COLOR REDUCTION BASED ON HUMAN CATEGORICAL PERCEPTION

R. Laganière[1], D. Pang[1], and A. Al-Kabbany[1,2]

[1] School of Electrical Engineering and Computer Science, University of Ottawa
[2]Faculty of Engineering, Arab Academy for Science and Technology

ABSTRACT

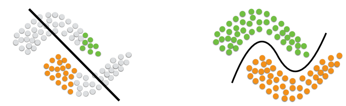
This paper addresses the problem of color reduction which aims at computing a compact representation of a color coordinate system. By capitalizing on studies that have suggested the existence of eleven focal colors, we conducted subjective experiments which exploited the categorical nature of human color perception. This paper describes a novel color reduction scheme based on human perception and graph transduction and proposes a look-up table that reduces the *RGB* color coordinate system into eleven salient colors. Objective results on standard datasets show improvements over previously proposed methods.

***Index Terms*—** Color, Reduction, Naming, Transduction, Propagation

1. INTRODUCTION AND RELATED WORK

Color information is used in many computer vision applications to characterize an image content. It is used in the detection, the recognition and the tracking of objects. Color also plays an important role in any content-based retrieval tasks. This paper addresses the problem of color reduction. It proposes a lookup table that reduces the space of color to 11 categories; these colors have been chosen to match the way humans perceptually classify colors. Indeed, several perceptual studies have shown that categorical perception plays an important role in the process of color discrimination and color memorization (e.g. [1]). Using human categorization of colors then appears to be an appropriate way of reducing the color space. A thus reduced color space will be particularly suitable to content-based image retrieval and classification tasks. The approach proposed here is based on the studies of [2] and [3] that both introduced a color reduction scheme based on 11 focal colors, the existence of which have been hypothesized in [4]. We extended these works by performing additional perceptual tests in order to come up with a color reduction that matches as much as possible human perception characteristics. In addition, we also introduced a fuzzy classification of the 11-color space in order to better take into account the subjective and contextual ambiguities in categorical perception.

This research is funded by NSERC.



(a) (b)

Fig. 1: The iconic illustration of transduction on the two half-moons configuration. Details are provided in the text.

Among the most recent methods in the literature is the probabilistic semantic analysis (PLSA)-based technique of van de Weijer et Al. [5, 3]. To overcome the drawbacks of chip-based methods [6, 7, 8, 9, 10] which cover some regions in the *RGB* cube only sparsely, they proposed to learn color names from real-world images, queried from Google Images. Towards this goal, they introduced two modifications in the standard PLSA [11] in order to take the image color label into account during learning. In [12], the authors addressed the problem of boosting the pixel-level color naming to the image/region level. Starting with an off-the-shelf color naming, they learned human prioritization of colors on natural images and then assigned to every region a dominant and an associative color name. Illumination-proof color naming was addressed by the method of [13] which relied on propagating color labels from regions under full illumination to abnormal regions with similar reflectance.

2. LEARNING BY TRANSDUCTION

In this research, we infer color names (labels) of *RGB* triplets using the graph Laplacian-based transduction method [14, 15]. Figure 1 shows a simplified case where all the observed data points (labelled and unlabelled) are available beforehand. The labelled data are shown as green and orange points in Fig. 1(a); the goal is to label the remainder grey points. Using just the labelled points to infer a predictive model may result in incorrectly fitting a hyperplane for example, as shown in Fig. 1(a). However, if the points are well-separated in the feature space, i.e., high-margin exists, transduction can learn a function from all the observed data points (both labelled and unlabelled) with the condition that it passes through regions of low-density in the feature space; this is shown in Fig. 1(b).

For discretization, graph Laplacian methods are adopted [16]; they are based on a discrete approximation of the sweighted Laplacian operator [15]. In these methods, a graph with nodes representing the data points (the X_is) is constructed, and the weights of the edges of that graph are induced using a kernel (often an exponential kernel) that represents the similarities (affinities) between the X_is in the feature space. The discrete approximation for the original optimization problem is given by

$$\min_{F \in \mathbb{R}^n} (F - Y)^T C (F - Y) + F^T L\, F, \tag{1}$$

where n is the total number of labelled and unlabelled data points, Y is the n-dimensional vector in which the i^{th} element is Y_i for a labelled point, and 0 for a test point, C is the diagonal $n \times n$ matrix in which the i^{th} diagonal element is c_i for a labelled point, and 0 for a test point, and L is the graph Laplacian. The n-dimensional vector F can then be obtained by solving the linear system given by

$$(L + C)\, F = C\, Y. \tag{2}$$

It should be noted that for binary labelling problems, such as the case in Fig. 1, the output vector F should be thresholded. The elements in F that correspond to the testing points in the constructed graph are the labels of the testing points. Transductive inference has been introduced to several computer vision problems including segmentation [15] and matting [17, 18].

3. PROPOSED METHOD

We propose a semi-supervised learning-based scheme for color reduction. Color names (Labels) for sparse triplets in the *RGB* cube are first provided by a group of volunteers to enforce a perceptually relevant labelling. Then, the acquired sparse labels are propagated to the whole *RGB* cube by means of the transduction of a graph whose nodes/vertices are all the triplets in the *RGB* cube. We start by detailing the setup and the procedure of the experiment used to gather sparse color labels in the *RGB* cube. Then, we present our graph transduction-based approach for constructing a lookup table that maps every *RGB* triplet to a fuzzy color naming vector. That vector indicates how likely a particular triplet is affiliated with each of the 11 Basic Color Labels. Using other color spaces (such as CIELAB) is discussed in section 4.

On voluntary basis, thirty-one students and researchers at the University of Ottawa, Canada, took part in the experiment. Prior to providing color labels, the participants were tested for color vision problems using the Ishihara color-blindness test [19]. Stimuli were displayed on a variety of monitors in average office environments. All computers were equipped with recent graphic card set at 16 millions colors. The procedure of our experiment was constrained by the number of available volunteers as well as the longest time period the volunteers could afford to participate in the experiment. Each of the thirty-one volunteers was assigned three images (a set of 93 training images, $\{\mathcal{I}_k^r\}$) to be labelled within a maximum of twenty minutes. Every image was shown as a set of SLIC [20] super-pixels (SPs) and the volunteers were instructed to give every super-pixel one of the 11 Basic Color Labels[1]. This results in a mapping from the *RGB* triplets – represented by the constituent pixels of a particular super-pixel – to the label set, the 11 Basic Color Terms.

Contrary to chip-based methods, we require the set $\{\mathcal{I}_k^r\}$ to sample the *RGB* cube efficiently, i.e., to ensure that the *RGB* triplets in the selected images are well distributed within the cube. Towards this goal, and starting with a super-set of 196 natural images featuring several object classes and a variety of scenes, we solved a knapsack problem to single out 93 training images so that we maximize *the coverage of the RGB cube*.

The last step in the procedure of the experiment is to compute the *Normalized Word Count* (as in the terminology of the authors of [5, 3]) for $\{\mathcal{T}_j^r\}$ – the set of *RGB* triplets contained in $\{\mathcal{I}_k^r\}$. A triplet in this set might have existed several times in $\{\mathcal{I}_k^r\}$, and might have been assigned different color labels by different/same participants as well since the *context* of an image was found to influence significantly the process of color naming. The normalized word count for $\mathcal{T}_j^r \in \{\mathcal{T}_j^r\}$, $\mathcal{C}_j^n$, is a vector with eleven elements, each of which is the number of times the triplet $\mathcal{T}_j^r$ was assigned the color name $l; l = 1, 2, \cdots, 11$ divided by the total number of occurrences of $\mathcal{T}_j^r$ in $\{\mathcal{I}_k^r\}$. The triplets in the *RGB* cube are given the symbol $\{\mathcal{T}_i\}$ where $i = 1, \cdots, 256^3$. The lookup table that realizes the proposed color reduction scheme maps $\{\mathcal{T}_i\}$ to $\{\mathcal{C}_i\}$ where every $\mathcal{C}_i := [c_1, \cdots, c_l, \cdots, c_{11}]$ with components indicating how likely $\mathcal{T}_i$ is affiliated to each of the 11 basic color labels.

We start the label propagation step by dividing the *RGB* cube into H sub-cubes. For each of these sub-cubes, we compute $\mathcal{T}_{jh}^r$ – *a pool of proposals* that is comprised of the subset of $\mathcal{T}_j^r$ that are not farther than a certain Euclidean distance (in the *RGB* color coordinate system) away from the centroid of the sub-cube under consideration. Figure 2 depicts an example of an *RGB* sub-cube and its corresponding *pool of proposals*, which is surrounded by a solid black circle. In this example, the proposals are mostly-red, mostly-pink, and mostly-orange, i.e., a proposal is mostly-red when the largest element in its normalized word-count vector is the element corresponding to the red color name. We call the red, the pink, and the orange: the *dominant color names* for the sub-cube in Fig. 2.

We loop over the sub-cubes, and for every dominant color name we construct a graph with nodes representing the *RGB* triplets in the sub-cube, $\{\mathcal{T}_h\}$, in addition to the proposals – the *RGB* triplets that are affiliated with the color name of the current loop. Then, we solve a graph transduction problem, by minimizing an objective function and solving a cor-

[1] We also allowed users to reject SPs if they contain more than one color.

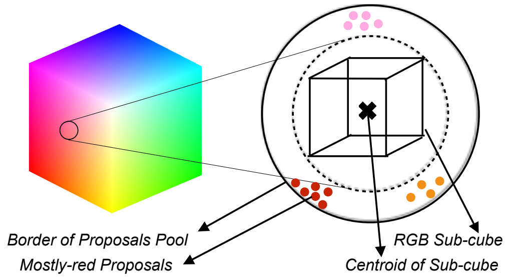


Fig. 2: An illustration of the proposals pool constructed for every sub-cube in the *RGB* cube. The normalized word counts of the shown proposals are propagated to the *RGB* triplets in the sub-cube by means of graph transduction. Please see text for more details.

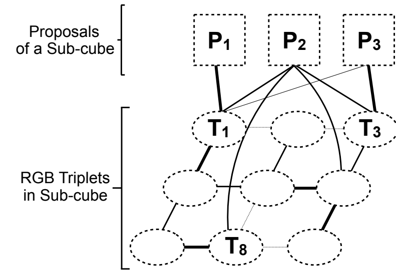


Fig. 3: An illustration of the graph constructed to propagate fuzzy color naming vectors from scarce manually-labelled proposals to every triplet in the *RGB* cube. Some edges are not shown for clarity of presentation. Please see text for more details.

responding linear system given by eqn.(1) and eqn.(2), to get a vector of scores whose length equals to the cardinality of $\{\mathcal{T}_h\}$, and whose elements indicate the *degree of acceptance* of the proposals by each member in $\{\mathcal{T}_h\}$. The constructed graph is shown in Fig. 3. The entries of the Laplacian matrix of this graph are calculated using the kernel function given by

$$k(X_i, X_j) = \frac{\widetilde{k}(X_i, X_j)}{[\widetilde{d}(X_i)\ \widetilde{d}(X_j)]^{\lambda}}, \qquad \text{where} \tag{3a}$$

$$\widetilde{k}(X_i, X_j) = \mathrm{e}^{-\frac{\|X_i - X_j\|^2}{2\sigma^2}}, \qquad \text{and} \tag{3b}$$

$$\widetilde{d}(X_i) = \sum_{j=1}^{n} \widetilde{k}(X_i, X_j), \tag{3c}$$

where n is the dimension of the Laplacian (square) matrix and X_i (and similarly X_j) is defined as the *RGB* feature vector.

After obtaining the vector F, by solving eqn.(2), we determine the vector F^s which is comprised of the elements in F that corresponds to the testing points in the constructed graph. F^s is the vector of scores which indicates the coherency between the triplets of a sub-cube and its proposals. We define $\mathcal{C}_{ih}$, the color naming vectors of the triplets $\{\mathcal{T}_h\}$ as:

$$\mathcal{C}_{ih} := \frac{1}{N} \sum_{n=1}^{N} F_n^s \times \mathcal{W}_n, \tag{4}$$

where N is the number of dominant color names in the sphere of proposals, F_n^s is F^s for the color name n and $\mathcal{W}_n$ is the 1×11 vector that is comprised of the maximum normalized word count that the proposals have for every color name. Accordingly, the fuzzy color naming vectors of the *RGB* triplets in a sub-cube are defined as the averages of the outer products of the vectors indicating their acceptance for the proposals and the vectors representing the maximum normalized word count for every color name across all the proposals.

A Bayesian Formulation for Color Naming: Following the authors of [5, 3], we aim to adapt our proposed approach so that an image 'color label' or 'region information' can be taken into consideration while assigning color names to pixels in that image or image region. The computed lookup table is equivalent to a distribution over *RGB* triplets representing the color names. In the terminology of the authors of [5, 3], this is the distribution over words representing the topics, $p(w|t)$, which can be used to compute the probability of a topic given a word by assuming a uniform prior over the topics, $p(t|w) \propto p(t)p(w|t)$. Thus, the probability of a color name (topic) given: 1) an *RGB* triplet (word) and 2) an image color label can be formulated as:

$$p(t|w, L) \propto p(w|t)p(t|L) \tag{5}$$

where L is an image color label, or generally, a prior over the frequency of color names in an image or a region in an image.

Counting on the intuition that *colors which often co-occur are more likely to affiliate to the same color name*, we computed a co-occurrence matrix, $\mathbb{M}$, for the color names. An entry in this 11×11 co-occurrence matrix is equal to the normalized count of the co-occurrence of those color names among the most likely K ($K = 3$ in our experiments) color names for all the triplets in the lookup table. For assigning a color name to a particular triplet $\mathcal{T}_i$ in the *RGB* cube, we have a likelihood term which is the color naming fuzzy vectors of the lookup table, and a prior term induced by the co-occurrence matrix. Thus, the color naming problem can be formulated in a Bayesian fashion as:

$$p(l|\mathcal{T}_i, L) \propto p(\mathcal{T}_i|l)p(l|L) \tag{6}$$

where $p(l|\mathcal{T}_i, L)$ is the probability of assigning a color name l to the triplet $\mathcal{T}_i$, given that the color label of the whole image is L. To bias the color naming with our prior, we determine *the index of the maximum value* in the vector $\mathcal{C}_i \times \mathbb{M}_L^{\lambda}$, where $\mathbb{M}_L$ is the column corresponding to L in $\mathbb{M}$ and λ is a weight for the prior. In both cases, we denote the maximum value in $\mathcal{C}_i$ as c_{il}^*.

4. RESULTS AND DISCUSSION

The results presented in this section were obtained using Matlab®, and were run on a PC with Intel Core2Quad

2.66GHz processor and 4GB of RAM. Since the PLSA-based methods of *PLSA-ind* and *PLSA-reg* have been shown to outperform the chip-based methods [6, 7, 8, 9], we compare the performance of our method with the former two methods only, in addition to the Dominant-Associative (DA) color naming of [12]. We had no access to the MCN dataset of [13]. It is worth mentioning that *PLSA-ind* and *PLSA-reg* represent two color naming algorithms that assign a color name for CIELab triplets with and without taking region information into consideration, respectively. Throughout this section, our proposed approaches of color naming (with and without considering prior information) will be referred to as *CTR* and *CTI* respectively. The performance of our method was found to be consistent across RGB and CIELAB color spaces. Hence, we show results for the former space only.

The performance of the proposed method is evaluated on the eBay image dataset[2]. This dataset is comprised of 4 object categories, namely, cars, shoes, dresses and pottery. For every category, there are 12 images for each of the 11 basic color names. Only pixel annotation is considered as an application of color naming. Evaluating the efficiency of the proposed method in object retrieval is left for future work.

Table 1 shows a comparison of the pixel annotation score for the aforementioned methods with the proposed method. Every figure represents the percentage of correctly-color-named pixels in every object category. We present our results for rank 1, 2, and 3 classifications. Results show that our CTR-Rk1 consistently outperforms PLSA-reg and DA color naming. In practice, a rank 'N' color-naming map can be obtained by optimizing a graph, using graph cuts for example, to ensure the smoothness of the assigned labels, i.e., every pixel is allowed a label just from its most probable 'N' labels.

Table 1: Comparison of pixel annotation scores of the methods proposed in [5] and [12] with our method. The third and the fourth sections of the table show the results of annotating pixels with and without considering the image label prior. For the two methods, we show the score for rank 1 (Rk1), rank 2 (Rk2), and rank 3 (Rk3) labellings.

Method	Cars	Shoes	Dresses	Pottery	Overall
PLSA-ind [5]	56	77	80	70	70.6
PLSA-reg [5]	74	94	85	82	83.4
DA [12]	63	88	85	79	78.8
CTI-Rk1	51.1	64	69.2	57.2	60.4
CTI-Rk2	70.7	81.7	86.5	77.5	79.1
CTI-Rk3	78.6	88.2	91.3	85.4	85.9
CTR-Rk1	88.1	94	94.9	92.3	**92.3**
CTR-Rk2	88.6	94.2	95	92.7	92.6
CTR-Rk3	89.1	94.4	95.1	93.4	93

Figure 4 depicts the results of the proposed method on eBay dataset. The 1st, 2nd, 3rd, and 4th columns show the original image, our CTI-Rk3 and CTR-Rk3 color maps obtained by solving a graph labelling problem using graph cuts [21, 22, 23, 24], and finally the color map of PLSA-bg* [3]. The upper row is for a white car while the lower row is for a grey pottery. While the pixel-level annotation of the proposed method, in some cases, assigns chromatic colors (e.g., green) to achromatic pixels (e.g., grey), the region-level annotation ameliorates this problem to a large extent, while achieving the highest score per object class and over the whole dataset as well (Table 1). Figure 5 depicts the results of the proposed method on two images of the BSD500 segmentation dataset. The 1st, 2nd, and 3rd columns show the original image, the CTI-Rk1, and the graph-cuts-smoothed CTI-Rk3 respectively.

[2]http://lear.inrialpes.fr/people/vandeweijer/color_names.html

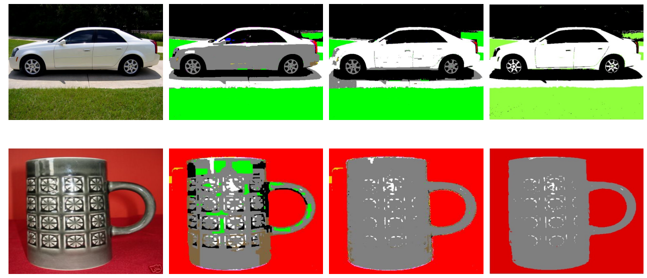


Fig. 4: Results of the proposed color naming method on the eBay dataset. From left to right: the original image, our graph cuts-smoothed CTI-Rk3 and CTR-Rk3 color maps, and the color map of PLSA-bg* [3].

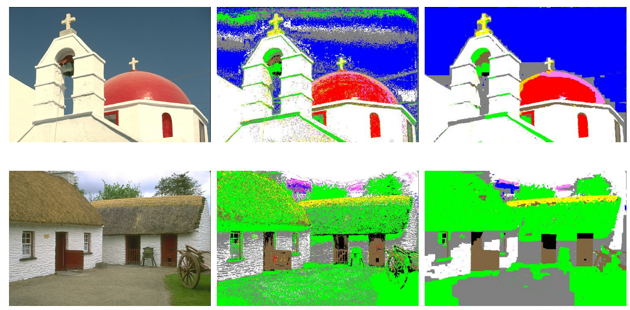


Fig. 5: Results of the proposed method on the BSD500 segmentation dataset. From left to right: original image, CTI-Rk1 and graph-cuts-smoothed CTI-Rk3 maps respectively.

5. CONCLUSION

In this paper, we propose a novel color reduction/naming scheme based on human categorical perception. Subjective experiments were conducted to sparsely sample and label the *RGB* cube and graph transduction was adopted to densely-label the whole cube. We contributed a look-up table for fuzzy color naming which has shown to achieve notable enhancements over previously proposed methods. Future directions include using the proposed method for object retrieval and tracking, and for scene classification.

6. REFERENCES

[1] Keiji Uchikawa and Hiroyuki Shinoda, "Influence of basic color categories on color memory discrimination," *Color Research and Application*, vol. 21, no. 6, pp. 430–439, 1996.

[2] E.L. van den Broek, Th.E. Schouten, and P.M.F. Kisters, "Modeling human color categorization," *Pattern recognition letters*, vol. 29, no. 8, pp. 1136–1144, June 2008.

[3] J. van de Weijer, C. Schmid, and J. Verbeek, "Learning color names from real-world images," in *IEEE TIP*, June 2007, pp. 1–8.

[4] Brent Berlin and Paul Kay, *Basic Color Terms: their Universality and Evolution*, University of California Press, Berkeley and Los Angeles, 1969.

[5] J. van de Weijer, C. Schmid, J. Verbeek, and D. Larlus, "Learning color names for real-world applications," *IEEE TIP*, vol. 18, no. 7, pp. 1512–1523, July 2009.

[6] M. Vanrell R. Benavente and R. Bladrich, "A data set for fuzzy colour naming," *COLOR Res. Appl.*, vol. 31, no. 1, pp. 48–56, 2006.

[7] K. Kelly and D. Judd, *Color: Universal Language and Dictionary of Names*, Number 440. MD: National Bureau of Standards, Gaithersburg, 1976.

[8] G. Menegaz, A. Le Troter, J. Sequeira, and J. M. Boi, "A discrete model for color naming," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, pp. 029125, 2006.

[9] G. Menegaz, A. L. Troter, J. M. Boi, and J. Sequeira, "Semantics driven resampling of the osa-ucs," in *Image Analysis and Processing Workshops, 2007. ICIAPW 2007. 14th International Conference on*, Sept 2007, pp. 216–220.

[10] Robert Benavente, Maria Vanrell, and Ramon Baldrich, "Parametric fuzzy sets for automatic color naming," *J. Opt. Soc. Am. A*, vol. 25, no. 10, pp. 2582–2593, Oct 2008.

[11] Thomas Hofmann, "Probabilistic latent semantic indexing," in *Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1999, pp. 50–57, ACM.

[12] Y. Liu, Y. Liang, Z. Yuan, and N. Zheng, "Learning to describe color composition of visual objects," in *ICPR*, 2012.

[13] Yuanliu Liu, Zejian Yuan, Badong Chen, Jianru Xue, and Nanning Zheng, "Illumination robust color naming via label propagation," in *ICCV*, 2015.

[14] Mikhail Belkin and Partha Niyogi, "Semi-supervised learning on Riemannian manifolds," *Machine Learning*, vol. 56, no. 1-3, pp. 209 – 239, 2004.

[15] O. Duchenne, J.-Y. Audibert, R. Keriven, J. Ponce, and F. Segonne, "Segmentation by transduction," in *CVPR*, 2008.

[16] Matthias Hein, Jean-Yves Audibert, and Ulrike von Luxburg, "From graphs to manifolds-weak and strong pointwise consistency of graph laplacians," in *the 18th Annual Conference on Learning Theory*, 2005.

[17] Jue Wang, "Image matting with transductive inference," in *Computer Vision/Computer Graphics Collaboration Techniques*, vol. 6930, pp. 239–250. 2011.

[18] Ahmad Al-Kabbany and Eric Dubois, "Matting with Sequential Pair Selection Using Graph Transduction," in *Vision, Modeling and Visualization*. 2016, The Eurographics Association.

[19] "Colorblind home page," `http://colorvisiontesting.com/ishihara.htm`, Accessed: 2017-2-7.

[20] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and Sabine Su, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE TPAMI*, vol. 34, no. 11, pp. 2274 – 2282, 2012.

[21] Yuri Boykov, Olga Veksler, and Ramin Zabih, "Fast approximate energy minimization via graph cuts," *IEEE TPAMI*, vol. 23, pp. 1222 – 1239, 2001.

[22] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?," *IEEE TPAMI*, vol. 26, pp. 147 – 159, 2004.

[23] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision," *IEEE TPAMI*, vol. 26, pp. 1124 – 1137, 2004.

[24] "The code page of the computer vision group at the University of Western Ontario," `http://vision.csd.uwo.ca/code/`, Accessed: 2017-2-7.