

Calculating Color Differences of Images via Siamese Neural Network

Yixuan Gao¹, Xiongkuo Min^{1*}, Xiaohong Liu², Lei Sun³, Yonglin Luo³, Zuowei Cao³ and Guangtao Zhai^{1*}

¹*Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, China*

²*John Hopcroft Center, Shanghai Jiao Tong University, China*

Email: {gaoyixuan,minxiongkuo,xiaohongliu,zhaiguangtao}@sjtu.edu.cn, {raylsun,vincenluo,ernestcao}@tencent.com

Abstract—Recently, the color difference (CD) of standard dynamic range (SDR) images has attracted the attention of researchers. It is worth noting that due to the development of high dynamic range (HDR) image generation technology, the CD of the SDR and HDR images is also worth in-depth research. This is because the HDR image generated from an original SDR image may have changes in color. Some color changes can give people a comfortable impression, but this may also change the information originally expressed in the SDR image. Therefore, this paper researches the CD of the original SDR image and the generated HDR image, and proposes a network to predict the CDs of SDR and HDR image pairs. Specifically, we first build a SDR-HDR image CD dataset. The dataset contains 504 SDR and HDR image pairs, where HDR images are generated from the SDR images using five HDR image generation methods. Second, we propose a siamese neural network to predict the CDs of SDR and HDR image pairs, which consists of three parts: space conversion, feature extraction, and CD calculation. Finally, experiments prove that the proposed network has a superior ability to predict the CDs of SDR and HDR image pairs.

Index Terms—Color difference, high dynamic range image, standard dynamic range image, siamese neural network.

I. INTRODUCTION

As an important physical attribute of images, color has attracted the attention of many famous scientists. Munsell defined color from the aspects of value, chroma, and hue. In the 20th century, the color space describing human color perception gradually emerged. On this basis, a large number of methods for calculating color difference (CD) have been proposed. For example, Robertson *et al.* [1] proposed CIE76 to calculate CD, Luo *et al.* [2] proposed CIEDE2000, and Imai *et al.* [3] proposed an image CD calculation method using Mahalanobis distance. Wang *et al.* proposed a network-based method to predict the CD of standard dynamic range (SDR) images [4].

Nowadays, with the development of multimedia technology, people hope that multimedia can display more comfortable images [5]–[12]. Therefore, to better restore the realistic world and improve the viewing experience, high dynamic range (HDR) image generation technology came into being [13]–[15]. In the process of traditional SDR image acquisition and processing, some of the scene information may be lost, while the HDR image can effectively reduce this loss. Compared

with SDR images, HDR images typically have wider luminance ranges and richer colors, which can give people a better viewing experience. There are two typical ways to generate HDR images from SDR images. The first is multi-exposure fusion [16], [17], which generates a HDR image from SDR images with different exposure levels. The second method is inverse tone mapping (ITM) [18], [19], which can generate a HDR image from a single SDR image by dynamic range expansion, over/under exposure area reconstruction, and color correction. In practice, it is difficult to obtain SDR images with different exposure levels for the existing content, so generating HDR images through ITM has become more and more popular.

However, rich colors can also bring some problems. For example, a HDR image generated by ITM may change the color semantics (such as green to yellow) and color tone (such as cold to warm) of the original SDR image. Some color changes can give people a comfortable impression, but this may also change the information originally expressed in the SDR image, thus providing the wrong information to people. Therefore, calculating the CD between the original SDR image and the generated HDR image is an urgent problem to be resolved. We can detect whether the generated HDR image has excessively changed the original information of the SDR image by calculating the CD, which can further guide the development of HDR image generation technology. As far as we know, no researcher has paid attention to the CDs between SDR and HDR image pairs. Therefore, this paper proposes a method to calculate the perceptual CDs between SDR and HDR image pairs.

The main contributions of this paper are as follows. First, we build a CD dataset for SDR and HDR image pairs, namely the SDR-HDR image CD dataset. Second, we propose a siamese neural network to predict the CDs of SDR and HDR image pairs. Finally, extensive experiments prove the effectiveness of the proposed method in predicting the CDs of SDR and HDR image pairs.

II. SDR-HDR IMAGE CD DATASET

In this section, we introduce the constructed SDR-HDR image CD dataset in detail.

A. Dataset Construction

1) *SDR Image Selection*: The SDR images in our SDR-HDR image CD dataset were selected from the MIT-Adobe

*Corresponding authors.

This work was supported in part by the National Natural Science Foundation of China under Grant 62271312, and in part by the Shanghai Pujiang Program under Grant 22PJ1407400.

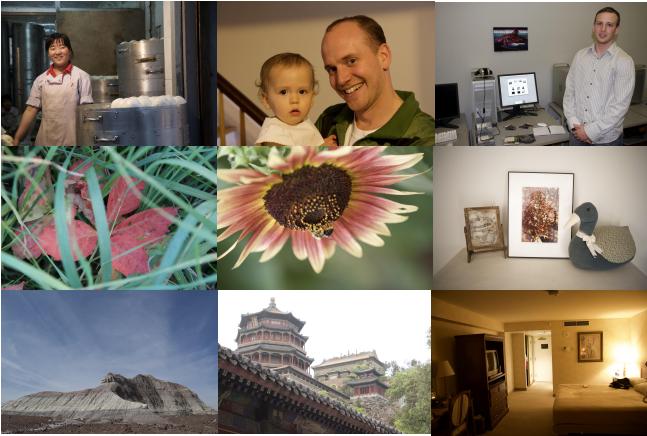


Fig. 1. Representative SDR images in the proposed SDR-HDR image CD dataset.

FiveK dataset [21], which consists of 5000 images taken by different photographers, covering a wide range of scenes. We selected 63 images from the MIT-Adobe FiveK dataset as SDR images in our SDR-HDR image CD dataset. These selected SDR images also contain rich scenes. Fig. 1 shows some representative SDR images.

2) *HDR Image Generation*: To obtain HDR images, we used some ITM methods to convert SDR images to HDR images, including four classical academic methods: Kovaleski-Oliveira [18], HuoPhys [19], Masia09 [22], Masia17 [23], and an industrial method. At the same time, to obtain HDR images with different CD levels, we modified the default parameters in these methods, such as bilateral filtering, color saturation, and contrast. Specifically, by modifying the bilateral filtering parameter in KovaleskiOliveira [18], we obtained 126 HDR images. For HuoPhys [19], Masia09 [22], and Masia17 [23], we found that the HDR images generated by these methods were not sensitive to the default parameters, so these three methods generated 63 HDR images, respectively. By modifying the color saturation and contrast parameters of the industrial method, we obtained 189 HDR images. Finally, we obtained 504 HDR images through these ITM methods. To sum up, our SDR-HDR image CD dataset contains 504 pairs of SDR and HDR images. Fig. 2 shows two SDR images and their HDR images generated by these ITM methods.

B. Subjective Experiment

1) *Subject*: We invited 18 subjects to participate in this subjective experiment, all of whom were students aged about 20. Before the experiment, we tested the visual conditions of all subjects. All 18 subjects had normal (corrected) visual acuity and color vision.

2) *Display Setting*: Since this experiment required subjects to observe a pair of SDR and HDR images at the same time, we needed two monitors. The first monitor was used to display SDR images. Here we used the monitor of the MacBook Air 2020, which has a peak luminance of 400 cd/m^2 , and a dual color gamut of 99% DCI-P3 and 100% sRGB. The second monitor was used to display HDR images. Here we

used the INNOCN M2U monitor with a peak luminance of 1000 cd/m^2 , and a dual color gamut of 99% DCI-P3 and 100% sRGB. However, if we want to observe HDR images normally, only having a HDR monitor is not enough. We also need a computer that supports HDR. Unfortunately, most popular computers only support playing HDR videos, but not displaying HDR images. Therefore, we converted each HDR image we obtained into a HDR10 video with only one frame.

3) *Experimental Procedure*: A pair of SDR and HDR images were simultaneously displayed on the SDR monitor and HDR monitor, respectively. Subjects were asked to score the CD between the two images and gave their scores on a continuous scale within the range of $[0, 5]$. The scoring criteria are as follows: a score in the area between 0 and 1 means that the two images have nearly the same color; a score in the area between 1 and 2 means that the two images have slightly different colors; a score in the area between 2 and 3 means that the two images have different colors; a score in the area between 3 and 4 means that the two images have significantly different colors; a score in the area between 4 and 5 means that the two images have completely different colors.

Before the formal test, we selected eight SDR and HDR image pairs with different contents from 504 SDR and HDR image pairs in our SDR-HDR image CD dataset as training image pairs. After that, 504 pairs of SDR and HDR images were randomly displayed to all subjects in the test stage. There was no time limit for the experiment. To avoid fatigue, subjects could suspend the experiment at any time.

C. Data Processing and Analysis

After the subjective experiment, each SDR and HDR image pair has 18 subjective CD ratings. After the subject rejection [24], each SDR and HDR image pair has 15 valid subjective CD ratings. We calculate the mean of the valid subjective CD ratings of each SDR and HDR image pair as the ground-truth perceptual CD. Fig. 3 shows the distribution of the ground-truth CDs of 504 SDR and HDR image pairs in the SDR-HDR image CD dataset.

III. PROPOSED METHOD

We design a siamese neural network to calculate the CDs of SDR and HDR image pairs, which includes three parts: space conversion, feature extraction, and CD calculation. The diagram of the proposed network is shown in Fig. 4.

A. Space Conversion

Typically, both SDR images and HDR images are stored in the RGB color space. However, HDR images have wider luminance ranges than SDR images, and they can store richer colors, so it is unfair to directly compare SDR images and HDR images. Therefore, we first convert SDR images and HDR images from the original color space to a new space. In this paper, we use two shared-weight residual networks to perform space conversion on a SDR and HDR image pair:

$$\text{SDR}_{\text{new}} = \text{SDR}_{\text{old}} + \mathcal{F}(\text{SDR}_{\text{old}}, w_{\text{old}}), \quad (1)$$

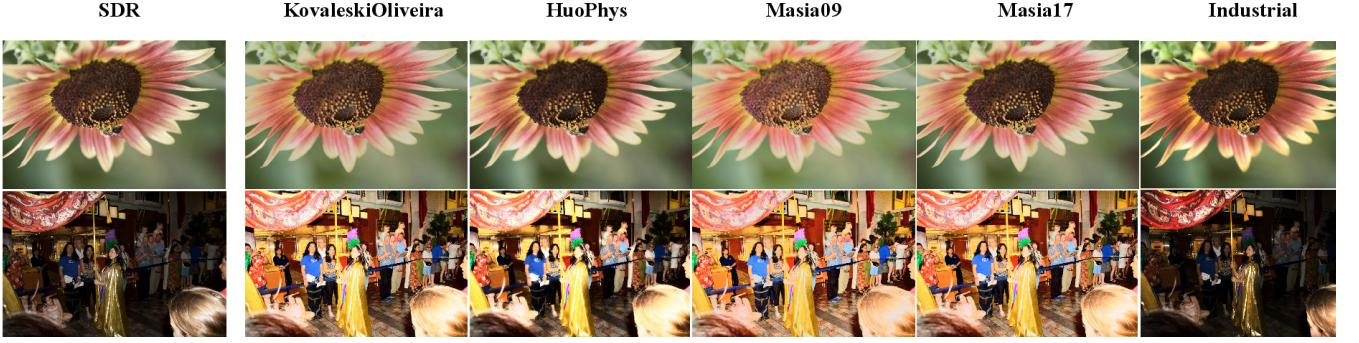


Fig. 2. SDR images and the generated HDR images. The first column shows SDR images, and the second to sixth columns show the generated HDR images. Note that the HDR images shown here are all tone mapped by Reinhard [20] for display purposes.

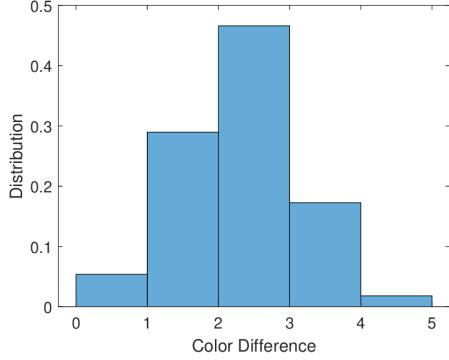


Fig. 3. Distribution of the ground-truth CDs for all image pairs in the proposed SDR-HDR image CD dataset.

$$\text{HDR}_{new} = \text{HDR}_{old} + \mathcal{F}(\text{HDR}_{old}, w_{old}), \quad (2)$$

where $\text{SDR}_{old} \in \mathbb{R}^{H \times W \times C}$ and $\text{HDR}_{old} \in \mathbb{R}^{H \times W \times C}$ denote the input SDR and HDR images respectively, $\text{SDR}_{new} \in \mathbb{R}^{H \times W \times C}$ and $\text{HDR}_{new} \in \mathbb{R}^{H \times W \times C}$ denote the transformed SDR and HDR images respectively, in which H , W , and C represent the height, width, and channel of the input image respectively. $\mathcal{F}(\cdot, \cdot)$ is the residual network with two convolution layers. w_{old} is the weight shared between the two residual networks.

B. Feature Extraction

We then use the pre-trained VGG16 networks without full-connection (FC) layers to extract the features of the transformed SDR and HDR images:

$$\text{SDR}_{map} = \text{VGG16}(\text{SDR}_{new}, w), \quad (3)$$

$$\text{HDR}_{map} = \text{VGG16}(\text{HDR}_{new}, w), \quad (4)$$

where $\text{SDR}_{map} \in \mathbb{R}^{\hat{H} \times \hat{W} \times \hat{C}}$ and $\text{HDR}_{map} \in \mathbb{R}^{\hat{H} \times \hat{W} \times \hat{C}}$ denote the feature maps of SDR and HDR images extracted from the pre-trained VGG16 respectively, in which \hat{H} , \hat{W} , and \hat{C} represent the height, width, and channel of the feature maps respectively. To reduce the network parameters to be trained, the two VGG16 networks also share weight, and w is the weight shared between them.

C. CD Calculation

Finally, we subtract the feature maps of the SDR image and HDR image to obtain the feature difference map DM:

$$DM = |\text{HDR}_{map} - \text{SDR}_{map}|. \quad (5)$$

Finally, we use DM and three FC layers to get the predicted CD of the SDR and HDR image pair. The loss function we used in this paper is the mean square error.

IV. EXPERIMENTS

Next, experiments are conducted to verify the effectiveness and feasibility of the proposed network in predicting the CDs of SDR and HDR image pairs.

A. Experimental Setups

Experiments are carried out on the SDR-HDR image CD dataset. In each experiment, we take 80% of the image pairs in the dataset as the training set and other image pairs as the test set. Specifically, we resize each image into a resolution of 224×224 and use the Adam algorithm [25] with a learning rate of 10^{-5} . The batch size in this paper is 8. To reduce the bias caused by randomness in training and test set splittings, we repeat each experiment ten times and report the mean value. We use three metrics for performance comparison: Spearman rank correlation coefficient (SRCC), Pearson linear correlation coefficient (PLCC), and root mean square error (RMSE).

B. Performance Comparison

We compare the proposed method with several state-of-the-art SDR image CD calculation methods: CIE76 [1], CIEDE2000 [2], CMC [26], Imai [3], Ouni [27], Zhang [28], Hong [29], Pedersen [30], Lee05 [31], Lee14 [32], Huertas [33], Simone [34], CD-Net [4], and two state-of-the-art HDR image CD calculation methods: HDRVDP [35] and HDRVQM [36]. The comparison results are shown in Table I. It is clear from the table that our proposed method achieves the best performance among all competing methods. In addition, it is also found that it is ineffective to directly use SDR or HDR image CD calculation methods to calculate the CDs of SDR and HDR image pairs, which may be due to the large differences in luminance ranges between them. This highlights the need to propose a suitable method to calculate the CDs of SDR and HDR image pairs.

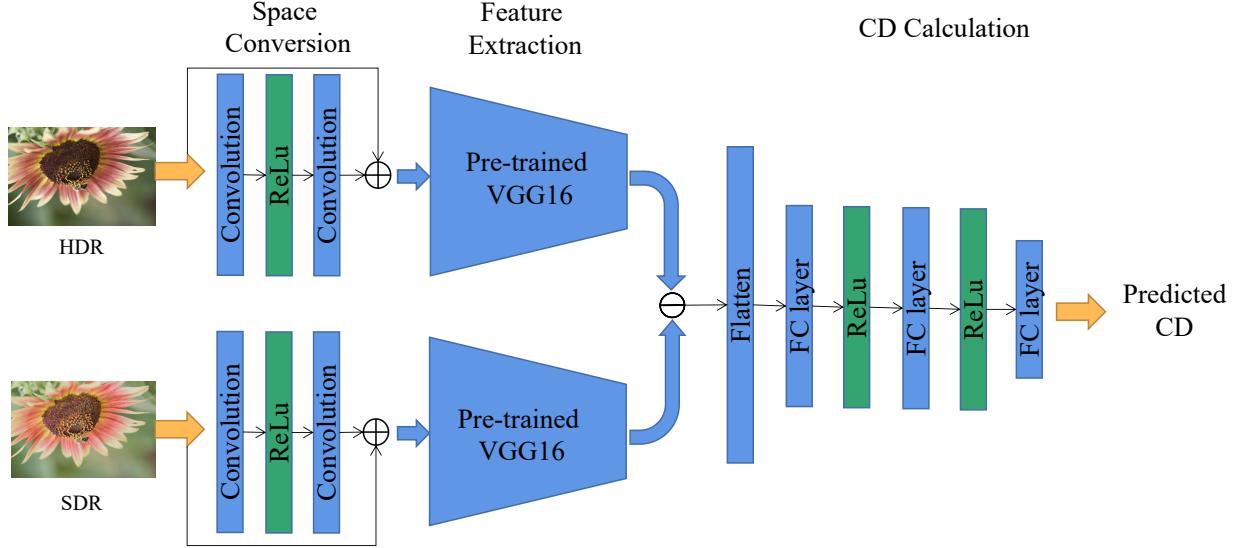


Fig. 4. Diagram of the proposed siamese neural network. The network includes a space conversion part, a feature extraction part, and a CD calculation part.

TABLE I

PERFORMANCE OF THE PROPOSED METHOD AND THE STATE-OF-THE-ART SDR/HDR IMAGE CD CALCULATION METHODS IN PREDICTING THE CDs OF SDR AND HDR IMAGE PAIRS. THE BEST PERFORMANCES ARE IN BOLD.

Method	SRCC	PLCC	RMSE
CIE76 [1]	0.2532	0.2948	0.7879
CIEDE2000 [2]	0.1934	0.2873	0.7897
CMC [26]	0.1610	0.1455	0.8157
Imai [3]	0.0908	0.1214	0.8184
Ouni [27]	0.1934	0.2873	0.7897
Zhang [28]	0.1876	0.2848	0.7904
Hong [29]	0.1908	0.2906	0.7889
Pedersen [30]	0.1725	0.2953	0.7877
Lee05 [31]	0.1091	0.1318	0.8173
Lee14 [32]	0.3515	0.3456	0.7737
Huertas [33]	0.3007	0.3336	0.7773
Simone [34]	0.3801	0.3541	0.7711
HDRVDP [35]	0.4010	0.4324	0.7434
HDRVQM [36]	0.5267	0.5556	0.6855
CD-Net [4]	0.4534	0.4807	0.7904
Proposed	0.8748	0.8772	0.5182

C. Ablation Study

The proposed siamese neural network consists of two important parts: space conversion and feature extraction. In this section, we construct comparison methods to verify the importance of space conversion and feature extraction for the proposed method. First, we delete the space conversion part of the proposed method to build the first comparison method. Second, we delete the feature extraction part of the proposed method to build the second comparison method. What's more, we also construct a comparison method whose weights are not shared. The comparison results are shown in Table II. It can be seen from the table that the performance of the proposed method is the best.

D. Computational Complexity

To compare the computational complexities of the proposed method and all state-of-the-art SDR/HDR image CD calculation methods, we calculate the mean running time of these

TABLE II
COMPARISON RESULTS OF THE ABLATION STUDY. THE BEST PERFORMANCES ARE IN BOLD.

Method	SRCC	PLCC
Without space conversion	0.8663	0.8677
Without feature extraction	0.3930	0.3786
Without weight sharing	0.8218	0.8213
Proposed	0.8748	0.8772

TABLE III
COMPUTATIONAL COMPLEXITIES OF THE PROPOSED METHOD AND STATE-OF-THE-ART METHODS.

Method	Time	Method	Time
CIE76 [1]	1.2356	Lee05 [31]	2.4515
CIEDE2000 [2]	4.3504	Lee14 [32]	15.4458
CMC [26]	3.0409	Huertas [33]	10.3563
Ouni [27]	4.4623	Simone [34]	14.0495
Zhang [28]	8.2529	HDRVDP [35]	48.8490
Hong [29]	16.0938	HDRVQM [36]	7.0634
Pedersen [30]	31.4687	CD-Net [4]	0.8164
Proposed	0.8419	-	-

methods when calculating the CDs of 100 SDR and HDR image pairs. We report the results in Table III. All method are tested on a computer with the Intel Core i7-12700F CPU @2.10 GHz and 16 GB RAM. From the table, it can be seen that our proposed method has a shorter running time than most state-of-the-art SDR/HDR image CD calculation methods.

V. CONCLUSION

In this paper, we propose to predict the CDs of SDR and HDR image pairs. First, we build the SDR-HDR image CD dataset. Specifically, we select 63 SDR images from the MIT-Adobe FiveK dataset and use five ITM methods to generate 504 HDR images. We then carry out a subjective experiment to obtain the subjective CD ratings of 504 SDR and HDR image pairs. We also propose a network to predict the CDs of SDR and HDR image pairs, which consists of three parts: space conversion, feature extraction, and CD calculation. Finally, extensive experimental results show that the proposed network is effective for CD prediction between HDR and SDR images.

REFERENCES

- [1] A. R. Robertson, "The CIE 1976 color-difference formulae," *Color Res. Appl.*, vol. 2, no. 1, pp. 7–11, 1977.
- [2] M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," *Color Res. Appl.*, vol. 26, no. 5, pp. 340–350, 2001.
- [3] F. H. Imai, N. Tsumura, and Y. Miyake, "Perceptual color difference metric for complex images based on Mahalanobis distance," *J. Electron. Imaging*, vol. 10, no. 2, pp. 385–393, 2001.
- [4] Z. Wang, K. Xu, Y. Yang, J. Dong, S. Gu, L. Xu, Y. Fang, and K. Ma, "Measuring perceptual color differences of smartphone photography," *arXiv preprint arXiv:2205.13489*, 2022.
- [5] X. Min, G. Zhai, J. Zhou, M. C. Farias, and A. C. Bovik, "Study of subjective and objective quality assessment of audio-visual signals," *IEEE Trans. Image Process.*, vol. 29, pp. 6054–6068, 2020.
- [6] X. Min, K. Gu, G. Zhai, J. Liu, X. Yang, and C. W. Chen, "Blind quality assessment based on pseudo-reference image," *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 2049–2062, 2017.
- [7] X. Min, G. Zhai, K. Gu, Y. Liu, and X. Yang, "Blind image quality estimation via distortion aggravation," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 508–517, 2018.
- [8] Y. Gao, X. Min, W. Zhu, X.-P. Zhang, and G. Zhai, "Image quality score distribution prediction via alpha stable model," *IEEE Trans. Circuits Syst. Video Technol.*, 2022.
- [9] Y. Gao, X. Min, Y. Zhu, X.-P. Zhang, and G. Zhai, "Blind image quality assessment: A fuzzy neural network for opinion score distribution prediction," *IEEE Trans. Circuits Syst. Video Technol.*, 2023.
- [10] Y. Gao, X. Min, Y. Zhu, J. Li, X.-P. Zhang, and G. Zhai, "Image quality assessment: From mean opinion score to opinion score distribution," in *ACM Int. Conf. Multimed.*, 2022, pp. 997–1005.
- [11] Y. Sun, X. Min, H. Duan, and G. Zhai, "The influence of text-guidance on visual attention," in *IEEE Int. Symp. Circuits. Syst.*, 2023, pp. 1–5.
- [12] Y. Cao, X. Min, W. Sun, and G. Zhai, "Subjective and objective audio-visual quality assessment for user generated content," *IEEE Trans. Image Process.*, 2023.
- [13] M. Daafallah, H. Yuan, S. Jiang, and Y. Yang, "An attention-based network for single image HDR reconstruction," in *IEEE Int. Symp. Circuits. Syst.*, 2022, pp. 732–736.
- [14] P.-H. Hsu, Y.-M. Yeh, C.-M. Yeh, and Y.-C. Lu, "A high dynamic range light field camera and its built-in data processor design," in *IEEE Int. Symp. Circuits. Syst.*, 2018, pp. 1–5.
- [15] M. White, S. Ghajari, T. Zhang, A. Dave, A. Veeraraghavan, and A. Molnar, "A differential SPAD array architecture in 0.18 μm CMOS for HDR imaging," in *IEEE Int. Symp. Circuits. Syst.*, 2022, pp. 292–296.
- [16] Y. Xu, Z. Liu, X. Wu, W. Chen, and Z. Li, "Restoration of HDR images for SVE-based HDRI via a novel DCNN," in *IEEE Int. Conf. Multimedia Expo*, 2021, pp. 1–6.
- [17] H. Wang, T. Zhang, and G. Lu, "Unsupervised HDR image reconstruction based on over/under-exposed LDR image pair," in *IEEE Int. Conf. Multimedia Expo*, 2021, pp. 1–6.
- [18] R. P. Kovaleski and M. M. Oliveira, "High-quality reverse tone mapping for a wide range of exposures," in *SIBGRAPI Conf. Graph., Patterns Images*, 2014, pp. 49–56.
- [19] Y. Huo, F. Yang, L. Dong, and V. Brost, "Physiological inverse tone mapping based on retina response," *Visual Comput.*, vol. 30, no. 5, pp. 507–517, 2014.
- [20] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 267–276, 2002.
- [21] B. Vladimír, P. Sylvain, C. Eric, and D. Frédéric, "Learning photographic global tonal adjustment with a database of input/output image pairs," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011.
- [22] B. Masia, S. Agustin, R. W. Fleming, O. Sorkine, and D. Gutierrez, "Evaluation of reverse tone mapping through varying exposure conditions," in *ACM Trans. Graph.*, 2009, pp. 1–8.
- [23] B. Masia, A. Serrano, and D. Gutierrez, "Dynamic range expansion based on image statistics," *Multimed. Tools Appl.*, vol. 76, no. 1, pp. 631–648, 2017.
- [24] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," *Rec. ITU-R BT.500-11*, 2002.
- [25] D. Kingma and B. Jimmy, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [26] Anon, "CMC: Calculation of small color differences for acceptability," *Text. Chem. Color.*, 1989.
- [27] S. Ouni, E. Zagrouba, M. Chambah, and M. Herbin, "A new spatial colour metric for perceptual comparison," in *Int. Conf. E-Syst. Eng., Commun. Inf.*, vol. 615, 2008, pp. 413–428.
- [28] X. Zhang and B. A. Wandell, "A spatial extension of CIELAB for digital color-image reproduction," *J. Soc. Inf. Disp.*, vol. 5, no. 1, pp. 61–63, 1997.
- [29] G. Hong and M. R. Luo, "New algorithm for calculating perceived colour difference of images," *Imag. Sci. J.*, vol. 54, no. 2, pp. 86–91, 2006.
- [30] M. Pedersen and J. Y. Hardeberg, "A new spatial filtering based image difference metric based on hue angle weighting," *J. Imaging Sci. Technol.*, vol. 56, no. 5, p. 1–12, 2012.
- [31] S. Lee, J. H. Xin, and S. Westland, "Evaluation of image similarity by histogram intersection," *Color Res. Appl.*, vol. 30, no. 4, pp. 265–274, 2005.
- [32] D. Lee and K. N. Plataniotis, "Towards a novel perceptual color difference metric using circular processing of hue components," in *IEEE Int. Conf. Acoust. Speech Signal Process Proc.*, 2014, pp. 166–170.
- [33] R. Huertas, M. Melgosa, and C. Oleari, "Performance of a color-difference formula based on OSA-UCS space using small-medium color differences," *J. Opt. Soc. Am. A-Opt. Image Sci. Vis.*, vol. 23, no. 9, pp. 2077–2084, 2006.
- [34] G. Simone, C. Oleari, and I. Farup, "An alternative color difference formula for computing image difference," in *Gjøvik Color Imaging Symposium (GCIS)*, no. 4, 2009, pp. 8–11.
- [35] R. Mantiuk, K. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 1–14, 2011.
- [36] M. Narwaria, M. P. Da Silva, and P. Le Callet, "HDR-VQM: An objective quality measure for high dynamic range video," *Signal Process.-Image Commun.*, vol. 35, pp. 46–60, 2015.