Toward Robust Categorical Color Perception

Giordano Beretta, Nathan Moroney, John Ludd Recker

HP Laboratories
HPL-2009-146

Keyword(s):
color difference; language; color naming

Abstract:
When implementing a software service for printing on commercial presses, it may be necessary to allow a casual user to modify the color rendering to remove a colorcast. In variable data printing, colored text on a colored background may no longer readable when the colors change or the text overflows. A high dynamic range photograph may require a large gamut mapping operation. These are examples where color matching and colorimetry are not adequate and a lexical color metric is more appropriate. We report on an experiment in progress to build a color name thesaurus. In particular, since lexical color is an acquired skill varying substantially from observer to observer, we present techniques we developed to increase robustness

External Posting Date: June 21, 2009 [Fulltext] Approved for External Publication
Internal Posting Date: June 21, 2009 [Fulltext]

To be published and presented at 11th Congress of the International Colour Association (AIC) 2009, Sydney, Australia, September 27th to October 2nd 2009.

EXTENDED ABSTRACT

Historically, a major focus in CIE colorimetry has been on color matching, i.e., on characterizing thresholds for local color differences near the just noticeable difference (JND) threshold. For some applications like user interfaces (communicating color attributes), readability, and gamut mapping (especially for wide gamut, high dynamic range devices), large global color differences are important.

Although from an algorithmic point of view perceptual metrics for small differences can be used for large differences, there is no psychophysical evidence that this is meaningful. Indeed, the human sensory systems are optimized for object/event recognition, not differences between stimuli. Therefore, it is relatively easy to devise a test for color matching, but it is harder to develop a metric for color difference. This is even more so for large color differences. Consequently, in applications like color readability and gamut mapping, one should not try to come up with thresholds based on metrics for color matching, but with a difference paradigm, like lexical distance.

The small color difference metrics are based on the physiology of color perception. Color names form a domain that allows a larger granularity and mediates a global metric through the lexical distance between colors. Linguists have long compiled color vocabularies, and color scientists have established links between such name compilations and color order systems and colorimetry. However, their use in applications like the three examples above (user interfaces, readability, gamut mapping), has never been completely satisfactory. One of the difficulties is that as an acquired skill: color names are culturally dependent and ephemeral.

One of the key characteristics of human cognition is categorization. The introduction of structure allows the distillation of data into knowledge through the reasoning about relationships. Categorical color perception, for example, manifests itself through the experimental effect that colors can be discriminated faster across color name categories than within color categories.

In this contribution we report on an experiment leading to the creation of a color thesaurus for commercial digital printing applications. We solve the problem of the requirement for a large data corpus due to the cultural and ephemeral characteristics by using crowd sourcing as a data collection and data validation tool. So far we have a large English data corpus that is the basis for an online color thesaurus. We have also collected data in other languages, but to date those corpora are still too small for a robust color thesaurus.

INTRODUCTION

Color Naming

In graphical user interfaces (GUI) it is difficult to communicate with the user through color coordinates[1], even when they correlate well with perceptual quantities. If a program can use expressions like *the lighter red*, it is easier to understand than for example `#ef3d47` or even (lightness 54, chroma 76, hue 29). Of course the computer and the color scientist are perfectly happy with tristimulus value (0.384044, 0.221477, 0.082137), but what is the mapping from this value to the name *light red*? And how is it different from *hot red*? Are they not just shades of *vermillion*? And if the context are flamingos, would that not just be *pink*?

Similarly, if the gamut-mapping algorithm changes a color with CIELAB coordinates (57, 48, 27) to one with coordinates (60, 46, 28), will the color consistency remain preserved or will the color cross a name boundary?

Although color consultants easily make up color names and then train their clients, as scientists we are constrained to algorithms based on sound universal principles. Linguists have long performed research on color names, because it presents many interesting research opportunities. However, linguists do not consider color appearance, which is the most important datum in color science.

One of the most fundamental contributions by linguists is the color ontogeny of languages by Berlin and Kay[2], which led to the concept of basic color terms. Unfortunately their research also had a negative impact on color naming research, because many color scientists have constrained their research to these basic color terms.

Our first motivation for this research has been the readability of colored text on a colored background. In variable data printing each printed copy is different from all others (for example magazine adverts are tailored to the subscriber income, in children's books the heroine is named after the child and is typed in its favorite color…), so visually checking for readability is not possible and we had to develop an algorithm that can be executed efficiently on a raster image processor (RIP).

The definition of *color discriminability* is clear — it refers to two colors that are above the matching threshold, and conventional colorimetry provides the necessary algorithms. *Color readability* refers to the ability of reading a colored text on a colored background without efficiency penalty, i.e., at the same speed or with the same number of reading errors[3] as black on white, i.e., with maximum contrast. Readability requires a metric for large color differences, as the efficiency threshold is well above the discriminability threshold.

Since any color combination can appear in a variable data piece (a work unit is called 'piece' in the graphic arts and 'job' in the print industry), a complete tiling of a color space is required, well beyond the basic color terms. Furthermore, since these colors are abstract, the experimental data must be based on aperture color (perceived color for which there is no definite spatial localization in depth, such as that perceived as filling a hole in a screen), so that memory color associations like the flamingos in the above example are ruled out.

Categorical perception refers to the perception of different sensory phenomena as being categorically different, while continuous perception refers to sensory phenomena located on a smooth continuum. Categorical perception occurs whenever perceived within-category differences are compressed and between-category differences are separated, relative to some unit comparison threshold.

Categorical perception was first studied in the context of color names, and Stevan Harnad[4] provided the following definition: A *categorical perception* effect occurs when (1) a set of stimuli ranging along a physical continuum is given one label on one side of a category boundary and another label on the other side and (2) the subject can discriminate smaller physical differences between pairs of stimuli that straddle the boundary than between pairs that are entirely within one category or the other.

A recent experiment at the University of Surrey[5] provides compelling evidence that categorical color perception is an acquired phenomenon, i.e., influences such as culture, education, socio-economic status (SES), etc. are important factors. Furthermore, there is a temporal effect as societies evolve.

A related concept is that of a *thesaurus*, which refers to a compilation of synonyms (and antonyms) with etymological and semantic information as well as examples to disambiguate the synonyms. We can think of a thesaurus as a categorization with the synonyms being categories and the disambiguation consisting in a geometric description of the perceptual relat[1]ion of two colors within a category, e.g. in terms of correlates for lightness, chroma, and hue.

Previous Work

Efforts based on the basic color terms, such as those by Boynton and Olson[6] and by Zollinger[7] are not adequate for our applications to commercial digital printing, for which we need a complete tiling of a color space.

The most well known color name thesaurus is that compiled by the ISCC and NBS[8]. Stone[9], Woolfe[1], and others have used this color naming system to build successful user interfaces for color specification. Unfortunately this system has several shortcomings, such as the categories being formed by committee consensus instead of psychophysics, and the color names derived from a small fixed number of color vocabularies available to the committee at a fixed point in time.

The work of Post[10] is based on psychophysics, but is limited to the basic color terms plus *peach*. The categorization is based on global statistics instead of performance within vs.

across categories. The results have been applied successfully to avionics applications, but the system is too limited for our more general applications in commercial digital printing.

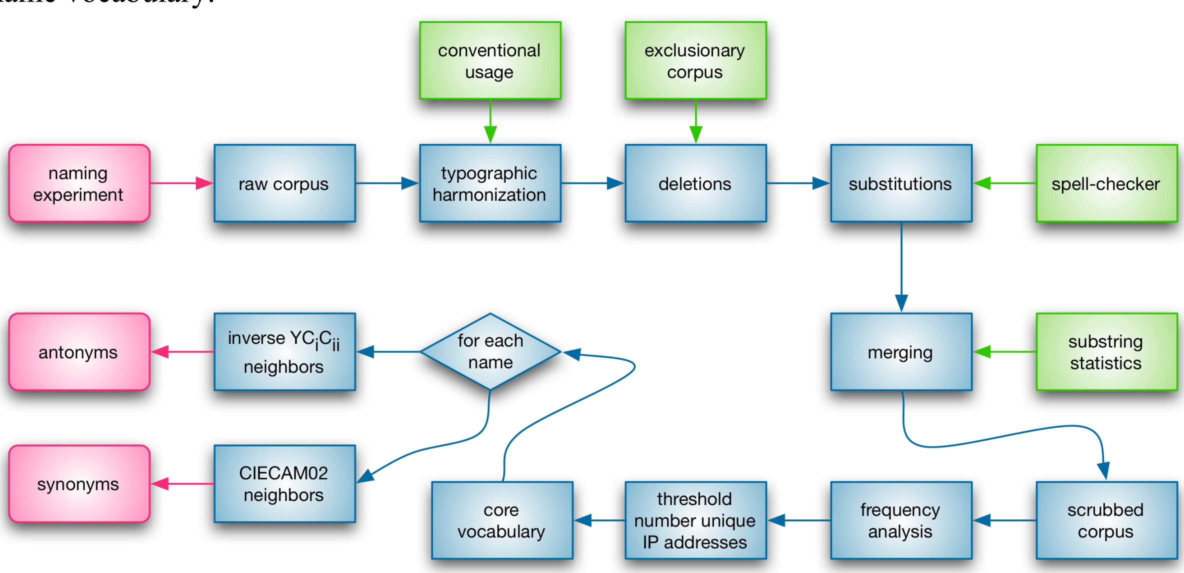
EXPERIMENT

The barriers for research in categorical color perception are the need for a large cross-cultural corpus of observers, and for the continuous update of the data corpus as societies evolve. Woolfe[1] has written about the difficulty in building such a corpus using traditional psychophysics experiments. We have designed an experiment that leverages on the World Wide Web for crowd sourcing.

The experiment consists in two steps. In a first step we collect color names by presenting to observers visiting our site seven random colors and asking them to name them[11]. In a second step we compile the collected names to a thesaurus, which provides value by giving the color coordinates for a color name, as well as a list of synonyms and antonyms.

This second step is important for two critical tasks: qualifying the corpus and expanding the corpus. Since categorical color perception is dependent on factors like SES and crowd sourcing always attracts a small number of disruptive visitors, vetting the data is important. This vetting itself can be performed through crowd sourcing in that the thesaurus users are asked to rate each result. This feedback-loop is used to weigh the data collected in the color naming experiment.

The expansion of the corpus is a subtle problem because statistically the distribution of color names is long tail, i.e., the basic color terms occur very frequently, while more literate names like ochre or emerald require many color naming instances to be collected within a given confidence interval. Again, we use the thesaurus as a crowd-sourcing tool. When a name is not found in the thesaurus with a sufficient confidence, the visitor is presented with a tool to specify a color for that name and the pair is added to the color name vocabulary.



The flowchart on the previous page illustrates the English implementation. At this time we have implemented the feedback loops only in the Italian version.

RESULTS

As of this writing, the English thesaurus site has served 182,079 color names, proving that it is possible to use crowd sourcing to perform color-naming experiments.

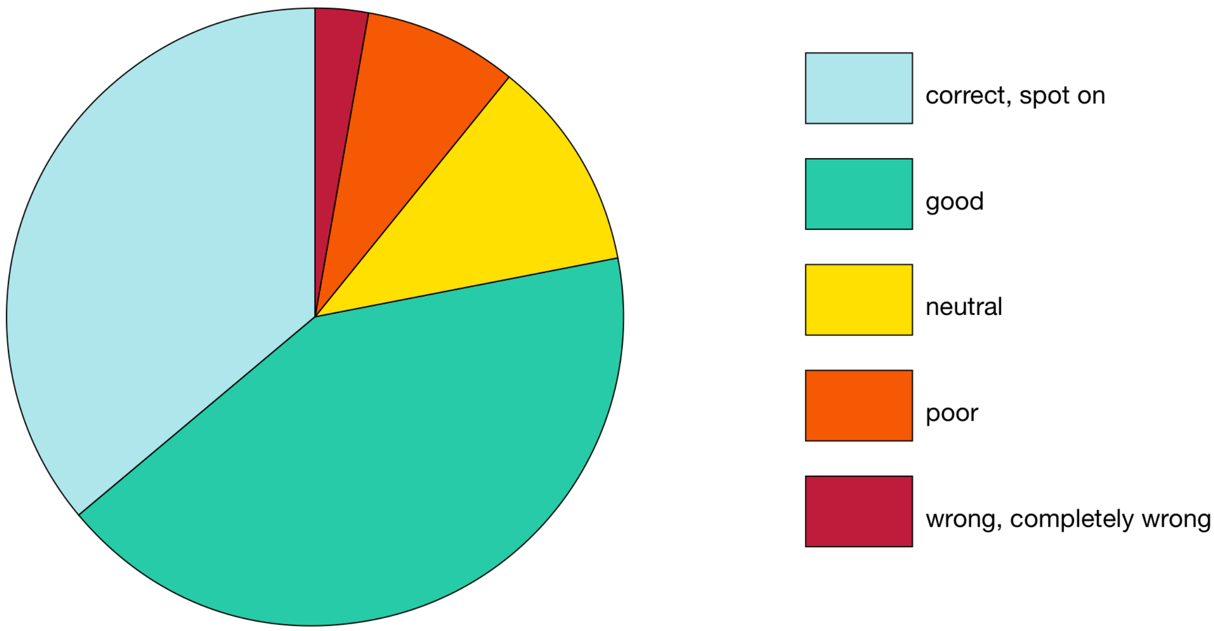
Our first implementation was achieved somewhat ad hoc to quickly provide a proof of concept that can attract funding and other support. In fact, the thesaurus is not based on categorical perception at all but the categories are faked at hoc by computing intervals of a fixed diameter in the perceptually uniform CIECAM02 color space. The antonyms are also computed ad hoc from the complementary color.

Of course, a delicate experiment as that designed by Franklin and her collaborators[5] cannot be performed through crowd sourcing. Instead, we have to design an experiment that can be administered in the form of a questionnaire. For example, in the color naming experiment we could ask for four color names instead of just one: a specific color name, a category name, an antonym, and an antonym category. Sophisticated statistical methods developed by linguists can be used to validate the collected data.

The Italian version of the color thesaurus is the most recent addition and has served almost 500 color names. This version of the tool also has two significant modifications relative to the English version. First is an optional rating of the found color names and corresponding color patch. The visitors used this optional rating in roughly ten percent of the number of names served.

A pie chart of these ratings is shown in graphical form on the next page. From this data we can see that the baseline quality of the names is quite good, with about 2.7% of the ratings being ‘Sbagliato, nemmeno per sogno’ or “Wrong, completely wrong.” Likewise almost 80% of the Italian color names are rated “correct, good, or spot on.” At the time of this writing, the English version of the Color Thesaurus does not yet have this type of user feedback on an individual name basis.

The second modification of the tool is a minimalist color picker for people to provide optional examples of missing color names. The result of this modification is that missing or new color names were harvested at a rate of about eight percent of the total number of names served. Again, the English version of the Color Thesaurus does not yet have a feedback mechanism for missing color names.



CONCLUSIONS

We have presented problems in commercial digital printing that cannot be solved with colorimetric color difference formulæ, such as GUIs, readability of colored text in variable data print jobs, and gamut mapping of high dynamic range images to printers. Categorical color perception can used to devise more suitable lexical metrics.

Current categorical perception data is not sufficient for industrial applications, but crowd sourcing on the World Wide Web shows promising results that scale well. Further research is required to tile the entire color space and build a true color thesaurus. Eventually, the tool should be so robust we can run it for a language none of us speaks.

REFERENCES

1. Woolfe G. Making Color Adjustmant Accessible to Non-Experts Through the Use of Language. In: Morovic J, Poynton C, editors. Fifteenth Color Imaging Conference. Albuquerque, New Mexico: IS&T; 2007. p 3-7.
2. Berlin B, Kay P. Basic Color Terms: Their Universality and Evolution. Berkeley: University of California Press; 1969.
3. Beretta GB. Color aspects of variable data proofing. In: Santini S, Schettini R, Gevers T, editors. Internet Imaging VI. Volume 5670. San Jose, CA; 2005. p 175-182.
4. Harnad SR. Psychological and cognitive aspects of categorical perception: A critical overview. In: Harnad SR, editor. Categorical Perception: The Groundwork of Cognition: Cambridge University Press; 1987. p 535-565.
5. Franklin A, Drivonikou GV, Bevis L, Davies IRL, Kay P, Regier T. Categorical perception of color is lateralized to the right hemisphere in infants, but to the left

hemisphere in adults. Proceedings of the National Academy of Sciences 2008;105(9):3221-3225.
6. Boynton RM, Olson CX. Locating Basic Colors in the OSA Space. Color research and applications 1987;12(2):94-105.
7. Zollinger H. Color: a multidisciplinary approach. Zürich: Helvetica Chimica Acta; 1999.
8. Kelly KL, Judd DB. The ISCC--NBS Method of Designating Colors and a Dictionary of Color Names. Washington: National Bureau of Standards; 1955.
9. Stone MC. Color, graphic design, and computer systems. Color Research & Application 1986;11 (Supplement):S75-S82.
10. Post DL. Applied Color-Vision Research. In: Widdel H, Post DL, editors. Color in electronic displays. Volume 3, Defense Research Series: Plenum Press; 1992.
11. Moroney N. Unconstrained web-based color naming experiment In: Eschbach R, Marcu GG, editors. Color Imaging VIII: Processing, Hardcopy, and Applications. Volume 5008. San Jose, CA; 2003. p 36-46.