CHAPTER 2

COLOR NAMES

with Robert Benavente, Maria Vanrell, Cordelia Schmid, Ramon Baldrich, Jakob Verbeek, and Diane Larlus

Within a computer vision context color naming is the action of assigning linguistic color labels to pixels, regions or objects in images. Humans use color names routinely and seemingly without effort to describe the world around us. They have been primarily studied in the fields of visual psychology, anthropology and linguistics [17]. Color names are for example used in the context of image retrieval. A user might query an image search engine for "red cars". The system recognizes the color name "red", and orders the retrieved results on "car" based on their resemblance to the human usage of "red'. Furthermore, knowledge of visual attributes can be used to assist object recognition methods. For example, for an image annotated with the text "Orange stapler on table", knowledge of the color name orange would greatly simplify the task of discovering where (or what) the stapler is. Color names are further applicable in automatic content labelling of images, colorblind assistance, and linguistic human-computer interaction [41].

Please enter \offprintinfo{(Title, Edition)}{(Author)}
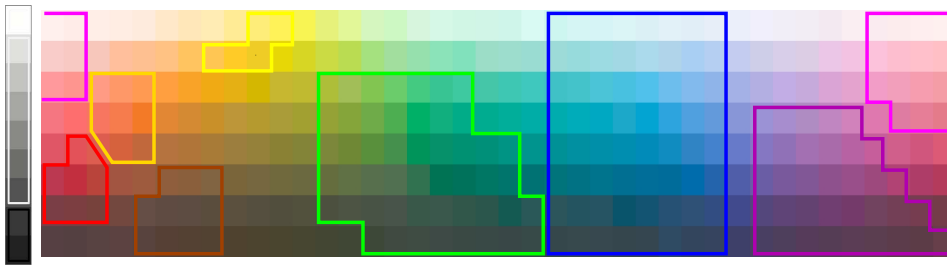at the beginning of your document.

Figure 2.1   *Categorization of the Munsell colour array obtained by Berlin and Kay in their experiments for English.*

In this chapter, we will first discuss the influential linguistic study on color names by Berlin and Kay [7] in section 2.1. In their work they define the concept of *basic color terms*. As we will see, the basic color terms of the English language are black, blue, brown, grey, green, orange, pink, purple, red, white, and yellow. Next, we will discuss two different approaches to computational color naming. The main difference between the two methods is the data from which the methods learn the color names. The first method, discussed in section 2.2, is based on *calibrated data* acquired from a psychophysical experiment. With calibrated we mean that the color samples are presented in a controlled laboratory environment on stable viewing conditions with a known illumination setting. The second method, discussed in section 2.3, is instead based on *uncalibrated data* obtained from Google Image. These images are uncalibrated in the worst sense. They have unknown camera settings, unknown illuminant, and unknown compression. However, the advantage of uncalibrated data is that it is much easier to collect. At the end of the chapter, in section 2.4, we compare both computational color naming algorithms on both calibrated and uncalibrated data.

2.1   BASIC COLOR TERMS

Color naming, and all the semantic fields in general, have been involved for many years in a discussion between two points of view in linguistics. On the one hand, relativists support the idea that semantic categories are conditioned by experience and culture, and therefore, each language builds its own semantic structures in a quite arbitrary form. On the other hand, universalists defend the existence of semantic universals shared across languages. These linguistic universals would be based on the human biology and directly linked to neurophysiological mechanisms. Color has been presented as a clear example of relativism since each language has a different set of terms to describe color.

Although some works had investigated the use of color terms in English [26], the anthropological study of Berlin and Kay [7] about color naming in different languages was the start point of many works about this topic in the subsequent years.

Berlin and Kay studied the use of color names in speakers of a total of ninety-eight different languages (20 experimentally and 78 through literature review). With their work, Berlin and Kay wanted to support the hypothesis of semantic universals by demonstrating the existence of a set of color categories shared across different languages. To this end,

they first defined the concept of "basic color term" by setting the properties that any basic color term should fulfil. These properties are:

- It is monolexemic, i.e. its meaning can not be obtained from the meaning of its parts.

- It has a meaning which is not included in that of other color terms.

- It can be applied to any type of objects.

- It is psychologically salient, i.e. it appears at the beginning of elicited lists of color terms, it is consistently used along time by speakers and across different speakers, and it is used by all the speakers of the language.

In addition, they defined a second set of properties for the terms that might be doubtful according to the previous rules. These properties are:

- The doubtful form should have the same distributional potential as the previously established basic terms.

- Basic color terms should not be also the name of an object that has that color.

- Foreign words that have recently been incorporated to the language are suspect.

- If the monolexemic criterion is difficult to decide, the morphological complexity can be used as a secondary criterion.

The work with informants from the different languages was divided in two parts. In the first part, the list of basic color names in each informant's language, according to the previous rules, was verbally elicited. This part was done in the absence of any color stimuli and using as little as possible of any other language. In the second part, subjects were asked to perform two different tasks. First, they had to indicate on the Munsell color array all the chips that they would name under any condition with each of their basic terms, i.e. the area of each color category. Second, they had to point out the best example (focus) of each basic color term in their language.

Data obtained from the 20 informants was completed with information from published works in other 78 languages. After the study of these data, Berlin and Kay extracted three main conclusions from their work:

1. *Existence of Basic color Terms.* They stated that color categories were not arbitrary and randomly defined by each language. The foci of each basic color category in different languages were all in a close region of the color space. This finding led them to define the set of eleven basic color terms. These terms for English are white, black, red, green, yellow, blue, brown, pink, purple, orange and grey.

2. *Evolutionary order.* Although languages can have different number of basic color terms, they found that the order in which languages encoded color terms in their temporal evolution was not random, but it followed a fixed order that defined seven evolutionary stages:

   - Stage I: Terms for only white and black.
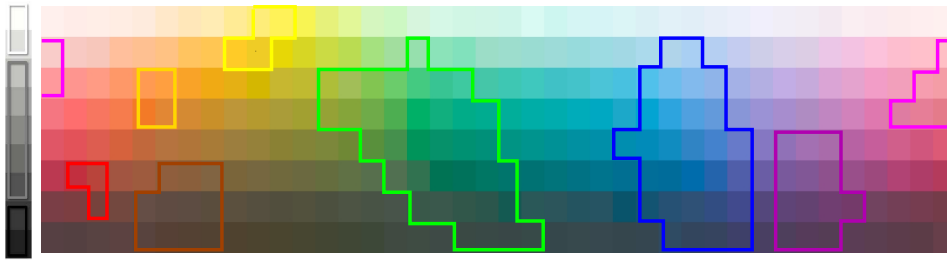   - Stage II: A term for red is added.

Figure 2.2   *Categorization of the Munsell colour array obtained in their experiment by Sturges and Whitfield for English.*

- Stage III: A term for either green or yellow (but not both) is added.
- Stage IV: A term for green or yellow (the one that was not added in the previous stage) is added.
- Stage V: A term for blue is added.
- Stage VI: A term for brown is added.
- Stage VII: Terms for pink, purple, orange and grey are added (in any order).

These sequence can be summarized with the expression:

$$[white, black] < [red] < [green, yellow] < [blue] < [brown] < [pink, purple, orange, grey]$$

where symbol '$<$' indicates temporal precedence, that is, for two categories A and B, $A < B$ means that A is present in the language before B, and order between terms inside '[ ]' depends on each language.

3. *Correlation with cultural development.* They noticed a high correlation of color vocabulary of a language with technological and cultural evolution. Languages from developed cultures were all in the last stage of color terms evolution, while languages from isolated and low-developed cultures were at lower stages of color vocabulary evolution.

Figure 2.1 shows the boundaries in the Munsell space of the eleven basic color categories for English that were obtained by Berlin and Kay. This categorization of the Munsell array has been used as a reference in later color-naming studies that confirmed Berlin and Kay results [9, 42]. Boundaries on the Munsell array obtained by Sturges and Whitfield in their experiment [42] are shown in figure 2.2.

Although the findings of Berlin and Kay were widely accepted for years, the debate about the existence of universals in color naming was reopened by Roberson et al. [39, 38] in 2000. On their works, Roberson and her colleagues defended that color categories were determined by language and, thus, boundaries between color categories were arbitrarily set by each language.

Despite additional evidences for the universalist theory were presented [35], an intermediate position between universalists and relativists has gained support in the last years

[34]. According to this new theory, the organization of the color naming systems in different languages is ruled by universal tendencies and, therefore, color categories are not arbitrarily located in the color space. On the other hand, the extension and the boundaries of the categories are more language dependent and more arbitrarily set by each language [36]. The explanation to this point could be found on the shape of the color space, where some areas seem to be more salient than others. Philipona et al. [33] showed that some properties of the human visual system could support the existence of such relevant parts in the color space. Moreover they showed that these areas coincide with the location of the focal colors for red, green, blue and yellow, found by Berlin and Kay . Hence, these salient parts of the color space would condition the formation of the color categories to reach a final configuration in each language that tends to be optimal in terms of information representation [37].

In this chapter we will discuss two approaches to computational color naming, which starting from a set of basic color terms, will learn models to predict color names in images. The experiments will be based on English language color terms. Although the debate about the universality of color categories is out of the scope of this book, this issue may have some influence on how the proposed models can be applied in images. If universalists are right these computational models are equally valid in other language. However, as we saw it is more probably that new models should be relearned for each language, especially to correctly place the boundaries between color names.

2.2   COLOR NAMES FROM CALIBRATED DATA

In this section we present a fuzzy color-naming model based on the use of parametric membership functions whose advantages are discussed later in this section.

Fuzzy set theory is an useful tool to model human decisions related to perceptual processes and language. A theoretical fuzzy framework was already proposed by Kay and McDaniel [19] in 1978. The basis of such a framework is to consider that any color stimulus has a membership value between 0 and 1 to each color category.

The model defined in this framework will be fitted to data derived from psychophysical experiments. These experiments are usually carried out in controlled environments under stable viewing conditions, with a known illuminant and by using samples with precisely measured reflectances. In such an environment, the naming judgements obtained from the subjects in the experiments are isolated from most other perceptual processes. By fitting a parametric model to data from psychophysical experiments we provide a set of tuneable parameters that analytically define the shape of the fuzzy sets representing color categories.

Parametric models have been previously used to model color information [1], and the suitability of such an approach can be summed up in the following points:

Inclusion of prior knowledge: Prior knowledge about the structure of the data allows choosing the best model on each case. However, this could turn into a disadvantage if a non appropriate function for the model is selected.

Compact categories: Each category is completely defined by few parameters and training data do not need to be stored after an initial fitting process. This implies lower memory usage and lower computation when the model is applied.

Meaningful parameters: Each parameter has a meaning in terms of the characterization of the data, which allows modifying and improving the model by just adjusting the parameters.

Easy Analysis: As a consequence of the previous one, the model can be analyzed and compared by studying the values of its parameters.

We have worked on the CIE $L^*a^*b^*$ color space since it is a quasi-perceptually-uniform color space where a good correlation between Euclidean distance between color pairs and perceived color dissimilarity can be observed. It is likely that other spaces could be suitable whenever one of the dimensions correlates with color lightness and the other two with chromaticity components. In this section, we will denote any color point in such space as $\mathbf{s} = (I, c_1, c_2)$, where I is the lightness and c_1 and c_2 are the chromaticity components of the color point.

Ideally, color memberships should be modelled by three-dimensional functions, i.e. functions defined $\mathbb{R}^3 \rightarrow [0, 1]$, but unfortunately it is not easy to infer precisely the way in which color naming data is distributed in the color space and hence, finding parametric functions that fit these data is a very complicated task. For this reason, in this proposal the three-dimensional color space has been sliced in a set of N_L levels along the lightness axis (see figure 2.3), obtaining a set of chromaticity planes over which membership functions have been modelled by two-dimensional functions. Therefore, any specific chromatic category will be defined by a set of functions, each one depending on a lightness component, as it is expressed later in eq. 2.12. Achromatic categories (black, grey and white) will be given as the complementary function of the chromatic ones but weighted by the membership function of each one of the three achromatic categories. To go into the details of the proposed approach we will first give the basis of the fuzzy framework and afterwards we will pose the considerations on the function shapes for the chromatic categories. Finally, the complementary achromatic categories will be derived.
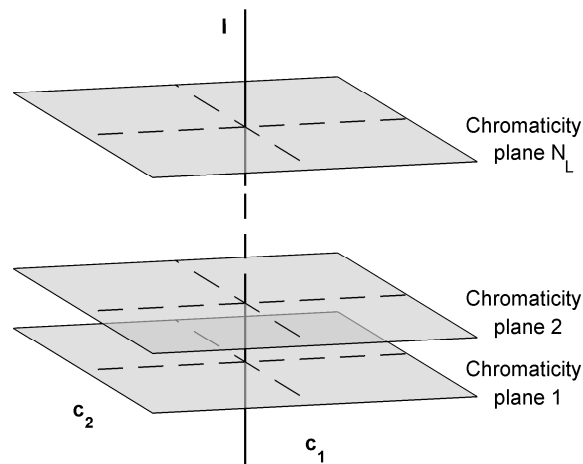

Figure 2.3   Scheme of the model. The color space is divided in N_L levels along the lightness axis.

2.2.1 Fuzzy Color Naming

A fuzzy set is a set whose elements have a degree of membership [50]. In a more formal way, a fuzzy set A is defined by a crisp set X, called the universal set, and a membership function, μ_A, which maps elements of the universal set into the $[0, 1]$ interval, that is, $\mu_A : X \rightarrow [0, 1]$.

Fuzzy sets are a good tool to represent imprecise concepts expressed in natural language. In color naming, we can consider that any color category, C_k, is a fuzzy set with a membership function, μ_{C_k}, which assigns, to any color sample $\mathbf{s}$ represented in a certain color space, i.e. the universal set, a membership value $\mu_{C_k}(\mathbf{s})$ within the $[0, 1]$ interval. This value represents the certainty we have that $\mathbf{s}$ belongs to category C_k, which is associated with the linguistic term t_k.

In the context of color categorization with a fixed number of categories, we need to impose the constraint that, for a given sample $\mathbf{s}$, the sum of its memberships to the n categories must be the unity

$$\sum_{k=1}^{n} \mu_{C_k}(\mathbf{s}) = 1 \qquad \text{with} \qquad \mu_{C_k}(\mathbf{s}) \in [0, 1], \quad k = 1, \ldots, n \qquad (2.1)$$

In the rest of the section, this constraint will be referred to as the unity-sum constraint. Although this constraint does not hold in fuzzy set theory, it is interesting in this case because it allows us to interpret the memberships of any sample as the contributions of the considered categories to the final color sensation.

Hence, for any given color sample $\mathbf{s}$ it will be possible to compute a color descriptor, $\mathcal{CD}$, such as

$$\mathcal{CD}(\mathbf{s}) = \left[\mu_{C_1}(\mathbf{s}), ..., \mu_{C_n}(\mathbf{s}) \right] \qquad (2.2)$$

where each component of this n-dimensional vector describes the membership of $\mathbf{s}$ to a specific color category.

The information contained in such a descriptor can be used by a decision function, $\mathcal{N}(\mathbf{s})$, to assign the color name of the stimulus $\mathbf{s}$. The most easy decision rule we can derive is to choose the maximum from $\mathcal{CD}(\mathbf{s})$:

$$\mathcal{N}(\mathbf{s}) = t_{k_{\max}} \quad | \quad k_{\max} = \arg \max_{k=1,\ldots,n} \{ \mu_{C_k}(\mathbf{s}) \} \qquad (2.3)$$

where t_k is the linguistic term associated with color category C_k.

In this case the categories considered are the basic categories proposed by Berlin and Kay, that is, $n = 11$ and the set of categories is:

$$C_k \in \{red, orange, brown, yellow, green, blue, purple, pink, black, grey, white\} \qquad (2.4)$$

2.2.2 Chromatic Categories

According to the fuzzy framework defined previously, any function we select to model color categories, must map values to the $[0, 1]$ interval, i.e. $\mu_{C_k}(\mathbf{s}) \in [0, 1]$. In addition, the observation of the membership values of psychophysical data obtained from a color naming experiment [6], made us hypothesize about a set of necessary properties that membership functions for the chromatic categories should fulfill:

- *Triangular basis*: Chromatic categories present a plateau, or area with no confusion about the color name, with a triangular shape and a principal vertex shared by all the categories.

- *Different slopes*: For a given chromatic category, the slope of naming-certainty toward the neighboring categories can be different on each side of the category (e.g. transition from blue to green can be different from that from blue to purple).

- *Central notch*: The transition from a chromatic category to the central achromatic one has the form of a notch around the principal vertex.

In figure 2.4 we show an scheme of the preceding conditions on a chromaticity diagram where the samples of a color-naming experiment have been plotted.
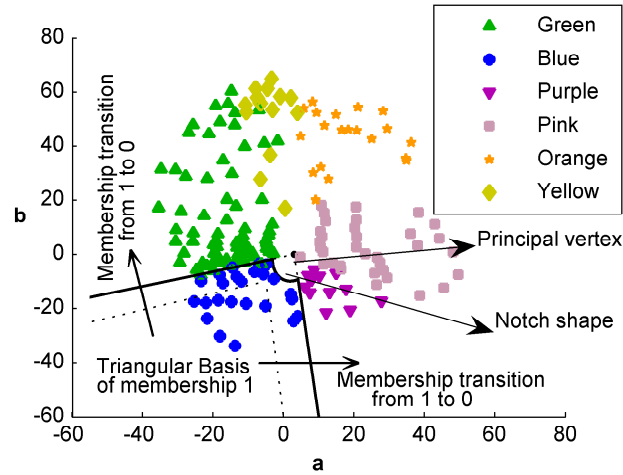

Figure 2.4   Desirable properties of the membership function for chromatic categories. In this case, on the blue category.

Here we will define the membership function, the triple sigmoid with elliptical center (TSE), as a two-dimensional function, $TSE : \mathbb{R}^2 \to [0, 1]$. Earlier works considered other membership functions [4, 3]. The definition of the TSE starts from the one-dimensional sigmoid function:

$$S^1(x, \beta) = \frac{1}{1 + \exp(-\beta x)} \tag{2.5}$$

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

This can be extended to a two-dimensional sigmoid function, $S : \mathbb{R}^2 \to [0, 1]$, as:

$$S(\mathbf{p}, \beta) = \frac{1}{1 + \exp(-\beta \mathbf{u_i p})}, \qquad i = 1, 2 \tag{2.6}$$

where $\mathbf{p} = (x, y)^T$ is a point in the plane, and vectors $\mathbf{u_1} = (1, 0)$ and $\mathbf{u_2} = (0, 1)$ define the axis in which the function is oriented.

By adding a translation, $\mathbf{t} = (t_x, t_y)$, and a rotation, α, to the previous equation, the function can adopt a wide set of shapes. In order to represent the formulation in a compact

matrix form, we will use homogeneous coordinates [15]. Let us redefine $\mathbf{p}$ to be a point in the plane expressed in homogeneous coordinates as $\mathbf{p} = (x, y, 1)^T$, and let us denote the vectors $\mathbf{u_1} = (1, 0, 0)$ and $\mathbf{u_2} = (0, 1, 0)$. We define S_1 as a function oriented in axis x with rotation α with respect to axis y and S_2 as a function oriented in axis y with rotation α with respect to axis x:

$$S_i(\mathbf{p}, \mathbf{t}, \alpha, \beta) = \frac{1}{1 + \exp(-\beta \mathbf{u_i} R_\alpha T_t \mathbf{p})}, \qquad i = 1, 2 \tag{2.7}$$

where T_t and R_α are a translation matrix and a rotation matrix, respectively:

$$T_t = \begin{pmatrix} 1 & 0 & -t_x \\ 0 & 1 & -t_y \\ 0 & 0 & 1 \end{pmatrix} \qquad R_\alpha = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) & 0 \\ -\sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.8}$$

By multiplying S_1 and S_2, we define the double-sigmoid (DS) function, which fulfills the first two properties proposed before:

$$DS(\mathbf{p}, \mathbf{t}, \theta_{DS}) = S_1(\mathbf{p}, \mathbf{t}, \alpha_y, \beta_y) S_2(\mathbf{p}, \mathbf{t}, \alpha_x, \beta_x) \tag{2.9}$$

where $\theta_{DS} = (\alpha_x, \alpha_y, \beta_x, \beta_y)$ is the set of parameters of the double-sigmoid function. Figure 2.5(a) shows a plot of a two-dimensional sigmoid oriented in the x-axis direction (S_1 in eq. 2.7). By multiplying two oriented sigmoid functions, the double-sigmoid function DS is obtained (figure 2.5(b)).
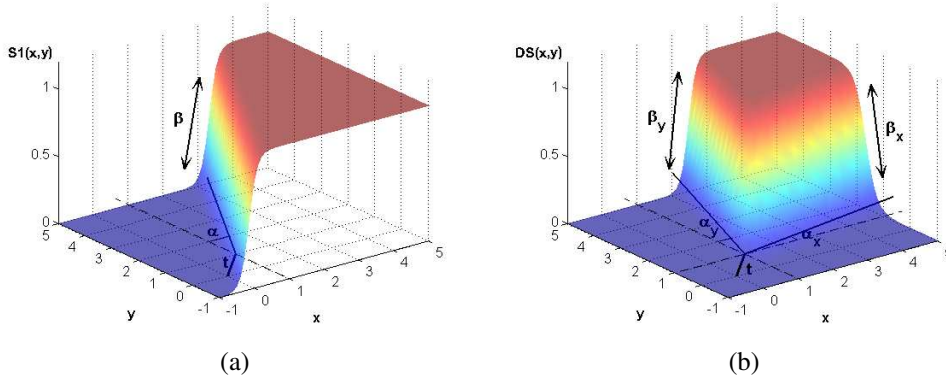

Figure 2.5   Two-dimensional sigmoid functions.   (a) S_1: Sigmoid function oriented in axis x direction (b) DS: The product of two differently oriented sigmoid functions generates a plateau with some of the properties needed for the membership function.

To obtain the central notch shape needed to fulfill the third proposed property, let us define the elliptic-sigmoid (ES) function by including the ellipse equation in the sigmoid formula:

$$ES(\mathbf{p}, \mathbf{t}, \theta_{ES}) = \frac{1}{1 + \exp\left\{ -\beta_e \left[\left(\frac{\mathbf{u_1} R_\phi T_t \mathbf{p}}{e_x} \right)^2 + \left(\frac{\mathbf{u_2} R_\phi T_t \mathbf{p}}{e_y} \right)^2 - 1 \right] \right\}} \tag{2.10}$$

where $\theta_{ES} = (e_x, e_y, \phi, \beta_e)$ is the set of parameters of the elliptic-sigmoid function, e_x and e_y are the semiminor and semimajor axis, respectively, ϕ is the rotation angle of

the ellipse, and β_e is the slope of the sigmoid curve that forms the ellipse boundary. The function obtained is an elliptic plateau if β_e is negative and an elliptic valley if β_e is positive. The surfaces obtained can be seen in figure 2.6.
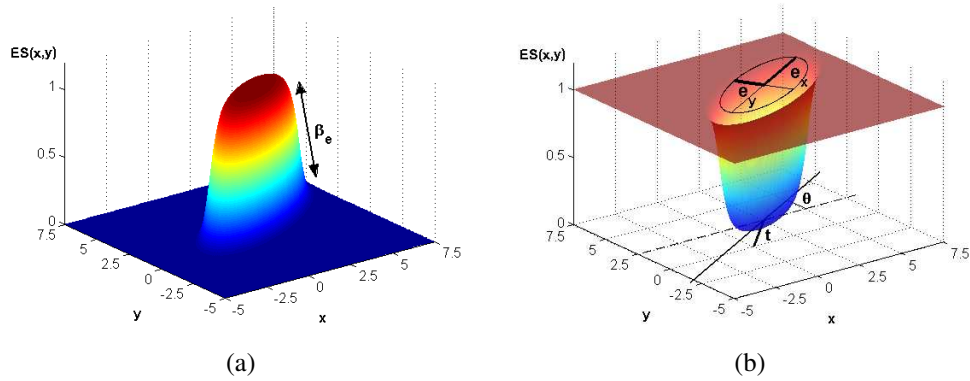

| (a) | (b) |

Figure 2.6   Elliptic-sigmoid function $ES(\mathbf{p}, \mathbf{t}, \theta_{ES})$. a) ES for $\beta_e < 0$ b) ES for $\beta_e > 0$

Finally, by multiplying the double-sigmoid by the elliptic-sigmoid with a positive β_e, we define the triple sigmoid with Elliptical center (TSE) as:

$$TSE(\mathbf{p}, \theta) = DS(\mathbf{p}, \mathbf{t}, \theta_{DS}) ES(\mathbf{p}, \mathbf{t}, \theta_{ES}) \qquad (2.11)$$

where $\theta = (\mathbf{t}, \theta_{DS}, \theta_{ES})$ is the set of parameters of the TSE.

The TSE function defines a membership surface that fulfills the properties defined at the beginning of Section 2.2.2. Figure 2.7 shows the form of the TSE function.
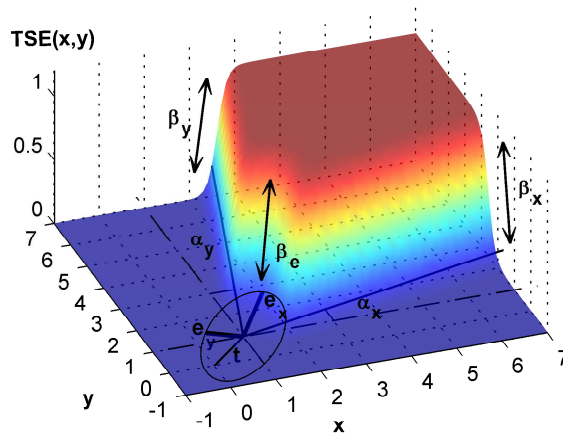

Figure 2.7   Triple sigmoid with Elliptical center (TSE).

Hence, once we have the analytic form of the chosen function, the membership function for a chromatic category μ_{C_k} is given by:

$$\mu_{C_k}(\mathbf{s}) = \begin{cases} \mu^1_{C_k} = TSE(c_1, c_2, \theta^1_{C_k}) & \text{if } I \leq I_1, \\ \mu^2_{C_k} = TSE(c_1, c_2, \theta^2_{C_k}) & \text{if } I_1 < I \leq I_2, \\ \vdots & \vdots \\ \mu^{N_L}_{C_k} = TSE(c_1, c_2, \theta^{N_L}_{C_k}) & \text{if } I_{N_L-1} < I, \end{cases} \tag{2.12}$$

where $\mathbf{s} = (I, c_1, c_2)$ is a sample on the color space, N_L is the number of chromaticity planes, $\theta^i_{C_k}$ is the set of parameters of the category C_k on the ith chromaticity plane, and I_i are the lightness values that divide the space in the N_L lightness levels.

By fitting the parameters of the functions, it is possible to obtain the variation of the chromatic categories through the lightness levels. By doing this for all the categories, it will be possible to obtain membership maps; that is, for a given lightness level we have a membership value to each category for any color point $\mathbf{s} = (I, c_1, c_2)$ of the level. Notice that since some categories exist only at certain lightness levels (e.g. brown is defined only for low lightness values and yellow only for high values), on each lightness level not all the categories will have memberships different from zero for any point of the level. Figure 2.8 shows an example of the membership map provided by the TSE functions for a given lightness level, in which there exist six chromatic categories. The other two chromatic categories in this example would have zero membership for any point of the level.
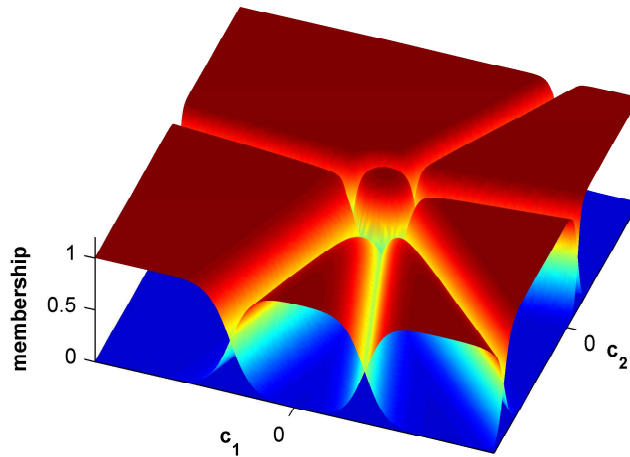
Figure 2.8   TSE function fitted to the chromatic categories defined on a given lightness level. In this case, only six categories have memberships different from zero.

2.2.3 Achromatic Categories

The three achromatic categories (black, grey and white) are first considered as a unique category at each chromaticity plane. To ensure that the unity-sum constraint is fulfilled (i.e. the sum of all memberships must be one) a global achromatic membership, μ_A, is

computed for each level as:

$$\mu_A^i(c_1, c_2) = 1 - \sum_{k=1}^{n_c} \mu_{C_k}^i(c_1, c_2) \tag{2.13}$$

where i is the chromaticity plane that contains the sample $\mathbf{s} = (I, c_1, c_2)$ and n_c is the number of chromatic categories (here, $n_c = 8$). The differentiation among the three achromatic categories must be done in terms of lightness. To model the fuzzy boundaries among these three categories we use one-dimensional sigmoid functions along the lightness axis:

$$\mu_{A_{black}}(I, \theta_{black}) = \frac{1}{1 + \exp[-\beta_b(I - t_b)]} \tag{2.14}$$

$$\mu_{A_{grey}}(I, \theta_{grey}) = \frac{1}{1 + \exp[\beta_b(I - t_b)]} \frac{1}{1 + \exp[-\beta_w(I - t_w)]} \tag{2.15}$$

$$\mu_{A_{white}}(I, \theta_{white}) = \frac{1}{1 + \exp[\beta_w(I - t_w)]} \tag{2.16}$$

where $\theta_{black} = (t_b, \beta_b)$, $\theta_{grey} = (t_b, \beta_b, t_w, \beta_w)$, and $\theta_{white} = (t_w, \beta_w)$ are the set of parameters for *black*, *grey*, and *white*, respectively.

Hence, the membership of the three achromatic categories on a given chromaticity plane is computed by weighting the global achromatic membership (see eq. 2.13) with the corresponding membership in the lightness dimension (see eq. 2.14, and eq. 2.16):

$$\mu_{C_k}(\mathbf{s}, \theta_{C_k}) = \mu_A^i(c_1, c_2)\mu_{A_{C_k}}(I, \theta_{C_k}), \qquad 9 \leq k \leq 11, \quad I_i < I \leq I_{i+1} \tag{2.17}$$

where i is the chromaticity plane in which the sample is included and the values of k correspond to the achromatic categories (see eq. 2.4). In this way we can assure that the unity-sum constraint is fulfilled on each specific chromaticity plane,

$$\sum_{k=1}^{11} \mu_{C_k^i}(\mathbf{s}) = 1 \qquad i = 1, \ldots, N_L \tag{2.18}$$

where N_L is the number of chromaticity planes in the model.

2.2.4 Fuzzy Sets Estimation

Once we have defined the membership functions of the model, the next step is to fit their parameters. To this end, we need a set of psychophysical data, D, composed of a set of samples from the color space and their membership values to the eleven categories,

$$D = \{< \mathbf{s_i}, m_1^i, ..., m_{11}^i >\}, \qquad i = 1, \ldots, n_s \tag{2.19}$$

where $\mathbf{s_i}$ is the ith sample of the learning set, n_s is the number of samples in the learning set, and m_k^i is the membership value of the ith sample to the kth category.

Such data will be the knowledge basis for a fitting process to estimate the model parameters taking into account the unity-sum constraint given in eq. 2.18. In this case, the model will be estimated for the CIE $L^*a^*b^*$ space since it is a standard space with interesting properties. However, any other color space with a lightness dimension and two chromatic dimensions would be suitable for this purpose.

Learning Set: The data set for the fitting process must be perceptually significant; that is, the judgements should be coherent with results from psychophysical color-naming experiments and the samples should cover all the color space.

To build a wide learning set, we have used the color-naming map proposed by Seaborn et al. in [40]. This color map has been built by making some considerations on the consensus areas of the Munsell color space provided by the psychophysical data from the experiments of Sturges and Whitfield [42]. Using such data and the fuzzy k-means algorithm this method allows us to derive the memberships of any point in the Munsell space to the eleven basic color categories.

In this way, we have obtained the memberships of a wide sample set, and afterwards we have converted this color sampling set to their corresponding CIE $L^*a^*b^*$ representation. The data set was initially composed of the 1269 samples of the Munsell Book of Color [29]. Their reflectances and CIE $L^*a^*b^*$ coordinates, calculated by using the CIE D65 illuminant, are available at the web site of the University of Joensuu in Finland [44]. This dataset was extended with selected samples to a total number of 1617 samples (see [5] for more details on how these extra samples were selected).

Hence, with such a data set we accomplish the perceptual significance required for the learning set. First, by using Seaborn's method, we include the results of the psychophysical experiment of Sturges and Whitfield, and, in addition, it covers an area of the color space that suffices for the fitting process.

Parameter Estimation: Before starting with the fitting process, the number of chromaticity planes and the values that define the lightness levels (see eq. 2.12) must be set. These values depend on the learning set used and must be chosen while taking into account the distribution of the samples from the learning set. In this case, the number of planes that delivered best results was found to be six, and the values I_i that define the levels were selected by choosing some local minima in the histogram of samples along the lightness axis: $I_1 = 31, I_2 = 41, I_3 = 51, I_4 = 66, I_5 = 76$. However, if a more extensive learning set were available, a higher number of levels would possibly deliver better results.

For each chromaticity plane, the global goal of the fitting process is finding an estimation of the parameters, $\hat{\theta}^j$, that minimizes the mean squared error between the memberships from the learning set and the values provided by the model:

$$\hat{\theta}^j = \arg \min_{\theta^j} \frac{1}{n_{cp}} \sum_{i=1}^{n_{cp}} \sum_{k=1}^{n_c} (\mu_{C_k}^j(\mathbf{s_i}, \theta_{C_k}^j) - m_k^i)^2, \qquad j = 1, \cdots, N_L \qquad (2.20)$$

where $\hat{\theta}^j = (\hat{\theta}_{C_1}^j, ..., \hat{\theta}_{C_{n_c}}^j)$ is the estimation of the parameters of the model for the chromatic categories on the jth chromaticity plane, $\theta_{C_k}^j$ is the set of parameters of the category C_k for the jth chromaticity plane, n_c is the number of chromatic categories, n_{cp} is the number of samples of the chromaticity plane, $\mu_{C_k}^j$ is the membership function of the color category C_k for the jth chromaticity plane, and m_k^i is the membership value of the ith sample of the learning set to the kth category.

The previous minimization is subject to the unity-sum constraint:

$$\sum_{k=1}^{11} \mu_{C_k}^j(\mathbf{s}, \theta_{C_k}^j) = 1, \qquad \forall \mathbf{s} = (I, c_1, c_2) \quad | \quad I_{j-1} < I \le I_j \qquad (2.21)$$

which is imposed to the fitting process through two assumptions. The first one is related to the membership transition from chromatic categories to achromatic categories:

Assumption 1: All the chromatic categories in a chromaticity plane share the same elliptical-sigmoid function, which models the membership transition to the achromatic categories. This means that all the chromatic categories share the set of estimated parameters for ES:

$$\theta^j_{ES_{C_p}} = \theta^j_{ES_{C_q}} \quad and \quad \mathbf{t^j_{C_p}} = \mathbf{t^j_{C_q}}, \qquad \forall p, q \in \{1, \ldots, n_c\} \tag{2.22}$$

where n_c is the number of chromatic categories.

The second assumption refers to the membership transition between adjacent chromatic categories:

Assumption 2: Each pair of neighboring categories, C_p and C_q, share the parameters of slope and angle of the double-sigmoid function, which define their boundary:

$$\beta^{C_p}_y = \beta^{C_q}_x \qquad and \qquad \alpha^{C_p}_y = \alpha^{C_q}_x - (\frac{\pi}{2}) \tag{2.23}$$

where the superscripts indicate the category to which the parameters correspond.

These assumptions considerably reduce the number of parameters to be estimated. Hence, for each chromaticity plane, we must estimate 2 parameters for the translation, $\mathbf{t} = (t_x, t_y)$, 4 for the ES function, $\theta_{ES} = (e_x, e_y, \phi, \beta_e)$, and a maximum of $2 \times n_c$ for the DS functions, since the other two parameters of $\theta_{DS} = (\alpha_x, \alpha_y, \beta_x, \beta_y)$ can be obtained from the neighboring category (eq. 2.23).

All the minimizations to estimate the parameters are performed by using the simplex search method proposed in [20] (see [5] for more details about the parameters estimation process). After the fitting process, we obtain the parameters that completely define the color-naming model and are summarized in table 2.1.

The evaluation of the fitting process is done in terms of two measures. The first one is the mean absolute error (MAE_{fit}) between the learning set memberships and the memberships obtained from the parametric membership functions:

$$MAE_{fit} = \frac{1}{n_s} \frac{1}{11} \sum_{i=1}^{n_s} \sum_{k=1}^{11} |m^i_k - \mu_{C_k}(\mathbf{s_i})| \tag{2.24}$$

where n_s is the number of samples in the learning set, m^i_k is the membership of $\mathbf{s_i}$ to the kth category, and $\mu_{C_k}(\mathbf{s_i})$ is the parametric membership of $\mathbf{s_i}$ to the kth category provided by the model.

The value of MAE_{fit} is a measure of the accuracy of the model fitting to the learning dataset, and in this case the value obtained was of $MAE_{fit} = 0.0168$. This measure was also computed for a test dataset of 3149 samples. To build the test dataset, the Munsell space was sampled at hues 1.25, 3.75, 6.25 and 8.75; values from 2.5 to 9.5 at steps of 1 unit; and chromas from 1 to the maximum available with a step of 2 units. As in the case of the learning set, the memberships of the test set that were considered the ground truth, were computed with Seaborn's algorithm. The corresponding CIE $L^*a^*b^*$ values to apply the parametric functions were computed with the Munsell Conversion software. The value of MAE_{fit} obtained was 0.0218, which confirms the accuracy of the fitting that allows the model to provide membership values with very low error even for samples that were not used in the fitting process.

Table 2.1 Parameters of the triple sigmoid with elliptical center model[a]

Achromatic axis

Black-Grey boundary	$t_b = 28.28$	$\beta_b = -0.71$
Grey-White boundary	$t_w = 79.65$	$\beta_w = -0.31$

Chromaticity plane 1

$t_a = 0.42$		$e_a = 5.89$		$\beta_e = 9.84$
$t_b = 0.25$		$e_b = 7.47$		$\phi = 2.32$
	α_a	α_b	β_a	β_b
Red	-2.24	-56.55	0.90	1.72
Brown	33.45	14.56	1.72	0.84
Green	104.56	134.59	0.84	1.95
Blue	224.59	-147.15	1.95	1.01
Purple	-57.15	-92.24	1.01	0.90

Chromaticity plane 2

$t_a = 0.23$		$e_a = 6.46$		$\beta_e = 6.03$
$t_b = 0.66$		$e_b = 7.87$		$\phi = 17.59$
	α_a	α_b	β_a	β_b
Red	2.21	-48.81	0.52	5.00
Brown	41.19	6.87	5.00	0.69
Green	96.87	120.46	0.69	0.96
Blue	210.46	-148.48	0.96	0.92
Purple	-58.48	-105.72	0.92	1.10
Pink	-15.72	-87.79	1.10	0.52

Chromaticity plane 3

$t_a = -0.12$		$e_a = 5.38$		$\beta_e = 6.81$
$t_b = 0.52$		$e_b = 6.98$		$\phi = 19.58$
	α_a	α_b	β_a	β_b
Red	13.57	-45.55	1.00	0.57
Orange	44.45	-28.76	0.57	0.52
Brown	61.24	6.65	0.52	0.84
Green	96.65	109.38	0.84	0.60
Blue	199.38	-148.24	0.60	0.80
Purple	-58.24	-112.63	0.80	0.62
Pink	-22.63	-76.43	0.62	1.00

Chromaticity plane 4

$t_a = -0.47$		$e_a = 5.99$		$\beta_e = 7.76$
$t_b = 1.02$		$e_b = 7.51$		$\phi = 23.92$
	α_a	α_b	β_a	β_b
Red	26.70	-56.88	0.91	0.76
Orange	33.12	-9.90	0.76	0.48
Yellow	80.10	5.63	0.48	0.73
Green	95.63	108.14	0.73	0.64
Blue	198.14	-148.59	0.64	0.76
Purple	-58.59	-123.68	0.76	5.00
Pink	-33.68	-63.30	5.00	0.91

Chromaticity plane 5

$t_a = -0.57$		$e_a = 5.37$		$\beta_e = 100.00$
$t_b = 1.16$		$e_b = 6.90$		$\phi = 24.75$
	α_a	α_b	β_a	β_b
Orange	25.75	-15.85	2.00	0.84
Yellow	74.15	12.27	0.84	0.86
Green	102.27	98.57	0.86	0.74
Blue	188.57	-150.83	0.74	0.47
Purple	-60.83	-122.55	0.47	1.74
Pink	-32.55	-64.25	1.74	2.00

Chromaticity plane 6

$t_a = -1.26$		$e_a = 6.04$		$\beta_e = 100.00$
$t_b = 1.81$		$e_b = 7.39$		$\phi = -1.19$
	α_a	α_b	β_a	β_b
Orange	25.74	-17.56	1.03	0.79
Yellow	72.44	16.24	0.79	0.96
Green	106.24	100.05	0.96	0.90
Blue	190.05	-149.43	0.90	0.60
Purple	-59.43	-122.37	0.60	1.93
Pink	-32.37	-64.26	1.93	1.03

[a] Angles are expressed in degrees and subscripts x and y are changed to a and b, respectively, in order to make parameter interpretation easier, since parameters have been estimated for the CIE $L^*a^*b^*$ space.

The second measure evaluates the degree of fulfillment of the unity-sum constraint. Considering as error the difference between the unity and the sum of all the memberships at a point, $\mathbf{p_i}$, the measure proposed is:

$$MAE_{unitsum} = \frac{1}{n_p} \sum_{i=1}^{n_p} |1 - \sum_{k=1}^{11} \mu_{C_k}(\mathbf{p_i}))| \qquad (2.25)$$

where n_p is the number of points considered and μ_{C_k} is the membership function of category C_k.

To compute this measure, we have sampled each one of the six chromaticity planes with values from -80 to 80 at steps of 0.5 units on both a and b axis, which means that $n_p = 153600$. The value obtained of $MAE_{unitsum} = 6.41e - 04$ indicates that the model provides a great fulfillment of that constraint making the model consistent with the proposed framework.

Hence, for any point of the CIE $L^*a^*b^*$ space we can compute the membership to all the categories and, at each chromaticity plane, these values can be plotted to generate a membership map. In figure 2.9 we show the membership maps of the six chromaticity planes considered with the membership surfaces labelled with their corresponding color term.

2.3 COLOR NAMES FROM UNCALIBRATED DATA

In the previous section, we saw an example where the mapping between RGB and color names is inferred from a labeled set of color patches. Other examples of such methods include [10, 16, 22, 28, 27]. In such methods, multiple test subjects are asked to label hundreds of color chips within a well-defined experimental setup. From this labeled set of color chips the mapping from RGB values to color names is derived. The main difference of these methods with the one discussed in this section is that they are all based on calibrated data. Color names from calibrated data have been shown to be useful within the linguistic and color science fields. However, when applied to real-world images, these methods were often found to obtain unsatisfactory results. Color naming chips under ideal lighting on a color neutral background greatly differs from the challenge of color naming in images coming from real-world applications without a neutral reference color and with physical variations such as shading effects and different light sources.

In this section, we discuss a method for color naming in uncalibrated images. More precisely, with *uncalibrated images* we refer to images which are taken under varying illuminants, with interreflections, coming from unknown cameras, colored shadows, compression artifacts, aberrations in acquisition, unknown camera and camera settings, etc. The majority of the image data in computer vision belongs to this category: even in the cases that camera information is available and the images are uncompressed, the physical setting of the acquisition are often difficult to recover, due to unknown illuminant colors, unidentified shadows, view-point changes, and interreflections.

To infer what RGB values color names take on in real-world images, a large data set of color name labeled images is required. One possible way to obtain such a data set is by means of Google Image search. We retrieve 250 images for each of the eleven basic color terms discussed in section 2.1 (see figure 2.10). These images contain a large variety of appearances of the queried color name. E.g. the query "red" will contain images with red
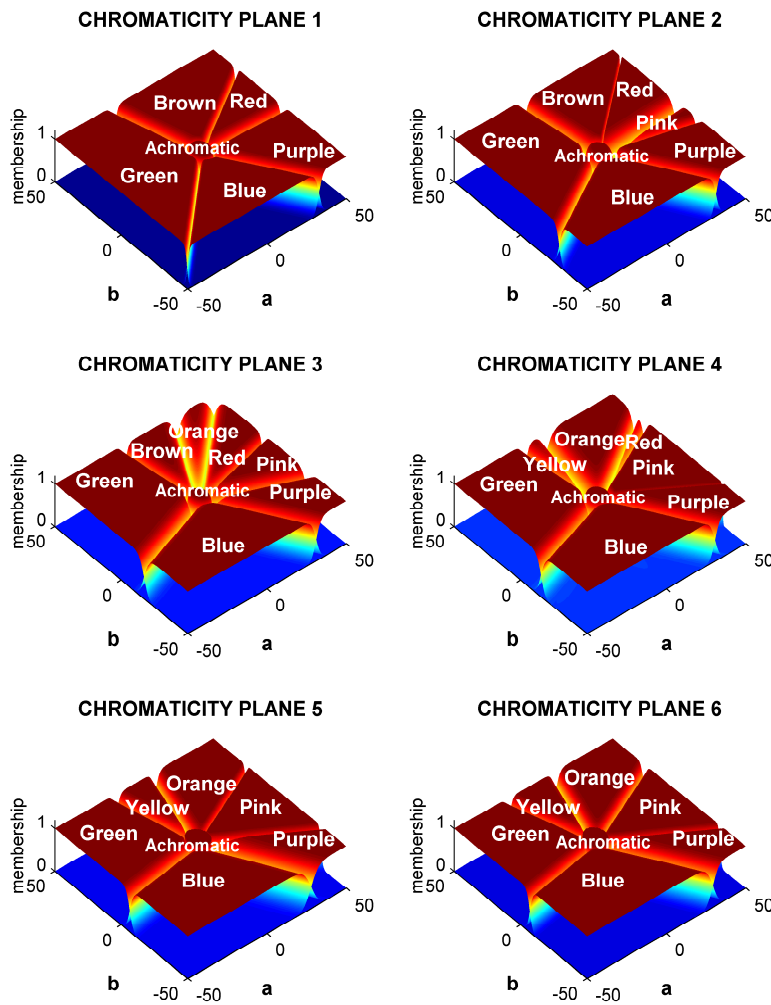
Figure 2.9 Membership maps for the six chromaticity planes of the model.

objects, taken under varying physical variations, such as different illuminants, shadows, and specularities. The images are taken with different cameras and stored with various compression methods. The large variety of this training set suits our goal of learning color names for real-world images well, since we want to apply the color naming method on uncalibrated images taken under varying physical settings. Furthermore, a system based on Google image has the advantage that it is flexible with respect to variations in the color name set. Methods based on calibrated data are known to be inflexible with respect to the

Figure 2.10   *Google-retrieved examples for color names. The red bounding boxes indicate false positives. An image can be retrieved with various color names, such as the flower image which appears in the red and the yellow set.*

set of color names, since adding for example new color names such as beige, violet or olive, would in principle imply to redo the human labeling for all patches.

Retrieved images from Google search are known to contain many false positives. To learn color names from such a noisy dataset, we will discuss a method based on Probabilistic Latent Semantic Analysis (PLSA), a generative model introduced by Hofmann [18] for document analysis. In conclusion, by learning color names from real-world images, we aim to derive color names which are applicable on challenging real-world images typical for computer vision applications.

2.3.1   Color Name Data Sets

As discussed Google Image is used to retrieve 250 images for each of the eleven color names. For the actual search we added the term "color", hence for red the query is "red+color". Examples for the eleven color names are given in figure 2.10. Almost 20 % of the images are false positives, i.e., images which do not contain the color of the query. We call such a data set *weakly labeled* since the image labels are global, meaning that no information to which particular region of the image the label refers is available. Furthermore, in many cases only a small portion, as little as a few percent of the pixels, represents the color label. Our goal is to learn a color naming system based on the raw results of Google image, i.e., we used both true and false positives.

The Google data set contains weakly labeled data, meaning that we only have a image-wide label indicating that a part of the pixels in the image can be described by the color name of the label. To remove some of the pixels which are not likely indicated by the image label, we remove the background from Google images by iteratively removing pixels which have the same color as the border. Furthermore, since the color label often refers to an object in the center of the image, we crop the image to be 70% of its original width and height.

The Google images will be represented by color histograms. We consider the images from the Google datasets to be in $sRGB$ format. Before computing the color histograms these images are gamma corrected with a correction factor of 2.4. Although images might not be correctly white balanced, we do not applying a color constancy algorithm, since color constancy was found to yield unsatisfying results for these images. Furthermore,

many Google images lack color calibration information, and regularly break assumptions on which color constancy algorithms are based. The images are converted to the $L^*a^*b^*$ color space, which is a perceptually linear color space, ensuring that similar differences between $L^*a^*b^*$ values are considered about equally important color changes to humans. This is a desired property because the uniform binning we apply for histogram construction implicitly assumes a meaningful distance measure. To compute the $L^*a^*b^*$ values we assume a D65 white light source.

2.3.2   Learning Color Names

Here we will discuss a method to learn color names, which is based on latent aspect models. Latent aspect models have received considerable interest in the text analysis community as a tool to model documents as a mixture of several semantic –but a-priori unknown, and hence "latent"– topics. Latent Dirichlet allocation (LDA) [8] and probabilistic latent semantic analysis (PLSA) [18] are perhaps the most well known among such models.

Here we use the topics to represent the color names of pixels. Latent aspect models are of interest to our problem since they naturally allow for multiple topics in the same image, as is the case in the Google data set where each image contains a number of colors. Pixels are represented by discretizing their $L^*a^*b^*$ values into a finite vocabulary by assigning each value by cubic interpolation to a regular $10 \times 20 \times 20$ grid in the $L^*a^*b^*$-space. An image (document) is then represented by a histogram indicating how many pixels are assigned to each bin (word).

We start by explaining the standard PLSA model, after which an adapted version better suited to the problem of color naming is discussed. Given a set of documents $D = \{d_1, ..., d_N\}$ each described in a vocabulary $W = \{w_1, ..., w_M\}$, the words are taken to be generated by latent topics $Z = \{z_1, ..., z_K\}$. In the PLSA model the conditional probability of a word w in a document d is given by:

$$p(w|d) = \sum_{z \in Z} p(w|z)p(z|d). \tag{2.26}$$

Both distributions $p(z|d)$ and $p(w|z)$ are discrete multinomial distributions, and can be estimated with an EM algorithm [18] by maximizing the log-likelihood function

$$L = \sum_{d \in D} \sum_{w \in W} n(d, w) \log p(d, w) \tag{2.27}$$

where $p(d, w) = p(d) p(w|d)$, and $n(d, w)$ is the term frequency, containing the word occurrences for every document.

The method in eq. 2.26 is called a generative model, since it provides a model of how the observed data has been generated given hidden parameters (the latent topics). The aim is to find the latent topics which best explain the observed data. In the case of learning color names, we model the color values in an image as being generated by the color names (topics). For example, the color name red generates $L^*a^*b^*$ values according to $p(w|t = red)$. These word-topic distributions $p(w|t)$ are shared between all images. The

---

Because the $L^*a^*b^*$-space is perceptually uniform we discretize it into equal volume bins. Different quantization levels per channel are chosen because of the different ranges: the intensity axis ranges from 0 to 100, and the chromatic axes range from -100 to 100.
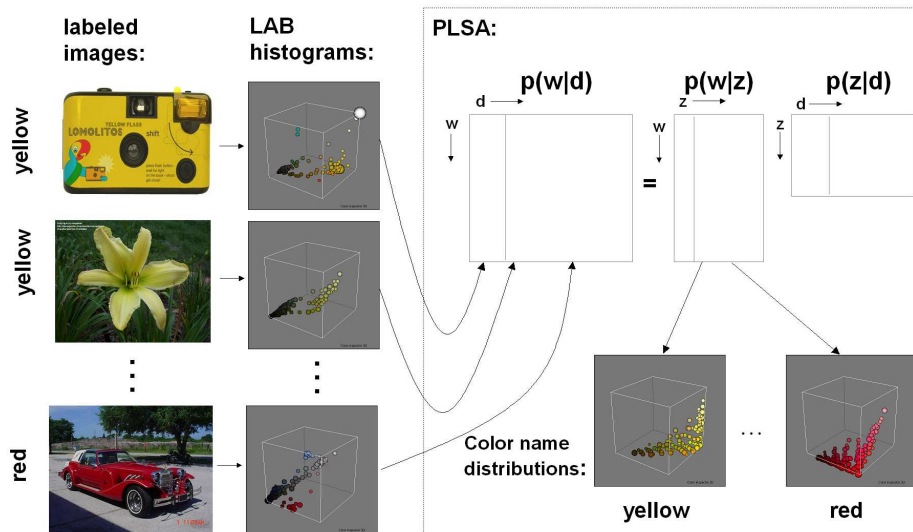
Figure 2.11   *Overview of standard PLSA model for learning color names. See text for explanation. Reprinted with permission, ©2009 IEEE.*

amount of the various colors we see in an image is given by the mixing coefficients $p(t|d)$, and these are image specific. The aim of the learning process is to find the $p(w|t)$ and $p(t|d)$ which best explain the observations $p(w|d)$. As a consequence, colors which often co-occur are more likely to be found in the same topic. E.g., the label red will co-occur with highly saturated reds, but also with some pinkish-red colors due to specularities on the red object, and dark reds caused by shadows or shading. All the different appearances of the color name red are captured in $p(w|t = red)$.

In figure 2.11 an overview of applying PLSA to the problem of color naming is provided. The goal of the system is to find the color name distributions $p(w|t)$. First, the weakly labeled Google images are represented by their normalized $L^*a^*b^*$ histograms. These histograms form the columns of the image specific word distribution $p(w|d)$. Next, the PLSA algorithm aims to find the topics (color names) which best explain the observed data. This process can be understood as a matrix decomposition of $p(w|d)$ into the word-topic distributions $p(w|t)$ and the document specific mixing proportions $p(t|d)$. The columns of $p(w|t)$ contain the information we are seeking, namely, the distributions of the color names over $L^*a^*b^*$ values. In the remainder of this section we discuss two adaptations to the standard model.

Exploiting image labels: the standard PLSA model cannot exploit the labels of images. More precisely, the labels have no influence on the maximum likelihood (eq. 2.27). The topics are hoped to converge to the state where they represent the desired color names. As is pointed out in [24] in the context of discovering object categories using LDA, this is rarely the case. To overcome this shortcoming we discuss an adapted model that does take into account the label information.

The image labels can be used to define a prior distribution on the frequency of topics (color names) in documents $p(z|d)$. This prior will still allow each color to be used in each image, but the topic corresponding to the label of the image—here obtained with Google—is a-priori assumed to have a higher frequency than other colors. Below, we use the shorthands $p(w|z) = \phi_z(w)$ and $p(z|d) = \theta_d(z)$.

The multinomial distribution $p(z|d)$ is supposed to have been generated from a Dirichlet distribution of parameter α_{l_d} where l_d is the label of the document d. The vector α_{l_d} has length K (number of topics), where $\alpha_{l_d}(z) = c \geq 1$ for $z = l_d$, and $\alpha_{l_d}(z) = 1$ otherwise. By varying c we control the influence of the image labels l_d on the distributions $p(z|d)$. The exact setting of c will be learned from the validation data.

For an image d with label l_d, the generative process thus reads:

1. Sample θ_d (distribution over topics) from the Dirichlet prior with parameter α_{l_d}.

2. For each pixel in the image

    (a) sample z (topic, color name) from a multinomial with parameter θ_d

    (b) sample w (word, pixel bin) from a multinomial with parameter ϕ_z

The distributions over words ϕ_z associated with the topics, together with the image specific distributions θ_d, have to be estimated from the training images. This estimation is done using an EM (*Expectation-Maximisation*) algorithm. In the Expectation step we evaluate for each word (color bin) w and document (image) d

$$p(z|w, d) \propto \theta_d(z)\phi_z(w). \tag{2.28}$$

During the Maximisation step, we use the result of the Expectation step together with the normalized word-document counts $n(d, w)$ (frequency of word w in document d) to compute the maximum likelihood estimates of ϕ_z and θ_d as

$$\phi_z(w) \propto \sum_d n(d, w)p(z|w, d), \tag{2.29}$$

$$\theta_d(z) \propto (\alpha_{l_d}(z) - 1) + \sum_w n(d, w)p(z|w, d). \tag{2.30}$$

Note that we obtain the EM algorithm for the standard PLSA model when $\alpha_{l_d}(z) = c = 1$, which corresponds to a uniform Dirichlet prior over θ_d.

Enforcing unimodality: the second adaptation of the PLSA model is based on prior knowledge of the probabilities $p(z|w)$. Consider the color name red: a particular region of the color space will have a high probability of red, moving away from this region in the direction of other color names will decrease the probability of red. Moving even further in this direction can only further decrease the probability of red. This is caused by the unimodal nature of the $p(z|w)$ distributions. Next, we discuss an adaptation of the PLSA model to enforce unimodality to the estimated $p(z|w)$ distributions.

It is possible to obtain a unimodal version of a function by means of greyscale reconstruction. The greyscale reconstruction of function p is obtained by iterating geodesic greyscale dilations of a marker m under p until stability is reached [49]. Consider the example given in figure 2.12. In the example, we consider two $1D$ topics $p_1 = p(z_1|w)$
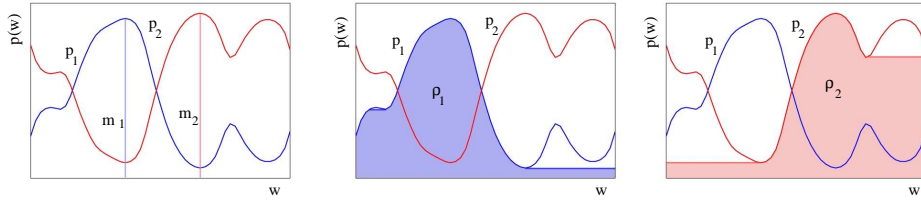
Figure 2.12   *Example of greyscale reconstruction. (left) Initial functions $p_1 = p(z_1|w)$, $p_2 = p(z_2|w)$, and markers m_1 and m_2. (middle) Greyscale reconstruction ρ_1 of p_1 from m_1. (right) Greyscale reconstruction ρ_2 of p_2 from m_2. Since ρ_1 is by definition a unimodal function, enforcing the difference between p_1 and ρ_1 to be small reduces the secondary modes of p_1. Reprinted with permission, ©2009 IEEE.*

and $p_2 = p(z_2|w)$. By iteratively applying a geodesic dilation from the marker m_1 under the mask function p_1 we obtain the greyscale reconstruction ρ_1. The function ρ_1 is by definition unimodal, since it only has one maximum at the position of the marker m_1. Similarly, we obtain a unimodal version of p_2 by a greyscale reconstruction of p_2 from marker m_2.

Something similar can be done for the color name distributions $p(z|w)$. We can compute a unimodal version $\rho_z^{m_z}(w)$ by performing a greyscale reconstruction of $p(z|w)$ from markers m_z (finding a suitable position for the markers will be explained below). To enforce unimodality, without assuming anything about the shape of the distribution, we add the difference between the distributions $p(z|w)$ and their unimodal counterparts $\rho_z^{m_z}(z)$ as a regularization factor to the log-likelihood function:

$$L = \sum_{d \in D} \sum_{w \in W} n(d, w) \log p(d, w) - \gamma \sum_{z \in Z} \sum_{w \in W} \left(p(z|w) - \rho_z^{m_z}(w) \right)^2, \qquad (2.31)$$

Adding the regularization factor in eq. 2.27 forces the functions $p(z|w)$ to be closer to $\rho_z^{m_z}(z)$. Since $\rho_z^{m_z}(z)$ is unimodal this will suppress the secondary modes in $p(z|w)$, i.e. the modes which it does not have in common with $\rho_z^{m_z}(z)$.

In the case of the color name distributions $p(z|w)$ the grey reconstruction is performed on the 3D spatial grid in $L^*a^*b^*$ space with a 26-connected structuring element. The markers m_z for each topic are computed by finding the local mode starting from the center of mass of the distribution $p(z|w)$. This was found to be more reliable than using the global mode of the distribution. The regularization functions $\rho_z^{m_z}$, which depend upon $p(z|w)$, are updated at every iteration step of the conjugate gradient based maximization procedure which is used to compute the maximum likelihood estimates of $\phi_z(w)$. The computation of the maximum likelihood estimate for $\theta_d(z)$ is not directly influenced by the regularization factor and is still computed with eq. 2.30.

In conclusion, two improvements of the standard PLSA model have been discussed. Firstly, the image labels are used to define a prior distribution on the frequency of topics. Secondly, a regularization factor is added to the log likelihood function which suppresses the secondary modes in the $p(z|w)$ distributions. The two parameters, c and γ, which regularize the strength of the two adaptations can be learned from validation data.
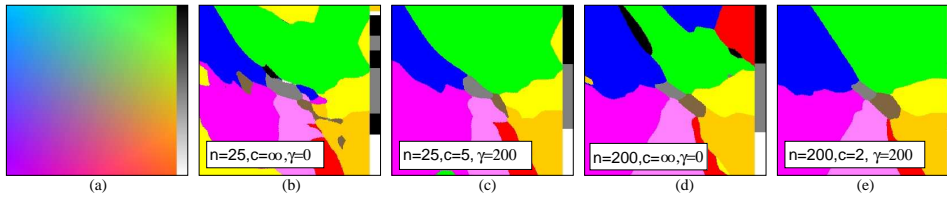
Figure 2.13   *(a) A challenging synthetic image: the highly saturated RGB values at the border rarely occur in natural images. (b-e) results obtains with different settings for c, γ and n the number of train images per color name. The figure demonstrates that the PLSA method, images (c) and (e), improves results. Reprinted with permission, ©2006 IEEE.*

2.3.3   Assigning Color Names in Test Images

Once we have estimated the distributions over words $p(w|z)$ representing the topics, we can use them to compute the probability of color names corresponding to image pixels in test images. As the test images are not expected to have a single dominant color, we do not use the label-based Dirichlet priors that are used when estimating the topics. The probability of a color name given a pixel is given by

$$p(z|w) \propto p(z)p(w|z), \tag{2.32}$$

where the prior over the color names $p(z)$ is taken to be uniform. The probability over the color names for a region is computed by a simple summation over all pixels in the region of the probabilities $p\,(z|w)$, computed with eq. 2.32 using a uniform prior.

The impact of the two improvements to standard PLSA discussed in section 2.3.2 is illustrated in figure 2.13 (for more detailed analysis see also [48]). The image shows pixels of constant intensity, with varying hue in the angular direction, and varying saturation in the radial direction. On the right side of the image a bar with varying intensity is included. Color names are expected to be relatively stable for constant hue, only for low saturation they change to an achromatic color (i.e. in the center of the image). The only exception to this rule is brown which is low saturated orange. Hence, we expect the color names to form a pie-like partitioning with an achromatic color in the center, like the parametric model which was introduced in section 2.2. Assigning color names based on the empirical distribution (figure 2.13(b)) leads to many errors, especially in the saturated regions. The extended method trained from only 25 images per color name (figure 2.13(c)) obtains results much closer to what is expected. If we look at the performance as a function of the number of training images from Google Image, we see that the difference between the PLSA method with optimal c-$γ$ settings and the empirical distributions becomes smaller by increasing the number of training images. However, the comparison shows that the extended method obtains significantly better results, especially in saturated regions (see figure 2.13(d) and (e)).

2.3.4   Flexibility Color Name Data Set

An advantage of learning color names from uncalibrated images, collected with Google image search, is that one can easily vary the set of color names. For the parametric

Figure 2.14   *First row: prototypes of the 11 basic color terms learned from Google images based on PLSA. Second row: prototypes of a varied set of color names learned from Google images. Third row: prototypes of the two Russian blues learned from Google images. Reprinted with permission, ©2009 IEEE.*

method described in section 2.2 changing the set of color names would mean that the psychophysical experiment needs to be repeated. Different sets of color names have for example been proposed in the work of Mojsilovic [28]. She asked a number of human test subjects to name the colors in a set of images. In addition to the eleven basic color terms beige, violet and olive were also mentioned.

In figure 2.14 we show prototypes of the eleven basic color terms learned from the Google images. The prototype w_z of a color name is that color which has the highest probability of occurring given the color name $w_z = \text{argmax}_w\, p\,(w|z)$. In addition, we add a set of eleven extra color names, for which we retrieve hundred images from Google image each. Again the images contain many false positives. Then a single extra color name is added to the set of eleven basic color terms, and the color distributions $p\,(w|z)$ are re-computed, after which the prototype of the newly added color name is derived. This process is repeated for the eleven new color names. The results are depicted in the second row of figure 2.14 and correspond to the colors we expect to find.

As a second example of flexibility of data acquisition we look into inter-linguistic differences in color naming. The Russian language is one of the languages which has 12 basic color terms. The color term blue is split up into two color terms: goluboi (голубой), and siniy (синий). We ran the system on 30 images for both blues, returned by Google image. Results are given in figure 2.14, and correspond with the fact that goluboi is a light blue and siniy a dark blue. This example shows internet as a potential source of data for the examination of linguistic differences in color naming.

2.4   EXPERIMENTAL RESULTS

In this section, we compare the two computational color naming methods discussed in section 2.2 and section 2.3. The most relevant difference between the two methods is the training data on which they are based, either calibrated or uncalibrated. The parametric method is based on color name labels given to colored patches which are taken in highly controlled environment with known illumination, absence of context, and a grey reference background. The second approach is based on real-world images of objects within a context with unknown camera settings, and illumination. We will test the two methods both on
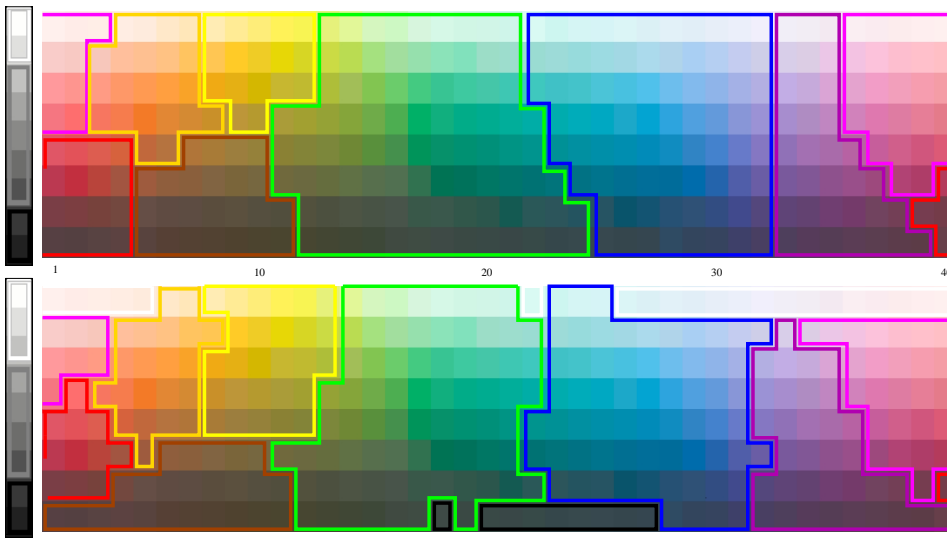
Figure 2.15 *top: color name categories on the Munsell color array obtained by parametric method. below: color names obtained with PLSA method. Note the differences in chromatic and achromatic assignments. The colored lines indicate the boundaries of the eleven color categories. Reprinted with permission, ©2009 IEEE.*

calibrated and on uncalibrated data. We will refer to the two methods as the parametric method and PLSA method.

Calibrated color naming data: First we compare both methods on classifying single patches which are presented under white illumination. We have applied both color naming algorithms to the Munsell color array used in the World Color Survey by Berlin and Kay [7]. The results are shown in figure 2.15. The top panel shows the results based on the parametric method, and the bottom panel shows the results obtained with the PLSA method. The color names are similarly centered, and only on the borders there are some disagreements. The main difference which we can observe is that all chromatic patches are named by chromatic color names by the parametric method, whereas the PLSA method names multiple chromatic patches with achromatic color names.

Model	Berlin and Kay data			Sturges and Whitfield data		
	Coincidences	Errors	% Errors	Coincidences	Errors	% Errors
parametric	193	17	8.10	111	0	0.00
PLSA	180	30	14.3	106	5	4.50
human	182	28	13.33	107	4	3.60

Table 2.2   Comparison of different Munsell categorizations to the results from color-naming experiments of Berlin and Kay [7], and Sturges and Whitfield [42].

method	cars	shoes	dresses	pottery	overall
parametric	56	72	68	61	64.7
PLSA	56	77	80	70	70.6

Table 2.3   Pixel annotation score for the four classes in the Ebay data set. The fifth column provides average results over the four classes.

To quantitatively compare the two methods on calibrated patches, we compare the outcome of the parametric and the PLSA method against the color name observations from two works of reference: the study of Berlin and Kay [7] and the experiments of Sturges and Whitfield [42] (figures 2.1 and 2.2). We count the number of coincidences and dissimilarities between the predictions of the models and their observations. The results are summarized in table 2.2. We see that the parametric model does significantly better than the PLSA model. This is what we expected since the parametric model is designed to perform color naming in calibrated circumstances. In addition, we compare to the categorization done by an English speaker presented by MacLaury in [25]. The results obtained by the English speaker show the variability of the problem, since any individual subject judgements will normally differ from those of a color-naming experiment, which are usually averaged from several subjects. Notice that the performance of the PLSA model is similar to that of an individual human observer, when compared to the averaged results from psychophysical experiments

Uncalibrated color naming data: To test the computational color naming methods on uncalibrated data, a human-labeled set of object images is required. For this purpose, we use a data set of images from the auction website Ebay [47]. Users labeled their objects with a description of the object in text, often including a color name. The data set contains four categories of objects: cars, shoes, dresses, and pottery (see figure 2.16). For each object category 121 images where collected, 12 for each color name. The final set is split in a test set of 440 images, and a validation set of 88 images. The images contain several challenges. The reflection properties of the objects differ from matt reflection of dresses to highly specular surfaces of cars and pottery. Furthermore, it comprises both indoor and outdoor scenes. For all images a hand-segmentation of the object areas which correspond to the color name is provided. The color naming methods are compared on the task of pixelwise color name annotation of the Ebay images. All pixels within the segmentation masks are assigned to their most likely color name. We report the *pixel annotation score*, which is the percentage of correctly annotated pixels.

The results are given in table 2.3. As can be seen, the PLSA method outperforms the parametric method by about 6%. This was to be expected since the PLSA method is learned from real-world Google images, which look more similar to the Ebay images. On the other hand, the parametric method based on calibrated data faces difficulties in the real-world where colors are not presented on a color neutral background under a known white light source. Figure 2.17 shows results on two real-world images. Both methods obtain similar results, but one can see that especially in the achromatic regions they differ (ground plane below the strawberry and the house). The parametric method assigns more chromatic

Figure 2.16   *Examples for the four classes of the Ebay data: green car, pink dress, yellow plate, and orange shoes. For all images masks with the area corresponding to the color name are hand segmented.*
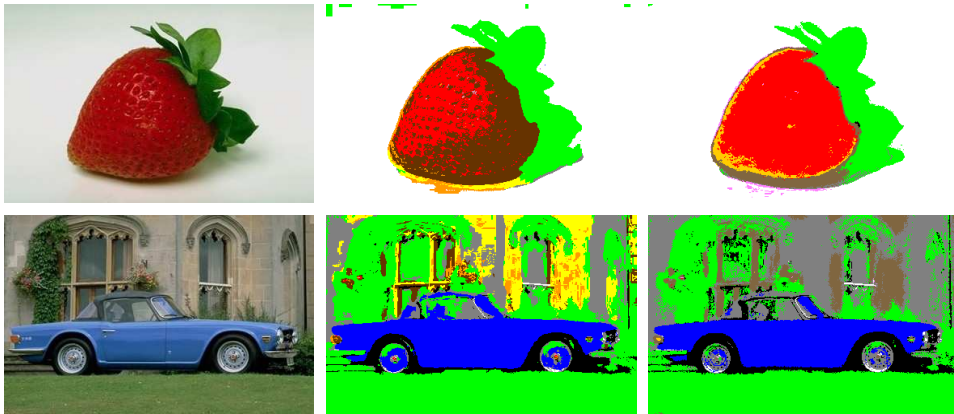

Figure 2.17   *Two examples of pixelwise color name annotation. The color names are represented by their corresponding color. In the middle the results of the parametric model are given, and on the right the results of the PLSA method.*

color names, whereas the PLSA method requires more saturated colors before assigning chromatic color names. Another difference is seen in some parts of the strawberry that are wrongly labeled as brown by the parametric model. Since the psychophysical experiments are carried out in controlled conditions the parametric model is not able to include the different shades that a color can adopt due to illumination effects. By contrast, the PLSA method labels correctly most of the strawberry because learning from real-world images allows the PLSA method to consider the different variations that any color can present.

2.5   CONCLUSIONS

In this chapter, we have discussed two approaches to computational color naming. Firstly, we have discussed a method for learning color names from calibrated data. The parametric

fuzzy model for color naming is based on the definition of the Triple-Sigmoid with Elliptical center as membership function. The main advantage of the parametric model is that it allows to incorporate prior knowledge about the shapes of the color name distributions. Secondly, we have seen a method to learn color names from uncalibrated images collected from Google image search. We have discussed a PLSA based learning to cope with the inherently noisy data retrieved from Google image search (the data contains many false positives). Learning color names from image search engines has the additional advantage that the method can easily vary the set of desired color names, something which is otherwise very costly.

Comparing the results of the two methods, we observed that the parametric method obtains superior results on calibrated data, and that the PLSA method outperforms the parametric model for real-world uncalibrated images. We use the term uncalibrated to refer to all kinds of deviations from the perfect setting of a single color patch on a grey background. Hence uncalibrated refers to unknown camera setting, unknown illuminant, but also the presence of physical events such as shadows and specularities. In the future, when color image understanding has improved, with improved illuminant estimation, and better object segmentations, the fact that the initial image is uncalibrated becomes less relevant. In such a scenario where we will be able to automatically calibrate uncalibrated data, the parametric models will become more important. The robustness which we observe now in the PLSA model will then be a disadvantage because it leads to reduced sensitivity. As a last remark we would like to point out that we have ignored the interactions between neighboring colors which can greatly influence the color sensation. With the usage of induction models the perceived color sensation can be predicted [32, 31]. Therefore, addition of such models in color naming algorithms is expected to improve results.

CITATION GUIDELINES

Throughout our research career we have collaborated with many people, without whose help the writing of this book would not have been possible. Here we provide the citation information for those chapters which are written with additional authors.

Chapter 2:
M. Lucassen. *Color Vision*, in: Th. Gevers, A. Gijsenij, J. van de Weijer, J.M. Geusebroek, Color in Computer Vision, Wiley.

Chapter 5:
J. van de Weijer, Th. Gevers, C. Schmid, A. Gijsenij. *Color Ratios*, in: Th. Gevers, A. Gijsenij, J. van de Weijer, J.M. Geusebroek, Color in Computer Vision, Wiley.

Chapter 6:
J.M. Geusebroek, J. van de Weijer, R. van den Boomgaard, Th. Gevers, A. W.M. Smeulders, A. Gijsenij. *Derivative-based Photometric Invariance*, in: Th. Gevers, A. Gijsenij, J. van de Weijer, J.M. Geusebroek, Color in Computer Vision, Wiley.

Chapter 7:
J.M. Álvarez, Th. Gevers and Antonio M. López. *Invariance by Learning*, in: Th. Gevers, A. Gijsenij, J. van de Weijer, J.M. Geusebroek, Color in Computer Vision, Wiley.

Chapter 13:
J. van de Weijer, Th. Gevers, A.W.M. Smeulders, A.Bagdavnov, A.Gijsenij. *Color Feature Detection*, in: Th. Gevers, A. Gijsenij, J. van de Weijer, J.M. Geusebroek, Color in Computer Vision, Wiley.

Chapter 14:
G.J. Burghouts, J.M. Geusebroek. *Color Descriptors*, in: Th. Gevers, A. Gijsenij, J. van de Weijer, J.M. Geusebroek, Color in Computer Vision, Wiley.
Chapter 15:
J.M. Geusebroek, Th. Gevers, G. Burghouts, M. Hoang. *Color Image Segmentation*. in: Th. Gevers, A. Gijsenij, J. van de Weijer, J.M. Geusebroek, Color in Computer Vision, Wiley.
Chapter 16:
K.E.A. van de Sande, Th. Gevers, C.G.M. Snoek, *Retrieval and Recognition of Digital Content*, in: Th. Gevers, A. Gijsenij, J. van de Weijer, J.M. Geusebroek, Color in Computer Vision, Wiley.
Chapter 17:
R. Benavente, J. van de Weijer, M. Vanrell, C. Schmid, R. Baldrich, J. Verbeek, and D. Larlus. *Color Names*, in: Th. Gevers, A. Gijsenij, J. van de Weijer, J.M. Geusebroek, Color in Computer Vision, Wiley.
Chapter 18:
Th. Gevers, H. Stokman, *Segmentation of Multispectral Images*, in: Th. Gevers, A. Gijsenij, J. van de Weijer, J.M. Geusebroek, Color in Computer Vision, Wiley.

REFERENCES

1. D. Alexander. *Statistical Modelling of Colour Data and Model Selection for Region Tracking*. PhD thesis, Department of Computer Science, University College London, 1997.

2. K. Barnard, L. Martin, B. Funt, and A. Coath. A data set for color research. *Color Research and Application*, 27(3):147–151, 2002.

3. R. Benavente and M. Vanrell. Fuzzy colour naming based on sigmoid membership functions. In *Proceedings of the 2nd European Conference on Colour in Graphics, Imaging, and Vision (CGIV'2004)*, pages 135–139, Aachen (Germany), 2004.

4. R. Benavente, M. Vanrell, and R. Baldrich. Estimation of fuzzy sets for computational colour categorization. *Color Research and Application*, 29(5):342–353, October 2004.

5. 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, 2008.

6. R. Benavente, M. Vanrell, and R. Bladrich. A data set for fuzzy colour naming. *COLOR research and application*, 31(1):48–56, 2006.

7. B. Berlin and P. Kay. *Basic color terms: their universality and evolution*. Berkeley: University of California, 1969.

8. D. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

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

10. D.M. Conway. An experimental comparison of three natural language colour naming models. In *Proc. east-west int. conf. on human-computer interaction*, pages 328–339, 1992.

11. G.D. Finlayson, M.S. Drew, and B. Funt. Color constancy: generalized diagonal transforms suffice. *Journal of the Optical Society of America A*, 11(11):3011–3019, 1994.

41

Please enter \offprintinfo{(Title, Edition)}{(Author)}
at the beginning of your document.

12. B.V. Funt and G.D. Finlayson. Color constant color indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):522–529, 1995.

13. Th. Gevers and A.W.M. Smeulders. Color-based object recognition. *Pattern Recognition*, 32:453–464, 1999.

14. Th. Gevers and H.M.G. Stokman. Robust histogram construction from color invariants for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):113–118, 2004.

15. W.C. Graustein. *Homogeneous Cartesian Coordinates. Linear Dependence of Points and Lines.*, chapter 3, pages 29–49. Introduction to Higher Geometry. Macmillan, New York, 1930.

16. L.D. Griffin. Optimality of the basic colour categories for classification. *R. Soc. Interface*, 3(6):71–85, 2006.

17. C.L. Hardin and L. Maffi, editors. *Color Categories in Thought and Language*. Cambridge University Press, 1997.

18. T. Hofmann. Probabilistic latent semantic indexing. In *Proc. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pages 50–57, 1999.

19. P. Kay and C.K. McDaniel. The linguistic significance of the meanings of basic color terms. *Language*, 3(54):610–646, 1978.

20. J.C. Lagarias, J.A. Reeds, M.H. Wright, and P.E. Wright. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal of Optimization*, 9(1):112–147, 1998.

21. R.L. Lagendijk and J. Biemond. Basic methods for image restoration and identification. In A. Bovik, editor, *The Image and Video Processing Handbook*, pages 125–139. Academic Press, 1999.

22. J.M. Lammens. *A computational model of color perception and color naming*. PhD thesis, Univ. of Buffalo, 1994.

23. E.H. Land and J.J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America A*, 61:1–11, 1971.

24. D. Larlus and F. Jurie. Latent mixture vocabularies for object categorization. In *British Machine Vision Conference*, 2006.

25. R.E. MacLaury. From brightness to hue: An explanatory model of color-category evolution. *Current Anthropology*, 33:137–186, 1992.

26. A. Maerz and M.R. Paul. *A Dictionary of Color*. McGraw-Hill, 1st edition, 1930.

27. G. Menegaz, A. Le Troter, J. Sequeira, and J. M. Boi. A discrete model for color naming. *EURASIP Journal on Advances in Signal Processing*, 2007, 2007.

28. A. Mojsilovic. A computational model for color naming and describing color composition of images. *IEEE Transactions on Image Processing*, 14(5):690–699, 2005.

29. Munsell Book of Color - Matte Finish Collection. Munsell Color Company, Baltimore,MD, 1976.

30. S.K. Nayar and R.M. Bolle. Reflectance based object recognition. *International Journal of Computer Vision*, 17(3):219–240, 1996.

31. X Otazu, C.A Párraga, and M Vanrell. Toward a unified chromatic induction model. *Journal of Vision*, 10(12)(6), 2010.

32. X. Otazu and M. Vanrell. Building perceived colour images. In *Proceedings of the 2nd European Conference on Colour in Graphics, Imaging, and Vision (CGIV'2004)*, pages 140–145, April 2004.

33. D.L. Philipona and J.K. O'Regan. Color naming, unique hues, and hue cancellation predicted from singularities in reflection properties. *Visual Neuroscience*, 23(3-4):331–339, 2006.

34. T. Regier and P. Kay. Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences*, 13(10):439–446, 2009.

35. T. Regier, P. Kay, and R.S. Cook. Focal colors are universal after all. *Proceedings of the National Academy of Sciences*, 102(23):8386–8391, 2005.

36. T. Regier, P. Kay, and R.S. Cook. Universal foci and varying boundaries in linguistic color categories. In B.G. Gara, L. Barsalou, and M. Bucciarelli, editors, *Proceedings of the 27th Meeting of the Cognitive Science Society*, pages 1827–1832, 2005.

37. T. Regier, P. Kay, and N. Khetarpal. Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, 104(4):1436–1441, 2007.

38. D. Roberson, J. Davidoff, I.R.L. Davies, and L.R. Shapiro. Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, 50(4):378–411, 2005.

39. D. Roberson, I. Davies, and J. Davidoff. Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, 129(3):369–398, 2000.

40. N. Seaborn, L. Hepplewhite, and J. Stonham. Fuzzy colour category map for the measurement of colour similarity and dissimilarity. *Pattern Recognition*, 38(1):165–177, 2005.

41. L. Steels and T. Belpaeme. Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Science*, 28:469–529, 2005.

42. J. Sturges and T.W.A. Whitfield. Locating basic colors in the Munsell space. *Color Research and Application*, 20(6):364–376, 1995.

43. M. Swain and D. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

44. Spectral database, university of joensuu color group. http://spectral.joensuu.fi. Last accessed on September 20, 2011.

45. J. van de Weijer and C. Schmid. Blur robust and color constancy image description. In *IEEE International Conference on Image Processing*, pages 993–996, 2006.

46. J. van de Weijer and C. Schmid. Coloring local feature extraction. In *European Conference on Computer Vision*, pages 334–348, 2006.

47. J. van de Weijer, C. Schmid, and J. Verbeek. Learning color names from real-world images. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Minneapolis, Minnesota, USA, 2007.

48. Joost van de Weijer, Cordelia Schmid, Jakob Verbeek, and Diane Larlus. Learning color names for real-world applications. *IEEE Transactions on Image Processing*, 18(7):1512–1524, july 2009.

49. L. Vincent. Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. *IEEE Trans. on Image Processing*, 2(2):176–201, April 1993.

50. L.A. Zadeh. Fuzz sets. *Information and control*, 8(3), 1965.