Color Name Applications in Computer Vision

Abstract

In Computer Vision, the association of names to colors is one of the fundamental problems in the field of image understanding. There are numerous computational applications (e.g. image retrieval, visual tracking, person identification, human–machine interaction, etc.) that require pixels to be labelled according to the color perceived by the user. This is relatively easy for focal colors under canonical illuminants, where the agreement is high, but becomes increasingly difficult as perceptions move away from these conditions. For these difficult cases, the traditional solution tends to be a collection of “ad-hoc” strategies, however, new approaches that combine knowledge from anthropology, linguistics, visual perception and machine learning have offered promising results. Specifically, deep neural networks appear to possess all the required building blocks to offer a color naming solution “in the wild”. This article reviews the current state of knowledge and discusses open challenges with a multidisciplinary (and non-specialized) readership in mind.

Color Name Applications in Computer Vision

C. Alejandro Parraga[1] and Arash Akbarinia[2]

1- Computer Vision Centre / Computer Science Department, Universitat Autònoma de Barcelona, 08203, Spain.
2- Department of Psychology, Justus-Liebig Universität Giessen, D-35394 Giessen, Germany.

Synonyms

Color categorization; Semantic-color interaction; Color category boundaries; Color names for image understanding; Chromatic description; Chromatic labelling; Color terms

Definition

Humans exchange color information through the use of language, associating special words (color names) to categorical regions of color space. Color Name Applications in Computer Vision refers to algorithms that process digital images, labelling each pixel or pixel region with a given word or combination of words that matches the color name a human would give to such pixel or region.

The problem of labelling image regions by color name

Items belonging to different categories are easier to discriminate from each other than items belonging to the same category along a given perceptual dimension. In the case of color, the human visual system enhances the separation between certain regions of the color space so that some colors are perceived sharply different from its neighbors, as is the case of the "bands" we see in the rainbow. These regions are then semantically categorized and assigned color names to aid communication. There are many computational applications where we need to accurately replicate these semantic categorizations to have a performance similar to humans.
The task of associating names to colors seems extremely easy for us humans, until we are presented with (semantic) categorical boundaries. This is an all too familiar situation: we ask for a yellow paint and the shopkeeper brings us a brown-looking sample, we strongly disagree with friends about the color of our new car, or we are dumbfounded by the amount of different names given to "pink" in the paint samples' catalogue. Providing that viewing conditions are the same, it is safe to assume that healthy individuals perceive the same

colors[1] but still, there are regions of the color space where individual observers often assign different names to the same color. In summary, there is strong agreement when naming some regions of the color space and strong disagreement when naming others. The former are often called “focal colors”, selected by observers to best represent a given color name (e.g. “red” or “purple”, etc.) and the latter are often referred to as “categorical boundary” colors. The precise source of these individual differences remains unknown, although some authors point out to differences in the physiology (macular and lens density, pigmentation density, etc.), environmental factors (geographical differences in the environment’s statistics), linguistic and cultural factors, etc. Perhaps the most popular and influential study of basic color names across cultures and languages is that of Berlin and Kay [2], who proposed the existence of 11 “basic color terms” (in English they are “black”, “blue”, “brown”, “green”, “gray”, “orange”, “pink”, “purple”, “red”, “white” and “yellow”). This study was later supported by others [3]. Although this and most of the research involves focal colors, some researchers have explored and mapped the boundaries between categories trying to relate their findings to the subjacent neurophysiology [4, 5, 6]. Other factors such as viewing conditions, illumination or the influence of surrounding areas can modify not only the perception of a given color but category assigned.

Why label image regions by color name?

There are many tasks we want computers to perform better than humans (e.g. driving; recognizing objects, gait or faces; tracking; translating languages; etc.) and there are others where human performance is the milestone. In other words, we want algorithms to include the variability which is intrinsic in humans (e.g. art appreciation, image understanding, content categorization, etc.) In many such tasks, extracting higher level descriptors that provide clues to image content such as the names of the colors is fundamental. These descriptors can be later combined with image segmentation, used to select objects by color, describe the appearance of the image, generate semantic annotations etc. A typical example is image retrieval, where the user enters a semantic description of the image (objects content, size, position, main colors, etc.) and the algorithm retrieves the most likely candidates from a large image database. Nowadays when, human-machine interaction is ubiquitous across a wide range of applications, its success ultimately requires a smooth integration with color naming: we do not want to miss possible candidates because the algorithm is constrained to a tight definition of color categories. The same applies when semantically describing the content of the picture or classifying it automatically (e.g. images of brown cars versus images of beige cars). Researchers have made use of many techniques to solve these problems.

Computational solutions

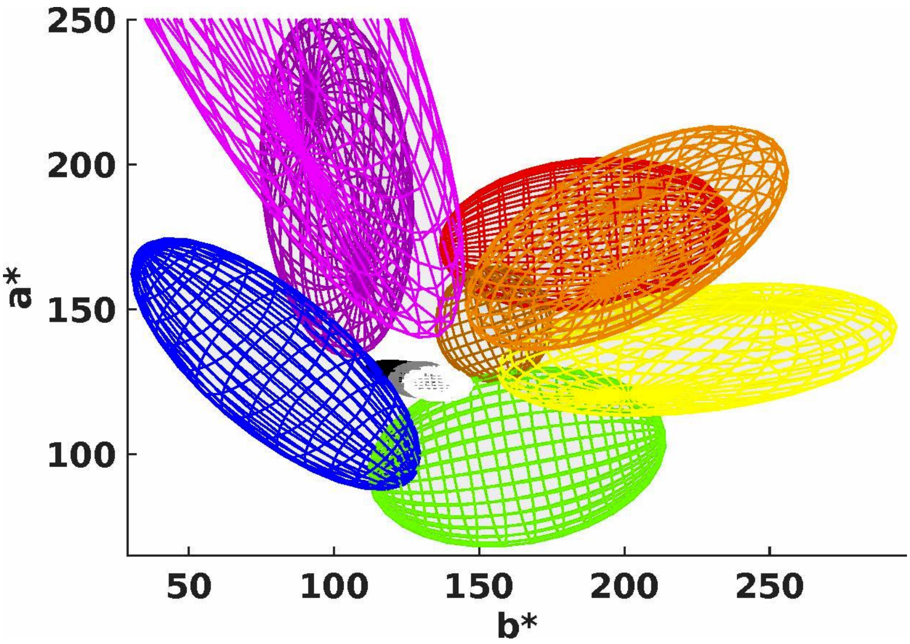
Broadly speaking, there are two kinds of computational solutions: those based on a psychophysically-based color-space partition and those developed by applying machine learning to natural images.

[1] This is not strictly true, there are some astonishing but very limited exceptions like #thedress [1]

Space partitioning-based models

The easiest computational approach to semantically labelling colors is to create “dictionaries” of exemplary colors and names [7] or to partition the color space in regions corresponding to an arbitrary number of color names. This fairly rigid approach works reasonably well for high agreement regions (focal colors) but is not good enough to capture the large variability of human percepts.

In 1978, Kay and McDaniel went further, proposing a model based in Fuzzy set theory (which in turn is a development from standard set theory) [4] that considers four fuzzy sets (“red”, “green”, “yellow”, and “blue”). The degree to which each pixel was a member of a particular semantic category was specified as a value between zero and unity. In other words, pixels were labelled with a numerical vector corresponding to their ‘belonging’ to these predefined semantic categories. In this framework, universal ‘focal’ color regions can be understood as regions of maximum color-category membership functions (i.e. unity). Correspondingly, non-focal colors have positive but non-maximal degrees of semantic category membership and are members of more than one basic category. For the English language, these include colors such as yellow-green, reddish-purple etc. These non-focal, multiple membership colors capture the variability in boundary judgments reported by Berlin and Kay [2]. The membership functions of the semantic color categories were derived from physiological neural response functions for wavelength and fuzzy set operations (union and intersection) . Using this model, they reinterpreted the evolutionary sequence of basic color names proposed by Berlin and Kay as the successive refinements of previously existing basic color names. A further extension was proposed by Lammens [8] who fitted Gaussian normal distributions to the 11 basic colors. This same paradigm was developed a step further by Benavente et al [9] who incorporated arbitrary “fuzzy” membership regions (combinations of Sigmoids and Gaussians) to fit B&K‘s categories to Surges and Whitfield’s psychophysical data [10] over three lightness levels in CIELab. Mojsilovic [11] proposed a geometrical solution in CIELab that assumed a well-represented set of prototypes (foci), computing the distance between a given color and all prototypes to obtain membership values. Seaborn et al [12] developed a model based on a Fuzzy C-Means clustering algorithm applied to psychophysical measurements by finding local minima within the group sum of squared error objective function. Menegaz et al [13] proposed a three-dimensional Delaunay triangulation of the color space to fit their own color name data to vertices of 3D tetrahedra in CIELab space. The final 11 fuzzy categories were obtained by a series of geometrical interpolations. All the previous models were adjusted to regions of high color-category membership (maximum agreement among observers), using linear interpolation to obtain membership values outside these. A completely different approach was adopted by Parraga and Akbarinia [6] who hypothesized that linear interpolation might not be the best way to obtain categorical boundaries in CIELab and decided to psychophysically measure the performance of observers in the transitional regions. They quantified the variability and fitted the surfaces of 11 3-dimensional Gaussians (3D ellipsoids, see Figure 1) to these boundaries. Table 1 provides a comparison of these color categorization models.



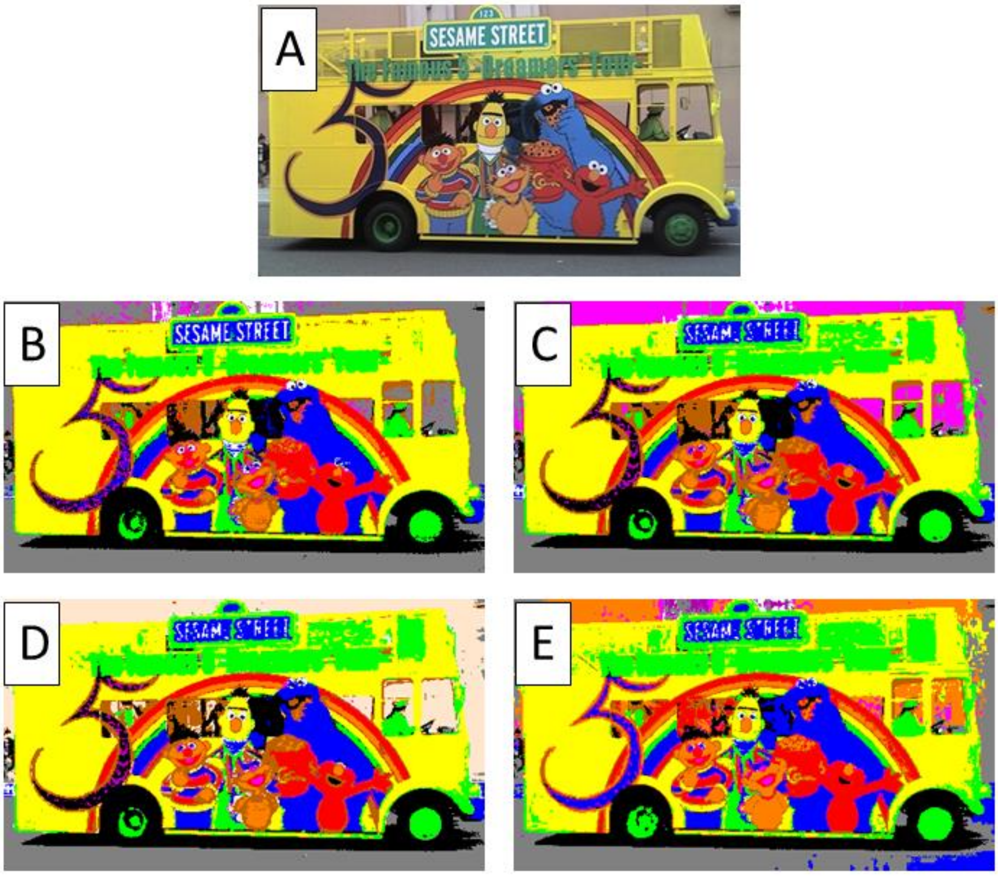
[Insert Figure 1 about here]
Figure 1. Schematics of the 3-dimensional ellipsoids used by Parraga and Akbarinia [6] to partition CIELab color space into 11 categorical regions.

Table 1. Comparative results of several color categorization models when applied to the Berlin & Kay [2] and Sturges & Whitfield [10] psychophysical results. Keys: LGM: Lammens's Gaussian model [8]; MES: MacLaury's English Speaker [14]; TSM: Benavente et al's Triple Sigmoid model [9]; SFKM: Seaborn's fuzzy k-means model [12]; TSMES: Benavente *et al*'s Triple Sigmoid-Eliptic Sigmoid model [15]; PLSA: van de Weijer et al's Probabilistic Latent Semantic Analysis [16] and NICE: Parraga and Akbarinia's Neural Isoresponse Colour Ellipsoids [6]. The data for LGM, MES, TSM, SFKM and TSEM was obtained from Table 2 in Benavente et al [15].

	Berlin and Kay results			Sturges and Whitfield results		
	Coincidences	Errors	% errors	Coincidences	Errors	% errors
LGM	161	49	23.33	92	19	17.12
MES	182	28	13.33	107	4	3.6
TSM	185	25	11.9	108	3	2.7
SFKM	193	17	8.1	111	0	0
TSEM	193	17	8.1	111	0	0
PLSA	187	23	12.3	109	2	1.8
NICE	206	4	1.9	111	0	0

Random initialization-based models

An alternative computational approach is by means of learning color names (categories) from "real-world" images. Following this line, Yendrikhovskij [17] reasoned that natural image's hues belonging to the same color name should cluster together in a perceptually uniform space. Consequently he applied k-means to acquire color categories from samples of pixels of natural images, using a minimum-distance criterion among members of the same color-category. In a later work, probabilistic latent semantic analysis (PLSA) was applied to learn color names from a collection of images obtained from search engines [16]. PLSA was originally introduced for document analysis and van de Weijer et al. [16] embraced this idea to model color pixels (words) in images (documents) through a combination of color names (topics). Akbarinia and Parraga [20] showed that the same set of images from [16] can be used to learn the parameters of their previously proposed ellipsoidal model, and furthermore to extend it to new color names. Figure 2 shows a comparison of some of these results.



[insert Figure 2 about here]

Results obtained by some of the computational color naming models mentioned above when applied to the image in panel A. Panels B, C D and E correspond to PLSA [16], NICE [6] (with 11 color terms), NICE [20] (with 12 color terms) and TSMES [15] respectively.

Color names have not yet been faithfully modelled by convolutional neural networks (CNNs), mainly because deep learning requires a large dataset for its training, which currently does not exist for color names. To date, there have been few attempts to create such datasets.

Cheng et al [21] used VGG16 to learn the color names of pedestrian's clothing items. Despite achieving a better performance in their own introduced fashion dataset, the resulting network lacks generalization to images of other environments. In summary, despite being a promising technological development, the use of CNNs to produce a more generic color naming model remains an open challenge.

Applications

Color names are actively used as a rich source of feature descriptors in a wide range of classical computer vision applications [19] that can be broadly categorized as: classification, recognition, tracking and retrieval.

Classification: a comparison of two sets of feature descriptors, one based on the classical eleven color names and another on raw hues, suggests that adding color names increase the robustness of the well-known bag-of-words method in the task of image classification [22]. Furthermore, the fusion of standard computer vision feature descriptors, e.g. HOG, with color names has been reported to improve the performance of object detection [23]. Overall, it can be argued that an explicit representation of color names is an effective method for improving object detection and classification.

Recognition: a similar fusion is proved to be successful in recognition tasks as well, Color names have been reported to provide a greater balance of photometric invariance and discriminative power. These properties make color names a suitable aid for action recognition [24]. Also, it has been shown that a feature representation grounded on 16 salient color terms facilitates person recognition and reidentification in video surveillance datasets [25].

Tracking: when it comes to visual tracking, it is well established that chromatic information provides a significantly aid to the task, however, it is computationally demanding. This complexity can be overcome by using color names, which have been shown to have the efficiency of raw chromatic information, while being computationally more affordable [26].

Retrieval: color names are one of the first attributes chosen by humans when describing an object. With this in mind, Zheng *et al* [27] supplemented a SIFT feature with basic color names, boosting the task of image retrieval by reducing false positives and improving query time. In order to reduce the semantic gap in algorithms of image retrieval, Liu *et al* [28] proposed to associate each region of an image with a color name. This technique better meets user expectations when they use color names to describe their query.

A recent online survey investigated which color names are commonly used without restrictions of choice [29]. The results suggest that average individuals use more than eleven color names. Based on these results, Yu et al [18] demonstrated that including a larger number of color names (28) instead of the traditional eleven, increases the discriminative power of the color descriptors in the tasks discussed above. Griffin and Mylonas [30] used the data from [29] as a color metric to geometrically categorize the color space. Interestingly, the number of emerging regions within the RGB cube corresponds well to the optimum names used in [18].

Within the past decade, the state-of-the-art in the above applications has shifted from classical computer vision algorithms to deep learning approaches where the optimization of the network's parameters (millions of them) is often best achieved in an end-to-end training procedure. In other words, the features learnt by these networks are mainly data driven. In most cases, imposing tailor-made features is likely to complicate the algorithms and impoverish the results. For this reason, incorporating color names to otherwise “pure” deep learning algorithms has been little explored.

Outline for future research

Although much has been done in the past, there is still large room for improvement before achieving human-like performance in color naming applications. The main problem is that humans are “noisy” when it comes to categorizing objects by color and this noise has both perceptual and cognitive sources. Another great limitation has to do with statically assigning a color term to a given RGB value. In this paradigm, an arbitrary RGB value will always have the same color name regardless of its context and surrounding. Very few works (see [20]) have tried to address this point by dynamically adjusting the categorical regions to the image context. Given that human color perception is strongly influenced by local and global context, this dynamic color naming approach seems promising. On the other hand, modern CNN models also promise to advance the field by expanding the number of categories and learning some of the cultural variability however, now there is a crippling lack of reliable datasets, although this is likely to improve in the future.

As a corollary, it is worth mentioning that the rewards for creating a robust color naming paradigm are enormous, since many important fields (e.g. social robotics, image understanding, etc.) rely on it. There are also important advances to be made in helping visually impaired people appreciate and navigate a colorful world.

References

1. Brainard DH and Hurlbert AC (2015) Colour Vision: Understanding #TheDress. Curr. Biol. 25: R551-R554.
2. Berlin B. and Kay, P.: Basic color terms: their universality and evolution. University of California Press, Berkeley; Oxford (1991)
3. Kay P and Regier T (2003) Resolving the question of color naming universals. Proc Natl Acad Sci U S A 100: 9085-9089.
4. Kay P, McDaniel CK (1978) The Linguistic Significance of the Meanings of Basic Color Terms. Language 54 (3):610-646
5. Olkkonen M, Hansen T and Gegenfurtner KR (2009) Categorical color constancy for simulated surfaces. J. Vis. 9: 6 1-18.
6. Parraga CA and Akbarinia A (2016) NICE: A Computational Solution to Close the Gap from Colour Perception to Colour Categorization. PLoS ONE 11: e0149538.

7. Maerz AJ and Paul MR (1930) A dictionary of color. McGraw-Hill Book Company, inc., New York.
8. Lammens JMG (1994) A computational model of color perception and color naming (PhD Thesis). In: Computer Science department, p. 253. State University of New York, Buffalo, NY.
9. Benavente R, Vanrell M and Baldrich R (2004) Estimation of fuzzy sets for computational colour categorization. Col. Res & Appl. 29: 342-353. (download model here)
10. Sturges J and Whitfield TWA (1995) Locating Basic Colors in the Munsell Space. Col. Res & Appl. 20: 364-376.2
11. Mojsilovic A (2005) A computational model for color naming and describing color composition of images. IEEE Transactions on Image Processing 14: 690-699.
12. Seaborn M, Hepplewhite L and Stonham J (2005) Fuzzy colour category map for the measurement of colour similarity and dissimilarity. Pattern Recognition 38: 165-177.
13. Menegaz G, Troter AL, Sequeira J and Boi JM (2007) A discrete model for color naming. EURASIP J. Appl. Signal Process. 2007: 113-113.
14. MacLaury RE (1992) From Brightness to Hue - an Explanatory Model of Color-Category Evolution. Curr Anthropol 33: 137-186.
15. Benavente R, Vanrell M and Baldrich R (2008) Parametric fuzzy sets for automatic color naming. J. Opt. Soc. Am. A 25: 2582-2593.
16. JOOST REFERENCE HERE
17. Yendrikhovskij SN (2001) A Computational Model of Colour Categorization. Col. Res & Appl. 26: S235-S238.
18. Yu L, Zhang L, van de Weijer J, Khan FS, Cheng Y and Parraga CA (2018) Beyond Eleven Color Names for Image Understanding. Machine Vision and Applications 29: 361-373.
19. Van De Weijer, J. and Khan, F.S., 2015, March. An overview of color name applications in computer vision. In International Workshop on Computational Color Imaging (pp. 16-22). Springer, Cham.
20. Akbarinia, A. and Parraga, C.A., 2017. Colour Terms: a Categorisation Model Inspired by Visual Cortex Neurons. arXiv preprint arXiv:1709.06300.
21. Cheng, Z., Li, X. and Loy, C.C., 2016, November. Pedestrian color naming via convolutional neural network. In Asian Conference on Computer Vision (pp. 35-51). Springer, Cham.
22. Khan, F.S., Van de Weijer, J. and Vanrell, M., 2012. Modulating shape features by color attention for object recognition. International Journal of Computer Vision, 98(1), pp.49-64.
23. Khan, F.S., Anwer, R.M., Van De Weijer, J., Bagdanov, A.D., Vanrell, M. and Lopez, A.M., 2012, June. Color attributes for object detection. In 2012 IEEE Conference on Computer Vision and Pattern Recognition (pp. 3306-3313). IEEE.
24. Khan, F.S., Anwer, R.M., Van De Weijer, J., Bagdanov, A.D., Lopez, A.M. and Felsberg, M., 2013. Coloring action recognition in still images. International journal of computer vision, 105(3), pp.205-221.
25. Yang, Y., Yang, J., Yan, J., Liao, S., Yi, D. and Li, S.Z., 2014, September. Salient color names for person re-identification. In European conference on computer vision (pp. 536-551). Springer, Cham.

26. Danelljan, M., Shahbaz Khan, F., Felsberg, M. and Van de Weijer, J., 2014. Adaptive color attributes for real-time visual tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1090-1097).
27. Zheng, L., Wang, S., Liu, Z. and Tian, Q., 2014. Packing and padding: Coupled multi-index for accurate image retrieval. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1939-1946).
28. Liu, Y., Zhang, D., Lu, G. and Ma, W.Y., 2005. Region-based image retrieval with high-level semantic color names. In 11th International Multimedia Modelling Conference (pp. 180-187). IEEE.
29. Mylonas, D. and MacDonald, L., 2016, Augmenting basic colour terms in English. Color Research & Application, 41(1), (pp. 32-42).
30. Griffin, L. D. and Mylonas, D., 2019. Categorical colour geometry. PLOS ONE, 14(5), e0216296.