Observer Variability in Color Image Matching on a LCD monitor and a Laser Projector

***Yuta Asano**[1,2]**, Mark D. Fairchild**[1]**, Laurent Blondé**[2]**, Patrick Morvan**[2]*
*[1] **Munsell Color Science Laboratory, Rochester Institute of Technology; Rochester, NY, USA***
*[2] **Technicolor; Rennes, France***

Abstract

Wide color gamut media are emerging in the market, and this trend has been accelerated by ITU recommendation, Rec.2020 in 2012. Wide color gamut media possess spectrally narrow primaries, which would potentially increase the degree of observer metamerism. In this study, it was investigated if observer metamerism could be a serious issue under practical viewing conditions. Namely, real images were used as a matching stimulus instead of uniform colors. We carried out the color image matching experiment on two different media: an Apple Cinema HD LCD monitor and a Microvision laser projector. The results from 28 color-normal observers were analyzed. The obtained inter-observer variability was large enough that observer metamerism would be a serious issue where the laser projector is viewed together with conventional media. Each observer had a match point that was significantly different from those of other observers. It was found that effective field size changes (and an observers CMFs change) depending on image contents. Complex images require smaller field size whereas simple images require larger field size.

Introduction

In motion picture industries, color grading (also known as color timing) is the process where the colors of raw movie contents are corrected by colorists to give artistic effects directed by cinematographers. The issue often encountered in such application is that media (usually displays or projectors) calibrated for the standard observer does not appear correct for a human observer. Another issue is that colors/images on media matched to a reference monitor by an observer do not appear matched for another observer. These phenomena are called observer metamerism, and result from individual differences in color matching functions (CMFs).

Some researchers investigated if observer metamerism poses any serious issues in practical applications. In 2010, Sarkar et al. conducted a color matching experiment on a CRT and a wide-gamut LCD displays aiming for color grading application [1]. Inter-observer variability was twice as large as intra-observer variability, and they concluded that observer metamerism could be a serious issue in practice. Similar work has been done in soft-proofing application (color matching on hard-copy and soft-copy). Alfvin and Fairchild [2] and Rich and Jalijali [3] used hard-copy (color patches illuminated by a light source) as reference colors and soft-copy shown on a CRT monitor as matching stimuli. Both of them confirmed large observer variability that could lead to a serious issue. On the other hand, Pobboravsky [4] used hard-copy and a CRT monitor for the experiment, and concluded that variability in normal color vision does not pose serious issues for the comparison of soft and hard proofs. Oicherman et al. [5] used hard-copy as reference colors and a CRT monitor and a LCD monitor as matching stimuli. The conclusion was the same as Pobboravsky. Note that the stimuli were only uniform colors in all the above-mentioned studies. Although its necessity was stated [4], no study has been made to investigate matching color of images, which would happen more often in reality. It is known that human CMFs change depending on the stimulus size (visual angle, or field size [6]). In practice, motion picture and imaging industries continues to use the CIE 1931 2° standard observer since its inception. Some point out that 2° CMFs are more relevant than 10° CMFs because people tend to focus on smaller areas of an image when judging colors although the image itself subtend more than 2°. Others oppose to this argument. Thus, it is relevant to investigate what would happen if real images instead of uniform colors were viewed. Moreover, the study by Oicherman et al. [5] is the only work so far which is close to practical viewing conditions in that origins of stimuli are known. Other studies used bipartite fields with minimum separation of stimuli, and/or presenting reference white, which would maximize the color discrimination ability, confidently define the visual angle, and stabilize the adaptation. Such well-controlled viewing conditions are advantageous to produce steady and reliable results. In this study, we are interested in the condition more typical for practitioners.

Observer metamerism would become a more serious issue because of the advent of wide color gamut displays/projectors which consist of spectrally narrow primaries. Fairchild and Wyble [7] and Ramanath [8] reported the spectrally narrow stimuli would magnify inter-observer variability. Furthermore, the wide color gamut media would be expected to gain more popularity soon since International Telecommunication Union (ITU) released the recommendation for ultra-high definition TV known as Rec.2020 (or BT.2020) in 2012. The color gamut proposed in Rec.2020 is extremely large, only achievable by monochromatic primaries such as lasers. In this study, we are interested in the new, emerging technology, laser projectors. Laser primaries can enlarge color gamut, but at the same time it would increase the possibility of observer metamerism. But how serious is it?

How much observer variability would be obtained under typical color grading situations? Is the obtained observer variability large to pose a serious issue? Would the average match points vary depending on image contents? Are they close to 2° or 10° prediction? The goal of this study is to evaluate potential answers to these questions. A color image matching experiment was designed and carried out using a LCD monitor and a laser projector to investigate observer variability.

Experiment

Overview

Two different media: a LCD monitor and a laser projector were used for the experiment. The LCD monitor was an Apple Cinema HD display. The laser projector was a Microvision SHOWWX+ Laser Pico Projector [9]. The spectral power distributions (SPDs) of the two media were measured by PhotoResearch 670 spectroradiometer and the maximum SPDs of red, green, and blue primaries for the two media are shown in Figure 1.

The temporal and spatial stabilities were evaluated for both media. In general, the Apple Cinema display exhibited excellent stabilities while the Microvision laser projector had relatively large temporal and spatial instabilities. Regarding temporal stability, measurements were repeated over two hours and standard deviations were taken. The standard deviations in Y, x, y [%] were 0.3, 0.05, 0.04 for the Apple Cinema display and 1.7, 0.2, 0.4 for the Microvision laser projector. Regarding the spatial stability, measurements were taken at 25 different locations uniformly sampled across a screen. The standard deviations in Y, x, y [%] within the area used in the experiment were 1.1, 0.3, 0.1 for the Apple Cinema display and 6.6, 0.3, 3.5 for the Microvision laser projector. Additionally, measurements were taken at the center of a screen, at three different viewing angles (-10°, 0°, +10°) to evaluate viewing-angle dependency. The standard deviations in Y, x, y [%] were negligible (less than 1 %) for both media. The temporal and spatial instabilities of the Microvision laser projector are in Y value (luminance) but not in chromaticities. As discussed later, humans become less sensitive to lightness (luminance) differences when a large separation between stimuli is introduced. Thus, we could assume that the temporal and spatial instabilities in Y value would not pose any serious issue.

Display models were developed to convert digital counts to colorimetric values and vice versa. The display model for the Apple Cinema display was similar to the one proposed by Day et al [10], which has 1D lookup tables and a matrix. The mean and maximum errors (CIEDE2000, ΔE_{00}) of the Apple Cinema display model for validation dataset were 0.2 and 0.5. The display model for the Microvision laser projector was the model with 3D LUTs. 3D LUTs were adopted because the projector exhibited significant additivity failures, which makes it difficult to apply the Day's model. Besides, the bit depth for the green primary of the projector was 6-bit and those for the red and blue primaries were 5-bits. In such case, the inverse display model (from colorimetric values to RGB values) for the projector was of little use. Therefore, the reference images were always shown on the Microvision laser projector and only the forward display model was used. The mean and maximum errors (ΔE_{00}) of the Microvision display model for validation dataset were 1.0 and 3.9. Given the relatively large temporal instability for the Microvision projector, this model performance is reasonable.

Setup

The experiment setup is shown in Figure 2. The projector screen was selected so as to minimize speckles that are major issues in laser projectors. Black cardboard frames were used for both the projector screen and the LCD screen to produce similar appearance. The two screens were not parallel but slightly tilted so that both of them became perpendicular to the observer's line of sight. The image width and height were 15° and 11° in visual angle (16cm and 12cm), respectively. The separation between two images was 13° in visual angle (14cm). The distance between the observer and the screens was approximately 60 cm. The surround was dark and there was no adaptation stimulus in the experiment setup, which is typical in color grading. The Microvision projector was used as a reference media; the reference images were shown. The Apple Cinema display was used as a matching media; the observer adjusted images shown on this media.

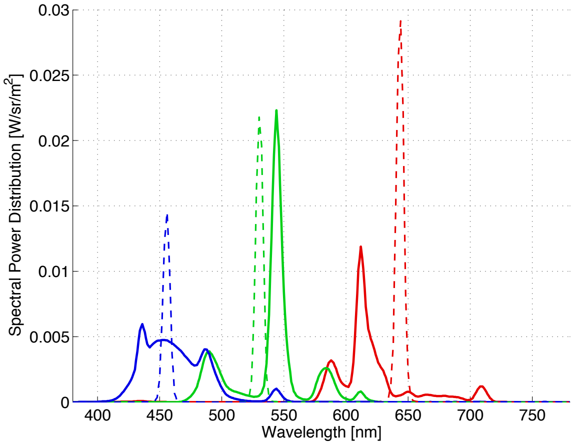
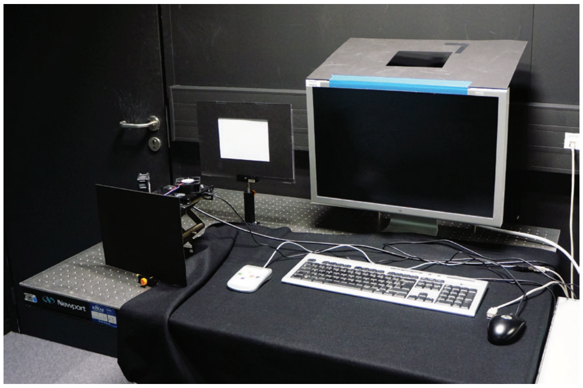


Figure 1. *SPDs of red, green, blue primaries for an Apple Cinema display (solid lines) and a Microvision laser projector (dashed lines).*

(a) Setup overview



(b) Setup during experiment

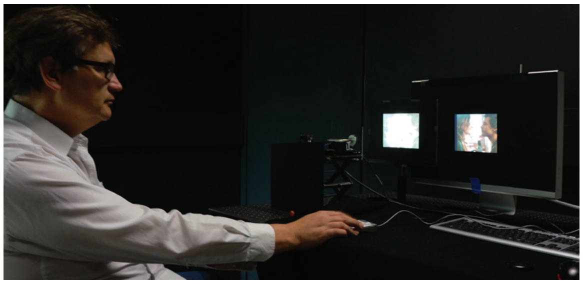


Figure 2. *Experiment setup.*

© 2014 Society for Imaging Science and Technology

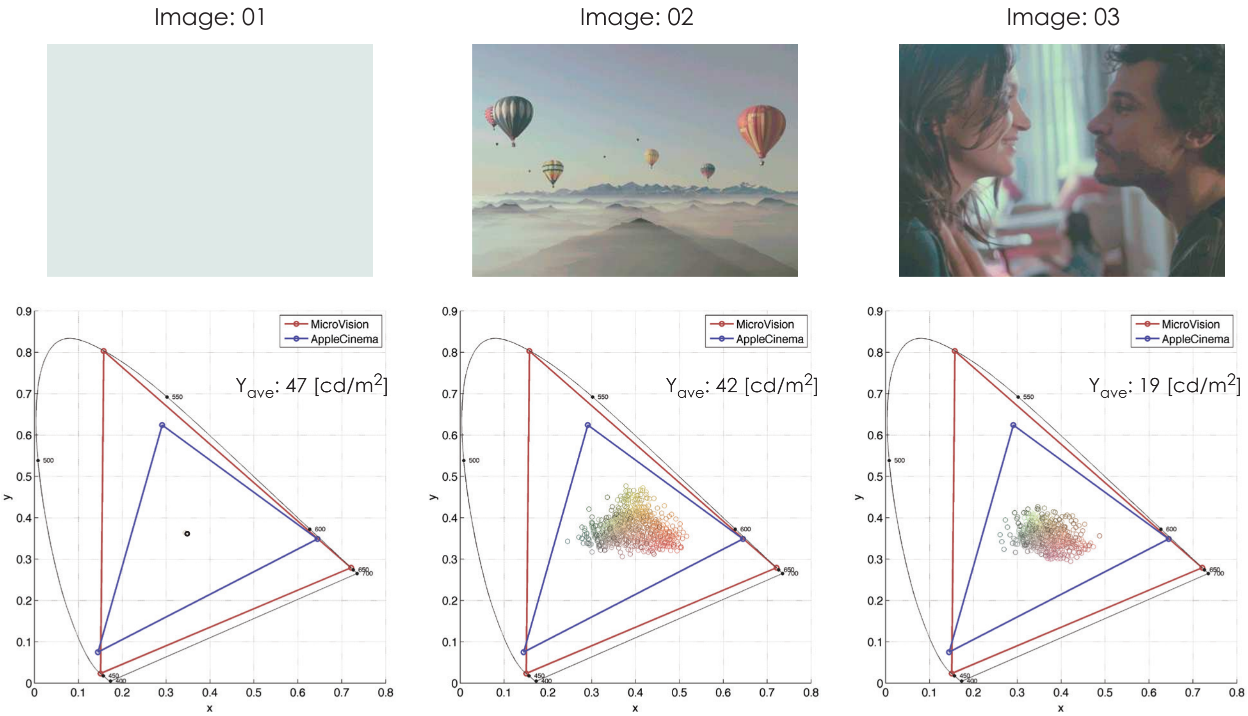


__Figure 3.__ Three test images used in the experiment. Luminances (averaged over an image) and Chromaticities of image contents are shown together with the gamut area of two media. Note that all of these computations were done using the CIE 1931 observer to comply with the conventions.

Test Images

Three test images were prepared for this study and shown in Figure 3. The aspect ratios of images were 4:3. Test image 01 is a blank (white) image as the simplest form of images. White (white point) correspondence is a key criterion in display color matching, and furthermore, human vision is sensitive to hue variations in near achromatic region. Test image 02 is an image with fewer components. There are large areas of sky, clouds and mountains. Test image 03 is an image where two people are facing each other. Image 03 was chosen since (1) faces are one of the most important features that people concern about its reproduction quality, and (2) humans possess dedicated brain circuitry to process faces and thus are able to perform more precise adjustment than other images [11]. All the images were adjusted so that the white point became D50. All the images were confirmed to be within the gamut of the two media for various CMFs after color matching. Namely, for each image, the average luminance and the saturations of all the pixels were adjusted such that (1) digital counts on the Microvision projector were within the gamut, and (2) digital counts on the Apple Cinema display after color matching by different target CMFs. The target CMFs we used for color matching simulation were the CIE 1931 observer, the CIE 1964 observer, Stiles and Burch 49 individual observers [12], and the CIE Physiological Observer 2006 (CIEPO06) [6] with the field size factor ranging from 2° to 10° and the age factor ranging from 20 to 60.

Procedure and Observers

The observer sat in a chair, approximately 60 cm away from the two media. The height of the observer's eyes was adjusted so that they were perpendicular to the media. The observer could move his/her head but was instructed to be perpendicular to the media when making decisions. The reference images were shown on the projector, and the observer adjusted images shown on the Apple Cinema display. The observer was instructed to adjust CIELAB L^*, a^*, and b^* values of the Apple Cinema display to match the overall appearance of the reference image. The adjustments were global. (e.g. Adding 1 value in L^* increased 1 L^* for all the pixels.) The computations from the user input (CIELAB) to RGB values were done using the above-described display models for the CIE 1964 observer. Since there was no adaptation stimulus, it had to be assumed to compute CIELAB. We calculated the reference white by averaging XYZ values of the image over all the pixels and multiplied it by five. This gives us approximately CIELAB L^* of 50 for the average. The initial matching image was the image matched by the CIE 1964 observer, randomized in the range of ± 10 a^* and b^* values. L^* was not randomized as we found it does not contribute to inter-observer variability (see Results and Discussion). The color adjustments were done through a keypad. There were six keys to increase or decrease L^*, a^*, b^*, and one key to randomize the starting values. This key was useful when the observer was lost in color space and wanted to initiate the matching process. After the observer finished adjustments, the experimenter saved the results and proceeded to the next trial. There were 3 repeated trials for each image for each observer. Before starting the experiment, the observer performed one trial using the test image 01 as a practice. Thus, there were 10 trials in total. The orders to present images were randomized. The experiment took about 20 minutes per an observer on average.

30 color-normal observers participated in the experiment. Their color vision was tested by Farnsworth D15 test. Ages

ranged from 21 to 53. The results from two observers were excluded from analyses due to relatively large intra-observer errors. Thus, the results from 28 observers were analyzed.

Analysis

Color image matching was simulated for different CMFs. A non-linear optimization was performed to optimize the three global adjustment values for a given observer function by minimizing the image difference. The procedure is briefly described as follows. First, the adjusted RGB image was reconstructed from the $L^*a^*b^*$ global adjustment values using the Apple Cinema display model for the CIE 1964 observer. Second, XYZ images were computed for the reference RGB image (shown on Microvision laser projector) and the adjusted RGB image (shown on Apple Cinema display) using the corresponding display models with a given set of CMFs. Third, the difference between the two XYZ images was taken assuming the reference white. We chose S-CIELAB [13] as an image difference metric. Ideally, the minimized image differences should be zero. However, since the global adjustments are CIELAB value for the CIE 1964 observer, the perfect match is less achieved as a set of CMFs deviates from the CIE 1964 observer. The simulation was performed for the same set of CMFs (CIE 1931, CIE 1964, CIEPO06, Stiles and Burch 49 observers) as used in Test Images section. The mean image differences for test image 01, 02, 03 were 0, 0.28, 0.50, respectively. Overall, the differences were small enough for valid analyses, and we could assume the global adjustments made by human observers would be satisfactory as well.

Results and Discussion

Table 1. Intra- and inter-observer variability expressed as MCDMs (ΔE_{00}, a^* and b^*) for each image

	Intra	Inter
Test Image 01	1.5	3.2
Test Image 02	1.8	3.8
Test Image 03	2.6	4.4

The preliminary experiment revealed that the intra-observer standard deviations in the lightness direction are much larger than (about twice as large) those in chromatic directions for any test image. In the preliminary experiment, four observers participated and there were at least five repeated matches for each test image. This is understandable since the luminance (or lightness) discrimination ability decreases with increasing separation between stimuli while the chromatic discrimination ability does not change [14, 15]. Oicherman et al. [5] also had large separations between media and the resultant inter- and intra-observer variability were larger in the L^* direction. In analyses, we would like to use the current standard color difference formula, CIEDE2000 (ΔE_{00}). However, the color difference space would be distorted by this large separation, and probably not reliable if used in a conventional way. To take the deteriorated lightness discrimination into account, the parametric factor for lightness in CIEDE2000 formula was set infinitely high to ignore the contribution from lightness differences. Thus, we analyzed the variability only in chromatic directions in the following discussion.

The inter- and intra-observer variability for test image 01 are shown on chromatic plane in Figure 4. As can be seen, each observer has an average match point that is significantly different from other observers. Similar variabilities were obtained for the other test images.

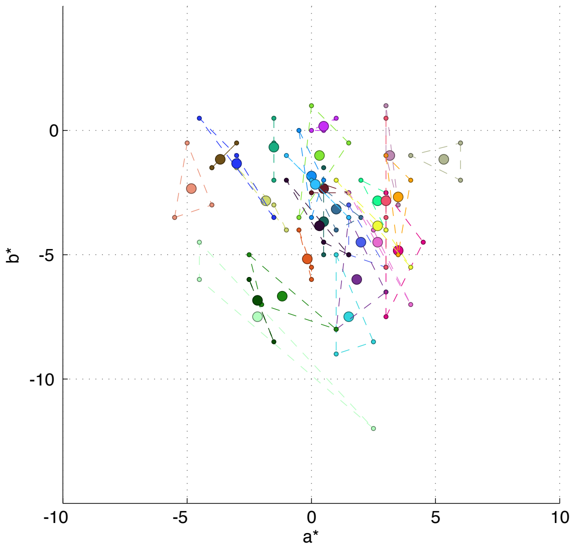


Figure 4. The matches made by 28 observers shown on a^*b^* plane for test image 01. Small circles with dashed lines represent three trials for each observer, and large circles represent average match point for each observer.

Mean color difference from the mean (MCDM) was used to express inter- and intra-observer variability quantitatively. MCDM is a metric often used to evaluate measurement precision [16]. In this study, increasing MCDM indicates increasing observer variability. As mentioned, ΔE_{00} without lightness contributions was used to compute color differences. For each test image, MCDM was calculated among observers (inter-) and each (intra-) observer. Inter-observer variability was calculated from observers' average match points. Intra-observer variability was calculated from computing a MCDM for three trials for each observer and then averaging the MCDMs over all the observers. The results are shown in Table 1. Inter-observer variability is always larger than intra-observer variability for all the images. It means one observer could match color images significantly differently from another observer. Overall, the ratio of inter- to intra-observer variability is almost two times. The more complex an image is, the more both inter- and intra- variability increase.

As a visual example, the test image 01 and 03 adjusted by some observers are shown in Figure 5 and 6, respectively. The matches made by these observers were at the edge of the population. Additionally, the CIE 1964 observer was also included. The adjustments were averaged over all the trials for each observer. It often happened that the color adjustments made by one observer could give an opposite impression to the adjustments made by another observer. For instance, the images adjusted by observer 20 appeared greenish and the images adjusted by observer 03 appeared reddish.

The matches made by 28 observers for all the test images are shown in Figure 7. Different test images produced similar variability but there are systematic shifts in the match points (averaged over all the observers for each image). The directions of

Figure 5. *sRGB rendered test image 01 adjusted by extreme observers and the CIE 1964 observer.*

Figure 6. *sRGB rendered test image 03 adjusted by extreme observers and the CIE 1964 observer.*

this systematic shifts are not exactly same but similar for most observers. Theoretically speaking, the match points would not change if an observer's color vision (or CMFs) stays constant. There would be two possible explanations for the match point shifts. The first explanation would be that the observer might use different effective visual angles for different images. In other words, the sizes of observers' regions of interest are different for different images (e.g. sky and mountains in test image 02 are larger than faces in test image 03). This would change the match points since CMFs change as visual angle (or field size) changes. The second explanation would be that the observer focused on specific contents (e.g. faces) rather than evaluating an entire image. However, given that the observer was instructed to match overall appearance between two images, the latter possibility is less likely.

To examine the first possibility, color image matching was simulated using CIEPO06 at a constant age with varying field size as illustrated in Figure 8. For a field size equals to 10° and an age at 30 year-old, the match of CIEPO06 is close to the CIE 1964 observer's match (a*, b* of [0, 0]). As the field size decreases, the match point shifts towards greenish and yellowish (from 10° to 6°), then towards reddish and yellowish (from 6° to 2°). Interestingly, this path caused by field size change is very similar to what it was obtained from the actual experiment (Figure 7). Directly comparing the results and the predictions (Figure 7 and 8), test image 01 requires a field size larger than 10° (in fact, the image size was 15° x 11°), test image 02 requires a field size smaller than 10°, and test image 03 requires a field size smaller than what required for test image 02. Obviously, the most complex image was test image 03, the simplest image was test image 01, and test image 02 was in-between. Thus, complex images seem to require smaller field sizes. It could be concluded that the match point shifts are well explained by effective field size (or visual angle) change.

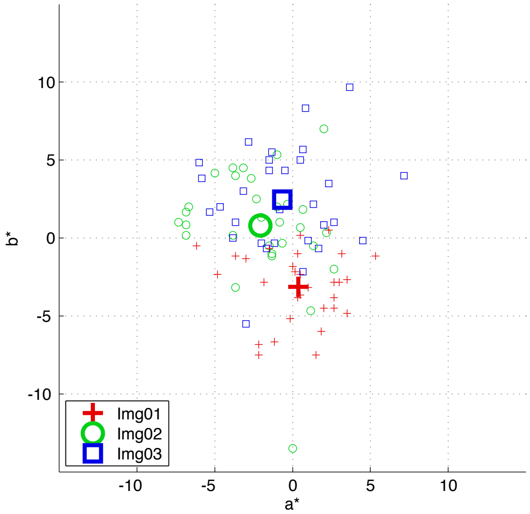


Figure 7. *Matches made by 28 observers for test image 01 (red crosses), 02 (green circles), and 03 (blue rectangles) represented by small markers. Large markers represent average match points over all the observers for a given image.*

Conclusion

Wide color gamut media are emerging, and observer metamerism is more and more concerned. We carried out the experiment on two different media involving color matching using real images to investigate how serious observer metamerism would be under a practical viewing condition. One media was a LCD monitor, and the other media was a laser projector. 30 observers participated in the experiment, and the data from 28 observers were analyzed. Under the viewing condition where images were widely separated (often the case in reality), the

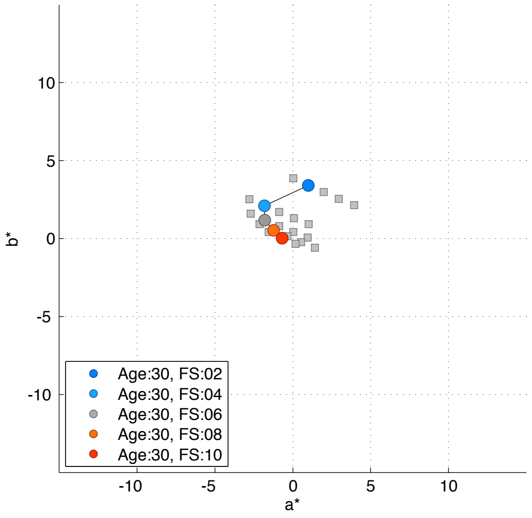


Figure 8. *Simulated color image matching results of CIEPO06 varying age (20-60 at every 10 year-old) and field size (2°-10° at every 2°) for test image 01. The highlighted ones are at a constant age (30 year-old) with varying field size. The simulation results for different test images yielded nearly same results.*

lightness discrimination was significantly decreased. It led to our decision to ignore the variability in the lightness direction when computing perceived color differences. The obtained inter-observer variability was larger than intra-observer variability for all the three images. The inter-observer variability was 3.2 to 4.4 in MCDM (ΔE_{00}, a^* b^*), which indicates observer metamerism would be significantly apparent when a laser projector is viewed together with conventional media. Image impression could be perceived in a opposite way (a person may see an image as 'warm' whereas another person may see it as 'cool'). It was found that effective visual angle (or field size) is dependent on image contents. The more complex an image is, the smaller the effective visual angle is. Solutions should be provided for observer metamerism in wide color gamut media, such as color management personalization (= Observer Dependent Color Imaging [17]).

Acknowledgments

The research was funded by Technicolor.

References

[1] A Sarkar, L Blondé, P Le Callet, F Autrusseau, P Morvan, and J Stauder. A color matching experiment using two displays: design considerations and pilot test results. In *Conference on Colour in Graphics, Imaging, and Vision*, volume 2010, pages 414–422. Society for Imaging Science and Technology, 2010.

[2] R L Alfvin and M D Fairchild. Observer variability in metameric color matches using color reproduction media. *Color Research & Application*, 22(3):174–188, 1997.

[3] D C Rich and J Jalijali. Effects of observer metamerism in the determination of human color-matching functions. *Color Research & Application*, 20(1):29–35, 1995.

[4] I Pobboravsky. Effect of small color differences in color vision on the matching of soft and hard proofs. In *TAGA Proceedings*, pages 62–79, 1988.

[5] B Oicherman, M R Luo, B Rigg, and A R Robertson. Effect of observer metamerism on colour matching of display and surface colours. *Color Research & Application*, 33(5):346–359, 2008.

[6] CIE. Fundamental chromaticity diagram with physiological axes - part 1. *CIE Publication No.170*, 2006.

[7] M D Fairchild and D R Wyble. Mean observer metamerism and the selection of display primaries. In *Final Program and Proceedings-IS&T/SID Color Imaging Conference*, pages 151–156, 2007.

[8] R Ramanath. Minimizing observer metamerism in display systems. *Color Research & Application*, 34(5):391–398, 2009.

[9] M Freeman, M Champion, and S Madhavan. Scanned laser pico-projectors: seeing the big picture (with a small device). *Optics and Photonics News*, 20(5):28–34, 2009.

[10] E A Day, L Taplin, and R S Berns. Colorimetric characterization of a computer-controlled liquid crystal display. *Color Research & Application*, 29(5):365–373, 2004.

[11] G Kindlmann, E Reinhard, and S Creem. Face-based luminance matching for perceptual colormap generation. In *Visualization, 2002. VIS 2002. IEEE*, pages 299–306. IEEE, 2002.

[12] Colour & Vision Research Laboratory, Stiles & Burch individual 10-deg colour matching data. Available at: `www.cvrl.org`, Accessed date: 2014.

[13] X Zhang and B A Wandell. A spatial extension of cielab for digital color-image reproduction. *Journal of the Society for Information Display*, 5(1):61–63, 1997.

[14] L T Sharpe and G Wyszecki. Proximity factor in color-difference evaluations. *Journal of the Optical Society of America*, 66(1):40–49, 1976.

[15] M V Danilova and J D Mollon. The comparison of spatially separated colours. *Vision research*, 46(6):823–836, 2006.

[16] R S Berns. *Billmeyer and Saltzman's Principles of Color Technology*. John Wiley & Sons, New York, 2000.

[17] A Sarkar. *Identification and Assignment of Colorimetric Observer Categories and Their Applications in Color and Vision Sciences*. PhD thesis, Université de Nantes, 2011.

Author Biography

Yuta Asano is a Ph.D. student in Color Science program at Rochester Institute of Technology in Rochester, NY. He obtained his B.Eng. in Architecture (2009) and M.Eng. in Architectural Engineering (2011), both from Kyoto University in Kyoto, Japan.

Mark Fairchild is Associate Dean of Research and Graduate Education in College of Science, and Director of the Program of Color Science and Munsell Color Science Laboratory at Rochester Institute of Technology.

Laurent Blondé is a Principal Scientist at Technicolor, Rennes, France. In the Research & Innovation group he supports image science projects since 1988 on various technical and scientific topics including optics, CGI, digital video processing, TV broadcast, digital cinema, color management, and visual perception.

Patrick Morvan is a researcher at Technicolor, Rennes, France. His main research area covers image processing, color science, and color management.