A Color Preserving Down-sampling Approach for 8K to 4K HDR Images

Hamid Reza Tohidypour, Yixiao Wang, Mahsa T. Pourazad, Panos Nasiopoulos, Alan Tong, Mohammadreza Saed, Mengya Zeng, and Ruixue Luo
Dept. of Electrical and Computer Engineering, The University of British Columbia, Vancouver, Canada
e-mail: {htohidyp, yixiaow, pourazad, panos}@ece.ubc.ca, {tongkaitai, mohammadrezasaed, echomaria98123}@gmail.com, {luorx1202}@outlook.com

***Abstract*— 8K High Dynamic Range (HDR) cameras have recently become available in the consumer market, capturing more accurate spatial and color information. Despite advancements in display technology, 4K displays continue to maintain their dominance in the market. In this paper, we propose an efficient approach for down-sampling 8K HDR content to 4K HDR content that maintains the spatial and color information of the former, to the maximum extent possible. In this regard, we converted the 8K HDR images into different commonly used color spaces, namely L*a*b*, YCbCr, and ICtCp. Then, we evaluated the performance of the Bilinear, Bicubic, Biquintic, and Lanczos down-sampling approaches on these color spaces. In addition, we investigated the effect of Gaussian and Bilateral filters. Our subjective evaluations showed that the combination of gaussian filtered RGB images and Biquintic down-sampling method resulted in the best performance.**

Keywords- 8K HDR; 4K HDR; down-sampling; color spaces.

I. INTRODUCTION

Advancement in camera sensor technology has increased the resolution of captured images. This advancement made the 8K cameras the successor of 4K cameras. 8K camera captures images and videos with more accurate spatial and color information from the environment. However, 4K displays are still dominant the market and it will take several years until 8K displays dominant the consumer market. Therefore, for backward compatibility purposes the captured 8K content need to be converted to its 4K version. Although this conversion will remove some details from the 8K content, it is expected to be better in terms of quality compared to the case of capturing the same scene using a 4K camera.

To the best of our knowledge, there is no previous work on converting 8K to 4K content, while attempting to preserve the spatial and color details. This paper addresses this problem for 8K HDR content by exploring the impact of four commonly used down-sampling methods on images in four well-known color spaces, and then decide which one has the best performance. We employed Bilinear, Bicubic, Biquintic, and Lanczos as the down-sampling methods and used the RGB, L*a*b*, YCbCr, and ICtCp color spaces. [1-3]. Two filtering methods, namely Gaussian and Bilateral, were also examined for our application [1]. More precisely, we conducted subjective tests to evaluate the performance of all the combinations in maintaining spatial details and color. Our results showed that the combination of Guassian filtered RGB images and Biquintic down-sampling method achieved the best performance.

The remainder of this paper is organized as follows. In Section II, we talk about our methodology. Section III discussed about our results. Section IV concludes our paper.

II. METHODOLOGY

A. Color spaces

Conducting down-sampling methods directly in RGB color space may not yield the most satisfying results given that it doesn't have a separate luminosity channel. Given that human vision is more sensitive to luma information than chroma information, it is important to explore down-sampling methods in color spaces that have a separate channel for luminance. The color spaces that we examined in this paper in addition to RGB include: 1) L*a*b* color space, also known as CIELAB, 2) YCbCr color space,3) ICtCp color space [2-3]. All these color spaces have separate luminance channels [2-3].

The reason to use L*a*b* color space is that unlike RGB color models, L*a*b* is intended to approximate human perception of color. L*a*b* color space still lacks perceptual uniformity, especially in blue hues. But, the L* component matches human perception of lightness closely, although the Helmholtz–Kohlrausch effect isn't taken into account. This makes it still useful for predicting small differences in color [2].

YCbCr is not an absolute color space, and it is a scaled and offset version of YUV. The main difference between YUV and YCbCr is that the former is for analog TV and the latter is for digital TV. Although YCbCr is not a perceptual color space, it is widely used in image and video compression [3].

ICTCP, ICtCp, or ITP is a color representation format specified in the Rec. ITU-R BT.2100 standard that is used as a part of the color image pipeline in video and digital photography systems for HDR and wide color gamut (WCG) imagery [4], which makes this color space a good candidate for our application. ICtCp has a near constant luminance, so it has a better result with chroma subsampling when compared with YCbCr. Also, ICtCp has an improved hue linearity compared to YCbCr, which is beneficial to compression performance and color volume mapping. All of the above qualities could make ICtCp a good choice for our application [3].

 ISBN: 978-1-68558-076-6

B. *Filters*

In addition to the color spaces, we investigate the impact of two frequently utilized filters, Gaussian filters and bilateral filters, on maintaining color accuracy.

Gaussian filters are among the most commonly used low-pass filters due to their effectiveness in removing high-frequency signals from input images when configured with appropriate standard deviation settings.

One of the drawbacks of Gaussian filters is that that they solely rely on the spatial relations between pixels within the kernel and do not take into account the image content. Bilateral filters, on the other hand, were developed based on Gaussian filters with considerations of the image content. As a result, bilateral filters have a desirable property of preserving edges [4].

C. *Down-sampling methods*

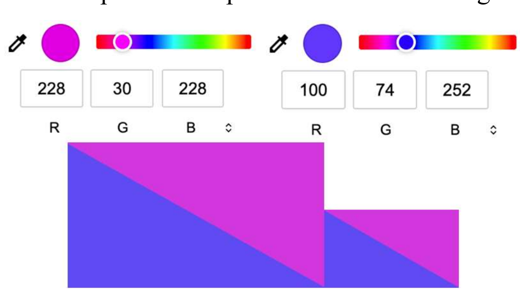
We chose four down-sampling methods including Bilinear, Bicubic, Biquintic, and Lanczos. Bilinear method linearly uses four neighbouring pixels to predict the pixels of down-sampled image. While the Bicubic method uses 16 pixel values instead of 4 and a third degree polynomial, which result in smoother images. In the case of Biquintic, a five degree polynomial function is used to approximate the pixels. This causes the resulting images to be smoother than bicubic. Lanczos uses a sinc function as its kernel to approximate the pixels. As the sinc function consists of positive and negative values, the negative values sharpen the images and increase the contrast of the images [1]. Four different color spaces, mentioned in the previous subsection, will be used to evaluate all the down-sampling methods.

III. EXPERIMENTS AND RESULTS

A. *Visualization on synthethic data*

As there is no ground truth 8K and 4K images dataset for our application, we generate a synthetic dataset to help visualize the effect of each color space and down-sampling combination on 8K raw images.

Two different colors are randomly picked in RGB color space, and each color is assigned to the upper and lower triangle in the 8K and 4K image separately. The diagonally split pattern represents an infinitely thin edge, as shown in Figure 1. This pattern is representative because edges are the



(a) (b)

Figure 1. a) Left: 8K Synthetic image; b) 4K Synthetic image.

most fundamental components in any image. Understanding the effects of each of the combinations of color spaces and down-sampling approaches on a synthetic edge will help to better analyze the real images.

Once the pair of input (8K image) and ground truth (4K image) are generated, the 8K image data is fed into the proposed 8K-to-4K converter approaches to get a set of 4K outputs for each combination. In order to compare our approaches, we use delta E as our error metrics, as shown below [5]:

$$\triangle E = \sqrt{(L_{gt}-L_o)^2 + (a_{gt}-a_o)^2 + (b_{gt}-b_o)^2}$$

where L_{gt}, a_{gt}, and b_{gt} represent the L*a*b* values of the ground truth, while the L_0, a_0, and b_0 show the L*a*b* values of the down-sampled image.

Figure 2 shows the error matrix $\triangle E$ for each color space and down-sampling combination. As it can be seen across each color space, the difference is very minimal, yet the patterns of the deltaE error matrix within a color space across four down-sampling methods differ significantly. In general, Bilinear tends to have the thinnest span of error, but its error values are much higher than the other 3 (lighter color means higher error). Bicubic and Lanczos both generate a medium span of error with medium error values. Biquintic has the lowest overall error value (close to gray color) and largest span size (close to Lanczos). In practice, when human eyes perceive these error patterns, a sharp, clear and high error edge like the one generated by Bilinear interpolation method has the most obvious artifact, because the gradient over the error region is extremely high, catching human eyes' attention. In the case of Biquintic, a relatively wider, yet low error value span can be observed, which means the artifact is less obvious.

We generate a normalized histogram of each error matrix to investigate the distribution of the errors within each combination, as shown in Figure 3. As it can be seen, the error matrices of Bilinear and Lanczos generated error matrices have high deltaE values in all ranges of errors. In Bicubic and Biquintic, on the other hand, error values are mostly in the low range. Since information loss is inevitable due to the nature of down-sampling, an ideal 8K-to-4K converter shall have most of the error values in the low range. Therefore, in this infinitely thin edge case, Bicubic and Biquintic outperform the other interpolation methods.

B. *Subjective results*

To further analyze the combinations and find the best set of combinations of color space and down-sample methods, subjective tests were conducted. It is worth mentioning that objective metrics (such as delta E and PSNR) can not be performed since the reference 4K HDR images do not exist for this study.

We chose 8K YUV from ITE videos with the resolution of 7680x4320 pixels as our test dataset [6]. The bit depth of the videos was 10, and the frame rate was 59.940. The color primary was BT 2020. We randomly chose 3 8K HDR images from three different videos of ITE dataset. We converted each frame from the original RGB format to L*a*b*, ICtCp, and YCbCr respectively. Moreover, to explore the impact that

 ISBN: 978-1-68558-076-6

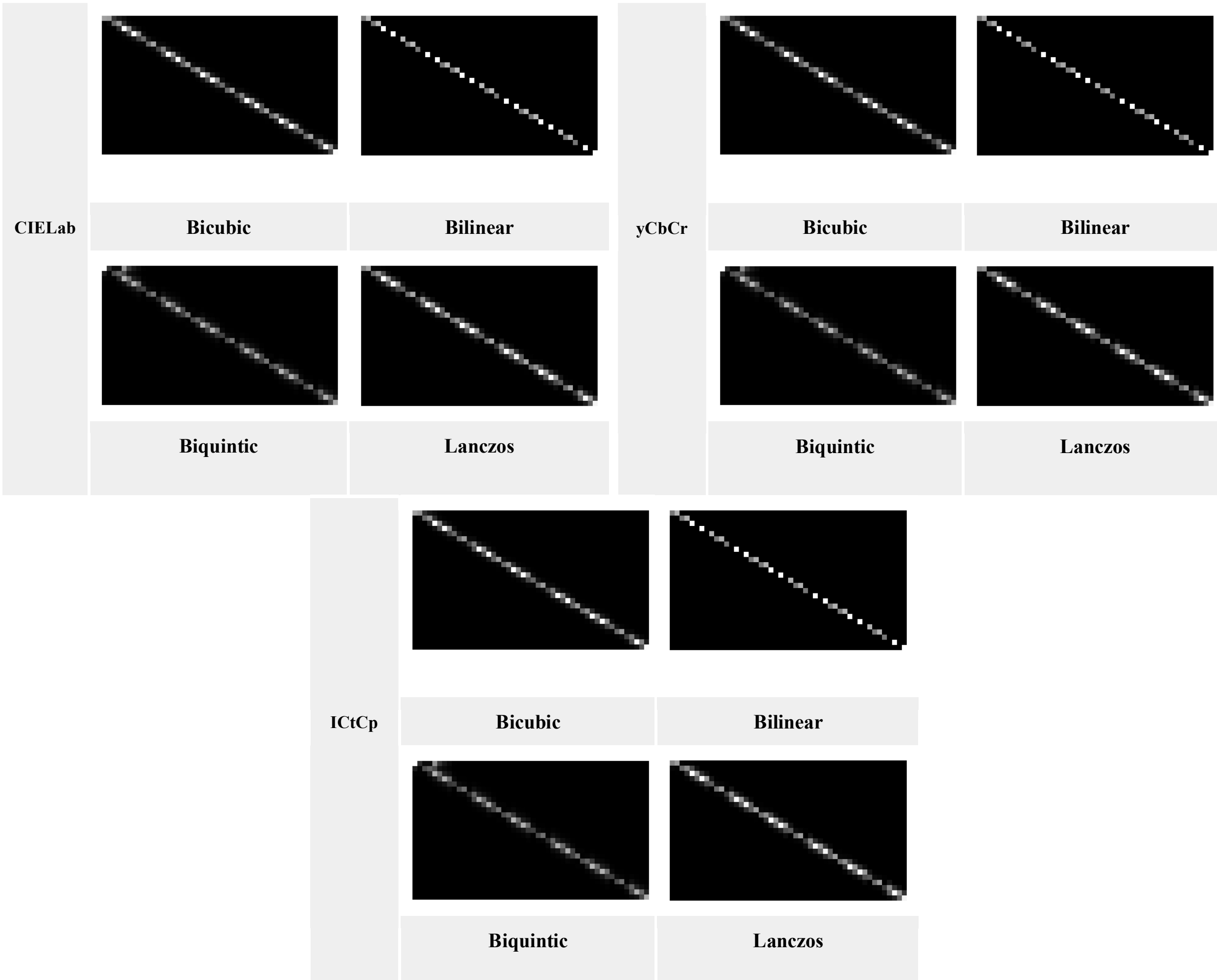


Figure 2. DeltaE error matrix for all colorspaces and down-sampling combinations on synthetic data. The axes of each sub-figure are the same as the axes of Figure 1 (vertical and horizental axes of the resulting deltaE image).

these filters have on our application, we applied two filters (Gaussian and Bilateral) to raw RGB files for comparison. Thus, finally we have 60 different frames.

We followed ITU-R BT.500-14 to run our subjective tests and used 11-point impairment scale recommended in [7]. 18 subjects participated in the test. Prior to the subjective test, all the participants successfully passed the color vision test and vision acuity test. Before starting the subjective tests, subjects were trained to become familiar with the test procedure. We used a professional 8K HDR TV for our subjective test. During the subjective test for each combination, the message indicating that the reference image would be shown was displayed for 2 seconds followed by the reference frame that was shown for 10 seconds. Afterwards, the message indicating that the 4K HDR image would be shown was displayed for 2 seconds, then one of that the generated 4K HDR was displayed at the center of 8K HDR TV was shown for 10 seconds. Then, the subjects were given 6 seconds to score the generated 4K HDR image compared to the reference image. The range of the score was between 0 to 10 according to ITU-R BT.500. The higher the number the better the down-sampling combination preserved the spatial and color information of the reference. Table I presents the interpretation of the 11-grade numerical quality scale, ranging from perceptible quality level to severely annoying quality level. It is worth mentioning that the combinations were shown to the subjects in random orders. Post processing resulted in finding one outlier for whom the related information was removed. Table II shows the Mean Opinion Score (MOS). As it can be seen, the combination of Gaussian and Biquintic achieved the highest MOS followed by the combination of Guassian and Bicubic.

IV. CONCLUSION

In this paper, we proposed an approach for converting 8K HDR images to 4K HDR images that maintains the spatial and color information as much as possible. In order to design our method, we investigated the performance of four commonly used down-sampling methods, namely Bilinear, Bicubic, Biquintic, and Lanczos. The color spaces that we examined

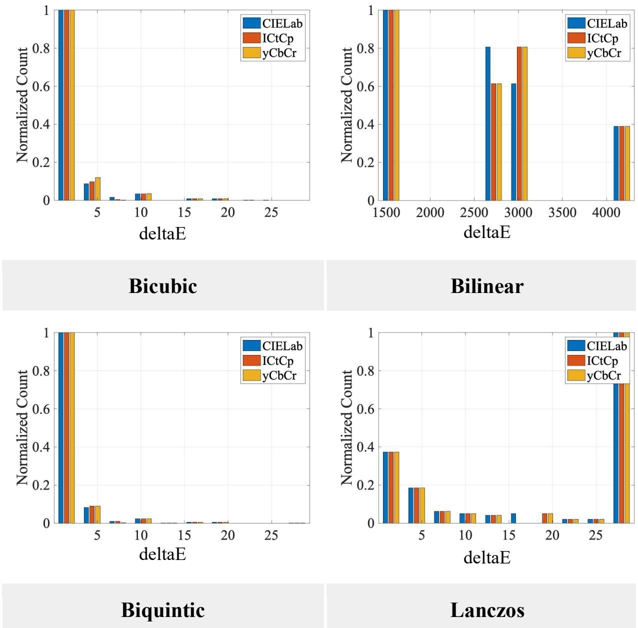


Figure 3. Normalized histogram of error values.

were RGB, L*a*b*, YCbCr, and IctCP. Moreover, we investigated the effects of two well-known filters including Gaussian and Bilateral. Our subjective results showed that the combination of the Gaussian filter for RGB color space and Biquintic achieved the best mean opinion score. Future work involves checking this combination for video applications.

TABLE I. MEANING OF THE 11 GRADES NUMERICAL SCALE [7].

Score	Impairment item	
10	Imperceptible	
9	Slightly perceptible	somewhere
8		everywhere
7	Perceptible	somewhere
6		everywhere
5	Clearly perceptible	somewhere
4		everywhere
3	Annoying	somewhere
2		everywhere
1	Severely annoying	somewhere
0		everywhere

TABLE II. AVERAGE MOS FOR ALL THE COMBINATIONS TESTED IN THIS STUDY.

Down-sampling Methods / Color Space or Filter	Bicubic	Bilinear	Biquintic	Lanczos
Lab	7.29	7.06	7.36	7.25
ICtCp	7.28	7.05	7.35	7.24
YCbCr	7.26	7.00	7.34	7.23
Bilateral	7.77	7.30	7.84	7.61
Gaussian	8.14	7.95	**8.52**	8.03

ACKNOWLEDGMENT

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC – PG 11R12450), and TELUS (PG 11R10321). This research was enabled in part by support provided by WestGrid (www.westgrid.ca) and Compute Canada (www.computecanada.ca).

REFERENCES

[1] CIE Colorimetry 15 (Third ed.). CIE. 2004. ISBN 3-901-906-33-9.

[2] Recommendation ITU-R BT.2100-2, “Image Parameter Values for High Dynamic Range Television for Use in Production and International Programme Exchange,” available online: https://www.itu.int/rec/R-REC-BT.2100 [retrieved: May 2023].

[3] R. C. Gonzalez and R. E. Woods, “Digital Image Processing (3rd Edition),” Prentice-Hall, Inc., 2006.

[4] S. Kumar, “A Straightforward Introduction to Image Blurring/Smoothing Using Python.” Medium, Spinor, 4 Dec. 2019, available online: https://medium.com/spinor/a-straightforward-introduction-to-image-blurring-smoothing-using-python-f8870cf1096 [retrieved: May 2023].

[5] G. Sharma, “Digital Color Imaging Handbook”, (1.7.2 ed.) CRC Press. 2003, ISBN 0-8493-0900-X.

[6] Test chart Ultra-high definition, wide color gamut HDR version standard moving image (C series). https://www.ite.or.jp/content/test-materials/uhdtv_hdr/ [retrieved: May 2023].

[7] Recommendation ITU-R BT.500-14, “Methodologies for the subjective assessment of the quality of television images,” 2019.