeley, and D. R. Bull, "BVI-HD: A video quality database for HEVC compressed and texture synthesized content," *IEEE Transactions on Multimedia*, vol. 20, no. 10, pp. 2620–2630, Oct. 2018.
- [39] A. Mackin, F. Zhang, and D. R. Bull, "A study of high frame rate video formats," *IEEE Transactions on Multimedia*, vol. 21, no. 6, pp. 1499–1512, Jun. 2019.
- [40] M. H. Pinson, M. Barkowsky, and P. Le Callet, "Selecting scenes for 2D and 3D subjective video quality tests," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, pp. 50:1–12, Aug. 2013.
- [41] L. Krasula, K. Fliegel, P. Le Callet, and M. Klíma, "Objective evaluation of naturalness, contrast, and colorfulness of tone-mapped images," in *Proceedings of the Applications of Digital Image Processing XXXVII*, vol. 9217, 2014, pp. 9217D:1–10.
- [42] P. Paudyal, J. Gutiérrez, P. Le Callet, M. Carli, and F. Battisti, "Characterization and selection of light field content for perceptual assessment,"

in *Proceedings of the International Conference on Quality of Multimedia Experience*, 2017, pp. 1–6.

- [43] M. Narwaria, C. Mantel, M. Perreira Da Silva, P. Le Callet, and S. Forchhammer, “An objective method for high dynamic range source content selection,” in *Proceedings of the International Workshop on Quality of Multimedia Experience*, 2014, pp. 13–18.
- [44] L. Krasula, M. Narwaria, K. Fliegel, and P. Le Callet, “Preference of experience in image tone-mapping: dataset and framework for objective measures comparison,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 64–74, 2017.
- [45] J. Froehlich, S. Grandinetti, B. Eberhardt, S. Walter, A. Schilling, and H. Brendel, “Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays,” in *Proceedings of the Digital Photography X*, vol. 9023, 2014, pp. 90 230X:1–10.
- [46] S. W. Zamir, J. Vazquez-Corral, and M. Bertalmío, “Gamut extension for cinema,” *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1595–1606, 2017.
- [47] J.-S. Lee, “On designing paired comparison experiments for subjective multimedia quality assessment,” *IEEE Transactions on Multimedia*, vol. 16, no. 2, pp. 564–571, 2014.
- [48] ITU, “ITU-R BT.500-13. Methodology for the subjective assessment of the quality of television pictures,” Tech. Rep., 2012.
- [49] B. Ortiz-Jaramillo, A. Kumcu, and W. Philips, “Evaluating color difference measures in images,” in *Proceedings of the International Conference on Quality of Multimedia Experience*, 2016, pp. 1–6.
- [50] A. Toet and M. P. Lucassen, “A new universal colour image fidelity metric,” *Displays*, vol. 24, no. 4, pp. 197–207, 2003.
- [51] G. Sharma, W. Wu, and E. Daa, “The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations,” *Color Research & Application*, vol. 30, pp. 21–30, 2004.
- [52] X. Zhang and B. A. Wandell, “A spatial extension of CIELAB for digital color-image reproduction,” *Journal of the Society for Information Display*, vol. 5, no. 1, pp. 61–63, 1997.
- [53] D. Hasler and S. Süsstrunk, “Measuring colourfulness in natural images,” in *Proceedings of the IST/SPIE Electronic Imaging 2003: Human Vision and Electronic Imaging VIII*, vol. 5007, no. 19, 2003, pp. 87–95.
- [54] M. H. Pinson and S. Wolf, “A new standardized method for objectively measuring video quality,” *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp. 312–322, 2004.
- [55] G. M. Johnson, “Using color appearance in image quality metrics,” in *Proceedings of the Second International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2006.
- [56] C. H. Chou and K. C. Liu, “A fidelity metric for assessing visual quality of color images,” in *Proceedings of the 16th International Conference on Computer Communications and Networks*, 2007, pp. 1154–1159.
- [57] U. Rajashekar, Z. Wang, and E. P. Simoncelli, “Quantifying color image distortions based on adaptive spatio-chromatic signal decompositions,” in *Proceedings of the 16th IEEE International Conference on Image Processing*, 2009, pp. 2213–2216.
- [58] ITU, “ITU-T J.149. Method for specifying accuracy and cross-calibration of Video Quality Metrics (VQM),” Tech. Rep., 2004.
- [59] J. Yan, J. Li, and X. Fu, “No-reference quality assessment of contrast-distorted images using contrast enhancement,” *arXiv preprint arXiv:1904.08879*, pp. 1–15, 2019.
- [60] K. Gu, J. Zhou, J. Qiao, G. Zhai, W. Lin, and A. C. Bovik, “No-reference quality assessment of screen content pictures,” *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 4005–4018, 2017.



**Junghyuk Lee** received his B.S. degree from the School of Integrated Technology at Yonsei University, Korea, in 2015, where he is currently working toward the Ph.D. degree. His research interests include multimedia signal processing and wide color gamut imaging.



**Toinon Vigier** obtained a PhD in July 2015 from the Ecole Centrale de Nantes in the Ambiances Architectures and Urbanity lab, where she focused on virtual reality for urban studies. She specifically studied the impact of rendering and color effects on the perception of urban atmospheres through VR subjective tests. She was then a postdoctoral fellow in the Image Video and Communication team at Université de Nantes. She worked mainly on video quality and eye-tracking studies in the European CATRENE project UltraHD-4U which aims at studying and implementing a complete chain for the broadcasting of UHD-4K videos. Since September 2016, she is an Associate Professor at Université de Nantes in the Image Perception Interaction research team of the Laboratory of Digital Sciences in Nantes (LS2N). Her research mainly focuses on the study, the analysis and the prediction of the quality of experience for immersive and interactive multimedia through subjective and objective measures. She is currently involved in various national and international interdisciplinary projects focusing on user experience in immersive VR media for various applications (health, cinema, architecture, design...). She is also active in the standardization working group IEEE 3333.1 and she served as reviewers in a lot of international conferences and journals (IEEE TIP, IEEE TCSVT, SPIE JEI, IEEE VR, IEEE QoMEX, ACM TVX, IEEE MMSP).



**Patrick Le Callet** (IEEE Fellow) is full professor at University of Nantes, in the Electrical Engineering and the Computer Science departments of Polytech Nantes. He is one of the steering director of CNRS LS2N lab (450 researchers). He is also the scientific director of the cluster “Ouest Industries Créatives”, gathering more than 10 institutions (including 3 universities). “Ouest Industries Créatives” aims to strengthen Research, Education & Innovation of the Region Pays de Loire in the field of Creative Industries. He is mostly engaged in research dealing with cognitive computing and the application of human vision modeling in image and video processing. His current centers of interest are AI boosted QoE Quality of Experience assessment, Visual Attention modeling and applications. He is co-author of more than 300 publications and communications and co-inventor of 16 international patents on these topics. He serves or has been served as associate editor or guest editor for several Journals such as IEEE TIP, IEEE STSP, IEEE TCSVT, Springer EURASIP Journal on Image and Video Processing, and SPIE JEI. He is serving in IEEE IVMSPTC (2015- to present) and IEEE MMSP-TC (2015-to present) and one the founding member of EURASIP TAC (Technical Areas Committee) on Visual Information Processing.



**Jong-Seok Lee** (M06-SM14) received his Ph.D. degree in electrical engineering and computer science in 2006 from KAIST, Korea. From 2008 to 2011, he worked as a researcher at Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland. Currently, he is an associate professor in the School of Integrated Technology at Yonsei University, Korea. His research interests include multimedia signal processing and machine learning. He is an author or co-author of over 150 publications. He serves as an editor for the IEEE Communications Magazine and Signal Processing: Image Communication.