A Computational Colour Naming Model Trained on Real-Life Images

Hany M. SalahEldeen, Robert Benavente and Maria Vanrell

Computer Science Department, Computer Vision Center
Computer Vision Center
Edifici O - Campus UAB, 08193 Bellaterra, Spain
E-mail:{ hsalah, robert, maria.vanrell } @cvc.uab.es

Abstract

Colour naming is the process done by human beings when they assign linguistic terms to objects to describe their colour. In computer vision, several colour naming models were developed and each has its own advantages and drawbacks. Models based on psychophysical data have a robust perceptual basis and obtain good results in ideal laboratory conditions, but they lack precision when applied to real images. On the other hand, models fitted using image data sets achieve better results when applied to real-life images, but they are not directly related to human perception. Thus, the goal of this paper is to merge two of these approaches to obtain a new model which includes the advantages of both methodologies. This modified model was tested against a data set of real-world uncalibrated images and the results exceeded the original model.

Keywords: Colour Naming, Context-Based, Parametric/Perceptual Model, Uncalibrated Data Set.

1 Introduction

Colour naming is one of the several visual tasks commonly done by humans involving colour. It is the process of making a decision (in linguistic terms) about which colour best describes a given region of homogeneous hue. It is the last step in human colour-processing and it is performed in the visual cortex [8]. The aim of studying colour naming is to try to reduce the semantic gap in the task of giving names to colours in images. The semantic gap is the lack of a direct link between the low-level colour features extracted by machines and high-level semantics humans use. This gap is even more significant in applications like image retrieval where users require systems to support queries in natural languages [6]. Given so, an urgent need evolved to automate the process of colour naming and accurately imitating human perception in assigning colours. An automated colour naming model is a model that can

(a) (b)

Figure 1: The Result of Color naming a picture of a car.

correctly assign a colour term to a specific pixel. Provided by any given image, a colour naming model is supposed to have the ability of analyzing each pixel in the image and successfully decide to which colour category it belongs (red, green,...etc). Several models were created to solve this problem of colour naming; some of these models will be shown shortly[7][5].

2 Related Work

On the way of shaping the current understanding of the colour naming process nowadays, a lot of experiments were done and several models were developed. Some of which are Psychophysical, Neuropsychological or Computational [1, chap. 2].

Benavente et al.[2] developed a parametric model to fit data samples based on psychophysical experiments. These samples were fitted using a Triple Sigmoid function in six lightness layers. The idea behind creating this model is to find a suitable function capable of presenting the shape of each colour in the CIELab space[1]. Given a point in the colour space, it is possible to decide the membership of this point to each of the 11 basic colour terms of Berlin and Kay[2][3].

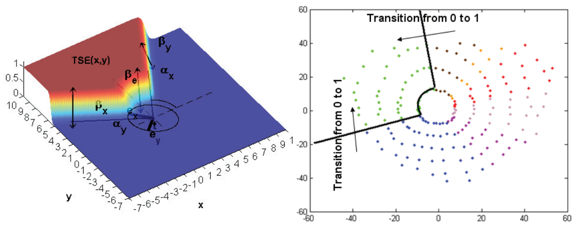
$$\mu_C(p,I) = \begin{cases} \mu_C = TSE(p, parL_1) & \text{if } I \leq I_1, \\ \mu_C = TSE(p, parL_2) & \text{if } I_1 < I \leq I_2, \\ . & \\ . & \\ \mu_C = TSE(p, parL_N) & \text{if } I_N < I, \end{cases}$$

μ_C is the membership of p to the chromatic[3] category C, I is the intensity level range and N is the number of lightness levels defined in the model while $parL_x$ are the parameters of level x. TSE stands for *Triple Sigmoid with Elliptical center* which is a variant of the one-dimensional sigmoid function as follows:

$$S^1(x,\beta) = \frac{1}{1 + exp(-\beta x)},$$

where β controls the slope of the transition from 0 to 1.

Each of the chromatics are represented by a Triple Sigmoid function with an elliptical centre (TSE) (see figure 2).



(a) (b)

Figure 2: The TSE function fitting the chromatics.

$$TSE(p,\beta,\alpha,e,\phi,t) = DS(p,\beta,\alpha,t) \cdot ES(p,e,\phi,t),$$

where DS is the Double Sigmoid function determining the separating boundaries between chromatics as follows:

$$DS(p,\beta,\alpha,t) = S_1(p,\beta_x,\alpha_x,t)S_2(p,\beta_y,\alpha_y,t),$$

$$\gamma_1 = (x - tx)cos(\alpha) + (y - ty)sin(\alpha),$$
$$S_1(p,\beta,\alpha,t) = \frac{1}{1 + exp(-\beta\gamma_1)},$$
$$\gamma_2 = (x - tx)(-sin(\alpha)) + (y - ty)cos(\alpha),$$
$$S_2(p,\beta,\alpha,t) = \frac{1}{1 + exp(-\beta\gamma_2)},$$

Vector α determines the axis in which the function is oriented, p is the (x,y) point investigated and t is where the origin was translated to.

Another type of sigmoid function is used to define the middle part in each layer where the achromatic colours[4] reside in the CIELab space. This

[1]The CIELab colour space is an approximately uniform colour space generated by optimal colour stimuli with respect to CIE standard illuminant D_{65}

[2]Pink, Red, Orange, Brown, Yellow, Green, Blue, Purple, Gray, Black and White

[3]Chromatics are colours Pink, Red, Orange, Brown, Yellow, Green, Blue and Purple

[4]Achromatics are colours Gray, Black and White

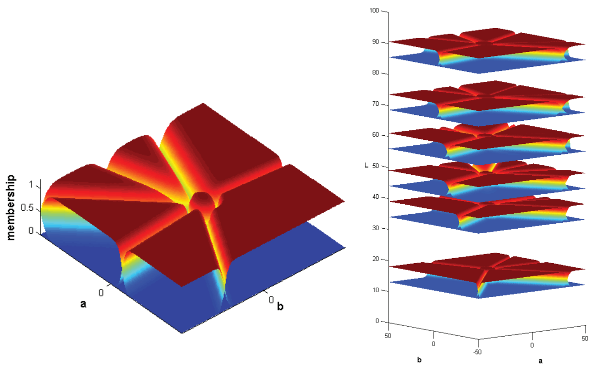
part takes an elliptical form and the function defining it is called the elliptic sigmoid function and it is illustrated as folows:

$$\gamma_1 = (\frac{(x-tx)cos\phi + (y-ty)sin\phi}{e_x}),$$

$$\gamma_2 = (\frac{(x-tx)(-sin\phi) + (y-ty)cos\phi}{e_y}),$$

$$ES(p,e,\phi,t) = \frac{1}{1+exp(-\beta_e(\gamma_1+\gamma_2))},$$

where e is the length of the axes of the central ellipse and ϕ is the rotation of this ellipse. Furthermore, within the elliptical sigmoid centre, the three achromatic colours reside and are separated by lightness through a one-Dimensional Sigmoid function.



(a) (b)

Figure 3: The Triple Sigmoid Elliptical centre model in one of the 6 intensity levels.

The Probabilistic Latent Color Naming Model was developed by van de Weijer et al.[10] based on the same concept of the *latent aspect models*. One of the most interesting characteristics of this model is that it was fitted using a data set of real-world uncalibrated images. This data set was obtained from Google images search engine and it is characterized by being weakly labelled. These weakly labelled Google images are represented by their normalized Lab histograms. These histograms form the columns of the image specific word distribution $p(w|d)$. See figure 4.

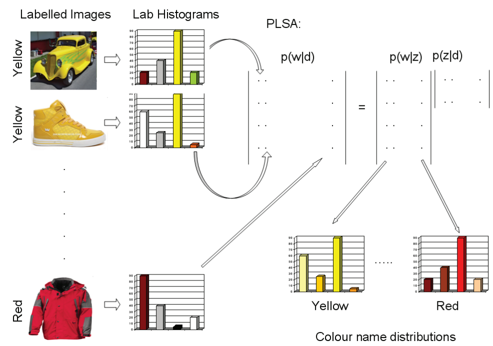


Figure 4: The PLSA color-naming model

After analyzing two of the most interesting approaches in solving the problem of colour naming automation some conclusions were drawn. Firstly, each of the two models possesses several advantages and suffers from other drawbacks. The TSE model is a parametric model, thus it is compact and easy to analyze. The compactness feature comes from the ability to fully describe a certain colour in few parameters. The ease of analysis and comparison comes from the ability of comparing different versions of the model, different colours and different intensity levels with each other by examining the corresponding parameters (see [1, chap. 4]). The PLSA model is a probabilistic model based on uncalibrated images from real-world. Thus, this model is more capable of correctly naming colours in images where acquisition conditions are unknown. The data set in which the model was trained contains built-in information about context as well. Secondly, the parametric model is powerful but lacks the ability of labelling uncalibrated images from real-world with high accuracy. These images are characterized by the variety of acquisition conditions from illuminant colour, angle of acquisition, shadows, reflectance...etc. on contrary to the psychophysical data obtained in an ideal controlled environment.

Given so, it is desired to create a model that enjoys both of the advantages of the two approaches. This model is required to be parametric and at the same time trained on context-based data from real-world uncalibrated images. The steps of developing this model will be illustrated in the next section.

3 A Parametric Colour Naming Model for Uncalibrated Real-World Images

In the end of the previous section, it was concluded that it is needed to create a context-based parametric model. In order to achieve this goal, the parametric model of Benavente et al will be fitted using the context-based data set utilized in the probabilistic semantic model of van de Weijer et al.

3.1 Training the TSE model on Uncalibrated Data

As it is illustrated in the TSE model of Benavente et al., a data set based on psychophysical experiments is utilized. These psychophysical experiments were commenced by Seaborn et al.[9] to model human perception of colour. A fuzzy colour category map is resulted from the analysis of these experiments. This map is sampled uniformly to be used in fitting the parametric model. The samples used in the fitting are gathered in what will be mentioned in the rest of the text as Lut (*Look-up table*).

For the fitting phase, the psychophysical-based Lut will be replaced by another resulted from the analysis of Google data set images by the probabilistic model of van de Weijer et al. mentioned earlier. The resulting Lut is considered the fitting set on which the model will be fitted. As this fitting set is obtained from uncalibrated images in the real-world, thus, it contains context-embedded information.

The modified Lut was provided to the TSE model in the fitting phase. Using the same number of intensity levels and the same ranges that those levels cover, the entire Lut was used in the fitting phase. The model was fitted successfully on the data provided as expected with some confusion areas. Fitted to the new data set, the fitting error of the model was calculated with acceptable error margin as demonstrated in table 1. The error is calculated using Mean Absolute Error (MAE).

$Method$ $name$	$Num\,of$ $samples$	MAE $fitting$	$\%\,of\,well$ $fitted\,samples$
Original TSE	1617	1.68%	96.60%
TSE_{uncal}	**32768**	**3.94%**	**85.92%**

Table 1: Statistics on the TSE model fitted to uncalibrated data.

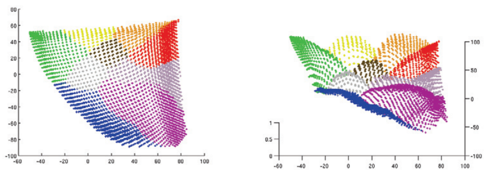
where TSE_{uncal} is the original TSE model fitted to the uncalibrated data set alone.

3.2 Bi-Elliptic Triple Sigmoid Model

It was proved that the model can represent context-based data from uncalibrated images. After analyzing the fitting error and backtracking the misfitted samples it was noted that the colour brown achieves a high misfitting error which required a further analysis. A hypothesis was proposed that the brown colour exhibits a different behaviour than the rest of the chromatics and it is better fitted by another function due to the high error rate that it produce. This odd behaviour could be due to the non-ideal acquisition conditions of the real-life uncalibrated images and the embedded context information in the data samples.

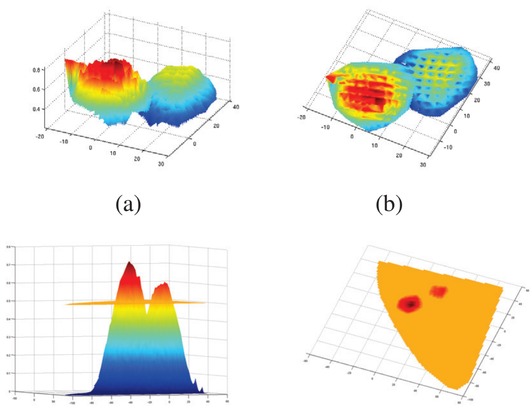
The samples from the full Lut were gathered, divided into several lightness levels and plotted on a 3D plot (see figure 5). These samples are interpolated into a smooth surface maintaining the same characteristics. This was done to give an idea about the functions that will be needed to fit these data samples.

The analysis of the membership distribution of the brown colour confirmed the feasibility of well-fitting the data samples of brown colour using another elliptical sigmoid function to be located in



(a) (b)

Figure 5: Samples plotted according to their memberships in space, the X-Y axes are the a-b coordinates and the Z axis is the membership value for this data sample (from 0 to 1), right figure is 2D prospective.



(a) (b)

(c) (d)

Figure 6: Analysis and surfing of the brown and achromatics-sum, right figures are 2D prospective.

the same vicinity of the achromatics-sum elliptical sigmoid. These two elliptical sigmoids would be surrounded by the triple sigmoid functions representing the rest of the chromatic colours. The rest of the chromatics maintained their expected triple sigmoid shape having the elliptic sigmoid of the achromatics to be the guiding centre. The work of Boynton[4] supports this hypothesis as well, and shows how the centroids of the colours are located in the perceptually uniform OSA space. This work illustrates how the chromatics surround the Grey-Black, White and Brown centroids uniformly.

By applying this hypothesis on the data samples and modifying the fitting functions, the new Bi-Elliptical Triple Sigmoid (BETS) model was developed. It is a modification to the original Triple Sigmoid model with two elliptical centres instead of one. One of the elliptical centres is used to define Brown while the other is used as before to define the achromatics. Figure 7 shows the new Bi-Elliptical Triple Sigmoid model and table 2 shows the fitting error resulted from testing the new model.

Method name	*Num of samples*	*MAE fitting*	*% of well fitted samples*
Original TSE	1617	1.68%	96.60%
TSE_{uncal}	32768	3.94%	85.92%
BETS	**32768**	**5.08%**	**88.97%**

Table 2: Statistics on the Bi-Elliptical Triple Sigmoid model.

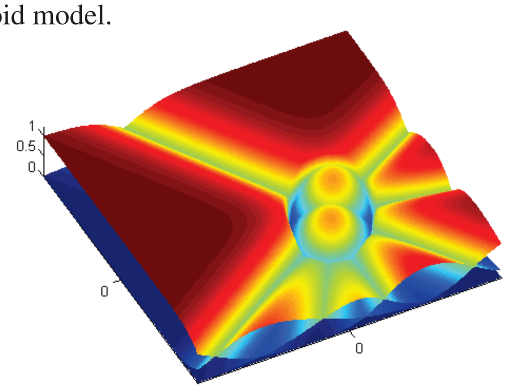


Figure 7: The Bi-Elliptical Triple Sigmoid model for an intensity level.

4 Context-Based Colour Naming in Real-World Images

After developing the new Bi-Elliptical TSE model and testing its fitting, it is needed to test it in reality as well. The fitting error is a good measure for the accuracy of the model in representing the fitting data set but it is not enough to give a whole picture about the model. It is also needed to prove that the model produces good results upon testing against real-life images having different and unknown acquisition conditions. For these reasons the models were tested upon each step against the eBay images data set[10].

The eBay data set contains 4 categories (cars, dresses, pottery and shoes). Each category contains 10 images from each of the 11 basic colours. Along with each image there is a mask specifying the region in the image that is labelled with this colour. The testing takes place by applying the model to the pixels of each image within its mask and calculate the percentage of pixels correctly labelled with the expected colour (e.g. measuring the percentage of pixels labelled as red by the model to all the pixels inside the mask of an image containing a red car). Table 3 shows these results.

$Method$	TSE	TSE_{uncal}	$Bi\_EllipticalTS$
Cars	53.34%	48.79%	51.58%
Dresses	73.25%	64.32%	78.18%
Pottery	61.5%	62.49%	71.29%
Shoes	72.19%	62.52%	70.61%
Total	**65.07%**	**59.53%**	**67.91%**

Table 3: Results of testing against the eBay data set.

5 Discussion and Future Work

From the experiments and analysis done so far, several conclusions could be drawn. The data from real-world uncalibrated images maintains a different shape, orientation and location from the data of psychophysical experiments. The parametric model fits the data but with minor modifications. After the analysis of this shifting and reformation, a conclusion was reached that the brown colour doesn't maintain the same behaviour of the other chromatics and tends to follow the behaviour of the achromatics in taking the shape of an elliptic sigmoid situated next to the achromatics and surrounded by the other chromatics. After successfully fitting the model to the context-based data set it was tested against uncalibrated images. The new model achieved higher results than the original TSE model tested against the same data set separately. Thus, proving the hypothesis stated in the beginning of the possibility of creating an improved colour naming model by incorporating context-based real life data in the fitting set of a perceptual model. As for future work, a thorough analysis must be commenced for the areas between the elliptical centers and the other triple sigmoids and a search for a modification to the functions to provide better fitting should be established.

References

[1] R. Benavente. *A Parametric Model for Computational Colour Naming*. PhD thesis, Universitat Autònoma de Barcelona, Bellaterra (Spain), Jun 2007.

[2] R. Benavente, M. Vanrell, and R. Baldrich. Parametric fuzzy sets for automatic color naming. *Journal of the Optical Society of America A*, 25(10):2582–2593, Oct 2008.

[3] B. Berlin and P. Kay. *Basic Color Terms: Their Universality and Evolution*. Center for the Study of Language and Inf, March 1969.

[4] R. M. Boynton and C. X. Olson. Locating basic colors in the osa space. *Color Research & Application*, 12(2):94–105, 1987.

[5] L. MacDonald A. Tarrant H. Lin, M. Luo. A cross-cultural colour-naming study. part iii–a colour-naming model. *Color Research and Application*, (26):270–277, 2001.

[6] E. S. Konak, U. Gudukbay, and O. Ulusoy. A content-based image retrieval system for texture and color queries, 2002.

[7] J. M. Lammens. *A Computational Model of Color Perception and Color Naming*. PhD thesis, University of New York, 1994.

[8] P. Lennie. Single units and visual cortical organization. *Perception*, 27:889–935, 1998.

[9] M. Seaborn, L. Hepplewhite, and T. John Stonham. Fuzzy colour category map for the measurement of colour similarity and dissimilarity. *Pattern Recognition*, 38(2):165–177, 2005.

[10] J. van de Weijer, C. Schmid, J. Verbeek, and D. Larlus. Learning color names for real-world applications. *IEEE Transactions on Image Processing*, 18(7):1512–1523, jul 2009.