The Impact of Illumination on the Perceived Quality of Spectral Reproductions

Jana Blahová, Steven Le Moan and Philipp Urban

Institute of Printing Science and Technology, Technische Universität Darmstadt
Magdalenenstraße 2, 64289 Darmstadt, Germany
eMail: blahova@idd.tu-darmstadt.de
URL: http://www.idd.tu-darmstadt.de/color

Abstract. We introduce the framework of a visual experiment to evaluate image quality under several illuminants, for (multi-/hyper-) spectral image quality assessment. We present the results obtained from 42 observers, 30 images and 5 illuminants.

1 Introduction

The advent of new image reproduction technologies, particularly spectral reproduction in printing, induces a growing demand on assessing the perceived quality of such reproductions. In contrast to typical metameric reproductions (as addressed by the International Color Consortium (ICC)), spectral reproductions are not only optimized for one viewing condition but aim to mimic the colors of the original image for a wide variety of viewing conditions, particularly different illuminants.

Various methods for evaluating spectral reproductions were developed so far. However, very few studies have been dedicated to the interplay between spectral reproduction capabilities and their effects on human perception under various illuminants [1, 2]. The *Spectral Image Quality* measures proposed in these studies take two spectral images and return a single value representing the perceived difference between these images, w.r.t. several illuminants. They may be used not only to judge the quality of spectral reproductions but may serve also as objective functions to optimize spectral reproduction workflows (e.g. spectral gamut mapping, spectral compression or spectral reconstruction of camera captures).

For judging the relevance of these quality measures, visual experiments are crucial. In this paper, we present an experimental framework designed to evaluate *Spectral Image Quality* measures, and motivated by the question: *How does the perceived image quality change w.r.t. different illuminants?* We first describe thoroughly the experimental setup, before presenting and discussing the results obtained.

2 Experimental Setup

2.1 Liquid crystal display

The experiments were conducted in a completely darkened room to minimize flare. The liquid crystal display, Eizo ColorEdge CG301W, used for the experiment was the only light source in the room. It has been warmed up for one hour before calibration and characterization.

The white point (R, G, B) = (255, 255, 255) was adjusted in the calibration to 120 cd/m^2 and x and y values corresponding to the investigated illuminant. The black level was set to the minimum and gamma to 2.2. The calibration measurements were performed by an X-Rite EyeOne Spectrophotometer.

To ensure that tristimuli shown to observers were similar to CIEXYZ values of the spectral image when observed under the investigated illuminant by a 2° observer, characterization method similar to that of Day et al. [3] was employed. This method showed average characterization errors bellow 1 CIEDE2000 in previous work [4]. All required measurements for characterization were performed by the Konica-Minolta Spectroradiometer CS-1000A. Note, that calibration and characterization had been carried out each day of the experiment dedicated to a particular illuminant.

2.2 Illuminants

Five different illuminants were considered for this experiment. They were selected from a database of 74 natural spectra [2], aiming to have minimum pairwise similarity of their spectral power distributions. This should allow a large variety of renderings for a single spectral image. The five chosen spectra were the following:

- *CIED65*: Standard CIE, daylight (CCT 6500K)
- *CIEA*: Standard CIE, tungsten filament (CCT 2856K)
- *CIEF11*: Standard CIE, fluorescent (CCT 5000K)
- *Lamina WW-NB*: Light-emitting diode (CCT 3236K) [5]
- *Luxina EXZ-CG-M250*: Tungsten halogen (CCT 2969K) [5],

where CCT stands for Correlated Colour Temperature [6]. All spectra were normalized to the same radiance range, as shown in Fig. 1.

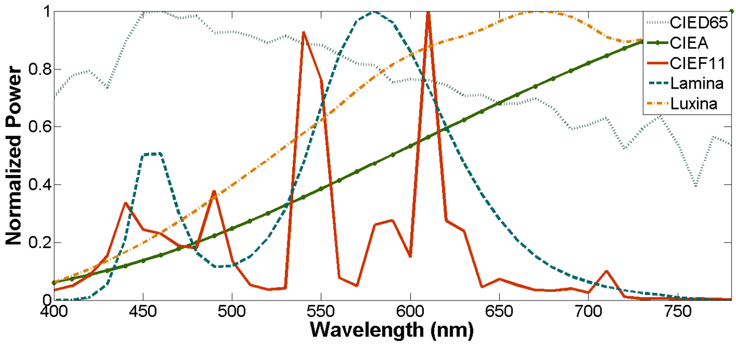

Fig. 1: Normalized spectral power distribution of used illuminants

One can observe that each spectrum has a quite different shape. The daylight peaks in the blue wavelengths (before 500nm), the fluorescent and LED-based lights peak around 600nm (green range), while the tungsten lights are much warmer and radiate mostly in the red range (after 650nm).

Each day of the experiment was dedicated to a particular illuminant, for which the display was calibrated and characterized. The tristimulus values shown to the observers were adjusted to match the one of the spectral image when illuminated by this illuminant.

2.3 Test Image

A 400 x 400 pixel square stimulus was displayed in the centre of the LCD to verify the x, y value of the white point for the CIE1931 2° standard observer. The measurement of x and y values of white digital counts ($d_r = d_g = d_b = 255$) was accomplished with the Konica Minolta Spectroradiometer CS-1000A. Ideal and measured x and y values of the white points are depicted in Tab. 1 and in the CIE 1931 xy chromaticity diagram (see Fig. 2).

Tab. 1: Ideal and measured x and y values of the white points

Illuminant	**Target**		**Measured Result**	
	x value	**y value**	**x value**	**y value**
CIED65	0.3127	0.3290	0.3130	0.3294
CIEF11	0.3457	0.3585	0.3483	0.3546
Lamina	0.4221	0.4019	0.4230	0.3985
Luxina	0.4381	0.4066	0.4428	0.4037
CIEA	0.4448	0.4080	0.4495	0.4042

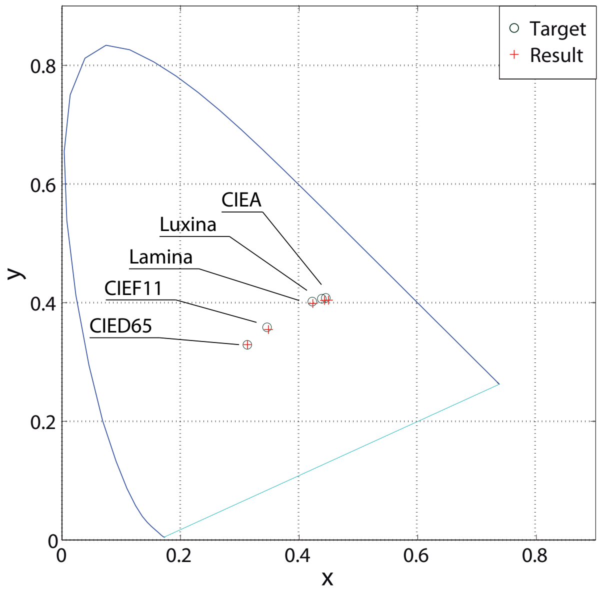

Fig. 2: CIE 1931 xy chromaticity diagram

2.4 Images

A total of 10 multispectral images of natural outdoor scenes were used. They were selected from Foster's 2002 and 2004 databases [7]. They contain between 30 and 32 spectral channels in the visible wavelength range.

In order to have a basis for comparison, we distorted the spectral images via three methods:

- *Paramer mismatch-based spectral gamut mapping* (Reproduction 1, referred as R1): With the spectral gamut of a Canon iPF5000 printer utilizing CMYKRGB inks, we applied the method presented in [8].
- *Wavelet-based compression* (Reproduction 2, R2): We quantized the tensor product of a 2D-wavelet transform and a cosine transform on the spectral cube [9] so as to perform a JPEG2000-like compression.
- *Spectral camera model* (Reproduction 3, R3): We simulated how a customized 6-channel Sinar camera would acquire the same scene. Reflectance curves were then reconstructed by the pseudo-inverse, using some samples of the MacBeth ColorChecker target.

Figures 3 and 4 show examples of a scene rendered under illuminants CIED65 and CIEA, respectively. The images were magnified for comparison. Under illuminant CIED65 (Fig. 3) most observers preferred R1 (67%) rather than R3 (33%). For illuminant CIEA (Fig. 4) it was vice versa (R1 11% vs. R3 89%). This confirms the importance of the scene illuminant for judging image quality.

Fig. 3: Scene example under illuminant CIED65
(top: reproduction 1, middle: original image, bottom: reproduction 3)

Fig. 4: Scene example under illuminant CIEA
(top: reproduction 1, middle: original image, bottom: reproduction 3)

2.5 Observers

Two groups of observers with color-normal vision participated in the experiment: a constant group of 12 observers attended the experiment every day and a group of 6 observers varying each day. A total of 42 observers were recruited for the experiment. Both groups included half women and half men with ages ranging from 21 to 57 years. The color-normal vision of each observer was tested using the Ishihara Test and Farnsworth-Munsell D15 Test in an X-Rite SpectraLight III viewing booth under day light (CIED65).

The visual resolution was set to 40 pixels/degree (of the visual field) by a chin rest.

2.6 Task

A paired comparison strategy was used to evaluate the global quality in terms of perception. For each trial, three images were shown on the display, the original in the middle and one reproduction on each side. All images were rendered for the same illuminant. Observers were asked to select the distorted image that is the most similar to the original. Tie decisions were allowed. All pairs of reproductions were compared under all five illuminants. A total of 2700 observer decisions were recorded in the experiment (5 illuminants x 10 original images x 3 comparisons/original x 18 observers/illuminant).

3 Results and Discussion

Figure 5 depicts the number of times that each reproduction was preferred, for each illuminant. Our results indicate that, by changing the illumination, the observers' preferences may change, suggesting that the perceived quality of the three reproductions differs. We can see a huge preference of R1 under illuminant CIED65. In this case, the images were optimized for the mentioned illuminant and very similar to the original image. On the other hand we note a lower preference for R1 under the other illuminants, mainly due to a worse chromatic fidelity.

R2 is characterized by structural distortions while keeping a fairly good overall color quality. Based on our results, R2 has been preferred in all illuminants except CIED65. We can summarize that observers prefer structural distortions in the image rather than large hue and chroma shift, for the considered visual resolution.

Also, there seems to be a correlation between the shape of the illuminants and the observer's preference for the reproduction methods considered. We observe that there are three trends in the following graph: CIED65, which has a maximal radiance in the blue wavelengths, ranks R1, R2 then R3, while CIEF11 and Lamina, which both are particularly reflective in the middle-range (greens), rank R2, R3 then R1. Finally, Luxina and CIEA both reach their maximal radiance in the red end of the spectrum, rank R2, R1 then R3.

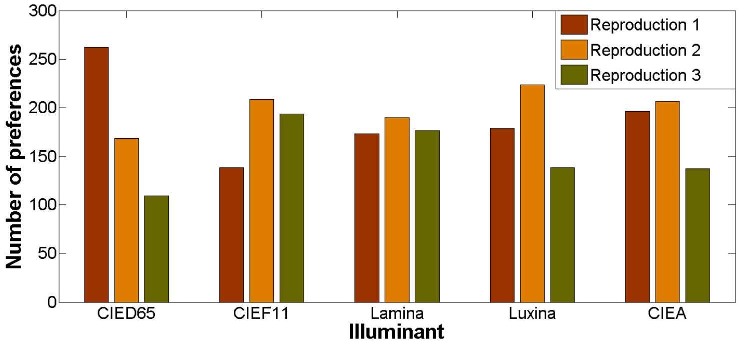

Fig. 5: Observers' ratings (number of times that a reproduction was preferred)

4 Application of Experimental Data

Our data can be used to evaluate the prediction performance of spectral image-difference measures (IDMs). One way of evaluation is provided through *hit rates* ***i*** [10]. The hit rate ***p*** is calculated as the ratio between the number of hits (the IDM predicts the observer choices correctly) and the total number of comparisons ***m***, i.e. $p = i/m$.

Tie scores can either be discarded or considered as half-hit (0.5). A high hit rate indicates a good ability of the IDM to predict the observers' choice. A rate of 0.5 corresponds to a random prediction. The maximum achievable rate is referred to as *majority hit rate* p_m. It corresponds to the rate that would be obtained if the measure constantly agreed with the majority of the observers. Therefore, it is possible to rescale the hit rates w.r.t. the range [0.5, p_m], in order to account for inter-observer uncertainties.

To test if two IDMs are significantly different on the experimental data a two-sample binomial test proposed by Yule (explained in [11]) can be used [10].

5 Conclusions and Future Work

Motivated by the recent advent of spectral image reproduction technologies, we introduced a psychophysical experimental setup for image quality assessment under various illuminants. Thirty images (ten scenes, three reproductions) were compared pairwise in random sequences under five selected illuminants. A total of 42 observers participated in the experiment, which was spread over 5 consecutive days, each one being dedicated to a single illuminant.

The results were evaluated using the aggregate judgment for each illuminant. They led us to conclude that the observer's preference is not constant across illuminants, which confirms that there is a need for considering multiple illuminants when it comes to spectral image quality and opens new challenges for further investigation. We also noted that, when an observer had to choose between a severe lightness-structural distortion and a severe chromatic distortion, he/she would most likely prefer the structural artifacts (such as the ones produced by a JPEG2000 compression for the visual resolution used in the experiment). Finally, we noted that there is an apparent correlation between the shape of the illuminant spectral power distribution curve and the observer ratings.

The results of this visual experiment may be used to evaluate the accuracy of spectral image quality measures, which is an emerging research field. In future work, we propose a deeper investigation of the relationship between the shape of the illumination spectral power distribution and the observer's judgments.

We publish all the data corresponding to this experiment, including Matlab code, images and observer decisions on our website [12].

References

[1] P. Morovič, J. Morovič, J. Arnabat, J. M. García-Reyero. Revisiting Spectral Printing: A Data Driven Approach. In: *Color and Imaging Conference*. Society for Imaging Science and Technology, 2012. p. 335-340.

[2] S. Le Moan, P. Urban. Evaluating the Perceived Quality of Spectral Images. In: *Image Processing, International conference on*. IEEE, 2013, to be published.

[3] E. A. Day, L. Taplin, R. S. Berns. Colorimetric characterization of a computer-controlled liquid crystal display. *Color Research & Application*, 2004, 29.5: 365-373.

[4] P. Urban, M. Fedutina, I. Lissner. Analyzing small suprathreshold differences of LCD-generated colors. JOSA A, 2011, 28(7), 1500-1512.

[5] "Spectral power distribution curves, The National Gallery: http://research.ng-london.org.uk/scientific/spd/?page=spd (last check: 2013/08/29)"

[6] G. Wyszecki, W. S. Stiles. Color Science: Concepts and Methods, Quantitative Data and Formulae. Second Edition, John Wiley & Sons, New York, 1982.

[7] D. H. Foster, K. Amano, S. M. C. Nascimento, M. J. Foster. Frequency of Metamerism in Natural Scenes. *JOSA A*, 2006, 23.10: 2359-2372.

[8] P. Urban, R. S. Berns. Paramer mismatch-based spectral gamut mapping. *Image Processing, IEEE Transactions on*, 2011, 20.6: 1599-1610.

[9] G. Peyré. Numerical tour - multispectral image compression: https://www.ceremade.dauphine.fr/~peyre/numerical-tour/tours/multidim_3_multispectral/#15 (last check: 2013/08/29),"

[10] I. Lissner, J. Preiss, P. Urban, M. S. Lichtenauer, P. Zolliker. Image-difference prediction: From grayscale to color. Volume 22, pages 435–446, 2013.

[11] L. Brown, X. Li. Confidence intervals for two sample binomial distribution, Journal of Statistical Planning and Inference, vol. 130, no. 1-2, pp. 359–375, 2005.

[12] "Supplementary material (visual experiment data) available at: http://www.idd.tu-darmstadt.de/color/papers (last check: 2013/08/29),".