Modeling human color categorization: Color discrimination and color memory

E.L. van den Broek [a] M.A. Hendriks [a] M.J.H. Puts [b]
L.G. Vuurpijl [a]

[a] Nijmegen Institute for Cognition and Information,
P.O.Box 9104 6500HE Nijmegen
[b] Visual Science Laboratories, University of Chicago,
940 East 57th Street, Chicago, IL 60637

Abstract

Color matching in Content-Based Image Retrieval is done using a color space and measuring distances between colors. Such an approach yields non-intuitive results for the user. We introduce color categories (or focal colors), determine that they are valid, and use them in two experiments. The experiments conducted prove the difference between color categorization by the cognitive processes color discrimination and color memory. In addition, they yield a Color Look-Up Table, which can improve color matching, that can be seen as a model for human color matching.

1 Introduction

The origin of the color *lilac* lays in the Sanskrit *nilla 'dark blue'*, of which the Persian made *nIlak 'bluish'*, from *nIl 'blue'*. In the Arabic the meaning evolved to a description of a plant with flowers of this color: the *Sering*. In 1560 the *Sering* was brought to Vienna, by an Austrian ambassador. From there the plant reached France. There the word's meaning evolved to *"a variable color averaging a moderate purple"*[1].

So, there is more with colors than one would think at a first glance. The influence of color in our everyday life and the ease humans use color are in strong contrast with the complexity of the phenomenon color, topic of research in numerous fields of science (e.g., physics, biology, psychology, computer vision, etc.).

In this paper, we focus on the use of colors in the field of *Content-Based Image Retrieval (CBIR)* [6, 7]. On the one hand, one has to take into account the RGB-color space used by the computer, the environmental conditions, etc. On the other hand, human color perception is of utmost importance. Since (human) users judge the retrieval results, the CBIR's matching algorithms need to provide a match that the user would accept. The complexity of this constraint is illustrated by the amount of available color spaces, such as: RGB, HSV, CIE[2] XYZ, and

[1] *Onze Taal Taalkalender 2003* (http://www.onzetaal.nl/) and Merriam Websters Online Dictionary (http://www.m-w.com/)
[2] http://www.cie.co.at/ciecb/

Munsell[3] [3]. However, none of these color spaces model human color perception adequately.

In our opinion, one should consider color in CBIR from another perspective, that of the *focal colors* or *color categories*: Black, white, red, green, yellow, blue, brown, purple, pink, orange, and gray [2, 4, 5, 8]. People use these categories when thinking, speaking, and remembering colors. Research from diverse fields of science emphasize the importance of them in human color perception. The use of this knowledge can possibly provide a solution for the problems of color matching in CBIR.

Most CBIR-engines distinguish two forms of querying, in which the user uses either an example image (*query-by-example*) or defines features by heart, such as: shape, color, texture, and spatial characteristics (query-by-content). In the latter case, we are especially interested in *query-by-color*. At the foundation of each of these queries lies a cognitive processes, respectively color discrimination and color memory. Let us illustrate the importance of the distinction between *query-by-example* and *query-by-color* by a simple example. Imagine you want to find images of brown horses.

In the case of *query-by-example*, the resulting images will be matched on the example image: a process of color discrimination is triggered. In this process the colors are (directly) compared to each other.

In the case of *query-by-color* we need to try to imagine the color brown. Probably, you will not have a clear color in mind, but a fuzzy idea or a fuzzy set of colors: a *color category*, based on your *color memory*. Each of the elements of this brown set (or category) are acceptable colors. There is no need for several types of brown. Providing the keyword "brown" or pressing a button resembling the fuzzy set brown is sufficient.

In both forms of querying the CBIR-system can use a *Color Look-Up Table (CLUT)* for the determination of the elements of this set, described by R, G, and B-values. The set is fuzzy due to the several influences on the color (of the object of interest), such as the color of the surrounding and the semantic context in which the object is present.

However, it is clear that a distinction should be made between color categorization by discrimination and color categorization by memory. An important distinction because humans are capable of discriminating millions of colors but when asked to categorize them by memory, they use a small set colors: *focal colors* or *color categories* [2, 4, 5, 8]. Despite the fact that the importance of such a distinction is evident, this differentiation is not made in CBIR-systems.

In the remainder of this paper a question posed and two experiments executed, will be discussed. The question posed to the subjects is: "Please write down the first 10 colors that come to mind.". With the experiments we prove the difference between color categorization by color discrimination and by color memory. Hence, this research will prove that:

- The use of *color categories* is valid in a CBIR context,
- The RGB-color space can be described using *color categories*,

[3] http://www.munsell.com

- There is a difference in color categorization using color discrimination or color memory.

Moreover, we will present markers, by which the color space is divided, on which a *CLUT* for CBIR can be employed. With that a new model of human color categorization is introduced.

2 Method

2.1 Subjects

Twenty-six subjects with normal or corrected-to-normal vision and no color deficiencies, participated. They participated either voluntary or within the scope of a course. The first group were employees and the latter were students of the University of Nijmegen. They were naive as to the exact purpose of the experiment.

2.2 Equipment

An attempt was made to create an *average office environment*. Stimuli were presented on a 17" CRT monitor (ELO Touchsystems Inc., model: ET1725C), with a resolution of 1024 x 768 pixels at a refresh-rate of 75Hz. The experiment was conducted in an room with average office lighting: a Cool White Fluorescent light source: TL84 was present, its color temperature: 4100K (Narrow Band Fluorescent), as used primarily in European and Asian office lighting.

The experiments ran on a PC with an Intel Pentium II 450 MHz processor, 128mb RAM, a Matrox Millennium G200 AGP card, and with a Logitech 3-button Mouseman (model: M-S43) as pointing-device. The experiments were conducted in a browser-environment with Internet Explorer 6.0 as browser and Windows 98 Second edition as operating system, using 16-bit colors.

2.3 Stimuli

The stimuli were the full set of *the 216 web-safe colors*[4]. These are defined as follows: The R, G, and B dimensions (coordinates) are treated equally. Their minimum value is 0, the maximum value of each of the dimensions is 255. For each dimension 6 values are chosen on equal distance, starting with 0. So, for the RGB-values 0 (0%), 51 (20%), 102 (40%), 153 (60%), 204 (80%), and 255 (100%) are chosen. Each of these 6 values is combined with each of the 6 values of the 2 other dimensions. This results in $6^3 (= 216)$ triple of coordinates in the RGB-space. These RGB-values result for both Internet Explorer and Netscape under both the Windows and the Mac operating system, in the same (non-dithered) colors iff the operating system uses at least 8-bit (256) colors.

The stimulus (width 9.5 cm and height 6.0 cm) was presented in the center of the screen, on a gray background. Below the stimulus 11 buttons were placed (width: 1.8 cm and height 1.2 cm; width between: 0.6 cm). In the color memory

[4] http://www.vu.msu.edu/pearls/color/1.htm

experiment the buttons were labeled with the names of the 11 focal colors; in the color discrimination experiment each of the buttons did have one of the 11 focal colors. The 11 *focal colors* were presented conform the sRGB standard of the World Wide Web consortium (W3C)[5]. The button of choice was selected with one click of the mouse upon it.

2.4 Design

Half of the participants started with the color discrimination experiment, the other half started with the color memory experiment. Each experiment consisted of 4 blocks of repetitions of all stimuli (in a different order), preceded by a practice session. Each block consisted of the same 216 stimuli, randomized for each block and for each participant. In addition, the 11 buttons were also randomized for block and for each participant. The practice session consisted of 10 stimuli. Block, stimulus, and button order was the same for both experiments.Between the stimuli a blank screen was provided for one second, with a gray color.

The participants were asked to take a short break between the blocks of repetition, within each experiment and to take a somewhat longer break between both experiments. The duration of the breaks was determined by the subjects. In total a complete session took on the average 70 minutes, including breaks.

2.5 Procedure

The global scope of the experiment was explained, in which the experiments were conducted. After that a small questionnaire was completed. The first task was to write down the 10 colors that arise from memory first. Next, the design of the experiments was explained. The subjects were instructed for the color memory experiment to categorize the stimulus into one of the color categories, represented by their names. In the color discrimination experiment the subjects were asked to choose one of the 11 focal-colors that best resembled the stimulus. Last, was emphasized that there were no wrong answers and that if questions would arise they could be asked during one of the breaks.

3 Results

3.1 Mentioning of color names

For the determination of the confidence intervals we have used *the modified Wald method* [1] that proved to work well with a limited number experiments and with proportions close to 0 or 1.0; both the case in the present research. The proportion or frequency of appearance was determined by:

$$p = \frac{S+2}{N+4}$$

[5]http://www.w3.org/Graphics/Color/sRGB.html

where p is the proportion, S is the number of times the color is mentioned, and N is the number of subjects (26 in the present research).

The confidence interval was determined by:

$$p - \phi\sqrt{\frac{p\,(1-p)}{N+4}} \quad to \quad p + \phi\sqrt{\frac{p\,(1-p)}{N+4}}$$

where ϕ is 2.58 or 1.96 (in literature frequently rounded to 2.5 and 2 respectively) for the critical values from the Gaussian distribution for respectively 99% and 95%. The (relative) frequencies as well as the confidence intervals (both 99% and 95%) for all colors mentioned, are given in Table 1.

Table 1: Frequency and confidence-intervals of color names mentioned.

Color name	Frequency (in %)	min.-max. p at 99% (in %)	min.-max. p at 95% (in %)
red	26 (100.0%)	81.6% - 100.0%	84.4% - 100.0%
green	26 (100.0%)	81.6% - 100.0%	84.4% - 100.0%
yellow	26 (100.0%)	81.6% - 100.0%	84.4% - 100.0%
blue	26 (100.0%)	81.6% - 100.0%	84.4% - 100.0%
purple	24 (92.3%)	70.6% - 100.0%	74.5% - 98.8%
orange	22 (84.6%)	61.2% - 98.8%	65.7% - 94.3%
black	20 (76.9%)	52.5% - 94.1%	57.5% - 89.2%
white	20 (76.9%)	52.5% - 94.1%	57.5% - 89.2%
brown	20 (76.9%)	52.5% - 94.1%	57.5% - 89.2%
gray	15 (57.7%)	33.4% - 80.0%	38.9% - 74.4%
pink	11 (42.3%)	20.0% - 66.6%	25.6% - 61.1%
violet	06 (23.1%)	5.9% - 47.5%	10.8% - 42.5%
beige	04 (15.4%)	1.2% - 38.8%	5.7% - 34.3%
ocher	03 (11.5%)	0.9% - 34.2%	3.3% - 30.0%
turquoise	02 (7.7%)	2.7% - 29.3%	1.1% - 25.5%
magenta	02 (7.7%)	2.7% - 29.3%	1.1% - 25.5%
indigo	02 (7.7%)	2.7% - 29.3%	1.1% - 25.5%
cyan	02 (7.7%)	2.7% - 29.3%	1.1% - 25.5%
silver	01 (3.8%)	4.1% - 24.1%	0.7% - 20.7%
gold	01 (3.8%)	4.1% - 24.1%	0.7% - 20.7%
bordeaux-red	01 (3.8%)	4.1% - 24.1%	0.7% - 20.7%

There were some observations of the experimenter of possible factors of influence on the data provided by the question of mentioning 10 colors:

- Most subjects were directly able to write down 7, 8, or 9 color names, but experienced it as difficult to mention the last.
- A considerable number of participants asked whether black, gray, and white were colors during their task of writing down 10 color names. This was confirmed by the researcher who conducted the experiment.
- Another group of subjects indicated after they had written down the color names that their opinion was that these black, gray, and white are no colors. With that as opinion they had chosen to not write down black, gray, and

white. This explains for a large part the less frequently mentioned colors, most written down last.

As presented in Table 1, every subject named red, green, blue, and yellow. With 11 occurrences, pink was the least mentioned *focal color*. Nevertheless, pink was mentioned almost twice as much than the most frequently mentioned *non-focal color*: violet (6). The other *non-focal colors* were mentioned even less. In addition, the three observations mentioned above only confirms the existence of the *focal colors* in human memory.

3.2 The color discrimination experiment and the color memory experiment separate

The main result of both experiments is a table of markers for a $CLUT$[6]. The table distinguishes the discrimination and memory experiment.

We have analyzed the color discrimination experiment on each of the three dimensions: R, G, and B. Block was, on the average, a strong factor of influence on all three dimensions ($R : F(33, 192.21) = 917.90, p < .000; G : F(33, 192.21) = 1143.350, p < .000; B : F(33, 192.21) = 600.28, p < .000$). This held for all 11 color categories.

The same was done for the color memory experiment. Again block appeared a strong factor of influence on all three dimensions ($R : F(33, 192.21) = 756.54, p < .000; G : F(33, 192.21) = 785.99, p < .000; B : F(33, 192.21) = 451.35, p < .000$). Again this held for all 11 color categories.

3.3 The color discrimination and the color memory experiment together

The analysis of the experiments, conducted on the three dimensions: R, G, and B, showed a strong difference between the experiments on each of the three dimensions ($R : F(11, 15) = 2.96, p < .027; G : F(11, 15) = 7.843, p < .000; B : F(11, 15) = 3.11, p < .022$).

A more detailed analysis for each color category separate on the R dimension revealed that only purple ($F(1, 25) = 6.49, p < .017$) and red ($F(1, 25) = 20.50, p < .000$) were clearly under influence of the difference in buttons between both experiments; blue ($F(1, 25) = 3.48, p < .075$) and brown ($F(1, 25) = 3.74, p < .065$) showed only a tendency of influence.

On the G dimension all color categories, except gray and yellow, were strongly influenced by the difference in buttons between both experiments ($blue : F(1, 25) = 35.46, p < .000; brown : F(1, 25) = 33.52, p < .000; green : F(1, 25) = 21.79, p < .000; orange : F(1, 25) = 30.12, p < .000; purple : F(1, 25) = 15.91, p < .001; red : F(1, 25) = 12.58, p < .002; white : F(1, 25) = 22.26, p < .000; black : F(1, 25) = 35.27, p < .001$).

[6] The full table of markers for the $CLUT$ can be found at: *http://eidetic.cogsci.kun.nl/egon/demos/vindx_colorselector/*

Last, on the B dimension 6 color categories were strongly influenced by the difference between the experiments ($blue : F(1,25) = 7.67, p < .010; brown : F(1,25) = 8.67, p < .007; yellow : F(1,25) = 7,67, p < .010; pink : F(1,25) = 9.82, p < .004; white : F(1,25) = 7.19, p < .013; black : F(1,25) = 12.89, p < .001$) and orange showed a tendency of being influenced ($F(1,25) = 4.02, p < .056$); gray, green, purple and red were not influenced at all.

However, it is much more interesting to consider the colors independent of their (R, G, and B) dimensions. In both experiments (the overlap), 62 of the same web-safe colors were categorized as blue, 69 were categorized as green, and 49 were categorized as purple. For the first two of these color categories the difference in categorization between the experiments was marginal, for the latter a clear difference between both experiments was present. The remaining colors were categorized to one of the other 9 color categories. The overlap between both experiments for these categories was much smaller (average: 12.89; range: 4-20). The differences were large (average: 6.78; range: 1-19).

4 Discussion

The questionaire proved that the 11 color categories exist. This validated not only the choice of the 11 buttons used for the categorization of stimuli in the experiment, but, more importantly, it validated the idea to describe the RGB-color space using these color categories. When people use color categories when thinking, speaking, and remembering colors [2, 4, 5, 8], why not use them for describing the color space and use this description for CBIR? Since the existence of color categories proved to be valid we used them for two experiments on color categorization: one by way of color discrimination and the other by way of color memory.

Conform the hypothesis, no consistent color categorization was found *over* the experiments. This, despite the fact that the same stimuli were presented in the same blocks with the same button order, for each of the experiments. So, this leaves as conclusion that the cognitive processes of discrimination and memory influence color categorization strongly. Such a distinction argues in favor of different algorithms for color matching in CBIR, using on the one hand *query-by-example* and on the other hand *query-by-color*.

Color matching using a *CLUT*, based on the markers derived from the experimental results, could enhance the color matching process significantly. Results based on such a CLUT would be more intuitive for users. This would yield for the user more satisfying results than when using non-intuitive color matching functions founded on a color space.

Furthermore, the strong effect of the stimulus order on their perception was remarkable. This again indicates the strong influence of color memory on color perception. However, this did not explain that the *CLUT* markers define fuzzy boundaries between the color categories. This is due to a wide range of variables influencing color perception: memory, illumination, object identity, culture, emotion, and language [2, 4, 5, 8].

So, we have presented a division of the RGB-colorspace, that can be employed as a model of human color categorization founded on two different cognitive processes: color discrimination and color memory. We propose to implement separate color matching algorithms for *query-by-example* and *query-by-color*. Each comprising their own sustained basic color categories as fuzzy clusters in the CLUT. Such an approach would yield perceptually intuitive retrieval, and with that, satisfying results for the user.

5 Acknowledgments

The Dutch organization for scientific research (*NWO*) is gratefully acknowledged for funding Eidetic (project-number: 634.000.001), a project within the ToKeN2000 research line, in which this research was done. Furthermore, we would like to thank Leon van den Broek for advice of all kinds and Frans Gremmen for advice of statistical nature.

References

[1] A. Agresti and B.A. Coull. Approximate is better than exact for interval estimation of binomial proportions. *The American Statistician*, 52:119–126, 1998.

[2] B. Berlin and P. Kay. *Basic color terms: Their universals and evolution.* Berkeley: University of California Press, 1969.

[3] S. Douglas and T. Kirkpatrick. Do color models really make a difference? In M. J. Tauber, V. Bellotti, R. Jeffries, J. D. Mackinlay, and J. Nielsen, editors, *Proceedings of CHI 96: Conference on Human Factors in Computing Systems*, pages 399–405. ACM, New York: ACM Press, 1996.

[4] R.L. Goldstone. Effects of categorization on color perception. *Psychological Science*, 5(6):298–304, 1995.

[5] E. Rosch Heider. Universals in color naming and memory. *Journal of Experimental Psychology*, 93(1):10–20, 1972.

[6] Y. Rui, T. S. Huang, and S.-F. Chang. Image retrieval: Past, present, and future. *Journal of Visual Communication and Image Representation*, 10:1–23, 1999.

[7] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[8] E.L. van den Broek, L. G. Vuurpijl, P. Kisters, and J. C. M. von Schmid. Content-based image retrieval: Color-selection exploited. In M.-F. Moens, R. de Busser, D. Hiemstra, and W. Kraaij, editors, *Proceedings of the Dutch-Belgian Information Retrieval Workshop (DIR2002)*, volume 3, pages 37–46, Belgium, Leuven, December 2002. Catholic University Leuven, Belgium.