What is constant in color constancy?

Jordi Roca-Vila, C.A. Párraga, Maria Vanrell; Computer Vision Center, and Computer Science Department, Universitat Autònoma de Barcelona; Barcelona, Catalonia, Spain

Abstract

Color constancy refers to the ability of the human visual system to stabilize the color appearance of surfaces under an illuminant change. In this work we studied how the interrelations among nine colors are perceived under illuminant changes, particularly whether they remain stable across 10 different conditions (5 illuminants and 2 backgrounds). To do so we have used a paradigm that measures several colors under an immersive state of adaptation. From our measures we defined a perceptual structure descriptor that is up to 87% stable over all conditions, suggesting that color category features could be used to predict color constancy. This is in agreement with previous results on the stability of border categories [1,2] and with computational color constancy algorithms [3] for estimating the scene illuminant.

Introduction

Color constancy (i.e. the stability of object color perception under illumination changes) is a phenomenon that arguably involves mechanisms spanning three levels in the brain: sensorial, perceptual and cognitive [4,5]. Its effects have been traditionally measured using *achromatic setting*, *color naming* or *asymmetric matching* psychophysical paradigms all adapted to tap into one of more of these mechanisms and to correspond to particular scene features, including 3D perception [6], movement [7], etc. However, few studies have measured more than one point under illuminant adaptation [8-11] and fewer have measured enough points to address the question of whether the subject's categorical perceptual *structure* (i.e. the interrelations among perceived colors) is kept constant under illuminant changes. The exceptions to this are color naming paradigms, where subjects are asked to name several colors under different adaptation conditions. In this case, one of the main limitations is the restriction of choices presented to the observer (for example, in two recent experiments [1,2], one restricted its measurements to an equiluminant plane with 417 testing samples and the other to a set of 469 tridimensional Munsell samples), which constraints the method's precision. In this work we overcome this problem by using a psychophysical paradigm [12], which instead of measuring category borders as before [1,2], allows subjects to select their own *memorable* colors (by manipulating a set of patches embedded on the CRT monitor's screen) from a set close enough to the focal colors so as to make them easy to memorize and reproduce. Focal colors are in this context the 11 universal basic color categories (white, grey, black, blue, yellow, red, green, purple, pink, orange and brown) defined by Berlin and Kay in their seminal study [13], which have been shown to be easier to memorize [14]. The memorable colors were selected from a choice of 52000 different (JND spaced) CIELab samples which look continuous to the subjects. This paradigm allows us to investigate changes in the perceptual "structure" of colors after adaptation to illumination changes by simultaneously measuring several points in the color space. The method used can be understood as an extension of achromatic setting to other colors, in the sense that observers have to set a test stimulus so that it appears of a given color, instead of "achromatic". Regarding the method's precision, previous unpublished measures showed that trained subjects are not necessarily worse at setting memorable colors than they are at setting achromatic patches.

Here we report a new experiment that measures the simultaneous perception of several colors under five different illuminants and two different backgrounds. We analyze the interrelations among these perceived colors after the subject was thoroughly adapted. The psychophysical paradigm used allows us to make use of more sophisticated techniques for describing the color constancy phenomenon.

The method's section describes the psychophysical experiment in detail and explains the roles played by the illuminant, backgrounds and other stimulus conditions. Following this, we model the interrelations in each set of measurements as a *graph*, which allowed us to derive conclusions about the stability of the categorical perception of the subject and its implications for computational approaches to color constancy.

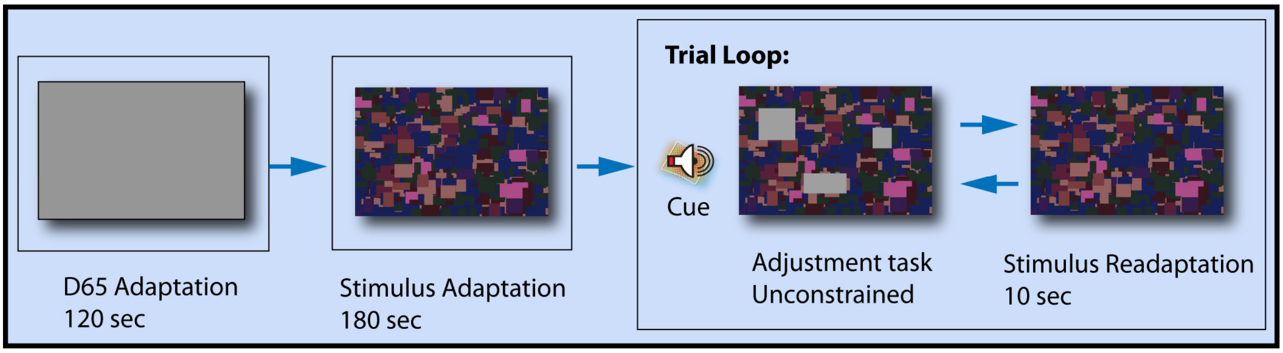

Figure 1. *Temporal sequence of the psychophysical method for a session. Each session consisted into 44 trial loops where the subject task consisted into reproduce the selected representatives stated in the reference session, i.e., one for each basic color category (red, green, blue, yellow, brown, orange, pink and purple).*

Methods

Overview

Subjects were presented on the screen with the written name of a focal color and asked to match it to their own internal representations by manipulating the color of a patch by means of a gamepad. After that, they were required to reproduce the very same colors on different days under different conditions of background and illumination.

Experimental setup

All adjustments/measures were done in CIELab space with D65 (Lum = 100 Cd/m^2) as a white point. The whole experiment was divided into 10 sessions and each session in 44 trials. Each trial lasted approximately 30 seconds and each session approximately 25 minutes. In order to avoid subject fatigue, no more than two sessions per day were allowed. All experiments were conducted inside a dark room (all the walls and the ceiling were lined black). The stimuli were displayed on a CRT Mitsubishi Diamond Pro 2045SU monitor (which was the only light source in the room) driven by ViSaGe graphics card from Cambridge Research Systems (CRS-www.crsldt.com) with 12 bits color resolution per channel. The screen (389x292mm and 1024x768 pixels, 100Hz) subtended approximately 22x17 deg. Viewing was binocular and the head was unrestrained. The monitor was calibrated regularly using a Minolta colorimeter. All experiments were run on Matlab (www.mathworks.com) using the COLORLAB toolbox [15] to get the color space conversions needed. Subjects adjusted the required colors by means of a gamepad. All adjustments were done in CIELab space using six different buttons, two for each color space dimension. At the end of each trial, the resulting CIELab coordinates were recorded by pressing a gamepad button.

Stimuli

The spatial structure of our stimuli (a "Mondrian" pattern) consisted of a set of overlaid colored rectangles randomly distributed across the image (i.e. flat, without highlights or mutual reflections) similar to others [16,17]. The rectangle size frequency distribution was similar for all stimuli (mean square size was 50x50 pixels) and its geometrical distribution was uniform across the digital image. We defined three types of Mondrian stimulus:

- *Type 0* built by randomly replacing all colors in the Mondrian by 7 intensity levels (from 40 to 70 L CIELab units in steps of 5) of the same (D65) chromaticity.
- *Type I* was constructed from the colors selected to be the best representative available of each category (see below).
- *Type II* was constructed by selecting hues in-between *selected representatives* but similarly saturated.

Type I and *Type II* were different for each subject and did not contain grey patches to avoid cueing the observer on the illuminant [5]. All Mondrians were in turn "illuminated" by performing the spectral product of each patch's reflectance times one of five simulated illuminations assuming a Lambertian reflectance model [18]. The illuminants were chosen so that the final product (the illuminated Mondrian chromaticities) was as saturated as possible while still inside the CRT monitor's gamut. There was no "central patch" to look at, but a set of randomly distributed patches that were simultaneously adjusted in color by manipulating the gamepad. These constituted up to 10% of the all patches and their positions were randomly selected in each

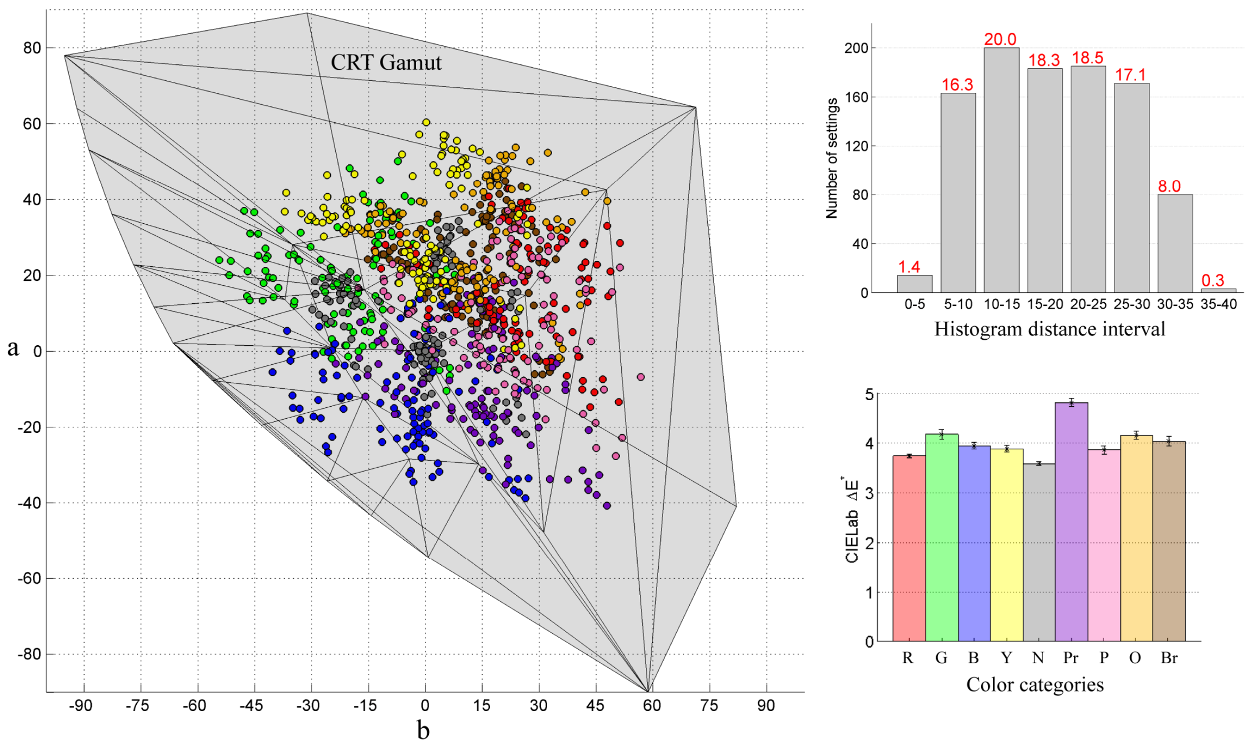

Figure 2. *Left: CIELab chromaticity plane where the CRT convex hull gamut is projected and 999 colored dots where each one corresponds to one different color measurement under different conditions. Each dot was computed as the average of 5 settings and its color indicates the perceived color by the subject when adjusted. Right top: Histogram where the number of measured colors is divided according to their distance to the convex hull CRT gamut boundary. Right bottom: Average chromatic setting precision over all conditions of experiment 2, depicted by color categories.*

trial. The object of this was to force the subject to average among patches that had different local surroundings, thus avoiding local chromatic induction effects. Each Mondrian was unique.

Table1: CIExy chromaticity for the illuminants.

Illuminant	*x*	*y*
D65	0.312	0.329
Purple	0.316	0.228
Green	0.296	0.453
Yellow	0.453	0.434
Orange	0.437	0.343

Experimental procedure

As illustrated in Figure 1, all sessions started by watching an achromatic stimulus (uniform D65 chromaticity and 30 Cd/m^2 luminance) for 120 seconds. Subsequently, subjects were adapted to a random illuminated Mondrian for a further 180 seconds. After this, they were prompted (auditorily and visually, by a black word written at the bottom of the screen) to the color category they had to produce. Their instructions were to operate the gamepad to select a categorically representative color (see particular instructions below) and then to continue to the next trial. This (prompting plus color selection) was repeated 44 times (5 times for each of the eight colors plus 4 times for the grey). The call order of each color was randomized for each session and subject and there were no time constraints on the trials.

In the *reference* session, subjects had to select the most representative samples for each of eight color categories (red, pink, purple, blue, green, yellow, orange and brown). The aim here was for subjects to select "best examples" of colors that could be easily reproduced throughout the rest of the experiment. The palette of possible colors in the *reference* session was limited in saturation and lightness by a cylinder whose axis was the "L" dimension of CIELab (radius 22 Lab-units, and lightness between 30 and 70 L-units). The purpose of this cylinder was strictly technical: we wanted subjects to find reasonably representative samples while still allowing these to be "illuminated" later without exceeding the CRT-monitor gamut limitations. To stop subjects from memorizing key presses all starting colors were randomized. The selection of best representative colors was done for every subject under the same illuminant (*D65*) and same *type 0* background. We termed this adjusted colors as *selected representatives (SR)*, and they were different for each subject.

The *regular* sessions were similar to *reference* sessions in every respect except that instead of selecting their best category representative, subjects were asked to produce the same colors they had selected in the *reference* sessions. The backgrounds used were Type I or II and they were illuminated with chromatic light without any saturation constraints (no cylinder).

Observers

Four subjects took part in our experiment. They were between 31 and 44 years old and their color vision was normal (or corrected to normal) as tested by the Ishihara colored plates and the Farnsworth-Munsell D15 Hue Test. Of these, two were volunteers and naïve to the experiment's purpose, and the other two were authors (JRV and CAP).

Results

Overview

The data presented in this section comes from two different experiments. The first one was done six months ago and used 10 subjects, three illuminations (*D65*, *yellowish* and *greenish*) and three backgrounds (*type* 0, I and II). The second one used 4

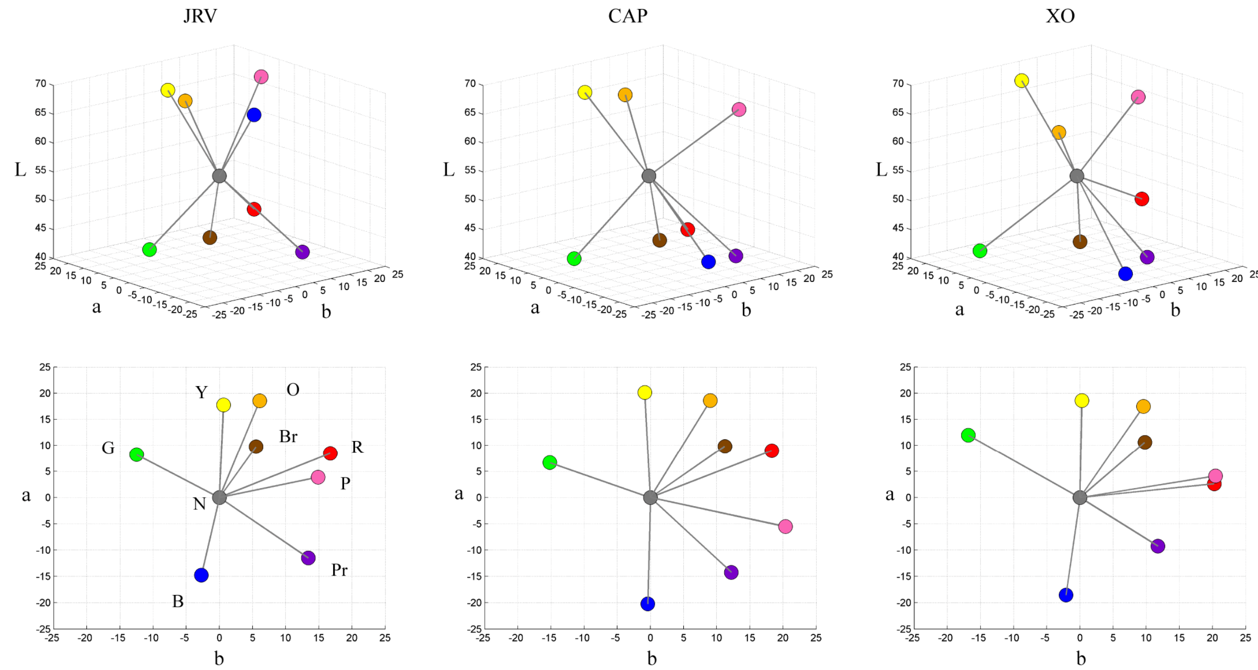

Figure 3. *Selected representatives for three subjects. Each column corresponds to one subject and it contains two views of the selected representatives in the reference session. Results are plotted in CIELab color space: the first row contains the isometric view and the second row the projection in the ab chromaticity plane. Key: G=green, B=blue, Pr=purple, P=pink, R=red, Br= brown, O=orange, Y=yellow, N= neutral (grey).*

subjects, two backgrounds (*type* I and II) and two different illuminations (*purplish* and *orange*). Our results reveal the invariance of perceived color interrelations under different illuminations.

Priors

Our own pilot studies revealed that subjects are able to remember the *selected representatives* over the experimental period of several weeks. Figure 2 shows all gathered data alongside the CRT gamut (all subjects, illuminants and

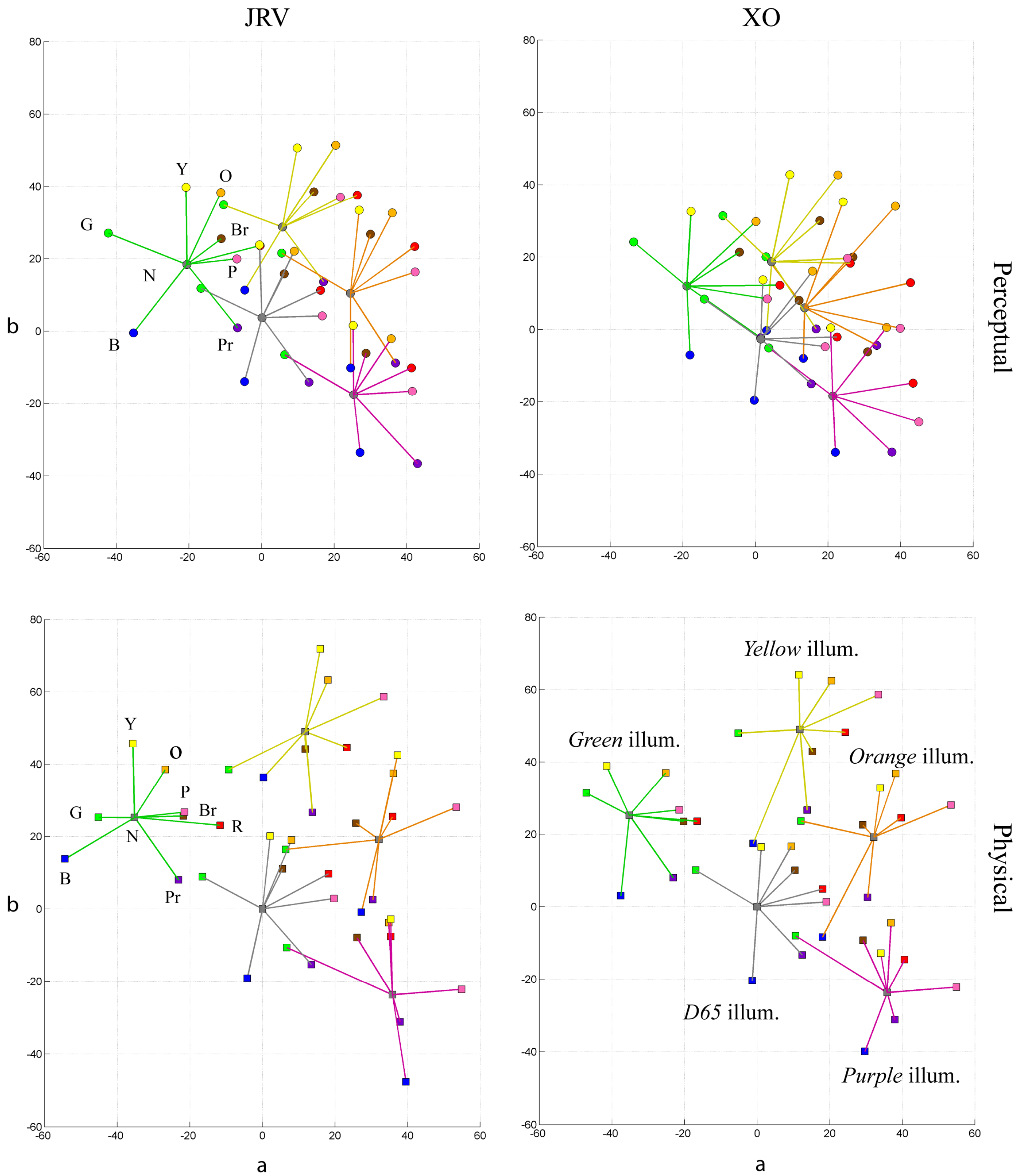

Figure 4. Selected representatives obtained for two subjects and five illuminants. Each plot shows the measures obtained for 5 illuminants (each line color corresponds to one illuminant). The top row shows the colors (SR) selected by two observers and the bottom row shows the corresponding projections of the original colors after illumination. Left plots correspond to subject JRV with type I background and right plots to subject XO with the type II background. All measures are shown as projections on the CIELab ab-chromaticity plane.

backgrounds: 999 color measurements obtained from adjusting 4880 settings over 111 regular sessions). Each point in Figure 2 is the average of 5 different trials. The bottom illustration in Figure 2 shows the experimental error computed as the standard deviation from the mean for each color category, averaged over subjects and sessions. The overall standard deviation was 4ΔE* units. In order to ensure that subjects did not use the CRT gamut as a reference when doing the adjustments requested, we computed for each of the 999 color samples the distance to the CRT gamut boundary and averaged this information at the top of each histogram bar in Figure 2. Our results show that only 1.4% of the 999 points were closer than 5ΔE* units from the CRT gamut boundary (only 16.3% were between 5 and 10ΔE* units). Figure 3 illustrates the *selected representatives* chosen by three different subjects on the *reference session*. Each colored circle shows the SR for the corresponding category and the joining lines help to illustrate their geometrical interrelations (Euclidean distances). Notice how this interrelations are different for each subject, for instance, subject JRV SR blue has high luminance while the other two subjects selected colors with low luminance values. Also, when comparing subject CAP and XO, notice how the selection of orange and red differs also in luminance level: subject CAP selected higher luminance than XO. Finally, notice how the red and pink SRs of CAP were different in hue and luminance from those of XO. Following this, we conclude that each subject had his/her particular choice of *selected representatives*, which is expected from previous studies [13,19,20], but at the same time the pattern conformed by those remained approximately invariant. In the following sections we studied whether this pattern was also invariant under different illuminant adaptations (over different stimuli, background and illuminant conditions) occurred.

Structural deformation

The psychophysical paradigm provided us with a set of 9 measures for each adaptation state. The first row of Figure 4 illustrates these measures for two subjects over the five illuminants tested. When considering a particular subject, notice how the same groups of measures (linked by colored lines) seem to keep their structures stable over illuminants changes. To formally describe the overall interrelations among sets of measures we modeled the CIELab coordinates from each set of measured data. For example, in Figure 4, each set of 9 linked circles is modeled as a *graph*, where each node represented a measured SR and edges were defined for all possible node combinations. The Euclidean distances (in CIELab ΔE*) between nodes act as weights to each of the edges and when all edges were considered, a graph distance matrix was defined. In order to extract inter-subject comparisons we normalized this distance matrix by the mean distance from all nodes to the node corresponding to grey. This allowed us to produce a distance matrix which enclosed the distance proportions to each node in terms of the distance to the central node. In summary, each session produced 9 SRs which were modeled as a graph, whose nodes were the corresponding measured CIELab coordinates and its distance matrix was build from the Euclidean distances among these CIELab coordinates. In this way the perceptual categorical structure of our subjects was captured by a graph allowing us to get a reliable comparison among adaptation conditions, i.e., we defined the distance between two graphs by the mean absolute difference between the corresponding graph distance matrices.

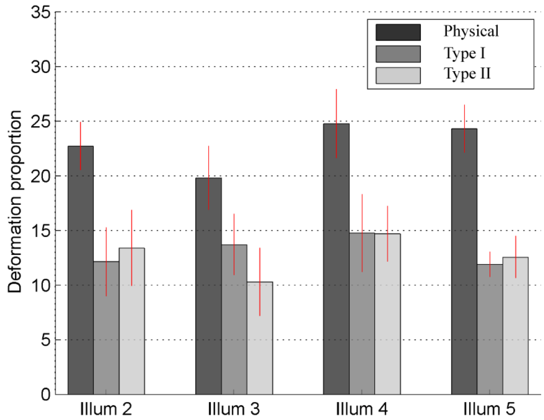

Figure 5. *Each group bar indicate the average distance, with D65 and the indicated illuminant, of the graphs used to assess the interrelations of measured data under each condition. Data is averaged over subjects and compared according to backgrounds and perceptual/physical.*

The top plots of Figure 4 show the graphs corresponding to the colors *perceived* by two subjects after a change of illumination. The bottom plots represent the *calculated* CIELab coordinates of the same stimuli after illumination (the *physical* colors). Looking at Figure 4 and comparing the top and bottom plots we can identify two trends: first, the perceptual representations of the top plots seem to have maintained the same proportions showing higher stability in terms of their interrelations, while their counterparts at the bottom plots have been slightly warped by the illumination. Second, the region spanned by the perceptual measures is more compact than the region spanned by the physical ones. The latter observation is wholly captured by standard color indexes (which measure distances between perceptual and physical grey) [5] while the former observation can only be captured by the graph representation approach. Figure 5 contains four sets of bars, each corresponding to one comparison between conditions under illuminant D65 and other illuminants indicated by the x axis label. They were produced averaging the results for all subjects. For instance, looking back at the top left plot of Figure 4, each of its 5 sets of measured data were modeled as graphs, and graph distances were computed from the graph centered at the achromatic locus (D65 illumination). When grouping over subjects and averaging graph distances we obtain the bar height and STD shown in Figure 5. Notice how the values corresponding to the physical graph distances are significantly higher than the perceptual ones (approx. 23% for the physical distances versus 13% for the perceptual distances in average).

The experimental error, which was mostly the product of the subjects' variability at producing SRs, was estimated to be about 4ΔE*. We propagated this error to our structural deformation index and obtained values of about 5.2%. These results indicate that subjects tend to maintain the same perceptual interrelations between colors (graph structures) despite variations in the structure of physical colors under an illuminant change.

Comparison to previous work

Previous work has also focused on the color appearance of multiple points under illumination changes. Some used real surfaces and a matching technique [8-11] or CRT-simulated

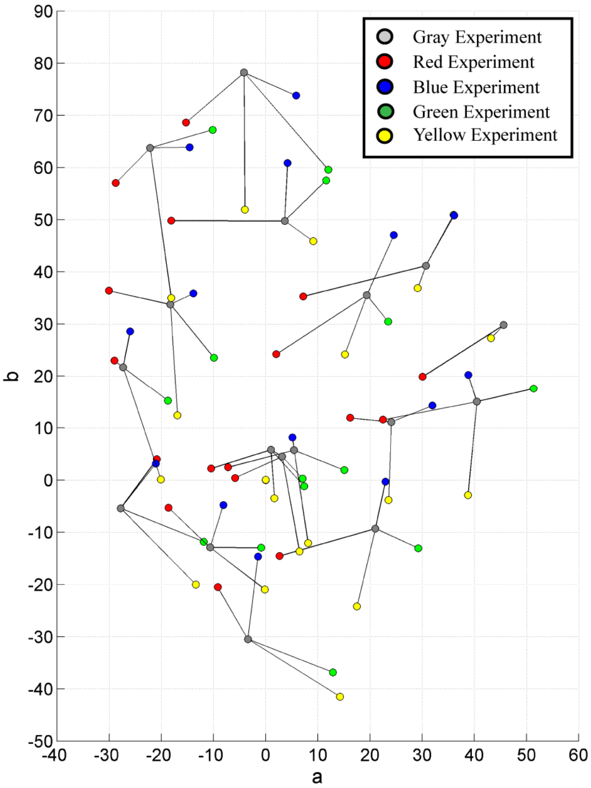

Figure 6. *CIELab ab plane projection of MacCann et al reported results. Each colored circle corresponds to the CIELab coordinates (D65 as a reference white point) of surface matches under a five different illuminants, which are linked by black lines.*

scenes [10,11]. Others, [1,2] measured the color appearance of multiple points indirectly by means of the deformation of categorical boundaries in color space, concluding that the categorical structure of color space remains roughly stable, which is in agreement with our results.

In order to further compare our results to others in the literature, we have considered pioneering work by McCann *et al* [8]. They reported the Munsell coordinates of 17 matches between 5 different illuminants. Figure 6 shows CIELab color space plots of MacCann *et al* results. Each color indicates the illuminant under which the matching was done. Notice how each grey dot is linked by means of a black line to its corresponding illumination matching. We applied our structural approach, comparing their 'grey' set of points to the other four colored sets and obtained a structural stability of 85% (2%STD) which is similar to our results and suggests a high degree of color constancy (but not perfect color constancy).

Implications for computational color constancy

Our results show the structural invariability of the interrelations between perceived colors under a change of illuminant. These have been exploited in the computational color constancy model of Vazquez *et al* [3], which estimates scene illumination by selecting the illuminant from a particular set of candidates according to their ability to map the transformed image onto a set of specific colors, such as those proposed here.

Conclusion

We have collected information on the perceptual interrelations of colored surfaces under illuminant changes and modeled our measurements using graphs. Our results show that these interrelations remained 87% constant under an illumination shift, in contrast with the structural deformation undergone by the physical colors. This suggests that categorical perception may be used to guide color constancy adaptation. This result is in accordance with previous studies, [1,2], that reported categorical stability using color naming techniques. Based on our previous results, we could state that categorical color perception maintains a high degree of structural invariance under illuminant changes.

Acknowledgments

This work has been partially funded by project TIN2010-21771-C02-01 and project Consolider-Ingenio 2010-CSD2007-00018 of the Spanish Ministry of Science. C. Alejandro Parraga was funded by the “Ramon y Cajal” program of the Spanish MEC (Ministry of Science)-Ref. RYC-2007-00484.

References

[1] T. Hansen, S. Walter, & K. R. Gegenfurtner, "Effects of spatial and temporal context on color categories and color constancy", J. Vis., 7, (2007).
[2] M. Olkkonen, T. Hansen, & K. R. Gegenfurtner, "Categorical color constancy for simulated surfaces", J. Vis., 9, (2009).
[3] J. Vazquez-Corral, M. Vanrell, R. Baldrich, & F. Tous, "Color constancy by category correlation", IEEE Trans Image Process, (2012).
[4] A. Hurlbert, & K. Wolf, "Color contrast: A contributory mechanism to color constancy", Prog Brain Res, 144, (2004).
[5] D. H. Foster, "Color constancy", Vis. Res., Volume 51: 674, (2011).
[6] M. Hedrich, M. Bloj, & A. I. Ruppertsberg, "Color constancy improves for real 3d objects", J. Vis., 9, (2009).
[7] A. Werner, "Color constancy improves, when an object moves: High-level motion influences color perception", J. Vis., 7, (2007).
[8] J. J. Mccann, S. P. Mckee, & T. H. Taylor, "Quantitative studies in retinex theory - comparison between theoretical predictions and observer responses to color mondrian experiments", Vis. Res., 16, (1976).
[9] J. J. Kulikowski, & H. Vaitkevicius, "Colour constancy as a function of hue", Acta Psychol (Amst), 97, (1997).
[10] K. Amano, & D. H. Foster, *Tracking categorical surface colour across illuminant changes in natural scenes*. CGIV 2008 / MCS/08 - 4th European Conference on Colour in Graphics, Imaging 289-292 (2008).
[11] D. H. Brainard, W. A. Brunt, & J. M. Speigle, "Color constancy in the nearly natural image. I. Asymmetric matches", J. Opt. Soc. Am. A, 14, (1997).
[12] J. Roca-Vila, C. A. Parraga, & M. Vanrell, "The fine structure of color constancy", Submitted, (2012).
[13] B. Berlin, & P. Kay, *Basic color terms: Their universality and evolution* (University of California Press, 1991, Berkeley ; Oxford, 1969)
[14] E. R. Heider, "Universals in color naming and memory", J Exp Psychol, 93, (1972).
[15] J. Malo, & M. J. Luque (2002). Colorlab: A color processing toolbox for matlab. **http://www.uv.es/vista/vistavalencia/software.html**
[16] L. Arend, & A. Reeves, "Simultaneous color constancy", J. Opt. Soc. Am. A-Opt. Image Sci. Vis., 3, (1986).
[17] L. E. Arend, "How much does illuminant color affect unattributed colors?", J. Opt. Soc. Am. A, 10: 2134, (1993).

[18] V. Walsh, & J. J. Kulikowski, *Perceptual constancy : Why things look as they do* (Cambridge University Press, Cambridge, UK ; New York, NY, USA, 1998)
[19] R. M. Boynton, & C. X. Olson, "Locating basic colors in the osa space", Color Research and Application, 12: 94, (1987).
[20] J. Sturges, & T. W. A. Whitfield, "Locating basic colors in the munsell space", Color Research and Application, 20: 364, (1995).

Author Biography

Jordi Roca-Vila has a first degree in mathematics, a MSc in Mathematical Engineering (specialized in Computer Vision) from the Univeritat Politècnica de Catalunya in 2009 and is currently finishing his Computer Science PhD in the Computer Vision Center of the Universitat Autònoma de Barcelona, where he studies color constancy using both psychophysical and computational approaches. In addition he worked for the industry, developing quality control processes based in the 3D scene reconstruction from range images and stereovision.