Perception Based Representations for Computational Colour

Maria Vanrell, Naila Murray, Robert Benavente, Alejandro Párraga, Xavier Otazu, and Ramon Baldrich

Computer Vision Center - Universitat Autnoma de Barcelona
Campus UAB, Edifici O, 08193 Bellaterra, Barcelona
{maria.vanrell,naila.murray,robert.benavente,
alejandro.parraga,xavier.otazu,ramon.baldrich}@uab.cat
http://cat.uab.cat

Abstract. The perceived colour of a stimulus is dependent on multiple factors stemming out either from the context of the stimulus or idiosyncrasies of the observer. The complexity involved in combining these multiple effects is the main reason for the gap between classical calibrated colour spaces from colour science and colour representations used in computer vision, where colour is just one more visual cue immersed in a digital image where surfaces, shadows and illuminants interact seemingly out of control.

With the aim to advance a few steps towards bridging this gap we present some results on computational representations of colour for computer vision. They have been developed by introducing perceptual considerations derived from the interaction of the colour of a point with its context. We show some techniques to represent the colour of a point influenced by assimilation and contrast effects due to the image surround and we show some results on how colour saliency can be derived in real images. We outline a model for automatic assignment of colour names to image points directly trained on psychophysical data. We show how colour segments can be perceptually grouped in the image by imposing shading coherence in the colour space.

Keywords: colour perception, psychophysical data, induction, saliency, naming, segmentation.

1 Introduction

Colour science has focused mainly on the study of colour representations, namely colour spaces, that allow to precisely describe the colour of a point. Its usual goal has been to define perceptual spaces where distance correlate with perceived dissimilarities. The dependency of colour with its surroundings has been partially introduced in the procedures to generate these colour spaces albeit in controlled conditions.

This approach is not very useful in computer vision where the inputs are digital images of unknown origin and therefore no information about the real scene

R. Schettini, S. Tominaga, and A. Trémeau (Eds.): CCIW 2011, LNCS 6626, pp. 16–30, 2011.

and the acquisition sensor exists. For example, it is usually assumed that the RGB vector component of each image pixel is the integration of three components over the visible wavelengths, that is

$$R = \int R(\lambda), E(\lambda), S_i(\lambda) d\lambda \;\; where \;\; i : R, G, B \tag{1}$$

where $R(\lambda)$ is the reflectance of the surface in the scene, $E(\lambda)$ is the scene illuminant and S_i are the corresponding RGB sensitivities of the camera. This formulation is a simplification of Shafer's dichromatic reflection model [26], after assuming that surfaces in the scene are Lambertian and that there are no reflection components (specularities), two assumptions that in general do not hold resulting in images usually full of shadows, highlights and specularities, unlike the actual appearance of real scenes.

The visual system has a tendency to keep its perceptions invariant to unimportant changes (i.e. illumination changes) and much effort was invested in researching for stable colour representations. Key contributions to this field were concerned with finding colour constancy algorithms capable of placing the image under the effects of a canonical illuminant [6,7], or invariant colour representations where the effects of the illuminant changes were removed from the image [8]. Although some of these approaches have been proven successful in controlled (calibrated) conditions, they are not widely used in common computer vision applications. In the last decade, some of the main advances in the computer vision field were based on the use of powerful machine learning techniques trained on large annotated image datasets. This general approach allowed computer vision scientists to achieve important results in real applications of automatic understanding of visual contents. The main contribution of colour research to this field has been to provide local features to be combined with shape descriptors in recognition tasks [10,9], or features to recover general scene shading [11] or the 3D shape of image objects [12].

2 Perception Based Representations

In this work we present several methods to deal with colour vision problems based on simple bottom-up approaches. The common point of these proposals is that they are not based on any previous learning step on large image-labelled datasets (supervised or unsupervised). Other than using such learning frameworks, we propose to solve a group of vision problems by inserting strong perceptual assumptions. This can be done by training the model (i.e. setting the parameters of the model) based on perceptual data acquired from psychophysical experiments. In this way, the data informs the model about the general behaviour of the underlying visual processes which are involved in performing the corresponding visual task. According to this, our ideas are articulated as follows:

First, we show how the data extracted from psychophysical experiments (based on setting a colour patch immersed in grating backgrounds with different spatial

frequency configuration and under different colour combinations) has allowed us to define a mathematical model of colour induction. This model uses the psychophysical data to fit the ECSF function that modulates the perception of colour in its surround and can form the basis of a general colour space that goes further than the colour of a point.

Secondly, we hypothesise that the modulation weights obtained in the induction model could form the basis of a bottom-up attention mechanism. We proved that building a saliency map just recovering the weights obtained by the induction model, we are able to correlate the obtained maps with the fixation data collected over a large image dataset. Sharing the low-level mechanisms trained on psychophysical data with induction effects, we achieve state-of-art results in saliency estimation.

Thridly, we present a general fuzzy set based model for colour naming. Similarly, as we do for induction, we fit specific functions based on a sigmoid basis to model colour naming judgements. The model can be fitted to different sets of naming data that can allow to introduce different perceptual conditions in the naming experiments. Some steps have been done in this direction by fitting the model with different backgrounds conditions [13].

Finally, we present an approach to segment image colour surfaces by modelling the ability of grouping colour on irregular surfaces by estimating the ridges of the colour distribution. In this case we hypothesise that the continuity of the perceived colour space form the basis to recover colour image segments. In this case the model is not based on a parametric function however, the computation of the distribution ridge is shown as a strong visual cue for segments which form the basis for higher level visual processes.

With these examples we try to sustain the view that robust visual cues in colour can be defined based on strong perceptual assumptions. Finding underlying processes of colour perception and inserting them in robust computational approaches may prove to be a valid approach to achieve powerful colour representations. This is the aim of this paper, which has been organised as follows. In section 3 we outline the induction model already developed in [2] and in section 4 we show how the model can be extended to be used for saliency estimation. In section 5 we show how membership functions for eleven basic colour terms have been fitted to a sigmoid based parametric model. Finally, in section 6 we explain how the colour distribution can represent the perceived coherence of the colour shading of a surface.

3 Colour Induction

Colour induction refers to the perceptual change in the colour of a stimulus due to the interactions with its surrounding region. When the perceived colour of a stimulus shifts towards the colour of its surround it is termed "assimilation". Conversely, contrast occurs when the perceived colour of the stimulus diverges from that of its surroundings. These two well-known effects are illustrated in Figure 1, which also shows the dependency of the effects on the local spatial frequency of the stimulus surround.

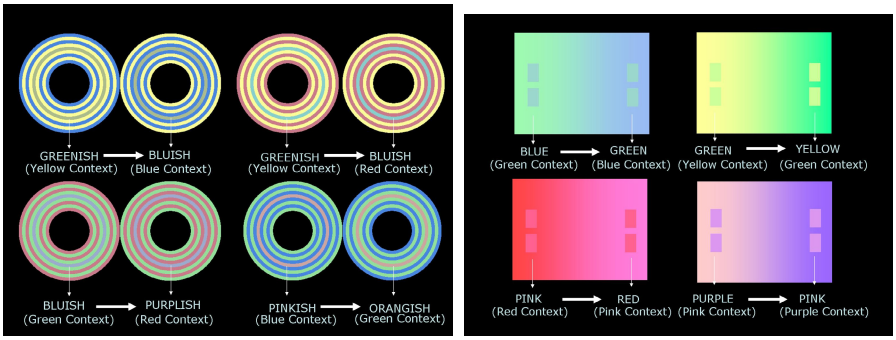


Fig. 1. Examples on induction effects, assimilation (on the left), contrast (on the right)

In a previous work [1,2] we showed that a multi-resolution framework was capable to predict both effects in a unified manner. Our model consisted of four stages in which different image representations were built. The final stage recovers a new image (referred here as the *perceived image*). The pipeline of the model can be summarised as follows:

$$I_c \xrightarrow{WT} \{\omega_{s,o}\} \xrightarrow{CS} \{z_{s,o}\} \xrightarrow{ECSF} \{\alpha_{s,o} \cdot \omega_{s,o}\} \xrightarrow{WT^{-1}} I_c^p \tag{2}$$

where I_c represents a colour channel of the input image, I, in an opponent colour space. The f stages of the model are:

- WT: a multi-resolution wavelet decomposition;
- CS: a center-surround mechanism developed as a divisive normalization [14];
- $ECSF$: a weighting with the extended contrast sensitivity function which was fitted to predict psychophysical data from assimilation and contrast experiments;
- WT^{-1}: an inverse wavelet transform that recovers the corresponding perceived image of the c channel, I_c^p.

In the next paragraphs we give a more detailed explanation of the model stages.

First stage (WT). The input image is convolved with a bank of filters using a multi-resolution wavelet transform. The resulting spatial pyramid contains wavelet planes oriented either horizontally (h), vertically (v) or diagonally (d). The coefficients of the spatial pyramid obtained using the wavelet transform can be considered as an estimation of the local oriented contrast. For an image I, the wavelet transform is denoted as:

$$WT(I_c) = \{w_{s,o}\}_{s=1,2,\ldots,n\,;\,o=h,v,d} \tag{3}$$

where $w_{s,o}$ is the wavelet plane at spatial scale s and orientation o. This wavelet transform contains Gabor-like basis functions and the number of scales used in

the decomposition is given by $n = log_2 D$ for an image whose largest dimension is size D.

Second stage (CS). At this level we simulate a center-surround mechanism based on computing a local contrast energy around each wavelet coefficient $\omega_{x,y}$ centered at position x, y. It is computed by convolving the coefficients with two filters, one for the center energy (small neighbourhood) and another for the surround (larger neighbourhood). By dividing the energy of the center by the energy of the surround window we obtain a measure of the surround contrast (denoted here as $r_{x,y}$). A non-linear scaling of $r_{x,y}$ is performed to produce the final center-surround energy measure $z_{x,y}$:

$$z_{x,y} = r_{x,y}^2/(1 + r_{x,y}^2). \tag{4}$$

As such, $z_{x,y}$ denotes the output of the second stage of the model or the center-surround energy.

Third stage (ECSF). In this stage the induction effects (assimilation and contrast) are introduced into the model using the *ECSF* function which was defined using psychophysical data. With this function we introduce a blurring effect to simulate assimilation, and a sharpening effect to simulate contrast. Both these effects are achieved simultaneously by using $ECSF$ as a weighting function that is parameterized by the z coefficients and the spatial frequency. $ECSF$ is defined as

$$ECSF(z, s) = z \cdot g(s) + k(s). \tag{5}$$

the function $g(s)$ is the combination of two exponential functions

$$g(s) = \begin{cases} \beta e^{-\frac{s^2}{2\sigma_1^2}} & s \leq s_0^g \\ \beta e^{-\frac{s^2}{2\sigma_2^2}} & otherwise \end{cases} \tag{6}$$

where s represents the spatial scale of the wavelet plane being processed, β is a scaling constant, and σ_1 and σ_2 define the spread of the spatial sensitivity of $g(s)$. The s_0^g parameter defines the peak spatial frequency sensitivity of $g(s)$. In Equation 5, the center-surround activity z of wavelet coefficients are modulated by $g(s)$. This $ECSF$ functions is used to weight the center-surround energy $z_{x,y}$ at a location, producing the final response of this stage $\alpha_{x,y}$:

$$\alpha_{x,y} = ECSF(z_{x,y}, s_{x,y}). \tag{7}$$

Fourth stage (WT^{-1}). This last stage uses the output of the previous stage, $\alpha_{x,y}$, as the weights that modulate the initial wavelet coefficient $\omega_{x,y}$. The perceived

Table 1. Parameters for $ECSF(z, s)$ obtained using least square regression

Param.	σ_1	σ_2	σ_3	β	s_0^g	s_0^k
Intensity	1.021	1.048	0.212	4.982	4.000	4.531
Colour	1.361	0.796	0.349	3.612	4.724	5.059

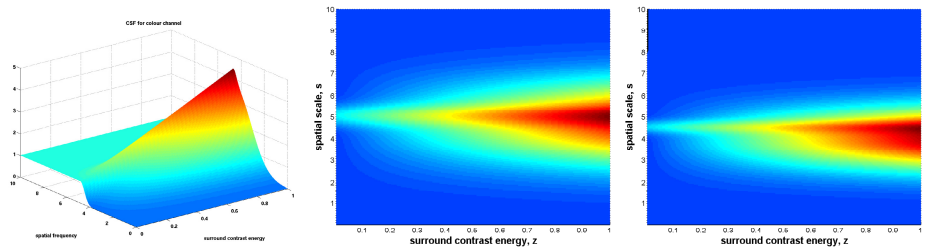


Fig. 2. $ECSF(z, s)$ function profile (left). 2D-plots of $ECSF(z, s)$ for chromaticity channels (center) and for intensity channel (right) (bluer colours represent lower values while redder colours indicate higher values).

image channel $I_c^{perceived}$ is obtained by performing an inverse wavelet transform on the wavelet coefficients $\omega_{x,y}$ at each location, scale and orientation, after the coefficients have been weighted by the $\alpha_{x,y}$ response at that location:

$$I_c^{perceived}(x, y) = \sum_s \sum_o \alpha_{x,y,s,o} \cdot \omega_{x,y,s,o} + C_r. \tag{8}$$

here o represents the orientation of the wavelet plane of $\omega_{x,y,s,o}$ and C_r represents the residual image plane obtained from WT.

The parameters of the $ECSF$ function (given in table 1) were estimated to predict psychophysical data obtained from two separate experiments. In the first experiment, by Blakeslee *et al* [15], observers performed asymmetric brightness matching tasks in order to match the illusions present in regions of the stimuli to a test patch. The second experiment was performed by Otazu *et al.* [2] in an analogous fashion, but with observers performing asymmetric colour and brightness matching tasks rather than tasks involving only brightness. The experiments were performed on stimuli such as those shown in figure 1. The resulting $ECSF$ functions are plotted in figure 2.

4 Colour Saliency

A great deal of research in computer vision is devoted to modelling attention mechanisms. To this end, models of bottom-up attention in image stimuli which construct saliency maps are popular. Given an image, the corresponding saliency map at each location estimates the probability of attracting the observer's gaze. There have been different approaches to create saliency maps that match the corresponding psychophysically-measured eye-fixation data [18,16,17]. Our contribution has been to extend the induction model defined in the previous section to produce a bottom-up, low-level image representation from which we can build saliency maps.

In the previous section we built a new image channel, I_c^p, that is a modified version of the original channel in which image locations may have been modified

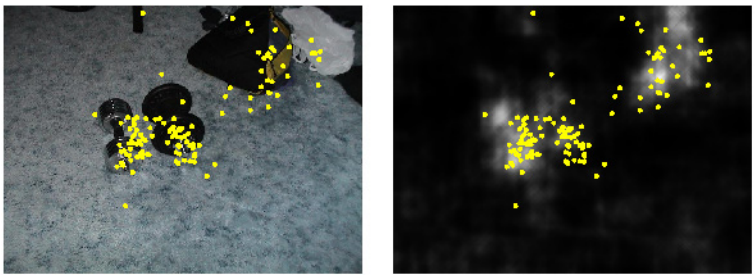


Fig. 3. Original image with fixation points (left). Our recovered saliency map (right).

by the α weight, either by a blurring or an enhancing effect. The colours of modified locations have either been assimilated (averaged) to be more similar to the surrounding colour or contrasted (sharpened) to be less similar to the surround.

To obtain predictions of saliency using this colour representation, we hypothesize that image locations undergoing enhancement are salient, while locations undergoing blurring are non-salient. In this sense we can directly define the saliency map of an specific image channel by the inverse wavelet transform of the α weight. Thus the saliency map, S_c, of the image channel I_c at the location x, y can be easily estimated as

$$I_c \xrightarrow{WT} \{\omega_{s,o}\} \xrightarrow{CS} \{z_{s,o}\} \xrightarrow{ECSF} \{\alpha_{s,o}\} \xrightarrow{WT^{-1}} S_c \tag{9}$$

where S_c denotes the saliency map of the image I. In figure 3(c) we show an example of one such saliency map. To evaluate the performance of a saliency estimation method, the predictions of the model are compared to eye fixation data. These psychophysical data are provided in large datasets that include image stimuli and eye-fixations, measured using eye-tracking hardware data, for multiple human observers.

We have assessed the accuracy of our model using the well-known receiver operating characteristic (ROC) and Kullback-Leibler (KL) divergence as quantitative metrics. The ROC curve indicates how well the saliency map discriminates between fixated and non-fixated locations for different binary saliency thresholds while the KL divergence indicates how well the method distinguishes between the histograms of saliency values at fixated and non-fixated locations in the image. For both of these metrics, a higher value indicates better performance.

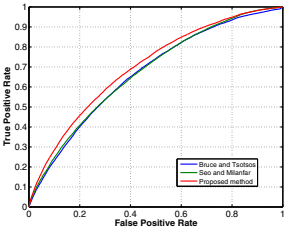
The dataset we used was provided by Bruce and Tsotsos in [16]. This popular dataset is commonly used as the benchmark for comparing eye-fixation predictions between methods. The dataset contains 120 colour images of indoor and outdoor scenes, along with eye-fixation data for 20 different subjects. The mean and the standard error of each metric are reported in Table 2. We performed this evaluation on two state-of-the-art methods as well as our proposed method and as Table 2 shows, our method exceeds the state-of-the-art performance as measured by both metrics.

Table 2. Performance in predicting human eye fixations from the Bruce and Tsotsos dataset (a) KL divergence and ROC Area (SE: Standard Error). (b) ROC curves for Bruce and Tsotsos, Seo and Milanfar, and the proposed method.

(a)

Model	KL (SE)	AROC (SE)
Bruce & Tsotsos [16]	0.2029 (0.0017)	0.6727 (0.0008)
Seo & Milanfar [17]	0.3432 (0.0029)	0.6769 (0.0008)
Our method	0.4265 (0.0030)	0.7013 (0.0008)

(b)



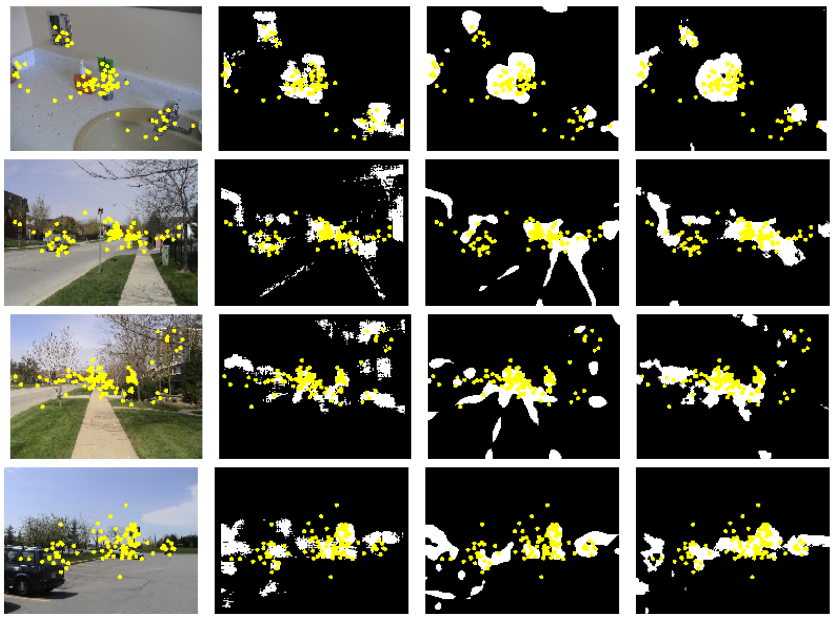


Fig. 4. Qualitative analysis of results for Bruce and Tsotsos dataset: Column A contains original image. Columns B, C, and D contain thresholded saliency maps obtained from Bruce and Tsotsos, Seo and Milanfar and our method, respectively. The saliency maps have each been thresholded to their top 10% most salient locations. Yellow markers indicate eye fixations. Our method is seen to be less sensitive to low-frequency edges such as street curbs and skylights, which is in line with human eye fixations.

5 Colour Naming

Colour naming relies on the assignment of a colour name label either to a point or to an image segment. This visual task has been studied from very different points of view. The anthropological study of Berlin and Kay [19] was a starting point that derived a lot of research about the topic in the subsequent decades. They studied colour naming in different languages and stated the existence of universal colour categories. They also defined the set of 11 basic colour categories that have the most evolved languages. These are white, black, red, green, yellow, blue, brown, purple, pink, orange and grey. Since then, several studies have confirmed and extended their results [20,21].

A computational model of colour naming can be very useful for several tasks such as segmentation, retrieval, tracking, or human-machine interaction. Although some models based on a pure tessellation of a colour space have been proposed [22,23], the most accepted framework has been to consider colour naming as a fuzzy process, that is, any colour stimulus has a membership value between 0 and 1 to each colour category. Kay and McDaniel [24] were the first in proposing a theoretical fuzzy model for colour naming. Later, some approaches from the computer vision field have adopted this point of view.

We proposed in [4] a fuzzy colour-naming model based on a family of membership functions that were fitted to psychophysical naming data. We worked on the CIELab space due to its perceptual properties. Likewise, other spaces could be suitable whenever one of the dimensions correlates with colour lightness and the other two with chromaticity components. Considering the psychophysical data on a chromaticity plane (see figure 5), we proposed to fit colour membership for the eight basic chromatic categories using a triple-Sigmoid function. With this function we are able to fit the configuration of the naming data obtained in a psychophysical experiments such as [25]. Data implied a set of necessary properties that membership functions for the chromatic categories should fulfil: a

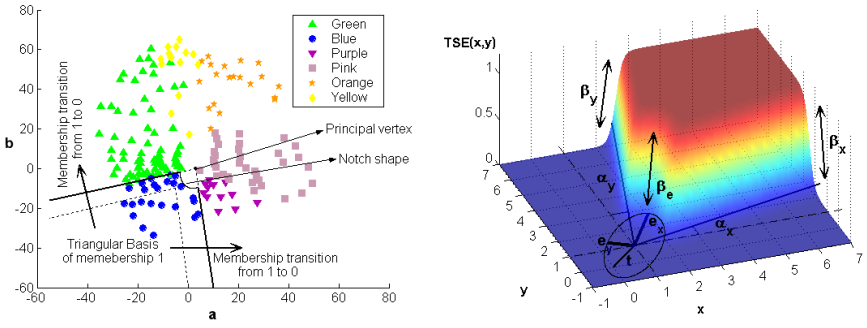


Fig. 5. Psychophysical naming data in CIELab space for a fixed L. Properties of the membership functions of a chromatic category (in this case, blue)(left). TSE function coping with expected properties.

triangular basis, two different slopes on both sides of the category, and a central notch to cope with the transition to the central achromatic category. To achieve these properties we defined the TSE function that for a given colour point $\mathbf{p}$ is defined as

$$TSE(\mathbf{p}, \theta) = DS(\mathbf{p}, \mathbf{t}, \theta_{DS}) ES(\mathbf{p}, \mathbf{t}, \theta_{ES}) \tag{10}$$

where $\theta = (\mathbf{t}, \theta_{DS}, \theta_{ES})$ is the set of parameters of the TSE function, which is defined as the product of a Double-Sigmoid function DS and an Elliptical-Sigmoid function ES. The DS function is defined as

$$DS(\mathbf{p}, \mathbf{t}, \theta_{DS}) = S_1(\mathbf{p}, \mathbf{t}, \alpha_y, \beta_y) S_2(\mathbf{p}, \mathbf{t}, \alpha_x, \beta_x) \tag{11}$$

where $\theta_{DS} = (\alpha_x, \alpha_y, \beta_x, \beta_y)$ is the set of parameters of the Double-Sigmoid function and function S_i is a sigmoid function defined as

$$S_i(\mathbf{p}, \mathbf{t}, \alpha, \beta) = \frac{1}{1 + \exp(-\beta \mathbf{u_i} R_\alpha T_t \mathbf{p})}, \qquad i = 1, 2 \tag{12}$$

where T_t and R_α are a translation matrix and a rotation matrix respectively, and u_i is a vector defining the axis on which the function is oriented. This function introduce the triangular basis of the function with two different slopes on both sides.

On the other hand, the ES function introduce the central notch allows to fit the boundary with the achromatic center. It is given by

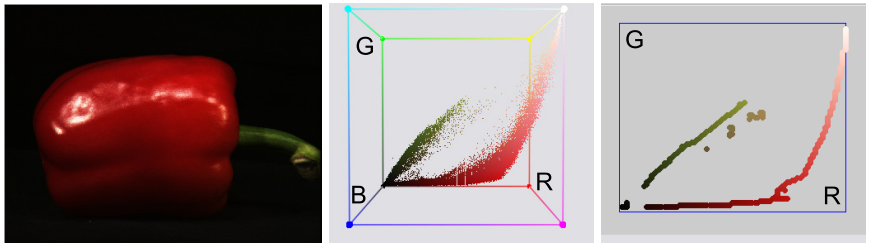
$$ES(\mathbf{p}, \mathbf{t}, \theta_{ES}) = \frac{1}{1 + \exp\left\{ -\beta_e \left[\left(\frac{\mathbf{u_1} R_\phi T_t \mathbf{p}}{e_x} \right)^2 + \left(\frac{\mathbf{u_2} R_\phi T_t \mathbf{p}}{e_y} \right)^2 - 1 \right] \right\}} \tag{13}$$

where $\theta_{ES} = (e_x, e_y, \phi, \beta_e)$ is the set of parameters, 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.

In figure 5(right) we can see how the TSE function adapts to the mentioned properties. By fitting the naming data of each chromatic category with this TSE function we can obtain the memberships of any colour sample in the CIELab space to the basic colour categories.

6 Colour Segmentation

Colour segmentation aims to partition an image into a set of non-overlapped regions corresponding to surfaces of a specific material. A robust and efficient colour segmentation is required as a preprocessing step in several computer vision tasks such as image classification or object detection and recognition. In real images changes due to illumination, shadow, shading and highlights provoke image measurements to vary significantly. These effects, are one of the main difficulties that have to be solved to yield a correct segmentation.



(a) (b) (c)

Fig. 6. (a) Original image (b) Colour Distribution. (c) Ridges extracted by RAD.

We proposed in [5] a pure bottom-up approach to recover a model of the material reflectance of the image objects by hypothesizing that our ability to perceive continuity of a coloured surface even the changes due to shading or highlights is a perceptual grouping mechanism that can be modelled by computing the connected ridges of the distribution in the colour space, we refer to it as *RAD* (Ridge-based Analysis of distributions).

Continuity of the material reflectance (MR) in the colour distribution is supported by the physical model defined by Shafer in [26]. A MR generates many image values due to geometrical and photometric variations that are likely to form a continous set in the histogram space. For this purpose, consider the distribution of a single MR as described by the dichromatic reflection model (DCM)[26] as

$$\mathbf{f}(\mathbf{x}) = m^b(\mathbf{x})\,\mathbf{c}^b + m^i(\mathbf{x})\,\mathbf{c}^i \tag{14}$$

where $\mathbf{f} = \{R, G, B\}$, $\mathbf{x}$ is the spatial image coordinate, $\mathbf{c}^b$ is the body reflectance, $\mathbf{c}^i$ the surface reflectance, m^b and m^i are geometry dependent scalars representing the magnitude of body and surface reflectance. Bold notation is used to indicate vectors. For one MR we expect both $\mathbf{c}^b$ and $\mathbf{c}^i$ to be almost constant, whereas $m^b(\mathbf{x})$ and $m^i(\mathbf{x})$ are expected to vary significantly. Hence, as for this definition, a MR, is formed by a single body reflectance $\mathbf{c}^b$ and a surface reflectance $\mathbf{c}^i$.

The two parts of the dichromatic reflection model are clearly visible in the histogram of figure 6(b). Firstly, due to the shading variations the distribution of the red pepper traces an elongated shape in histogram-space. Secondly, the surface reflectance forms a branch which points in the direction of the reflected illuminant. In conclusion, the distribution of a single MR forms a ridge-like structure in histogram space.

To extract this perceived MR, we used the multilocal creaseness MLSEC-ST operator introduced by Lopez *et al.* in [27] to enhance ridge points. Afterwards, the structure tensor computes the dominant gradient orientation in a neighbourhood of size proportional to σ_d. Basically, this calculus enhances those situations where either a big attraction or repulsion exists in the gradient direction vectors. Thus, it assigns the higher values when a ridge or valley occurs. Given a

distribution $\Omega(\mathbf{x})$, (colour histogram in the current context), and a symmetric neighbourhood of size σ_i centered at point $\mathbf{x}$, namely, $N(\mathbf{x}, \sigma_i)$ the structure tensor field S is defined as:

$$S(\mathbf{x}, \sigma) = N(\mathbf{x}, \sigma_i) * (\nabla\Omega(\mathbf{x}, \sigma_d) \cdot \nabla\Omega^t(\mathbf{x}, \sigma_d)) \tag{15}$$

where $\sigma = \{\sigma_i, \sigma_d\}$, and the calculus of the gradient vector field $\nabla\Omega(\mathbf{x}, \sigma_d)$ has been done with a gaussian kernel with standard deviation σ_d.

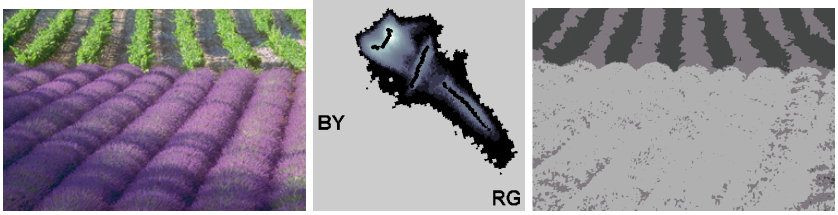
This operator assigns high values to those point of the distributions more likely to belong to a ridge. This scores comes from the divergence of the main orientation of the gradient in a given neighbourhood against the normal vector in it. The main orientation is extracted using the egeinvectors of the structure tensor ($S(\mathbf{x}, \sigma)$). An example of this resultant distribution is shown in figure 7(b) projected on a chromaticity space.

Once the ridge structure of the distribution has been enhanced with creaseness operator, next step is to extract the exact ridge points that describe the different MRs. As a result only those points necessary to maintain the connectivity of a MR remain. These points form the ridges of Ω^σ. The extraction is essentially based on a zero-crossing detection onto the MLSECT-ST output that maintains the spatial coherence, the whole ridge point detection is given by

$$RP(\Omega^\sigma) = LMP(\Omega^\sigma) \cup TRP(\Omega^\sigma) \cup SP(\Omega^\sigma) \tag{16}$$

that is the union of all different characteristic points found in the ridge of the distribution, these are the local maxima (LMP), the transitional ridge points (TRP) and the saddle points (SP). The final output of this set of point is depicted in figure 7(b) as black dots. The final segmentation is obtained assigning each point in the image to the closest ridge in the colour distribution (figure 7c). The details are explained in [5].

To evaluate quantitatively the performance the method it is compared to four state of art segmentation algorithms using the Berkeley image database (Table 3). Figure 8 shows qualitative results of the method applying different parameters to obtain from fine to coarse segmentation. In all cases the segments behave consistently.



(a) (b) (c)

Fig. 7. Example of MR extraction (a) Original image. (b) Result of the creassenes operator on a RG/BY chromaticity space. Extracted ridge points are given in black on top of the distribution. (c) Segmented image.

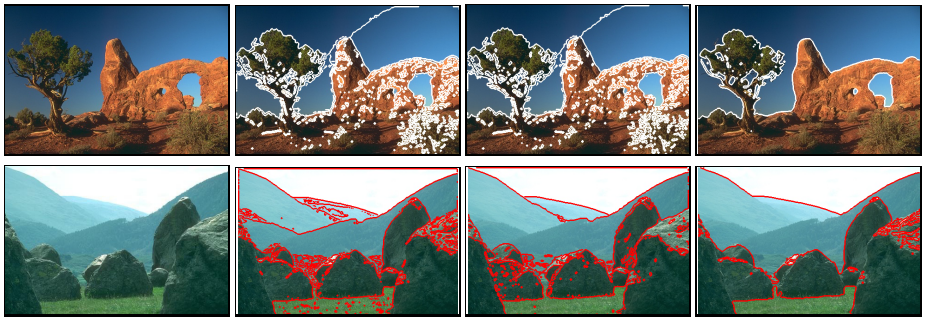


Fig. 8. Original image. Columns from 2 to 4: RAD-based segmentation on RGB with (σ_d,σ_i)={(1.5,0.05),(2.5,0.05),(2.5,1.5)}.

Table 3. Global Constancy Error: seed [28], fow [30], and mean-shift [31]

	human	RAD	seed	fow	mean-shift
GCE index	0.080	0.2048	0.209	0.214	0.2598

7 Conclusion

As we already mentioned in the introduction this paper reviews a methodological approach to tackle the problem of defining useful computational representations of colour. Our aim is to propose colour representations that go further from the basic three-dimensional spaces by exploring the perceptual processes that underly the role of colour in general visual tasks. To sustain the methodological proposal we have shown some examples of colour representations based on specific perceptual hypothesis.

References

1. Otazu, X., Vanrell, M., Párraga, C.A.: Multiresolution wavelet framework models brightness induction effects. Vision Research 48, 733–751 (2008)
2. Otazu, X., Párraga, C.A., Vanrell, M.: Toward a unified chromatic induction model. Journal of Vision 10(12) (2010)
3. Murray, N., Vanrell, M., Otazu, X., Párraga, C.A.: Non-Parametric Saliency Estimation based on low-level vision mechanism. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2011) (in press)
4. Benavente, R., Vanrell, M., Baldrich, R.: Parametric fuzzy set for automatic color naming. Journal of the Optical Society of America 25(10), 2582–2593 (2008)
5. Vázquez, E., Baldrich, R., van de Weijer, J., Vanrell, M.: Describing Reflectances for Color Segmentation Robust to Shadows, Highlights and Texture. IEEE Trans. on PAMI (2010)
6. Hordley, S.: Scene illuminant estimation: Past, present, and future. Color Research & Application 31, 303–314 (2006)

7. Finlayson, G.D., Hordley, S.D., Hubel, P.: Color by correlation: A simple, unifying framework for color constancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 23, 1209–1221 (2001)
8. Gevers, T., Smeulders, A.: Color based object recognition. Pattern Recognition 32, 453–464 (1997)
9. Khan, F.S., wan de Weijer, J., Vanrell, M.: Top-down color attention for object recognition. In: International Conference on Computer Vision, pp. 979–986 (2009)
10. van de Sande, K.E.A., Gevers, T., Snoek, C.G.M.: Evaluating color descriptors for object and scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(9), 1582–1596 (2010)
11. Grosse, R., Johnson, M.K., Adelson, E.H., Freeman, W.T.: Ground-truth dataset and baseline evaluations for intrinsic image algorithms. In: International Conference on Computer Vision, pp. 2335–2342 (2009)
12. Zickler, T., Mallick, S.P., Kriegman, D.J., Belhumeur, P.N.: Color Subspaces as Photometric Invariants. International Journal of Computer Vision 79(1), 13–30 (2008)
13. Benavente, R., Párraga, C.A., Vanrell, M.: European Conference on Visual Perception. Perception Suppl. Series, vol. 38, p. 36 (2009)
14. Heeger, D.H.: Normalization of cell responses in cat striate cortex. Visual Neuroscience 9(2), 181–197 (1992)
15. Blakeslee, B., McCourt, M.E.: Similar mechanisms underlie simultaneous brightness contrast and grating induction. Vision Research 37(20), 2849–2869 (1997)
16. Bruce, N.D., Tsotsos, J.K.: Saliency based on information maximization. In: Advances in Neural Information Processing Systems, vol. 18, pp. 155–162. MIT Press, Cambridge (2006)
17. Seo, H.J., Milanfar, P.: Nonparametric bottom-up saliency detection by self-resemblance. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009, pp. 45–52 (2009)
18. Itti, L., Koch, C., Niebur, E.: A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. IEEE Trans. Pattern Anal. Mach. Intell. 20(11), 1254–1259 (1998)
19. Berlin, B., Kay, B.: Basic Color Terms: Their Universality and Evolution. University of California Press, Berkeley (1969)
20. Boynton, R., Olson, C.: Salience of chromatic basic color terms confirmed by three measures. Vision Research 30, 1311–1317 (1990)
21. Sturges, J., Whitfield, T.: Salient features of munsell color space as a function of monolexemic naming and response latencies. Vision Research 37, 307–313 (1997)
22. Tominaga, S.: A color-naming method for computer color vision. In: Proceedings of IEEE International Conference on Cybernetics and Society, pp. 573–577. IEEE, Los Alamitos (1985)
23. Wang, Z., Luo, M., Kang, B., Choh, H., Kim, C.: An algorithm for categorising colours into universal colour names. In: Proceedings of the 3rd European Conference on Colour in Graphics, Imaging, and Vision, Society for Imaging Science and Technology, IS&T, pp. 426–430 (2006)
24. Kay, P., McDaniel, C.: The linguistic significance of the meaning of basic color terms. Language 3, 610–646 (1978)
25. Benavente, R., Vanrell, M., Baldrich, R.: A data set for fuzzy colour naming. Color Research and Applications 31, 48–56 (2006)
26. Shafer, S.: Using color to seperate reflection components. Color Research and Application 10(4), 210–218 (1985)

27. López, A.M., Lumbreras, F., Serrat, J., Villanueva, J.J.: Evaluation of methods for ridge and valley detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(4), 327–335 (1999)
28. Micusık, B., Hanbury, A.: Automatic image segmentation by positioning a seed. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 468–480. Springer, Heidelberg (2006)
29. Pantofaru, C., Hebert, M.: A comparison of image segmentation algorithms. Technical Report CMU-RI-TR-05-40, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA (September 2005)
30. Fowlkes, C., Martin, D., Malik, J.: Learning affinity functions for image segmentation combining patch-based and gradient-based approaches. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2003)
31. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(5), 603–619 (2002)