Universitat Autònoma de Barcelona

Computational framework for the white point interpretation based on nameability

A dissertation submitted by **Francesc Tous Terrades** at Universitat Autònoma de Barcelona to fulfil the degree of **Doctor en Informàtica**.

Bellaterra, May 26, 2006

Director: **Dr. Maria Vanrell i Martorell**
Universitat Autònoma de Barcelona
Dept. Ciències de la Computació & Computer Vision Center
Co-director: **Dr. Ramon Baldrich i Caselles**
Universitat Autònoma de Barcelona
Dept. Ciències de la Computació & Computer Vision Center

Centre de Visió per Computador

This document was typeset by the author using LaTeX 2_ε.

The research described in this book was carried out at the Computer Vision Center, Universitat Autònoma de Barcelona.

als meus pares

Colour is a power which directly influences the soul. Colour is the keyboard, the eyes are the hammers, the soul is the piano with many strings. The artist is the band that plays, touching one key or another to cause vibrations in the soul.

Wassily Kandinsky

Acknowledgements

Previous to the presentation of this work, I would like to acknowledge some people that, in many ways, have helped me in its realisation. I will use different languages for a closer acknowledgement.

Primerament agraeixo a la Maria Vanrell, la meva directora, tot el suport i l'esforç continu durant aquests anys. Sobretot per confiar en mi en un principi, i per l'empenta final, que entenc que ha estat esgotador. Una gran part d'aquest treball, sinó tot, és seva. Sincerament, moltes gràcies per tot.

També vull agrair a en Ramon Baldrich, no diré els ànims, perquè qui el conegui mínimament sabrà que això li costa una mica, però sí tot l'esforç i ajuda constant que m'ha donat. Ha estat un contrapunt perfecte a la Maria, i fins i tot en algun moment s'ha posat de part meva. Sense la seva inestimable ajuda això no hauria pogut tirar endavant. Moltes gràcies.

Vull donar les gràcies a tota la gent del CVC pel suport i ànims. Sobretot a aquells que fan que el centre funcioni dia a dia, la seva tasca tant de gestió com de suport ha estat imprescindible perquè ara mateix estigui acabant d'escriure aquest treball. Vull destacar en Josep Lladós per donar-me l'oportunitat d'entrar a treballar al centre i de realitzar un doctorat (per això últim no sé si donar-li les gràcies o culpar-lo). També vull fer una menció especial de la Mari, pel seu suport, ànims i professionalitat, i, sobretot, de la Montse per la seva ajuda i ànims constants i per ser una font d'energia positiva de la qual, crec, tot el CVC se'n nodreix.

I would like to acknowledge Prof. Graham Finlayson for welcoming me at UEA and also for the interest, advice and kind help he has provided me the two times I have been there. Also, I would like to acknowledge Steve Hordley for his help and advice. I remember the rest of the people at the colour group: Jeff, Michal, Julia, David, Clement and Elisabetta. Also the important italian comunity, and non-italians, in Norwich: Luca, Sergio (crazy man), Sergio2, Johan, Onema, Marco, Federica, Key and Alessandra. Last, but not least, I would like to acknowledge Peter Morovic for many things: welcome, help, advice, slokavian cuisine, etc. and most important, friendship. Life is hard when you go abroad, but if you live with a happy family it's not that bad.

Vull tornar al CVC per agrair a la gent del grup de color i textura el suport,

ànims i idees durant aquests anys de seminaris intermitents: Robert, Xavier, Susana, Eduard i Anna. A l'Eduard li vull donar les gràcies per l'ajuda en aquest darrer tram. I a l'Anna li hauria d'agrair moltes coses, entre elles suportar el mal gust de les meves bromes, però només li agrairé haver-me presentat el Javier, el Sam, el Bruno, la Daniela, etc. i deixar-me compartir amb ells una experiència molt especial.

No em vull oblidar la gent amb qui he compartit sala i/o bar durant aquests anys. Sobretot a l'Agnés, companya de batalles que m'ha donat un suport i ànims incondicionals, molt per sobre del que em mereixo. També vull recordar la Carme (there is a light that will never go out..), en Jaume Garcia (amb qui espero poder compartir algun altre correllengua), en Batlle, en David, l'Àgata, l'Oriol, l'Aura, en Misael, en Xevi, l'Anton (home! tu per aquí!), l'Eva (per saber perdonar i oblidar), i a la resta que potser em deixo.

Gracias sinceras, con los dedos en las manos, a Jordi Sequero, de los Sequero de toda la vida, Lidia, Mónica y Vicente. També a l'Agustí, l'Eduard, en Gustau, la Sònia, l'Enric, la Marta, en Dalmau, la Yaiza, en Pol, en Torra i en Miquel. A tots, pel suport i ànims incondicionals.

No em vull deixar a en Chris Peterson, Ignatius J. Reilly i Morrissey per ser font d'inspiració contínua. I a Gomaespuma per fer-me riure dia sí, i dia també.

Finalment vull acabar donant les gràcies a la meua família. A la meua iaia Teresa, al meu iaio Luís, als meus tios i a les meues cunyades Vicenta, Sandra i Ceci. Als meus germans Vicent, Luís i David pels seus ànims constants. També als meus nebots David, Josep i Anna, que m'han fet molt feliç estos últims anys. I sobretot, done les gràcies a ma mare per animar-me, ajudar-me, cuidar-me i soportar-me en els moments més difícils, que no és poc. Jo no ho hauria fet.

Moltes gràcies a tots.

Abstract

In this work we present a framework for white point estimation of images under uncalibrated conditions where multiple interpretable solutions can be considered. In this way, we propose to use the colour matching visual cue that has been proved as related to colour constancy. The colour matching process is guided by the introduction of semantic information regarding the image content. Thus, we introduce high-level information of colours we expect to find in the images. Considering these two ideas, colour matching and semantic information, and existing computational colour constancy approaches, we propose a white point estimation method for uncalibrated conditions which delivers multiple solutions according to different interpretations of the colours in a scene. However, we present the selection of multiple solutions which enables to obtain more information of the scene than existing colour constancy methods, which normally select a unique solution. In this case, the multiple solutions are weighted by the degree of colour matching between colours in the image and semantic information introduced. Finally, we prove that the feasible set of solutions can be reduced to a smaller and more significant set with a semantic interpretation.

Our study is framed in a global image annotation project which aims to obtain descriptors which depict the image, in this work we focus on illuminant descriptors. We define two different sets of conditions for this project: (a) calibrated conditions, when we have some information about the acquisition process and (b) uncalibrated conditions, when we do not know the acquisition process. Although we have focused on the uncalibrated case, for calibrated conditions we also propose a colour constancy method which introduces the relaxed grey-world assumption to produce a reduced feasible set of solutions. This method delivers good performance similar to existing methods and reduces the size of the feasible set obtained.

Resum

En aquest treball presentem un marc per a l'estimació del punt blanc en imatges sota condicions no calibrades, on considerem múltiples solucions interpretades. D'aquesta manera, proposem la utilització d'una cua visual que ha estat relacionada amb la constància de color: aparellament de colors. Aquest aparellament de colors està guiat per la introducció d'informació semàntica referent al contingut de la imatge. Així doncs, introduïm informació d'alt nivell dels colors que esperem trobar en les imatges. Tenint en compte aquestes dues idees, aparellament de colors i informació semàntica, i les aproximacions computacionals a la constància de color existents, proposem un mètode d'estimació de punt blanc per condicions no calibrades que lliura múltiples solucions, en funció de diferents interpretacions dels colors d'una escena. Plantegem l'extracció de múltiples solucions ja que pot permetre extreure més informació de l'escena que els algorismes clàssics de constància de color. En aquest cas, les múltiples solucions venen ponderades pel seu grau d'aparellament dels colors amb la informació semàntica introduïda. Finalment demostrem que la solució plantejada permet reduir el conjunt de solucions possibles a un conjunt més significant, que és petit i fàcilment interpretable.

El nostre estudi està emmarcat en un projecte d'anotació d'imatges que pretén obtenir descriptors que representen la imatge, en concret, els descriptors de la llum de l'escena. Definim dos contextos diferents per aquest projecte: condicions calibrades, quan coneixem alguna informació del sistema d'adquisició, i condicions no calibrades, quan no coneixem res del procés d'adquisició. Si bé ens hem centrat en el cas no calibrat, pel cas calibrat hem proposat també un mètode computacional de constància de color que introdueix l'assumpció de 'món gris' relaxada per a generar un conjunt de solucions possibles més reduït. Aquest mètode té un bon rendiment, similar al dels mètodes existents, i redueix el tamany del conjunt de solucions obtingut.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this chapter, we present colour as an important cue to deal with visual tasks for image understanding. Then, we present an image annotation project in which we will aim to obtain illuminant invariant descriptors from the coloured surfaces of an image. We propose a frame of different low-level visual tasks concerning this annotation process, and we will focus on the module that aims to obtain these invariant descriptors, labelled as white point estimation. This module directly influences the colour naming task, which subsequently aims to give a name to the colours in the image. We will propose to introduce information of this colour naming process to guide the white point estimation process and, therefore, to obtain useful descriptors from the image. Finally, we present the thesis outline that will follow this dissertation.

Colour plays an important role in the understanding of the world that we see. When looking at a real scene, the colour of an object sometimes defines its properties and allows us to interact within it. For instance, the colour of a traffic light indicates whether we must stop or go on. Also, colour is an important feature to identify types of mushrooms and to distinguish the edible from the toxic ones. More examples could be found to depict the significance of colour to understand the world and interact with it. Thus, colour is a powerful visual cue that delivers important information of a scene and, in this way, it helps to infer properties of its content. Colour perception in the human visual system is a trichromatic process base on the three different types of photoreceptors we have in the retina. If we had a single type of photoreceptors, then we would live in a grey world. Therefore we could only perceive intensity of some light, which is true for some animals that are not able to process the multispectral information of the light. The interaction with the world would be harder, since our visual system should have to further hypothesise about the real world than actually has to. In fact, this situation happens under small amounts of light, normally at night, when only one type of photoreceptors, rods, are excited, and it is more complicated to perform some tasks, since we get less information of what surrounds us.

Therefore, colour perception makes the interaction with the real world easier. Also it is considered to make life more exciting.

In this work we study how colour information can be considered in computer vision, which pursues the final goal of automatic image understanding. Computer vision is a subfield of artificial intelligence which research methods which allow computers to understand image content, and focuses on theoretical studies of visual information processing [71]. Computer vision aims to understand the context in which a system is placed, using a visual sensor. Also, and therefore, computer vision aims to understand the human visual system as an information processing system. The model presented by Marr in [65] presents vision as a bottom-up process where the goal is to develop vision modules which obtain image attributes, that are able to infer information of the scene it represents. Considering that this way of working implies to solve an ill-posed problem, since an image is just a projection of the real scene, then we need some constraints to solve this kind of problems. Thus, the general bottom-up approach has not led to the design of successful vision systems and in [1] a new paradigm for computer vision is proposed: purposive vision. In this paradigm, methods are designed task-oriented and avoid creating an accurate description of the scenes. Our proposal will be given in this frame where high-level information will reduce the uncertainty of the multiple solutions. But we will come later, at the end of the thesis, to this point.

Colour is an important cue in vision and it delivers significant information about the real world. Therefore, colour should be considered if we want to improve our computer vision algorithms and go further in the computational understanding of the world, but it must be done carefully. A coloured image is usually obtained using a camera that acquires the light reflected by a given scene through three different filters, emulating the three types of photoreceptors involving colour in the human visual system, known as: red, green and blue filters. However, since light reaching the camera depends on the illumination of the scene, to obtain reliable colour image descriptors this influence from the illuminant should be removed. For this reason, colour needs a processing when used in computer vision. In this dissertation we study what does a *processing* mean, and we also present different computational approaches to deal with colour, along with some tasks in which they are used.

1.1 Colour

Colour must be carefully taken into account, since it is a perception and the physical stimulus is altered by more complex processes. The sensation of colour is produced in the human visual system, HVS from now on, through the integration of the information obtained by three kinds of photoreceptors, cones, which have different response curves (they respond to variation in light in the visible spectrum in different ways). Visible light is a small part of the electromagnetic radiation, composed of streams

of photons, which are visible by the human visual system (figure 1.1). The HVS combines the information from each type of photoreceptor to give rise to different colour perceptions for different wavelengths of light. Thus, the multidimensional information contained in the visible light reaching the retina is reduced to a three dimensional representation, which gives place to what is known as colour description. Actually, the perception of colour involves more than dimensionality reduction to a three-dimensional space. There are different works that have studied how the human visual system perceives colour [98, 51, 54, 58, 72, 57]. In [51, 54, 72] mathematical models are proposed to predict the colours perceived by a subject under given conditions. Land in [57] proposes a theory to explain how the constant perception of colours despite changes in the illumination, i.e. colour constancy, is performed.

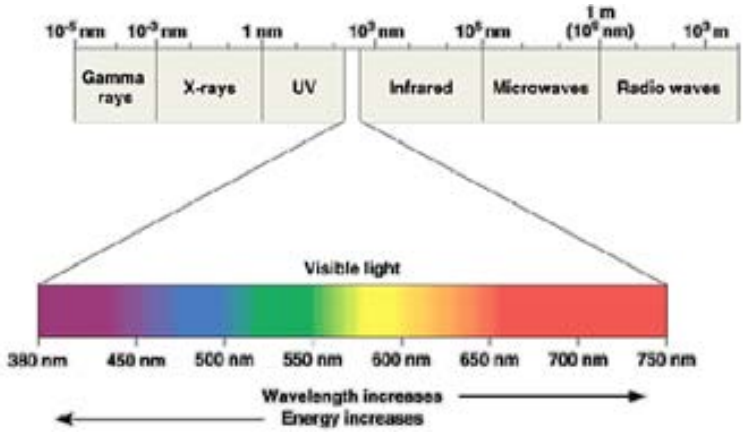


Figure 1.1: Visible light is a small part of the electromagnetic radiation, which includes from gamma rays up to radio waves, and that stimulates the photoreceptors in the human visual system through wavelenghts comprised from 400 to 700nm.

In computer vision a large amount of methods have been defined for grey-level images, since it is easier and therefore they involve less complex algorithms, and in a lot of visual tasks the intensity information is sufficient to solve the task. However, in some cases intensity information is not enough to perform some visual tasks and colour has to be considered to achieve better results. The three-dimensional representatioin of colour can be separated into two components: intensity and chromaticity. Intensity is the brightness of the colour and chromaticity is the quality of colour independent of brightness. In figure 1.2 we show a plain example of the significance of colour information. We have built two similar synthetic images consisting of three different colours: in the first image the colours have different intensity component, whereas in the second image the three colours have the same intensity value. We show the colour image and its intensity component, which corresponds to its grey-level representation. If we just consider the grey-level image, unless there is some intensity variation, no shape is recognisable, while in the colour image we can distinguish the different shapes in both images. It is clear, that if we aim to find shapes in this type of images,

colour information should be considered to avoid the problem of images with constant intensity.

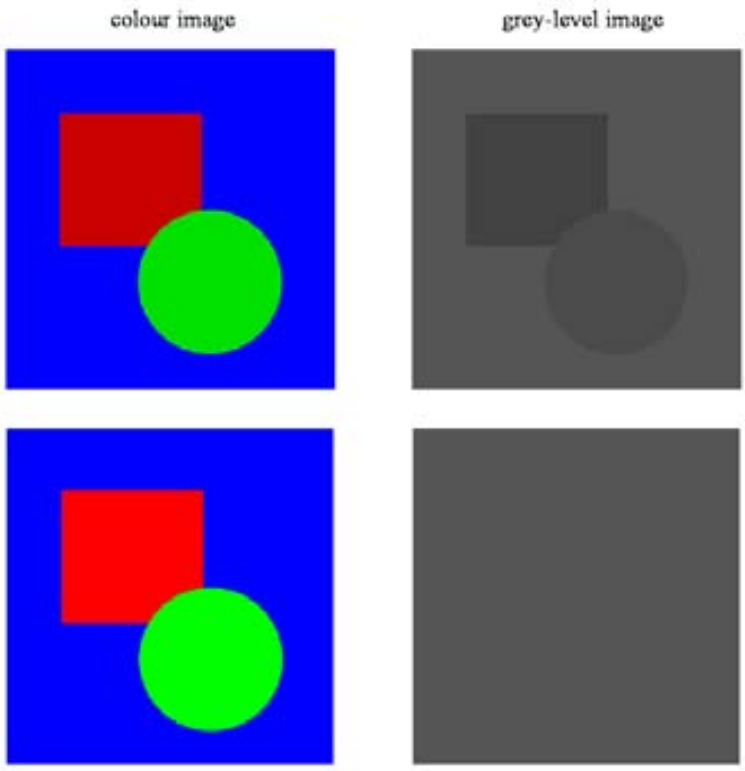


Figure 1.2: Two similar synthesised images consisting of three colours along with its corresponding grey-level representation (where just the intensity component is presented): in the first image the three colours have a different intensity component, but in the second image the intensity component is the same for the three colours and no shape is recognisable if we just consider intensity information.

On the other hand, the colour of a digital image is always depending on the illuminant of the scene. In order to deal with this dependency, different approaches have been proposed. They can be classified as follows:

- Invariance normalisations [33, 48].
- Colour constancy methods:
  - Illuminant transformations to a reference illuminant [40, 31, 38].
  - Recovery of spectral information of illuminant and surfaces of a scene [64, 27, 18, 74].

These approaches will be presented in more detail in following chapters. Depending on the visual task we have to perform and the amount of available information about a scene, we will choose within them. For this reason, introducing colour information to an image processing method is not a trivial step. The different approaches must be analised and adapted to the task that we are dealing with.

1.2 Automatic image annotation

When we look at a picture we figure out the scene that it represents by considering the shapes and the colours in it we interpret how was the scene when the picture was taken. This is one of the goals of computer vision, to understand a scene and to be able to interact within it. A system that completely understands a scene is a highly ambitious goal, and may not be affordable. Even to understand the same as human beings understand through HVS is not manageable. For this reason, and in a similar way to what has happened in artificial intelligence, computer vision has been subdivided in different tasks, each of them performs a concrete visual task: tracking [59], object recognition [49], colour segmentation [70], shape extraction [47], etc. Each of these processes aim to create a model of a scene for a given image. This model should be a representation that enables us to interact with the real world in a useful way.

Image understanding is a high-level visual task, which comprises some of the visual tasks defined before. To interpret a scene is to describe it in terms of descriptors, i.e. to give meaning to its elements. There can be different levels of complexity in the description, depending on the descriptors used and our goal in the problem.

Image annotation is an example of scene interpretation (figure1.3). We are given an image and we are asked to detail the different shapes that can be found in it, and also colour and texture properties for each different shape. And sometimes the relations between the shapes. Automatic image annotation is a process by which a computer system automatically assigns metadata in the form of captioning or keywords to a digital image. For instance, it is used in image retrieval systems to automatically organise and locate images of interest from a database. This image annotation process can have different levels of complexity, which depend of the dictionary of descriptors that we are dealing with. We could consider a dictionary with just 'round', 'squared' and 'triangle' shapes and 'red', 'green', 'blue', 'yellow', 'black' and 'white' colours, which is a simple model, or a dictionary with 'car', 'tree', 'dog', 'house', etc. and a much larger set of colours, which is more complex, and therefore involves a more complex process to analyse images.

In any case, some processing of colour information should be performed. This processing can be performed in different ways, depending on the information we know and the goal of our system. Invariant normalisations look for illuminant invariant im-

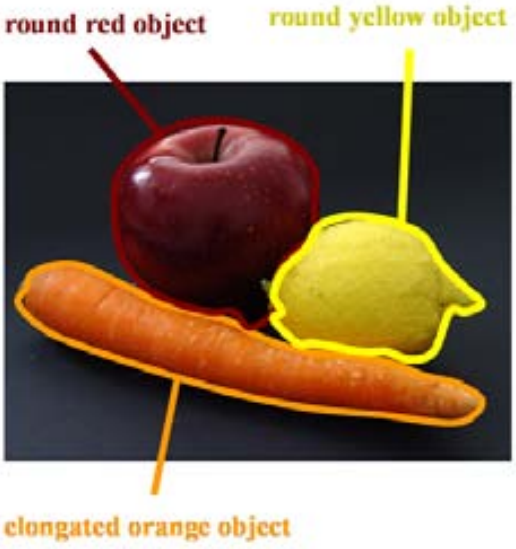


Figure 1.3: Plain image annotation goal: describe objects and colours in images.

age descriptors, and they do not need to know information of the acquisition process of the image. If we know the spectral properties of the illuminant of the scene, the sensitivity curves of the camera and the response of the camera for a white under the illuminant, colour correction can be perfectly performed. Sometimes with just the response of the camera under a white illuminant it is enough. We might know nothing of the illuminant and the white response, but know the sensitivity curves of the camera. And so on. Different methods have been proposed according to the information available. The worst case is when we do not dispose of any kind of information and we have to do some colour correction or process to be able to *trust* the colours in an image. This worst case is quite common when dealing with images from unknown origin, such as internet images. This worst case is the case we are more interested in. What can we say about the colours in an image that we do not know how has it been acquired? This is one of the questions that we will try to answer in this work.

1.3 Framework

The thesis work we are presenting here is framed in a context of automatic image annotation project. We are working in a system that aims to assign different descriptors to an image according to its content. This can be seen as an image understanting process, since we give meaning to what can be seen in it. In our system, we define four types of descriptors for the annotations of an image: illuminant, colour, texture and location. Our colour and texture group is not dealing with shape descriptors,

which deliver significant information of the images but are out of the scope of our research lines. Each of these descriptors has a set of dictionaries that can be used. For instance, a dictionary for colour could be:

$$c_1 = \{red, green, blue, yellow, ...\}$$

Illuminant might have two different dictionaries :

$$d_1 = \{sunnyday, cloudyday, tungsten, fluorescent, ...\}$$

and

$$d_2 = \{D65, A, B, C, ...\}$$

And so on. These dictionaries are not fixed and should be customised to work for concrete contexts. In figure 1.4, we show an example of the annotation aimed by our system.

Image	IdAnnot.	Illumin ant	Colour	Texture	Location	Other descriptors (shape)
	Idpolygon Idlight Idmaterial Idmaterial	- Red - -	Yellow - - -	- - Sea Sky	Left Global Bottom Top	Circle - - -
	Idlight Idobjects Idobjects Idmaterial	Daylight - - -	- - Whitish -	- - Blob-pattern Brick	Global Center Center Left&Right	- Woman Dress -
	Idtexture	-	Bluish	Semi-random &Blob-pattern	Global	-

Figure 1.4: Example of image annotation aimed by our system, where we do not deal with shape descriptors.

To simplify this automatic annotation system we have divided it in modules that perform distinct visual tasks. In figure 1.5 we present a diagram of the modules for

the low-level tasks that must be performed and the different relations between them. The input of the diagram is a digital image and the output is a set of descriptors figured out, arranged for subsequent higher-level visual processes. Each of the modules performs tasks that cannot be considered independently, since they are closely related to the other modules as it is marked in the figure.

Some work related to the local induction effects module and its relation with colour and texture representation is presented in [96]. Also, some work related to colour combined with texture information to obtain clothes descriptors is presented in [16]. However, in this work we will center in the colour constancy module (yellow box in figure 1.5), which aims to extract illuminant descriptors from the image and also to balance its colours for a better processing in the next modules. To properly perform a colour naming task it is required a colour constancy module. Also, in this work, we propose that in some cases colour naming can help to solve the colour constancy problem.

The image annotation system can be developed to work under different conditions. We define two types of contexts, according to the information that we know of the acquisition conditions:

- **Calibrated conditions:** we do know some information regarding the sensor, and the scene is, more or less, constant (e.g. a surveillance sytem that controls the entrance of a building).
- **Uncalibrated conditions:** we do not know the sensor and the scene is completely unpredictable (e.g. images from the internet).

In this work, we will consider both contexts, and we will present different computational approaches to obtain reliable colour descriptors from images independent from the illuminant conditions. For calibrated conditions, we will present methods which aim to recover the illuminant of the scene or to obtain colour invariant descriptors from the image. For uncalibrated conditions, since it is a less restricted problem, we will not look for a perfect recovery of the illuminant, but for a white point estimation, i.e. response of a white patch as it would be in the scene, which delivers meaning to the elements in the scene. In this way, the colour constancy module will be guided by the colour naming module, which will introduce high-level information of expected colours in the image, i.e. colours with a name. This colour naming information will be used in a process of colour matching. The colour naming module will not be limited to classical colour naming names, which are synthetic colours such as red, yellow and green, and different contexts will be able to be considered to define expected colours in our image annotation system. Hence, if we deal with natural images, we will be able to use a colour naming module that delivers colour names such as tree color, sky colour, etc. and it will guide the colour constancy module that we are working in to obtain different meaningful colour transformations.

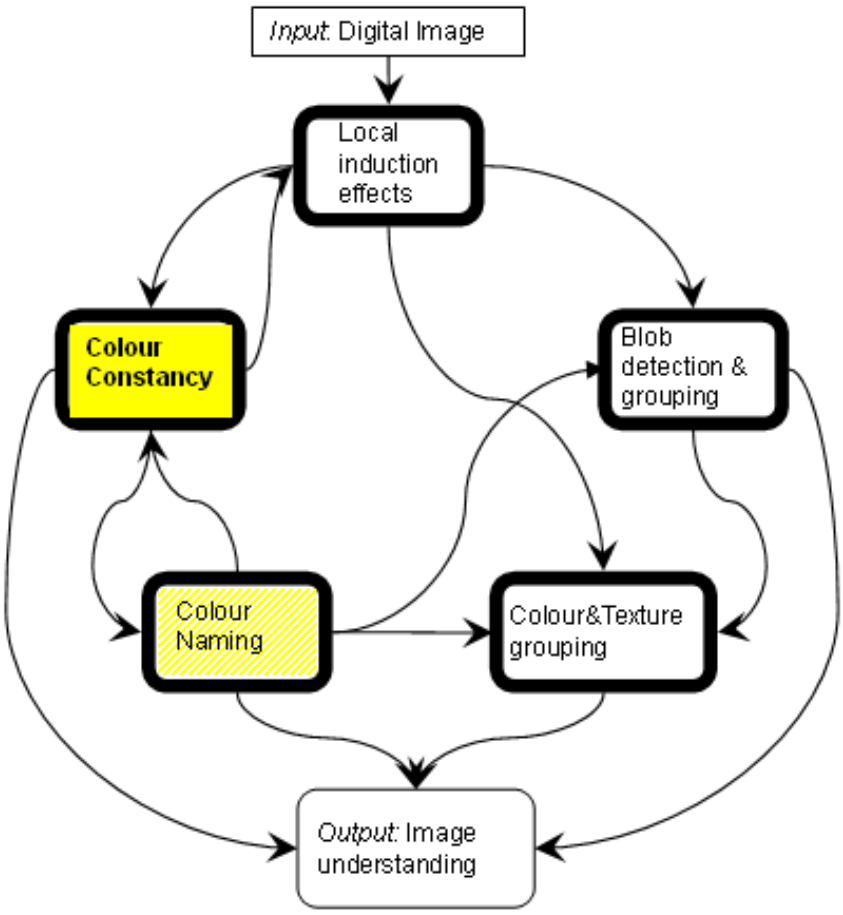


Figure 1.5: Diagram of the different modules/visual tasks of the annotation system, and their links, that take place in low-level processes. In this work we focus on the colour constancy module (yellow box), and in the introduction of colour naming information to help to solve it.

1.4 Thesis outline

This dissertation has been structured in six chapters. In chapter 2 we detail the process of image formation, the different elements involved in it and how does colour is obtained. It is necessary to study the formation of a coloured image to be able to subsequently deal with colour. In chapter 3 we present colour invariance approaches to deal with colour in images, which do not require additional information apart from the image and deliver image descriptors which aim to be invariant to the illuminant of the scene. An application in colour modelling for segmentation is presented along with the performance of some of the invariant normalisations. In chapter 4 computational colour constancy methods are presented. These methods aim to recover the illuminant of an image and use it to colour correct it. We propose a new colour constancy method, combining ideas of the existing approaches, and some experiments are shown to prove its performance. Also, we study the space of feasible solutions of these methods and propose some hints to improve the selection step. Next, in chapter 5, we reformulate the problem for the image annotation frame under uncalibrated conditions and we propose a colour correction method that considers different white point interpretations according to the scene content. Finally, in chapter 6 we present some conclusions derived from the work, along with our contribution in the colour correction field and the future directions that we think that might be interesting to go further.

Chapter 2

Colour and imaging

To understand how colours we perceive from a scene are affected by the illuminant conditions, first we need to study the process of image formation. In this chapter, we present the process of how HVS perceives colour and also how colour can be represented considering this process. This will lead us to look at the image formation model, that is the physical process that allows to build digital images. Then, we present the colour correction process that is performed when acquiring digital images to obtain a white balanced image. Since in computer vision we do not normally know any sensor information, we propose a framework where both illuminant and sensor are unknowns, to perform some colour correction to obtain a white balanced image. Finally, we present different ways to proceed to process colour for image interpretation according to the known information of the acquisition system and the goal of our computer vision system.

2.1 Introduction

The process of producing a coloured digital image of a scene involves different steps. The light reaching a sensor device is a combination of the scene illumination and the reflectance properties of the objects in the scene. A colour digital camera is normally composed by a lens, three colour filters, an image sensor (CCD or CMOS) and a storage memory. The lens focuses light rays on the image sensor, through the colour filters. The colour filters deliver different responses for a given range of wavelengths. An image is formed in the image sensor, and then it is stored in a memory. This architecture is not arbitrary, but it has been inspired by the HVS. Therefore, to understand it, it is useful to previously study the HVS, how a colour image is formed, and also the different elements involved in it.

2.2 Human visual system

The structure of the human visual system (HVS) has been a source of inspiration to create cameras for imaging. HVS is the part of the nervous system which allows humans to see. It interprets the information from visible light reaching the eyes to build a representation of the world surrounding the body.

The eye (figure 2.1) is a biological device, whose operation has inspired digital cameras, taking visible light and converting it into a stream of information that can be transmitted via nerves. HVS is a very complex system that processes these signals and subsequently generate the image representation that we perceive from our environment.

Light entering the eye is refracted as it passes through the cornea. Then, it passes through the pupil, controlled by the iris by contracting or dilation in order to regulate the intensity of light entering the eye, and is further refracted by the lens. The lens inverts the light and projects an image onto the retina.

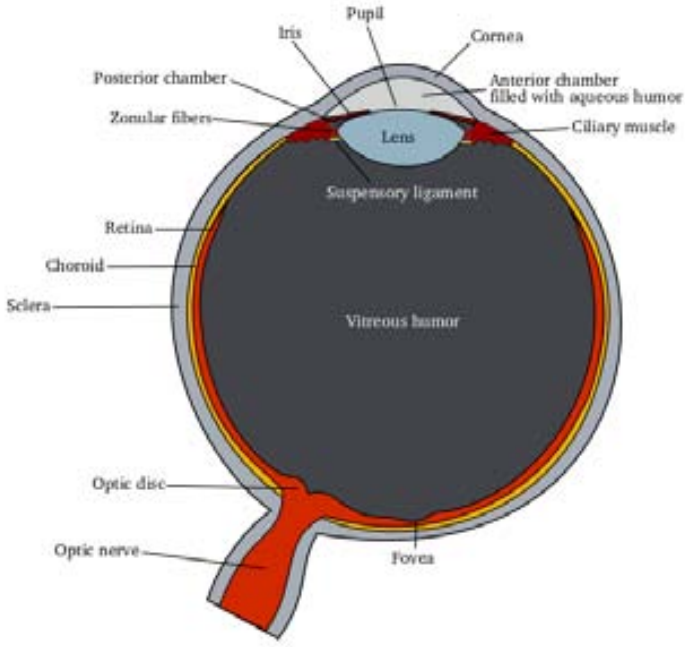


Figure 2.1: Structure of the eye: the retina is the sensor sensible to light.

The retina (figure 2.2) consists of a large number of photoreceptor cells which contain a particular protein molecule: the photopigment called rhodopsin [100, 73].

When rhodopsin is struck by a photon (a particle of light) it transmits a signal to the cell: the more photons strike the cell, the stronger the signal will be. There are two types of photoreceptor cells: rods and cones. In some animals, like humans, cone cells contain cone opsin molecules attuned to specific wavelengths of light; i.e., a blue cone cell contains opsin most attuned to blue-wavelength light and will most strongly be stimulated by blue-wavelength light, while a yellow-red cone cell will only be weakly stimulated by blue-wavelength light. This gives the ability to distinguish colour.

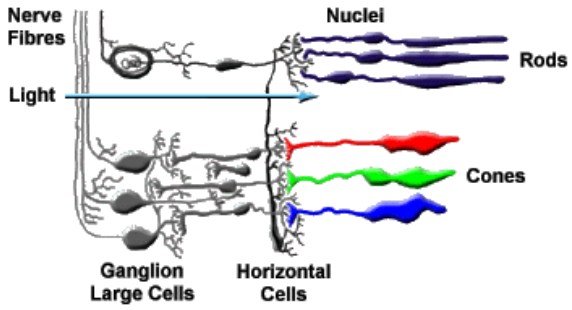


Figure 2.2: In the retina there are two types of photoreceptor cells: cones and rods. There are three types of cones, depending to the specific wavelengths of light to which they are attuned.

The HVS has three types of cones ($k = 3$) which are associated with red, green and blue: short (S), medium (M) and long (L) wavelengths respectively. For this reason HVS is known as a trichromatic system. Rods are stimulated at low intensity light and get saturated before the cones, since they are sensitive enough to respond to a single photon of light, whereas cones require tens to hundreds of photons to become activated. Thus, they are normally active under low intensity environments and give place to achromatic images. For this reason they are not normally considered in colour perception. The spectral absorption curves of the three types of cones and of the rods are shown in figure 2.3.

The study of the perception of colour of the HVS arrived at an important stage in 1931 when the International Comission on Illumination (CIE) proposed one of the first mathematically defined colour spaces: CIE XYZ colour space [100]. The fact that the human eye has three types of receptors means that there are needed, in principle, three parameters to describe a colour sensation. Therefore, they created a method for associating three numbers (or tristimulus values) with each colour, based on direct measurements of the human eye. Experiments were performed using a circular split screen 2 degrees in size. On one side of the field a test colour was projected and on the other side, an observer-adjustable colour was projected. The adjustable colour was a mixture of three primary colours, each with fixed chromaticity, but with adjustable

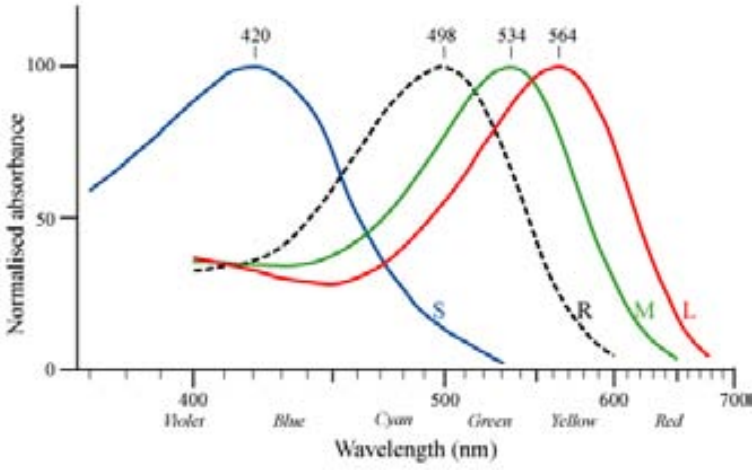


Figure 2.3: Spectral absorption curves of the short (S), medium (M) and long (L) wavelength pigments in human cone and rod (R) cells.

brightness.

The observer would alter the brightness of each of the three primary beams until a match to the test colour was observed. Not all test colours could be matched using this technique. When this was the case, some of the primaries could be added to the test colour, and a match with the remaining primaries was carried out with the variable colour spot. For these cases, the amount of the primary added to the test colour was considered to be a negative value. In this way, the entire range of human colour matching could be measured. When the test colours were monochromatic, a plot could be made of the amount of each primary used as a function of the wavelength of the test colour. These experiments gave place to three functions, $\overline{r}$, $\overline{g}$ and $\overline{b}$, that are known as CIE 1931 RGB colour matching functions (figure 2.4).

Having developed an RGB model of human vision using the CIE RGB matching functions, the members of the commission aimed to develop another colour space, CIE XYZ, which should be related to the CIE RGB colour space, but chosen to have the following desirable properties:

- The new colour matching functions, $\overline{x}$, $\overline{y}$ and $\overline{z}$, were to be positive everywhere.
- The $\overline{y}$ colour matching function would be exactly equal to a previously defined luminous efficiency function $V(\lambda)$ for the "CIE standard photopic observer" [22]. The luminance function described the variation of perceived brightness across the different wavelengths. The fact that the luminance function could be constructed by a linear combination of the RGB colour matching functions was not guaranteed by any means but might be expected to be nearly true due to the near-linear nature of human sight.

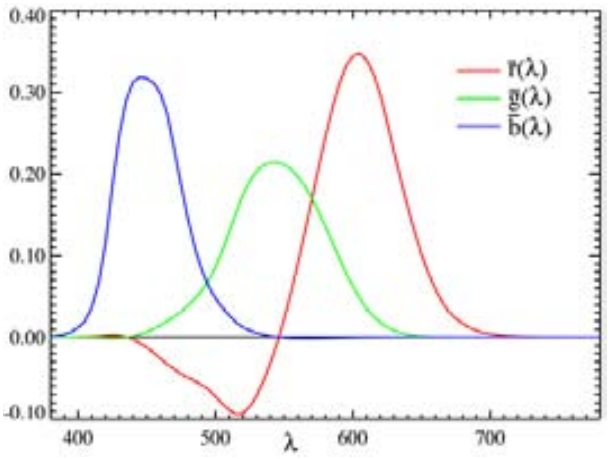


Figure 2.4: CIE 1931 RGB Colour matching functions.

- For the constant energy white point, it was required that X=Y=Z=1/3.
- Since the requirement of positive values of x and y (normalised values of X and Y), it can be seen that the xy gamut of all colours would lie inside the triangle defined by coordinates (0,0), (1,0) and (0,1). It was required that the gamut fill this space practically completely.
- It was found that the $\overline{z}$ colour matching function could be set to zero above 650 nm while remaining within the bounds of experimental error. Again, it was specified that this would be so, for computational simplicity.

Therefore, following these desirable properties, the CIE XYZ colour space was defined. In the CIE XYZ colour space, the tristimulus values are not the S, M, and L stimuli of the human eye, but rather a set of tristimulus values called X, Y, and Z, which are also roughly red, green and blue, respectively. Two light sources may be made up of different mixtures of various colours, and yet have the same colour representation (metamerism). If two light sources have the same apparent colour, then they will have the same tristimulus values, no matter what different mixtures of light was used to produce them. In figure 2.5 we show the $\overline{x}$, $\overline{y}$ and $\overline{z}$ functions proposed to be taken as a standard observer.

This colour space was designed so that the Y parameter was a measure of the brightness of a colour. Therefore, a chromaticity space was defined by two parameters which are functions of all three tristimulus values X, Y and Z,

$$\begin{aligned} x &= \frac{X}{X+Y+Z} \\ y &= \frac{Y}{X+Y+Z} \end{aligned} \tag{2.1}$$

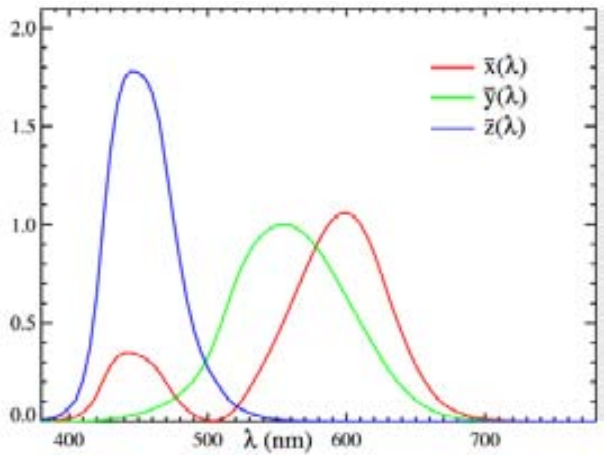


Figure 2.5: CIE 1931 XYZ colour matching functions.

Thus, a chromaticity diagram is derived (figure 2.6). The diagram represents all of the chromaticities visible to the average person. These are shown in colour and this region is called the gamut of human vision. The gamut of all visible chromaticities on the CIE plot is the tongue-shaped or horseshoe-shaped figure shown in colour. The curved edge of the gamut is called the spectral locus and corresponds to monochromatic light, with wavelengths listed in nanometers. The straight edge on the lower part of the gamut is called the purple line. These colours, although they are on the border of the gamut, have no counterpart in monochromatic light. Less saturated colours appear in the interior of the figure, with white at the centre. All visible chromaticities correspond to positive values of x and y. If one chooses any two points on the chromaticity diagram, then all colours that can be formed by mixing these two colours lie between those two points, on a straight line connecting them. It follows that the gamut of colours must be convex in shape. All colours that can be formed by mixing three sources are found inside the triangle formed by the source points on the chromaticity diagram, and so on for multiple sources. The mixture of two equally bright colours will not generally lie on the midpoint of that line segment. In more general terms, a distance on the xy chromaticity diagram does not correspond to the degree of perceived difference between two colours.

CIE colour matching functions have been very important in the study and development of computational colour, since colour is a subjective (referred to the HVS) perception of the physical properties of the surfaces.

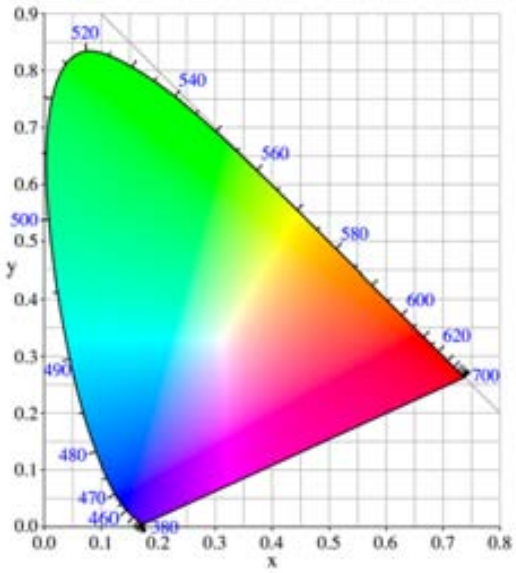


Figure 2.6: The CIE 1931 colour space chromaticity diagram. The outer curved portion is the spectral (or monochromatic) locus, with wavelengths shown in nanometers.

2.3 Colour image formation model

Once we have explained how the HVS works with colour, now we are ready to explain how colour digital images are built to simulate colour perception. To form a digital image of a scene, we need a receptor device that captures the light reflected by it. Geometrical and reflectance properties of surfaces are combined along with spectral information of the scene illuminant and captured by the receptor.

Shafer in [76] proposes the dichromatic reflection model to describe reflections in object surfaces. He describes surface reflectance as a linear combination of diffuse and specular components. Thus, a surface reflects the incident light in two different components: specular and diffuse reflections, also known as surface albedo and Fresnel reflection. The formulation of the model is,

$$\rho_k = m_b(\vec{n}, \vec{s}) \int E(\lambda) C_b(\lambda) R_k(\lambda) d\lambda + m_s(\vec{n}, \vec{s}, \vec{v}) \int E(\lambda) C_s(\lambda) R_k(\lambda) d\lambda \qquad (2.2)$$

where $E(\lambda)$ is the spectrum of the incident light, $R_k(\lambda)$ are the sensitivity curves of the acquisition device (for the k-th sensor) and C_b and C_s are the surface albedo and Fresnel reflectance respectively. $\vec{n}$ is the surface patch normal, $\vec{s}$ is the direction

of the illumination source and $\vec{v}$ is the direction of the viewer. Geometric terms m_b and m_s denote the geometric dependencies on the body and surface reflection component respectively. Specular reflection in a surface point is given at a concrete viewing angle, whereas diffuse reflection is normally constant at all angles of view (figure 2.7).

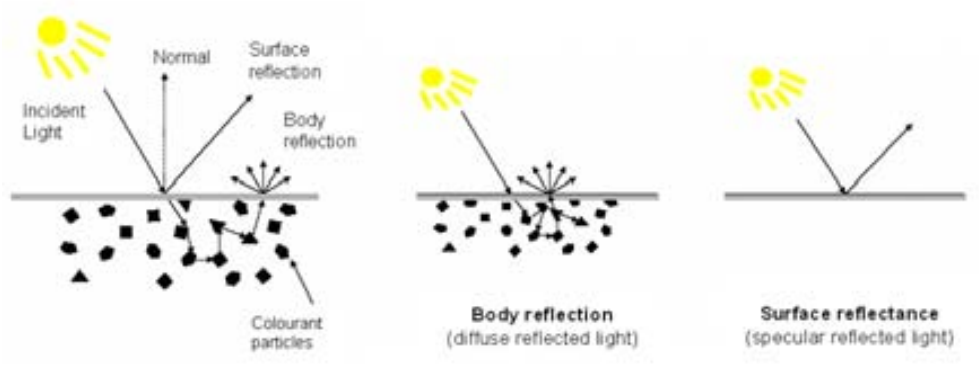


Figure 2.7: The dichromatic reflection model characterises reflectance as two components: diffuse and specular.

Frequently, in computer vision the dichromatic model is simplified and the specular component is taken out, since most of the colour information in images come from the diffuse reflection. This simplification is known as the Lambertian reflectance model, which just considers diffuse reflection. When dealing with colour, highlights might seem interesting to be considered, because they are a direct reflection of the incident light, and they could be used as illuminant descriptors of the scene [60, 81]. Nevertheless, highlights usually produce saturated regions in images, due to an excess of exposure to the light. Also, to study colour, we will not consider the geometry terms of the surfaces, since they do not normally affect to chromaticity perception.

Hence, in the simplified colour image formation model, an image is the result of the integration of the spectrum of the illuminant, the reflectance functions of the surfaces in the scene and the sensitivity curves of the sensor through wavelenghts corresponding to visible light, usually from 400 to 700nm. The reflectance function is the ratio of reflected power by a surface. The colour response of an imaging device is given by eq. 2.3,

$$\rho_k = \int E(\lambda)S(\lambda)R_k(\lambda)d\lambda \tag{2.3}$$

where $E(\lambda)$ is the spectra of the illuminant, $S(\lambda)$ is the reflectance function of the surface (surface albedo) and $R_k(\lambda)$ are the sensitivity response curves of the ac-

quisition device (with k sensors). When $k = 1$ we have a single type of sensor and a grey-level intensity image if formed, and when $k = 3$ we refer of coloured images, since it is the number of photoreceptors that the HVS uses to perceive colour. In this way, a colour RGB image is formed,

$$\begin{aligned} \rho_1 = R &= \int E(\lambda)S(\lambda)R(\lambda)d\lambda \\ \rho_2 = G &= \int E(\lambda)S(\lambda)G(\lambda)d\lambda \\ \rho_3 = B &= \int E(\lambda)S(\lambda)B(\lambda)d\lambda \end{aligned} \tag{2.4}$$

where $R(\lambda)$, $G(\lambda)$ and $B(\lambda)$ are the sensitivity curves of the RGB camera. In figure 2.8 we show the image formation for RGB images. Finally, when $k > 3$ we have multispectral images, which contain more information of the physical properties of the surfaces in the scene than the previous images. There are acquisition devices to acquire these different types of images.

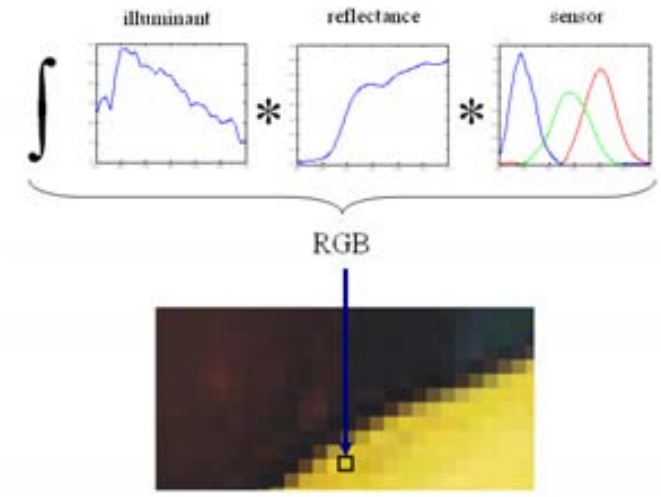


Figure 2.8: Colour image formation for RGB images: each rgb pixel in an image is the resut of the integration of an illuminant spectra a surface reflectance and a sensor device. In this case, a yellow trichromatic representation is obtained from a yellowish reflectance function, a blue-whitish illuminant and a white-balanced sensor.

2.4 White balancing in acquisition devices

The image formation model presented in previous section might not give the expected response when acquiring a white surface, where we normally expect the same maximum response in all the channels. Depending on the illumination conditions in which the white surface is viewed the model will deliver different coloured responses. Nevertheless, HVS perceives a white surface as white, whether viewed under daylight, a fluorescent or a light bulb. This ability of the HVS is called chromatic adaptation or colour constancy. Colour constancy ability ensures that the perceived colour of objects remains relatively constant under varying illumination conditions. This means, basically, to remove the effect of the illuminant when viewing a scene. A simple theory about how the HVS performs it is represented by the von Kries coefficient rule algorithm [98], where the gain of each class of photoreceptor is adjusted independently to obtain surface colour descriptors which are invariant to changes of illuminant. The factors, by which the gains are adjusted, are called coefficients, and they can be represented by a 3x3 diagonal matrix. Land in [57] proposed a theory of how colour constancy is achieved by the HVS. We will review it in chapter 4.

Therefore, for digital imaging it is interesting to perform a similar adaptation process to correct the colours of an acquired image, so that white is a balanced maximal response. In controlled systems, to obtain reliable colour images, sensor devices are normally calibrated for the given illuminant, in order to make that a white surface delivers the maximum response for all the channels. This process is known as calibration or white balancing, and normally consists of a scaling the response of each channel.

In figure 2.9 we show an example of three images of the same scene [8] under three different illuminants, where the calibration or white balancing has been performed for the illuminant of the first image. In the other two images, colours appear biased and should be subsequently corrected to obtain a perceptually good coloured image of the scene.

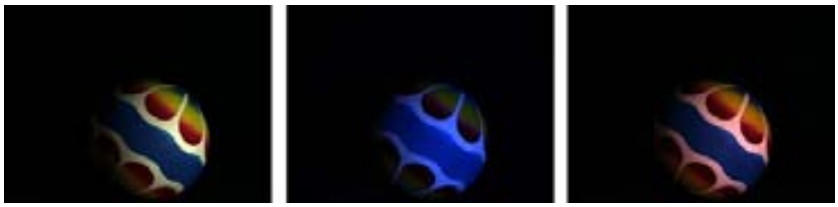


Figure 2.9: Three images of the same scene under three different illuminants, where the camera is calibrated for the illuminant of the first image.

The problem is that we do not always deal with a controlled system, and devices cannot be calibrated. In these cases, there exist a wide literature of methods to deal with this problem and therefore perform some illuminant estimation. These methods

will be reviewed in next chapters. When the white balancing process is not properly done, or there is no process of device calibration, the colours in an image do not represent physical properties of the surfaces in the scene. It is important to know, when possible, if the images we have to work with have been acquired with a calibrated device or not, or if we know the response of a camera for a white surface (in which case we can perform the white balancing process after the image has been acquired). Otherwise, we will have to consider just the different existing approaches to work with uncalibrated systems.

We will fix a framework to deal with images from unknown origin. In the colour image formation model proposed in section 2.3, there are three different elements to consider: the scene illuminant, the reflectance of the surfaces in the scene and the sensor curves of the acquisition device. In computer vision, we usually do not know either the illuminant or the spectral sensitivity of the sensor used in the acquisition. In this conditions we deal with a frame of image formation model like shown in figure 2.10, where the only information we know is the acquired image.

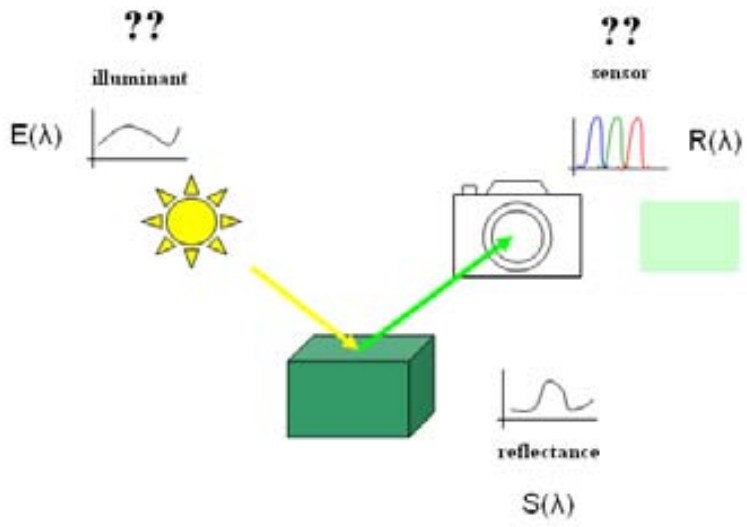


Figure 2.10: Frame of model of image formation for images of unknown origin where both illuminant and sensor are unknown. We propose to unify $E(\lambda)$ and $R(\lambda)$ in a single unknown to recover, which we name acquisition conditions, $A_k(\lambda)$.

The pixels in the image could be surface descriptors of the scene if the acquisition device was calibrated, but both unknown illuminant and sensor bias their values. To properly white balance the image we should know the two of them, or at least the sensor responses of a white surface under the illuminant. Since in this frame we know none of them, we reformulate the colour image formation model to unify these two

unknown parameters (illuminant and sensor) into a single unknown parameter, which we name acquisition conditions. Hence, for this frame we propose to *guess* this single parameter to subsequently correct the images. We define the acquisition conditions of an image, $A_k(\lambda)$, which integrate the illuminant and the sensor information, in eq. 2.5,

$$A_k(\lambda) = E(\lambda)R_k(\lambda) \tag{2.5}$$

Therefore, the colour image formation model becomes,

$$\rho_k = \int S(\lambda)A_k(\lambda)d\lambda \tag{2.6}$$

This reformulation aims to reduce the number of unknowns in the problem, since in this frame there are no ways to recover both unknown parameters. Therefore we will assume that the bias of the colours in an image might be due to a biased illuminant or to a wrong balanced sensor. In this way we perform a white point estimation process, where we aim to obtain the response of a white patch for the unknown acquisition conditions. This reformulation will be referred in chapter 5. In figure 2.11 we show the unification of these two parameters in the colour image formation model. Figure 2.11(a) represents the image formation model for calibrated conditions, where we aim to estimate the illuminant of the scene, and figure 2.11(b) the model for uncalibrated conditions, where we aim to estimate a white point, i.e. the response of a white patch under these acquisition conditions.

2.5 Colour in computer vision

In computer vision we can work with images of different origins. When we deal with calibrated devices in a controlled system, we can rely on the colours given by acquired images, and they can be used in further analysis where colour information is needed.

When we deal with uncalibrated devices, i.e. the sensor curves are not known to be white balanced or in an uncontrolled environment where we know them but we do not know the spectra of the illuminant, the colours might be biased. This latter case of uncalibrated conditions is quite common, since we are not always able to procure a controlled context with constant illumination or, even worse, we might be dealing with images from unknown origin, such as images from the internet. Normally, with images obtained from the internet colours appear to be white balanced and the HVS can interpret them, even though if we compare the values of similar perceived coloured regions in them, or at least with the same meaning for us, they result to be quite different. This might be either because different sensors have been used for

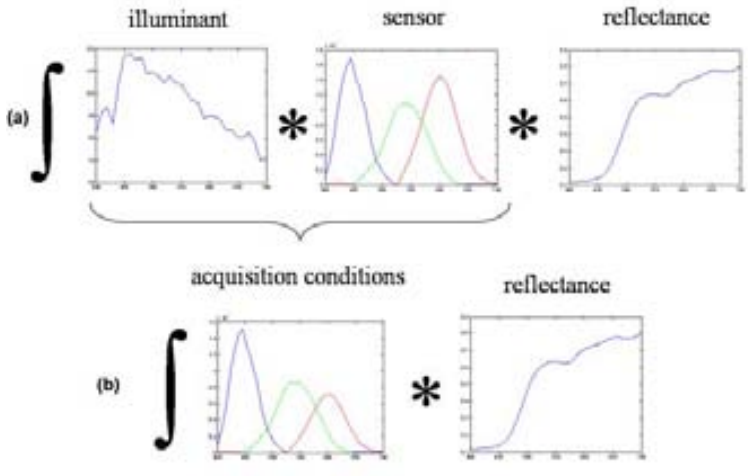


Figure 2.11: Illuminant and sensor information comprise adquisition conditions, therefore the colour image formation model results in the integration of these adquisition conditions and the given reflectance.

their acquisition or because different white balancing processes have been performed. In figure 2.12 we show different images obtained from internet, where there are skin and sky colour regions. If we take the skin regions out of the image, we can observe that the colours are very different, but for us they have the same meaning. The same happens with the sky regions. The images appear white balanced, but colours are very different, which might be due to different acquisition conditions.

There is a wide literature of colour processing methods which aim to use colour with reliability in computer vision methods, since it is an important and useful source of information. There are two main approaches to deal with colour: colour invariant normalisations and colour constancy methods.

Colour invariant normalisations [33, 48, 34, 97, 35, 36] look for image descriptors invariant to some image features such as illuminant intensity, illuminant colour, surface geometry or highlights. Hence, a normalisation of the pixel values in the image is performed to achieve this invariance (figure 2.13). Images are transformed to such invariant colour spaces, and sometimes the structure of the colours in the image is lost. Surface descriptors obtained with these methods are related to the rest of surfaces in the image and cannot be considered as reliable descriptors of the reflectance properties of the surfaces.

Colour constancy methods try to recover the scene illuminant, in order to be able to remove it from the image, that is, to put the image under a white balanced illuminant 2.14, known as canonical illuminant. Land in [55] proposed a theory of how the HVS performs colour constancy, from which a computational colour constancy method

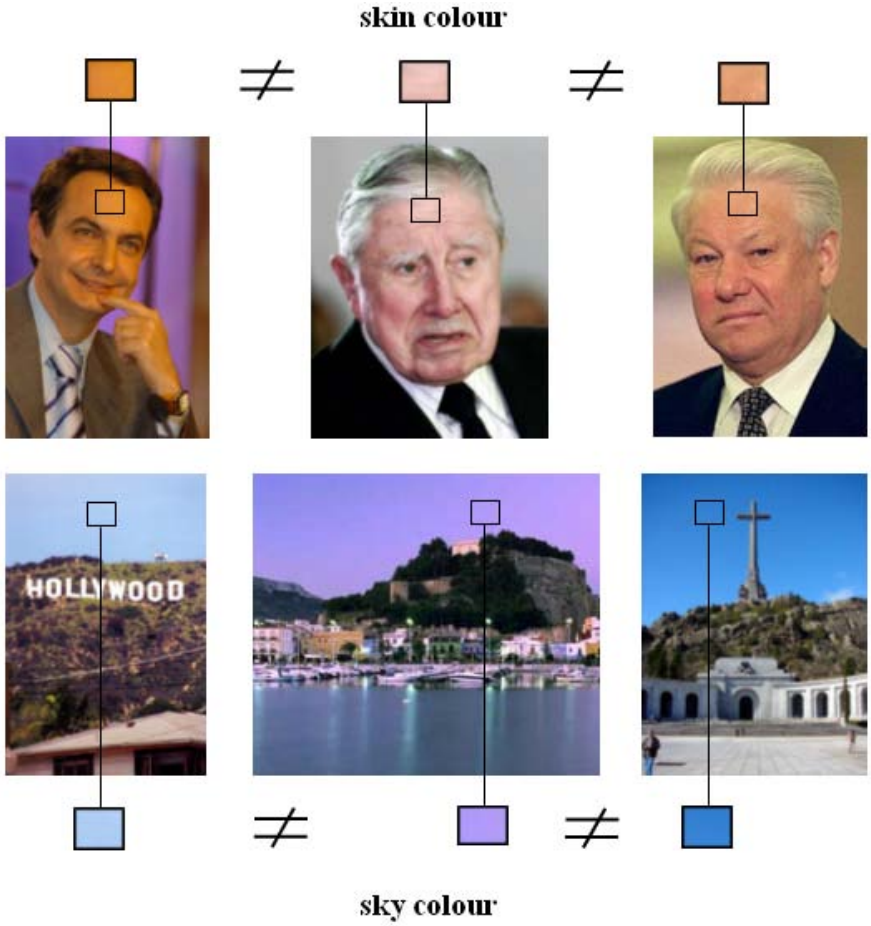


Figure 2.12: Images from the internet with very different colour representations.

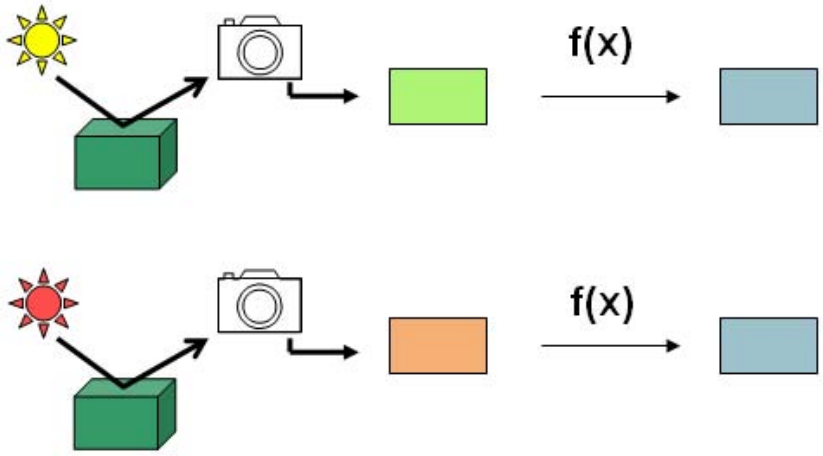


Figure 2.13: Colour invariant normalisations pursue to obtain colour image descriptors which are invariant to the illumination conditions of a scene through the use of normalisations (f(x)).

can be derived [56]. Then, there are methods that try to recover descriptors of the spectral power distribution of the illuminant of a scene [64, 27, 28, 29, 18, 74], which afterwards can be used to colour correct the illumination of the image. These methods do not normally perform well for rgb images. And there are methods which aim to estimate illuminant transformations to the canonical illuminant [19, 40, 31, 45, 38]. These methods can be classified following different criteria, in chapter 4 we propose a classification according to the information we know from the system. Surface descriptors obtained with all these methods should be considered as reliable descriptors of the reflectance properties of the surfaces, when colour constancy is perfectly performed.

In the following chapters we will show these approaches in more detail.

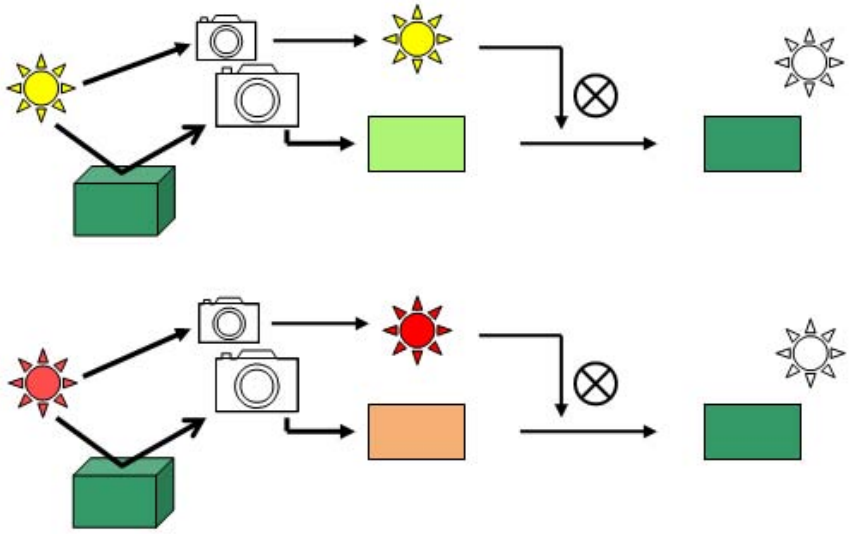


Figure 2.14: To perform perfect colour constancy the illuminant in a scene must be measured to properly correct the colours in the image.

Chapter 3

Colour invariance in practice

In this chapter, we present some invariant normalisations, which aim to obtain image descriptors invariant to the illuminant conditions of the scene. First, we propose a hypothesis to recommend when to use them. After presenting the normalisations, we use them for a problem of skin colour segmentation, which is useful for detecting faces in images, a common problem in computer vision. We present two different experiments to test the normalisations: one modelling skin colour under changes of illuminant intensity and another modelling it under changes of illuminant colour. Finally, we present a discussion and some conclusions of using invariant normalisations for dealing with colour images under unknown acquisition conditions.

3.1 Introduction

In colour invariance approaches we do not pursue to obtain a realistic image representation of the scene by estimating the illuminant of the image, but to obtain an illuminant invariant representation of it. Therefore surface descriptors obtained with these methods aim to be invariant to some features of the acquisition conditions. This is achieved by the use of normalisation functions, which transform images to invariant colour spaces. In this invariant spaces, general colour modelisations which operate under different imaging conditions can be built, for instance, for colour segmentation.

We have been working in a surveillance system where we need to detect people for extracting its appearance features. This project is explained in detail in apendix A. In this system we deal with a context of acquiring images where the sensor is known and with a more or less constant backgroud. The main problem in this system are the illumination conditions, which are not controlled.

This context, which is quite common in surveillance systems, can be considered fairly controlled. The main drawback is the uncontrolled lighting conditions. We propose a hypothesis to solve this problem for this conditions:

Hypothesis 1 *For acquisition conditions with a constant background and where the sensor is known, we will refer to them as semi-controlled conditions, the use of invariant normalisations can be enough for a colour segmentation task, such as skin detection and possibly for a general colour naming task.*

Next, we present the main invariant normalisations and also a skin colour detection experiment to prove the proposed hypothesis.

3.2 Colour invariant normalisations

There is a wide literature in invariant normalisations that cancel some features of the acquisition conditions of an image. Next, we present the most common invariance methods, along with the type of invariance do they aim. The model of illuminant change assumed by most of the invariant normalisations is given by:

$$(R \quad G \quad B)_{il_1} = s \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}_{il_2} \tag{3.1}$$

where s is the intensity component and $\alpha\beta\gamma$ the colour component of the illuminant change. This model will be used to explain the invariance proposed.

3.2.1 Chromaticity coordinates

The first normalisation that has long been used is chromaticity coordinates. This normalisation removes the dependency from the intensity of the illuminant and geometry of objects in colour images. This normalisation is quite often considered, and is used in many different problems as a space to build colour modelisations, e.g. skin colour [82].

Given that a change in the intensity of the light reflected by objects (either by changing of the illuminant or by geometry of the objects) increase or decrease equally the values for the different channels of an image, we can remove the intensity illuminant dependency by dividing pixel values at each channel by the sum of the values of all the channels for the given pixel. This is how chromaticity coordinates are defined,

$$nr = \frac{R}{R+G+B} \qquad ng = \frac{G}{R+G+B} \qquad nb = \frac{B}{R+G+B} \tag{3.2}$$

If we consider the model of illuminant change in eq. 3.1, we see that for the red channel this normalisation cancels the intensity component:

$$nr = \frac{sR}{sR + sG + sB} = \frac{sR}{s(R + G + B)} = \frac{R}{R + G + B} \tag{3.3}$$

And so on for the green and blue channels.

As we can see, $nr + ng + nb = 1$, which means that the normalised values remain in a plane. In this way, we only need to work with the two first coordinates, since the third coordinate is redundant and can be obtained from the two given. Usually nr and ng are the coordinates selected, and nb is obviated. As said before, chromaticity coordinates values remain in a plane, which is obtained by the intersection of the rgb space with the $r + g + b = 1$ plane, and also it is possible to transform nr, ng and nb values to the 2D space defined by this plane. This space is named (in [11]): uv colour space. Equation 3.4 defines the transformation to this space,

$$u = \frac{1 + nr - nb}{\sqrt{2}} \qquad v = \frac{\sqrt{6}nb}{2} \tag{3.4}$$

This transform gives place to a normalised chromaticity diagram (figure 3.1) in a uv space that has been used in [11] for a colour naming task.

In figure 3.2 we show the result of this normalisation for five images of the same scene under five different illuminants. The mean of the distance between the normalised images is 0.6984 and the standard deviation 0.5174.

3.2.2 Comprehensive colour normalisation

Finlayson in [33] proposes a comprehensive colour normalisation to avoid the dependency due to changes of the geometry of the illumination or to changes in its colour. Since with chromaticity coordinates we only removed the illuminant intensity dependence, he proposes a grey-world normalisation step that can be performed together and iteratively with chromaticity coordinates normalisation to avoid dependency on the illumination geometry and colour.

Therefore, a function which normalises each pixel of the image invariant to intensity, chromaticity coordinates, is defined. Let $\mathcal{I}$ be the pixels of an image with N pixels and k channels. The normalisation function, $R()$, can be defined as,

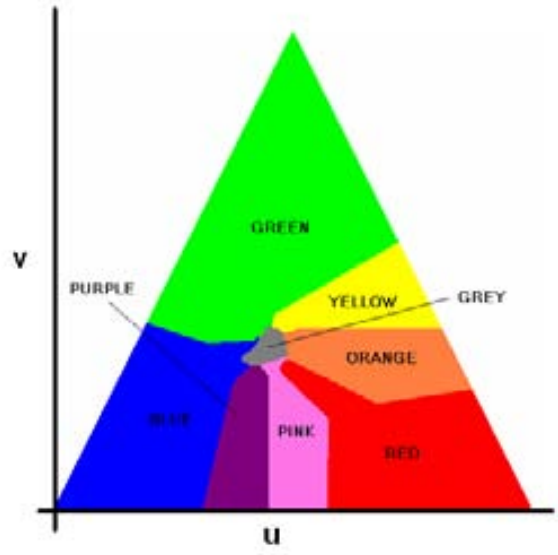


Figure 3.1: Chromaticity diagram obtained by the transformation of the chromaticity coordinates to a 2D space.

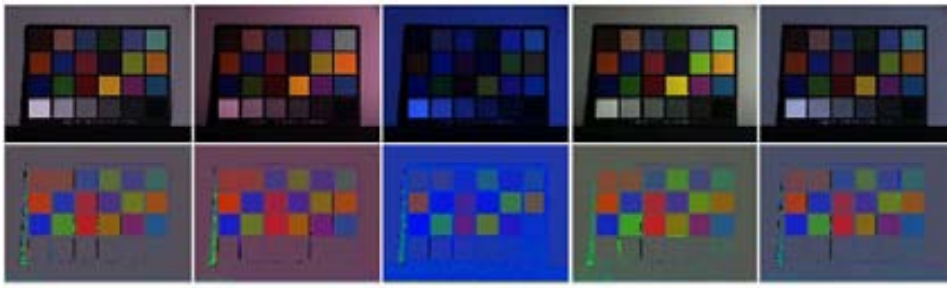


Figure 3.2: The intensity normalisation for five images of the same scene under five different illuminants, where the sensor is calibrated for the illuminant of the first image.

$$R(\mathcal{I}_{i,j}) = \frac{\mathcal{I}_{i,j}}{\sum_{c=1}^{k} \mathcal{I}_{i,c}} \tag{3.5}$$

where $i = 1, .., N$ and $j = 1, .., k$. This function is the general expression for chromaticity coordinates. To avoid the dependence on the illuminant colour, another normalisation can be applied, a grey-world normalisation that deals with channel information, which is where illuminant colour information lies. Another normalisation function, $C()$, can be defined as,

$$C(\mathcal{I}_{i,j}) = \frac{N\mathcal{I}_{i,j}}{\sum_{n=1}^{N} \mathcal{I}_{n,j}} \tag{3.6}$$

This $C()$ function normalises each channel of $\mathcal{I}$. If we consider the model of illuminant change in eq. 3.1, we see that for the red channel this normalisation cancels the colour component:

$$C(\mathcal{I}_{i,1}) = \frac{\alpha N\mathcal{I}_{i,1}}{\sum_{n=1}^{N} \alpha\mathcal{I}_{n,1}} = \frac{\alpha N\mathcal{I}_{i,1}}{\alpha \sum_{n=1}^{N} \mathcal{I}_{n,1}} = \frac{N\mathcal{I}_{i,1}}{\sum_{n=1}^{N} \mathcal{I}_{n,1}} \tag{3.7}$$

And so on for the green and blue channels. Thus, the dependence from the illuminant colour is avoided. The comprehensive colour image normalisation proposed by Finlayson makes the two normalisations through an iterative process that converges and has unique solution [33]. The procedure of the algorithm is the following:

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

This algorithm successively performs intensity and colour illuminant normalisations until the image converges. In this way, we obtain image descriptors invariant to intensity and colour of the illuminant. In figure 3.3 we show the result of this normalisation for five images of the same scene under five different illuminants. The mean of the distance between the normalised images is 0.3477 and the standard deviation 0.2984.

3.2.3 Comprehensive colour normalisation without foreground

Comprehensive colour normalisation considers all the pixels in the image to perform the channel normalisation. In specific cases, in situations where we have an object in front of a camera on a fixed background, we can think of a channel normalisation

Figure 3.3: The comprehensive normalisation for five images of the same scene under five different illuminants, where the sensor is calibrated for the illuminant of the first image.

which only takes into account background pixels of the image. Vanrell in [97] proposes an adaptation of Finlayson's comprehensive normalisation in order to use information of the background about the illuminant colour on the scene.

In many computer vision applications, there is an object in the foreground of a scene with uncontrolled lighting conditions and the background is constant, which means that we might be able to subtract the background from the image (segment the object in the foreground). We can consider that the light reflected by the object is affected by the colour of the background of the image, if we take into account that the light on the scene is reflected on the background and reach the object. The object that we have to analyse is affected by the light which directly reaches the object and by the light that comes from the background. In figure 3.4 we show an image of the surveillance system we have worked on, where we can perform a background subtraction process. We are interested in analysing the colours of the clothes of the person that enters a building. Since the background is always the same, we can segment the person from the background, and therefore use only background pixels for a channel normalisation.

Figure 3.4: Example of an application where we can segment an object (person) in the foreground of a scene, since the background is always the same. Therefore only the pixels in the background might be considered in the channel normalisation process.

Suposing that we can segment the object in the foreground of the image, this modification proposes to use only the background of the scene in the channel normalisation step, and hence remove the influence of the background colour from the colour perceived from the object. Therefore, the algorithm proposed by Finlayson is only modified in the channel normalisation step in the following way,

$$C(\mathcal{I}_{i,j}) = \frac{N_b \mathcal{I}_{i,j}}{\sum_{n=1}^{N_b} \mathcal{I}_{n,j}} \tag{3.8}$$

where N_b is the number of pixels in the background and n iterates only on the pixels in the background. In figure 3.5 we show the result of this normalisation for five images of the same scene under five different illuminants. The mean of the distance between the normalised images is 0.3611 and the standard deviation 0.2916.

Figure 3.5: The comprehensive normalisation without foreground for five images of the same scene under five different illuminants, where the sensor is calibrated for the illuminant of the first image.

3.2.4 Non-iterative comprehensive colour normalisation

Finlayson has proposed a variation of his comprehensive colour image normalisation [35], which performs the same invariant normalisation previously presented but in a single step. He avoids the iterations on the previous algorithm, which in some cases and for some images it was long to converge and could take quite computation time.

He proposes to work in the log-space and therefore turn geometry and light colour dependency, which is assumed as multiplicative, into additive processes. In log colour space two simple projection operators lead to invariance to geometry and light colour. And since these projection operators are idempotent, illuminant invariance is achieved in a single step.

In RGB space, when the lighting conditions change the effect on a pixel is multiplicative. Comprehensive normalisation removes multiplicative dependencies using

division. In this way, in log RGB space the multiplication is turned into addition. If the image is processed in its matrix representation, a projection matrix can be defined to perform a projection to a subspace which is orthogonal complement of the space spanned by lighting geometry change. Another projection matrix can be defined to project to a subspace which is orthogonal complement of the space spanned by light colour change. These two projection matrices represent a subtraction of the mean of the log rgb to each log r, g and b values, and a subtraction of the mean channel for each component in that channel. With this two subtractions the dependency on lighting geometry and colour is removed. Therefore, the normalisation is defined,

$$L(\mathcal{I}_{i,j}) = log(\mathcal{I}_{i,j}) - \frac{\sum_{n=1}^{c} log(\mathcal{I}_{i,n})}{c} - \frac{\sum_{m=1}^{N} log(\mathcal{I}_{m,j})}{N} \tag{3.9}$$

where N is the number of pixels of the image and c is the number of channels.

The projection matrices have been proven to be idempotent [35], and therefore the same result is obtained if we perform more than a single projection. Let I be the matrix representation for an image, and P_r and P_c the projection matrices corresponding to these normalisations. The iterative normalisation would remain as,

$$I_{norm} = P_c \dots P_c I P_r \dots P_r \tag{3.10}$$

Since the projection matrices are idempotent, $P_r P_r = P_r$ and $P_c P_c = P_c$. Therefore, just one multiplication with each of the matrices is enough to perform the normalisation. Thus, the normalisation process remains as,

$$I_{norm} = P_c I P_r \tag{3.11}$$

The matrix formulation of the normalisation is used to prove that no iteration is needed, due to the idempotency properties of these matrices. However, the normalisation is performed by just a subtraction of mean pixel and mean channel in the log space. In figure 3.6 we show the result of this normalisation for five images of the same scene under five different illuminants. The mean of the distance between the normalised images is 4.7270 and the standard deviation 5.1475.

In table 3.1, we show a summary of the distance mean and distance standard deviation for the normalisations studied until now.

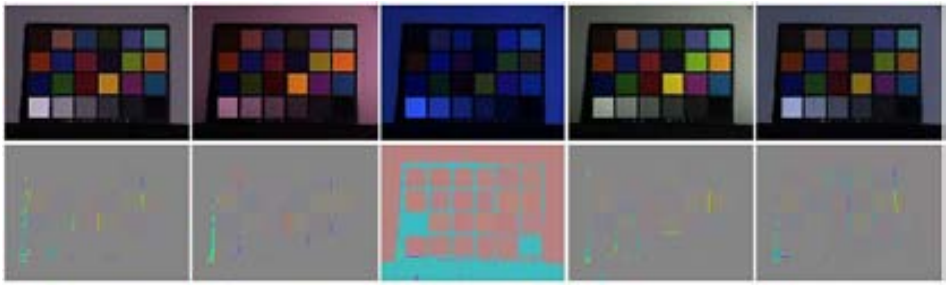


Figure 3.6: The non-iterative comprehensive normalisation for five images of the same scene under five different illuminants, where the sensor is calibrated for the illuminant of the first image.

Normalisation	Distance mean	Distance standard deviation
CC	0.6984	0.5174
CCN	0.3477	0.2984
CCNWF	0.3611	0.2916
NICCN	4.7270	5.1475

Table 3.1: Distance mean and standard deviation for the five macbeth images using CC, CCN,CCNWF and NICCN normalisations.

3.2.5 $l1l2l3$ normalisation

Gevers in [48] proposes two interesting colour normalisations invariant to different illuminant features. The first, $l1l2l3$, delivers image descriptors in a three dimensional circular space which is invariant to highlights,

$$
\begin{aligned}
l1 &= \frac{(R-G)^2}{(R-G)^2+(R-B)^2+(G-B)^2} \\
l2 &= \frac{(R-B)^2}{(R-G)^2+(R-B)^2+(G-B)^2} \\
l3 &= \frac{(G-B)^2}{(R-G)^2+(R-B)^2+(G-B)^2}
\end{aligned}
\tag{3.12}
$$

The subtraction of two components of the same pixel removes the specular factor in the dichromatic reflection model, since the specular factor is the same for the two of them, as it is shown in [48]. Thus, image descriptors obtained are dependent on sensors and surface albedo only.

In figure 3.7 we show the normalisation for an image with specular reflections. We

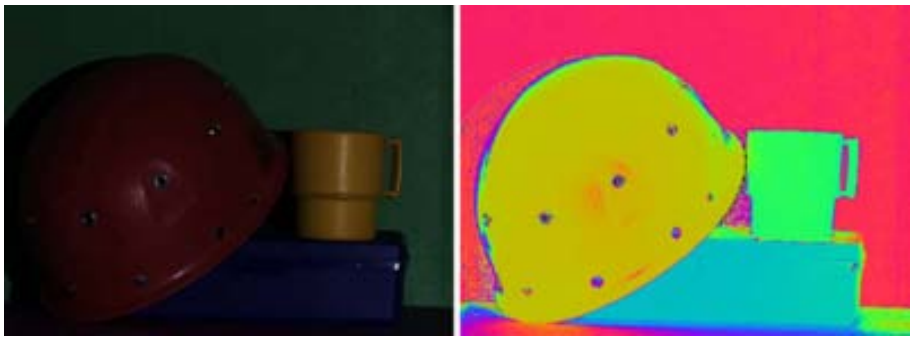


Figure 3.7: The $l1l2l3$ normalisation for an image with highlights.

can see that most of the specular reflections are avoided, such as the one in the yellow cup, and the others are minimised. We do not show the result of this normalisation for the five macbeth images previously presented, since this normalisation is mainly designed to remove highlights, which are not present in these images.

3.2.6 $m1m2m3$ normalisation

In [48] there is a second interesting normalisation proposed, $m1m2m3$. This normalisation involves two neighbouring pixels and assumes that the colour of the illuminant is locally constant. Equation 3.13 defines the normalisation,

$$\begin{aligned} m1 &= \frac{(R^{x_1}G^{x_2})}{(R^{x_2}G^{x_1})} \\ m2 &= \frac{(R^{x_1}B^{x_2})}{(R^{x_2}B^{x_1})} \\ m3 &= \frac{(G^{x_1}B^{x_2})}{(G^{x_2}B^{x_1})} \end{aligned} \tag{3.13}$$

where x_1 and x_2 refer to two adjacent pixels. If we consider the model of illuminant change in eq. 3.1, and assume a single illuminant over all the scene, we see that for the $m1$ component this normalisation cancels the colour component of the illuminant change:

$$m1 = \frac{(\alpha R^{x_1}\beta G^{x_2})}{(\alpha R^{x_2}\beta G^{x_1})} = \frac{(R^{x_1}G^{x_2})}{(R^{x_2}G^{x_1})} \tag{3.14}$$

And so on for the $m2$ and $m3$ components. Also, image descriptors obtained are invariant to interreflections, as it is shown in [48].

We do not show the result of this normalisation for the five macbeth images previously presented, since this normalisation is not designed to work with images of 'flat' scenes. It is designed to deliver invariance for varying geometry scenes and detect colour descriptors in edges, since for constant flat patches the normalisation cancels itself when processing neighbouring pixels.

3.3 Experiments for skin colour detection using invariant normalisations

To test the performance of some of these normalisations we have considered a typical problem in computer vision: skin segmentation in images with varying illumination. In the experiments that we have performed, different skin colour models have been computed in the different colour spaces given by the invariant normalisations. In this way, their invariance properties are evaluated for colour modelling. This is a different approach to test invariant normalisations, which normally have been tested for image indexing [48]. There are many previous works that have considered normalised colour spaces to model and detect skin colour [102, 101, 50, 21, 89, 91, 90, 78, 79, 77, 67, 68, 83, 84, 85, 82, 87].

In [102, 101], Yang and Waibel propose a real-time face tracker and they use rg-colour space, obtained using chromaticity coordinates normalisation, to characterise skin colour distribution using a multivariate normal distribution. In [50], Greenspan proposes face colour modelling and segmentation based on skin colour. In a similar way to Yang and Waibel, he builds a statistical model in rg-colour space, but in this case uses a mixture of gaussians to model skin colour. Cho in [21] proposes a plain colour model in HSV space with dynamic adaptiveness. The model uses just some thresholds in this 3D colour space to model skin colour and automatically adapts adjusting these thresholds values. In [89, 91, 90], test different colour spaces for skin colour modelling using single and mixture of gaussian models. He concludes that a single gaussian in rg-colour space is the most efficient option for skin colour segmentation. In [78, 79, 77, 67, 68], Soriano and Martinkauppi use histograms to model skin colour again in rg-colour space. Störring in [83, 84, 85, 82, 87] proposes a physics-based approach to model skin colour in rg-colour space, again, and first using histogram models and then a single gaussian model. We have presented some previous work of skin colour detection based on the preceding approaches in [95]. Recently, some works have considered infrared information to improve existing skin colour detection methods [26, 86].

Considering these previous works it seems interesting to test other invariant nor-

malisations to model skin colour, since chromaticity coordinates are widely used to get a normalised colour space, but other normalisations could deliver better results. Also, a single gaussian has been proved to be enough to model skin colour in normalised colour spaces. Therefore, we have performed different experiments of skin colour detection. These experiments have been performed in two steps: learning of the colour model and test it. In the learning step we have computed a skin colour gaussian model for the different normalisations, using a learning set of skin colour samples. This means that we have fitted a multivariate normal distribution for the skin colour samples in the different colour spaces defined by the invariant normalisations. We have not considered $l1l2l3$ and $m1m2m3$ normalisations, since their colour spaces do not maintain the normal distribution property for skin colour and more complex mathematical models should be used. Also, the comprehensive colour normalisation without foreground has not been tested because we do not dispose of a background subtraction process for the considered images.

For these experiments we have used the OULU face database [66]. This database is composed of images of faces of 125 different people of different races. The images were taken in a dark room with a gray background. Four different illuminants were used along with the four corresponding white balancing processes. The combination of the illuminants and white balancing processes delivered 16 different images per person, which could be interpreted as different illuminant conditions (since images are colour biased through these configurations). In figure 3.8 we show a sample of the image database, where we can see the different images obtained for a single person.

3.3.1 Experiment 1

For the first experiment, we have selected a large amount of skin and non-skin coloured regions within these images, considering the 16 different configurations, which will be used in the learning and test steps. A set of 4000 skin and non-skin regions (2000 of each) have been used in the learning step to fit the gausian model. On the other hand, 8000 skin and non-skin regions (4000 of each) have been used to test the fitted model.

In the learning step, the sample set has been used to fit a multivariate normal distribution [95], eq. 3.15, that models skin colour in the different colour spaces given by the normalisations,

$$P(\mathbf{x}|\mu,\mathbf{\Sigma}) = \mathbf{e}^{-\frac{\mathbf{1}}{\mathbf{2}}(\mathbf{x}-\mu)^{\mathbf{T}}\mathbf{\Sigma}^{-\mathbf{1}}(\mathbf{x}-\mu)} \tag{3.15}$$

The mean, μ, and covariance matrix, Σ, values have been fitted to enclose 80% of the skin samples. In the test step, we have used the test set of skin and non-skin coloured regions for a classification process using the gaussian models learned previ-

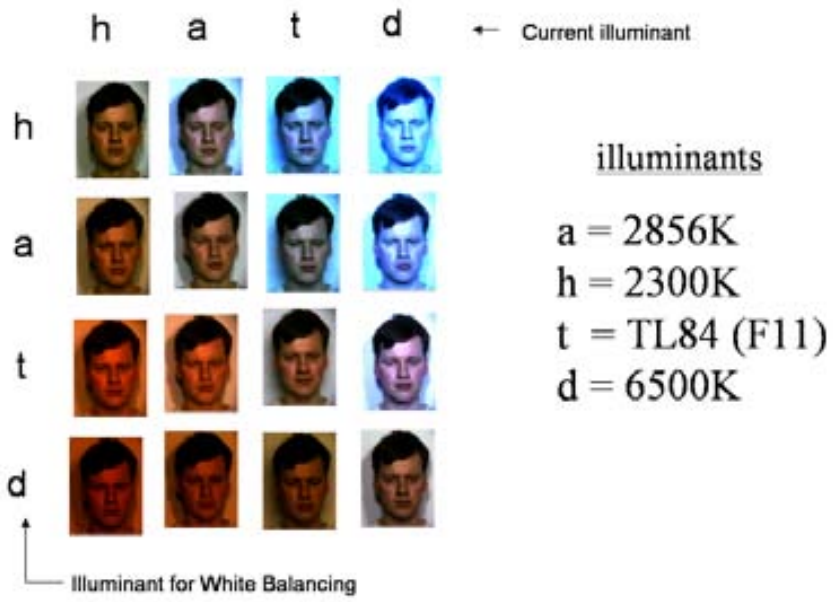


Figure 3.8: Different images obtained for a single person in the OULU face database, along with the different illumination and white balancing process.

ously. The classification results for the test set can be seen in table 3.2.

Normalisation	Success rate (%)
CC	62.93
CCN	85.32
NICCN	86.02

Table 3.2: Results obtained with the invariant normalisations: chromaticity coordinates (CC), comprehensive colour normalisation (CCN) and non-iterative comprehensive colour normalisation (NICCN).

If we look at these results, we observe that comprehensive normalisations (both iterative and non-iterative) perform much better than chromaticity coordinates. The models have been learned and tested using images from the 16 different configurations, which correspond to 16 illuminants with different intensity and spectra. Since comprehensive normalisations have been proposed to this end, they cancel the dependency of intensity and colour of the illuminant much better than chromaticity coordinates (which are designed to remove just intensity of the illuminant dependency) and, thus, the gaussian model fits better in their normalised colour spaces and deliver better performance. In this experiment we have tested the performance of invariant normalisations for skin colour modelling in images with strong changes

of illumination and the results have been that comprehensive normalisations deliver better colour invariance. However, we do not know if when dealing with light changes in illumination (just intensity changes) comprehensive normalisations performs better as well, and are the best option to choose when modelling skin colour in any illumination conditions. To this end we have performed a second experiment.

3.3.2 Experiment 2

To extend the conclusions obtained with the previous experiment, and check what happens if we only deal with images with changes in the intensity illuminant, we have performed a second experiment in which we have previously grouped the 16 different configuration type of images into five groups of illuminant conditions, if they have a similar illuminant colour. In figure 3.9 we show the different groups created according to the simulated illuminant colour.

Figure 3.9: Five different groups for the different configurations of camera/white balancing, according to a similarity in the simulated illuminant colours.

Therefore, we have processed each of these groups independently. A skin gausian model has been computed for each groups using skin and non-skin samples from the corresponding images. Thus, we obtain five different skin colour models. This situation would be possible if we had a classifier of the colour of the illuminant of the scenes or if we dealt with a system where there are only intensity changes in the illuminant (and we made five independent experiments, one for each group). It is interesting to perform this experiment to see how well do the gaussians adapt in the different colour spaces given by the normalisations, and therefore compare it to the unified processing of the samples, seen in previous experiment. In figure 3.10 we show the skin samples

used in the learning step and the subsequently fitted gaussians for the three normalisations considered, along with the plot of the density skin samples for each colour space.

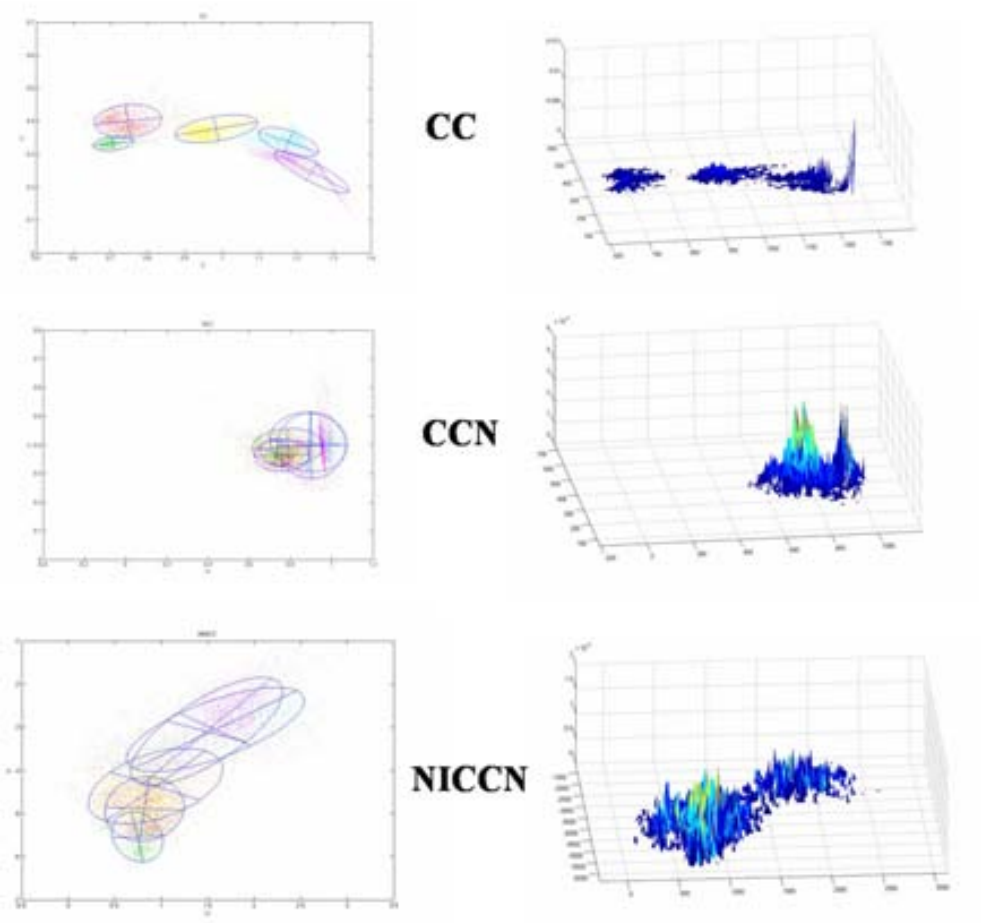


Figure 3.10: Fitting of the gausian models for the five different groups and using the three normalisations.

If we look at the chromaticity coordinates modelling, we can observe that the sample sets arrange in five distinguishable groups, according to the previously given groups. Also, these groups are fitted properly by the gaussian models. On the other side, the comprehensive normalisations deliver superposed gaussian models for the five different groups. This expected result is because comprehensive normalisations deal with the colour of the illuminant as well, and skin colour tend to be normalised to the same sub-space. In table 3.3 we show the results for the classification for skin and non-skin regions in the test set, combining the results from the five groups.

Normalisation	Success rate (%)
CC	95.82
CCN	88.42
NICCN	89.35

Table 3.3: Results obtained grouping by illuminant with the normalisations: CC, CCN and NICCN.

We can observe that the performance in this second experiment has improved for the three normalisations, which was an expected result since we have introduced some information about the illuminant. There is a slight improvement in the comprehensive normalisations. However, the best result are obtained with chromaticity coordinates. This shows that, when dealing with illuminant intensity changes, chromaticity coordinates remain as the best option. Comprehensive normalisations deals well with illuminant colour changes, but not as good as chromaticity coordinates for illuminant intensity changes.

3.3.3 Discussion

In these two experiments, we have tested the normalisations for changes in illuminant colour and illuminant intensity changes. Many previous work has considered just illuminant intensity changes and chromaticity coordinates. We have introduced comprehensive normalisations for skin colour segmentation and dealt with images with strong changes in illuminant conditions. In figure 3.11 we show results of skin colour segmentation for some images for the OULU face database using the models derived in the first experiment, where there is no previous illuminant colour grouping. Table 3.2 showed that comprehensive normalisations achieved best results for these conditions, and the segmentation results obtained reaffirm the numerical results. Better skin colour detection is achieved than the segmentation obtained with the chromaticity coordinates.

Finally, we have tested the normalisations for skin segmentation in images obtained from the internet, where we do not know the acquisition conditions. In the selected images, there are some skin coloured regions and we have tested how well do the different models obtained with the normalisations segment them. Some results are shown in figure 3.12.

In this case, chromaticity coordinates seem to achieve a better segmentation than comprehensive normalisations. This can be due to the fact that images we usually find in internet seem white balanced. Despite they have most probably been acquired with different sensors, the skin colour in the images is not strongly biased. Under these circumstances, chromaticity coordinates achieve a better skin colour segmentation than comprehensive normalisations, which deliver worse performance.

Figure 3.11: Results of skin segmentation for images from the OULU face database.

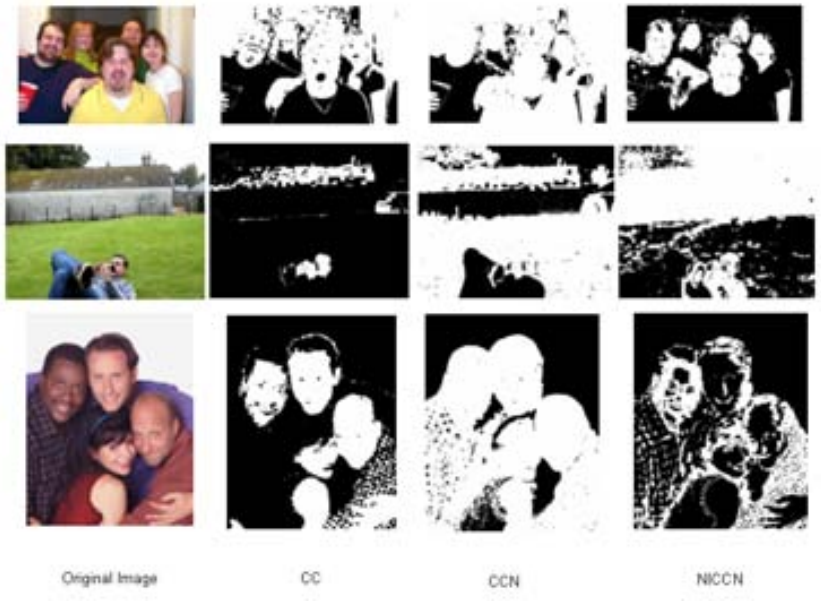


Figure 3.12: Results of skin segmentation for images from internet with unknown origin.

3.4 Conclusions

In this chapter we have presented invariant colour normalisation methods and tested the performance of some of them for the skin colour segmentation problem. We have proved that when dealing with images strongly colour biased the best option is to use comprehensive normalisations, since they deal quite well with changes in the colour and intensity of the illuminant. Otherwise, dealing with images with small changes of illumination (normally illuminant intensity) or images slightly colour biased the best option is to use chromaticity coordinates, since comprehensive normalisations take into account information of all the image to normalise its colours, and thus the entire image affects the colour of a concrete region. Sometimes, this produces indesirable normalisation results, which under strong changes of illumination are acceptable, but when working with colour unbiased images results are not as good as the ones achieved with chromaticity coordinates. This effect can be seen in figure 3.13 where we show the intensity normalisation and the comprehensive normalisations for two images with a different colour mean.

In both images we can observe how the achromatic background is maintained when using the chromaticity coordinates and is biased when using the comprehensive normalisation, where the colour mean of the whole image affects the normalisation for a single pixel. This effect is not a problem when dealing with images with strong colour illumination changes (see table 3.2), but when this is not the case, it is better

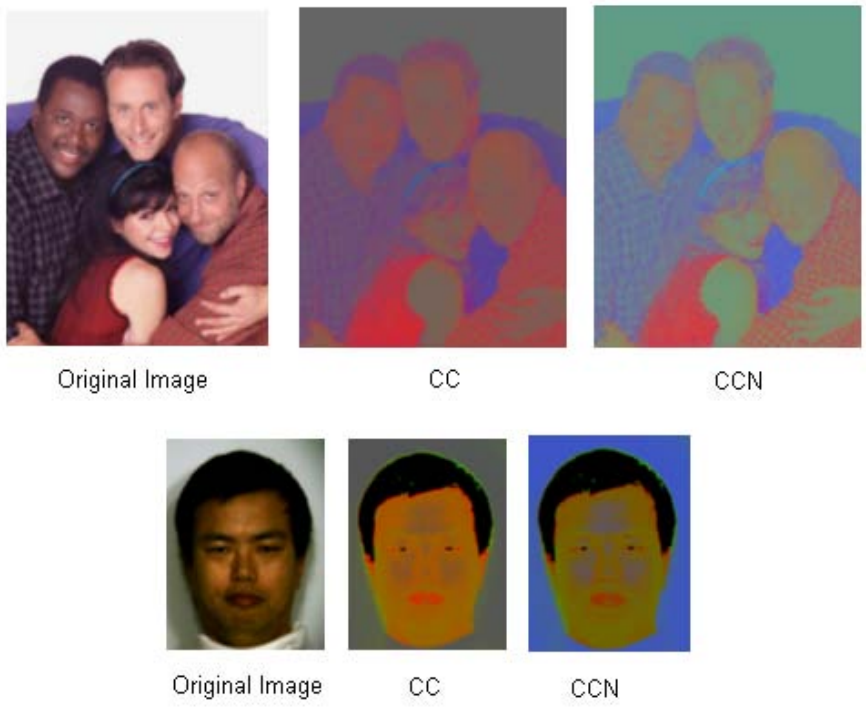


Figure 3.13: Intensity normalisation and comprehensive normalisation for two different images.

to use chromaticity coordinates. Images from unknown origin tend to be somewhat white balanced. In this case, we will choose chromaticity coordinates as well.

In the begining of this chapter, hypothesis 1 was proposed to use colour invariance normalisations for colour segmentation tasks for acquisition conditions with slight changes in illumination. We have proved that they deliver good performance under these, more or less, calibrated conditions for skin colour segmentation, and that they can also be considered for images with strong changes in illumination (chromaticity coordinates and comprehensive normalisations respectively). This has demonstrated that skin colour can be completely defined by its chromaticity component. However, when we move to a more general image interpretation context, some problems arise. Normally, all the invariant normalisations remove the intensity component of the images, and therefore it is difficult to perform a general colour naming task, which needs intensity information to work out colours. For this reason, and after considering invariant normalisations, we have derived the following conclusion:

Conclusion: under uncalibrated conditions, if we aim to perform image interpretation, and therefore a colour naming task, we will have to consider other approaches than colour invariance.

Colour constancy methods aim to recover the illuminant of a scene, and generally propose colour transformations without losing image information. In next chapter we present these approaches to deal with colour in the image interpretation problem along with some experiments to explore how might they be improved.

Chapter 4

Colour constancy in practice

To deal with general colour naming for image understanding we have discarded invariant normalisation methods and get into colour constancy methods. Colour constancy aims to recover the illuminant of a scene to properly white balance it without losing information. In this chapter we review computational colour constancy methods and propose a new method to deal with colour constancy for calibrated conditions, through a surface matching process. With this method we obtain similar performance to previous methods, and also we reduce the space of feasible solutions, which is good, since we reduce uncertainty in the problem. We use this method to present an exploration of colour constancy solutions in this feasible space and propose some hints to improve the selection step of some methods. Finally, we present a discussion about the evaluation of performance of colour constancy methods that proves that some performance results depend on the data set used. We conclude that for uncalibrated conditions we should avoid this evaluation and pursue some colour transformation that could be guided by the goal of our image understanding system.

4.1 Introduction

In the previous chapter, we have discarded the use of invariant normalisations for a general colour naming task, since they normally remove intensity information which is needed. The alternative is to use colour constancy methods. Colour constancy can be defined in two different contexts: perception and computer vision. In the frame of perception, colour constancy is the perceptual mechanism of the HVS which provides humans with colour vision which is relatively independent of the spectral content of the illumination of a scene. In this way, colour of surfaces are perceived invariant under changes in global illumination. A lemon for instance looks yellow to us at noon, when the main illumination is *bluish* sunlight, and also at sunset, when the main illumination is *reddish*. This ability helps us to identify objects and interact with them. On the other hand, in the frame of computer vision, colour constancy is a desirable ability of a machine vision system, and therefore a set of different techniques

to recover the illumination of images have been proposed. When processing images in a computer system, many of the same problems encountered by the HVS are present, also colour constancy. These methods aim to obtain illuminant independent descriptors of a scene, without losing information about its physical content.

There is a wide variety of computational methods, which try to achieve colour constancy from different points of view. Land in [57] and [55] proposed the retinex theory of colour vision, a theory of how HVS achieves colour constancy. This theory has been origin of methods that try to white balance images [40, 31, 45, 38]. Also, there are methods that are based on the recovery of spectral descriptors of the illuminant of a scene [64, 27, 28, 29, 18, 74]. The recovery of the spectral descriptors of illuminants and surfaces is out of the scope of this work, and thus they will not be considered. Also, we do not aim to reformulate the theory of how HVS performs colour constancy. In this chapter, we will center on the colour constancy problem under calibrated conditions.

Existing computational methods tend to solve the colour constancy problem by giving a single solution. This single solution proposed aims to estimate a white point in an image where we do not know if there is any, or to estimate the illumination through its effect on the scene surfaces. These methods deliver a solution in the form of a change of illumination, which pursues to obtain a white balanced image. Colour constancy is an ill-posed problem, since there exists a set of different possible solutions. Ill-posed problems must be regularised in order to reduce the set of solutions (and in the optimal case to obtain a unique solution), that is additional assumptions must be included to constrain the space of possible solutions [71]. To choose a single solution from this set of different possible solutions is a whole problem by itself. We can either introduce more information of the acquisition conditions, which is not always possible, introduce some assumptions of the scene, or both.

4.2 Colour constancy review

Next, we will see that even though some of the methods propose a feasible set of solutions, they normally tend to choose a single one. However, when it might interest us, they could be readapted to select a set of different solutions. To select a single solution within the feasible solution space, existing methods introduce assumptions to the problem in many different ways. Grey-world methods [19] assume a grey average of the image chromaticity, which can be considered realistic when images contain a wide enough range of colours. 'White patch' methods assume the brightest value for each channel as a white patch [6]. Obviously, in this case, saturated regions should not be considered, since they do not represent reflectance properties of the surfaces, and might bias the results. These two methods deal with uncalibrated conditions, as they do not need to know any information related to the acquisition process. Therefore, they can be used in any context, when other methods cannot, and they give a single solution by introducing these strong assumptions. Gamut mapping methods [40, 31]

introduce some assumptions to select a solution within a feasible set of solutions, such as take the mean of the feasible set or take the solution that maximises the image gamut, and bayesian methods [18, 38] introduce information about chromaticities under given illuminants also to select the most probable illuminant according to the colours which appear in the image. These two latter approaches need information from the acquisition device, which therefore must be known, and means that they cannot deal with images from uncalibrated conditions. Also, there are neural net methods [45, 20], which use a multilayer Perceptron to learn the estimation of the illuminant chromaticity considering the chromaticities in an image. Next, considering its relevance, we will present Land's retinex theory, and later we will get into detail of gamut mapping and bayesian approaches.

4.2.1 Retinex

Before we present some colour constancy methods in detail, we will get into a study of how colour constancy is achieved by the HVS, which gave place to many of the existing methods. Land in [57, 55] proposed a general theory for colour vision where he studied how the HVS achieves colour constancy. This theory is known as **retinex**, which is formed from the words 'retina' and 'cortex', suggesting that both the eye and the brain are involved in the process.

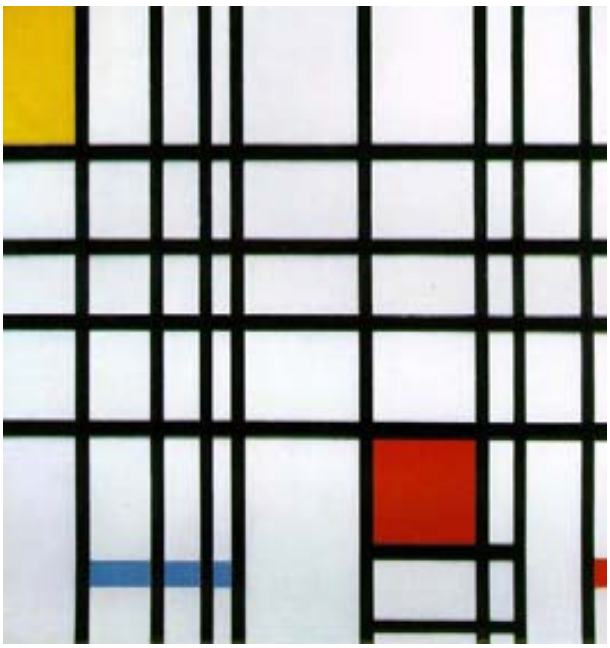


Figure 4.1: Composition with Red, Yellow and Blue 1921. Piet Mondrian. Oil on canvas. 72.5 x 69 cm.

He demonstrated this effect experimentally as follows. He created a 'mondrian' composition consisting of numerous coloured patches (after the dutch painter, Piet Mondrian, whose compositions are conceptually similar, figure 4.1) and showed it to a person. The 'mondrian' was illuminated by three white lights, one projected through a red filter, one projected through a green filter, and one projected through a blue filter. The person was asked to adjust the intensity of the three lights so that a particular patch in the 'mondrian' *appeared* white. This emulates chromatic adaptation proposed by Von Kries in [98], where a three coefficient rule models colour correction (this model has been used for most colour constancy algorithms). The intensities of red, green and blue light reflected from this *white-like* patch were measured. Then, the person was asked to identify the colour of a neighbouring patch, which appeared red. The lights that illuminated the 'mondrian' were adjusted so that the intensities of red, green and blue light reflected from the red patch were the same as were measured from the previous white patch. The person showed colour constancy since the red patch continued to appear red, the white patch continued to appear white, and all the remaining patches continued to have their original colours. In figure 4.2 we can see the scheme and the different elements used in these experiments.

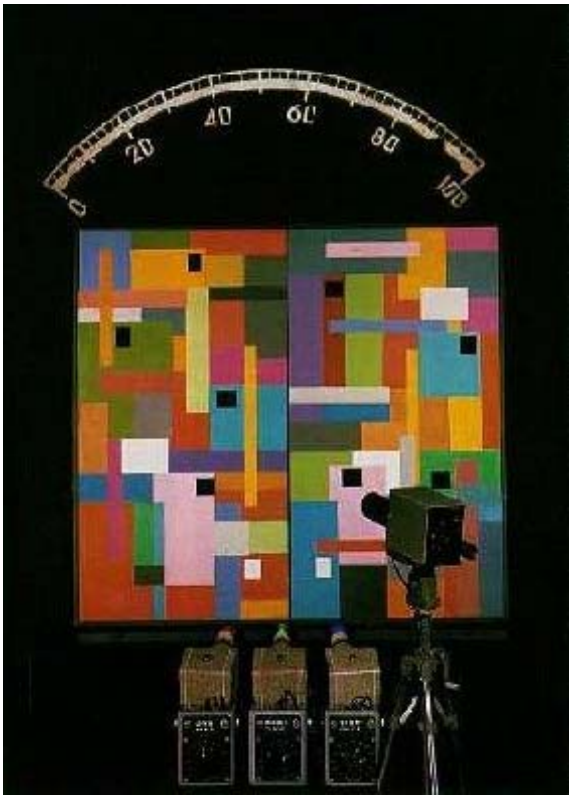


Figure 4.2: Scheme and different elements used in Land's colour constancy experiments.

Following these experiments, he developed a physics-based mathematical model for colour constancy based on the obtained results. This model was based on the use of random paths and the analysis of the luminance ratios measured through the three different filters of the coloured surfaces in the paths. The maximum luminance for each filter was taken as a white reference. Several algorithms have been developed based on this idea, even though most of them simplify the process described by Land. The max responses at each channel are joined to deliver an estimation of the illuminating light.

4.2.2 Gamut mapping approaches

Gamut mapping is a general assumption to deal with computational colour constancy and it was introduced by Forsyth in [40]. These methods introduce the concept of a canonical illuminant (and therefore a canonical color gamut) and build a feasible set of solutions. A canonical illuminant is a reference illuminant, for which a given sensor is normally calibrated. Assumptions are introduced to select a single solution within this feasible set, such as taking the average solution from the set or the solution which maximises the volume of the image gamut. Gamut mapping methods are depicted through three single steps:

1. Computing of the canonical color gamut, for a given sensor and a canonical illuminant, using an exhaustive set of reflectances.
2. Building the feasible set of solutions, that for a given image takes its color gamut into the canonical gamut.
3. Selecting a single solution within this feasible set of solutions.

Forsyth introduces an important assumption widely used, which we name the gamut mapping assumption:

Assumption 1 *The space of feasible solutions can be bounded if we require that any solution must guarantee that it takes the image gamut within the canonical gamut. In this way we obtain the feasible set of solutions.*

The canonical illuminant should be a standard illuminant which produces a balanced response for a white surface using a given sensor. Hence a canonical gamut, Γ_c, can be defined as the close convex set of the rgb sensor responses to all possible surfaces under this canonical illuminant. Image gamut, Γ_i, is defined as the convex set of the recorded camera responses for a given image. In figure 4.3 we show different transformations that take the image gamut into the canonical gamut following eq. 4.1, all these transformations form the feasible set,and in figure 4.4 we can see that the feasible set of solutions is a convex set in the space of illuminant changes and some heuristics need to be used to select a single solution within this set.

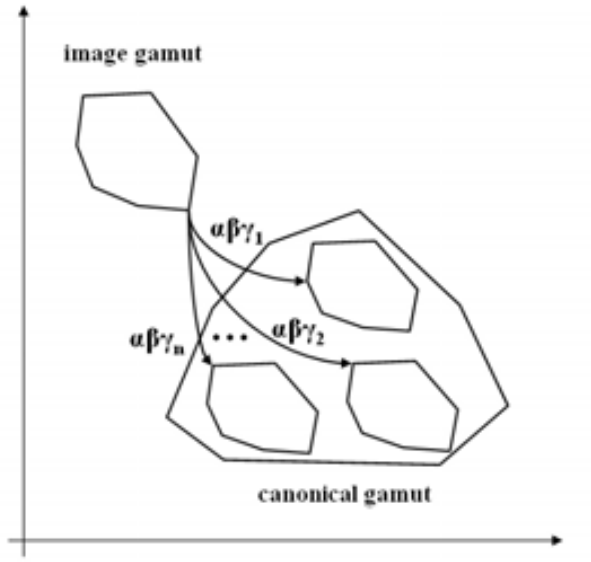


Figure 4.3: To build the feasible set of solutions, we consider the illuminant changes, $\alpha\beta\gamma_n$, that take the image gamut within the canonical gamut.

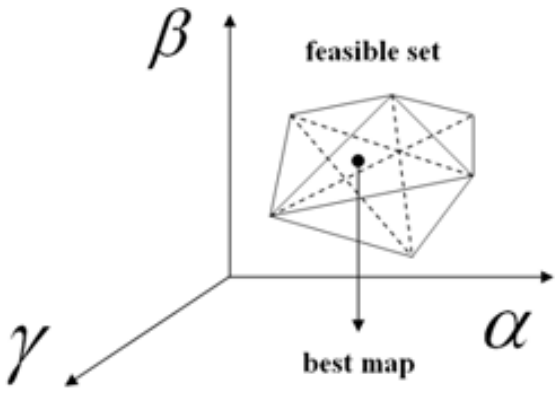


Figure 4.4: The feasible set is a convex set in the space of illuminant changes. Any solution within this set is, by definition, possible and some heuristics need to be used to select a single solution.

Forsyth in [40], following Von Kries three coefficient rule [98], proposes to use diagonal matrices, D, as the linear transforms which represent illuminant maps, since full 3×3 linear transforms introduce high computational complexity to the problem and it is unclear that, in a 3D colour space, the full model is necessary [30]. The feasible set of solutions is the convex set of all diagonal matrices, D which take the image gamut within the canonical gamut:

$$\forall \underline{\rho} \in \Gamma_i \qquad D\underline{\rho} \in \Gamma_c \qquad \text{where} \qquad D = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix} \tag{4.1}$$

where $\underline{\rho}$ denotes a colour of the image gamut. Once the feasible set is computed, some heuristic must be introduced to select a single map within it: average of the feasible set, max volume map, etc [37].

In [31], Finlayson presents a variation of Forsyth's method that performs a gamut mapping process in a 2D space, considering only chromaticities and removing intensity information. Hence, this 2D gamut mapping process estimates the chromaticity of unknown illuminants. Also, he introduces restrictions on the feasible illuminants, to subsequently further constrain the set of solutions.

All these methods use information of the sensor to operate, in the process of building the canonical gamut, and they cannot be used in problems where we know nothing of the acquisition conditions, such as images taken from the internet. It is important to notice that the canonical gamut in these methods is computed for a given sensor. Therefore, they are used to perform colour constancy when acquiring images with the corresponding sensor. For this reason, they are normally used for image correction, in problems where this information is known.

4.2.3 Bayesian approaches

There are probabilistic colour constancy methods, known as bayesian methods. Bayesian methods, such as Brainard's bayesian colour constancy [18] and Finlayson's colour by correlation [32, 38, 39], use prior knowledge of chromaticities under some illuminants as restrictions to constrain the problem.

In the case of the colour by correlation method [38], the probabilities of chromaticities under a set of given illuminants are used to compute a probability for each illuminant to be the scene illuminant. A correlation matrix is built to correlate possible image colours with each of the possible scene illuminants. This method deals with the recovery of the chromaticity of the illuminant, therefore intensity information is discarded and just reflectance chromaticity is considered. The chromaticity space is partitioned into uniform regions. For each illuminant, the range of posssible image chromaticities that can be observed under that light is characterised. This information

is used to build a probability distribution with the likelihood of observing an image colour under a given light: the probability distributions of each illuminant comprise the columns of the correlation matrix. To estimate the illuminant of an image, first, we determine which image chromaticities are present in the image by binarising a histogram of the chromaticities of the image. This gives us a vector of ones and zeros corresponding to whether or not a given chromaticity is present in the image. The product of the obtained vector with the correlation matrix produces a correlation of the vector of chromaticities in the image with the probability distribution of each illuminant in each column. Therefore, we obtain a vector with the correlation values for each illuminant. To find an estimate of the unknown illuminant, we can choose the illuminant which is most correlated with the image data, i.e. the maximum in the vector with the correlation values. In figure 4.5 we show a diagram of how the correlation matrix is built and in figure 4.6 how it is used to estimate an unknown illuminant by multiplying it by a binary vector containing the chromaticities present in an image. The result of this multiplication is a vector with the probability for each illuminant according to its chromaticities.

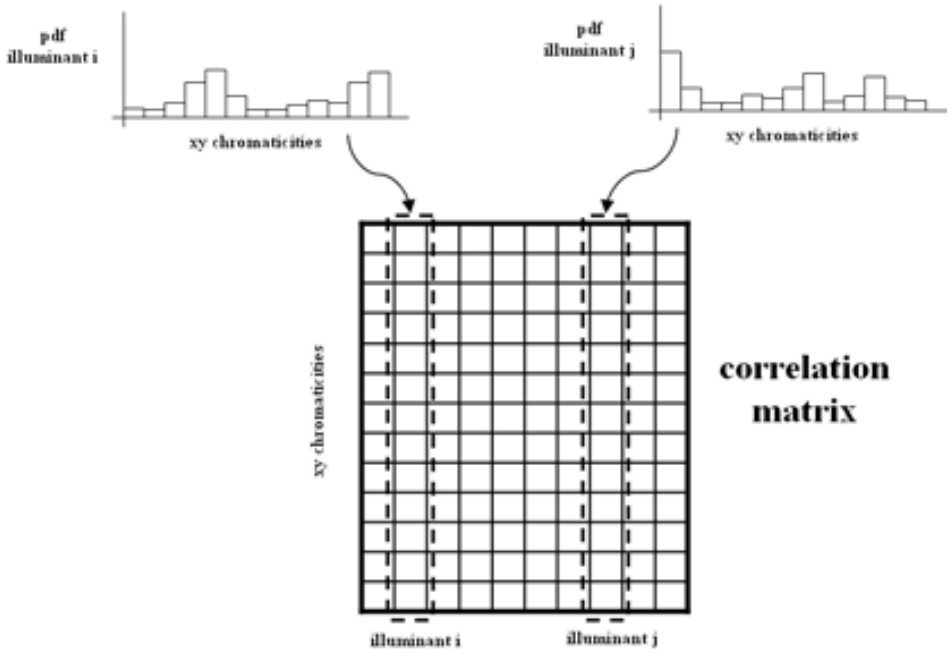


Figure 4.5: In the correlation matrix, information of probability distributions of chromaticities for different illuminants is combined.

Colour by correlation can only estimate illuminants considered in the computation of the correlation matrix, which could be a restriction when considering problems with no control of expected illuminants. Therefore a large set of assorted illuminants should be considered.

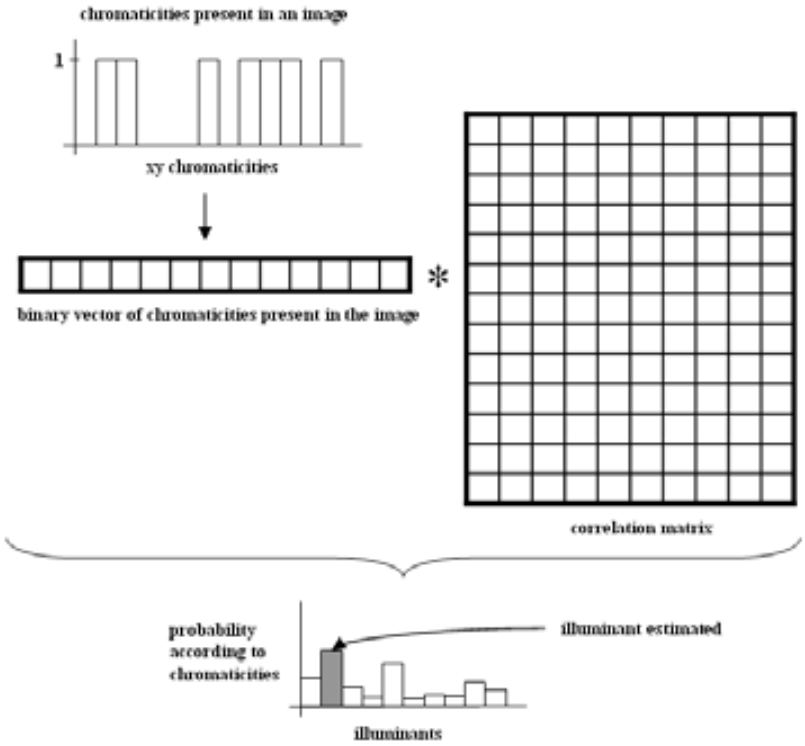


Figure 4.6: Information of chromaticities present in an image are combined with the correlation matrix to obtain a vector with illuminant probabilities. The illuminant with maximum probability is selected as the estimation of the illuminant.

In [9], Barnard proposed a modification to the proposed colour by correlation method to work in a three dimensional colour space, instead of working with just the chromaticities.

Brainard's bayesian colour constancy proposes to use prior information about the world to estimate the most probable illuminant and reflectance for a given scene. The basic idea is to use Bayes rule to estimate the spectra of the illuminant and the reflectance of the surfaces given the values of an image. Determining the prior probability of the scene parameters, spectra of the illuminant and reflectance of the surfaces, is a crucial step in the Bayesian estimation approach, and they are derived from standard illuminants and the Munsell chips respectively. This method deals with the recovery of illuminant spectra and reflectance of surfaces, which is not in the interest of this work, but it is interesting the idea of introducing prior information of illuminants and surfaces in the scenes.

Sapiro in [74] proposed an algorithm for estimating the scene illuminant which uses Bayes rule as well. The method is based on the probabilistic Hough transform, which is used for a voting process, and needs the computation of a matrix that contains expected reflectance information. Hence, this method also introduces prior information on expected surfaces.

Like gamut mapping approaches, bayesian methods need information of the sensor device used in the acquisition to work, and therefore they can only be used in problems where this information is known. Again, either the correlation matrix or the prior information of expected surfaces are computed for a given sensor, which constrains the images in which they can be used.

4.3 Relaxed grey-world

Considering existent colour constancy methods and after having analysed them, we have arrived at a conclusion: either under calibrated or uncalibrated conditions, the problem of the recovery of the illuminant of a scene is an ill-posed problem, and therefore the number of valid feasible solutions is bounded but infinite. Only by introducing more assumptions or constraints, can the space of solutions be reduced, and in this way we can select feasible solutions to the problem.

Gamut mapping assumption, seen in section 4.2.2, has been widely used to constrain the space of feasible solutions. Under calibrated conditions, the canonical gamut is the gamut of the known sensor under a white illuminant, considering the maxim number of feasible surfaces. Under uncalibrated conditions, we will assume that it is the colour space defined by the HVS. However, we should consider more assumptions to further constrain the feasible set of solutions. To this end, we propose the following hypothesis:

Hypothesis 2 *Considering that current methods for colour constancy introduce several physical constraints to reduce the number of solutions, it is possible to introduce high-level visual cues to complement and relax them to further constrain the problem and improve the performance of the algorithms.*

Considering existing colour constancy approaches, we will try to improve them by the introduction of a visual cue related to colour constancy: colour matching. Colour matching, also known as assymetrical surface matching [42], is a visual task where the HVS pairs coloured surfaces in two different scenes under different illumination conditions. In next chapter we will get into it with more detail.

Since colour matching has been proved to be an important visual cue for colour constancy, we propose to introduce a colour matching process to improve computational colour constancy [94]. The surface matching process we present wants to match every image surface with a set of canonical surfaces, i.e. surfaces seen under canonical acquisition conditions. The initial idea was to generate all the possible combinations of matchings. Even when using a reduced and significant set of image surfaces, what could mean to consider only surfaces in the convex hull of the image gamut, and a small set of canonical surfaces, the set of pairs of matches that need to be computed was too large and also lots of non-consistent pairs of matchings were introduced: if a reddish image surface was matched with a bluish canonical surface, it was not coherent to match another bluish image surface with a reddish canonical surface. This constraint is known as relational colour constancy [43, 69, 42], i.e. the relation between the colours remain the same despite changes in the illumination conditions (figure 4.7).

In order to introduce this high-level visual cues, we choose the existing grey-world assumption, which is used for uncalibrated conditions methods. The grey-world assumption, as depicted before in [19], supposes that the average of chromaticities of an image is grey. Even though it is a strong assumption, it has been widely used since it works in different situations, when considering images with a wide range of chromaticities. Thus, it can help us to introduce a constriction that maintains the colour structure of the image gamut when performing the surface matching process. In order to relax the assumption we propose another one, the relaxed grey-world assumption:

Assumption 2 *The image gamut under the canonical illuminant contains grey or its average is close to grey.*

Considering this new assumption the set of canonical surfaces that can be paired with each image surface is reduced to the canonical surfaces which are close to the image surfaces when the grey-world map is applied to the image, figure 4.8. A subset of canonical surfaces are selected to be matched with each image surface. Since a perfect match is not possible, we match with the nearest canonical surfaces. In this way, the relaxed grey-world assumption is introduced in order to find the solutions near the grey-world, enabling some sort of flexibility near this solution.

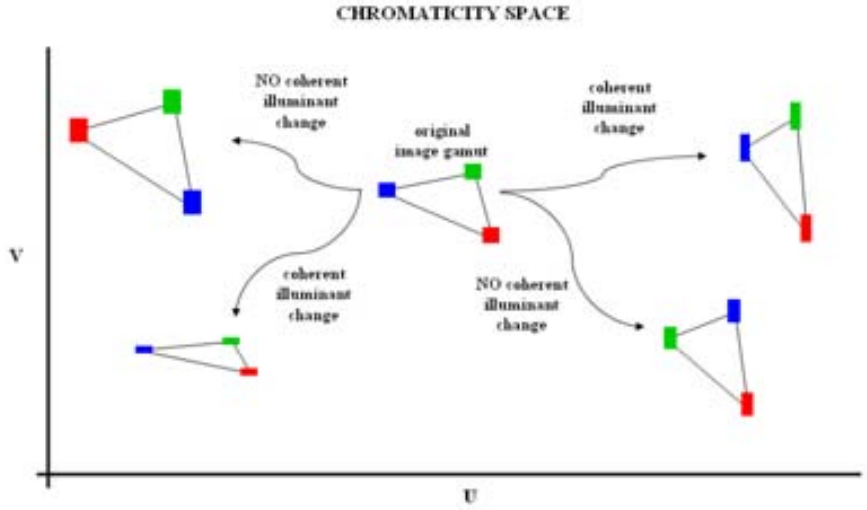


Figure 4.7: Relational colour constancy assures that relation between colours remain the same despite changes in the illuminant conditions: an illuminant change cannot alter the relation between colours in the image.

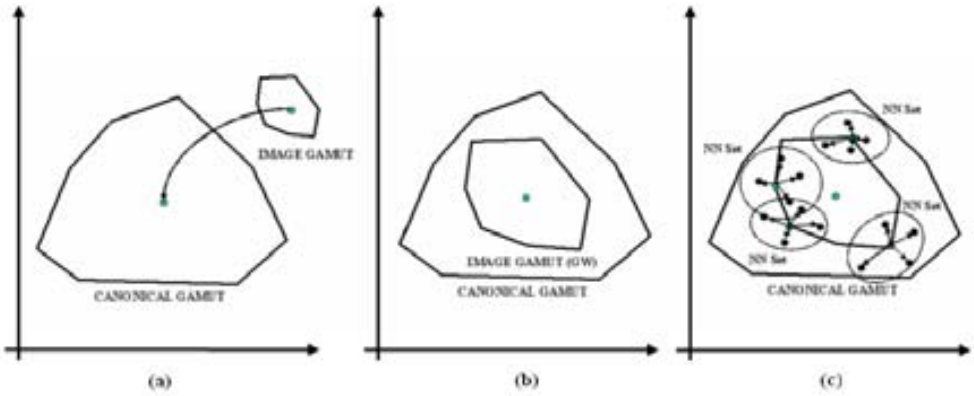


Figure 4.8: With the relaxed grey-world assumption, we have to find a set of nearest-neighbour canonical surfaces for each image surface, when the grey world map is applied. The image is maped to the center of the canonical gamut (a),(b) and the nearest-neighbour canonical surfaces for each image surface are selected as possible pair for the matching process (c).

The combination of the relaxed grey world asumption with the surface matching process motivate the new approach we present here. The proposed method matches each image surface with a subset of canonical surfaces that we have previously selected. The selection criteria has been to get a subset of essential colours that are likely to appear in the real world. One good set of colours are those presented in the Macbeth Color Checker Chart, since are close to fulfill this criteria.

Since after the grey-world map we do not have a unique perfect match with the canonical surfaces, we get the set of nearest neighbours to the image, i.e. the canonical colours that are near a neighbourhood region in the grey world transform. The method can be divided in two parts:

- **canonical surface selection**

  Computing the canonical surfaces as prior knowledge in the surface matching process, that is to select a representative set of surfaces and compute their RGB values for the canonical illuminant, which is selected to be well balanced with the sensor used. Hence we obtain a set of k canonical surfaces, denoted as $S^C = \{S_1^C, S_2^C, \ldots, S_k^C\}$.

- **colour matching**

  For a given image, I, acquired under an unknown illuminant U, the matching algorithm is carried out through the following steps:

1. Getting RGB values of significant surfaces from the image I, denoted as $S^U(I) = \{S_1^U, S_2^U, \ldots, S_n^U\}$, where n is the number of significant surfaces. Significant surfaces are those that depict the image gamut.

2. Applying the grey world transform to $S^U(I)$, which places the center of the image gamut in the center of the canonical gamut (figure 4.8 a,b). It is denoted as $S^{GW}(I)$.

3. For each surface, $i = 1 \ldots n$, in $S^{GW}(I)$ we select the m nearest neighbours surfaces from the canonical surfaces (figure 4.8 c), S^C, we denote each of this subsets as S_i^{NN}.

4. Computing the set of all possible matchings between each S_i^U with the surfaces in the corresponding S_i^{NN}, we bring together the matchings for all S_i^U and name this set $RCorr = \{S_1^U = S_{1,p_1}^{NN}, S_2^U = S_{2,p_2}^{NN}, \ldots, S_n^U = S_{n,p_n}^{NN}; \forall p_i = 1, \ldots, m\}$, where $\#RCorr = m^n$.

5. For each element of $RCorr$, the corresponding $\alpha\beta\gamma$ map (eq. 4.1) is computed, and we obtain a set of maps, $MAP_{\alpha\beta\gamma}^{RCorr}$.

6. All the maps in $MAP_{\alpha\beta\gamma}^{RCorr}$ out of the feasible set are removed, as we do not want to deal with impossible maps.

Once we have generated the set of maps, $MAP^{RCorr}_{\alpha\beta\gamma}$, the last step is to use one of the existing heuristics to select a map within this sub-feasible set of solutions. Next we will show the results using the heuristics of maximising image gamut volume and selecting the average map of the set. A simplificated scheme of the process proposed by the method can be seen in figure 4.9.

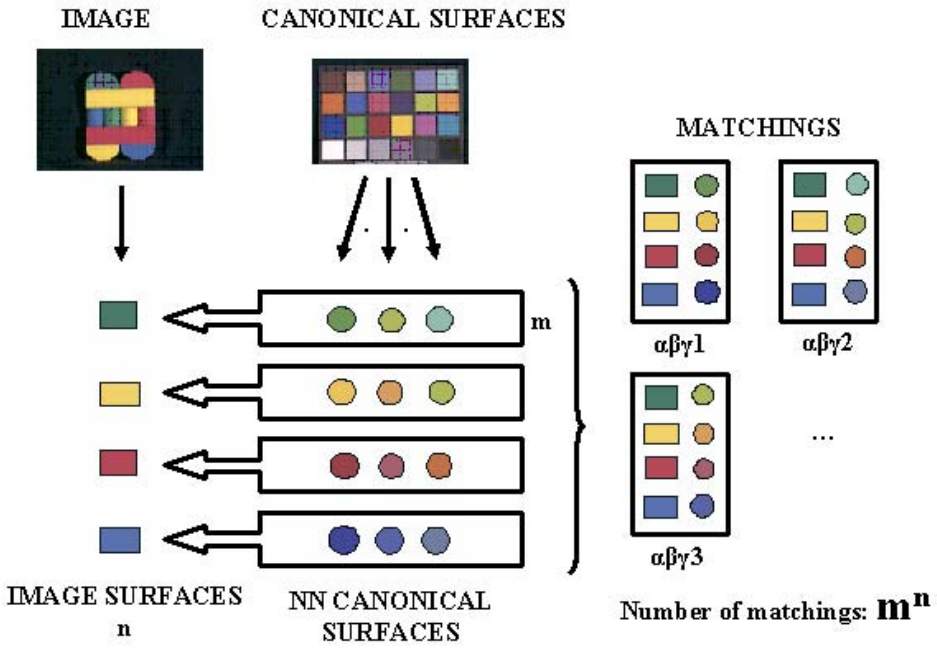


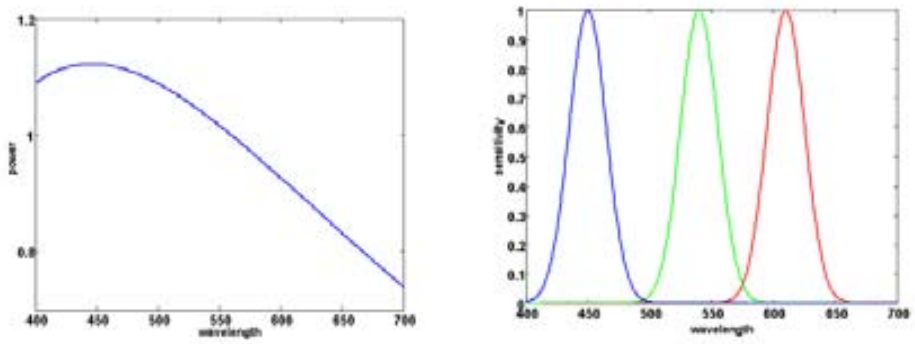
Figure 4.9: An illustration of how the relaxed grey world algorithm proceeds.

The method uses concepts introduced by gamut mapping methods, such as considering $\alpha\beta\gamma$ diagonal maps as illuminant changes and the restriction of a feasible set of solutions, and also by bayesian methods by introducing prior knowledge of canonical surfaces, which are the coloured surfaces that we will expect to found in the images.

4.3.1 Experiments and Results

To evaluate the method we have looked at its performance on an experiment using synthetic data. This is a first way to evaluate algorithms because performance is not affected by image noise and we are able to evaluate performance over a large amount of synthetic images and thus obtain a reliable performance statistic. Otherwise, with real data these problems arise, and also we consider that the available datasets to evaluate computational colour constancy are not large and assorted enough to extensively test the method.

First, we had to select the canonical acquisition conditions, that is a standard illuminant and a sensor calibrated for this illuminant, which therefore is white balanced. We have chosen a synthetic planckian illuminant with CCT=6500K (fig. 4.10 (a)) and built a gaussian narrow-band sensor, with centers in 450, 540 and 610 nm (fig. 4.10 (b)). Then, we had to choose the reflectances to be used as canonical surfaces. In our case we have selected the 1995 reflectances of the Munsell chips [23]. Taking the canonical illuminant, we have synthesised the RGB values of these reflectances to obtain our canonical set of surfaces.



(a) (b)

Figure 4.10: The synthetic illuminant (a) and sensor (b) used in the experiments.

Once we have built the base of prior knowledge of canonical surfaces, we have generated synthetic images to test the method: 400 images consisting of 10 reflectances per image (from Munsell chips randomly selected) under a random illuminant, chosen from a widely used selection of 11 different illuminants [6]. As remainder parameters of the method, we have selected 6 significant surfaces from each image (considering the 6 most representative surfaces from the convex hull of its gamut) and taken their 5 nearest neighbours surfaces from the canonical surfaces for the grey world transformation, that is $n = 6$ and $m = 5$.

To measure the performance of the method, we have used as recovery error the angular error between the RGB of the estimated illuminant, $\widehat{RGB}_w^C$, and the RGB of the canonical illuminant used, RGB_w^C. This evaluation of computational colour constancy methods has been widely used previously, and it is explained in [6]. The RGB values of the illuminants are usually unknown in real images, but they can be computed easily working with synthetic data.

$$\text{recovery error} = angle(\widehat{RGB}_w^C, RGB_w^C)$$

Heuristic	CRULE	Relaxed Grey-World
Maximum Volume map	7.09º	7.55º
Average map	9.35º	6.62º

Table 4.1: Comparison of the performance of the two methods. The value shown is the root mean square of the angular errors computed for the 400 synthetic images.

In table 4.1 we show the performance of the proposed method and compare it with Forsyth's CRULE [40], since it is considered a powerful colour constancy method that usually achieves best results [6]. We have considered two heuristics to select the optimal map within the feasible set of solutions generated in each case: maximum volume map and average map. The heuristic of the maximum volume map is often used, since it takes the map within the feasible set that delivers the most colourful solution and it delivers good, in some cases best, performance. The results obtained vary depending on the heuristic used. In the table, we can see that the best performance is achieved with the proposed relaxed grey world method, taking the average map of the computed maps. Therefore, our method implies a slight improvement of the colour constancy achieved by CRULE, which reinforces the use of the assumptions proposed.

However, we are interested in the type of solutions generated with our method. In addition, we show in figure 4.11 a comparison of the different sets of maps generated with the two algorithms considered, they are plotted acording to their recovery angular error (x-axis) and the volume of the image gamut generated (y-axis). Yellow dots are the solutions generated with the relaxed grey world approach and blue dots are the feasible set of solutions generated with CRULE. We can observe that comparing with CRULE, with our method we avoid to generate a large set of maps, and also to include maps with an important recovery error. We look for a reduced set of maps which usually includes the best solutions. In this sense we have computed the average value of the best angular error produced by the generated solutions for each of the 400 images and it has resulted to be 1.9º, which means that an optimal map is included in our subset of feasible maps in most cases. Both results, improving of the recovery angular error and reduction of the feasible set of maps generated, are important results that justify the proposed method.

Summarising, we have presented the relaxed grey-world method, with a similar performance to CRULE, which improves the original grey-world method and computes a reduced feasible set, and that delivers an acceptable error. This proves hypothesis 2 presented at the beginning of section 4.3.

4.3.2 Discussion

It has been proven that the introduction of the relaxed grey world approach to solve computational colour constancy opens a new line of research in this problem, that can

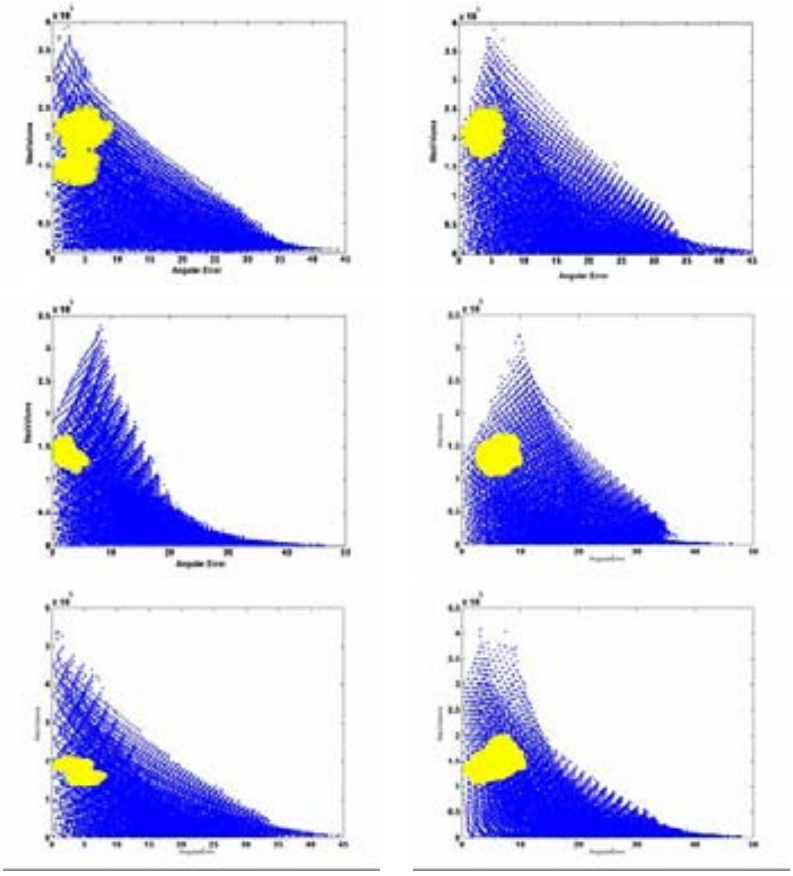


Figure 4.11: Comparison of the sets of maps generated with CRULE (blue dots) versus the set of maps generated with our method (yellow dots) for 6 different images. In the x-axis it is represented the recovery angular error and in the y-axis the volume of the image gamut generated.

help in reducing the recovery error of current methods, which ignore image information that can be introduced by surface matching. The selection of the canonical surfaces is an important step to pay more attention and to be focus of a deep study, but it is proposed in the future research lines of this work. Indeed, the selection of canonical surfaces used in our experiments has been roughly done, but it is just a first approach to introduce the colour matching cue, and it has delivered promising results. We think that further work needs to be done in the selection of the set of canonical surfaces, as they should represent more trustworthily the knowledge of more likely colours.

4.4 Performance on computational colour constancy

In [93] we have presented an exploration of the space of solutions, using the reduced feasible set obtained with the method proposed. To improve the existing heuristics in the selection step, we have explored different measures, existing and new ones, such as volume map, distance to the grey world solution, etc. and the experiments lead us to conclude that other criteria, such as a combination of the explored measures, could be defined in order to better approximate the optimal solutions. However, it is a complex problem wich would require a much deeper study. In this section we analyse the existing performance of colour constancy methods, to see whether it is interesting to consider it in an uncalibrated frame.

Many computational colour constancy methods have been proposed in the literature, along with an evaluation of their performance. Most of the existing evaluations looked at how acurately the illuminant of the scene is recovered or how well it can be converted to a reference illuminant, which is normally calibrated. Normally, the methods proposed are first evaluated with synthetic data [6], which is noise free and experiments can be performed exhaustively through a large amount of data, and after they are tested with real image data [7], which normally is an assorted but not exhaustive set of images of scenes acquired under different illuminants. Recently, Hordley et al. in [53] have proposed some interesting modifications for evaluating colour constancy to present more reliable results, using the median rather than the most common used root mean square (RMS) of the angular error and, also, using the Wilcoxon sign test [52] to compare the performance of the methods.

At this point, we arrived to a conclusion: evaluation of colour constancy methods has been based on measuring whether the proposed assumption agreed with the actual solution the image was acquired. In practice, images do not fulfill a concrete criterium, and it is usual to deal with images that do not 100% fulfill the assumptions:

- Scenes with a non average grey chromaticity.
- White do not correspond to maximum intensity.
- The scene does not present a maximum diversity of colour.
- Etc.

The evaluation method usually used depends on the data set used. When most of the images in the data set fulfill the assumption introduced by the method, the method works and delivers good performance, otherwise not. If we want to evaluate a colour constancy method, it is important to consider this dependency of the data set. To answer this question we have proposed the following hypothesis:

Hypothesis 3 *The performance of colour constancy methods is determined by the number and chromaticity distribution of illuminants in the data sets, and also by the type of images or reflectances considered.*

In experiments with synthetic data, normally are used the reflectances of the Munsell chips, a set of measured spectra of illuminants (as in [6]) and some sensor calibrated for one of these illuminants, which will be considered as the canonical. If we want to perform colour constancy experiments with real image data, there only exist two widely used image data sets [46, 8] that come along with information of the acquisition conditions, such as spectra of the illuminants and sensibility curves of the sensor used, that helps in the evaluation process of the methods. The construction of these image data sets is rather laborious and complex, and normally it must be focused to evaluate some concrete type of performance. For this reason, we think it is difficult to find exhaustive enough real image data sets to test colour constancy methods.

If we focus on the evaluation for synthetic data, images are composed of a set or RGB values, which represent the surfaces in the image. The RGB values are built from a randomly selected set of reflectances and also a random illuminant is selected from the available data. In many previous works it has been shown that, the more surfaces are in the image the lower is the error. Then, to test a method, we measure how well the chromaticity of the selected illuminant is estimated. To this end, we have analysed the illuminant sets proposed in [6], since an extensive comparison of colour constancy methods can be found in it, and its data has been widely used in other experiments. We thought that performance measures of the methods tested might depend also on the number and chromaticity distribution of illuminants in the data sets.

In figure 4.12, we show the xy chromaticity coordinates of the first set of illuminants proposed, consisting of 11 different real illuminants. We can observe that they do not enclose a large amount of different chromaticities, and they are mostly limited to the planckian locus chromaticity region.

In figure 4.13, we show the xy chromaticity coordinates for the two other sets of illuminants in the Simon Fraser data set. This other illuminants have been obtained by linear combination of the 11 initial illuminants. In the figure we can see that since these two sets have been obtained by linear combination of the previous ones, the chromaticity region enclosed by them is roughly the same as by the initial set.

In favour of these illuminant sets, we have to say that even though they do not represent all the range of chromaticities, and are just enclosed to a small chromaticity region, this chromaticity region, the planckian locus, is where most natural and

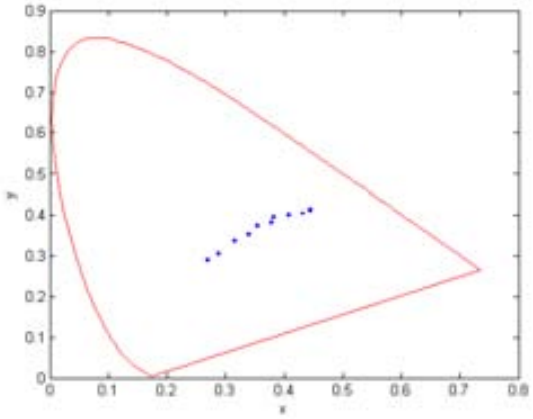
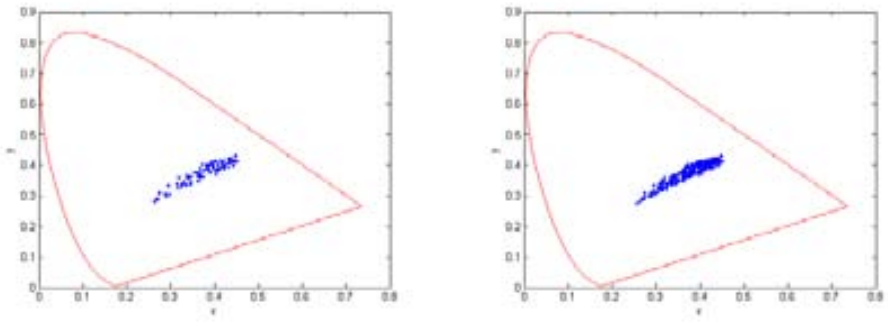


Figure 4.12: xy chromaticities of the set of 11 illuminants of the Simon Fraser data set.



(a) (b)

Figure 4.13: xy chromaticities of the set of 87 illuminants (a) and the set of 287 illuminants (b) of the Simon Fraser data set.

Synthetic illuminant set	Chromaticity area enclosed	Recovery angular error
1	0.0001	7.9294
2	0.0341	10.7785
3	0.0729	11.7745
4	0.1111	12.3573
5	0.1508	17.1378
6	0.1858	17.2140
7	0.2176	16.2176

Table 4.2: Results of the recovery angular error (RMS over 1000 images in each set) for the different illuminant sets proposed using CRULE.

artificial illuminants lay. However we might be interested on the performance of the methods in extreme illuminant conditions, i.e. when a colour filter is placed in front of the illuminant, and unknown illuminants can have any chromaticity. In this case, we think that the performance of the methods will be limited by the area of the chromaticity region enclosed by the illuminants used.

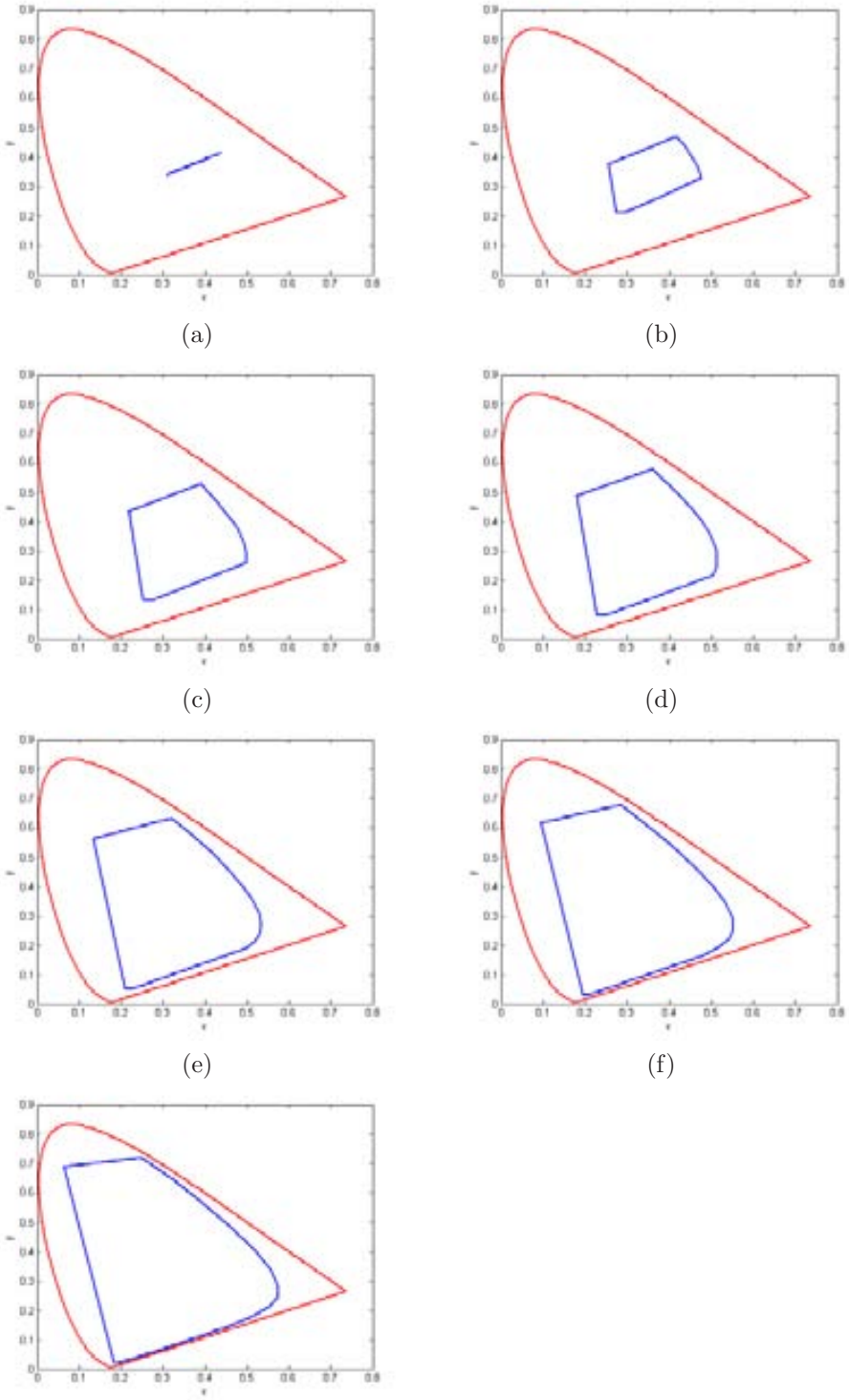
To this end we have performed an experiment with synthetic data to see if the chromaticity range of the illuminants considered has some effect on the performance of the colour constancy methods. We have just used Forsyth's CRULE method for the experiment, since we want to focus on the evolution of a single method performance and CRULE has been extensively tested and proven to be a robust method. Then we have created different synthetic illuminants in the range of all the possible chromaticities and have arranged different sets which enclosed increasing chromaticity area. In figure 4.14 we show the chromaticity area enclosed by the 7 different illuminant sets generated.

Then we have generated 1000 synthetic images consisting of 8 surfaces for each of the 7 illuminant sets and measured the RMS of the recovery angular error using CRULE. In table 4.2 the results of the angular error along with the chromaticity area covered by the illuminants are presented.

We can observe that the angular error increases when the range of illuminants, and its chromaticity range, is increased. Hence, considering the dependence of evaluation measures on the data used in the experiments, we have derived the following conclusion:

Conclusion: The solutions within the feasible set are equally good, unless we introduce high-level information. Therefore, the performance of existing selection methods has been determined by the image contents and the chromaticity of the illuminants.

The measure of the recovery angular error of estimated unknown illuminants is a proper way to test colour constancy when we aim to achieve perfect colour con-



(a) (b) (c) (d) (e) (f) (g)

Figure 4.14: Chromaticity area enclosed by the 7 different synthetic illuminant sets created for the experiment (a-g).

stancy and it is useful to evaluate methods that are designed to be used in camera calibration, since it delivers a measure of how well colours are white balanced in an image. However, this goal might not be that important in computer vision, which is the interest of our work. In computer vision we do not always, not to say rarely, know the sensor information of the acquisition devices used, information which is needed in most colour constancy methods, what clearly indicates that they are aimed to deal with camera calibration. We might deal with a computer vision system with a camera, whose sensitivity curves are not known and no calibration process is possible, or deal with images from unknown origin, such as internet images, and we would like to perform a colour constancy process to balance the colours in an image, when possible, for a post-processing of image understanding. In image understanding, we think that it might be more interesting to obtain useful solutions for this post-processing, rather than to achieve a perfect white balanced image.

The experiment performed proves hypothesis 4: if we alter the illuminant sets used in the experiment, the performance of the methods changes. This dependency from the test data set demonstrates that, under uncalibrated conditions, it is nor reliable nor useful to work with unique optimal solutions. With this goal in mind, in next chapter we present a method for white point estimation under uncalibrated conditions, focused on delivering practical solutions, with a colour interpretation of the surfaces in the image, according to the problem we are dealing with.

Chapter 5

Semantic white point estimation

In the frame of computer vision under uncalibrated conditions, we present a white point estimation method which proposes multiple solutions that can deliver different coherent interpretations of the scene. In this chapter we explore furhter colour matching as a visual cue to solve white point estimation. Firstly, we present the necessity of considering multiple white point estimations within the context of image understanding. Then we propose a semantic matrix to introduce previous knowledge about the scene content. This previous knowledge will be used in a process of colour matching to perform a voting within the set of feasible solutions. A cross-correlation is computed to generate a weighted feasible set, which will be used to select interpretable solutions. Finally, we present some results considering local maxima in this feasible set and the computing of ridges to compact the weighted feasible set losing minimum information.

5.1 Introduction

Following the two contexts for image annotation proposed in section 1.3, calibrated and uncalibrated conditions, we differentiate colour constancy methods in two groups regarding the context in which they can work. Therefore, we separate methods in two groups, those which need information of the acquisition system and those which do not require any previous knowledge of the imaging process. We place gamut mapping and bayesian methods in the first group, since sensor information is needed to build the canonical gamut or the correlation matrix used in the process, whereas maxRGB and grey world methods are in the second group. Thus, we name methods in the first group calibrated methods, regarding the necessity of information which is normally used for calibration, and methods in the second group uncalibrated methods.

In last chapter, we concluded that performance on colour constancy methods must be carefully taken into account. Also, we arrived at a more interesting conclusion: under uncalibrated conditions and when we aim to interpret an image, it is not prac-

tical to look for a unique optimal solution, but for different interpretable solutions. In this way, high level information needs to be introduced to find these white point estimations related to a colour interpretation of the images. This high level information could be introduced in the form of assumptions of the scene content. Foster in [42] shows that colour constancy is aided by visual cues, such as colour matching. Also colour naming [10], a particular case of colour matching which involves higher level information, helps in colour constancy. Therefore, we propose the next hypothesis to estimate the white point of an image:

Hypothesis 4 *In a frame of uncalibrated conditions, it is possible to obtain coherent interpretations of the illumination of an image by introducing assumptions of the scene content.*

This information about the context of the scene can be introduced in many different ways. We propose to use a colour matching process considering coloured surfaces with a name. Hence, we introduce an assumption, nameability assumption, to guide the process:

Assumption 3 *The adequacy of the white point estimation of an image is based on the assignment of a number of known names to the surfaces within it.*

We propose a method which considers gamut mapping and bayesian methods, and readapt them to a frame of uncalibrated conditions. Since for uncalibrated conditions, colour constancy is less constrained than for calibrated conditions, there is more freedom regarding the unknown sensor, we have considered to introduce more assumptions to the problem to reduce the space of possible solutions.

In this way, we present an approach to deal with uncalibrated white point estimation using colour matching, which psychophysics relate to colour constancy [42], as a way to introduce assumptions and to constrain the problem. We are not interested in looking for a single solution, but to give different possible interpretations regarding the coloured surfaces that we *expect* to find in images, trying to perform a similar process to HVS during colour matching. The degree of colour matching of the different possible solutions is measured through a correlation process, in a similar way to the voting process proposed in [74].

Existing colour constancy methods try to estimate the illuminant of a scene, $E(\lambda)$ from equation 2.3, and in our framework we propose to estimate a white point, considering that we do not know the acquisition conditions, which comprise both illuminant and sensor information, $A_k(\lambda)$ from equation 2.6. In figure 5.1 we show a problem that can be found in computer vision, in the frame of image understanding. The question 'what is the colour of an object in an image?' might have different answers if we consider the colour constancy problem, and different configurations might be valid.

In the figure 5.1 we can observe a synthetic image of two apples. We do not know the acquisition conditions of the image, and hence, there can be different interpreta-

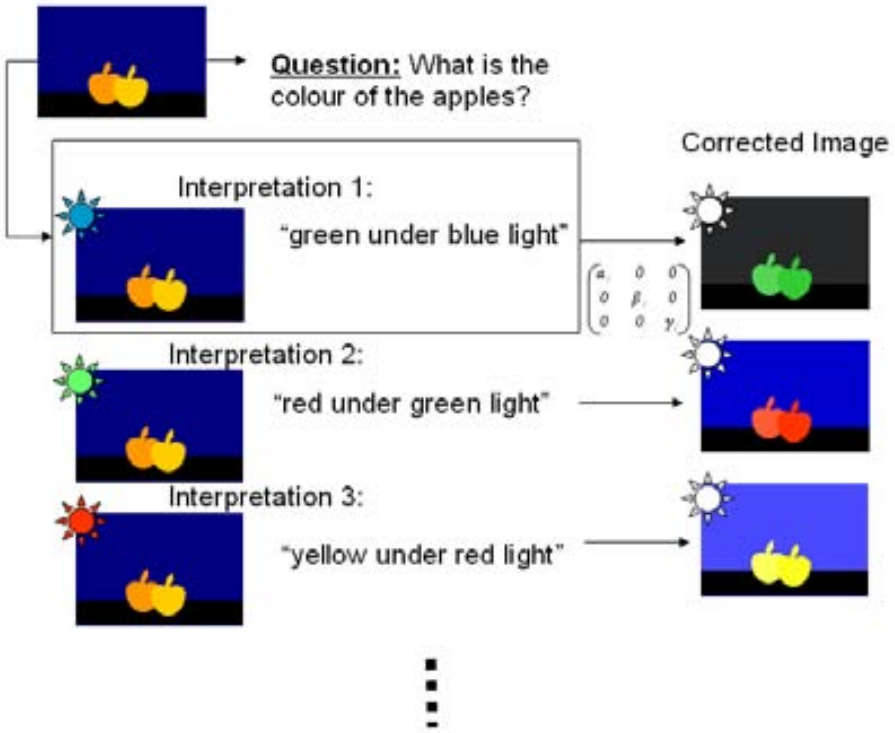


Figure 5.1: Problem in computer vision: 'what is the color of the apples?'

tions of the scene. We show three different possibilities, but there could be more. In the first interpretation we suppose a bluish illuminant and thus the apples would be green. In the second interpretation it is supposed a greenish illuminant and the apples would be red. Finally, in the third interpretation we suppose a reddish illuminant with yellow apples. These three interpretations are valid for us when we look at the images, since we know that there exist apples of these colours. More interpretations could be considered, which might give unrealistic colours of apples (such as purple, blue, etc). In any case, interpretations delivered must maintain the relation between colours to be feasible, which is a restriction introduced by relational colour constancy, i.e. relation between colours remain the same despite changes in illumination [42].

With this approach, we do not aim to achieve perfect colour constancy, since there exist more than one solution equally feasible. Nevertheless, under uncalibrated conditions we think that the colour constancy problem should be considered and we propose a method to interpret the acquisition conditions of a scene under this conditions, using a procedure in the frame of gamut mapping methods and introducing some ideas of the bayesian approaches to restrict the problem. We propose to perform a colour matching process, which introduces restrictions on the surfaces in the scene, in order to further constrain the problem. We will not consider just a single solution, or interpretation, but a set of different likely estimations of the white point, which agree with the surfaces that we expect to find in the scene. Also, more restrictions could be introduced and combined in the method , such as illuminant restrictions, to further constrain the problem, but they are out of the scope of this work.

5.2 Colour matching for image interpretation

HVS performs colour constancy with the help of several visual cues [80, 44, 42]. We propose to introduce one of these visual cues in the method we present: colour matching. The colour matching process consists of matching pairs of surfaces under different illuminants, which are perceived as the same surface. The problem is usually presented in the way that we have to pair a set of surfaces under an illuminant with a set of surfaces under another illuminant (figure 5.2). Other variations of the surface matching problem have been proposed [3, 17]. In this visual process we seek to ignore the effect of the illuminant and match the surfaces according to their reflectance properties. Colour matchings are constrained by relational colour constancy, which assures that the relations between the colours are maintained within the matchings. The performance of HVS in colour constancy for the surface matching problem has been evaluated in [42, 3, 17]. Colour matching has been previously considered for dealing with varying illumination in [14] for an object recognition problem.

To perform a colour matching process in our method, we have to deal with computational data. In this way, we will use colour surface descriptors instead of the physical coloured surfaces used in psychophysical experiments. We propose to define a set of coloured surface descriptors, and restrict the possible solutions so that we have

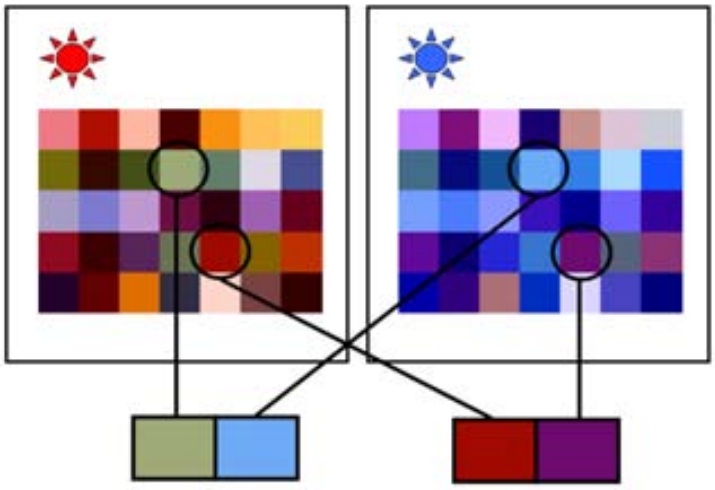


Figure 5.2: In the colour matching problem, we have to pair surfaces of the same scene that are perceived as the same colour under different illuminants, in the example a red and a blue light.

as many surface descriptors as possible of the image within this set. We will consider those white point estimates as likely candidates which give a high degree of colour matching. We will perform a colour matching process between the coloured surfaces in an image and the set that we have defined. Hence, images with coloured surfaces which are not actually within our set of coloured surfaces might not be solved properly.

The method we propose can be adapted to work with different colour sets, depending on the type of images we are dealing with, and it is a way to impose restrictions on what we expect to find in the image. Therefore the method is in the paradigm of purposive vision [1], where the process is task-oriented and where, to solve the problem, information about what do we want to solve is added. In our case, we introduce assumptions on the coloured surfaces that we expect to find in scenes we want to analyse.

In this work, we will consider, as previous knowledge, surfaces of synthetic colours (red, green, blue, etc.) and deal with images with these types of coloured surfaces. Other colour sets, depending on the scenes that we want to deal with, might be used to work with other sort of images, e.g. natural colors, where we could have 'leave' colour, 'sky' colour, 'stone' colour, etc.

5.3 The semantic colour matrix

To introduce the set of colours for the colour matching in the method we use a semantic colour matrix. To include the colour information in this matrix we will only

consider values of colours with an unambiguous name or meaning, i.e. coloured surfaces which have probability 1 to belong to a colour category and probability 0 to belong to the rest of the colours.

To build the semantic colour matrix, denoted as SM, we need a set of colour focals, which are the unambiguous colour descriptors defined before. Then, we need to define the canonical acquisition conditions, i.e. the conditions within which we want to match the surfaces and the conditions that we want to colour correct our images to. If we want to consider a general frame for colour understanding, we should take as canonical acquisition conditions a standard illuminant i_s, e.g. daylight, and a standard observer sensor o_s, e.g. the colour matching functions. The sensor must be calibrated for the corresponding illuminant. Thus, the canonical acquisition conditions $A_k^c(\lambda)$, following eq. 2.5, can be defined as

$$A_k^c(\lambda) = E^{i_s}(\lambda) R_k^{o_s}(\lambda) \tag{5.1}$$

Also we can define Γ_c as the canonical colour gamut for these acquisition conditions. Let $S_i^{cn}(\lambda)$ be a set of n reflectances that correspond to colour focal surfaces, $i = 1..n$. We define sensor responses of these focal surfaces under canonical acquisition conditions as

$$\rho_{k,i}^{cn} = \int S_{k,i}^{cn}(\lambda) A_k^c(\lambda) d\lambda \tag{5.2}$$

Now we can define the semantic surface matrix, SM, which stores the points of Γ_c that belongs to a semantic category with probability 1.

$$SM(\vec{c}) = \begin{cases} 1 & \text{if} \quad \vec{c} \quad \exists \quad in \quad \rho^{\mathbf{cn}} \\ 0 & \text{elsewhere} \end{cases} \tag{5.3}$$

where $\vec{c}$ is a colour vector, $\vec{c} \in \Gamma_c$. SM is a $k-$dimensional matrix, where k is the dimension of the colour space (i.e. the number of sensors of the acquisition device), which is usually 3 in colour images. For $k = 3$, a normal dimension of Γ_c is $255 \times 255 \times 255$, and SM has the same dimension as Γ_c. SM introduces the restrictions of expected colour in images. Thus, we can build different SM matrices according to the types of images we have to deal with. In figure 5.3, we show some possible matrices that could be constructed for synthetic images, natural images, and so on, depending on the categories we want to use. In the following section we explain how to use this matrix SM to give different interpretations of the white point of a scene.

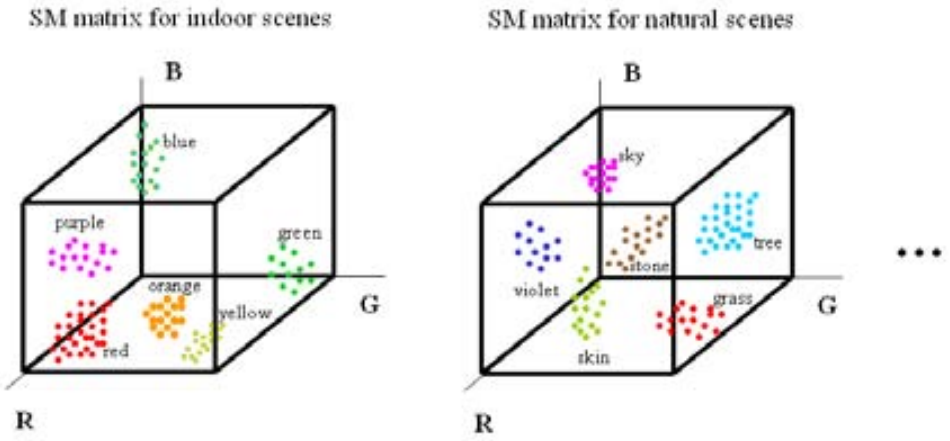


Figure 5.3: Possible SM matrices target oriented: depending on the types of images we have to deal with we have to construct the corresponding SM matrix.

5.4 A Weighted Feasible Set

The SM matrix is used to perform a colour matching process to find different white point interpretations for a given image. The idea of the method is to build a weighted feasible set of solutions, correlating SM with the gamut of an image, which must be transformed to estimate each possible white point. We will not only consider the acquisition conditions corrections that take the image gamut inside the canonical gamut, but inside the rgb cube, since it will reduce the complexity of the problem. Though in this way the problem is less constrained, we restrict it subsequently when matching only with the colour focals represented in our matrix. The weight of each map indicates a degree of colour matching for the given estimation. The local maxima in this weighted feasible set indicate likely surface matchings, which we will consider as likely white point interpretations. Next, we give the procedure steps of the depicted method.

The correlation between SM and the image gamut we propose is complex if we notice that the image gamut varies in its size and is rescaled for each different diagonal illuminant change (see eq. 4.1). This means that to compute each correlation value a different image gamut should be calculated. In [13] a similar problem is solved by introducing the log space. The acquisition conditions change in log space is a translation of the image gamut instead of a scaling, and therefore the shape of the image gamut remains the same. In rgb space, rgb values of an image, rgb_1, are scaled by an acquisition conditions change, $\alpha\beta\gamma$, to obtain the corrected rgb values, rgb_2 (eq. 5.4),

$$\begin{pmatrix} r_2 \\ g_2 \\ b_2 \end{pmatrix} = \begin{pmatrix} r_1\alpha \\ g_1\beta \\ b_1\gamma \end{pmatrix} \tag{5.4}$$

In log rgb space, $log(rgb_1)$ values are translated by $log(\alpha\beta\gamma)$ to obtain the corrected $log(rgb_2)$ values (eq. 5.5),

$$log \begin{pmatrix} r_2 \\ g_2 \\ b_2 \end{pmatrix} = log \begin{pmatrix} r_1 \\ g_1 \\ b_1 \end{pmatrix} + log \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} \tag{5.5}$$

This is an interesting property that can be used in the correlation process we want to perform. Thus, we can define a log semantic colour matrix with log responses, SM', and correlate it with the log image gamut. Then, we can compute a direct cross-correlation between them, which will give us the correlation values for all the white point estimation changes. The correlation value in this log space has the same meaning as in the previous space: a degree of colour matching. We will only have to notice that the acquisition conditions changes proposed by each correlation value is a translation in the log space instead of a scaling in the non-log space.

Therefore, we define the log semantic colour matrix, SM', to contain information in the log space,

$$SM'(\vec{c'}) = \begin{cases} 1 & \text{if} \quad \vec{c'} \quad \exists \quad in \quad log(\rho^{\mathbf{cn}}) \\ 0 & \text{elsewhere} \end{cases} \tag{5.6}$$

where $\vec{c'}$ is the log of a colour vector and $\vec{c'} \in \Gamma'_c$ (Γ'_c is the gamut of the log responses under the canonical illuminant). The range ot the SM' matrix must be quantized in order to compute the matrix to work with.

Once we have built the SM' matrix, we can use it for colour image interpretation. Let I be an image to interpret, we define the histogram of the image in the log space as

$$H'(I) = hist(log(I)) \tag{5.7}$$

where $hist()$ is the histogram function. $H'(I)$ contains statistical information of the coloured surfaces in the image and we propose to combine it with SM' to obtain a surface matching correlation map with the colour focals. This operation is defined as

$$CM(I) = H'(I) \star SM' \tag{5.8}$$

The cross-correlation, $\star$, between $H'(I)$ and SM' in the log space delivers high values when colours in $H'(I)$ adapt to the chromaticities in SM'. Assuming Von Kries' diagonal model assures us to maintain relational colour constancy, and the relation between the coloured surfaces is preserved. Each correlation value is related to a different white point estimation change, considering the shift of $H'(I)$ within SM'. To perform the cross-correlation depicted in eq. 5.8, we propose to use the fast fourier transform [24] and perform a dot product in the frequency space, to reduce the complexity, $O(NlogN)$, and the computation time.

Thus, a diagonal acquisition conditions change, $\vec{q_i}$, in the non-log space is given by,

$$\vec{q_i} = e^{\vec{c_i}' - \vec{c_m}'} \tag{5.9}$$

where $\vec{c_i}'$ is a colour vector in the log-space with a correlation value and $\vec{c_m}'$ is the center of Γ'_c. In the case of a 3 channel sensor the diagonal change is $\vec{q_i} = (\alpha, \beta, \gamma)$.

In this way, a correlation value is related to a diagonal white point estimation change. The cross-correlation computed, $CM(I)$, delivers a voting map of surface matching, that is a measure of how well the coloured surfaces in the image do adapt to the focal colours in SM'. The peaks in this correlation map can be taken as white point estimations that give a high degree of colour matching. Hence, the different likely acquisition conditions interpretations can be found by taking the local maxima in this correlation map, $\widehat{u_i}$, $i = 1..n_s$, where n_s is the number of local peaks.

$$\widehat{u_i} = max(CM(I)) \tag{5.10}$$

The function $max()$ returns a local maximum in the correlation space. A point u_i is a local maximum of $CM(I)$ if there exists some $\epsilon > 0$ such that $CM(u_i) \geq CM(u)$ for all u with $|u - u_i| < \epsilon$. To find the local maxima we propose to use a *morphological tophat* transform [75, 25] to get regions with peaks, and within them we have taken the chromaticities with maximum correlation value.

Each $\widehat{U_i}$ must be related to an acquisition conditions change, $\widehat{q_i}$, using eq. 5.9. The set of $\widehat{q_i}$ are likely white point estimations. Thus, the different image interpretations, I_i, can be computed using the different estimations of acquisition conditions,

$$I_i = \widehat{q_i} I \tag{5.11}$$

and therefore obtain the result of the different sets of feasible colour matchings.

Summarising the method, first we have to select the semantic focals and build SM'. Then, for a given image I, the algorithm is carried out through the following steps:

1. Computing the histogram of the image in the log space, $H'(I)$ (eq. 5.7).
2. Computing the cross-correlation between $H'(I)$ and SM', using the fast fourier transform, obtaining a correlation map $CM(I)$ (eq. 5.8).
3. Extraction of the peaks in $CM(I)$, $\widehat{u}_i$, as different white point interpretations (eq. 5.10).

5.5 Experiments

We have tested the method with some images of the Simon Fraser database [6]. We have considered images with surfaces of synthetic colours. It is difficult to evaluate the performace of the method, as we do not aim to give a single solution, but a set of likely interpretations when it is possible. For the tested images, we will show the corrected image acording to the white point estimation and also their interpretations, i.e. the segmentation of the coloured surfaces that lead us to that interpretation.

5.5.1 Implementation details

Previously, we explain the parameters of the method used for the experiments. As canonical acquisition conditions we have used a D65 light as illuminant and the RGB Colour Matching Functions as a standard observer sensor. Then, as colour focals we have used those from colour naming derived from an exhaustive psychophysical experiment in [12] to build SM' (eq. 5.6). These focals comprise eleven synthetic colours most used in the language: red, orange, brown, yellow, green, blue, purple, pink, white, gray and black. We have also added the focals for skin colour, in order to deal with common indoor scenes. We have used 50 bins per channel in order to generate the histograms used in the processing. This SM' matrix proposed can be used to process any uncalibrated image, as it is given in CIE rgb (eq. 5.1), i.e. a standard space. We assume that any image can be transformed to these canonical acquisition conditions through a diagonal change. In figure 5.4 we show the SM' matrix created with the proposed configuration, and that will be used in the subsequent experiments.

5.5.2 Interpretation of images

In the first experiment, figure 5.5, we can observe that for a given image, a scene of some apples, our method delivers two different interpretations, along with the original

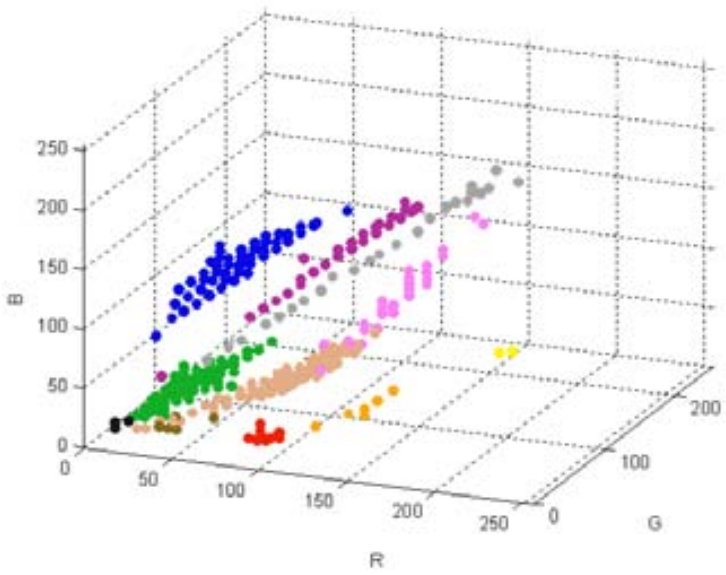


Figure 5.4: *SM* matrix obtained considering the 11 focals of the colour naming experiment, plus skin colour focals, and the canonical acquisiton conditions defined.

image which is another possible interpretation as well. In this case, the three interpretations are consistent with the real world, since there exist apples of these three colours: yellow (original image), red and green.

These three interpretations are given due to the surface matching process. In figure 5.6 we present the illuminant interpretations for another image of apples and we can observe how coloured surfaces of the image have been matched with our set of coloured surfaces to deliver these different interpretations.

Results for other images from the same database are shown in figures 5.7, 5.8, 5.9, 5.10, 5.11 and 5.12. In the figures we show the original images, followed by the illuminant estimations, the recovered images and their semantic interpretation, where a labelled image is shown according to the set of matchings that has given a high degree of correlation. Each correction proposed is a coherent interpretation with a set of colour matchings, which can be useful in a higher level process where we would like to interpret not only colours, but objects in the image.

Even though our method is not designed for calibration but to give coherent interpretations, we have also tested it with the 321 images in the Simon Fraser set, which offer calibrated information as well, with a white reference for each one, to see how well does it estimate the illuminant. Our method is not designed to estimate illuminants, but acquisition conditions. Also, we impose a canonical illuminant and

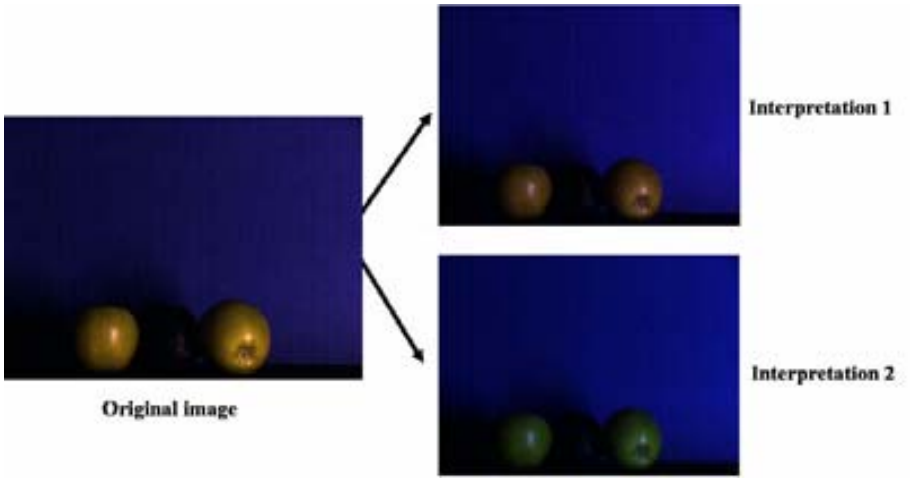


Figure 5.5: Results of an image with apples. The method derives 2 different interpretations: red apples and green apples, whereas in the original image the apples appear yellow.

a sensor which are different from the canonical illuminant and sensor from the Simon Fraser image set. However, the sensor in both acquisition conditions are white balanced, therefore we found interesting to test the performance of our method. The experiments have been performed as in [8] and we have computed the recovery angular error of the illuminant which can be added to the results in the first column in Table II in Barnard et al. paper. Our method has given a 10.39 angular error, which we consider is good enough since we do not use sensor information and our goal is not to perform colour constancy for device calibration. The best results shown in [8] are achieved by gamut mapping methods, with an angular error of 5.6, which is good performance, but they need the sensor information used in the acquisition.

5.6 Selection of significant solutions

In the experiments presented, we have looked for local maxima within the weighted feasible set in order to obtain the different semantic white point estimations. However, it might be interesting to keep more significant information of the weighted feasible set than just the local maxima. In this way, we could reduce the weighted feasible set of solutions in a useful form that could be properly used in following high level processing. To this end we propose the following hypothesis:

Hypothesis 5 *It is possible to obtain a reduction of the weighted feasible set, keeping the relevant information for interpretation.*

Considering the significant information in the computed correlation map, the maxima, we propose to compact it by extracting the ridges within it. Ridges define local

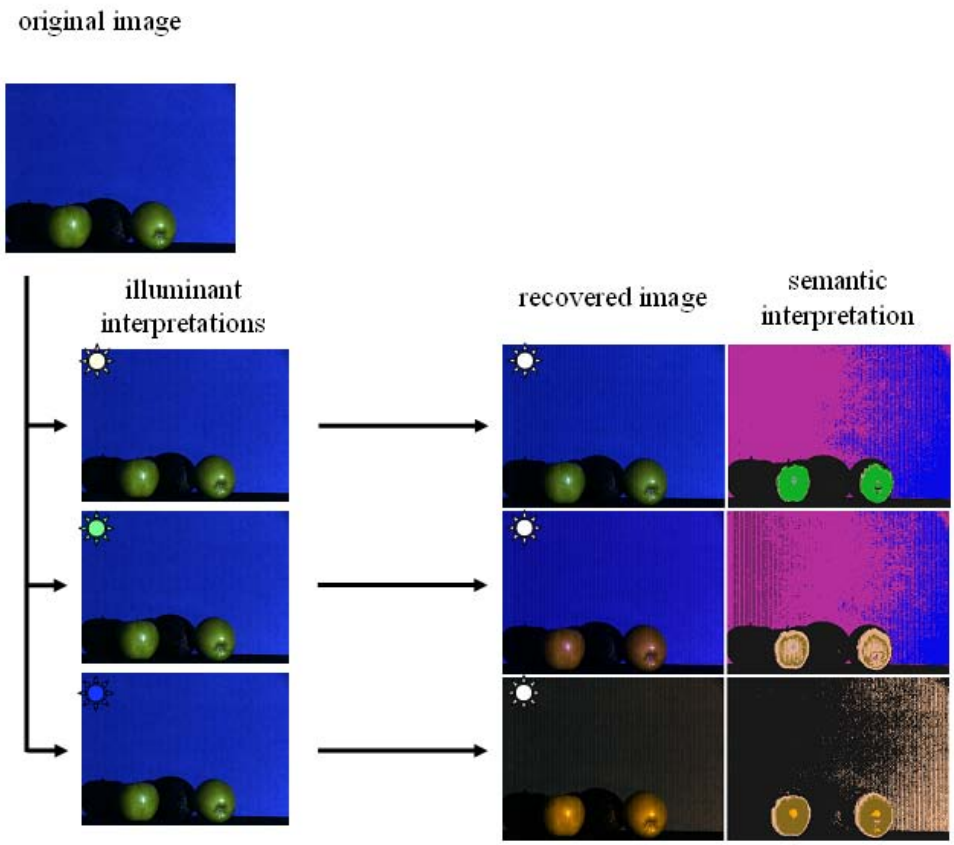


Figure 5.6: Illuminant interpretations of an image along with the semantic interpretations that have delivered them.

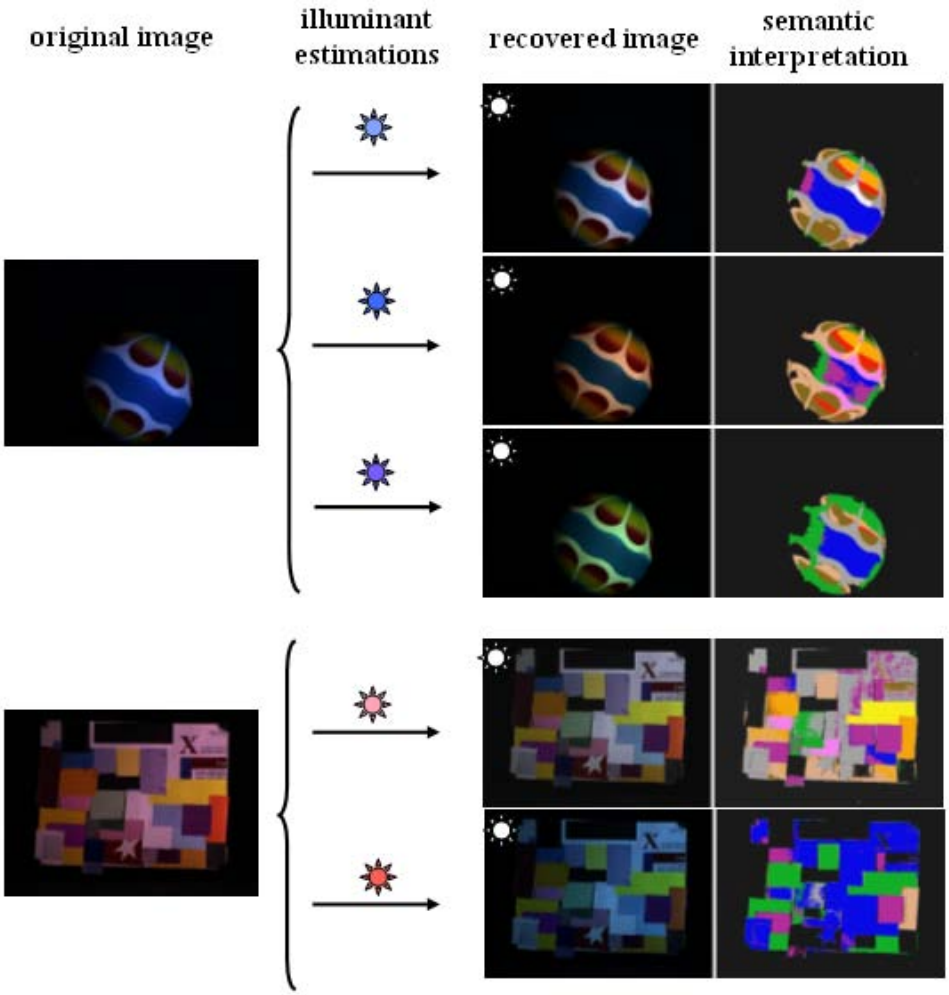


Figure 5.7: Results for images of the Simon Fraser database. For each image we show a set of possible illuminant estimations along with their semantic interpretation.

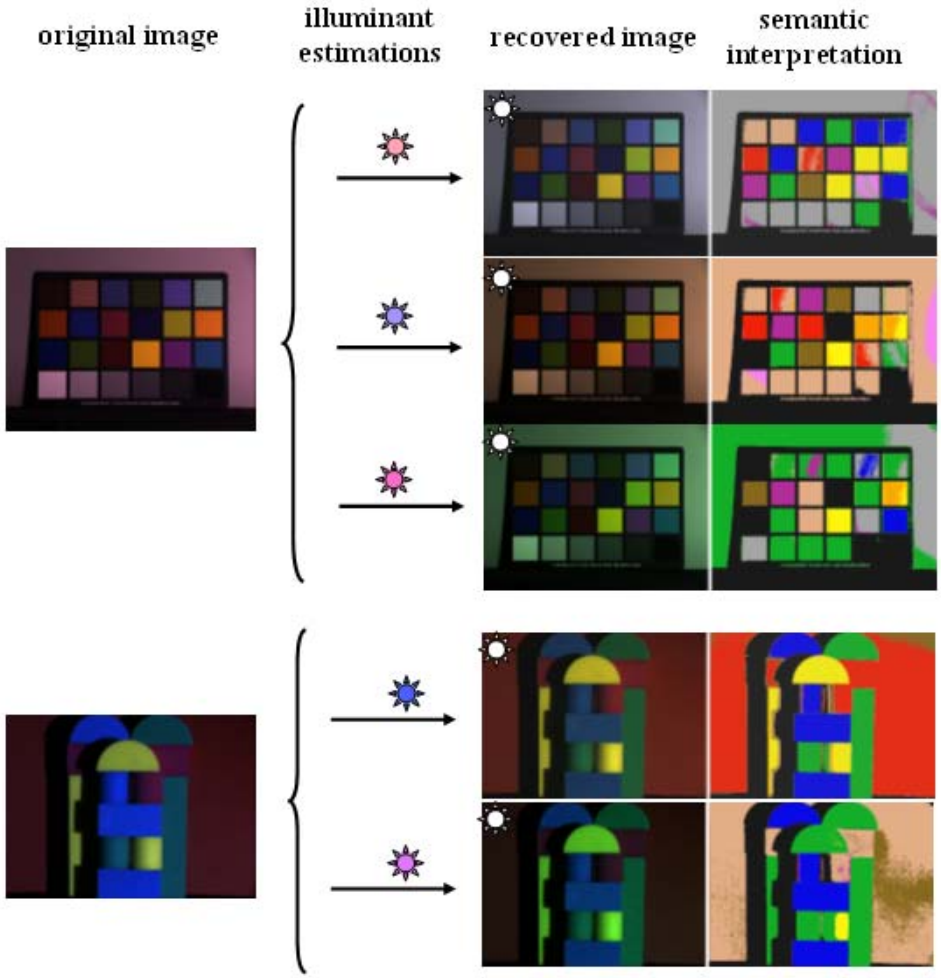


Figure 5.8: Results for images from the Simon Fraser data set (2).

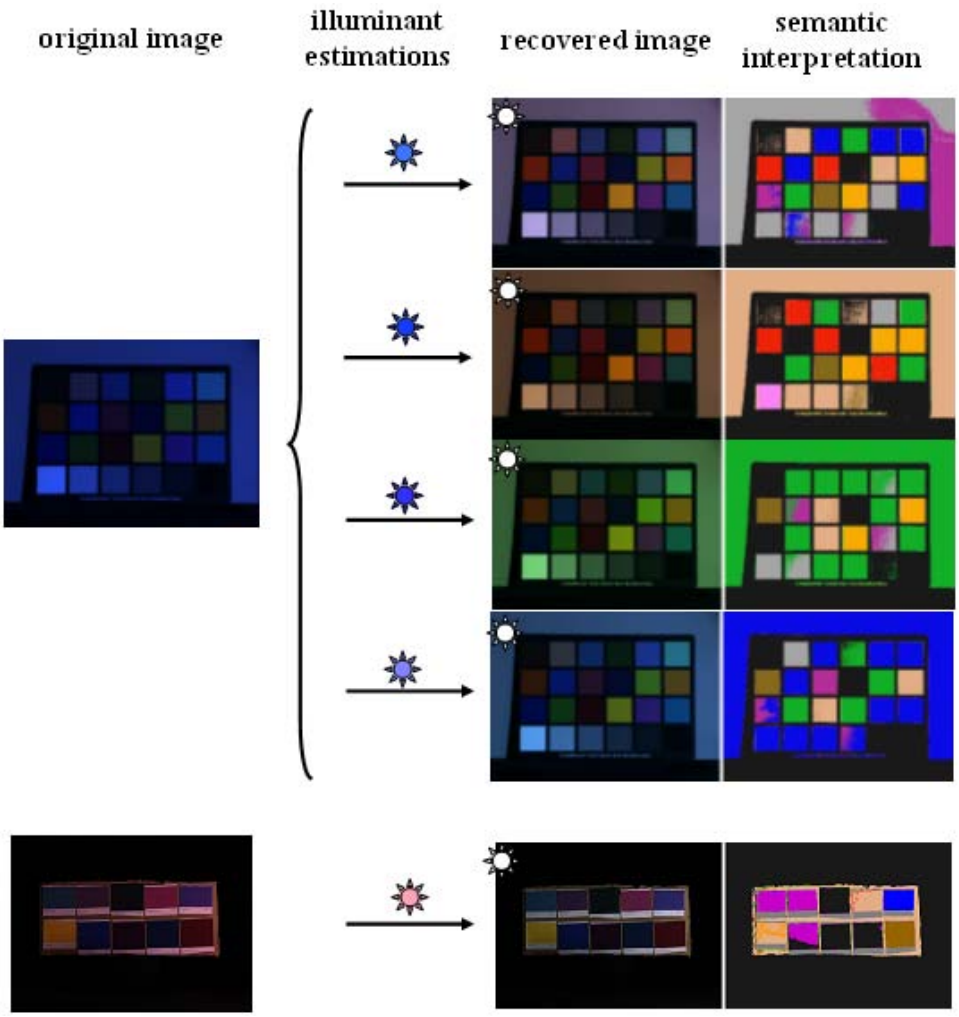


Figure 5.9: Results for images from the Simon Fraser data set (3).

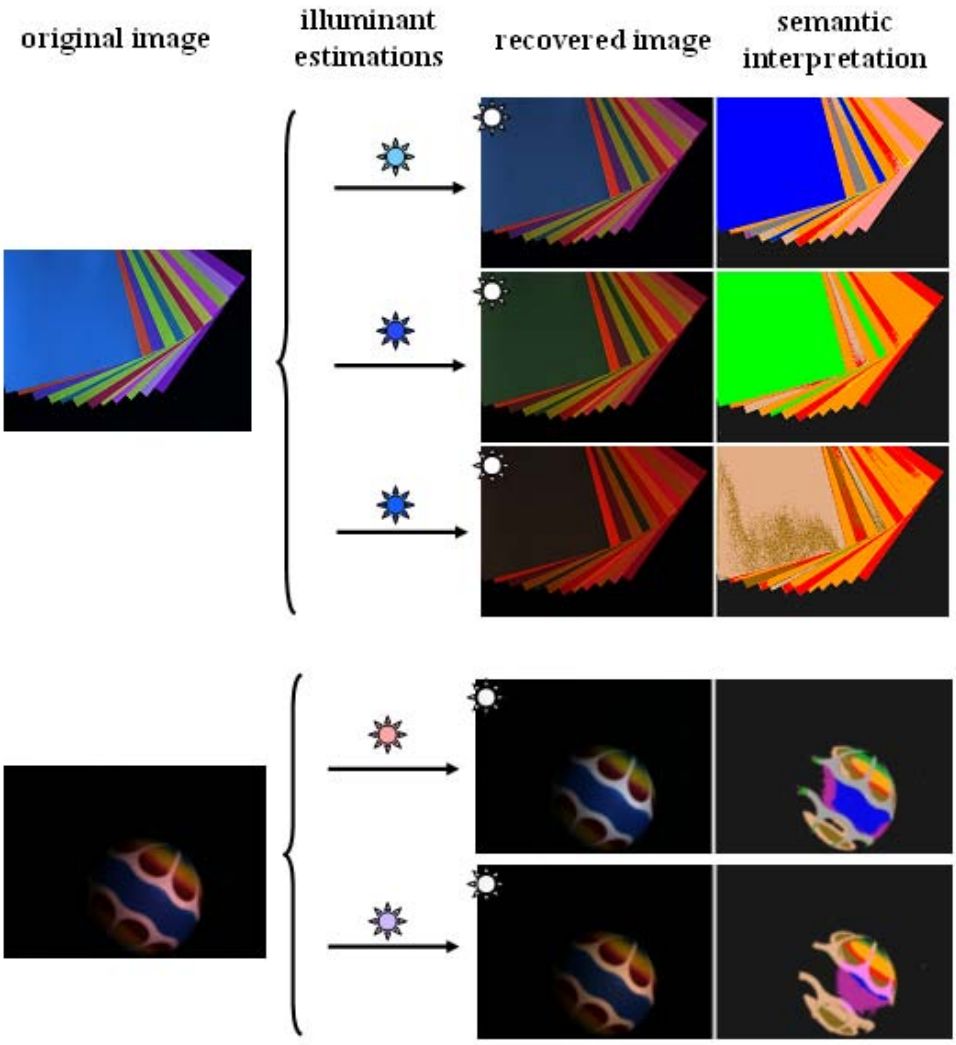


Figure 5.10: Results for images from the Simon Fraser data set (4).

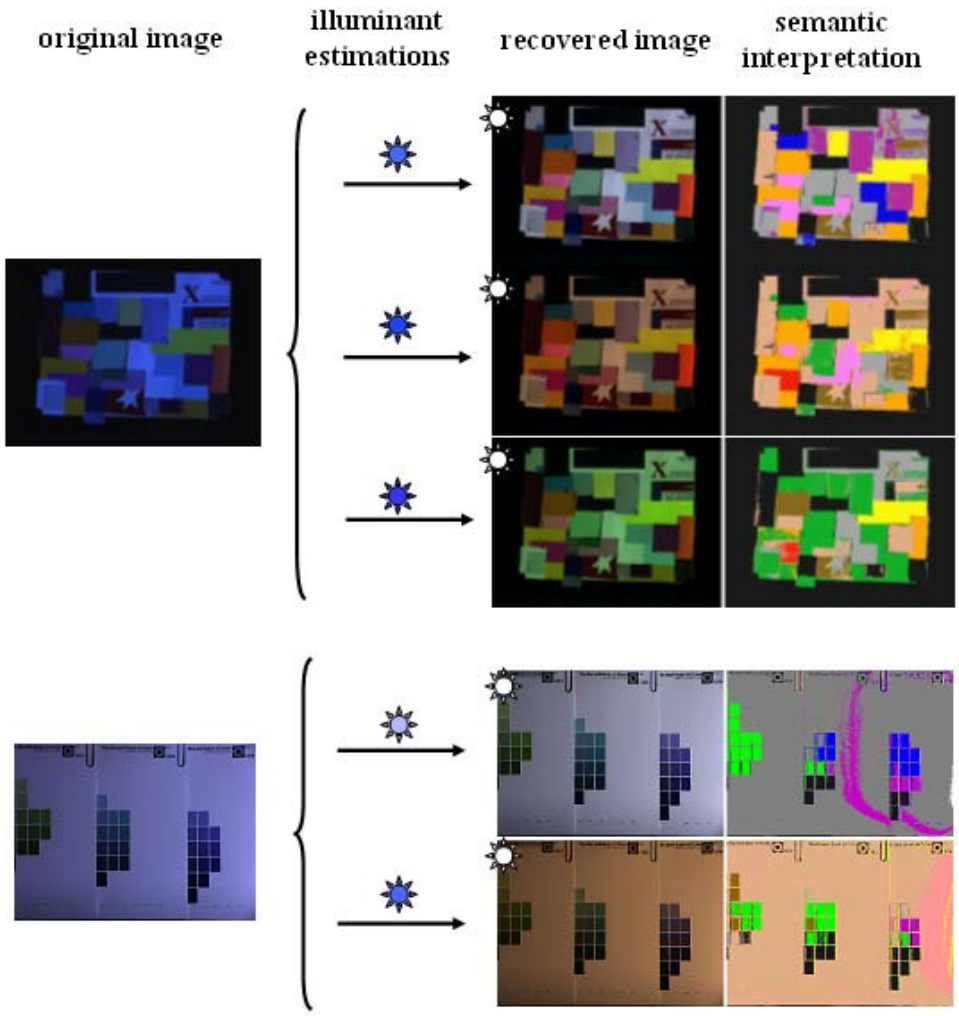


Figure 5.11: Results for images from the Simon Fraser data set (5).

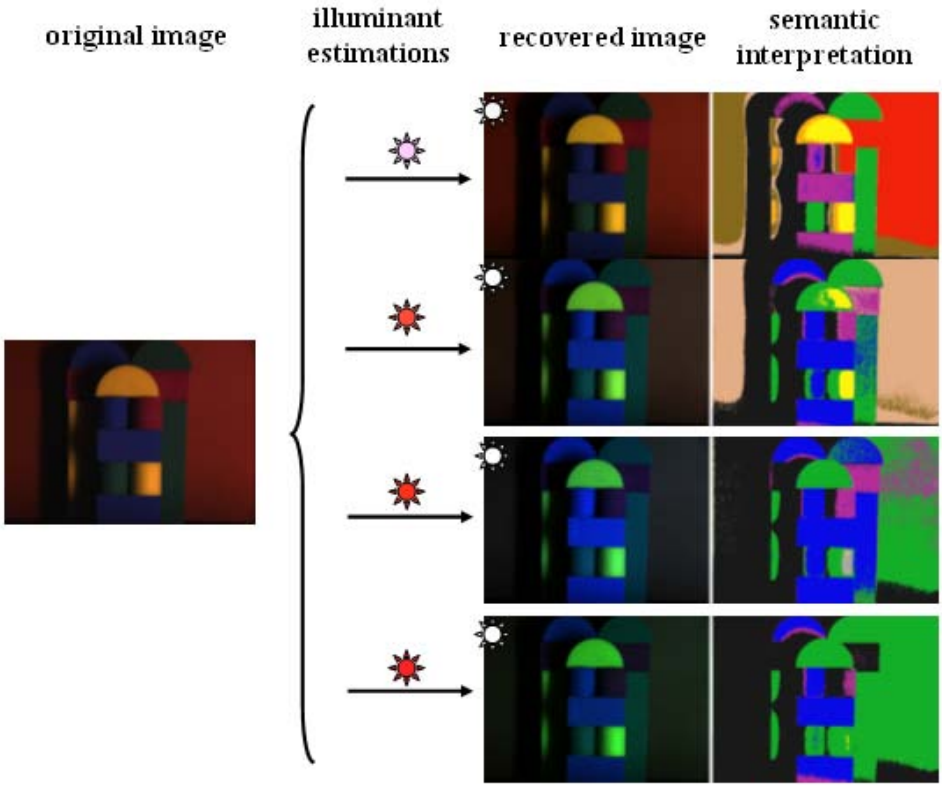


Figure 5.12: Results for images from the Simon Fraser data set (6).

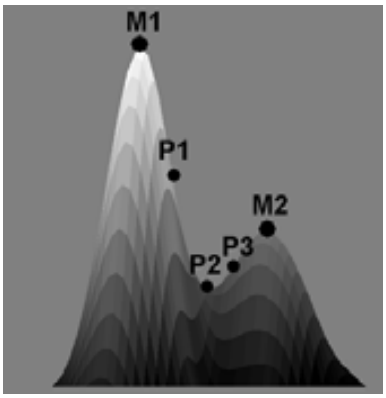


Figure 5.13: A one dimensional distribution with 2 local maxima, where other points, P_i, are useful to establish a relation between the local maxima.

maxima and paths with a unique direction of increase. Since we look for high correlation estimations we propose to extract the ridges of the weighted feasible set. In this way, feasible interpretation information can be compacted and also mathematically modelised for a proper subsequent processing. Next we present the ridge detection process performed to obtain the relevant information within a correlation map.

5.6.1 Ridge detection

In order to achieve a correct spatial reduction of a d-dimensional distribution, we must keep just the most important and representative information. The information retrieved for local maxima is useful to this end, since when a local maximum occurs, there is significant information related to its surroundings. If we keep just the local maxima, the information about relationship between different maxima is lost. The main problem of this approach is that local maxima are one-dimensional information, and it is not enough to achieve a valid description of distributions with more than two dimensions.

In figure 5.13, we show a distribution with two local maxima, M_1 and M_2, and a collection of points $P_i = P_1, P_2, P_3$. Considering just M_1 and M_2, the shape of the distribution is not enough described. Without knowing points P_i, it is not possible to say if M_1 and M_2 are isolated points or if they are connected.

The problem of finding the P_i points is directly related with the problem of ridge extraction [15, 99]. In the frame of ridge extraction, M_1 and M_2 are the peaks of the distribution, and P_i are ridge points. Hence, the ridge extraction accomplishes two requirements: first, the local maxima appear, and second, the spatial relationship between these maximums is represented by the line which joins the peaks across the path with maximum height. However, not all distributions have the shape of a regular relieve. And if we directly apply a ridge extraction algorithm, we might

keep some non-relevant information. Thus, we should find a method to measure how much significant each point is, in order to remove the non-representative ridges. Then, we will apply a ridge extraction algorithm which allows us to extract just these representative ridges.

Creaseness analysis

To measure *how much important* are the points in a distribution, we will take into account the creaseness analysis proposed in [62]. This creaseness analysis delivers a likelihood value to be part of a ridge to every point.

Let $G(x, \sigma)$ be a d-dimensional Gaussian centered at x, with normal deviation σ, and $\tilde{w}$ the dominant gradient vector in a neighbourhood with size proportional to σ, e.g., the dominant direction of this neighbourhood. The structural tensor assumes that every point has a preferred orientation. When the structural sensor is combined with the normal vector, it delivers a magnitude of how the surface orientation and the surface tension agree. Hence, if $B = x_1, ..., x_r$, are the set of points that form a discrete boundary, C, of a neighbourhood centered on x, and $N = n_1, ..., n_r$ are the set of unit normal vectors of B, then the multilocal creaseness measure $\tilde{k}_d$ at x_i in the discrete case can be defined as,

$$\tilde{k}_d = -\frac{d}{r} \sum_{k=1}^{r} \tilde{w_k}^t \cdot n_k \tag{5.12}$$

Therefore, the creaseness value is the average of the dot product between the dominant gradient vector of a neighbourhood and the normal vector of a point belonging to this neighbourhood. That is, the creaseness measures the degree of similarity between the direction of a point and the dominant direction of this point's neighbourhood.

Since the dominant gradient vector of a neighbourhood is the direction where the most meaningful change on orientation happens, the structural tensor will be a good description of the hypersurface shape, because the shape of a surface can be related to its changes. For instance, the shape of a cube can be related to its twelve edges. Thus, in order to find the most *representative* information of a distribution, we must just keep the points with a high creaseness value. These points belong to the curves that follows the gradient direction from one maximum to another through a saddle point, the so-called watershed curve [63].

The watershed algorithms are based on an immersion process. Imagine we make a hole in each minimum of the landscape and we plunge it into a lake with a constant vertical speed. The water entering through the holes will flood the surface of the landscape. During the flooding, two or more floods coming from different minima may merge. We build a dam on each of these points. At the end of the process only the different dams emerge and they constitute the watershed of the landscape. In the practice, in order to achieve good results, it is not enough to pierce the minimums. Given difficulties to find the local minima and the irregularities of the landscape, we must to put marks on the areas that will be pierced. Also, independently to the se-

lection criteria to find the potential marks, there exist some problems that arise some undesired results with this flooding process.

A common way to find the marks is to work with the negative values, the valleys, which divide the different existing mountains. However, the isolated ridges with negatives values just in one side or even just with positive values became a problem for this criterion. The watershed definition coincides with this from the ridge, hence this problem is just a ridge extraction problem. In existing literature there are a lot of methods oriented to ridge extraction, in special, on medical image papers [4, 5]. We will consider an approach based on these algorithms.

Ridges are commonly defined as a local maximum in one direction. Intuitively, a ridge can be defined as the path you follow on a mountain, where there is always a downward slope both to your left and to your right. In image analysis, that is a connected sequence of pixels having intensity values which are higher in the sequence than those neighbouring the sequence. Formally, to find a ridge we must search the points which reach a local maximum in the gradient direction. The algorithm proposed on a digital image is as follows:

Let p_1 be a point of a ridge. The height of p_1 is bigger than the other $n_d - 2$ neighbours, where n_d are the number of neighbours on a d-dimensional space. For instance, for $d = 2$, $n_d = 8$, and for $d = 3$, $n_d = 26$. Because a ridge is a line, p_1 has two neighbours, p_2 and p_3, which belong to the ridge. Then, with neighbours p_2 and p_3, the next cases can occur:

a) p_1 is a local maximum.

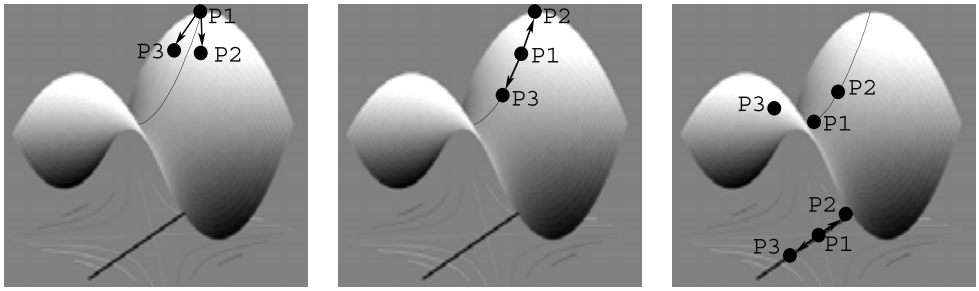
b) p_1 is not a local maximum. Then, either p_2 or p_3 can be strictly higher, than p_1.

c) p_1 is not a local maximum and p_2 and p_3 are higher (p_1 is a singularity).

A graphical representation of each case is shown in figure 5.14.

Since the third case is just a singularity, the first step is to find the points of the first type. If we take the problem as *how many higher neighbours can have a ridge point*, the problem is reduced towards an easy point classification as follows: points with not more than one higher neighbour will be a ridge point. As a result, each point will be labelled with the number of higher neighbours. Then, we keep the points labelled with a value lower than two, giving point in a) and b). Finally, each different sequence of connected points is labelled as a different ridge in order to distinguish from one ridge to another.

But as we have introduced before, there exist singularities that this classification method does not take into account. The problem arises when two or more ridges converge. In this case, the convergence point, which effectively is a ridge point, has more than one higher neighbour. Thus, we must define some criteria in order to include these singularities on the final result. This criteria has to distinguish between



a) b) c)

Figure 5.14: Different cases of a ridge point: (a) local maximum, (b) p_1 is a ridge point but not a peak, then, just p_2 can be higher, and (c) p_1 is a local minimum of the ridge because both p_2 and p_3 are higher.

a convergence point and a point that is just a neighbour of a ridge. For a convergence point different paths are from different ridges. Hence, it is easy to know that, because each ridge has a different label.

5.6.2 Results

We have used the ridge extraction procedure presented, to select significant solutions within the weighted feasible set of solutions obtained with our semantic correlation method. In figure 5.15, we show the ridges obtained for an image from the Simon Fraser database. In figure 5.16 we present the profile of the ridges versus the correlation value. Above each ridge profile we show two numbers: first a ratio between the portion of correlation included in the ridge and the compression of the correlation map (ratio of ridge solutions versus feasible solutions), and second the portion of correlation included in the ridge. Then, in figure 5.17 we show the corrected images for the maximum correlation in each ridge, along with a label of the illuminant colour of the scene given by our method. The label of the illuminant colour has been obtained through the computing of the inverse of the diagonal map, which represents the estimated illuminant colour, and the use of a voronoi diagram of the focals in SM. The global compression ratio of the correlation map for this image is 0.0039, and the global ratio of correlation included in the ridges is 0.0881.

In figure 5.18, we show the ridges obtained for another image from the Simon Fraser database. Again, in figure 5.19 we present the profile of the ridges versus the correlation value. And in figure 5.20 we show the corrected images for the maximum correlation in each ridge, along with a label of the illuminant colour of the scene given by our method. The global compression ratio of the correlation map for this image is 0.0026, and the global ratio of correlation included in the ridges is 0.0873.

Finally, in figure 5.21, we show the ridges obtained for a third image from the Simon Fraser database. Again, in figure 5.22 we present the profile of the ridges ver-

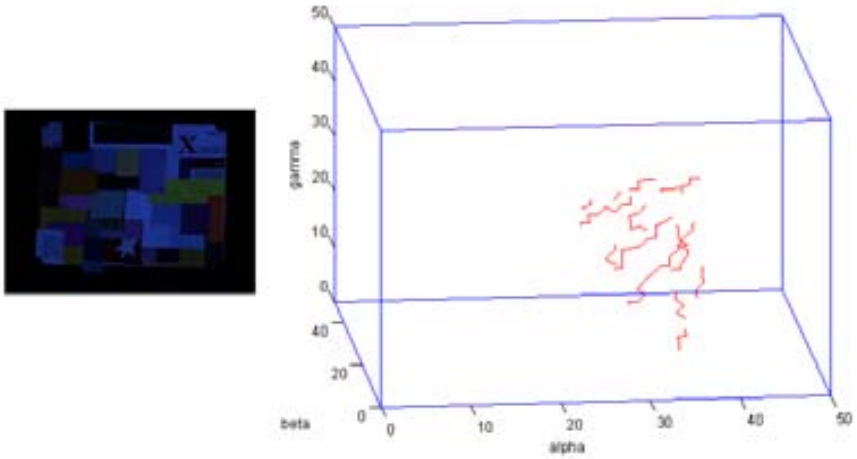


Figure 5.15: Ridges extraction from the weighted feasible set for a Simon Fraser image.

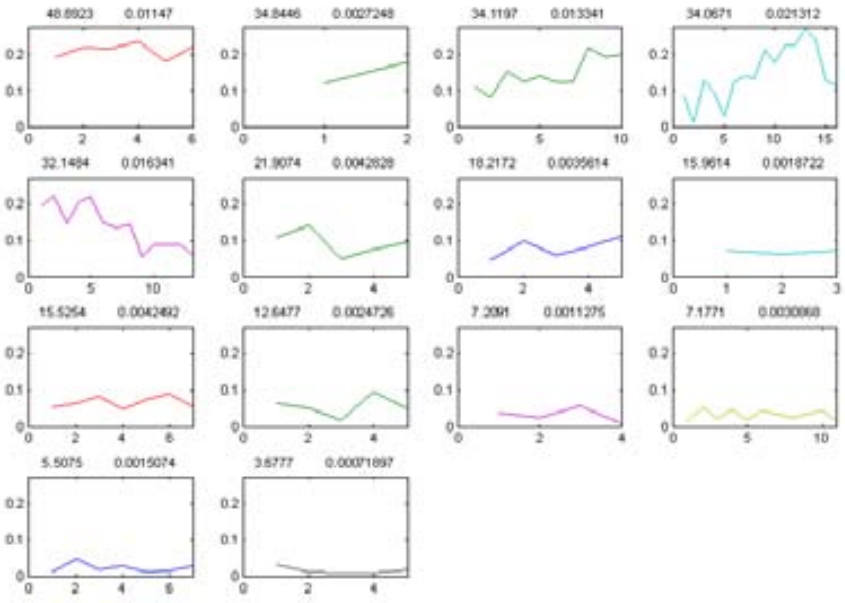


Figure 5.16: Profile of the ridges extracted from the weighted feasible set for a Simon Fraser image.

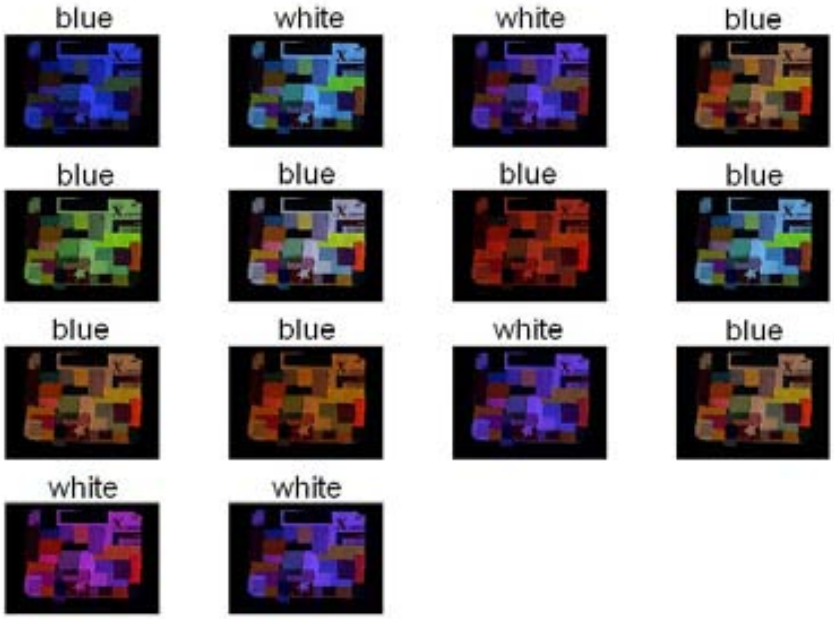


Figure 5.17: Corrected images for the maximum correlation of each ridge and a label of the illuminant colour of the scene.

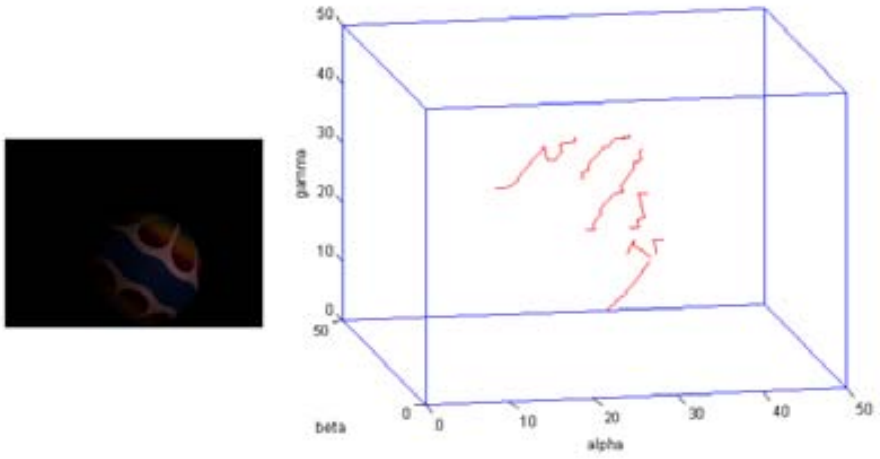


Figure 5.18: Ridges extraction from the weighted feasible set for a second Simon Fraser image.

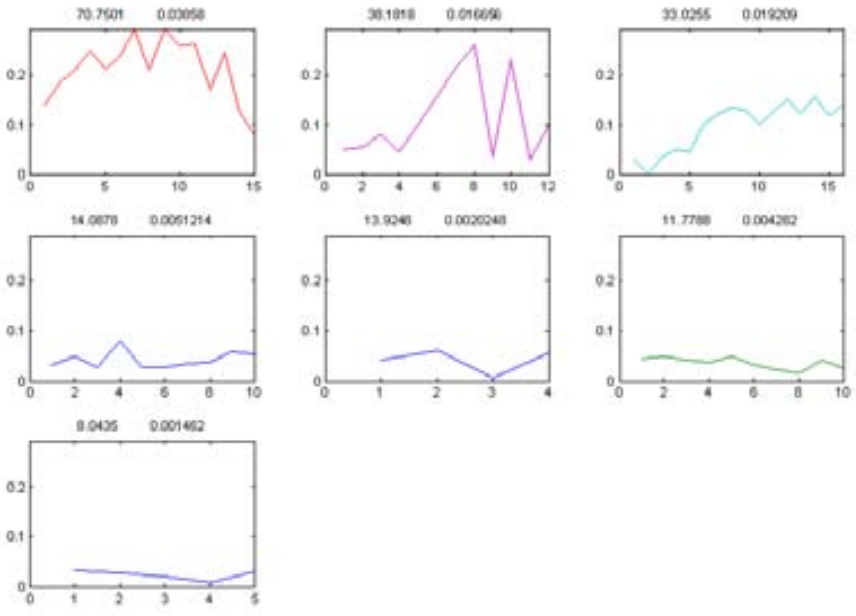


Figure 5.19: Profile of the ridges extracted from the weighted feasible set for a second Simon Fraser image.

sus the correlation value. And in figure 5.23 we show the corrected images for the maximum correlation in each ridge, along with a label of the illuminant colour of the scene given by our method. The global compression ratio of the correlation map for this image is 0.0017, and the global ratio of correlation included in the ridges is 0.0970.

The mean of the compression ratio of the correlation map for the 321 Simon Fraser images is 0.0025 and the mean of the ratio of correlation included for these images is 0.0897. This confirms that through ridge extraction we significantly reduce the amount of information considered in the correlation map, keeping the important information, since ridges by definition include the different local maxima. Furthermore, ridges can be mathematically modelised, which could lead us to obtain the function which represent the interpretations of an image. This modelisation could be interesting for a subsequently interpretation processing, such as the introduction of illuminant constraints, etc.

5.7 Discussion

The proposed method and the following selection of the most significant solutions, enables us to reduce the number of interpretations.

In this chapter we have not proposed a new method for colour constancy, but

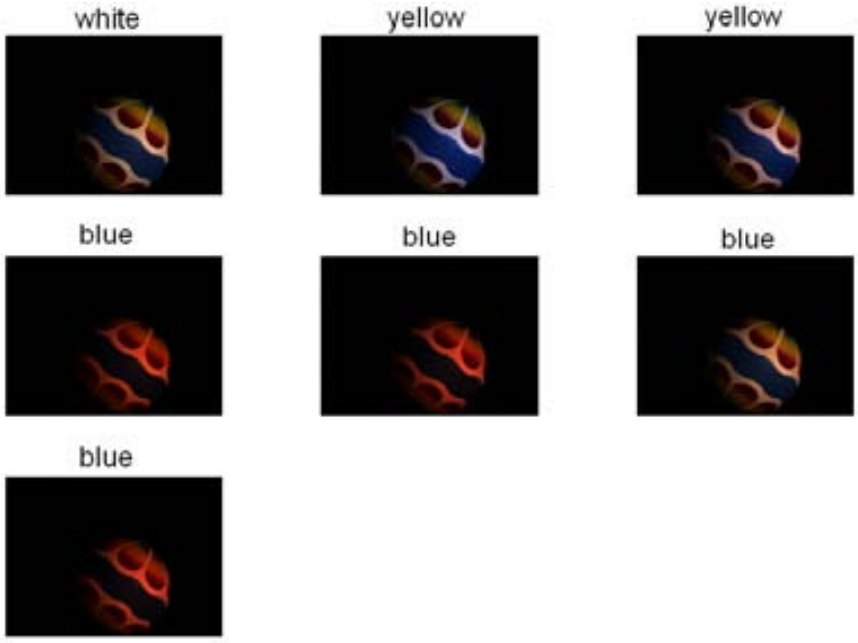


Figure 5.20: Corrected images for the maximum correlation of each ridge and a label of the illuminant colour of the scene.

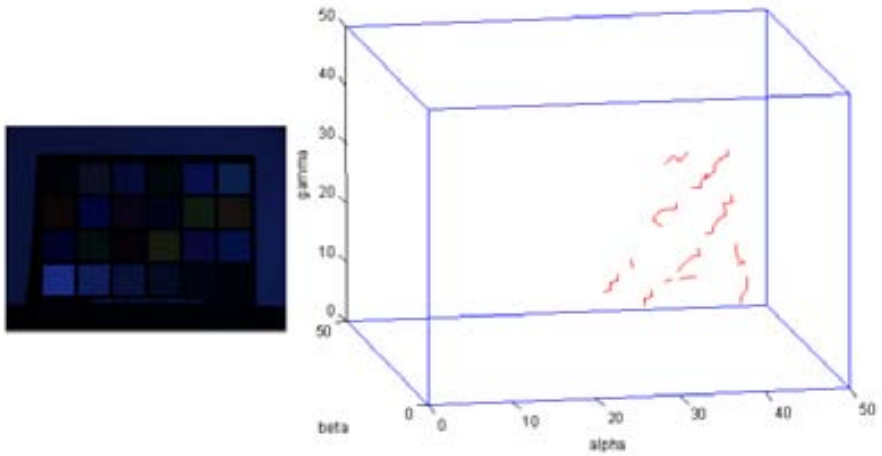


Figure 5.21: Ridges extraction from the weighted feasible set for a third Simon Fraser image.

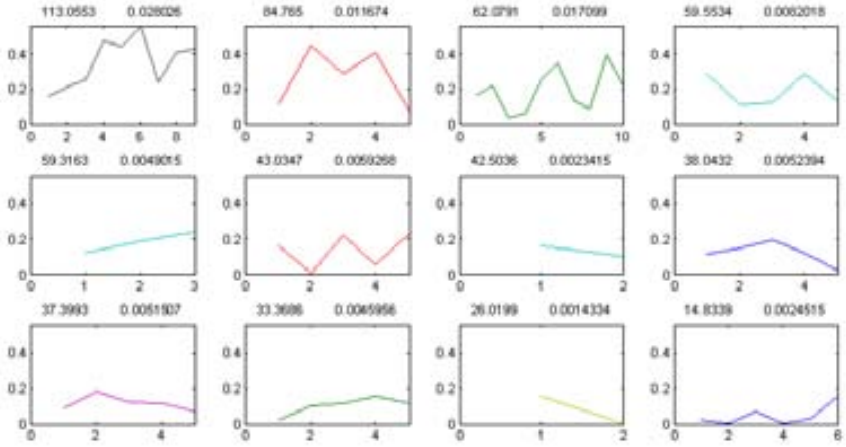


Figure 5.22: Profile of the ridges extracted from the weighted feasible set for a third Simon Fraser image.

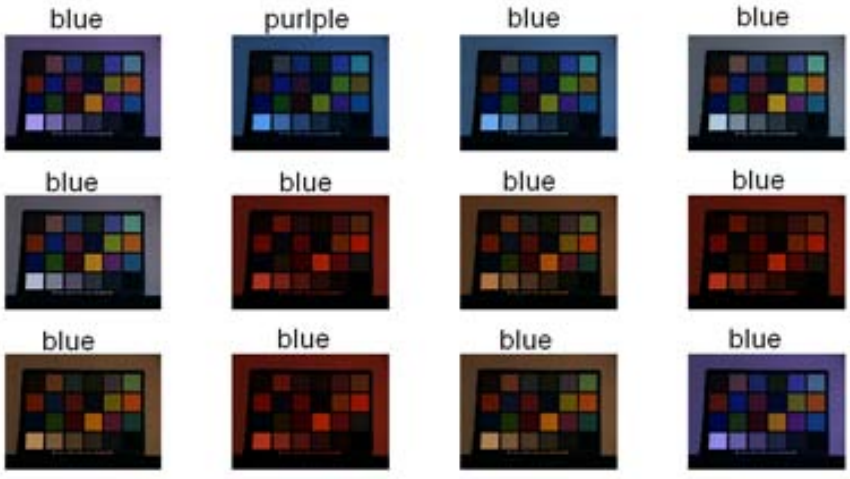


Figure 5.23: Corrected images for the maximum correlation of each ridge and a label of the illuminant colour of the scene.

a method for image interpretation under uncalibrated conditions. Therefore, we do not aim to perform perfect colour constancy in the classical point of view, but use the frame of the existing colour constancy methods to deal with the problem in the uncalibrated case. Uncalibrated colour constancy is a less constrained problem than the calibrated case, and therefore more feasible solutions exist. Effective colour constancy algorithms are designed for calibrated conditions, where the sensor information is known. Also, most of them only consider a single solution, which in the uncalibrated case might be too restrictive, and to consider different feasible interpretations can be useful in an image understanding frame.

To sum up our proposal, what we do is to adapt the existing calibrated methods to the uncalibrated case, by changing the usual assumptions used. It has been done in this way:

- We have proposed to replace the classical estimation of the illuminant by the estimation of the acquisition conditions, $A_k(\lambda)$, which comprise both illuminant and sensor.
- We have introduced a visual cue related to colour constancy: colour matching. Through colour matching, we introduce information of the expected coloured surfaces in the images, and these will be the surfaces to match with in the colour matching process.
- We have introduced restrictions to the problem to constrain it, and, subsequently, a method to estimate different feasible interpretations of the white point of an image has been proposed.

In the frame of image understanding, further knowledge about the scene content might be introduced in a higher level process to select between the set of possible interpretations given (e.g., if we expect an apple in an image we would only choose those interpretations in which the apples are red, green or yellow, and discard any interpretation where the apples are blue or some other unrealistic colour). The correlation map obtained with the method could be reduced for interpretation by computing its medial axis, since it would represent the different feasible solutions in a compacted way.

Also, the proposed method might be extended to introduce more restrictions to the problem, just by adding a restriction matrix that could be easily combined with the proposed SM' matrix. For example, if we would like to restrict the set of feasible illuminants, we could build an illuminant probability matrix, P^I, where the probability of finding different illuminants might be mapped to the illuminant change they produce. This probability matrix P^I can easily be introduced to the method through eq. 5.13.

$$CM^I(I) = CM(I) \cdot P^I \tag{5.13}$$

Other restrictions might be introduced in the same way, to further constrain the problem and, in the end, find a single solution when possible. In this frame of uncalibrated conditions, the introduction of knowledge of what we expect to find in images is a way to find a unique feasible solution.

Chapter 6

Summary and conclusions

In this work we have studied how to represent colour information for an image annotation project. The RGB values acquired by a camera need to be processed if we aim to use them as descriptors of the surfaces in a scene, since colours in an image tend to be biased by the illuminant of the scene. Image annotation imply to assign colour names to image regions, therefore we firstly have worked on skin detection and afterwards for more general colour naming. Finally, we have proposed a new line of research for more general semantic interpretation. There exist two main approaches to deal with the colour variability problems, these are:

Colour invariant methods are interesting, since they do not need to know any information of the acquisition conditions and can be used in any type of images. They aim to obtain image descriptors invariant to some features of the scene, such as intensity of the illuminant, colour of the illuminant, highlights, etc. We have performed a skin colour modelisation and segmentation experiment for images under varying illumination, considering intensity and colour changes, using some of these normalisations and compared their performance. We have proved that when we deal with just illuminant intensity changes, chromaticity coordinates deliver best performance, but under illuminant colour changes it is better to use comprehensive normalisations. However, invariant normalisations are not good if we have to perform a general colour naming task, since they normally remove intensity information, which is important to distiguish some colours.

Colour constancy methods aim to recover the illumination of images and to obtain a balanced representation of the image, where colours can be taken as physical descriptors of the surfaces in the scene. We have proposed a colour constancy method that works for calibrated conditions based on colour matching. However, since in computer vision we do not always know the sensitivity curves of the acquisition device, we have proposed a new frame to work under uncalibrated conditions. In this way, we have proposed a method to deal with images of unknown origin based on existing colour constancy methods for calibrated conditions, but where no previous knowledge of the acquisition conditions is required. In this new approach, our goal is

not to achieve perfect colour constancy but to find meaningful white point estimations which deliver different interpretations of the colours in the images. In this way, we do not look for a unique solution but consider any solution which gives some meaning to the coloured surfaces in the image, since in the image annotation problem under uncalibrated conditions a single solution might be too restrictive. These interpretations are guided by some previous knowledge, that we have introduced as *expected* colours in the images. Therefore, each white point estimation will be given by a concrete interpretation of the colours in the image.

6.1 Contributions

The essential contributions of this work are:

- A reformulation of white point estimation for uncalibrated conditions which looks for **multiple and most interpretable solutions**.
- The introduction of the **colour matching** visual cue to guide the estimation of the white point.
- Introduction of high-level information in the colour matching process, that is, **semantic information** regarding expected colours in the images.
- A **computational algorithm to select the white point estimation** for uncalibrated conditions which delivers multiple interpretable solutions through a colour matching process which introduces semantic high-level information.
- The **relaxed grey-world assumption** which bounds the location of optimal solutions under calibrated conditions.

The proposed method is based on existing colour constancy approaches and it delivers feasible semantic white point estimations according to different interpretations of the colours in the images. The colour matching process has been proposed to constrain the feasible set of solutions through the introduction of high-level information. Thus, we compute a weighted feasible set and within it we can choose solutions which deliver a high-level interpretation of the coloured surfaces in the scene. The weight in this feasible set represents a degree of colour matching for the corresponding solution. With this approach we do not aim to achieve perfect colour constancy, since for uncalibrated conditions is not feasible, but to give different feasible interpretations of the acquisition conditions of an image, according to the colours that are present in the image. Therefore, we deliver a recovered image and its colour interpretation, which might be useful for a higher level image understanding processing, such as object recognition.

Other contributions of this work are:

- Demonstration that it is possible to **select a reduced number** of interpretable solutions, through ridges extraction.
- A **computational colour constancy algorithm** based on the relaxed grey-world assumption, which performs a colour matching process to deliver a sub-sampling of the feasible set of solutions.
- Experiments to *test existing evaluation procedures* of colour constancy methods.
- Use of **comprehensive normalisations for skin colour detection**.

The method proposed for colour constancy based on the relaxed grey-world assumption delivers a reduced feasible set of solutions, which has been proven to deliver, on average, better performance than the feasible set obtained by Forsyth's CRULE, and similar results have been obtained when using existing heuristics with some experiments with synthetic data. Also, existing evaluation methods might not be useful for colour constancy for uncalibrated conditions.

6.2 Future directions

In this work we have left open different lines of research, that we find interesting to go further:

- Building other semantic matrices (SM) in the semantic white point estimation method, regarding other semantic colour sets, such as natural colours or hair colours, and an application to test it.
- Introduction of other restrictions, appart from colour matching, to constrain the problem and improve the results. For instance, we could introduce a matrix of illuminant restrictions, where we could add probability information to find different feasible illuminants.
- Use of the semantic white point interpretations in a high-level image understanding system, which could introduce further assumptions to decide from the different colour interpretations proposed. This system could introduce more high-level information, i.e. more knowledge about the real world, which could lead us to select a single meaningful solution.
- Comprehensive interpretation of ridges in order to deliver semantic illuminant descriptors, such as *sunny*, *cloudy*, *indoor*, etc. The first and easiest experiment would be to calculate the intersection between the obtained ridges with the Planckian locus.
- An application under calibrated conditions to use the semantic white point estimation method, where other semantic colour sets can be introduced.

The essential contribution of this work has been presented and justified but not completely proved. A proper evaluation would require specific and complex psychophsyical data, we are already working on:

- Definition of a set of psychophysical experiments to determine in which specific conditions the nameability assumption holds.
- Construction of a database of natural images with a set of likely white point estimations which can be obtained by the consensus from a wide group of observers.

Appendix A

Fisonomies project

We have been working on a surveillance system which needs the ability of automatic people detection in images with a complex background and varying illumination. This system acquires images from people entering a building to subsequently extract textual descriptors based on people appearance. This information is stored in a database for later queries of people inside the building.

A.1 People detection for appearance description

The people description module, where we have considered colour invariant normalisations for skin colour detection, is a part of a general surveillance system. Images of people entering a building are processed while they are checking-in (figure A.1). Textual descriptors based on people appearance are extracted from these images and this information is saved in a global database where the security staff of the building can perform some queries.

In this way, if they see in a camera inside the building someone who is causing problems, and can perform these queries to obtain information that identifies this person. Here is where the description module acquires importance, because in our database we have information about the appearance of the people who have entered the building and that has been extracted. With this purpose, the system allows the user to make queries formulated in terms of textual descriptors, to retrieve those images from the database agreeing with the descriptors of the query. Queries are formulated in terms of colour, texture and structural properties of clothes that people are wearing. The system aims to automatically build an appearance feature vector from an image acquired while people is checking-in in front of the desk. In figure A.2 we show real images of the entrance of the building we have been working in.

Retrieving images from large databases using image content as a key is a largely problem studied in computer vision. Two major approaches can be stated. First, similarity retrieval consists in looking for images in a database using a reference image as query. The second approach concerns browsing applications and consists in retrieving

Figure A.1: Scheme of the registration process which involes our system.

images by pictorial content, i.e. using symbolic descriptors as indices. Regarding the features used as the basis to formulate queries, usually early visual primitives such as colour and texture are used. Sometimes, structure of objects in the image is important. A number of works combine low level visual cues,such as color and texture, with higher level information such as structure (e.g. [41, 88, 61]). Ous system follows this approach. Queries are formulated in terms of textual descriptors like 'we are looking for a man in a red shirt' that are compared with descriptions stored in the database that were previously extracted from the input images. In figures A.3, A.4 and A.5 we show different steps of our description/retrieval system: constructing the apppearance feature vector in the registration (A.3), formulating a query on the people that have been registered entering the building (A.4) and the results of the query (A.5).

Our approach to the people description module focus on a computational extraction of clothes features that is based on a four-step process. First, the person must be located in the center of the image. Our skin segmentation module will detect skin colour regions in the image. In this way, people will be detected and the regions that must be analised will be given. Then, a colour feature vector of the pixels in the regions of interest is computed. After a first region initialisation and considering colour properties of regions of interest plus edges information, a merging process is proposed, which will join any neighbouring image regions with similar colour properties. Finally, a high level interpretation of these image regions will allow to model a structural description on the clothes that people are wearing. Some examples of content-based queries can be seen in [92]. They help to illustrate how the descriptor

Figure A.2: Sample images from the entrance of the building we have to deal with.

proposed in the people description module can behave on the system presented. Also, a more detailed explanation of how the high-level descriptor operates can be found in [2].

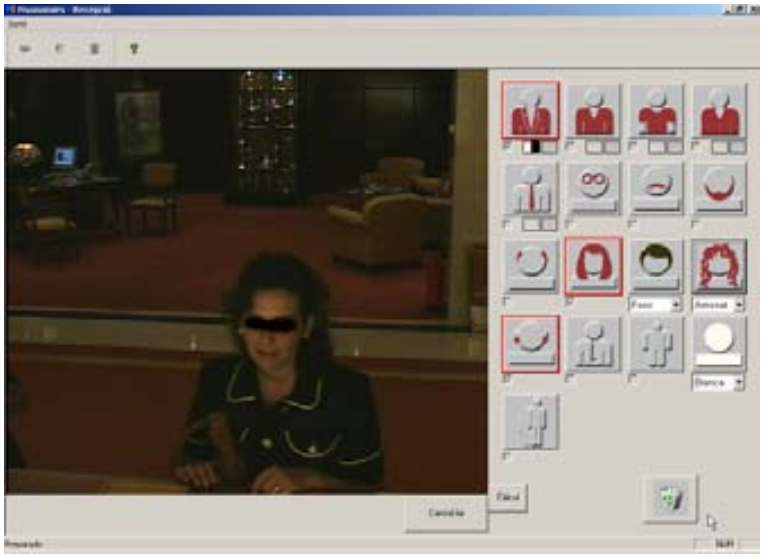


Figure A.3: The different steps in the description/retrieval system: constructing the appearance feature vector in the registration.

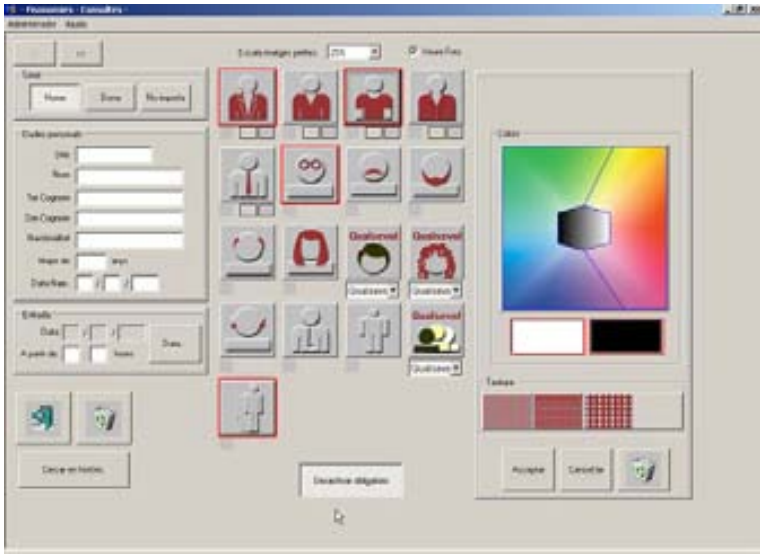


Figure A.4: The different steps in the description/retrieval system: formulating a query on the people who have been registered entering the building.

Figure A.5: The different steps in the description/retrieval system: the results of the query.

Bibliography

[1] Y. Aloimonos and A. Rosenfeld. Computer vision. *Science*, 253:1249–1254, September 1991.

[2] Agnés Borràs Angosto. High-level clothes description based on color-texture features. Technical report, Computer Vision Center, UAB, September 2002.

[3] L. Arend and A. Reeves. Simultaneous color constancy. *Journal of the Optical Society of America A*, 3(10):1743–1751, 1986.

[4] S. Aylward and E. Bullitt. Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction, 2002.

[5] Stephen R. Aylward, Julien Jomier, Sue Weeks, and Elizabeth Bullitt. Registration and analysis of vascular images. *Int. J. Comput. Vision*, 55(2-3):123–138, 2003.

[6] Kobus Barnard, Vlad Cardei, and Brian Funt. A comparison of computational colour constancy algorithms. part one. methodology and experiments with synthesized data. *IEEE Transactions on Image Processing*, 11(9):972–984, 2002.

[7] Kobus Barnard, Lindsay Martin, Adam Coath, and Brian Funt. A comparison of computational colour constancy algorithms. part two. experiments on image data. *IEEE Transactions on Image Processing*, 11(9):985–996, 2002.

[8] Kobus Barnard, Lindsay Martin, Brian Funt, and Adam Coath. A data set for color research. *Color Research and Application*, 27(3):147–151, 2002.

[9] Kobus Barnard, Lindsay Martin, and Brian V. Funt. Colour by correlation in a three-dimensional colour space. In *ECCV '00: Proceedings of the 6th European Conference on Computer Vision-Part I*, pages 375–389, London, UK, 2000. Springer-Verlag.

[10] R. Benavente, F. Tous, R. Baldrich, and M. Vanrell. Statical modelling of a colour naming space. In *Proc. of the First European Conference on Color in Graphics, Image and Vision*, pages 406–411, 2002.

[11] Robert Benavente, Maria Vanrell, and Ramon Baldrich. Estimation of fuzzy sets for computational colour categorization. *COLOR Research and Application*, 29:342–353, 2004.

[12] Robert Benavente, Maria Vanrell, and Ramon Baldrich. A data set for fuzzy colour naming. *COLOR Research and Application*, (In Press), 2005.

[13] J. Berens and G.D. Finlayson. Log-opponent chromaticity coding of colour space. In *ICPR00*, pages Vol I: 206–211, 2000.

[14] Daniel Berwick and Sang Wook Lee. A chromaticity space for specularity-, illumination color- and illumination pose-invariant 3-d object recognition. In *ICCV*, pages 165–170, 1998.

[15] Arijit Bishnu, Partha Bhowmick, Sabyasachi Dey, Bhargab B. Bhattacharya, Malay K. Kundu, C. A. Murthy, and Tinku Acharya. Combinatorial classification of pixels for ridge extraction in a gray-scale fingerprint image. In *ICVGIP*, 2002.

[16] A. Borràs, F. Tous, J. Lladós, and M. Vanrell. High-level clothes description based on colour-texture and structural features. *Lecture Notes in Computer Science*, 2652:108 – 116, 2003.

[17] David H. Brainard, Wendy A. Brunt, and Jon M. Speigle. Color constancy in the nearly natural image. i. asymmetric matches. *Journal of the Optical Society of America A*, 14(9):2091–2110, 1997.

[18] D.H. Brainard and W.T. Freeman. Bayesian colour constancy. *J. Opt. Soc. Am.-A*, 14:1393–1411, 1997.

[19] G. Buchsbaum. A spatial processor model for object colour perception. *J. Franklin Inst.*, 310:1–26, 1980.

[20] V. Cardei. *A Neural Network Approach to Color Constancy*. PhD thesis, Simon Fraser University, 2000.

[21] K.M. Cho, J.H. Jang, and K.S. Hong. Adaptive skin-color filter. *Pattern Recognition*, 34(5):1067–1073, May 2001.

[22] CIE. *Commission Internationale de l'Éclairage Proceedings*. Cambridge University Press, 1924.

[23] J. Cohen. Dependency of the spectral curves of the munsell color chips. *Psychonomic Science*, 1:369–370, 1964.

[24] J. W. Cooley and J. W. Tukey. An algorithm for the machine computation of the complex fourier series. *Mathematics of Computation*, 19:297–301, 1965.

[25] Michel Coster and Jean-Louis Chermant. *Précis d'analyse d'images*. Editions du Centre National de la Recherche Scientifique, Paris, 1985.

[26] Jonathan Dowdall, Ioannis Pavlidis, and George Bebis. Face detection in the near-ir spectrum. *Image Vision Comput.*, 21(7):565–578, 2003.

[27] M. d'Zmura and G. Iverson. Color constancy. i. basic theory of two stage linear recovery of spectral descriptions for lights and surfaces. *JOSA-A*, 10:2148–2165, 1993.

[28] M. d'Zmura and G. Iverson. Color constancy. ii. results for two-stage linear recovery of spectral descriptions for lights and surfaces. *JOSA-A*, 10:2166, 1993.

[29] M. d'Zmura and G. Iverson. Color constancy. iii. general linear recovery of spectral descriptions for lights and surfaces. *JOSA-A*, 11:2389–2400, 1994.

[30] G. D. Finlayson, M. S. Drew, and B. V. Funt. Spectral sharpening: Sensor transformations for improved color constancy. *JOSA-A*, 11:1553–1563, 1994.

[31] G.D. Finlayson. Color in perspective. *Transactions on Pattern Analysis and Machine Intelligence*, 18:1034–1038, 1996.

[32] G.D. Finlayson, P.H. Hubel, and S. Hordley. Color by correlation. In *IST/SID Fifth Color Imaging Va.*, pages 6–11, 1997.

[33] G.D. Finlayson, B. Schiele, and J.L. Crowley. Comprehensive colour image normalization. In *ECCV-A 98*, pages 475–490, 1998.

[34] G.D. Finlayson and G.Y. Tian. Perfect colour constancy vs normalisation for object recognition. *Journal on Pattern Recognition and Artificial Intelligence*, 1999.

[35] G.D. Finlayson and R. Xu. Non-iterative comprehensive normalization. In *CGIV 2002*, pages 159–163, 2002.

[36] G.D. Finlayson and Ruixia Xu. Illuminant and gamma comprehensive normalisation in logrgb space. *Pattern Recognition Letters*, 2003.

[37] Graham Finlayson and Steven Hordley. Selection for gamut mapping colour constancy. *Image and Vision Computing*, 17(8):597–604, 1999.

[38] Graham D. Finlayson, Steve Hordley, and Paul M. Hubel. Colour by correlation: A simple, unifying framework for colour constancy. *Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1209–1221, November 2001.

[39] Graham D. Finlayson, Steve Hordley, and Paul M. Hubel. Illuminant estimation for object recognition. *COLOR research and application*, 27:260–270, august 2002.

[40] D.A. Forsyth. A novel algorithm for color constancy. *Int. J. Comput. Vision*, 5(1):5–36, 1990.

[41] D.A. Forsyth, J. Malik, M.M. Fleck, H. Greenspan, T.K. Leung, S. Belongie, C. Carson, and C. Bregler. Finding pictures of objects in large collections of images. *Object Representation in Computer Vision*, pages 335–360, 1996.

[42] David H. Foster. Does colour constancy exist? *TRENDS in Cognitive Sciences*, 7(10):439–443, 2003.

[43] D.H. Foster and S.M.C. Nascimento. Relational colour constancy from invariant cone-excitation ratios. In *Proceedings of the Royal Society of London, Series B*, volume 257, pages 115–121, 1994.

[44] D.H. Foster, S.M.C. Nascimento, B.J. Craven, K.J. Linnell, F.W. Cornelissen, and E. Brenner. Four issues concerning colour constancy and relational colour constancy. *Vision Research*, 37:1341–1345, 1997.

[45] B. Funt, V. Cardei, and K. Barnard. Learning colour constancy. In *Proc. Fourth IST/SID Colour Imaging Conf*, pages 58–60, 1996.

[46] Brian Funt, Kobus Barnard, and Lindsay Martin. Is colour constancy good enough? In *5th European Conference on Computer Vision*, pages 445–459, 1998.

[47] B.V. Funt, M.S. Drew, and M. Brockington. Recovering shading from color images. In *In Proceedings of 2nd ECCV 92*, pages 124–132, 1992.

[48] Th. Gevers and A. W. M. Smeulders. Color based object recognition. *Pattern Recognition*, 32:453–464, March 1999.

[49] 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):30033010, 1994.

[50] H. Greenspan, J. Goldberger, and I. Eshet. Mixture model for face-color modeling and segmentation. *Pattern Recognition Letters*, 22(14):1525–1536, December 2001.

[51] H. Helson. Fundamental principles in color vision. i. the principle governing changes in hue, saturation, and lightness of non-selective samples in chromatic illumination. *J. Exp. Psychol.*, 23:439–471, 1938.

[52] R.V. Hogg and E.A. Tanis. *Probability and Statistical Inference*. Prentice Hall, 2001.

[53] Steven D. Hordley and Graham D. Finlayson. Re-evaluating colour constancy algorithms. In *ICPR (1)*, pages 76–79, 2004.

[54] D.B. Judd. Hue saturation and lightness of surface colors with chromatic illumination. *Journal of the Optical Society of America*, 30:2–32, 1940.

[55] Edwin H. Land. The retinex theory of color vision. *Scientific American*, pages 108–128, 1977.

[56] Edwin H. Land. Recent advances in retinex theory. *Vision Res.*, 26:7–21, 1986.

[57] Edwin H. Land and John J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, January 1971.

[58] E.H. Land. Experiments in color vision. *Scientific American*, 200:84–99, 1959.

[59] James L.Crowley and Franois Berard. Multi-modal tracking of faces for video communications. In *In IEEE Proceedings of CVPR 97*, 1997.

[60] H. Lee. Method for computing the scene illuminant chromaticity from specular highlights. *Journal of the Optical Society of America A*, 3(10):1694–1699, 1986.

[61] P. Lipson, E. Grimson, and P. Sinha. Configuration based scene classification and image indexing. *Computer Vision and Pattern Recognition*, 1997.

[62] Antonio M. López, David Lloret, Joan Serrat, and Juan J. Villanueva. Multilocal creaseness based on the level-set extrinsic curvature. *Computer Vision and Image Understanding*, 77(2):111–144, February 2000.

[63] Antonio M. López, Felipe Lumbreras, Joan Serrat, and Juan J. Villanueva. Evaluation of methods for ridge and valley detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(4):327–335, 1999.

[64] Laurence T. Maloney and Brian A. Wandell. Color constancy: a method for recovering surface spectral reflectance. *Journal of the Optical Society of America*, 3(1):29–33, January 1986.

[65] David Marr. *Vision*. W. H. Freeman and Company, 1982.

[66] E. Marszalec, B. Martinkauppi, M. Soriano, and M. Pietikinen. A physics-based face database for color research. *Journal of Electronic Imaging*, 9(1):32–38, 2000.

[67] B. Martinkauppi and M. Soriano. Basis functions of color signal of skin under different illuminants. In *Proc. 3rd International Conference on Multispectral Color Science*, pages 21–24, 2001.

[68] B. Martinkauppi, M. Soriano, and M. Laaksonen. Behaviour of skin color under varying illumination seen by different cameras in different color spaces. In *Proc. SPIE Machine Vision in Industrial Inspection IX*, volume 4301, pages 249–258, January 2001.

[69] S.M.C. Nascimento and D.H. Foster. Relational color constancy in achromatic and isoluminant images. *Journal of the Optical Society of America A*, 17(2):225–231, 2000.

[70] Y.I. Ohta, T. Kanade, and T. Sakai. Color information for region segmentation. *Computer Graphics and Image Processing*, 13:222241, 1980.

[71] T. Poggio. Computer vision. In *Image Pattern Recognition: Algorithm Implementations, Techniques, and Technology, SPIE Proceedings Vol. 755. 1987, p.54*, pages 54–61, January 1987.

[72] W. Richards and E.A. Parks. Model for color conversion. *Journal of the Optical Society of America*, 61:971–976, 1971.

[73] Hazel Rossotti. *Colour: Why the World Isn't Grey*. Princeton University Press, 1st princeton pbk. print., with corrections edition, 1983.

[74] Guillermo Sapiro. Color and illuminant voting. *Transactions on Pattern Analysis and Machine Intelligence*, 21(11):1210–1215, November 1999.

[75] J. Serra. *Image analysis and mathematical morphology*. Academic Press, London, 1982.

[76] S. A. Shafer. Using color to separate reflection components. *COLOR Research and Application*, 10:210–218, 1985.

[77] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen. Skin color modeling under varying illumination conditions using the skin locus for selecting training pixels. In *Proc. Workshop on Real-Time Image Sequence Analysis*, pages 43–49, 2000.

[78] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen. Skin detection in video under changing illumination conditions. In *ICPR00*, pages Vol I: 839–842, 2000.

[79] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen. Using the skin locus to cope with changing illumination conditions in color-based face tracking. In *IEEE Nordic Signal Processing Symposium*, pages 383–386, 2000.

[80] J.M. Speigle and D. H. Brainard. Is color constancy task independent? In *Proceedings of the 4th IST/SID Color Imaging Conference*, pages 167–172, 1996.

[81] Harro Stokman. *Robust Photometric Invariance in Machine Color Vision*. PhD thesis, University of Amsterdam, 2000.

[82] M. Storring, H. J. Andersen, and E. Granum. Physics-based modelling of human skin colour under mixed illuminants. *Journal of Robotics and Autonomous Systems*, 3-4(35):131–142, June 2001.

[83] M. Storring, H.J. Andersen, and E. Granum. Skin colour detection under changing lighting conditions. In *Proc. of 7th International Symposium on Intelligent Robotic Systems*, pages 187–195, 1999.

[84] M. Storring, H.J. Andersen, and E. Granum. Estimation of the illuminant colour from human skin colour. In *Proc. of 4th International Conference on Automatic Face and Gesture Recognition*, pages 64–69, 2000.

[85] M. Storring, H.J. Andersen, and E. Granum. Estimation of the illuminant colour using highlights from human skin. In *Proc. of 1st International Conference on Color in Graphics and Image Processing*, pages 45–50, 2000.

[86] M. Storring, H.J. Andersen, and E. Granum. A multispectral approach to robust human skin detection. In *Proc. of the Second European Conference on Color in Graphics, Image and Vision*, pages 110–115, 2004.

[87] M. Storring and E. Granum. Adapting a statistical skin color model to illumination changes. In *Proc. of the First European Conference on Color in Graphics, Image and Vision*, pages 16–21, 2002.

[88] M. Stricker and A. Dimai. Color indexing with weak spatial constraints. *Storage and Retrieval for Image and Video Databases*, 2670:29–40, 1996.

[89] J.C. Terrillon, M. David, and S. Akamatsu. Automatic detection of human faces in natural scene images by use of a skin color model and of invariant moments. In *Proc. of 3rd International Conference on Automatic Face and Gesture Recognition*, pages 112–117, 1998.

[90] J.C. Terrillon, Y. Niwa, and K. Yamamoto. On the selection of an efficient chrominance space for skin color-based image segmentation with an application to face detection. In *Proc. of International Conference on Quality Control by Artificial Vision*, pages 409–414, 2001.

[91] J.C. Terrillon, M.N. Shirazi, H. Fukamachi, and S. Akamatsu. Comparative performance of different skin chrominance models and chrominance spaces for the detection of human faces in color images. In *Proc. of 4th International Conference on Automatic Face and Gesture Recognition*, pages 54–61, 2000.

[92] F. Tous, A. Borràs, R. Benavente, R. Baldrich, M. Vanrell, and J. Lladós. Textual descriptors for browsing people by visual appearance. In *CCIA '02: Proceedings of the 5th Catalonian Conference on AI*, pages 419–430. Springer-Verlag, 2002.

[93] F. Tous, M. Vanrell, and R. Baldrich. Exploring colour constancy solutions. In *Proc. of the Second European Conference on Color in Graphics, Image and Vision*, pages 24–29, 2004.

[94] F. Tous, M. Vanrell, and R. Baldrich. Relaxed grey-world: Computational colour constancy by surface matching. In *2nd Iberian Conference on Pattern Recognition and Image Analysis*, pages 192–199, 2005.

[95] Francesc Tous. Study of colour normalisation for skin detection. Master's thesis, Universitat Autònoma de Barcelona, 2002.

[96] M. Vanrell, R. Baldrich, A. Salvatella, R. Benavente, and F. Tous. Induction operators for a computational colour-texture representation. *Comput. Vis. Image Underst.*, 94(1-3):92–114, 2004.

[97] M. Vanrell, F. Lumbreras, A. Pujol, R. Baldrich, J. Lladós, and J.J. Villanueva. Colour normalisation based on background information. In *Proceedings of the 8th ICIP*, volume 1, pages 874–877, October 2001. , Thessaloniki, Greece.

[98] J. von Kries. Beitrag zur physiologie der gesichtempfindung. *Arch. Anat. Physiol.*, 2:505–524, 1878.

[99] L. Wang and T. Pavlidis. Direct gray-scale extraction of features for character recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(10):1053–1067, 1993.

[100] G. Wyszecki and W.S. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formula*. R.K. Printing & Publishing Company, seconde edition, 1982.

[101] J. Yang, W. Lu, and A. Waibel. Skin-color modeling and adaptation. Technical Report CMU-CS-97-146, Carnegie Mellon University, School of Computer Science, May 1997.

[102] Jie Yang and Alex Waibel. A real-time face tracker. In *Proceedings of WACV'96*, 1996.