HDRi 2014 - Second International Conference and SME Workshop on HDR imaging,
Chalmers A., and Bouatouch K. (Editors)

Colorimetric calibration of high dynamic range images with a ColorChecker chart

D. Varghese, R. Wanat[1] and R.K. Mantiuk[1]
[1]Bangor University, U.K.

Abstract

In this paper we assess the quality of colorimetric calibration of HDR images captured using the multi-exposure technique. Three cameras were used to take HDR photographs of 4 scenes containing a color checker chart (X-Rite) and custom color patches. These were also measured using a spectrometer to provide ground truth data. One of the scenes was then used as a training for 2 camera calibration models, one including and one without the black level. The parameters of the models were fitted by optimizing an error function based on the CIE2000 ΔE color error measure. For the remaining 3 scenes, the accuracy of the fit was verified using the CIE2000 ΔE error between the measured and predicted color values. Our results indicate that a very good fit can be achieved, in the range of 0.8 *to* 2.1 ΔE *using the model without black level.*

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation

1. Introduction

High dynamic range (HDR) images are often claimed to contain accurate photometric and colorimetric measurements of the captured scenes. Since the HDR merging process removes the non-linearity introduced by a camera, the values are indeed approximately linearly related to luminance. However, there are also other factors that can distort captured color values, such as lens glare [MR07], vignetting, noise and spatially variant image processing of the camera. Moreover, the spectral sensitivity of most cameras sensors is different from the color matching functions (such as CIE XYZ 1931), therefore some inaccuracies in captured color can be expected.

In this work we test how accurately HDR images can be color-calibrated. We capture HDR images using the multi-exposure technique [MN99] with camera RAW images. Because we use RAW images, there is mimimum camera image processing introduced in the captured images. The calibration is performed using a single HDR image containing an X-Rite color checker, in which color values have been measured using a photospectrometer. The accuracy of the calibration is tested in three other scenes, in which we placed diffuse color targets, which were also measured with the photospectrometer. The color calibration is tested for three cameras (Canon 1000D, Canon 550D and Panasonic Lumix DMC-LX7), and two camera calibration models.

2. Related Work

The feasibility of using HDR images as a cheap alternative to measuring luminance has been studied before. Inanici [Ina06] measured how accurately an HDR image merged from several photographs taken at different exposure can capture the luminance in a scene, as compared with the values measured with a luminance meter. They used a 5.1 megapixel Nikon Coolpix 5400 camera with a 7.18 x 5.32 mm sensor and a Minolta LS110 luminance meter. Their results for color patches photographed under varying illumination indicate that greyscale targets generally produce lower luminance error than colored targets when photographed. Altogether, they reported luminance error in the range of 0 to 40% of the measured luminance.

An another study that evaluated the accuracy of luminance measurement using HDR images was done by Anaokar and Moeck [AM05]. The study measured the dependence of the error of captured luminance on the measured color. The authors photographed Munsell chips with a Nikon Coolpix 5400 camera. Their results indicated that the measurement error increases as the Munsell value decreases. The error was independent of the illuminance level and ranged between 0 and 60% of the expected illuminance, depending on the used illuminant.

As opposed to the previous studies, which used cameras equipped with a CCD array sensor, a study by Moeck [Moe07] used a camera with a CMOS sensor (Canon EOS 350 D) to find the error in luminance measurement in HDR images. The test was conducted by photographing 16 matte gray cards and 140 Munsell color checkers. Just like in the previously mentioned studies, Photosphere was used to merge the photographs into an HDR image. The results indicate that the error in luminance reproduction is the highest when dark colors with high chroma are captured. Hues with the largest error were blue, purple blue and blue green. The lowest errors were measured for yellow red, yellow and yellow green hues. On average, the luminance errors reported vary between 6% and 53% depending on the hue and value of the Munsell sample.

Krawczyk et al. [KGS05] described a photometric calibration of 3 HDR cameras: HDRC VGAx, Silicon Vision Lars III and Jenoptik C14. The authors fitted the response curves of the tested cameras and inverted them to recover the luminance values of color patches on 3 scenes spanning 8 orders of magnitude. The reported relative error varied between cameras but was generally higher for lower luminance levels (up to 120%) and below 20% in the higher luminance range. The error was attributed to the sensor noise and, at least partially, the ND filters used.

The dissertation of Tyukhova [Tyu12] assessed the suitability of using HDR photography for measuring the luminance of a Cree XP-E light emitting diode (LED). Two cameras were used in the study, a Canon EOS 7D and a Canon EOS Rebel T1i (500D). After extracting their response curves, both cameras were used to take photographs with different exposures and merge the resulting images (RAW images using *raw2hdr* [War08] and JPEG images using Photosphere) into HDR images. The HDR images fused from RAW photographs showed lower error stemming from vignetting. The final LED luminance measured with a camera showed a relative error of 1.5% compared to the value obtained through illuminance measurements. The same result could not be achieved with a luminance meter due to the small size od the LED, causing an error of about 35%.

Goesele et al. in [GHS01] explored possibility of color-calibrating HDR images in the situation when the ICC profile of the camera is known so the tone curve need not be extracted from the photographs and the color primaries are known. They used a set of images of the ANSI IT8 color target captured with a Kodak DCS 560 camera. An ICC camera profile was then used to convert the images into XYZ colorspace with 16 bits per pixel. After testing whether the XYZ space created using the ICC profile was linear the authors noted that the resulting response curves are not smooth and can cause non-linearities and noise. Finally, all exposures were merged into an HDR image using the ICC profile and the error in color reproduction was measured using the CIE76 ΔE^*_{ab} measure. Their results showed that 50% of image pixels had a CIE76 ΔE^*_{ab} error value lower than 2, with the highest error greater than 12.

3. Camera Models

We consider two models, which explain trichromatic CIE XYZ 1931 coordinates (measured with the photospectrometer) from the RGB values registered in HDR images. The first model has the form:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{13} \\ m_{31} & m_{32} & m_{13} \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

The second model extends the first one by incorporating an additional noise or black level:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \\ 1 \end{bmatrix} \tag{2}$$

Our goal is to find the transformation matrix M, given RGB values from an HDR image and XYZ values measured with the photospectrometer. Such a matrix can be easily found by solving a linear least-squares problem. However, the least square error in the XYZ color space is not the best error metric for perceived differences in color or luminance. It is more desirable to minimize the error in terms of the color-difference metric, such as CIE2000 ΔE. Therefore, to find the best transformation matrix we minimize the error:

$$\begin{aligned} E = \sum_{k=1}^{N} \Delta E(XYZ_m[k], XYZ_c[k]) + \\ + \lambda \left(log(Y_m[k]) - log(Y_c[k]) \right)^2 , \end{aligned} \tag{3}$$

where $XYZ_m[k]$ are the XYZ colorimetric values of the color patch k, as measured by the photospectrometer. $XYZ_m[k]$ are the XYZ coordinates predicted from RGB values using the current model M (right-hand-side of Equation 1 or 2). ΔE is the CIE2000 color difference metric. We use the white patch in each image as the white point required by the metric. The second term, which is weighted by $\lambda = 0.01$, is the error in log luminance values. Such an error term is necessary, as CIE2000 ΔE metric operates on relative colorimetric values and disregards absolute luminance levels. To find the model, we run a non-linear solver using Equation 3 as an error function. The solver is initialized with the least-square solution.

4. Measurements

To collect data for calibration (training) and validation (testing), we photographed four different scenes with three cameras. Each scene contained several diffuse color patches, which were measured with a photospectrometer.

Scenes. The first scene contained X-Rite ColorChecker Chart and custom made color patches (see Figure 2). The only source of illumination were two photographic lights

Scene	*Camera*	*Shutter speeds*	*ISO*	*Aperture*
ColorChecker	550d	1/16; 1/8; 1/4; 1/32; 1/64	100	f/5.7
	1000d	1/16; 1/8; 1/4; 1/32; 1/64	100	f/5.7
	Lumix	1/13; 1/50; 1/3.2	80	f/5.6
Conference	550d	1/5.2; 1/2.6; 1/1.3; 1/10.4; 1/20.7	100	f/8.0
	1000d	1/5.2; 1/2.6; 1/1.3; 1/10.4; 1/20.7	100	f/8.0
	Lumix	1/13; 1/50; 1/3.2	80	f/5.0
Library	550d	4; 8; 16; 32; 2; 1; 1/2; 1/49.4; 1/197.4; 1/1024	100	f/8.0
	1000d	4; 8; 16; 32; 2; 1; 1/2; 1/49.4; 1/197.4; 1/1024	100	f/8.0
	Lumix	1/13; 1/100; 1/1.6	80	f/5.0

Table 1: *Capture settings for the exposures.*

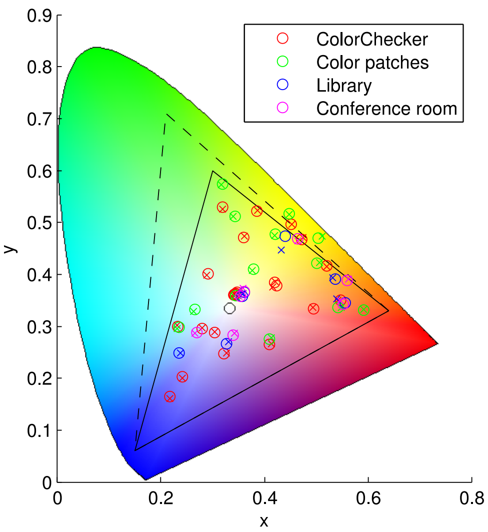

Figure 1: *Chromatic coordinates of all measured colors for four test scenes. Circles correspond to photospectrometer measurements and crosses to the calibration results for Canon 550D. Calibration for other cameras produced similar results. The solid line denotes the sRGB color gamut and dashed line the Adobe RGB gamut.*

(Tricolor 1400W Ultra Cool Day Light), positioned at a 45 degree angle to the chart on both sides. The camera was mounted in the middle on a tripod. We used that scene for two data sets: "ColorChecker", which was used for training, and "Color patches", which was used for testing. Note that we did not use any non-diffuse patches in the "Color patches" data set.

The two other scenes used for testing, "Conference room" (Figure 3) and "Library" (Figure 4), were illuminated by one of the photographic lights as well as the existing light fixtures. We placed in these scenes seven measurement targets, which were made from a diffuse colored paper glued to cardboard pieces.

Cameras. HDR multi-exposure photographs were taken with three cameras: *Canon 550D, Canon 1000D and Panasonic Lumix DMC-LX7*. The goal was to find how accurately each camera can be color-calibrated. Each camera was in turn mounted on a tripod to take RAW images at 5 different exposures at 1 f-stop interval in case of Canon cameras, and 3 exposures at 2-fstop interval in case of the Panasonic camera. Because of very high contrast in "Library" scene, we captured 10 exposures to avoid any over- or under-exposed pixels. The middle exposure was selected using camera's metering. The list of camera setting for each scene can be found in Table 1. Canon cameras were controlled from a PC using *gphoto2* software and auto-bracketing was used to take pictures with the Panasonic camera.

HDR merging. Multiple exposures were merged into HDR images using *pfstools* software [MKMS07]. An example command that was used is shown below:

```
pfsinme *.CR2 | pfshdrcalibrate -r linear -v
      --bpp 16 | pfsout 550d_cc1.hdr;
```

The script `pfsinme` executed `DCRAW` software [Cof20] on each exposure with the arguments `-c -o 0 -4 -w`. The arguments ensured that each RAW photograph was decoded as a 16-bit linear (non-gamma corrected) image in a native camera color space. Then, a custom matlab script was used to manually select center of each patch, which was used to calculate an average RGB color value.

XYZ measurements. The reference CIE XYZ values were measured using a Specbos 1211 spectroradiometer connected to a PC. The measuring spot of the spectroradiometer was adjusted so as to fit completely within each individual color patch. The device took 5 readings, averaged them and the results were stored in a CSV file. The chromatic coordinates for all measured colors are shown in Figure 1. The measurements well cover the sRGB color gamut except deep blues and purples.

5. Calibration Results

Following the calibration procedure explained in Section 3, we compute a color transform matrix for each camera using the ColorChecker scene and use these matrices to color-calibrate the remaining scenes. The calibration errors for all the cameras and scenes are shown in Tables 2–5. Note that the error for individual patches is given only for Model A (Equation 1). For brevity, we report only average errors for

#	M	C_1	C_2	C_3	ΔE_1	ΔE_2	ΔE_3
1					0.2 (0.0)	0.1 (0.1)	0.8 (1.7)
2					1.1 (2.5)	1.2 (3.6)	1.1 (2.6)
3					0.9 (1.8)	0.7 (2.0)	0.1 (0.3)
4					0.6 (1.2)	0.4 (0.7)	1.1 (0.4)
5					1.0 (0.1)	0.8 (0.2)	0.3 (0.4)
6					0.6 (1.2)	0.7 (2.0)	1.4 (4.5)
7					1.1 (2.3)	0.3 (1.2)	0.6 (1.8)
8					0.2 (0.9)	0.6 (0.6)	0.4 (2.1)
9					0.3 (0.3)	0.4 (0.5)	0.4 (0.4)
10					0.5 (0.1)	0.6 (0.7)	0.5 (1.2)
11					0.3 (0.6)	0.4 (1.0)	0.0 (0.0)
12					0.9 (3.5)	1.1 (3.7)	0.7 (0.7)
13					0.5 (4.1)	0.9 (3.3)	0.5 (4.2)
14					0.5 (2.3)	0.3 (1.1)	1.3 (0.3)
15					0.7 (2.4)	0.5 (1.6)	1.2 (3.9)
16					0.4 (0.6)	0.7 (2.1)	0.7 (0.8)
17					0.6 (1.9)	0.5 (1.4)	0.7 (0.5)
18					1.1 (5.0)	1.6 (5.4)	2.1 (7.7)
19					0.6 (2.2)	1.1 (1.2)	0.4 (1.8)
20					0.2 (0.2)	0.4 (0.0)	0.3 (0.1)
21					0.6 (0.5)	0.1 (0.0)	0.5 (1.6)
22					0.4 (0.4)	0.7 (0.1)	0.7 (0.2)
23					0.7 (0.5)	0.6 (0.0)	0.7 (3.1)
24					0.6 (3.8)	0.5 (0.0)	1.8 (13.5)
Averaged Model A					0.6 (1.6)	0.6 (1.4)	0.8 (2.2)
Averaged Model B					0.7 (2.1)	0.8 (1.9)	0.7 (1.8)

Table 2: *Calibration results for the ColorChecker, which was a training data set. M column shows the color as measured by the spectrometer (after converting to the sRGB color space). C columns show the color after camera colorimetric calibration. C_1 is Canon 1000D, C_2 is Canon 550D and C_3 is Lumix LX-7. The colors shown in the table have been transformed into the sRGB color space without correcting for white ballance. Therefore, the white patch (#19) is different from the D65 white point. ΔE is CIE2000 DeltaE color difference between the color measured with the photospectrometer and the color captured with a camera . The numbers in paranthesis are the relative error in luminance in percent ($\Delta L/L$). The errors for each color patch are provided only for Model A (Equation 1).*

Figure 2: *"Color Checker" on the left and "Color patches" on the right. Only the bottom target was used for testing as the top one contained non-diffuse materials with a complex BRDF. The HDR image was converted to the sRGB color space using the luminance of the white patch $\times$ 1.5 as the peak luminance.*

#	M	C_1	C_2	C_3	ΔE_1	ΔE_2	ΔE_3
1					0.5 (1.1)	0.7 (0.8)	0.8 (2.3)
2					0.4 (0.8)	0.7 (1.4)	1.0 (2.1)
3					0.8 (1.2)	0.5 (0.1)	1.8 (7.5)
4					0.8 (1.8)	0.4 (1.4)	1.0 (4.7)
5					1.3 (5.4)	1.1 (4.6)	1.8 (2.3)
6					0.5 (1.2)	0.5 (1.6)	0.5 (1.4)
7					0.5 (1.9)	1.0 (3.6)	1.2 (0.9)
8					0.5 (0.3)	0.9 (2.7)	1.7 (0.7)
9					1.0 (4.3)	1.2 (4.9)	0.4 (1.0)
10					1.3 (3.6)	1.7 (3.0)	2.6 (1.3)
11					1.1 (3.9)	1.2 (4.7)	0.5 (1.2)
12					1.2 (3.6)	1.7 (6.3)	1.1 (2.8)
13					0.6 (1.7)	1.3 (4.7)	1.8 (1.2)
Averaged Model A					0.8 (2.4)	1.0 (3.1)	1.2 (2.3)
Averaged Model B					0.8 (2.4)	1.0 (2.8)	1.2 (2.2)

Table 3: *Calibration results for the "Color patches" scene. The labels are the same as in Table 2.*

Model B (Equation 2). The color patches corresponding to each row in these tables can be found in Figures 2–4.

Both models result in very similar calibration error (refer to Table 2). However, Model B gives worse predictions for scenes "Conference room" and "Library". This is unexpected given that this model offers more degrees of freedom. Its worse predictions are most likely caused by overfitting. The ColorChecker may not provide enough samples for accurate calibration. If the calibration is to be limited to 24 patches in the ColorCheker, there seems to be no benefit of a more complex model B with a black level.

The color errors are relatively small across all images, though they are larger for the more complex scenes "Library" and "Conference room". Only in case of a few patches CIE2000 ΔE error exceeds 4 (marked with red color in the tables), which is often considered a just-noticeable-difference (JND). Note, however, that actual ΔE error corresponding to 1 JND may differ between colors as CIE color difference formula gives only approximately perceptually uniform distance measure. The visual similarity of the measured and calibrated colors is clearly seen in all the tables, where it is difficult to discern between colors shown in each row, which were captured by the photospectrometer and each camera. The same can be said about the images that were calibrated using the computed matrices. The bot-

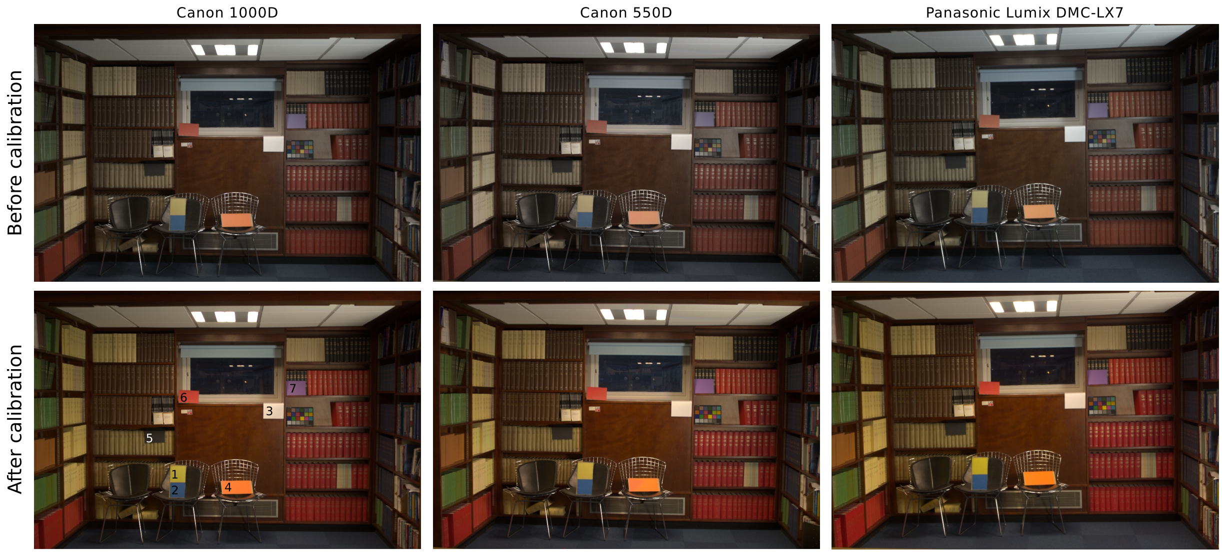

Figure 4: *Comparison of images before and after calibration for the "Library" scene. The numbers correspond to sample numbers in Table 5.*

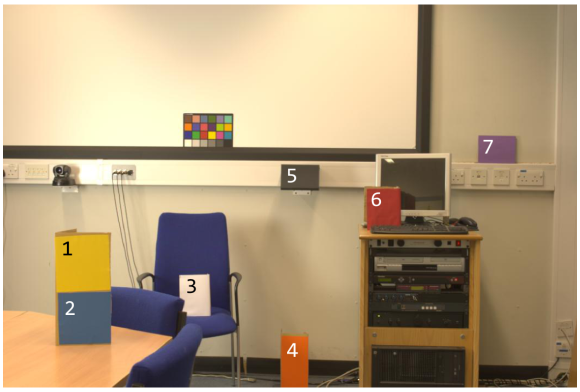

Figure 3: *"Conference room" scene. The numbers correspond to sample numbers in Table 4.*

#	M	C_1	C_2	C_3	ΔE_1	ΔE_2	ΔE_3
1					0.8 (2.4)	3.2 (12.4)	1.5 (4.5)
2					1.4 (5.1)	2.7 (12.5)	1.9 (9.8)
3					1.1 (4.5)	3.0 (11.7)	2.4 (8.7)
4					1.6 (7.1)	2.2 (8.6)	2.6 (9.8)
5					2.1 (12.2)	2.5 (9.3)	1.9 (12.4)
6					2.8 (15.6)	2.2 (8.5)	2.8 (15.5)
7					0.5 (0.3)	4.6 (19.8)	0.9 (2.5)
Averaged Model A					1.5 (6.8)	2.9 (11.8)	2.0 (9.0)
Averaged Model B					1.2 (3.8)	3.2 (14.3)	2.5 (4.7)

Table 4: *Calibration results for the "Conference room" scene. The labels are the same as in Table 2.*

#	M	C_1	C_2	C_3	ΔE_1	ΔE_2	ΔE_3
1					4.1 (10.9)	5.3 (7.4)	1.3 (4.8)
2					1.1 (5.9)	1.5 (8.0)	0.6 (0.4)
3					2.8 (7.6)	3.6 (5.2)	2.7 (4.5)
4					4.2 (14.8)	1.8 (5.6)	3.0 (8.3)
5					0.7 (2.9)	0.7 (8.6)	0.7 (3.4)
6					0.8 (1.9)	2.0 (0.6)	1.4 (0.6)
7					2.2 (8.7)	0.7 (0.8)	1.4 (5.5)
Averaged Model A					2.3 (7.5)	2.2 (5.2)	1.6 (3.9)
Averaged Model B					4.5 (17.3)	3.6 (5.5)	4.4 (14.0)

Table 5: *Calibration results for the "Library" scene. The labels are the same as in Table 2.*

tom row in Figure 5 shows little color difference between images captured with different cameras.

The relative difference in luminance (values given in parenthesis in the tables) are lower than those reported in previous studies. For the scenes with uniform and well controlled illumination ("ColorChecker" and "Color patches"), the relative error was mostly below 5%. The two exceptions were the black color (ColorChecker #24) for the Panasonic camera and orange (Color patches #5) for Canon 1000D. The relative luminance error increases significantly (up to 15%) in more complex scenes. Orange, red and black appear to be the most problematic colors. This could be due to glare, vignetting, imprecise measurements and other factors.

We could not find an explanation as to why colors captured by Canon 550D in the "Conference room" scene (column C_2 in Table 4) were darker than for other cameras. For this shot, all camera paramaters were the same for both

Canon 1000D and Canon 550D. The photographs were also taken in the order C_1, C_2 and C_2, so we could not attribute this darkening to the change of illumination.

Although cardboard color targets were identical in "Library" and "Conference room" scenes, the colors of some of the patches are very different, especially when seen in isolation (compare corresponding colors is Tables 5 and 4). This demonstrates the strong influence of illumination and shading on measured colors.

In overall, ΔE values indicate very good quality of calibration for all tested cameras. High quality monitors can be calibrated with the mean ΔE between 2 and 3, which is a higher error than that achieved for our measurements. Interpreting these results, however, requires some caution. The test colors gave good coverage of the sRGB color gamut in terms of chromatic coordinates, but they lacked large luminance variation and deeply saturated colors available in extended color gamuts, such as Adobe RGB. Our calibration matrices contain negative values, indicating that some primaries lie outside the visible gamut. We could not find an explanation for this result.

6. Conclusions

Three consumer cameras were color-calibrated to capture accurate colorimetric values in HDR images using the multi-exposure technique. The color calibration tested whether introducing black level in the camera model improves the accuracy of the fit but no improvement was found. Additionally, a new camera model fitting approach was used, one that minimizes the perceived color error rather than the objective error used in the least square fitting. The results show that when camera RAW images are used for merging into an HDR image, the colors can be calibrated with high accuracy. As opposed to the previous studies, the accuracy of not only luminance but also color capture was measured. Most of the differences between the colors measured with photospectrometer and those captured by the camera were below the perceiveable threshold.

In the future work, we would like to validate our calibration using more color samples, especially those that are outside the sRGB color gamut. We will use scenes with larger luminance variation. We will also elimitate lens distortions, such as vignetting and glare, which have been shown to decrease the quality of the fit in previous studies.

References

[AM05] ANAOKAR S., MOECK M.: Validation of high dynamic range imaging to luminance measurement. *Leukos 2*, 2 (2005), 133–144.

[Cof20] COFFIN D.: DCRAW software, 2013, version 9.20. http://www.cybercom.net/ dcoffin/dcraw/.

[GHS01] GOESELE M., HEIDRICH W., SEIDEL H.-P.: Color calibrated high dynamic range imaging with icc profiles. In *Color and Imaging Conference* (2001), vol. 2001, Society for Imaging Science and Technology, pp. 286–290.

[Ina06] INANICI M.: Evaluation of high dynamic range photography as a luminance data acquisition system. *Lighting Research and Technology 38*, 2 (2006), 123–134.

[KGS05] KRAWCZYK G., GOESELE M., SEIDEL H.-P.: *Photometric calibration of high dynamic range cameras*. Tech. rep., Max-Planck-Institut für Informatik, 2005.

[MKMS07] MANTIUK R., KRAWCZYK G., MANTIUK R., SEIDEL H.-P.: High dynamic range imaging pipeline: perception-motivated representation of visual content. In *Human Vision and Electronic Imaging* (2007), pp. 649212–12.

[MN99] MITSUNAGA T., NAYAR S.: Radiometric self calibration. In *Proc. of CVPR* (1999), CVPR, pp. 374–380.

[Moe07] MOECK M.: Accuracy of luminance maps obtained from high dynamic range images. *Leukos 4*, 2 (2007), 99–112.

[MR07] MCCANN J. J., RIZZI A.: Camera and visual veiling glare in HDR images. *Journal of the Society for Information Display 15*, 9 (2007), 721.

[Tyu12] TYUKHOVA Y. I.: *The Assessment of High Dynamic Range Luminance Measurements with LED Lighting*. Master's thesis, University of Nebraska, 2012.

[War08] WARD G.: raw2hdr software. `http://www.anyhere.com/gward/pickup/HDRcapOSX.tar.gz`, 2008.