2000
Robert Benavente / Ramon Baldrich / Maria C. Olivé / Maria Vanrell
COLOUR NAMING CONSIDERING THE COLOUR VARIABILITY PROBLEM
Computación y Sistemas, julio-septiembre, año/vol. 4, número 001
Instituto Politécnico Nacional
Distrito Federal, México
pp. 30-43

Computación y Sistemas Vol. 4 No. 1 pp. 30 - 43
© 2000, CIC - IPN. *ISSN 1405-5546* Impreso en México

Colour Naming considering the Colour Variability Problem

Robert Benavente, Ramon Baldrich, Maria C. Olivé and Maria Vanrell
Computer Vision Center - Dept. Informàtica (UAB)
Edifici O - Campus UAB 08193 Bellaterra, SPAIN.
email: {robert,ramon,maria}@cvc.uab.es
phone: 34 93 581 24 40

Article received on March 31, 2000; accepted on June 1, 2000

Abstract

Colour is an important visual cue to understand scenes. However, most of the previous research in computer vision has been mainly focused on grey level images. At present, this tendency has changed and colour has become an active research topic in computer vision. In this work, we firstly make a review of colour perception and how to deal with computational colour. The variation of the illumination conditions of scenes leads to serious difficulties when working with colour images. We have classified the different approaches to deal with the variability problem into three groups: colour constancy algorithms, definition of colour invariants and computation of normalized images. Secondly, we present a simple method for colour naming which is based on a normalization approach, that is the comprehensive colour normalization of (Finlayson et al., 1998). In order to evaluate it, our method has been tested on a real surveillance application.

Keywords: Computer Vision, Colour Perception, Constancy, Normalization, Naming.

1 Introduction

Colour is a perceptual phenomenon due to the human visual system processing of the electromagnetic radiation that reaches the retina (Levine, 1985). This process can be seen as a change in representation which, in general, implies a dimensionality reduction that will be explained in the following sections.

Colour is a visual cue that is normally associated to surface appearance. Although colour has not been given much importance by researchers up to present (since most of the previous work in computer vision has been made for grey level images), the situation is now changing. Colour is becoming a very important visual cue for most of the vision tasks, such as object recognition (Healey and Slater, 1994), image indexing (Swain and Ballard, 1991), tracking (Crowley and Berard, 1997), shape extraction from colour variations (Funt *et al.*, 1992), etc.

The objective of this paper is twofold. The first goal is to make a brief introduction to the colour perception problem seen from the computer vision point of view. The second objective is to apply a normalization technique to a specific vision problem such as color naming.

To introduce colour cue in the visual tasks we must take into consideration the variability of this visual stimulus. Colour perception is always dependent on the context: the illuminant, the receptor sensibility and the scene geometrics have an important influence on the perceived scene. The human visual system presents a chromatic adaptation ability which allows us to avoid, in some sense, those context influences over the final perception. Any system doing a visual task involving colour processing should always take into account the colour constancy problem. This problem has been the goal of previous works ((Maloney and Wandell, 1986), (Finlayson *et al.*, 1998), (Funt and Finlayson, 1995), (Healey and Slater, 1994), (Healey and Wang, 1995), (Forsyth, 1990)).

One of the most common human visual tasks is the colour naming. The aim of this task is, given a region from a scene with a more or less homogeneous colour, to take a decission in natural language about what is the hue or colour that best describes the region. Obviously, a lot of shades of the same colour may be given in natural language descriptions, but we are oriented to just associate the name of the predominant colour in each region of the scene. An interesting review about this problem can be seen in (Lammens, 1994).

This paper has been organized in the following way. Firstly, we give a brief description of the biological basis of colour perception in the human visual system. Then, some of the most important experimental results about colour perception undertaken in the field of colorimetry are summarized. In these two previous sections we try to give the basic formulation of colour representation and then, in section 4, we present the colour problem from the computer vision point of view. In section 5, we show different approaches to deal with the colour constancy problem. In section 6, we present our proposal to apply a colour normalization technique in order to solve the colour naming problem. In section 7, we apply our method to a real surveillance problem and we give some results. Finally, in section 8, we present the conclusions of this application and explain the lines for future work.

2 Biological Colour Perception

Colour perception phenomenon begins when a luminous stimulus reaches the retina. However, only a part of the whole radiant energy that receives the retina causes a visual stimulus. This is what is called the visible spectrum referring to those signals containing wavelengths between 380nm and 730nm. Outside this band of frequencies the human visual system has no sensitivity.

The retina contains two kinds of receptors: rods and cones, which receive their names from their shape. While rods are sensitive to low intensity stimulus, cones give response to higher intensities than rods and also provide the information used by the human visual system to produce the colour perception sensation.

Colour perception is related to the presence in the cones of three different photopigments. Depending on these photopigments there exist three types of cones. Each one of them presents different sensitivity and concentrates its response on certain wavelengths. Thus, the three cone types are called *l* (*large*), *m* (*medium*) and *s* (*small*) related to the part of the spectrum where they present maximum sensitivity. Normally, cones are also called *red*, *green* and *blue* respectively, although the maximum sensitivity wavelength does not exactly correspond to these colours.

The spectral sensitivity curves of the three cone types in the retina have been measured by different authors and with different techniques. In all cases, the results have been very similar (Levine, 1985). In figure (1) the sensitivity curves estimated from psycophysical data can be seen.

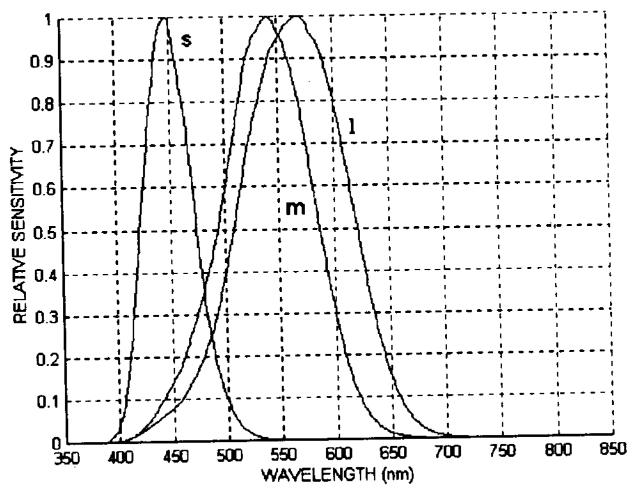


Figure 1: Sensitivity curves of the three cone types of the human visual system

Given an input stimulus $I(\lambda)$ trichromatic colour perception theory considers that there are three independent parallel processes in the three types of cones. Each process is taken in the retina and it is done in two different stages. The first part of the process can be seen as a linear transformation of the input stimulus into a three-dimensional space that is normally called trivariant space. This transformation is formulated by equations (1), (2) and (3).

$$A_l = \int_{\lambda_{min}}^{\lambda_{max}} I(\lambda)l(\lambda)d\lambda \tag{1}$$

$$A_m = \int_{\lambda_{min}}^{\lambda_{max}} I(\lambda)m(\lambda)d\lambda \tag{2}$$

$$A_s = \int_{\lambda_{min}}^{\lambda_{max}} I(\lambda)s(\lambda)d\lambda \tag{3}$$

where A_l, A_m and A_s are the coordinates values in the trivariant space, λ_{min} and λ_{max} are the limits of the visible spectrum and $l(\lambda)$, $m(\lambda)$ and $s(\lambda)$ are the sensitivity functions of the retina receptors.

According to trichromatic theory, the second stage of the process, which consists on a non-linear transformation, is taken from input stimulus representation in the trivariant space. The way this second step works is still almost unknown, but it is supposed to be related to chromaticity adaptation and other intensity and temporal effects. This second stage of the process can be described by equations (4), (5) and (6).

$$l_0(t) = f_l[A_l(t), t] \tag{4}$$

$$m_O(t) = f_m[A_m(t), t] \quad (5)$$

$$s_O(t) = f_s[A_s(t), t] \quad (6)$$

where $l_O(t)$, $m_O(t)$ and $s_O(t)$ are the values corresponding to input stimulus in the resulting three-dimensional space and f_l, f_m and f_s are the transformations applied to each one of the input stimulus coordinates in the trivariant space.

There are physiological and psycophysical evidences (Ebenhoh and Hemminger, 1981) that there exists a second level of processing based on colour differences. According to the opponent-colour model, this process is done in the high levels of the lateral geniculate body of the visual cortex.

A spatial opponent system is that in which two concentric regions of the receptive field demonstrate antagonistic behavior. Thus, each one of the regions is activated or inhibited by opposite stimulus.

The opponent-colour model considers the existence of four different types of opponent cells. One type is excited by a red signal and is inhibited by a green signal, while a second type has the opposite behavior. This pair is normally referred as the "*RG*" system. The two other types of opponent cells have the same behavior as the "*RG*" system but with yellow and blue signals. This pair is referred as the "YB" system. The model is completed by two types of non-opponent cells which are only sensitive to intensity changes and that are referred as the "$Wh - Bl$" system.

There are several models that try to explain the relationship between the cones output and the opponent-colour system, but the mechanism followed by the human visual system is not understood yet.

It is also unknown the way in which the human visual system manages the information from the opponent-colour system to codify the colour names and to label each stimulus with its corresponding name. It seems quite clear that the "$Wh - Bl$" system provides the intensity information, but the way in which chromaticity information from "*RG*" and "YB" systems is used to codify hue and saturation has not been determined up to now. Moreover, the colour naming mechanism is also influenced by the knowledge coming from the illuminant properties and the surface properties of the object being viewed. The human visual system has colour constancy mechanisms to controlate this dependency from the illuminant. Thus, the same surface under different illuminants is normally perceived as being of the same colour.

3 Colorimetry

Once the causes of the colour perception in the human visual system have been introduced, we will see how this phenomenon has been studied by the physics. Colorimetry is the field of the physics which has mathematically specified the colour perception phenomenon.

Isaac Newton is considered the first who established the relationship between light and colour. In 1704, Newton obtained the colour spectrum that forms white light using a glass prism to decompose sunlight.

In the early 19th century, Thomas Young set one of the basis of colorimetry when he suggested that in the human visual system colour processing there were only involved three independent variables related with three primary colours. In 1854, Grassman suggested that colour matches are based on linear operations and enunciated the colour additivity law.

Taking into account the two previous assumptions (three variable representation and additivity law) several matching colour experiments have been carried out. These experiments are based on mixing three monochromatic lights **R**, **G** and **B** in order to obtain the equivalent of a given colour light **Q** by only adjusting the intensities r, g and b of the three monochromatic lights.

In 1931, the CIE (*Comission Internationale de l'Eclairage*) determined the colour-matching functions for a standard observer $\bar{r}(\lambda)$, $\bar{g}(\lambda)$ and $\bar{b}(\lambda)$, which define the amount of each wavelength that is needed to obtain a certain colour of the spectrum (Figure 2).

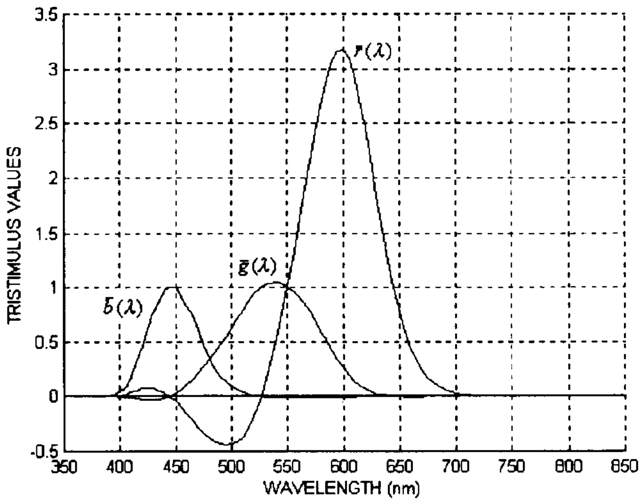


Figure 2: Colour matching functions for the RGB system

From the results obtained on these experiments the trichromatic generalization has been accepted. That is, under a wide range of conditions, most of the colour stimulus can be specified with only three coordinates in a three-dimensional subspace generated from three primary vectors **R**, **G** and **B**. This relationship can be expressed as:

$$Q(\lambda) = R\mathbf{R} + G\mathbf{G} + B\mathbf{B} \quad (7)$$

where R, G and B are the tristimulus coordinates and they are obtained from equations (8), (9) and (10).

$$R = \int_{\lambda} Q(\lambda)\bar{r}(\lambda)d\lambda \quad (8)$$

$$G = \int_{\lambda} Q(\lambda)\bar{g}(\lambda)d\lambda \quad (9)$$

$$B = \int_{\lambda} Q(\lambda)\bar{b}(\lambda)d\lambda \quad (10)$$

Nevertheless, the RGB space presents some problems:

- The $\bar{r}(\lambda)$ function has negative values for some wavelengths, which can be an impediment in the design of colour measurement instruments.
- There is no axis directly related to light intensity.

In order to solve these problems the CIE defined another colour space referred as XYZ, which is obtained from the application of a linear trasformation to the RGB space. This new system is based in the use of three imaginary primitives that are positive for all λ.

In this new space, the tristimulus values for a given colour are obtained from equations (11), (12) and (13).

$$X = \int_{\lambda} Q(\lambda)\bar{x}(\lambda)d\lambda \quad (11)$$

$$Y = \int_{\lambda} Q(\lambda)\bar{y}(\lambda)d\lambda \quad (12)$$

$$Z = \int_{\lambda} Q(\lambda)\bar{z}(\lambda)d\lambda \quad (13)$$

where $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ are the colour-matching functions for the XYZ space. Their graphic representation can be seen in figure (3).

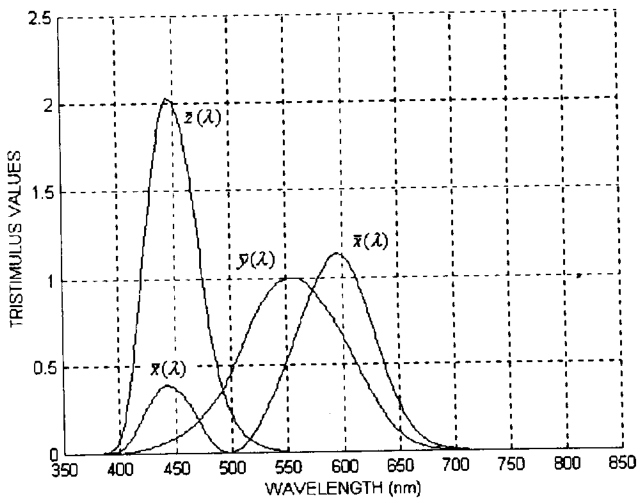


Figure 3: Colour matching functions for the XYZ system

As well as in the previous case, in this new space there not exist a perceptual interpretation of the coordinates, with the exception of the approximated relationship between Y-coordinate and colour intensity. However, the XYZ system has been accepted as a standard colour system.

Therefore, we can state that colorimetry has contributed with the definition of a psycophysical model to define colour as it is viewed by humans and, moreover, has created a standard colour space to have a measure of colour which is the base of most of the colour measurement instruments.

4 Colour in Computer Vision

Up to this point, we have analysed the human visual processing of the luminous stimuli that reach the retina. These stimulus will normally be the result of a light reflected by the surface of an object (Figure 4).

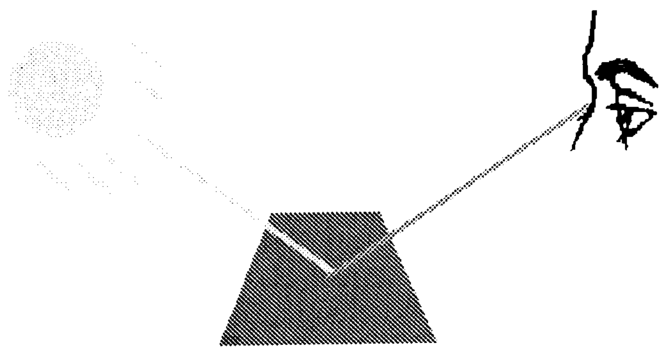


Figure 4: Luminous stimulus that reach the retina, are normally the result of the reflected light by a surface

The radiant energy reflected by an object surface can be expressed as a function of the incident light, $P(\lambda)$, and the surface reflection, $S(\lambda)$. Therefore, XYZ coordinates can be expressed as (14), (15) and (16).

$$X = \int_{\lambda} P(\lambda)S(\lambda)\bar{x}(\lambda)d\lambda \quad (14)$$

$$Y = \int_{\lambda} P(\lambda)S(\lambda)\bar{y}(\lambda)d\lambda \quad (15)$$

$$Z = \int_{\lambda} P(\lambda)S(\lambda)\bar{z}(\lambda)d\lambda \quad (16)$$

In computer vision, the receptor is not the human eye and image is formed by an acquisition device (Figure 5).

The way an image acquisition device works is based on the use of prisms to decompose the input signal into a set of signals only containing wavelengths that correspond to each one of the camera sensors.

To work in colour, three sensor devices are normally used. The three sensors provide a decomposition of the input signal into three channels R_c(red), G_c(green) and B_c(blue). These values can be expressed as (17), (18) and (19).

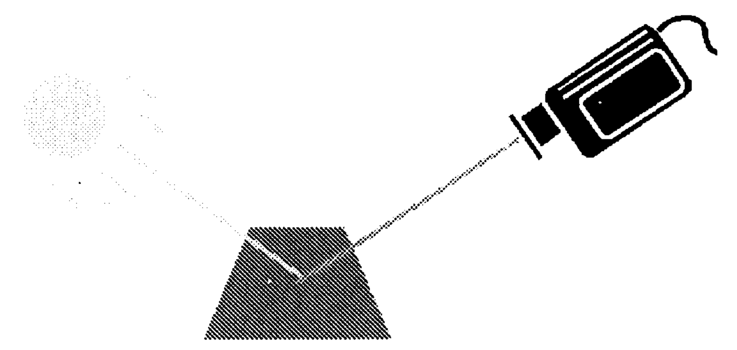


Figure 5: In computer vision the receptor is not the human eye, but an acquisition device

$$R_c = \int_\lambda P(\lambda)S(\lambda)R_r(\lambda)d\lambda \quad (17)$$

$$G_c = \int_\lambda P(\lambda)S(\lambda)R_g(\lambda)d\lambda \quad (18)$$

$$B_c = \int_\lambda P(\lambda)S(\lambda)R_b(\lambda)d\lambda \quad (19)$$

where $R_r(\lambda)$, $R_g(\lambda)$ and $R_b(\lambda)$ are the spectral response functions of sensors for red, green and blue channels , respectivelly. Given that, R_c, G_c and B_c values depend on the camera spectral response, which implies that the RGB space above defined is specific for each device.

Nevertheless, colour in computer vision is a surface property that allows obtaining information useful for different visual tasks: object segmentation, object recognition and classification, shape extraction from colour, colour naming, etc... The input to all those visual tasks is always a digital image represented in the RGB space of the acquisitions device which has been used.

4.1 Colorimetry vs. Computational Colour

Once the visual tasks in computer vision that involve colour have been introduced (that is what we call computational colour) let us see how the colour bases set by colorimetry can be exploited for colour vision.

From colorimetry point of view, any colour representation is based on the fact of relating colour to the standard XYZ space. In computer vision, the starting-point is always the RGB space which is dependent on the acquisition device used to acquire the images.

It is easy to work with the XYZ space if the reflectance functions of the surfaces and the illuminant spectral composition are known, since $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ functions are tabulated. However, the problem is that the illuminant, the reflectance functions of the surfaces and the spectral responses of the camera are normally unknown. Thus, the process to obtain the tristimulus coordinates of the XYZ system is further complicated.

In order to solve the problem of how to pass from the device RGB space to the standard XYZ system, two alternatives are possible:

- To work in controlled environments which allow measuring the illuminant spectrum and the surface reflectances.
- To assume that the transformation from the camera RGB space to the standard XYZ space is linear and try to estimate it from the knowledge of the XYZ and RGB coordinates of a basic set of acquired surfaces.

Above alternatives are based on the fact of being working in controlled environments where it is possible to measure the illuminant spectrum and the surface reflectances. Most of the research in computer vision does not present these possibilities and researchers work on standard sets of images for which no acquisition condition is known. Hence, it is normal the use of the acquired RGB coordinates followed by computational processing in order to solve the derived problems.

Up to this point, above arguments make us to conclude that colour can be a very different problem depending on the way in which it will be handled. Given that, in computer vision, any problem is normally posed from a set of RGB images, taken from a non-controlled scene and with an unknown camera. Thus, we are not able to use standard colour spaces from colorimetry and we have to dessign algorithms able to overcome the uncertain conditions of the acquisition process of any scene.

4.2 The Colour Constancy Problem

The main problem that any colour application directly working with the RGB values will find is the variation of the illuminant in intensity and/or colour. As it has been presented in the previous sections, colour perception is related to the illuminant spectral composition. Thus, the same surface may present very different appearances under different illuminants or under different intensities of the same illuminant. In Figure (6) an example of the variation that the same scene can suffer when there are changes in the illuminant characteristics is shown.

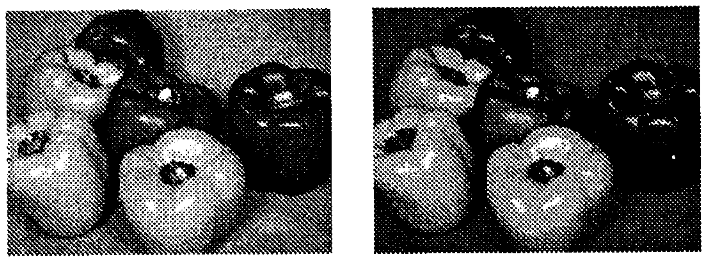


Figure 6: The same scene viewed under different light conditions may present very different appearance

However, a human observer will still be able to infer that the colours of the two kind of peppers in figure 6 are yellow and red. That is because the human visual system has an adaptative mechanism that allows us to avoid the spectral variations of the light of the scene and assign stable colour names to objects. This perceptual ability is called colour constancy (Figure 7).

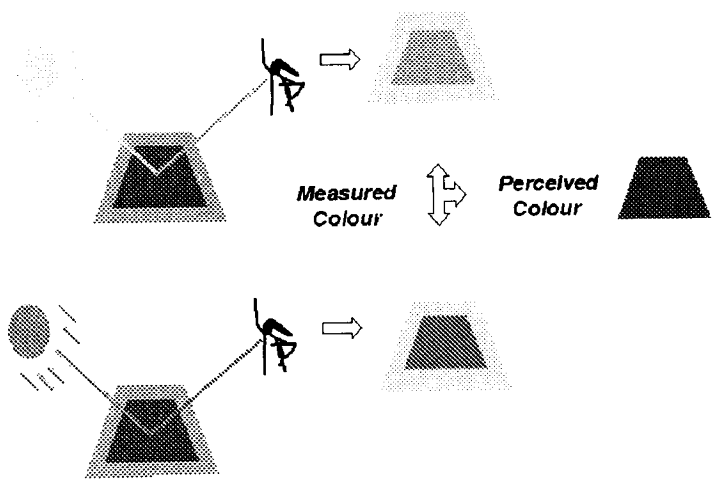


Figure 7: Colour constancy ability of the human visual system allows us to perceive the same colour although the viewed surface is illuminated by very different lights

The way in which the human visual system performs its colour constancy ability is still unknown. The most accepted hypothesis is Von Kries model (Worthey and Brill, 1986), also called the coefficient rule, which suggests that a change on the chromatic adaptation is caused by the sensitivity reduction or extension in each kind of cone and without affecting their relative spectral sensitivity.

Since standard cameras do not have colour constancy ability, the illumination variability is one of the main problems in computer vision. Simple changes in the illuminant intensity or colour may dramatically decrease the performance of RGB-based colour algorithms.

In order to solve this limitation, computer vision has developed a set of techniques that try to introduce the chromatic adaptation ability of the human visual system. These algorithms can be classified into three main groups, and they will be used depending on the visual task we want to solve. The three gropus of computation techniques developed in order to deal with the colour constancy problem are presented in the next section.

5 Techniques to deal with colour variability

Before studying the different computational approaches to the problem presented in the preceeding section, we must consider the fact that the light reflected by a surface not only depends on the spectral properties of illumination and surface reflectance. It also depends on other factors such as specularities and mutual illumination (Finlayson *et al.*, 1993). In order to simplify colour analysis it is usual to set some restrictions over real world.

One of the most extended models is to consider real world as a simplified Mondrian world. That is to consider the image as a planar surface composed of overlapping Lambertian surfaces. A Lambertian surface is a uniformly diffusing surface for which the luminous intensity I_ϵ in any given direction varies as the cosine of the angle ϵ between that direction and the normal ($\epsilon = 0$) to that surface, that is,

$$I_\epsilon = I_{\epsilon=0} \cos\epsilon$$

This implies that a Lambertian surface presents the same brightness in all its points despite the direction from which it is viewed (Wyszecki and Stiles, 1982). Moreover, in Mondrian world model, a spatially uniform intensity and a constant spectral power distribution of the illuminant are assumed. Hence, the only factor affecting the surface colour appereance is the variation of illumination across the time.

Once the general assumptions for the image model have been presented, we will see a generic classification of the different solutions for the illuminant variations that have been proposed during the last twenty years. Even though a lot of very different computational techniques have been developed, actually most of them can be classified into three main groups:

- **Colour constancy methods**
  These techniques attempt to estimate the characteristics of the illuminant used to acquire the images and then remove the colour cast due to the estimated illuminant.

- **Invariant measures**
  These methods are based on the definition of measures that have the property of being invariant to illumination changes and that extract the image properties to be analysed. When a colour invariant is applied, the conventional image structure is lost, since invariant measures values do not directly represent pixel's colour values.

- **Normalization techniques**
  These algorithms do a representation change of the image to a space where illuminant influence has been removed, even though the new space might not be a standard space and might not let recover the surfaces reflectances.

Recent work (Finlayson *et al.*, 1999) group these two last approaches together under the name of 'invariant normalizations'. However, we have separated them in terms of the two objectives they pursuit.

In the following sections (5.1, 5.2 and 5.3) this classification of the methods is presented in more detail.

5.1 Colour Constancy methods

Colour constancy methods are based on the definition of a standard illuminant usually referred as the 'canonical illuminant'. This canonical illuminant is used to obtain a stable descriptor for object surface properties, that is, the RGB values of each pixel are transformed to the canonical RGB descriptor relative to the canonical illuminant. Such descriptor is illuminant independent and it represents the colour that would have been perceived if the surface had been acquired under the canonical illuminant.

Hence, colour constancy problem in computer vision can be expressed as the problem of parametrizing the transformations which allow obtaining an image of the scene as it would appear under the canonical illuminant, from the image of the scene under an unknown illuminant. That is, for each RGB vector $\overline{p}_i$ corresponding to a pixel of an image $\mathcal{I}$ a mapping $\mathcal{M}$ to the illuminant independent descriptor $\overline{d}_i$ is applied:

$$\forall \overline{p}_i \in \mathcal{I}, \overline{d}_i = \mathcal{M}\overline{p}_i \tag{20}$$

where $\mathcal{M}$ is a linear transformation.

Finlayson et al. have demonstrated that perfect colour constancy can be achieved by a 3×3 diagonal matrix transform for typical scene illuminants (Finlayson *et al.*, 1993). Then, equation (20) becomes (21).

$$\forall \overline{p}_i \in \mathcal{I}, \overline{d}_i = \mathcal{D}\overline{p}_i \tag{21}$$

where $\mathcal{D}$ is the 3×3 diagonal matrix that maps the RGB vector $\overline{p}_i$ to the illuminant independent descriptor $\overline{d}_i$

This model can be considered valid if we suppose we are working with narrow-band response cameras. Otherwise, a sensor transformation $\mathcal{T}$ must be applied prior to the application of a diagonal matrix. In that case, equation (21) becomes (22).

$$\forall \overline{p}_i \in \mathcal{I}, \mathcal{T}\overline{d}_i = \mathcal{D}\mathcal{T}\overline{p}_i \tag{22}$$

As it can be seen, when the mapping is a 3×3 diagonal matrix, it is just the application of the coefficient rule or Von Kries Law since each image channel is scaled by a factor that do not depend on the other two.

Another important fact to point out is that the mapping must only account for the relative spectral power distribution change between the unknown illuminant and the canonical (Finlayson, 1996; Forsyth, 1990) but not for intensity. In order to be intensity independent one of the most used normalizations is to pass the image to chromaticity coordinates (see section 5.3) before applying the colour constancy algorithm.

Many different colour constancy methods have been developed. Although there were early implementations in the 60's and 70's, the real colour constancy interest has come in the last twenty years: greyworld algorithm (Buchsbaum, 1980), white-patch retinex (Wandell, 1986), gamut-constraint methods (Finlayson, 1996; Forsyth, 1990) and neural network-based approaches (Funt *et al.*, 1996).

Unfortunately, current colour constancy methods work well on some images but they obtain very poor results when methods constraints are not completely fulfilled. Colour constancy methods have been tested for a typical computer vision task such as colour indexing in (Funt *et al.*, 1998), where results indicate that current colour constancy methods are far from being applicable to real problems.

5.2 Colour Invariants

Another possibility to achieve the scene illumination invariance is to define image descriptors based on measures which are independent from the intensity and/or the spectral composition of the illuminant, instead of estimating the illuminant of the scene.

Illuminant-invariant measures are usually defined in terms of relationships between image pixels or channels. Hence, the typical image structure is lost since the result of the measure is not normally directly related to the RGB values in the original image. However, the loss of the individual pixels colour information is not important if the illumination-invariant descriptor provides the appropiate information for the concrete task in which the invariant measure is applied.

Hence, an illuminant-invariant measure f is a function:

$$f : \mathbf{R}^{3 \times n} \longmapsto \mathbf{R}^m \tag{23}$$

where f has the property of being illuminant-independent, n is the number of pixels in the input image and m is the dimension of the result. Let $\mathcal{I}_1$ and $\mathcal{I}_2$ be two colour images of the same scene viewed under two different illuminants, then:

$$f(\mathcal{I}_1) = f(\mathcal{I}_2) \tag{24}$$

That kind of measures have been applied to different tasks such as object recognition (Funt and Finlayson, 1995; Finlayson *et al.*, 1996) and texture representation (Healey and Wang, 1995).

5.3 Colour Normalizations

Colour normalizations make a representation change of the image to another space which is not affected by illuminant variations or scene geometry. Thus, two images of the same scene seen under different illuminants should be represented by two normalized images with similar chromatic properties.

Just like invariant measures, colour normalizations do not estimate the illuminant of the scene, but they tend to remove the dependence of the image from the illumination conditions. However, while illuminant-invariant measures do not always output an image, the result of a colour normalization is another image but represented in a different space which have the property of being intensity and/or colour illuminant independent.

The transformation of the representation space from RGB to the new space is calculated only using the image information, that is, given an image $\mathcal{I}$,

$$\forall \overline{p}_i \in \mathcal{I}, \overline{d}_i = \mathcal{N}orm(\overline{p}_i, \mathcal{I}) \tag{25}$$

where $\overline{d}_i$ is the representation of pixel $\overline{p}_i$ in the new space and $\mathcal{N}orm$ is the normalization function:

$$\mathcal{N}orm : \mathbf{R}^3 \times \mathbf{R}^{3 \times n} \longmapsto \mathbf{R}^3 \tag{26}$$

As image variations are mainly due to intensity and chromaticity illuminant variations, intensity and chromaticity image normalizations have been widely used in colour computer vision.

Pixel based normalization

The aim of the pixel based normalization (Swain, 1990) is to remove the variation due to intensity changes. To achieve it, the image acquisition system is assumed to have a linear response. Therefore, when light intensity is scaled by a factor s, the image is scaled by the same factor on its three channels. That is, each pixel (r, g, b) under the original illuminant becomes (sr, sg, sb).

The pixel based normalization consists on dividing each pixel of the image by the sum of its three RGB values:

$$\left(\frac{sr}{sr+sg+sb}, \frac{sg}{sr+sg+sb}, \frac{sb}{sr+sg+sb}\right) \tag{27}$$

Hence, factor s is cancelled:

$$\left(\frac{r}{r+g+b}, \frac{g}{r+g+b}, \frac{b}{r+g+b}\right) \tag{28}$$

This normalization is, indeed, a change on the space of representation. The new space does not take into account the illumination intensity since the factor s is cancelled. Moreover, pixel based normalization brings a dimensional reduction since the three new values, normally called chromaticity coordinates, sum one:

$$\frac{r}{r+g+b} + \frac{g}{r+g+b} + \frac{b}{r+g+b} = 1 \tag{29}$$

Hence, each one of the three chromaticity coordinates can be obtained from the other two.

Channel based normalization

The objective of the channel based normalization (Finlayson *et al.*, 1996) is to deal with the illuminant colour changes. In that case, it is assumed a Von Kries model for the chromatic adaptation. This model can be considered valid in computer vision if we suppose we are working with narrow-band response cameras. In that case, illuminant chromaticity variation can be represented by scaling each one of the pixel's channels (r, g, b) by a different factor, this is, $(\alpha r, \beta g, \gamma b)$.

Channel based normalization works as follows:

$$\frac{\alpha \mathbf{R}}{\alpha \sum_{i=1}^{N} R_i} = \frac{\mathbf{R}}{\sum_{i=1}^{N} R_i} \tag{30}$$

$$\frac{\beta \mathbf{G}}{\beta \sum_{i=1}^{N} G_i} = \frac{\mathbf{G}}{\sum_{i=1}^{N} G_i} \tag{31}$$

$$\frac{\gamma \mathbf{B}}{\gamma \sum_{i=1}^{N} B_i} = \frac{\mathbf{B}}{\sum_{i=1}^{N} B_i} \tag{32}$$

where $\mathbf{R}$, $\mathbf{G}$ and $\mathbf{B}$ represent the image channels and R_i, G_i and B_i represent the i-th pixel of the corresponging channel.

After the normalization, the pixel's new values do not depend on the illuminant chromaticity since factors α, β and γ are cancelled.

Comprehensive colour normalization

In the real world, it will be quite normal to find illuminant intensity and chromaticity variations at the same time. The problem with the above normalizations is that none of them is able to deal with both variations at the same time. In order to solve this problem, comprehensive colour normalization (Finlayson *et al.*, 1998) iteratively applies pixel and channel based normalizations until each normalization stage is idempotent, that is when each normalization does not affect the image. The algorithm always converges and the solution provided is unique. Comprehensive normalization has been evaluated and it has been demonstrated that it obtains better results than colour constancy methods for common computer vision tasks such as object recognition (Finlayson *et al.*, 1999). This algorithm is explained in more detail in section 6.1.

6 A method for Colour Naming

One of the most common visual tasks is colour naming. This process consists in the labelling of a scene region with a colour name and it is easily done by the human visual system. Nonetheless, colour naming is a complex problem, from a computational point of view, due to the variation a scene can suffer as a consequence of the illumination conditions. The mechanism used by the human visual system to implement colour naming is not known yet. Although there exist some models on how to assign colour names to input stimulus (Uttal, 1973), it is still an open research topic.

However, what seems to be clear is that the chromaticity adaptation plays an important role in the colour naming process performed by the human visual system. since the same surface under a wide range of illuminants is always perceived as being of the same colour.

The colour naming method that we propose in this paper is based on the application of a normalization technique in order to avoid the illuminant and sensor influence on the scene. The use of this normalization implies to do the following assumption:

Assumption: *Two similar colours under different illuminant conditions always have aproximately similar normalized coordinates.*

Above assumption implies to consider a normalization technique as a colour constancy technique. To be able to consider it we have to stablish some restrictions on the processed scenes. These are:

- The illuminant and sensor models are those assumed by the normalization process.
- The scene context does not vary dramatically between images, that is, there is a common background in all the images.

In section 7, we will see that these constraints are fullfilled in our application. Hence, in the following subsections, we explain the normalization technique used in our algorithm and give a basic scheme of our colour naming method.

6.1 Comprehensive Colour Image Normalization

In the definition of the comprehensive colour image normalization (Finlayson *et al.*, 1998) two important assumptions are made. On one hand, it is assumed that the image acquisition system has a linear response. Therefore when light intensity is scaled by a factor s, the image is scaled by the same factor on its three channels. Thus, each pixel (r, g, b) under the original illuminant becomes (sr, sg, sb).

The second assumption is to consider a Von Kries model for the chromatic adaptation. This model can be considered valid in computer vision if we suppose we are working with narrow-band response cameras. In that case, illuminant chromaticity variation can be represented by scaling each one of the pixel channels (r, g, b) by a different factor, this is, $(\alpha r, \beta g, \gamma b)$.

If we put the N pixels of an image into a matrix $\mathcal{I}$ of dimension $N \times 3$, where each row are the values (r, g, b) of an image point, then we can represent the model of the image colour variation as:

$$\mathbf{D}^s \mathcal{I} \mathbf{D}^c$$

where $\mathbf{D}^s$ represent the $N \times N$ diagonal matrix of the intensity variation factors on the whole image (lighting geometry) and $\mathbf{D}^c$ represents the 3×3 diagonal matrix of the colour illuminant variation. This expression can be seen as:

$$\begin{pmatrix} s_1 & 0 & \dots & 0 \\ 0 & s_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & s_N \end{pmatrix} \mathcal{I} \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix}$$

To avoid the illuminant intensity dependence, the most common transformation is to pass to chromaticity coordinates space:

$$\frac{r}{r+g+b}, \frac{g}{r+g+b}, \frac{b}{r+g+b} \tag{33}$$

Hence, a function $R()$ which normalizes each pixel of the image can be defined. Function $R()$ follows equation (34).

$$R(\mathcal{I})_{i,j} = \frac{\mathcal{I}_{i,j}}{\sum_{k=1}^{3} \mathcal{I}_{i,k}} \tag{34}$$

where subscripts i, j are referred to the ij-th element of a matrix. It must be noticed that after applying that normalization, each row of matrix $\mathcal{I}$ sums 1. Thus, the total sum of the pixels in the image is N.

To avoid the dependence from the illuminant colour, a normalization over each one of the channels can be applied according to equations (35), (36) and (37).

$$\frac{\alpha N/3\mathbf{R}}{\alpha \sum_{i=1}^{N} R_i} = \frac{N/3\mathbf{R}}{\sum_{i=1}^{N} R_i} \tag{35}$$

$$\frac{\beta N/3\mathbf{G}}{\beta \sum_{i=1}^{N} G_i} = \frac{N/3\mathbf{G}}{\sum_{i=1}^{N} G_i} \tag{36}$$

$$\frac{\gamma N/3\mathbf{B}}{\gamma \sum_{i=1}^{N} B_i} = \frac{N/3\mathbf{B}}{\sum_{i=1}^{N} B_i} \tag{37}$$

The factor $N/3$ which multiplies each fraction of equations (35), (36) and (37) is applied to obtain a $N/3$ sum on each column. Thus the total sum of the pixels from image $\mathcal{I}$ is N.

As was done before with the intensity normalization, a function $C()$ which normalizes each column of matrix $\mathcal{I}$ is defined. Such normalization is ruled by equation (38).

$$C(\mathcal{I})_{i,j} = \frac{N/3\mathcal{I}_{i,j}}{\sum_{k=1}^{N} \mathcal{I}_{k,j}} \tag{38}$$

The comprehensive colour image normalization makes the two normalizations with an iterative algorithm that always converges and has a unique solution (Finlayson *et al.*, 1998). The steps of the algorithm are:

1.	$\mathcal{I}_0$	=	$\mathcal{I}$	*Initialization*
2.	$\mathcal{I}_{i+1}$	=	$C(R(\mathcal{I}_i))$	*Iteration step*
3.	$\mathcal{I}_{i+1}$	=	$\mathcal{I}_i$	*Termination condition*

This algorithm succesively applies intensity and illuminant colour normalizations until the termination condition is reached, that is when both normalizations hardly affect the obtained image. Normalization algorithm may be seen as the process of finding the image representation change that fulfil both row and column constraints (rows sum one and columns sum $N/3$) at the same time.

6.2 The Colour Naming Algorithm

Once we have explained the normalization method, let us give a basic scheme of our method, based on two main phases: learning and naming.

Learning.

The learning step has the goal of obtaining a chromaticity diagram tessellated according to the regions defined by the points of the diagram which have the same colour label. The input to the learning step must be a set of images containing different regions labelled with the name of the predominant colour in the region. Thus, the learning process performs in this way:

1. For each image in the learning set:

1.(a) Compute the normalized representation of the image.

1.(b) For each labelled region in the image, map the average of the normalized coordinates of the region onto a chromaticity diagram. (For the rest of this paper, we will refer to it as the 'Normalized Chromaticity Diagram').

2. Compute the convex hull of each set of samples labelled with the same colour name.

3. Tessellate the normalized chromaticity diagram according to the convex hulls computed in 2.

Naming.

The naming process of a region from an image is a simple mapping between the region and a colour name. This process performs as follows:

1. Compute the normalized representation of the image which contains the region to be named.

2. Average the normalized coordinates of the region.

3. Give the colour label corresponding to the average computed in 2 within the normalized chromaticity diagram.

Once we have introduced the proposed method, we will explain, in more detail, the construction process of the chromaticity diagram.

6.3 Constructing a normalized chromaticity diagram

As has been seen in the previous sections, the application of the Comprehensive Colour Image Normalization implies a representation change from the RGB original image coordinates to chromaticity coordinates. From (33) can be seen that:

$$\frac{r}{r+g+b} + \frac{g}{r+g+b} + \frac{b}{r+g+b} = 1 \qquad (39)$$

As each one of the resulting coordinates can be obtained from the other two, it is not necessary to work in a three-dimensional space. The effects of the normalization can be observed over a chromaticity diagram which is obtained from the intersection of the rgb space with the $r+g+b=1$ plane (Figure 8).

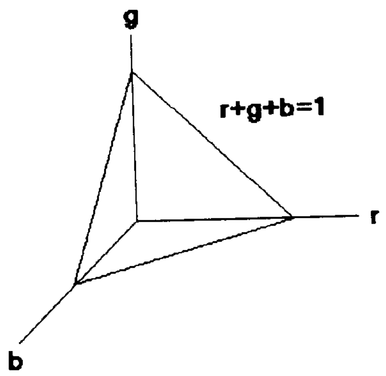


Figure 8: The chromaticity diagram is obtained by intersecting the rgb space with the $r+g+b=1$ plane

To create the normalized chromaticity diagram from the normalized images a set of 55 different colour patches samples were used as learning set. We selected the most common colours for our specific application and different samples of each colour to be named were included in the learning set. Each sample was acquired in eight different conditions and was labeled with one of the eight selected names to be the basic colours, these are, grey, green, blue, purple, red, brown, orange and yellow. Notice that, as colour normalization implies a loose of the intensity information, all the colours in the grey scale are labelled as 'grey'.

Following the learning step of our algorithm, for each sample we compute the average of the region coordinates in order to obtain the representative point of that sample in the normalized chromaticity diagram.

Once all the representatives of the samples in the learning set have been represented in the normalized chromaticity diagram, it can be tessellated according to the regions defined by the convex hulls of each set of points labelled with the same name. The resulting normalized diagram is shown in figure (9).

As was seen in section 6.1, the comprehensive image normalization is context dependent. Hence, the regions of the

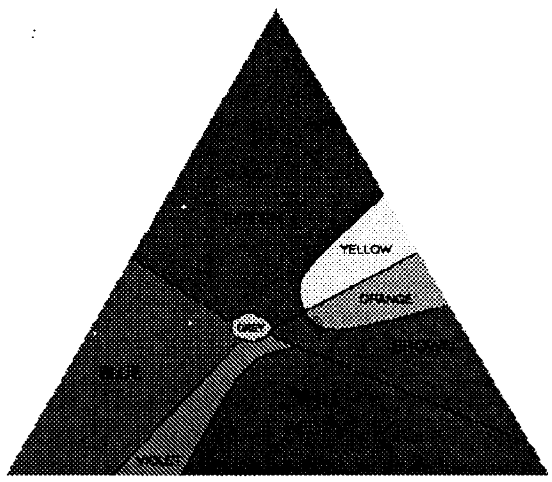


Figure 9: Normalized chromaticity diagram obtained after the application of the learning step of the colour naming algorithm

normalized chromaticity diagram are dependent on the image content. That is, the space occupied by each colour can vary if the learning set is changed.

Our method has obtained very good results in controlled experiments with a Macbeth colourchecker (above 97% of correct labellings). However, in order to evaluate the performance of the method in a real problem, it has been applied to a real surveillance problem. The results obtained in that case are presented in the next section.

7 Colour naming on a real application

The application in which our colour naming method has been tested is a computer system to automatically describe people in natural language and index that information and the person's identity in a database. The description is based on physiognomical and clothing features and it is done in the same terms as it would be done by a human observer. The goal of this application is to obtain an accurate description of the person and, thus, be able to retrieve the right identity from the database corresponding to a given description.

The system is thought to work on controlled stablishments or restricted areas where people must give their identification before coming in. However, in such stablishments, if anyone has a problematic behavior, it is normally quite difficult to obtain a fast identification of the subject. To overcome this problem, the application makes a personal description based on the content of an image of the subject which is acquired when the person is giving his/her identification. The image and the text-description of the person is attached to the corresponding identity data and it is indexed in a database. Hence, if there is any problem with anyone inside the stablishment, the security staff can achieve a fast identification of the subject by making a query on the database. In that database, a query consists on the description of a person and the system will retrieve the images corresponding to the most similar people to the given description. The final identification among the retrieved images will be done by the security staff and the system will provide the identity data of the final election.

The first step of the system is to obtain a good image of the person to be described. With this objective, the camera is fixed behind the reception desk and images are acquired when the person is standing up in front of the reception desk. Hence, the system is able to obtain frontal images of the person and with the subject approximately placed on the same position of the image. An scheme of the acquisition system can be seen in figure 10.

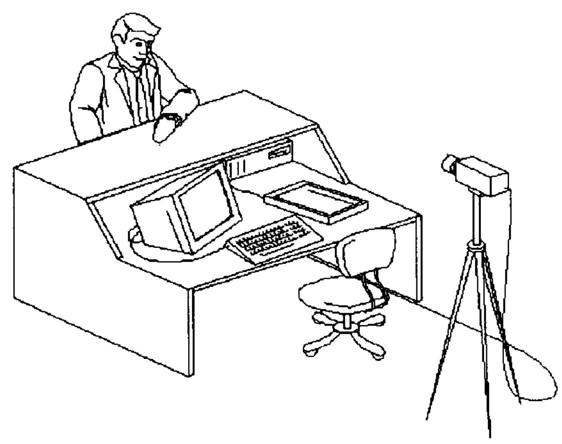


Figure 10: Scheme of the working environment used to acquire the subject's image. The camera is fixed behind the reception desk in order to obtain a frontal image of the subject with the same background for all the images

The way the images are acquired (with a fixed camera) allows the system to accomplish the restriction specified in section 6, since all the acquired images present a similar background and the only important variation in the image is the person to describe. Thus, the context variation is small and the normalized chromaticity diagram is quite stable with well defined regions for the different colours.

Once the image has been acquired, it is processed in order to obtain the subject's description. Apart from clothing colour, other features considered by the system are height, hair colour, presence of glasses, beard or moustache, ... However, in this work, we only deal with the problem of the clothing colour description.

Using the real environment, a set of 165 images were acquired as a test set. From these images a total of 330 colour regions were selected. In the final system, the regions to be named will be automatically segmented before the application of the colour naming process over each region. In figure (11) some examples of selected regions are shown.

All the selected regions were labelled with a colour name by the security staff and that labelling was compared with the results obtained with the different methods that were tested in order to obtain the rate of correct labellings achieved by each method. However, any result has to consider the subjec-

Figure 11: Examples of regions selected on the acquired images. As can be seen in the images, there are few restrictions over the working environment illumination conditions

tivity of colour naming problem, since a region is not always named with the same colour term by two different people.

In order to see the improvement apported by the use of our colour naming method based on normalized chromaticity coordinates, the results obtained were compared with the ones obtained by different approaches to the problem. The other methods used in this test were based on the use of RGB coordinates and chromaticity coordinates.

In the case of the RGB-based method, the labelling of the regions was done following a nearest neighbour strategy: the RGB average of the region was labelled with the colour name of the nearest sample from the learning set.

In the case of the use of chromaticy coordinates (CC), a chromaticity diagram was also built. Thus, the region labelling was done using both classification methods: the nearest neighbour and the mapping over the regions defined in the chromaticity diagram.

Finally, the results obtained from the application of the normalized chromaticity diagram were compared to the results obtained by the application of the nearest neighbour rule over the normalized chromaticity coordinates (NCC).

The results obtained by the different methods are shown on table (1). As can be seen on the table, the results obtained by our method considerably improve previous ones based on RGB and chromaticity coordinates (CC) methods. In the left column, that corresponds to the nearest neighbour strategy, it can be noticed that the use of normalized chromaticity coordinates performs more than 10% better than chromaticity coordinates and more than 30% better than RGB nearest neighbour. The normalized chromaticity coordinates also improve very much the results obtained mapping over a chromaticity diagram based only on chromaticity coordinates (right column of the table). Finally, it can be seen that our colour naming algorithm also improves the nearest neighbour classification in the normalized chromaticity coordinates space (18% approximately) achieving a correct naming rate of 83.33% .

	Nearest Neighbour	Chromaticity Diagram
RGB	34.24%	
CC	54.54%	63.33%
NCC	65.45%	**83.33%**

Table 1: Results of colour naming experiments for 330 regions. The nearest neighbour strategy assigned the label of the learning sample that had minimum euclidean distance to the average values of the test sample. The right column of the table correspond to the mapping of the average values of the test region over the Chromaticity Diagram in the colour space of choice

8 Conclusions and future work

The contribution of this work is twofold. The first one is to give a brief introduction and a review about colour perception. The second one is to propose a new computational method for colour naming that is tested on a real application.

Firstly, we explain colour as a visual cue and we give two points of view: biological and physical. We conclude this first review with the tricromatic mathematical formulation of colour. Secondly, we show a brief review of techniques to deal with colour in computer vision. We have grouped the different approaches in three different groups of techniques: colour constancy methods, invariant measurements and normalization techniques. All of them try to solve the variability of colour representations in real problems, trying to implement the chromacity adaptation performed by the human visual system.

The second part of the paper is devoted to show the application of a normalization technique to solve the colour naming visual task. The goal of this task is to map an image region to a colour label. The proposed algorithm is based on two steps: learning and naming. The learning step constructs a normalized chromaticity diagram from a set of normalized images and the naming step maps the normalized coordinates to a label of the constructed diagram. In order to apply this algorithm, several constraints have to be assumed: linear and narrowband sensors, non-radical illuminant variations and a permanent background in the scene. Assuming all this conditions the algorithm has been applied to a real system for a surveillance problem and some results have been showed.

As a future work, all the experiments should be tested on a larger set of images. We also have to validate the robustness of the method by testing the range of condition variations that allows to correctly work with a given normalized chromaticity diagram. Finally, we should make a more suitable comparison with other colour naming methods.

Acknowledgments

This work is supported by projects 2FD97-1800 and TAP98-0618 of Spanish CICYT, by Universitat Autonoma of Barcelona FIPD grant and by Casinos de Catalunya.

References

G. Buchsbaum, A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310:1–26, 1980.

J. Crowley and **F. Berard**, Multi-modal tracking of faces for video communications. In *IEEE Proceedings of CVPR'97*, pages 640–645, 1997.

A. Ebenhoh and **H. Hemminger**, Scaling of color sensation by magnitude estimation: A contribution to opponent-colors theory. *Biological Cybernetics*, 39(3):227–237, 1981.

G. Finlayson, Color in perspective. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10):1034–1038, 1996.

G. Finlayson, S. Chatterjee, and **B. Funt**, Color angular indexing. In *Proceedings of the 4th ECCV'96*, pages 16–27, 1996.

G. Finlayson, M. Drew, and **B. Funt**, Diagonal transforms suffice for color constancy. In *Proceedings on the 4th ICCV '93*, pages 164–171, 1993.

G. Finlayson, B. Schiele, and **J. Crowley**, Comprehensive colour image normalization. In *Proceedings of the 5th ECCV '98*, pages 475–490, 1998.

G. Finlayson, B. Schiele, and **J. Crowley**, Perfect colour constancy vs colour normalization for object recognition. In *Proceedings of EUROPTO/SPIE Conference on Colour and Polarisation Techniques in Industrial Inspection (Printing)*, 1999.

D. Forsyth, A novel algorithm for color constancy. *International Journal of Computer Vision*, 5(1):5–36, 1990.

B. Funt, K. Barnard, and **L. Martin**, Is machine colour constancy good enough? In *Proceedings of the 5th ECCV '98*, pages 445–459, 1998.

B. Funt, V. Cardei, and **K. Barnard**, Learning color constancy. In *IS&T Fourth Color Imaging Conference*, pages 58–60, 1996.

B. Funt, M. Drew, and **M. Brockington**, Recovering shading from color images. In *Proceedings of the 2nd ECCV '92*, pages 124–132, 1992.

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

G. Healey and **D. Slater**, Global color constancy: Recognition of objects by use of illumination invariant properties of color distributions. *Journal of the Optical Society of America*, 11(11):3003–3010, 1994.

G. Healey and **L. Wang**, Illumination-invariant recognition of texture in color images. *Journal of the Optical Society of America*, 12(9):1877–1883, 1995.

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

M. Levine, *Vision in Man and Machine*. McGraw-Hill, 1985.

L. Maloney and **B. Wandell**, Color constancy: a method for recovering surface spectral reflectance. *Journal of the Optical Society of America*, 3(1):29–33, 1986.

M. Swain, *Color Indexing*. PhD thesis, University of Rochester, 1990.

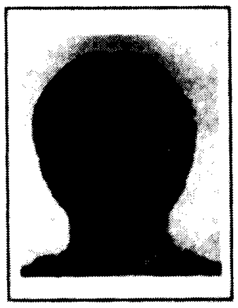
M. Swain and **D. Ballard**, Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

W. Uttal, *The Psychobiology of Sensory Coding*. Harper&Row, 1973.

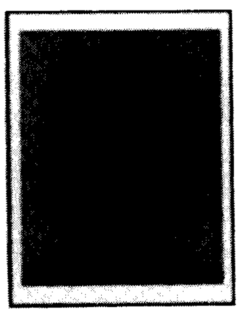
B. Wandell, Analysis of the retinex theory of color vision. *Journal of the Optical Society of America*, 3(10):1651–1661, 1986.

J. Worthey and **M. Brill**, Heuristic analysis of Von Kries color constancy. *Journal of the Optical Society of America*, 3(10):1708–1712, 1986.

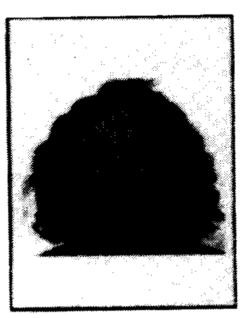
G. Wyszecki and **W. Stiles**, *Color science: concepts and methods, quantitative data and formulae*. John Wiley & Sons, 2nd edition, 1982.



***Robert Benavente** is a research fellow in the Computer Science Department of the Universitat Autònoma de Barcelona and is affiliated of the Computer Vision Center. He received his MSc degree in Computer Vision from the Computer Vision Center in 1999. In addition, he participates in the development of industrial applications related with quality control processes based on computer vision. His research interests include colour perception and representation, and colour naming.*



***Ramon Baldrich** is an Associate Professor at the Computer Science Department of the Universitat Autònoma de Barcelona and is affiliated to the Computer Vision Center as a researcher. He received his MSc degree in computer science from the Universitat Autònoma de Barcelona in 1996. He is currently pursuing a PhD in Computer Science and is involved in the development of several vision systems for*
Industrial quality control. His research interests include texture representation and colour perception.



***Maria Vanrell** is an Associate Professor of Artificial Intelligenge in the Computer Science Department of the Universitat Autònoma de Barcelona. She received an MSc in Graphics, Computer Vision and Artificial Intelligence from the UAB. In 1996, she received a PhD at the UAB for work in texture analysis and representation. She is an affiliate researcher in the Computer Vision Center, where she leads several industrial projects on colour vision. Her research interest includes texture and colour representation and artificial intelligence.*