**Universitat
Autònoma
de Barcelona**

Perceptual approach to a computational colour texture representation for surface inspection

A dissertation submitted by **Ramon Baldrich i Caselles** at Universitat Autònoma de Barcelona to fulfil the degree of **Doctor en Informàtica**.

Bellaterra, November 19, 2001

Director: **Dr. Maria Vanrell i Martorell**
          Universitat Autònoma de Barcelona
          Dept. Informàtica & Computer Vision Center

**Centre de Visió
per Computador**

This document was typeset by the author using $\text{\LaTeX}\,2_\varepsilon$.

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

a la Montse, l'Alba i el Xavier

Acknowledgement and preliminary note

I know that when using a foreign language it is nearly impossible to use it fluently and perfectly, some times even moderately acceptable is difficult. However I decided to write this thesis in English instead of using my own language because, even with mistakes, it is easier to reach more people. Thus, I will like to acknowledge the people that will read this thesis for being kindly and to read the content and not so much the form, if it is possible. Now, with your consent I will switch to Catalan for a moment.

No perque, de forma establerta, sempre és el primer en aquestes llistes d'agraïments, sóc menys sincer. Gràcies Juanjo per haver-me obligat a fer la tesi. Amb aquesta última història et treus un pes de sobre, i jo també.

No hi ha manera possible d'agrair la tasca de na Maria en aquesta tesi, directora i part implicada. Però del que n'estic més satisfet és d'haver acabat aquesta tesi essent més amics que al principi, i això val una tesi i més.

No enumeraré tots aquells amb qui em sento amb deute per haver-me ajudat, o simplement haver estat en el meu entorn, ara i abans. El meu agraïment és per a tots vosaltres.

He d'agrair la empresa Alcalagres per les facilitats donades en les estances entre màquines polidores, especialment a Alfonso Remiro i Jose Antonio Trigeros, i alhora fer menció a en Jaime López-Krahe pels seus contactes inicials.

Suposo que aquesta tesi tampoc servirà per fer entendre als meus pares i germanes a que coi em dedico, però de totes formes una part també és d'ells.

Per acabar vull fer palés que sense el suport de la Montse, la vitalitat de l'Alba i la simpatia del Xavier hagués estat molt més difícil, si no impossible, fer la tesi. Es mereixen un doctorat en paciència. El millor agraïment és prometre que faré lo impossible per no fer-los-hi tornar a passar una altra època com aquesta.

Abstract

The main goal of this thesis is to deal with the colour texture representation problem from a computer vision point of view. It is easy to demonstrate that the extension of classical grey level methods for texture processing to the three channels of the corresponding colour texture does not succeed in having a human-like behaviour on this visual task. Chromatic induction mechanisms of the human visual system, that has been widely studied in psychophysics, plays an important role on the dependency of the colour perception from its surround. Chromatic induction includes two complementary effects: chromatic assimilation and chromatic contrast. While the former has been psychophysically measured and lately extended to computer vision, some aspects on the last one still remain to be measured. The main contribution of this thesis is a computational operator that simulates the contrast induction phenomena that has demonstrated a coherent behaviour on different texture colour perception problems, since it allows to emphasise colour differences on almost-unimodal colour distributions and consequently improving the segmentation of colour regions. An open problem that will remain open from this work is the psychophysical measurement of the operator parameters, in the same sense as it was done with the s-cielab for the assimilation process.

A perceptually-consistent colour texture computational representation is a goal of extreme importance in automatic colour-textured surface inspection problems, where the classic colorimetric tools does not succeed in given good colour appearance measurements. In this scope, a second contribution is a colour-texture representation based on global colour features considering colour assimilation and local features based on properties of colour blobs considering colour contrast. This representation is applied to an automatic tile classification problem.

Since an important accuracy is needed to measure such small differences, we have devoted a great part of this work to the colour acquisition issue, and to the problem of achieving good colour constancy properties on the acquired images. In this sense, a last contribution of this work has been to define an on-line colour constancy algorithm for a high colour precision scan line camera based on a diagonal linear colour constancy model previously guaranteed by linear transform changing the camera sensitivity properties.

Resum

El principal objectiu d'aquest treball de tesi és tractar el problema de la representació de la textura en color des del punt de vista de la visió per computador. No és difícil demostrar que l'extensió dels mètodes classics de processament de textura per imatges en nivells de grisos a cada un dels tres canals d'una imatge en color no és sinònim d'assolir resultats semblants als de la percepció humana en aquesta tasca. Els mecanismes d'inducció cromàtica del sistema visual humà, que han estat àmpliament estudiats en psicofísica, tenen un paper molt important en la dependència que crea l'entorn en la percepció del color. La inducció cromàtica inclou dos efectes complementaris: l'assimilació cromàtica i el contrast cromàtic. Mentre el primer ja ha estat mesurat des de la psicofísica i extés a la visió per computador, molts aspectes del segon encara queden per fer. La contribució principal d'aquesta tesi és la definició d'un operador computacional que simula el fenòmen del contrast cromàtic i que té un comportament coherent amb el del sistema visual humà en diferents problemes de la percepcció de la textura en color, ja que permet enfatitzar les diferències de color en distribucions que són quasibé unimodals i conseqüentment millorar la segmentació de les petites regions de color. El problema que encara queda obert després d'aquest treball, és la realització de mesures psicofísiques pels paràmetres de l'operador definit, tal com es va fer amb l's-cielab per al procés de l'assimilació.

La definició de representacions computacionals de textura i color que siguin perceptuals és un objectiu de gran importància en els problemes d'inspecció automàtica de superfícies en els que els dispositius de la colorimetria clàssica no permeten donar bones mesures d'aparença de color. La segona contribució d'aquesta tesi, s'emmarca en aquest àmbit, i defineix una representació computacional basada en mesures globals de color que inclouen l'assimilació de color i mesures locals de les propietats de les regions segmentades considerant el contrast cromàtic. Aquesta representació és aplicada al problema de la classificació automàtica de gres porcelànic.

Tenint en compte que s'han de realitzar mesures molt acurades de petites diferències, s'ha dedicat una gran part d'aquest treball al tema de l'adquisició d'imatges en color, i en concret al problema d'aconseguir bones propietats de constància de color a les imatges adquirides. En aquest sentit, la darrera contribució d'aquest treball ha estat la definició d'un algorisme de contància de color en línia per a una càmera lineal amb alta precisió de color. Aquest mètode s'ha basat en el model lineal diagonal de constància de color prèviament garantit amb una transformació lineal que canvia les propietats de la sensibilitat de la càmera.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The aim of this thesis is the computational representation of two surface properties: colour and texture. To build computational representations of visual information is an essential goal in the computer vision field in which this thesis is framed. A computational texture-colour representation has to allow building automatic descriptions of surfaces that can help in a wide range of computer vision tasks. A large number of works have been reported in the last decades on these two properties separately. But, for the last years the number of works dealing with both properties at the same time is increasing considerably. In this chapter we give a brief review of previous works on colour and texture to put the scope of this thesis within the computer vision field.

1.1   Surface properties

Any natural scene in the world is projected on our retina as a map of different regions that are the projections of 3D surfaces. The properties of these projected surfaces are concrete perceptions derived from specific positioning conditions of the surfaces in the scene and the observer, and the lighting conditions that provoke the neuronal excitation of the visual system. In computer vision, people usually work with the following set of surface properties: shape, orientation, colour and texture. In this work we will only deal with the last two.

Up to this point, we have only revealed one of the two goals of this thesis. The second goal of this thesis is also to develop the engineering background to take the computational texture colour representation to make it works on a real system ready to solve problems of automatic measurement of surface properties in the industry. A wide range of automatisation of industrial problems requires the measurement of coloured surfaces. These measurements are easily solved using calibrated colorimetric devices specially developed to measure colour-homogeneous surfaces. This solution fails when surfaces are coloured textures. Colorimetric measurements on colour texture surfaces give a quantitative measurement that is the result of a colour integration over the surface, and two very different colour textures can give similar measurements even though they have a very different spatial appearance, this is the case for the three
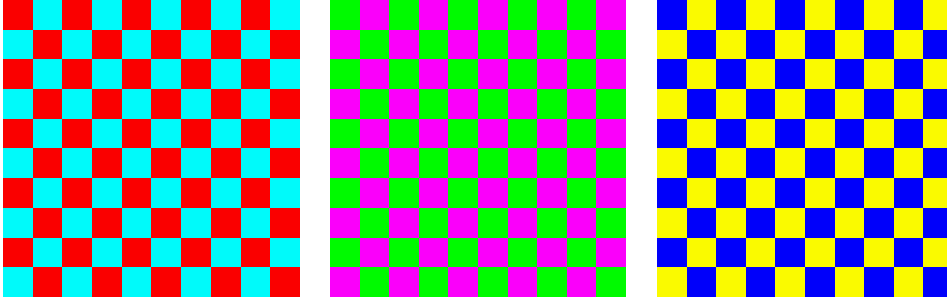
Figure 1.1: Images with the same colour mean but different appearance

images shown in figure 1.1 that share the same colour mean but a very different colour texture appearance.

In the next sections we will give a brief outline on how colour and texture have been treated in computer vision.

1.1.1   Colour

Colour is the visual cue derived from the human visual processing of the electromagnetic radiation that reaches the retina [78]. This process can be seen as a change in representation, which, in general, implies a dimensionality reduction. Although colour has not been given much importance in the first decades of computer vision, since most of the previous work in computer vision has been made for grey level images, the situation has changed and colour has become a very important visual cue for most of the vision tasks, such as object recognition [47], image indexing [100], tracking [66], shape extraction from colour variations [16], etc.

To introduce colour cue in the visual tasks we must take into consideration the variability of this visual stimulus. Colour perception is always dependent on the context: the illuminant, the receptor sensibility and the scene geometry have a great influence on the perceived scene. The human visual system presents a chromatic adaptation ability, which allows avoiding in some sense those context influences over the final perception. Any system doing a visual task involving colour processing should always take into account the colour constancy problem. This problem has been the topic of a lot of research that will be reviewed in chapter 3.

1.1.2   Texture

Texture is the visual cue derived from non-homogeneous surfaces in the scenes. Depending on the surface reflectance, positioning of the observer and lighting conditions, we can obtain different texture images from the same surface. Although there are some recent works dealing with the recovery of the physical reflectance properties of a texture [25, 48] and some other works that have recovered 3D shape information from texture [115, 44], the most traditional approach in computer vision has been the analysis of the texture images without taking considerations on the image formation

process. Extensive reviews can be found in [49, 111, 103, 90], where it is shown that texture has been studied for different purposes such as segmentation, classification or synthesis. Despite the large number of works, there is still a lack of a standard texture definition and does not exist a widely accepted texture representation space, as it exists for colour. Interesting works directed to define a standard texture space based on perceptual considerations has to be considered [89, 88, 101], since this kind of work could be the basis to establish a standard computational representation. Before to go deeply on computational representations we will do a short inside on psychophysics theories on texture perception, that have been the basis for some of the works in computer vision.

In psychophysics, the aim has been to understand how the human visual system represents textures and which are the mechanisms used for texture segregation. Texture is one of the most complex visual cues and for the moment there is not a unique accepted theory. Two basic approaches are confronted as being the basis for a visual internal representation of texture. On one hand, feature extraction processes have received a hard support from the Julesz's texton theory [62], and on the other hand a global spatial frequency analysis seems to be indispensable as it has been demonstrated by J. Beck et al in [7]. Let us go deeply in these two approaches.

The first approach, the Julesz's texton theory is based on the fact that differences between two textures, are due to differences in the first order statistics, or densities, of the texton attributes, it ignores the positional relationships between adjacent textons. Texton attributes are defined as the blob properties, that is, size and contrast for general blobs, and orientation for elongated blobs. Other textons can be line endings or terminators, but a more exhaustive list of texton has not been developed yet. Although all the texton theory conclusions are based on psychophysical experiments, Julesz associates the feature extractors with simple or complex cortical receptive fields described by Hubel and Wiesel in 1968.

The second approach, leaded by J. Beck [7] and supported by other researchers [52, 51] advocates that, differences on first order statistics of local properties independently of the blob arrangement is not enough to be able to capture the segregation of textures, since in a wide range of cases, differences are due to patterns emerging from the different arrangements of image blobs. In these cases a global spatial-frequency analysis is needed in order to represent different textures.

In figure 1.2 we demonstrate the complementary character of these two approaches. While the textures (a) and (b) can be easily differentiated in the frame of the Julesz's texton theory due to differences on blob contrast; textures (b) and (c) are equals from this theory, since there is no difference in terms of texton attributes. Differences between textures (b) and (c) can be easily derived in the frame of a of a global frequency analysis, for which a difference in emergent orientations can be considered.

Considering the conclusions from psychophysical theories, different approaches have been followed to solve problems involving different visual tasks in computer vision. We briefly summarise the taxonomy proposed by Tuceryan et al in [103]:

Geometrical approach Texture is described by the set of textural primitives that composes the image, therefore a texton isolation step is always needed. Once the basic elements have been extracted, two approaches are essentially used. One

(a)   (b)   (c)
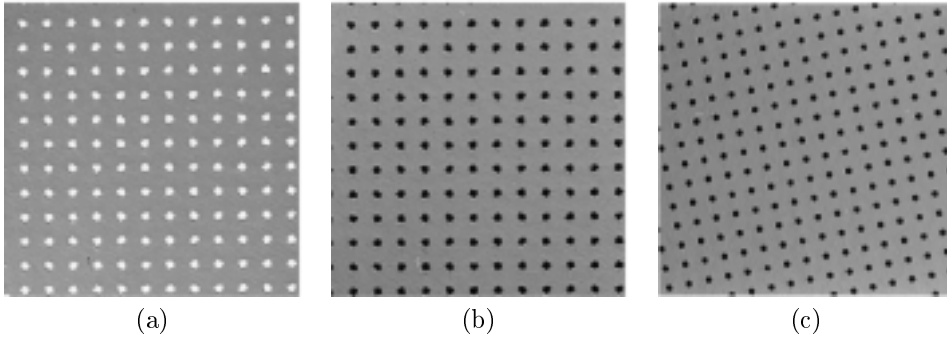
Figure 1.2: Examples of textures formed by simple blobs and their emergent patterns.

computes statistical properties of the extracted elements and their attributes [105]. The second one extracts the placement rules that organise these shapes in the texture [41], this last approach is called a structural approach.

Model-based approach Texture is considered as the realisation of a concrete mathematical model, hence it is defined by the model parameters. From a methodological point of view this is the most well defined solution, problems can arise from generality, it does not exist a unique model that can represent any natural texture. Interesting texture models can be seen in [1, 54, 61, 87, 23]

Filtering approach Texture is described by the responses of convolving a set of filters with the image. This approach is based on the previous introduced idea of the existence of an spatial-frequency global analysis of the textures in the human visual system. Malik and Perona in [71] proposed a global preattentive texture perception model based on neuro-physiological and psychophysical considerations. A global Fourier-based analysis of textures has been recently proposed in [39] and when spatial dependency is needed the Gabor transform has been used [59, 73].

From all these approaches, different visual tasks can be carried out. In texture segmentation, region-based or edge-based mechanisms have been used, all these methods try to evaluate when two small regions have a uniform texture or, on the contrary, have different textures. The general problem of texture representation has been mainly developed for image retrieval, image annotation or image classification. In all these problems the final goal is to build a feature vector expressing an enough quantitative measure of the image content. The rest of this work will be devoted to study how texture and colour can be combined considering these previous experiences on gray level textures.
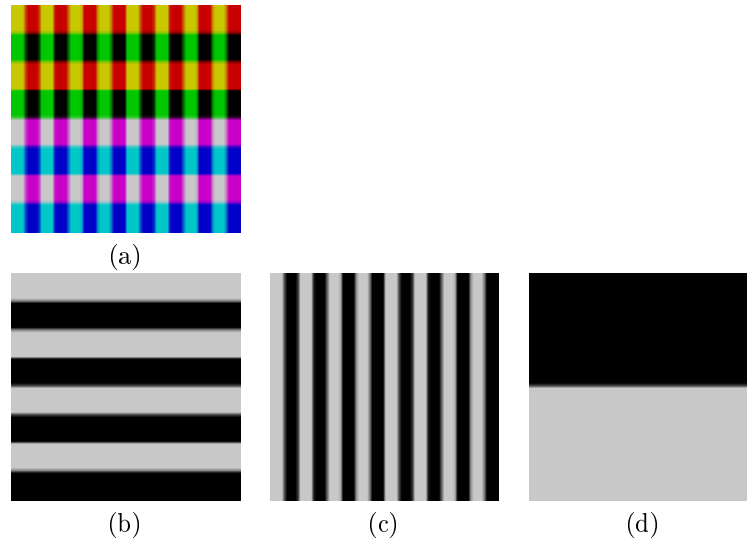
Figure 1.3: (a) A colour image. (b) Red channel of (a), (c) Green channel of (a), (d) Blue channel of (a).

1.2 Colour and Texture

Colour texture representation is a current topic in computer vision. Although both are properties of a surface as we have just introduced, these two visual cues have been usually studied separately. One reason is that while colour is a point feature given by the value of a pixel in several bands or channels, texture has to be modeled as a spatial relationship of the point with its neighbours. The trichromatic representation of colour images taken from common imaging devices has provoked an important dependency, that is probably not the best to deal with these two dependent properties. In figure 1.3 we can see the RGB channel representation of a colour image, where we can observe that the spatial information of the colour image is not present in the separate channels and therefore specific representations have to be constructed in order to deal with both cues at the same time.

The study of colour texture representations has received an increasing attention. The objective of many researchers has been to find co–joint representations of spatial and chromatic information which capture the spatial dependence (in particular, correlation) *within and among* spectral bands. One of the most frequent approaches is the construction of a feature vector mixing grey level texture features and colour features [19, 102]. Another one is to extend classical texture models, such as Markov Random fields and the autocorrelation function, in order to deal with multichannel images [82, 53]. Other works, like [42], convert RGB values into a single code from which texture measurements are computed as if it were a grey scale image. Spatio–chromatic representations are computed in [17, 37] over the smoothed Laplacian of the image, and the structural tensor that is usually used to represent local texture

properties is extended to colour images in [113].

Finally, there are some works that have been influenced by known perceptual mechanisms of the human visual system, where the interaction of colour with the spatial frequency of the coloured patterns is considered [84, 80]. These works have considered some important conclusions from psychophysical experiments on colour texture interaction which are the conclusions of some works [2, 85, 109, 119, 118]. The contributions of these works and its application to computer vision will be reviewed in more detail in chapter 4. This perceptual mechanism simulates the colour assimilation phenomenon of the human visual system that is affected by a spatial blurring of the colour representation when looking at colour textures presenting high spatial frequencies.

In this work we will present a complementary operator that will allow simulating the colour contrast phenomenon that appears in the visual system when looking at colour textures presenting low spatial frequencies.

1.3   Thesis Outline

The content of this thesis work has been organised in five chapters. Chapter 1 is the introduction we have done above. We have introduced the thesis goals and a brief introduction on how colour and texture have been studied in computer vision.

Chapter 2 is devoted to explain the design of a colour image acquisition system. Since one of the final goals of this work is to design a vision system able to measure colour appearance on textures as colorimetry does on homogeneous surfaces, we will need to take an special attention to the accuracy and to the stability of the designed system. Is for this reason we will dedicate a complete chapter to the problem of acquiring colour images with a CCD-based sensor.

In chapter 3, we give the basis of a colour image formation and the laws underlying colour constancy theories. Afterwards, a brief review of the most important methods for colour constancy is given. In order to be able to apply a linear diagonal model the spectral sharpening transform is computed once the sensitivities of the camera have been recovered. In the last part of this chapter an on–line colour constancy algorithm for scan line cameras is proposed.

Chapter 4 begins with a review on psychophysical literature, it is directed to establish the basis for the most common colour induction phenomena: assimilation and contrast. Considering the most important conclusions from the previous review, a computational pattern-colour separable model based on the opponent-colour space is derived. The chromatic assimilation model based on a perceptual blurring is introduced, and all the details of the Spatial–CIELAB are explained. In the following sections we propose a chromatic contrast model that is based on a perceptual sharpening. Three types of perceptual sharpening are proposed: local, region and spread. Finally, we show the behaviour of the spread sharpening on some natural textures.

The goal of chapter 5 is to build a computational colour texture representation based on the previous considerations, and to apply it to a pair of automatic surface inspection problems, these are: tile classification and printing quality evaluation. In both cases we will see how the perceptual blurring proposed allows improving the

results.

In the last chapter we sum up all the conclusions of this thesis work, and after a short discussion on the results we describe the open research directions that have been outlined from this work.

Chapter 2

Colour image acquisition

As we aim to work on machine vision problems, where colour is one of the main cues to take in regard, the acquisition system is the first issue to deal with. It is not the same to construct a system to get geometrical or structural measures than defining the best way to acquire a surface under spatially–homogeneous conditions. Another feature that the system has to obey is the ability to measure small colour differences. We will analyse the relevant parameters of the CCD cameras and their involved problems and how to solve them. We also focus on the construction of a lighting architecture for a surface inspection system. Taking into account these requirements let us examine the different alternatives, and the proceedings that should be done.

2.1   Dark current

In this chapter we are assuming that the image acquisition is done by means of a CCD device. It could be done using other type of sensors, such as CMOS devices but, up the moment, CCD's are the most suitable for high performance and *low cost* imagery. In such devices the pixel information comes from a photo-sensor and the circuitry involved in transforming emitted light to a digital or analog quantity. In an ideal case [43], given a point in the light emitted $L(x, y)$ the final result $Z(x, y)$ should be

$$Z(x, y) = tL(x, y)$$

where t includes the exposure time and the factors involved in the manipulation of the signal. However it is not the case. Many problems affect the acquisition process; some of them are noise, black current and sensitivity irregularities. Combining these effects the final representation is given by

$$Z(x, y) = D_t(x, y) + tF(x, y)L(x, y) + N(x, y)$$

where $D_t(x, y)$ (*dark field*) is the signal from the *dark current signal* (also called *thermic noise*) [55]. It is induced by thermal excitation of the CCD components and

9
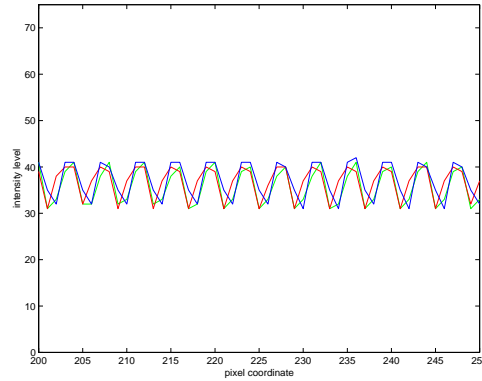
Figure 2.1: Example of a dark field for a TVI line scan camera. When acquiring an image without dark current signal correction, the colour information has 35 intensity levels above what it should be.

may vary with position in the array of photo-sensors, and exposure time. $F(x, y)$ is the function that captures the relative sensitivity of the device in each position of the CCD array in relation to its ideal behaviour. $F(x, y)$ is known as the *flat field*. $N(x, y)$ is the intrinsic noise of the system. Both $D_t(x, y)$ and $F(x, y)$ can be avoided but not $N(x, y)$. Usually $N(x, y)$ is defined as a constant amount in *db* all over the sensor and related to the input signal (*signal–to–noise–ratio*), and it should be provided by the manufacturer. Our first choice is the camera with the less noise level. Later in this chapter we will describe how to solve the dark current signal problem. Although the flat field defines the CCD sensitivity it could be analysed with other effects introduced by the optics and lighting conditions as a whole. This will be studied in the next chapter.

As $D_t(x, y)$ is an additive signal and supposing that the $N(x, y)$ signal is small enough, it could be easily subtracted from $Z(x, y)$. Getting $D_t(x, y)$ is as simple as setting $L(x, y) = 0$, that is covering the entrance of the light, so that $Z(x, y)$ only respond to the dark current signal. However, and in order to minimise the effect of the intrinsic noise, we need to average several dark fields. Another parameter that is significant in the resulting dark field is the effect of the exposure time. The coefficient t represents the fact that the dark current depends on the amount of time the CCD is active. This means that a set of dark currents corrections should be recorded for each different exposure time our system will work. Fortunately, most of the industrial vision problems work at a fixed cadence, and we only need one such a measure.

The last point that we should be aware of is the effect of the device temperature. Dark current increases when the temperature increases. To solve this problem the sensor is cooled using liquid nitrogen (in the most expensive case) or using forced–air cooling which is the cheapest solution, among other possibilities [15]. In industrial application the forced–air cooling is the most used solution. This solution stabilises the temperature after some working time. It should be considered, and take the correction dark field information after one hour and a half the camera comes into

operation, as an average. Figure 2.1 shows an example of the dark field calculated from a small section of a TVI line scan camera line response. It should be noticed there is a difference of 35 intensity levels on average between the real acquired image and the corrected one. That means colours become overemphasised in some cases. Moreover, this particular case seems to include problems with the synchronisation between the camera and the frame grabber, although the manufacturers state that it is correct.

2.2 Camera architecture

This section tries to answer to the question of which is the best camera architecture to acquire colour flat surfaces, obtaining the best colour fidelity. The two main architectures used in industrial vision are the matrix array cameras and the linear array cameras. Whenever it is possible it is much better to use a matrix camera than a linear one. Some of the reasons for this choice are:

- There is no need of camera synchronisation with the conveyor–belt. Even in the case when the conveyor–belt can not stop, we can use non–interlaced matrix cameras with high shutter speed to remove the motion blurring effect. In the case of line scan cameras the exposure time, the horizontal synchronisation and the vertical sync signal have to be adjusted to the cadence of the production line. The last one of these parameters is the most sensitive to produce a good result.

- It is easier to focus the scene. As what we see in the monitor is what we get it is quite obvious to focus the scene without using any artifice. In linear array cameras what we see is just a line of the scene and it is difficult to visually focus the image from a scene that could be complex. Some simple and a priori known patterns should be used for this purpose.

- There is no need to perfectly align the acquisition system with the transporter system. Whereas when using a matrix camera we get an image of the scene in a flash (stoping the conveyor or any other means), it is not possible with a line scan camera.

When choosing a colour matrix camera it is important to use a 3 CCD instead of a 1 CCD camera. The colour sensitivity is much better in the first case because when using a 1 CCD camera, the responses on the red, green and blue channels are mixed in the same CCD, obtaining less spatial colour resolution.

However, there are some cases where matrix cameras are not suitable to the inspection problem. One set of them is that involved in this work: industrial vision problems where the solution needs a very accurate degree of colour representation on the acquired image. The problem of obtaining a good image is not from the camera itself but from the lighting system. Now, we are going to examine which are the options and the pros and cons for both cases.
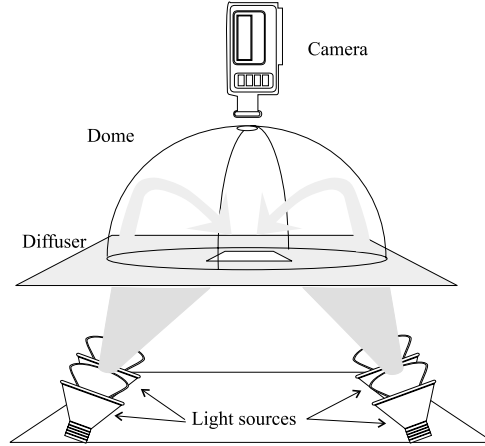
Figure 2.2: Lighting architecture for matrix camera acquisition. The light is diffused using a reflective dome and a diffuser material to avoid predominant light directions from bulbs.

2.2.1   Lighting for a matrix camera

In this section we will explain why the matrix camera performs badly in colour accuracy. Although there have been interesting developments on how to place lights to get uniform frontal illumination [45], we have not succeeded in getting good homogeneity properties. Therefore, we decided to go to a more directed architecture. The lighting architecture we have used is shown in the the diagram in figure 2.2. The light source comes from below the sample a hits a white dome. The semi–spherical shape of the dome makes the light to reach the sample from all possible directions and angles. This is what makes the illumination to be homogeneous. In the theoretical case it should be a flat back–light surface source but if the dimensions of the sample are greater than a few centimetres there is not such a device. One of the solutions is to diffuse the light sources before entering the dome. In the experiment four 250 watts bulbs where placed at the corners. In this way there should not be predominant directions of light. The dome is opaque to avoid light entering downward. The camera is placed at the top of the dome using a hole as small as possible.

To work with applications that request a very sensitive ability to deal with colour differences, the main point is to get the surface illuminated as much homogeneously as possible. If a constant colour surface is acquired and there are high differences among pixels all over the image, colour processes will not be reliable. We tested the configuration on figure 2.2 to its spatial light homogeneity. Figure 2.3(a) is a 3D display of the intensity level of the red channel of a constant very light brown tile. It should be noticed that there are significative differences between the corners of the acquired image. The differences on the corners of the image are ranged from 4% to 6% of the dynamic range of the camera used in this experiment. Another effect of the dome lighting configuration is that it appears a *hole* in the centre of the image

(a)   (b)
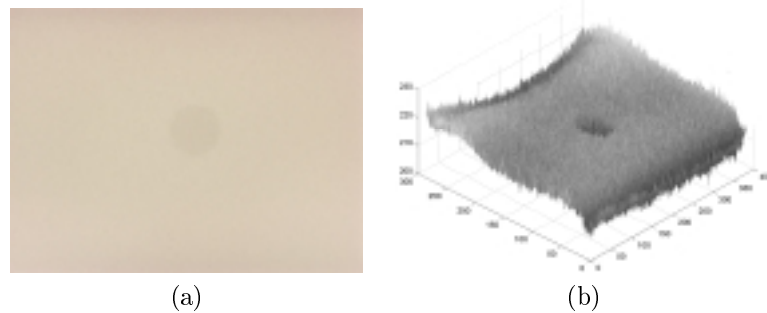
Figure 2.3: Lighting non-uniformity for matrix camera acquisition. (a) a sample acquired using schema 2.2 (b) is a 3D representation of the red channel of a constant colour surface.

(a)   (b)   (c)
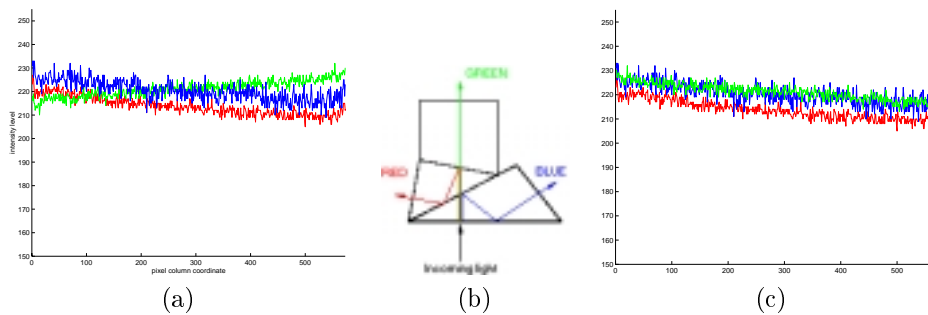
Figure 2.4: Lighting non-uniformity for matrix camera acquisition. (a) is the profile of one of the columns of a white surface. All channels should maintain its ratios between them on the whole image. (b) is a diagram of the set of prism to guide the light to the respective CCD of a 3 CCD colour camera. (c) an curious result where the profile run parallel when plotting upside down the green response.

that corresponds to the hole in the highest point of the dome where the camera lens is placed.

Apart from the lighting conditions, other problems are derived from the use of a matrix camera. When acquiring a constant colour surface it is expected that the relationship between channels remains unchanged across the image. There might be changes in the intensity level but, even in this case what is red (or any other colour) should be seen red in any place of the image. We tested two cameras (Sony XC–003P and an equivalent JAI M–90) and both of them behave wrongly. Figure 2.4(a) plots the red, green an blue profile for a certain column of an image of a white surface. While red and blue channels go by parallel, the green channel crosses both of them. The top of the image looks like more reddish than the bottom that looks like greenish. We tested the light changing the position of the bulbs and the behaviour remained the same. The lens was also tested having no change on the profiles. Although it is
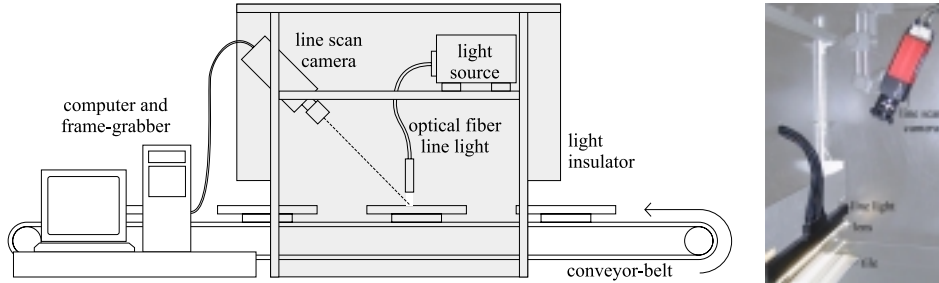
Figure 2.5: Lighting architecture for line scan camera acquisition.

not documented it seems that the reason could be a combination of the prisms used
to direct the light to the CCDs and the CCD themselves (figure 2.4(b). A very small
de–correlation between the prism and the corresponding CCD changes the behaviour
dramatically. But, all in all, what it is more astonishing is the fact than when plotting
the vertical profiles of the red and blue channels in the correct way, but the green
upside down all the profiles run parallel, as shown in figure 2.4(c). It seems that the
green prism has been calibrated in the inverse way that it had should be done. No
camera manufacturer has reported such problem. However, this is a point that can
not be confirmed and what we can do is to correct this defect or to avoid 3CDD colour
cameras presenting this problem.

2.2.2   Lighting for a line scan camera

All the homogeneity problems derived from the use of matrix array cameras could
be eluded changing the camera architecture to a linear array one. The basics of the
camera are the same except that what you acquire in each step is only one line. This
makes a lot of differences when designing the acquisition system. The first problem
is that the signal from the camera is not as standard as a matrix camera. It involves
more complex and more expensive frame grabbers. Supposing that it is a minor
problem let us concentrate on the architecture of the system and the methods to
make it work. Although most of them can be easily deduced, it is worthwhile to
comment some of them here.

  This new architecture will always need a conveyor–belt or any similar mechanism
to make the inspection scene run under the camera, an example of a typical design
is presented in figure 2.5. Therefore, it is one of the solutions for those applications
where the target can not be stopped or it is a non–end production line. But what it
is of interest to a surface inspection system is the fact that only one line is acquired
at each frame. This is a great advantage when designing a homogeneous lighting
system. Instead of worrying about getting a rectangular homogenous–lighted area it
is only necessary to achieve a thin homogeneous light strip, which is simpler. There
are several manners to do this:

- The first option we have considered is the use of a line optic fibre connected to
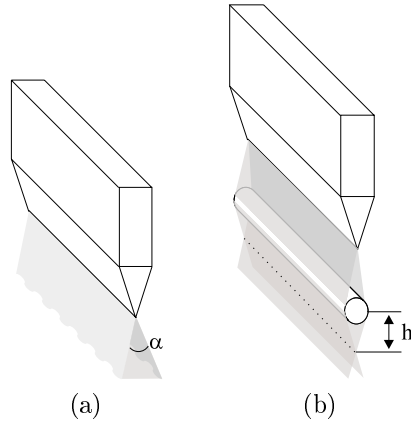
Figure 2.6: Lighting architecture for line scan camera acquisition.

a light source, as can be seen in figure 2.5. The type of light source will be an important choice. We have considered two options:

– A tungsten halogen lamp is the most common choice, but its spectral colour distribution has taken different problems. This lamps has a colour temperature of 3200°K, that is a very reddish light. Considering that digital cameras without any IR filter tend to have much more sensitivity on the red sensor than on the other two, we needed to use a set of blue filters that makes to loose a great amount of light, and you need to increase the number of light sources in four times.

– A second option was the use of a metal halide lamp. This type of lamp is used by professional photographers when they want to simulate indoor sun light. It has a colour temperature around 6000°K, that is a bluish light that does not increase red sensitivity of sensors and there is no need to add more light sources. Furthermore, its life is 5 times the life of tungsten halogen lamps.

Following with this architecture, when using an optic fiber line light the light beam is not parallel but spreads as a cone, like the diagram of figure 2.6(a). In this way a great amount of light is wasted, and considering that the aperture of the optic should be as much closed as possible for increasing the depth of view this is an important factor in the final design.

Moreover, line scan cameras work at a very high frame rate which translates to a very short exposure time interval and so more light is needed. To solve this problem we can use a lens to focus the light beam wherever we like as in the figure 2.6(b). This schema was also used in [83]. One more point to take into account in this architecture is the distance of the sample to inspect to the lens. Ideally it should be at a distance h from the lens to maximise the amount of light available, where h is the focus distance of the selected lens. Nonetheless,

(a)   (b)
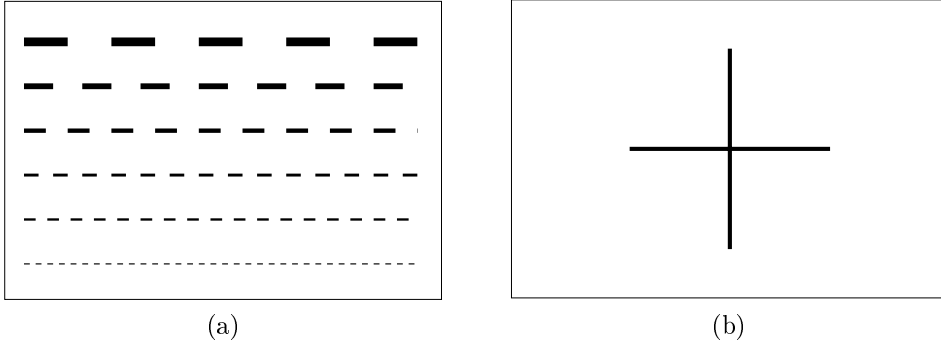
Figure 2.7: The patterns used to calibrate a line scan camera. (a) pattern used to align the array sensor and the transporter device, and to focus the optic system. (b) simple pattern to guide the synchronisation of camera an conveyor–belt.

this is not the correct answer. The line acquired with the camera is very narrow, and if the light beam is very thin then it would be very difficult to maintain a constant illumination all over the line. Therefore, the sample should be placed at a distance h' smaller than h. As smaller the h' the easier will be getting a homogenous illuminated surface and the darker it will be. The best h' will depend on each single application.

- Another option is to use white light emitting diodes arrays. The LED array is a very long life system but is must be build ad hoc for each problem. In such a system each diode should be controlled individually to equilibrate the amount of light on the line. We do not know of any commercial system with these characteristics.

- One more flexible solution is the use of fluorescent lamps that have the inconvenient of its low lighting power. This is solved by using special lamps that do not coat a small strip of the glass. Thus, the light in this strip is more intense and can be directed to the sample. This solution is the one that is currently working in laboratory conditions.

In the preceding section it has been noted that some artifices have to be used to align, focus and synchronise the line scan camera. The methods most widely used and that we have applied to our inspection line are explained below.

Aligning and focusing a line scan camera

The first step once the camera is set up for acquiring is to align the sensor array with the conveyor–belt. It is an operation that on most cases must be done in static mode. While in the case of matrix array cameras what we will acquire is exactly the scene under inspection and it can be easily tested for focusing, in the case of line scan cameras we do not have this intuitive feedback. It is difficult to understand what you are seeing when only one line of the scene is visible. To fix this obstacle we use a very

simple line pattern (figure 2.7(a)). It is very useful not to use a fixed pattern but a set of patterns from coarse to fine. In a first moment the coarsest one is used for an initial configuration, then the pattern is changed for a finer one until the desired precision is achieved.

The second step involves setting up the optics focus. In fact it could be done at the same time that aligning is done. Some manufactures suggest using an oscilloscope to monitor the output signal from the camera. This approach needs some skills on electrical engineering and is not very intuitive because, apart from the sample acquired signal, there are electrical signals as line–transfer, back porch, end porch, etcetera that make difficult its interpretation. One simple way to do it is to plot the profile of the response of one single line from the image acquired. The steeper the changes are in the transitions between white and black the more focused image we get. This step can be automatised by any auto–focus process as for example, maximising the energy from a contour detector.

Synchronising a line scan camera

The last problem is to make the pictures have a 1:1 spatial ratio, *i.e.* they should maintain the proportion of the real scene. The best way is to use an encoder to synchronise the speed of the conveyor–belt with the line rate of the camera. Sometimes it is not possible, difficult or unnecessary because the speed is always maintained constant. In these cases an initial set up is essential. It could be fixed using a known simple geometric pattern as a cross, like in figure 2.6(b). The extreme points of the cross are detected by means of morphological operators (for example: *hit or miss* operator) or blob analysis, and the ratio between the vertical and the horizontal line length evidence whether the line rate should be augmented (or conveyor–belt speed down) or viceversa. This is an iterative process until the following ratio is 1.

$$R = \frac{horizontal\ line\ length}{vertical\ line\ length},$$

$$\text{synchronisation action} = \begin{cases} \text{increase line rate} & \text{if} & R > 1 \\ \text{decrease line rate} & \text{if} & R < 1 \\ \text{no action} & \text{if} & R = 0 \end{cases}$$

2.3   Temporal stability

In 2.1 we have commented the effect of the temperature on the dark field current intensity, hence whichever it is the camera architecture we have to test the temporal stability of the whole system, camera, optics and lighting. What we have done is to take images of an homogenous sample every 5 minutes until the system stabilises. We tested two configurations to know which part of the system is more sensitive to the warming up. In the first case (figure 2.8(a)) the camera and the light had been switched on at the same time when they are at ambient temperature. In the second case (figure 2.8(b)) the camera was warmed up for two hours before turning on the light. The tests show that the effect is more intense when the camera is cool, but it is more lengthy when the light that is not warmed up. The camera needs one hour

(a)   (b)
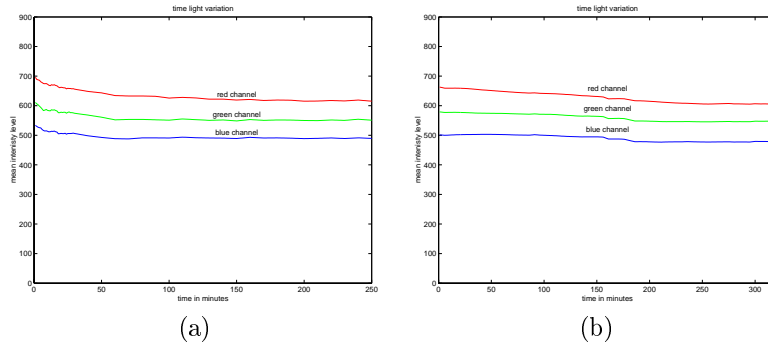
Figure 2.8: Stability of the acquired signal. (a) Evolution of the signal when switching on the camera and the light at the same time. (b) evolution of the signal when the light is turned on after two hours the camera has been switched on.

approximately to become thermally stable whereas the light needs up to two hours. In any case the maximum of these times will be the warming up time. Although other cameras and illuminations have been tested, the case shown as an example is on a TVI lines scan camera with halogen tungsten lamp. The times may vary if camera or lighting is changed. In each particular case this test has to be done.

2.4   Discussion

As a conclusion we want to note the importance of the dark current signal calibration in any industrial application with high colour (or gray) detail needs.

On the other hand, we think it is advisable to use a linear camera whenever it is possible because of its ability to deal with uniform lighting and avoiding transversal colour aberrations of some 3CCD matrix cameras.

However, the use of linear cameras involves some extra work to be done. More tricky methods are needed to focus the image, and additional hardware or extra methods have to be designed to synchronise the whole system.

As the last point to mention is the necessity of a pre–warming time to assure the correct operation of the system, especially when most of the calibration processes are done at the launch of the inspection (or whatever task) system.

Chapter 3

Colour constancy for inspection problems

In this chapter we introduce the phenomena of colour adaptation and some of the approaches to computational resolution. It is fundamental to take it into account in an inspection system where reproducibility is basic. As the method used needs to assure that camera sensor responses are independent, we show the result of the existing literature to transform a set of sensor that do not have to hold these properties to ones that they do. Finally, we explain the problems on assuring a temporal and spatial constancy of the colour representation and the approach adopted in our case.

3.1   Introduction

Colour constancy is one of the phenomena that human vision system performs when processing a visual stimulus from a scene of the real world. It is also called colour adaptation in the psychophysics field. It can be defined as the ability to perceive the same colour perception from a given surface even with changes in the illuminant [30]. As an example, suppose the daylight scene of figure 3.1(a) has been taken with a blue filter. The result will be an image like the one in figure 3.1(b). In both cases the visual system perceives the top of the woman to be yellow. In fact if we superimpose the woman's shirt of the second image over the shirt on the first image (figure 3.1(c)) we will perceive it as green. In computer vision, there are many situations where light changes and so does the stimulus acquired. If the goal of the vision system is to deal with colour information of the scene, colour constancy is a major issue. A lot of work has been done and is being done in this subject. There are various approaches to the problem for different conditions and using different methods. Some of them will be summarised in the next section.

The work done in this field starts from a specific model of colour image formation. In this process there are three main elements, which work together to compose a colour representation. These are the surface being seen, the light under the surface

<div align="center">(a)   (b)   (c)</div>
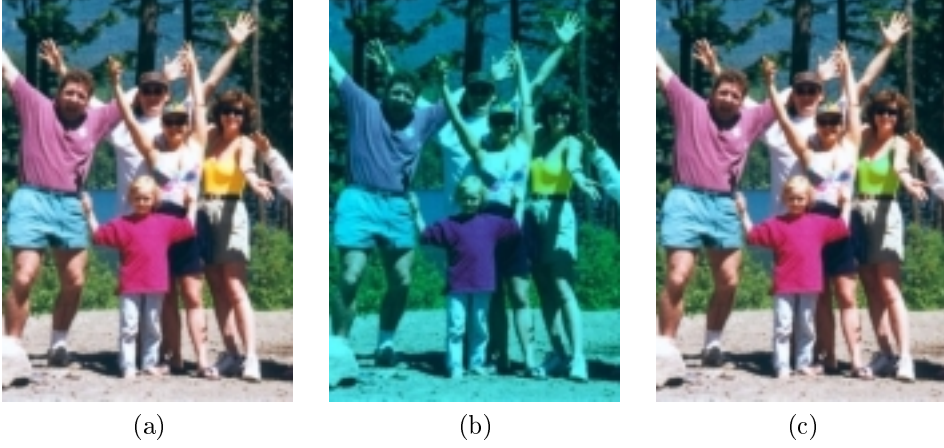
Figure 3.1: Colour constancy example. (a) a scene taken with daylight. (b) simulated blue light of the same scene. In both case the woman's top are perceived as yellow. (c) this is the original scene mixed with the shirt in the second image. In fact, the top on the second one is green although the perception is different.

is seen, and the device used to see the surface. The equation that models the colour formation is

$$\rho_x^k = \int_w L_x(\lambda, \Theta) R^k(\lambda)\, \mathrm{d}\lambda \qquad k = 1, 2, \dots, p \qquad (3.1)$$

$L_x(\lambda, \Theta)$ is the spectral power distribution emitted by the surface at a certain location x, $R^k(\lambda)$ is the spectral sensitivity of the k-th sensor of the receptor, ρ_x^k is the response obtained from the position x on the scene for the sensor k, and w is the visible spectrum. In the Visual Human System there are three types of sensors and so it is called a trichromatic model([117]) or Young-Helmholz theory, but there are many other possibilities. Some animals can see in 4 basic colours whereas others can only see in 2 or 1 colour. In computational vision p also may vary. It is the case of gray cameras or multi-spectral band cameras, normally used in remote sensing. As our purpose is the analysis of colour surfaces we will use the trichromatic model and set $p = 3$.

We will consider the dichromatic reflection model introduced by Shafer [92] to model the interaction of light with a surface. In short, it states that for a certain location x, $L(\lambda, \Theta) = L^s(\lambda, \Theta) + L^b(\lambda, \Theta)$, where Θ define the geometry of the light, the surface and the sensor. $L^s(\lambda, \Theta)$ corresponds to the specular light emitted by the surface, and can be omitted if we can guarantee that it will never occur. When using controlled conditions this is the case, and we will ignore it in our study. In these conditions such configuration is called the Lambertian diffusion model. $L^b(\lambda, \Theta)$ is defined as the light that is not absorbed by the body (surface) and will cause a certain colour stimulus to hit the sensor. It can be divided into two factors:

$$L^b(\lambda, \Theta) = m_b(\Theta) c_b(\lambda)$$

where $m_b(\Theta)$ captures all the geometric information and $c_b(\lambda)$ the physical properties of the light and body. If we use an homogeneous illumination all over the scene, for practical usage $m_b(\Theta)$ can be ignored, and consequently we have $c_b(\lambda) = I(\lambda)S(\lambda)$, where $I(\lambda)$ is the light spectral power distribution and $S(\lambda)$ the surface reflectance. At the end, equation 3.1 can be rewritten as

$$\rho_x^k = \int_w I(\lambda)S_x(\lambda)R^k(\lambda)\,\mathrm{d}\lambda \qquad k = 1, 2, \ldots, p \tag{3.2}$$

where the colour representation of point x on sensor k is given by the incident light, $I(\lambda)$, the reflectance of the surface at this point, $S_x(\lambda)$, and the sensitivity of the k–th sensor, $R^k(\lambda)$ at every single wavelength.

3.2 Basis of computational colour constancy

We will describe the colour constancy general basis that apply on most methods. The aim of computational colour constancy is to get a representation of the acquired stimulus as it has been acquired under a known illuminant. This definition does not include those methods that reach constancy obtaining a representation invariant to colour and/or intensity light changes. This representation can be quite abstract without an evident human interpretation. Inspection system are more concerned with the first set of methods as their intention is, usually, to reproduce the same conditions in the whole inspection process. However, in some restricted situations the second class could be a good solution and it will be explored in this work later on. The general approach in both cases is based on Grassman's Laws of additive colour mixture and they are the basis for most of these methods:

First law: any colour, c, can be matched by a linear combination of three primary colours if none of those three can be matched by a combination of the other two,

$$c = \alpha\mathcal{R} + \beta\mathcal{G} + \gamma\mathcal{B}$$

where $\mathcal{R}$,$\mathcal{G}$ and $\mathcal{B}$ are the primaries and α, β and γ are the amount of the respective stimulus to obtain c. $\mathcal{R}$,$\mathcal{G}$ and $\mathcal{B}$ do not stand for the usual *red*, *green* and *blue* camera system. They could be any whereas they fulfil the independence restriction.

Second law: when mixing two colours, c_1 and c_2, the result can be matched adding together the mixtures of the primaries that individually match the two initial colours,

$$c_3 = c_1 + c_2 \quad \wedge$$
$$c_1 = \alpha_1\mathcal{R} + \beta_1\mathcal{G} + \gamma_1\mathcal{B} \quad \wedge$$
$$c_2 = \alpha_2\mathcal{R} + \beta_2\mathcal{G} + \gamma_2\mathcal{B} \Longrightarrow c_3 = (\alpha_1 + \alpha_2)\mathcal{R} + (\beta_1 + \beta_2)\mathcal{G} + (\gamma_1 + \gamma_2)\mathcal{B}$$

This law holds for any number of colours.

Third law: the colour matching holds under changes on luminance conditions,

$$\lambda c_3 = \lambda c_1 + \lambda c2,$$

that is, when changing the illuminant, all colours will vary proportionally.

These laws are valid when working with an additive colour system, as it is the Human Vision System and colour cameras under some conditions. A necessary condition for a camera to hold the additive colour mixture properties is to disable any automatic settings. Specially gamma correction that introduces an exponential factor, which breaks the linearity properties of the model. The use of the Grassman model of colour mixing makes highly convenient the use of digital cameras against the analog ones. The main reason is that most of the frame–grabbers perform colour manipulation in digitalising the signal, as for example conversions to the PAL system, etc.

Since Grassman's Laws assure linear properties of colours, colour representation can be modeled by linear algebra. By these properties, given an acquired stimulus $\mathbf{s}^i$ under a certain illuminant, and the same stimulus under the illuminant taken as the canonical one $\mathbf{s}^c$, the transform can be written as

$$\mathbf{s}^c = \mathbf{G}^i \mathbf{s}^i, \tag{3.3}$$

where bold symbols denote vectors when lowercase and matrices when uppercase. Both stimulus are trichromatic stimulus and $\mathbf{G}^i$ is a full 3×3 matrix representing the linear transform between the canonical illuminant and the illuminant on the scene. Many of the methods we will review in this section simplify the use of a full matrix by a diagonal matrix. It is widely accepted that this assumption is enough for an approximate solution [38]. Thus the equation 3.3 does not hold the equality,

$$\mathbf{s}^c \approx \mathbf{D}^i \mathbf{s}^i. \tag{3.4}$$

When using the 3.3 equation model, methods will be called full linear transform models (FTM), and when follow equation 3.4 they will be called diagonal transform models (DTM). The diagonal model was first proposed by von Kries as a model for human adaptation. Although it had been some controversial discussion about its validity [114], it has been revisited and now is a widely accepted approach [38]. Most of the colour constancy methods can be viewed as reformulations of the von Kries model.

The computational approach to colour constancy is usually broken in two processes. The first one implies to estimate the illuminant information and the second process to use the precedent process to transform the response of the camera to independent illuminant descriptors. The methods differ on how the illuminant parameters are extracted and related to an independent illuminant representation. The following sections will enumerate and briefly describe some of these methods. This is not an exhaustive review of colour constancy and many other taxonomies can be done, and it is based on the work of Barnard [6] . The methods presented are those that we consider most representative or have been widely used in computer vision.

3.3   Direct transform based methods

In this section we will describe two methods of colour constancy that work on the ratios between the observed measures and the canonical descriptors. We call them direct transform because they are based on a simple processing step to infer the illuminant.

3.3.1   Grey world

This is considered the simplest approach to the colour constancy problem. It is based on a calculation of a single description for the whole scene. It assumes that lighting is uniform all over the scene and uses an statistical descriptor to discount the effect of the illuminant. It assumes a physical model where scenes in real world are grey in average, what is called the grey world assumption. From this point, the obvious statistic is the mean of the image for every channel as a descriptor of the illumination changes. That is, if there is a change in the colour of the light with respect to daylight (under it the scene should average to grey), the different channel means will be the correction ratio.

This definition does not take into account the luminance, as the grey could be thought to be from very dark to very light, being all of them different grades of grey. As an example, if we define our world to be an average grey, we could think the statistic descriptor as the response to a stimulus of 50% of a pure white. Using the diagonal model an (r, g, b) response will be transformed to $(r/2m_r, g/2m_g, b/2m_b)$ where m_x is the mean of channel x. The grey world assumption is very restrictive, even in the case that it holds for a given scene this does not guarantee that it holds for all regions of the scene. The method will act differently when applied to the entire scene or to its parts.

3.3.2   Retinex method

The main work of this method is presented by Land in [65]. It was initially conceived as a computational theory of human vision, but it has been applied on computer vision as well. The method assumes that slight spatial changes in the response are due to changes in the illumination or noise, whereas large changes correspond to surfaces changes. The idea is to run random paths from each pixel. When following the paths the ratio of the responses in each channel is computed. If it is near 1 then it is noise or light change and is set to 1, if not it remains as it is. The ratios are combined (multiplied) while the path is followed, obtaining at each step the percentage of light of the starting point that the current point has for a given channel. If the ratio is greater than 1 at a given point this point is taken as the start of the path, that is, the maximum luminance point is selected as reference. The ratios from different path are average and taken as a descriptor of the pixel. In this way a diagonal model is being assumed. Another approximation is to take the average of the ratios in the path. To simplify the process logarithms have been used reducing the problem to a differentiation to follow the path and integration to recover the descriptors. When considering a uniform illumination taking the maximum of the image is also called

the *white patch* algorithm or *white world* because it is assuming that the light colour descriptor is the maximum of its channel and the method will work if white is present in the scene. And, when taking the average it is equivalent to the *grey world*.

3.4   Gamut based methods

Another kind of methods are those based on the observation of the population of image pixels and their transformation to a plausible non illuminant dependent distribution. In other words, all the pixels values of an image at the same time are considered to be plausible only under a restricted set of illuminants. If the values of all possible surfaces are known for a specific illuminant, a suitable transform between the plausible illuminants to the canonical is calculated. The set of all possible tristimulus representations for a certain imaging system (camera, scanner, printer, monitor,...) is called its gamut. When we are acquiring an image its gamut is associated to the scene illuminant. For example, we will not get strong red response if the light is blue. The way gamuts are processed and the *guess* about the best transform is the difference between these methods.

3.4.1   3D gamut

Forsyth was the first author introducing a gamut based method in [40] , the idea behind his method is very intuitive. Once the canonical illuminant is fixed the set of all possible *rgb* observations is calculated, this will form the canonical gamut. The pixel values from an input image acquired under an unknown illuminant form an approximation of the gamut of this illuminant. Following the colour additivity law it is not possible to obtain a colour outside the convex hull of the gamut. It permits to simplify the complexity reducing the gamut to its 3D convex hull. Then all plausible mappings that make the unknown convex hull gamut polygon to lie inside the canonical are computed and considered. Although working with the convex hull instead of the complete hull reduces the algorithm complexity, the fact that the transform has 9 freedom degrees (it is an FTM) implies a large number of possibilities. The author reduces this complexity using a diagonal model. This variation is named CRULE algorithm. The method has an important weakness; it is based on the assumption that the light is constant all over the image. When light varies within the image the results are poor.

3.4.2   2D gamut

The weakness of the above method drove Finlayson to modify the CRULE algorithm [34]. The idea is to simplify the model transforming the 3D convex gamut to a 2D convex gamut. The way to do it is with a perspective transform. A point in the 3D space (r, g, b) is transformed to $(r', g', 1) = (r/b, g/b, b/b)$, and the third component is omitted. With this transform the new 2D space is independent of intensity changes of the light. He demonstrates that this representation can be used to apply the CRULE algorithm. From the set of all plausible transforms the one that maximises the colourfulness of the solution is selected. Other selections, like the mean illuminant

of all plausible illuminants are considered in other works. This leads to observe some incongruities in the selected illuminant.

In [32] this work has been revisited reversing the 2D gamut to a 3D gamut over the plane $b = 1$ and taking the mean illuminant and its corresponding transform in the 3D space.

3.4.3 Statistical gamut

The idea of plausible illuminants from the pixels in the image is in the core of a statistical approach of the gamut–based approach presented in [33]. It follows a voting mechanism filling a table with n columns indicating n illuminants and m rows, one for each descriptor. Each descriptor is a value in the two-dimensional chromaticity space. In short, a position in the matrix will be set to 1 if the respective chromaticity coordinate is plausible from the corresponding illuminant, it will be set to 0 otherwise. Then, each chromaticity point of the image will increase the associate counter of an illuminant if this observation is possible with the corresponding illuminant. The illuminant/s with larger number of votes are selected as the plausible illuminants of the scene.

3.5 Other methods

There are many methods that do not fit in the previous sections, and among them there is the well–known Maloney–Wandell algorithm. It is important for the way they approach the colour constancy problem, and it will be presented separately here for its elegant mathematical development.

Other methods that broach the problem from other perspectives are based on: Neural Networks [18], image specularity [28], bayesian approaches [14], illuminant spectra recovery [76], etc.

3.5.1 Maloney-Wandell algorithm

In the work by Maloney and Wandell [72] a linear method with rigourous mathematical posing was introduced. The main idea is to approximate reflectance and illuminant by a linear model of $n-1$ and n dimensions respectively, being n the number of sensors. The method search for a transform of the $n-1$–dimensional space of reflectance descriptors to the n–dimensional space of sensor responses. This transform will yield an hyperplane in the nD space passing through the origin whose orientation will describe the ambient light. The process is to derive the hyperplane from the responses of the image to estimate the vector that describes the illuminant and to calculate the inverse transform from nD to $n-1$D giving a set of surface descriptors. When applied to camera sensors, n is 3 and so the surfaces are described in a 2–dimensional space, which is not sufficient. Despite its elegance the method does not perform very well because of this strong assumptions in the model.

3.6   Invariance methods

In this section a set of invariant techniques are enumerated and briefly described. Their purpose is not to perform colour constancy but extract some properties in which intensity or colour of the light are not considered. Although most of the invariance methods are not colour constancy methods it is worthy to comment them.

These methods transform the input responses to a new representation in which some kind of invariance is achieved. The difference with the previous approaches is that their purpose is not to recover the image as seen under controlled circumstances but assure some useful features.

One of the simplest invariants is chromatic normalisation. Each pixel (r, g, b) of an image is normalised as: $(\frac{r}{r+g+b}, \frac{g}{r+g+b}, \frac{b}{r+g+b})$, which is an invariant representation to changes on the intensity of light.

Some other invariants had been proposed as for example the description of each pixel using a set of ratios between the signal in each channel and the signal in the respective channel of a set of neighbour pixels. This is useful when considering object recognition because its local invariance to changes on light properties. Another invariance used in image indexing is the angles defined by the covariance between channels of an image. The Frobenius distance is used to compare to distributions. It is also invariant to change on colour light [37].

Other works exists that consider invariance to changes on the colour of the light. One of them is to divide the pixel response in a channel by the mean of its channel on the image. In fact this is the same as the grey world method. So, it can be seen as invariant method or a canonical recovery method. This invariance and chromatic invariance has been merged in an iterative process in [31] and it would be used and analysed in this work.

The methods presented above are not a complete enumeration of colour constancy or invariance methods. We only have intended to give a brief review of the most important ones.

3.7   Adapting the camera system to VonKries theory

In section 3.3 we have introduced the use of the diagonal transform model instead of the full transform model for its simplicity. Although our aim is to use a DTM, for real applications where an accurate representation of colour can be important, we want to assure the colour precision. In order to do this and to avoid computing all the requirements for a FTM, we will transform our acquisition system to fully hold a DTM. This approach has been proposed by Finlayson in [35].

The starting point of this work is that for a DTM to suffice to identify the change of the illumination it has to fulfil

$$\frac{\int_w I^i(\lambda) S^q(\lambda) R^c(\lambda)}{\int_w I^i(\lambda) S^p(\lambda) R^c(\lambda)} = \frac{\int_w I^j(\lambda) S^q(\lambda) R^c(\lambda)}{\int_w I^j(\lambda) S^p(\lambda) R^c(\lambda)} \qquad \forall c = 1\dots 3. \qquad (3.5)$$

where I^i and I^j are two different illuminants, R^c is the sensor sensitivity of the c

channel, and S^q and S^p are two any surfaces. One way to guarantee that equation 3.5 holds is to construct sensors as narrow-band as possible. The ideal case is to be delta functions, ie: each sensor is only sensitive to one single wavelength. Instead of constructing narrow-band sensors, the camera sensor sensitivities can be narrowed by sharpening the spectral curve. That is the same that sharpening their responses. This is done by a linear transform that is independent of the illuminants. Therefore, equation 3.4 becomes

$$\mathcal{T}\mathbf{s}^c \approx \mathbf{D}^i \mathcal{T}\mathbf{s}^i \tag{3.6}$$

where $\mathcal{T}$ is the sharpening transform of the original sensor sensitivities. The method defined in the work minimises an expression over a set of three fixed wavelength intervals. These intervals have to be set a priori. The problem is how to fix them since they will vary with each different camera. The advantage of the method used is that it does not take into account the illuminants and surfaces, only the sensor sensitivities, and hence it is not data dependent. In his analysis [35], a comparison between the sensor–based sharpening and the data–based sharpening is done. The conclusion is that the results from both approaches are nearly identical. From this conclusion we decide to obtain the sharpening transform using the data-based method. In this way, no guess should be done. The process starts from the observation of samples viewed under two different illuminants. One of them is taken as the canonical illuminant, the images will be transformed as they would be seen under it. Two $3 \times n$ matrices are constructed, $\mathbf{S}^c$ and $\mathbf{S}^i$. The first one represents the sensor responses of n samples under the canonical illuminant, and the second one the responses of the same samples under another illuminant. Then, using equation 3.6 we have

$$\mathcal{T}^i\mathbf{S}^c = \mathbf{D}^i\mathcal{T}^i\mathbf{S}^i \tag{3.7}$$

The equality is true if $\mathbf{D}^i$ is considered to be the least-square solution, which is obtained by $\mathbf{D}^i = \mathcal{T}^i\mathbf{S}^c[\mathcal{T}^i\mathbf{S}^i]^+$ where A^+ is the pseudo-inverse Moore-Penrose inverse $(A^+ = A'[AA']^{-1})$. Developing this expression yields to

$$\mathbf{S}^c\mathbf{S}^i = (\mathcal{T}^i)^{-1}\mathbf{D}^i\mathcal{T}^i. \tag{3.8}$$

Since the eigenvector decomposition of equation 3.8 is $\mathbf{S}^c\mathbf{S}^i = \mathbf{U}\mathbf{D}\mathbf{U}^{-1}$ and as $\mathbf{D}^i$ is diagonal then $\mathcal{T}^i = \mathbf{U}^{-1}$.

The task to do is to acquire all the samples under the two illuminants of interest. The number of samples used in [38] was 462 from the Munsell set of colours. This is a very tedious task, and needs of the Munsell charts. Another way to deal with data-based sharpening is to do all the development synthetically. The data of the Munsell spectra and the illuminants spectra are known. To apply equation 3.2 the camera sensitivities must be known. Supposing that this is the case, applying the sharpening will be straightforward.

For most of the cases the dealer does not supply sensor sensitivity spectra. Although it was the case, the system optics will change the sensor specifications. This implies that a method to recover the sensor spectral properties is needed.
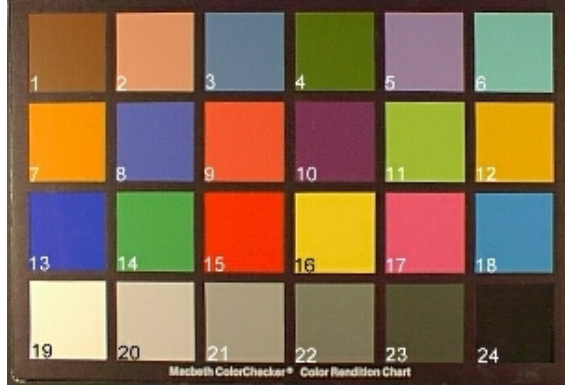
Figure 3.2: Macbeth Color Checker Chart used in sensor recovering.

3.7.1 Sensor sensitivity recovering

Fortunately there are various possibilities to do it, a short review can be found in [6]. The most reliable is the use of monochromators. Such devices can emit light on a very narrow interval of the visible spectrum. Illuminating a white surface and measuring the camera response the sensitivities can be recovered. This approach was used by Vora et al in [107, 106],where they recover the sensors of two digital cameras and prove their linearity. This method is very accurate when is done carefully. The problem is that devices capable of generating narrow band light at the desired intervals are expensive and not readily available. Therefore various authors have attempted to solve this problem without using this equipment [93, 67, 57, 36, 5]. The starting point is equation 3.2, but as it is a computational approach based on measured data we need to rewrite it in the discrete domain as,

$$\rho_x^k = \sum_{w=1}^{W} I(\lambda_w) S_x(\lambda_w) R^k(\lambda_w), \qquad k = 1, 2, 3 \tag{3.9}$$

that implies to know the information for W wavelengths of the visible spectrum. Usually, it is enough with $W = 31$. Equation 3.9 in vector form will be

$$\rho_x^k = (\vec{L}_x)' \vec{R}^k \qquad k = 1, 2, 3 \tag{3.10}$$

where $L_x(w)$ is the energy emitted by the pixel x at the w-th wavelength.

The general idea is to measure a number of input spectra and its camera response for each sensor from a set of samples. If $\boldsymbol{\rho}^k$ is the responses vector of the k-th sensor for m surfaces and $\mathbf{L}$ an $m \times W$ matrix where each row is the spectral response of the respective stimulus then the problem reduces to find the spectral sensitivity of the sensor as the vector $\mathbf{R}^k$,

$$\boldsymbol{\rho}^k = \mathbf{L}\mathbf{R}^k \tag{3.11}$$

which can be solved by minimising the RMS error, $\|\boldsymbol{\rho}^k - \mathbf{L}\hat{\mathbf{R}}^k\|_2$. But if m is large then the method will be analogous to the monochromator method. The intention is
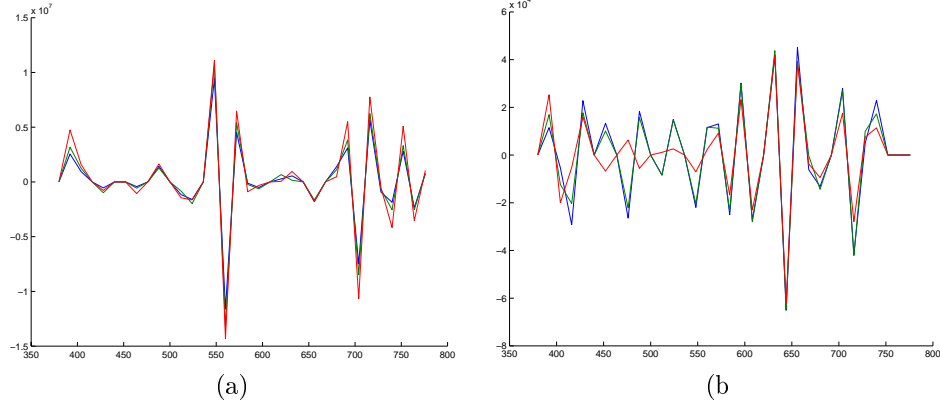
Figure 3.3: Recovered camera sensitivities without restrictions. (a) are the results for a 3CCD Sony XC–003P camera and (b) are the sensor sensitivities for a 12 bits line scan TVI camera.

to make the process as simple as possible and use as few samples as possible. That yields **L** to be rank–deficient because of its dimensionality. The most used samples are those from a *Macbeth Color Checker Chart* in figure 3.2 [77]. It consists of 24 colour patches representing 18 natural colours an 6 achromatic stimuli. The test presented here was done on a digital TVI line scan camera and a 3 CCD Sony XC-003P. The light used is irrelevant because the data used includes it. The spectral measures were collected using a PhotoResearch PR-650 spectroradiometer. If we try to recover the sensitivities using the direct approach the results are those in figure 3.3, where it is plotted the spectra for all sensors of the Sony and the TVI camera respectively. Is it obvious that there not exists a camera with such sensors. The expression minimised is the RMS error of the following set of linear equations:

$$
\begin{pmatrix} \rho_1^k \\ \rho_2^k \\ \vdots \\ \rho_{24}^k \end{pmatrix} = \begin{pmatrix} I(\lambda_1)S_1(\lambda_1) & I(\lambda_2)S_1(\lambda_2)\cdots I(\lambda_W)S_1(\lambda_W) \\ I(\lambda_1)S_2(\lambda_1) & I(\lambda_2)S_2(\lambda_2)\cdots I(\lambda_W)S_2(\lambda_W) \\ \vdots & \vdots \quad \cdots \quad \vdots \\ I(\lambda_1)S_{24}(\lambda_1) & I(\lambda_2)S_{24}(\lambda_2)\cdots I(\lambda_W)S_{24}(\lambda_W) \end{pmatrix} \begin{pmatrix} \hat{R}^k(\lambda_1) \\ \hat{R}^k(\lambda_2) \\ \vdots \\ \hat{R}^k(\lambda_K) \end{pmatrix}
$$

(3.12)

As it becomes under–determined ($W > 24$) the best fitting can be reached by a wide range of configurations. What the methods do is to impose some constraints on the solution $\hat{\mathbf{R}}^k$ of equations 3.12, in order to improve the solution to a more realistic one. In our work we have used the constraints imposed by by Finalyson et al. in [36], for being one of the last works on this subject when the problem was set up, and for its simplicity.

The simplicity of this method relies on the fact that all the restrictions made can be included in the minimisation problem using quadratic programming which solves the equation 3.12 subject to a set of q linear constraints:

$$a_{1,1}R^i(\lambda_1) + \cdots + a_{1,K}R^i(\lambda_W) \leq b_1$$
$$a_{2,1}R^i(\lambda_1) + \cdots + a_{2,K}R^i(\lambda_W) \leq b_2$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$a_{q,1}R^i(\lambda_1) + \cdots + a_{q,K}R^i(\lambda_W) \leq b_q$$

The different constraints imposed are: positivity, modality and smoothness. Positivity states that it is not feasible to get a sensor with negative response. The behaviour of sensors is always additive, never subtractive. Modality comes from the fact that most of the sensors tend to be unimodal, i.e.: there is a unique peak in the spectral response. And smoothness from the fact that there are no abrupt transitions in the spectrum as is the case in figure 3.3. In the following lines we summarise the constraints introduced in each case:

Positivity:

$$R^k(\lambda_w) \geq 0 \qquad w = 1, \dots, W \tag{3.13}$$

Modality: If we want the peak of the sensitivity to be in the v–th wavelength then the constraints are:

$$
\begin{aligned}
R^k(\lambda_{v+1}) &\geq R^k(\lambda_v) & v = 1, \dots, w-1 \quad (3.14)\\
R^k(\lambda_{v+1}) &\leq R^k(\lambda_v) & v = w, \dots, W-1 \quad (3.15)
\end{aligned}
$$

These constraints should be tested for all plausible v's. In our experiments we divide the range of spectral response in as many subranges as number of sensors. Each sensor is imposed to be unimodal with the peak in one subrange. That is, we approach the k sensor W/p times (p is the number of sensors) with different locations of the unimodal peak. From these W/p approaches, the one with minimum error is selected.

Smoothness: To achieve the smooth sensitivity curves Finlayson proposes to approximate the sensitivities to a set of Fourier basis:

$$\mathbf{B}_1 = c, \quad \mathbf{B}_2 = sin(x), \quad \mathbf{B}_3 = cos(x), \quad \mathbf{B}4 = sin(2x), \quad \mathbf{B}_5 = cos(2x), \cdots$$

and combining them following the equation:

$$\mathbf{R}^i = \sigma_1 \mathbf{B}_1 + \sigma_2 \mathbf{B}_2 + \dots + \sigma_l \mathbf{B}_l$$

taking c an arbitrary constant (i.e: $c = 1$) and $x = (\lambda - \lambda_{min})\pi/((\lambda_{max} - \lambda_{min})/2)$, for $\lambda = \lambda_{min} \cdots \lambda_{max}$. Then the problem is to find the vector $\boldsymbol{\sigma}^k$ that minimises

$$\|\mathbf{LB}\boldsymbol{\sigma}^k - \boldsymbol{\rho}^k\|^2$$

Figure 3.4: Recovered camera sensitivities with positivity, modality and smoothness constraints. (a) are the results for a 3CCD Sony XC–003P camera and (b) are the sensor sensitivities for a 12 bits line scan TVI camera.

the matrix equation involved in this minimisation and equivalent under this formulation to equation 3.3 is:

$$
\underbrace{\begin{pmatrix} \rho_1^k \\ \rho_2^k \\ \vdots \\ \rho_{24}^k \end{pmatrix}}_{\boldsymbol{\rho}^k} = \mathbf{L} \underbrace{\begin{pmatrix} B_1(\lambda_1) & B_2(\lambda_1) \cdots B_l(\lambda_1) \\ B_1(\lambda_2) & B_2(\lambda_2) \cdots B_l(\lambda_2) \\ \vdots & \vdots \cdots \vdots \\ B_1(\lambda_W) & B_2(\lambda_W) \cdots B_l(\lambda_W) \end{pmatrix}}_{\mathbf{B}} \underbrace{\begin{pmatrix} \sigma_1^k \\ \sigma_2^k \\ \vdots \\ \sigma_l^k \end{pmatrix}}_{\boldsymbol{\sigma}^k}
$$

To impose positivity and unimodality to this approach the constrains have to be reformulated, and we can write

$$ R^k(\lambda_v) = \sigma_1 B_1(\lambda_v) + \sigma_2 B_2(\lambda_v) + \ldots + \sigma_l B_l(\lambda_v) $$

it follows that the unimodality constraints 3.14 and 3.15 are, repectively:

$$ \sigma_1(B_1(\lambda_v) - B_1(\lambda_{v+1})) + \ldots + \sigma_l(B_l(\lambda_v) - B_l(\lambda_{v+1})) \leq 0 \quad v = 1, \ldots, w-1 $$
$$ \sigma_1(B_1(\lambda_{v+1}) - B_1(\lambda_v)) + \ldots + \sigma_l(B_l(\lambda_{v+1}) - B_l(\lambda_v)) \leq 0 \quad v = w, \ldots, W-1 $$

and the same for positivity:

$$ R^k(\lambda_v) \geq 0 \equiv -\sigma_1 B_1(\lambda_v) - \sigma_2 B_2(\lambda_v) - \ldots - \sigma_l B_l(\lambda_v) \leq 0 \qquad v = 1, \ldots, W $$

With this method the sensor of two cameras were recovered. The spectral sensitivities are shown in figure 3.4, at the left is the analog 3CCD matrix camera, and at the right the digital line scan camera. In the case of the matrix camera we do not have enough information and we can not validate the results. We do not know the theoretic spectral distribution of the TVI camera but we know the spectral transmittance of the prism used to split the light to the sensors. The results obtained are

Table 3.1: The relative error on recovering the sensors of two cameras: Sony XC-003P and TVI line scan camera.

Mean relative error of samples			
Camera	red sensor	green sensor	blue sensor
TVI	0.0244	0.0271	0.0310
Sony	0.0570	0.0581	0.0580

congruent with them. Moreover, to test the level of error we get the mean relative error for each sensor k,

$$\frac{\sum_{j=1}^{n} \left(\frac{\mathbf{L}_j \mathbf{R}^k - \rho^k(j)}{\rho^k(j)} \right)}{n} \tag{3.16}$$

where $\mathbf{L}_j$ denotes the j–th row of the matrix $\mathbf{L}$ and n the number of samples used. The results are presented in table 3.1 for both cameras. The error we have in the worst case is of 3% in the line scan camera and 5.8% for the matrix camera. Although it is not perfect, they are quite good results.

3.8   Taking Spectral Sharpening into practice

To take spectral sharpening into practice without having to acquire 462 samples (as it is suggested in [35]) the solution is to simulate the process. From equation 3.2 we need to know the spectral power distribution of the light, the spectral reflectance of surfaces and the spectral sensitivities of sensors. Now, we know all of them. Surfaces and light are tabulated and sensors are just recovered. It is immediate to apply the above equation 3.8 giving us a matrix $\mathbf{S}^i$ for any known illuminant i. Because data–based sharpening is conceived as a method of validation of the sensor–based sharpening some extra considerations have to be done. The method is applied from a canonical illuminant against another one. To test its validity it should be done for various pairs of illuminants maintaining the canonical one. When it is done the resulting transforms, although they are nearly identical they are not the same and some kind of fusion among them is needed. Let us examine this last point a little bit later.

The process will be done only for the line scan camera since the matrix camera is not suitable for accurate colour inspection as explained in section 2.4. The set of samples are the 1269 ones from the *Munsell Book of Color* matte samples [22]. The illuminants used are the CIE standards: A, B, C, D55, D65 and D75. Illuminant A relates to a tungsten lamp at 2856°K and the others to various approximations of the daylight [117]. We took illuminant A as the canonic and we calculated the inverse of the eigenvectors for every possible pair $\mathbf{S}^A \mathbf{S}^X$ where $X = B, C, D55, D65, D75$. All transform matrices are compiled in table 3.2. As mentioned previously the results are nearly identical. However, we need to extract only one transform. The spectral sharpening algorithm does not take the data-based sharpening as the way to obtain

Table 3.2: The 3×3 sharpening transforms for the TVI line scan camera and for all the considered illuminants, taking A as the canonic.

A vs B
$$\begin{pmatrix} 1.014 & 0.138 & 0.020 \\ 0.081 & 1.030 & 0.048 \\ -0.033 & 0.096 & 1.005 \end{pmatrix}$$
A vs C
$$\begin{pmatrix} 1.015 & 0.129 & 0.016 \\ 0.085 & 1.029 & 0.041 \\ -0.029 & 0.105 & 1.005 \end{pmatrix}$$

A vs D55
$$\begin{pmatrix} 1.014 & 0.124 & 0.019 \\ 0.084 & 1.028 & 0.052 \\ -0.032 & 0.098 & 1.006 \end{pmatrix}$$
A vs D65
$$\begin{pmatrix} 1.014 & 0.112 & 0.016 \\ 0.087 & 1.028 & 0.047 \\ -0.029 & 0.106 & 1.006 \end{pmatrix}$$

A vs D75
$$\begin{pmatrix} 1.014 & 0.117 & 0.015 \\ 0.088 & 1.028 & 0.042 \\ -0.027 & 0.112 & 1.005 \end{pmatrix}$$
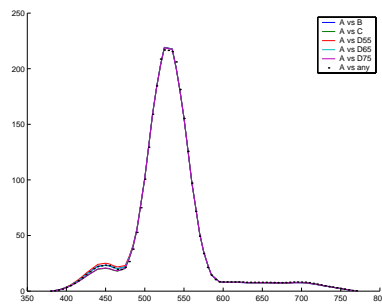
Figure 3.5: Modified camera sensor sensitivity to improve DTM. All transform are similar but not identical. The dotted line is the best fit to all transforms..

it. Our proposal is to extract the transform that best fits all the above ones. To obtain the new spectral sensitivities, given an illuminant X, we only need to apply the linear transform $\boldsymbol{\mathcal{T}}^X$ to the sensor sensitivities as in equation 3.6,

$$(\boldsymbol{\mathcal{R}}^X)' = \boldsymbol{\mathcal{T}}^X \mathbf{R}'$$

where the k-th column of $\mathbf{R}$ is the sensitivity of k-th sensor, and $\boldsymbol{\mathcal{R}}^X$ the new spectral sensitivities. As an example figure 3.5 shows the results of the reconstruction of the middle spectrum channel. The coloured solid lines are the applied transforms. Now, to obtain the final transform we will minimise the RMS error of the needed transform from the original sensitivities to the mean taken at each wavelength $\overline{\boldsymbol{\mathcal{R}}}$, i.e: to minimise

$$\|\overline{\boldsymbol{\mathcal{R}}} - \mathbf{R}\boldsymbol{\mathcal{T}}\|^2$$

At the end the resulting transform is:

$$\boldsymbol{\mathcal{T}} = \begin{pmatrix} 1.012 & 0.123 & 0.017 \\ 0.092 & 1.018 & 0.047 \\ -0.031 & 0.109 & 0.996 \end{pmatrix} \tag{3.17}$$

We will discuss the effect of the sharpening transform in the final system at the end of the next section.

3.9 Colour constancy for on–line inspection

In chapter 2 we defined a hardware system for accurate colour vision inspection, but some computational efforts should be dedicated to assure stable colour acquisition. It is specially important when the underlying application relies on past measures. This is the case of many industrial processes that maintain a catalog (or they should do) of their production and the current output is related to it. If the system can not reproduce the initial conditions then this reference to the past is not possible. There are several factors that can make acquisition to differ from time to time. As an example: the aging of the filters (if used), the soiling of the optics and lighting system, small changes on the relative positioning of the triad sample, light and sensor, etc. All of them can be solved with a periodical maintenance protocol that any industrial computer vision application must define. Apart from these, there are some troublesomeness when handling light stability, and they can not be settled by a human operator:

Non-homogeneous spatial illumination: Due to the use of non homogeneous filters (when needed) or to optic effects the response on the image varies between locations. The optics introduce an attenuation of the signal when moving around the image due to geometric effects, principally vignetting and a fall-off proportional to the fourth power of the cosine of the off axis angle [56]. The vignetting effect is stressed when working with high apertures. These effects will always appear when considering a computer vision system, with the consequent outcome to make difficult to get an homogeneous illumination through the acquired
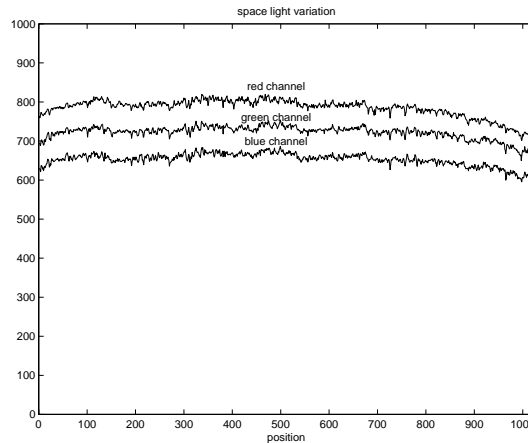
Figure 3.6: Light spatial variation: it shows the fall–off at the edges of the image due to the cosine–4th law.

image. As we are centred on the use of line scan cameras (see section 2.4) the effect is the same for all the lines of the image, because it is formed joining lines coming from a single line CCD sensor. An example of a particular configuration of this system and to show the spatial inhomogeneity figure 3.6 plots the intensity profile of a random line from an constant colour surface. Under this circumstance no comparison can be made between different regions of the image, thus testing spatial coherence or methods that apply all over the image to extract information are not possible, or at least very weak.

Time varying illumination: This is a normal problem on any acquisition system where a high degree of stability is required in order to do a colour based inspection. The aging of lamps changes the equivalent colour temperature, and so do the acquired images. Under this circumstance absolute colour measures (and in some degree illuminance measures) are not reliable if it is not corrected.

Both cases differ from its causes and its consequences, but both of them can be analysed from colour constancy. In the first case we want all positions of the image to be referenced to a known canonical light. In the second case this is also true but at any time. Our interest is to achieve, at the same time, spatial and time stability.

We will start from the results in section 3.8 applying the matrix 3.17 to all the image pixels. In this way we can rely on algorithms performing a DTM.

3.9.1   Merging spatial and temporal colour constancy

In the course of our work different colour problems have come to us. All of them needed to treat the colour constancy problem and different approaches have been developed that helped us to deal with it.

Firstly we will explain the way we have broached the problem with a surface inspection problem in mind. Afterwards we present what it was the firs attempt to

get an illuminant invariant representation of the image. Although it is not useful for generic surface inspection it demonstrates good capabilities to solve the variability conditions when only chromatic segmentation is needed.

Diagonal transform approach

We will now assume that diagonal transform model suffices to cope with the problem of colour constancy. Afterward, we will add the sharpening transform we have computed in the previous section. Then for the moment we will assume the use of a DTM as being absolutely reliable.

The diagonal matrix has been computed by using a constant colour pattern sample, $\mathcal{C}$ that is acquired periodically. This forms a set $\mathbf{w}^t$ of reference images, where t stands for the time they are acquired.

As we have mentioned above, we need to compute space and time corrections. We will do these two corrections in separate steps:

1. The first step is to correct the spatial distortions of $\mathcal{C}$. The distortions, due mainly to optic effects and uneven line light, can be modeled by a set of diagonal transforms $\{\mathbf{S}_x\}$, that is, one diagonal 3×3 matrix for each x position along the x axis, where the spatial variation occurs. For each triad of photo–sensors (photogate) of the CCD we will calculate the corresponding DTM.

2. The second step is to correct light variations due to time. They will be corrected in a similar way. We calculate another set of diagonal transforms, $\{\mathbf{T}_x^{t_i}\}$, which models the changes at time t_i with respect to instant t_0, resulting in a temporal DTM.

Now, we can attack the problem separately, i.e: to define $\{\mathbf{S}_x\}$ and $\{\mathbf{T}_x^{t_i}\}$, and then we will merge both sets in a single set of DTM.

In order to extract the diagonal transforms $\{\mathbf{S}_x\}$, we fix a canonical colour descriptor of $\mathcal{C}$, $\mathbf{w}^c$, which is the rgb vector that will represent the canonical colour of the reference pattern, $\mathcal{C}$. Its value will be derived from the first acquired reference pattern at time t_0, denoted as $\mathbf{w}^{t_0}$. We want to transform each triplet $\mathbf{w}_x^{t_0}$ to the canonical descriptor, and use this transform for the subsequent images.

Changes in colour representation can be due to intensity or chromatic changes on the illuminant. The former is constant for all channels and the latter can vary for each channel. In a line scan camera all the pixels on the same column came from the same spatial position, that is all of them have the same illumination conditions. For a given column x, these conditions are defined by a constant s_x representing intensity changes and a matrix, $\mathbf{C}_x$, for the chromatic changes.

Considering this separate model of illuminant changes we can define $\mathbf{w}^c$ as follows:

$$\mathbf{w}^c = s_x^{t_0}\,\mathbf{C}_x^{t_0}\,\mathbf{w}_x^{t_0}\,,$$

where $\mathbf{w}_x^{t_0}$ is the *rgb* vector at position x of the white reference image, as well as, s_x and $\mathbf{C}_x^{t_0}$ represent the shading factor and the diagonal light colour transform, respectively, for the x image position. We have to recall we are assuming a specific transformation for each image column, given that we are working with a 3CCD line

scan sensor. As we do not need to know s_x and $\mathbf{C}_x^{t_0}$ separately we can rewrite the expression as:

$$\mathbf{w}^c = \mathbf{S}_x \mathbf{w}_x^{t_0}$$

and, from it, it follows that:

$$(\mathbf{S}_x)_{kk} = \frac{(\mathbf{w}^c)_k}{(\mathbf{w}_x^{t_0})_k} \quad \forall k = 1 \ldots 3 \, , \tag{3.18}$$

where $(\mathbf{S}_x)_{kk}$ is the lighting and colour correcting factor for the sensor k and $(\mathbf{w}_x^{t_0})_k$ is the k channel value of the x pixel on the initial reference image at position x. At this point, $\mathbf{w}^c$ has not been defined yet. As our objective is to have all images in terms of a canonical descriptor, and which is the descriptor is not relevant, we choose it as the mean value on each sensor:

$$(\mathbf{w}^c)_k = \frac{1}{N} \sum_{x=1}^{N} (\mathbf{w}_x^{t_0})_k \, . \tag{3.19}$$

Substituting equation 3.19 in 3.18, we obtain the set of spatial DTM $\{\mathbf{S}_x\}$. The following step is to compute the diagonal transforms $\{\mathbf{T}_x^{t_i}\}$. Doing the same reasoning as in the previous step, that is, assuming the same lighting model we can write

$$\mathbf{w}_x^c = s_x^{t_i} \mathbf{C}_x^{t_i} \mathbf{w}_x^{t_i} \, .$$

In this case the descriptors we want to refer to are the *rgb* values of the reference pattern $\mathcal{C}$ at time t_0. We are transforming the outputs of the camera to those that would be obtained at a reference time. As we are not doing spatial correction, the transform is applied to each individual element of the reference array. Because $\mathbf{w}^{t_0}$ is the first known output from the camera, it will be taken as the reference time. We can rewrite the expression in a compact style as:

$$\mathbf{w}_x^{t_0} = \mathbf{T}_x^{t_i} \mathbf{w}_x^{t_i} \, ,$$

and for each channel, k, of the image we have

$$(\mathbf{T}_x^{t_i})_{kk} = \frac{(\mathbf{w}_x^{t_0})_k}{(\mathbf{w}_x^{t_i})_k} \, .$$

Now what remains is to extract a single set of diagonal transforms from the spatial set and the temporal set, $\mathbf{S}_x$ and $\mathbf{T}_x^{t_i}$ respectively. The process is depicted in the following schema

$$
\begin{array}{ccc}
 & \mathbf{S}_x & \\
\mathbf{w}^c & \longleftarrow & \mathbf{w}_x^{t_0} \\
\mathbf{S}_x \mathbf{T}_x^{t_i} \nwarrow & & \uparrow \mathbf{T}_x^{t_i} \\
 & & \mathbf{w}_x^{t_i}
\end{array}
$$

Finally, what we need to do is to combine both transformations on the acquired image. Given the *rgb* vector $\mathbf{p}_x$ from the position x of an image taken at an instant t where

(a)

(b)
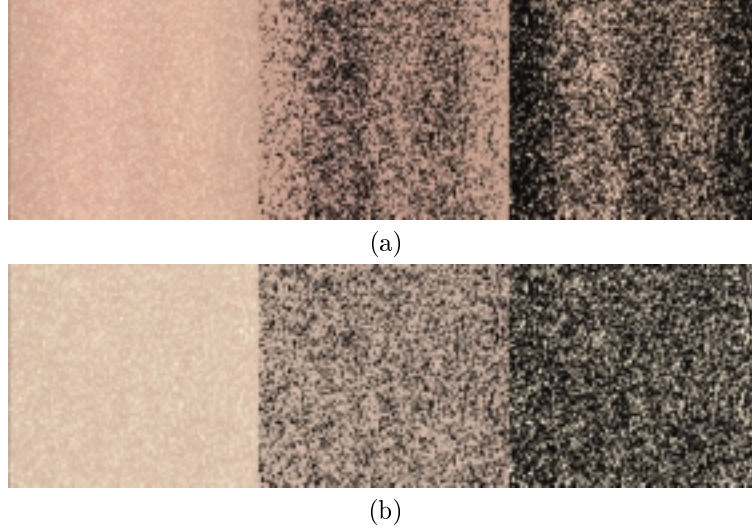
Figure 3.7: Test of the DTM colour correction approach. (a) are the original image, the cluster image of pixels belonging to class 1, and the cluster image of pixels belonging to class 2 when no correction is performed. (b) is the same configuration but using the spatio–temporal DTM.

$t_i < t < t_j$, being t_i and t_j the times when two consecutive white reference images have been taken, we can express the canonical descriptor of $\mathbf{p}_x$ as

$$\mathbf{p}_x^c = \mathbf{S}_x \mathbf{T}_x^{t_i} \mathbf{p}_x \,, \qquad (3.20)$$

being $\mathbf{p}_x^c$ the descriptor vector of $\mathbf{p}_x$ that is illuminant independent. The final set of diagonal transforms $\{\mathbf{D}_x^{t_i}\}$ are $\mathbf{D}_x^{t_i} = \mathbf{S}_x \mathbf{T}_x^{t_i}$, thus

$$(\mathbf{D}_x^{t_i})_{kk} = \frac{(\mathbf{w}^c)_k}{(\mathbf{w}_x^{t_0})_k} \frac{(\mathbf{w}_x^{t_0})_k}{(\mathbf{w}_x^{t_i})_k} = \frac{(\mathbf{w}^c)_k}{(\mathbf{w}_x^{t_i})_k} \,. \qquad (3.21)$$

As a test we used a sample of a ceramic tile with a random isotropic texture. Once the correction is done we apply a *k–means* clustering algorithm with $k = 2$ to get two different images. Briefly, this algorithm groups pixels by its colour similarity, it will be explained later in section 5.2. Since the texture is randomly distributed it is supposed to have the same amount of a certain colour in each column of the image. If the colour correction is correct the segmented images should be spatially homogeneous. The results on figure 3.7 confirm that point. The top row of the figure presents the origianl image without any correction, and the corresponding segments of two different colours. In both segments we can appreciate the effects of the spatial non–homogeneity on the segmentation. These images are very conclusive, but it can be seen more intuitively in figure 3.8. Each graphic plots the percentage of pixels in the respective column belonging to the first or second segment. Because of the properties of the texture, all columns have similar amounts of particles of each colour. Thus,

(a)   (b)

Figure 3.8: DTM colour correction. The image is segmented in two clusters and the cumulative profile for each cluster is computed ((a) and (b)). The red lines are the results for the uncorrected image and the black line are the colour corrected results.

the resulting profile should be almost flat except for small variations. If this profile is not nearly constant is due to non uniformity of the light and optic system, which correspond to different colours along the line sensor. In (a) and (b) the results of the cumulative profiles for both segments respectively are plotted when applying de DTM approach in black, and without it in red. It is clear that the process is a must. Temporal stability is also tested and the profiles obtained are practically the same.

Up to the moment no sharpening transform has been applied. The last step is to extent the colour correction to the sharpened responses of the camera as it has been explained in section 3.8. To do this the rgb outputs from the camera are multiplied by the transform $\mathcal{T}$ in equation 3.17.

We have measured the difference between the colour correction with and without this sensor sharpening for a line scan camera. The error between both transforms is

$$\mathcal{E}^k = \mid \mathbf{I}_D^k - \mathbf{I}_D T^k \mid$$

where k is the channel being analysed, $\mathbf{I}_D^k$ is the corrected image using the diagonal transform model, $\mathbf{I}_D T^k$ denotes the correction applying *Spectral Sharpening*. The $\mathcal{E}^k$ means were 0.13%, 0.11% and 0.06% in the red, green and blue channels respectively for a set of 274 images with different colour distributions. The changes in illuminant are not dramatic but are the real conditions in a industrial inspection problem. The standard deviations were 0.01%, 0.009% and 0.005% that means that the obtained coefficients are very stable. These small differences evidence the good properties for colour constancy of the camera sensitivities, which are quite narrow band.

Colour normalisation approach

This approach is based on the change of representation of the colour space, eliminating the information that is not referring to intensity and colour of light. This is a

(a)

(b)
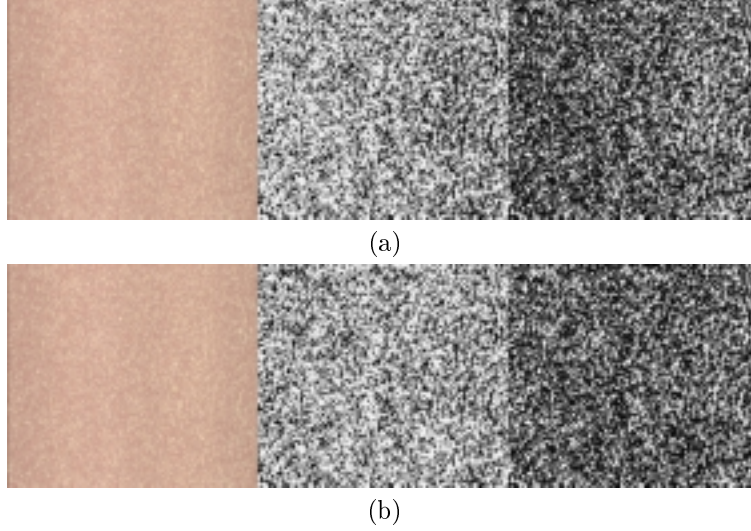
Figure 3.9: Test of the column comprehensive normalisation. (a) are the original image, the cluster image of pixels belonging to class 1, and the cluster image of pixels belonging to class 2 when applying the original method. (b) is the same configuration but using the modified version. The second case has no agglomerative areas.

work based on the *Comprehensive Normalisation* of Finlayson presented in [31]. The comprehensive colour normalisation is an iterative algorithm, which tries to remove shading and light colour, successively. One of the assumption of this normalisation is that light colour is constant all over the scene, thus, given an image of $N \times M$ pixels, represented as a $NM \times 3$ matrix, $\mathbf{I}$, where rows are *rgb* values of a pixel, the normalisation is computed by considering an iterative process

$$\mathbf{I}'_{t+1} = \mathbf{D}^s \mathbf{I}'_t \mathbf{D}^c \, , \tag{3.22}$$

where $\mathbf{D}^s$ is a $NM \times NM$ diagonal matrix that represents lighting geometry of the image, and $\mathbf{D}^c$ is a 3×3 diagonal matrix assuming a diagonal model for the colour of the illuminant. The iterative normalisation tries to remove the factors introduced by $\mathbf{D}^s$ and $\mathbf{D}^c$, by transforming the image to its chromatic coordinates, and fixing the magnitude of the image channels, respectively.

Given that, we can not assume that the colour light is exactly the same all over the image, we have introduced a modified version of this algorithm. As we have already commented, the $N \times M$ images from a line scan camera are formed from a 3CCD sensor array of length M, then all pixels of the same column come from the same sensor and the same point light source. We made use of this fact to modify the algorithm. Instead of normalising the entire image, it is split into M sub–images, one for each column. These images are separately normalised and merged again.

This column-comprehensive normalisation allows avoiding spatial illuminant variations that can be important to get a good starting point for an inspection system.

(a)   (b)

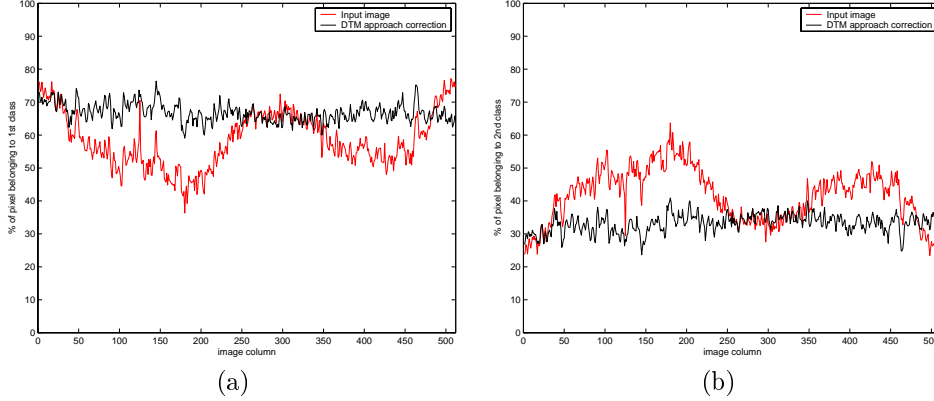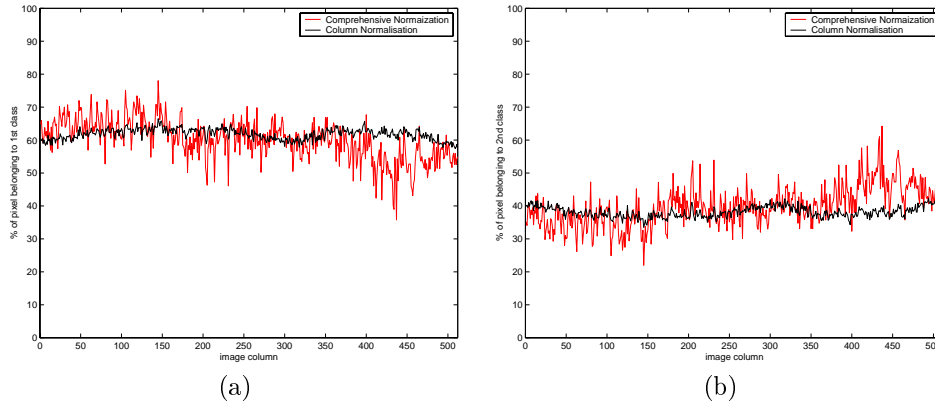**Figure 3.10:** Colour normalisation correction. The image is segmented in two clusters and the cumulative profile for each cluster is computed ((a) and (b)). The red lines are the results for the *Comprehensive Normalisation* and the black line are the modified version of the colour normalisation.

However, it presents some problems on specific applications where the lightness is an important cue for the inspection. Comprehensive colour normalisation removes lightness reducing the image to its chromatic information. Those applications that are concerned to relative chromatic content are candidates to use this approximation. A variation of this approach has been used to calibrate colour acquisition in a real inspection environment [104, 9, 8]. But, in most cases this is not enough, and lightness should be corrected together with chromatic information.

We used the same test to compare the results from the original *Comprehensive Normalisation* and the modification that we suggest. Figure 3.9(a) is the result of applying this process with the original comprehensive normalisation, at the left the original image and at the middle and right the masks of the cluster images. The real values are not shown because the nature of the normalisation is not intended to give a visually interpretable space. In figure 3.9(b) the same process is applied with the modified version of the colour normalisation. It can be perceived that the second case results in a more distributed segmentation.

Figure 3.10 shows the results for the two segments of the example. It is the comparison between colour normalisation and the proposed variation. Although the images depicted in fig. 3.9 are visually very similar, there exist spatial variation when using the original form of the *Comprenhensive Normalization*. The column normalisation comes to be very stable.

3.10 Discussion

We have made a concise introduction to computational colour constancy methods, with the basis of colour formation that explains most of the algorithms intended for this purpose.

An obvious and important conclusion of this chapter is that one of the key factors to succeed in colour inspection based on computer vision techniques, is the need of some kind of colour correction.

We have defined a colour constancy method adapted to a line scan camera. The method computes a linear transform that combines both spatial and temporal colour variations. It is based on a diagonal model improved by a linear sensor modification. When applying sensor sharpening to the line scan camera used in this work, we realised that it presents very good sharpening properties on its sensors.

As a lateral contribution, we have also modified an existing method of chromatic invariance that could work in industrial vision and treats both spatial and temporal variations at the same time.

Chapter 4

Computational operators for colour texture perception

Most of the previous works dealing with computational representations for colour texture have been directed to extend gray level representations to every one of the RGB channels. As we have already seen in the introduction chapter, to deal with colour texture we need operators that combine co–jointly the spatial and the colour information in a way that simulates the especial behaviour demonstrated by the human visual system. In this chapter we will analyse colour induction as the most important phenomena that acts on the colour texture perception, and we will propose a computational operator for a perceptual sharpening that allows to complement previous results on perceptual blurring, providing both a general model for colour induction, the first one in chromatic contrast and the last one in chromatic assimilation.

4.1 Colour Induction

Colour induction is a colour phenomena that changes the colour appearance of a stimulus due to the influence of the scene contents in the field of view. In this category we have to include the colour adaptation phenomena introduced in chapter 3, which is always involved in any scene interpretation. Adaptation models or colour constancy methods usually are global visual mechanisms.

In this section we will deal with other induction phenomena that depend on the surrounding colour of a certain stimulus. The surrounding colour is called the inducing stimuli or inductor [116]. Depending on the direction of the chromatic change provoked by the inductor, we will distinguish two types of colour induction:

Chromatic Assimilation occurs when the chromaticity of the test stimulus changes towards the chromaticity of the inducing stimulus. An example of assimilation phenomena is shown in figure 4.1.(a).

Chromatic Contrast occurs when the chromaticity of the test stimulus changes

away from the chromaticity of the inducing stimulus. An example of this effect can be seen in figure 4.1.(b).

In figure 4.1.(c) and (e), we can see a plot of the chromaticity coordinates of the stimuli presented in images (a) and (b). We denote the test stimulus as TS, that is, the image region that is affected by an inducing surround. These inductors are denoted as S1 and S2.

In the first column of figure 4.1 we see the effects of the assimilation, the test stimulus moves its appearance towards the appearance of its own surround. The TS is yellow, and it appears pink when surrounded by S1, that is red, i.e. yellow moves toward red and becomes pinkish. The same TS becomes greenish when it is surrounded by S2 that is green.

In the second column of figure 4.1, we see the effects of the colour contrast, the test stimulus moves its appearance away from the appearance of its own surround. The TS is grey and it appears yellowish when it is surrounded by the S1 bluish surround. Complementary, the same TS appears bluish when it is surrounded by a S2 yellowish surround. In this case, the induction phenomena is behaving inversely as it behaves in assimilation. Chromaticities of the perceived stimuli are going far from the surround chromaticity. This phenomena is called simultaneous contrast when it is given on achromatic images. A typical example of simultaneous contrast or brightness contrast is shown in figure 4.2, where the same stimulus seems darker when surround is lighter and lighter when the surround is darker.

Considering the given definitions and examples, it is obvious that any perceptual approach towards a colour texture representation should take into account the colour induction effects we have introduced above.

In psychophysics we find a wide range of works dealing with the induction phenomena or the influence of surrounding chromaticities on the appearance of colour [98, 81, 96, 97, 95, 24, 94, 2, 86, 109, 20, 85, 112, 110]. In all these works, authors present different aspects of colour human induction measurements. The influence from direct surrounds or remote inducers, the asymmetry of the measurements due to changes from luminance or the dependency on spatial frequency of patterns are some of the aspects that are measured and analysed. Conclusions from all these measurements pursue to give answers about how this perceptual mechanisms are organised in the human visual system. They help in building a more precise model on how human visual system acts from the retinal representation of colour to the final judgements on colour appearance. Considerations are done in terms of different physiological aspects as cone absorption rates and their retinal distribution, optical chromatic aberrations or the existence of opponent-colour signals in the visual pathways.

The most interesting conclusions from all these works from a computer vision point of view can be summarised in the two following points:

1. Changes on colour appearance due to the spatial frequency of patterns can be described by a two-step pattern-colour separable model [85, 109]:

   - First step, a colour transformation to a new coordinate space that is independent of the image content. The best correspondence of the derived data is given by the opponent-colour transformation.

(a)   (b)

(c) Assimilation induced by S1   (d) Contrast induced by S1

(e) Assimilation induced by S2   (f) Contrast induced by S2
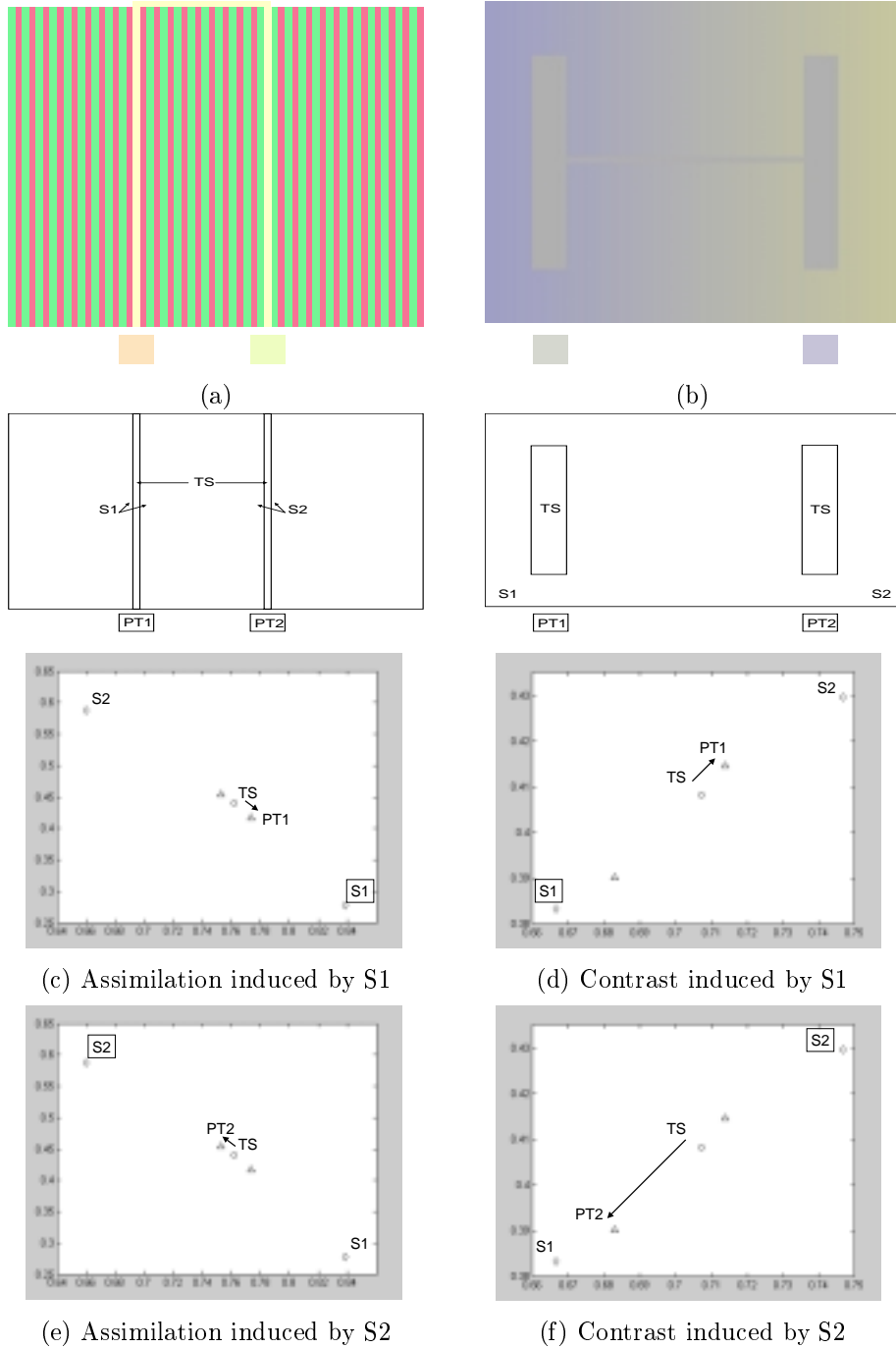
Figure 4.1: Colour Induction. (a) Colour Assimilation. (b) Colour Contrast. (c), (d), (e) and (f) plot chromaticity coordinates of the RGB values of the images (a) and (b) denoted as given in the below graphics. (c) and (e) Chromaticity moves towards the inducing surround.(d) and (f) Chromaticity moves away from the inducing surround.

Figure 4.2: Simultaneous Contrast

- Second step, in the previous coordinate frame, colour representation is transformed by a gain factor that is dependent of the image content.

2. The relationship between spatial frequency and the two types of colour induction can be summed up as follows [98, 29]:

  - A spatial frequency of 4 cpd. is a transition frequency between assimilation induction to contrast induction.
  - Spatial frequencies at 9 cpd. and 0.7 cpd. assures assimilation and contrast induction respectively for any inductor.

Frequency measures are given in cpd units (cycles per degree), that represents the number of cycles for 1 degree of visual angle. The visual angle is a common way to express a spatial measure that allows to adjust the observer distance and the displayed window size to different possibilities. In figure 4.3, we can see coloured square-wave patterns at different spatial frequencies. These plots are given on image size corresponding to the diameter of 6 degrees of visual angle when observed at 30cm. From 0.5 cpd tp 2 cpd we can perceive images with two coloured types of blobs, blue and yellow. As the frequency increases we tend not to perceive separate blobs but a global colour that is the result of the two basic colours plus the frequency effect.

Considering the above conclusions, we can derive a computational model for colour texture image representation based on the pattern-colour separable model shown at figure 4.4. Where Opp represents the opponent-colour transformation and A_ϕ and C_γ are respectively the assimilation and contrast operators, that represent the induction effects on each colour channel and for the corresponding range of spatial frequencies in the image. We have also indicated the possibility to insert other special phenomena that has been referred in the bibliography. A combination step of the resulting signals is represented by a P transformation.

This model has to allow to derive colour texture properties from the set of perceptually defined images. While the perceptual blurring has to allow defining global colour properties, the perceptual sharpening has to allow a better segmentation of different coloured blobs and the computation of their attributes.
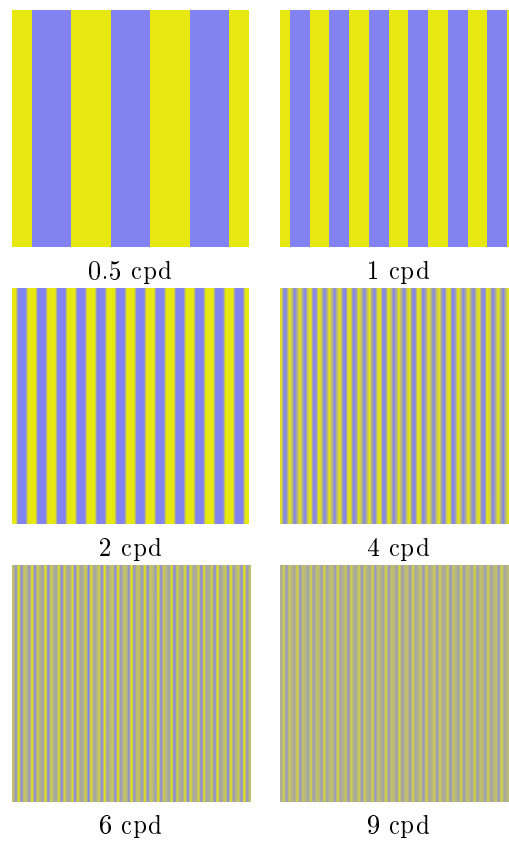
Figure 4.3: Colour Induction at different spatial frequencies. Frequencies are computed by considering observer position at 30cm from the image. Images are displayed on 6 degrees of visual angle.
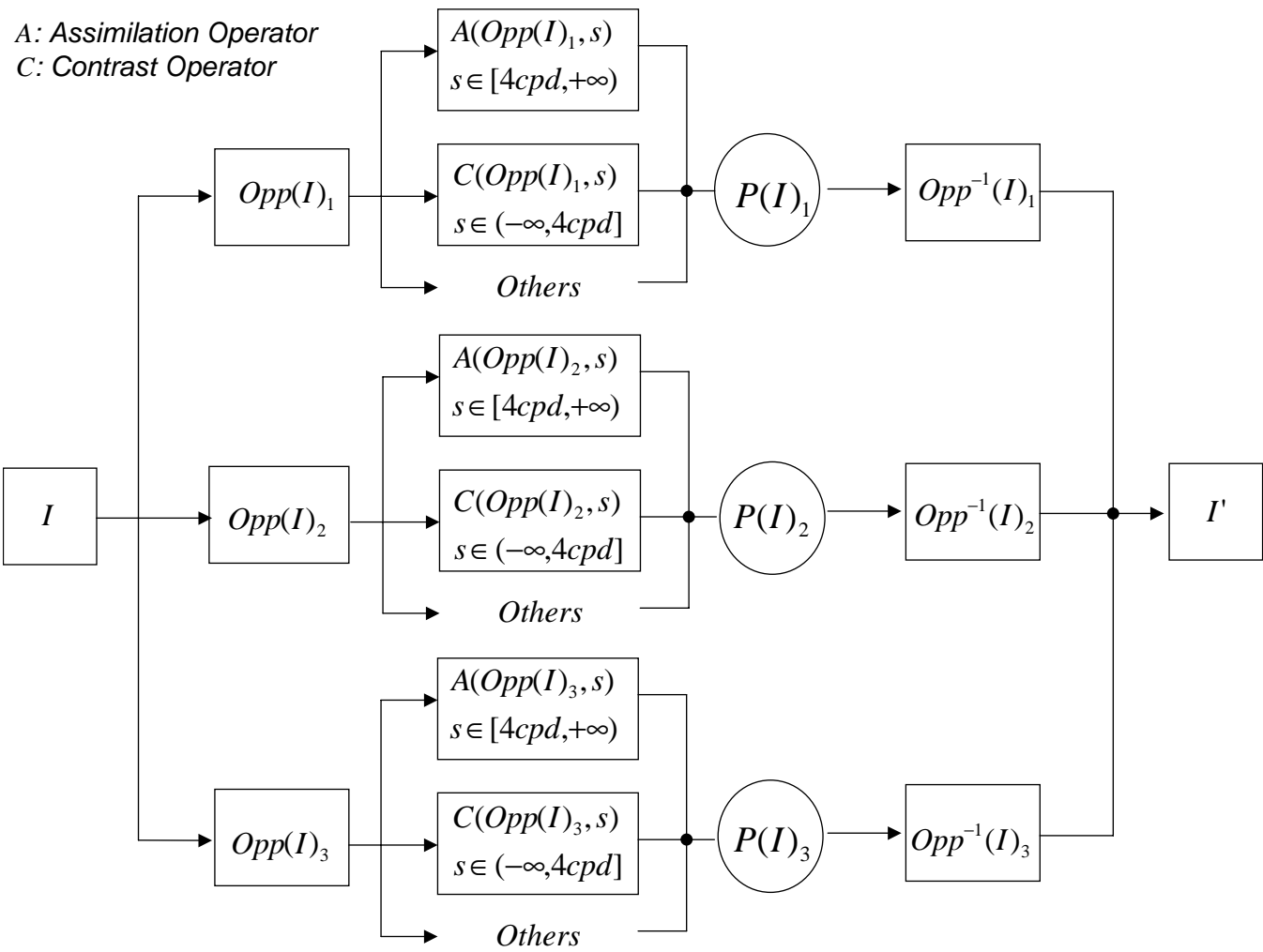
Figure 4.4: A pattern-colour separable model for colour induction.

In the next sections we will go deeply on how to define computational operators implementing the induction operators, but before to do it, we will introduce the opponent-colour space.

4.1.1   Opponent-Colour Space

The concept of opponent colours was first described by Hering in 1878, he made some interesting observations about some pairs of colours one never sees together at the same place and at the same time. While we are able to see a reddish or a yellowish orange, and a bluish or a greenish cyan, we never can observe a greenish red or a bluish yellow neither the opposite. These two hue pairs, red-green and blue-yellow are called opponent colours.

From this observation Hering hypothesised the existence of a unique visual pathway to encode red and green, and a unique visual pathway to encode blue an yellow. The same hypothesis was done for a visual pathway encoding achromatic black and white signals. It takes to formulate a neural representation of colours.

This opponent process model was left behind while the trichromatic theory of colour was stabilising the basis of the modern colorimetry based on the colour-matching experiments and all the derived standard spaces. It was resurrected when a hue-cancellation method was defined by Hurvich and Jameson [58] to quantitatively measure colour-opponency.

Due to the efforts of Hurvich and Jameson with the hue-cancellation experiment and plus the quantitative data provided by direct neurophysiological responses obtained from some measurements in the retinal neurons of a fish and in the lateral geniculate nucleus of non human primates, the opponent processing has been no longer questioned.

From the Hurvich and Jameson measurements a general opponent model schema can been derived, we show a computational approach of it in figure 4.5. There are some variations of the transform to this space from a trichromatic Young-Helmholz space, all of them follow the same schema of colour incompatibility the difference lies in the coefficients, α_i, β_j and γ_k, that combine the input signals.

In computer vision we usually only have a colour image representation in a RGB space of an unknown camera and under undefined conditions. Among others, a common representation of the opponent colour space is the one used in [100] that is defined as:

$$Opp(p) = p \cdot \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 0 & -2 \end{pmatrix}, \tag{4.1}$$

$$RGB(p) = Opp^{-1}(p) = p \cdot \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & -1/2 & 0 \\ 1/6 & 1/6 & -1/3 \end{pmatrix} \tag{4.2}$$

where p is a 3D-vector of the RGB coordinates of the given space.

To be able to better establish the parameters for the spatial operators we will define in the next sections, we will use an orthonormal basis, given by:
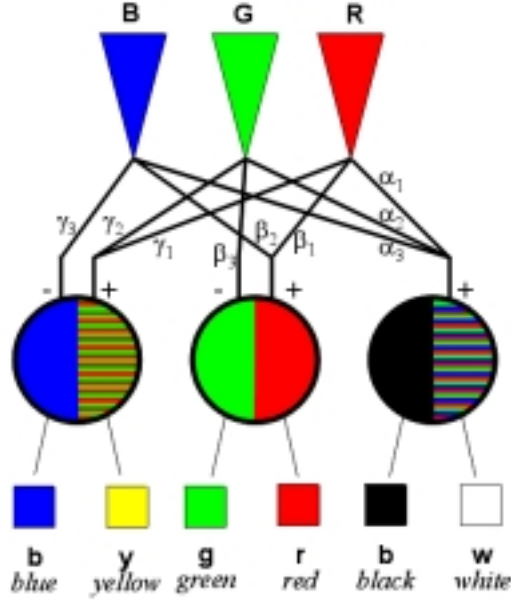
Figure 4.5: An Opponent colour vision model for a computational approach.

$$Opp(p) = p \cdot \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \end{pmatrix}, \qquad (4.3)$$

$$RGB(p) = Opp^{-1}(p) = p \cdot \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} \end{pmatrix} \qquad (4.4)$$

In both cases the first dimension represents the intensity or dark-white channel, the second dimension represents the red-green chromaticity channel and the third dimension represents the yellow-blue chromaticity channel.

4.2 Colour Assimilation as a perceptual blurring

As has been previously introduced, colour assimilation is the perceptual mechanism that takes chromaticities of regions with very high frequencies towards the chromaticities of the neighbouring regions. This effect is the result of a spatial blurring, that is usually implemented in computer vision with the convolution of the image with a gaussian spatial filter [75, 68]. However, in this case the spatial filter will not be applied on the RGB space as it is usually done, it will be applied to the opponent-colour space.

The idea of building a perceptual tower for a multiscale representation simulating different views of the same scene from different observer distances and considering the human colour perception has been taken to a computer vision model for the first time in [11, 79, 84, 80]. All these works have been based on pshychophysical measurements of colour appearance on human subjects given by the Spatial–CIELAB space defined in [119], this measurements have given the parameters of the spatial filters needed to simulate human assimilation on a CIELAB colour space. We will go deeply on this space in the next section.

In order to correctly apply Spatial–CIELAB blurring in images, the sensor blurring should be removed and be substituted by the perceptual one. In the thesis of Boukoubalas [12] there is an interesting explanation on how to do this.

4.2.1 S–CIELAB: Spatial CIELAB

S–CIELAB is a spatial extension to the CIELAB[1] colour metric that is used for measuring the quality of colour reproduction in digital images. It has been defined to improve the error computation on non-uniform spatial regions.

The Spatial–CIELAB representation is based in the two-step model defined by Wandell et al in [85, 109]. Firstly, a step to an opponent-colour space from the CIELAB representation is done, and secondly a convolution with a kernel whose shape has been psychophysically determined for each colour dimension. Finally, the filtered channels are transformed again to the standard CIELAB [117], that is actually representing the Spatial–CIELAB.

In this case, the opponent representation is built from the standard XYZ colour space, and is given by:

$$Opp(p) = p \cdot \left(\begin{array}{ccc} 0.279 & -0.449 & 0.086 \\ 0.72 & 0.29 & -0.59 \\ -0.107 & -0.077 & 0.501 \end{array} \right) \tag{4.5}$$

where p is given by (X, Y, Z) following the standard CIE 1931.

The spatial filters for each opponent channel are built as a sum of gaussian functions, that is:

$$f_k = m_i \sum_i \omega_i E_i \tag{4.6}$$

where k represent every one of the three opponent channels, m_i is a scale factor chosen to make that the kernel sums to one, and

$$E_i = k_i \exp^{\frac{x^2 + y^2}{\sigma_i^2}} \tag{4.7}$$

again the k_i factor scale is selected to make that E_i sums to 1. The measured values to substitute the parameters ω_i and σ_i are given for each opponent channel in table 4.1. Where the spreads are given in degrees of visual angle. Depending on

---

[1]The CIELAB space is an important international standard for colour measurement. The main property of CIELAB space is the uniformity with respect to human colour judgements.

Opponent Channel	Weights ω_i	Spreads σ_i
1 (I)	0.921	0.0283
	0.105	0.133
	-0.108	4.336
2 (R-G)	0.531	0.0392
	0.330	0.494
3 (Y-B)	0.488	0.0536
	0.371	0.386

Table 4.1: Parameters of the Spatial–CIELAB spatial kernels.
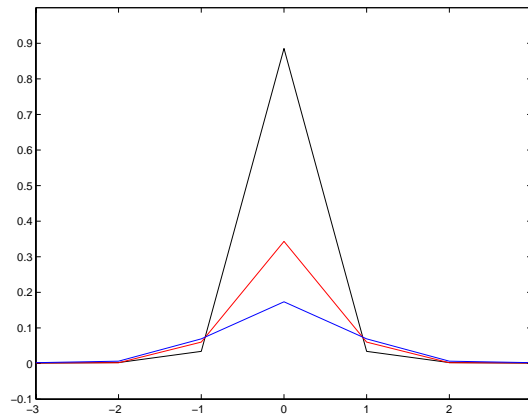
Figure 4.6: Profiles of the two-dimensional symmetric kernels for the Spatial–CIELAB. Black, Red and Blue colour lines represent the kernel for the Intensity, Red-Green and Yellow-Blue channel respectively.

the observing conditions the equivalence in pixels is easily computed by the following expression:

$$\sigma_{pixels} = d \cdot \tan\left(\sigma_{degrees}\right) \cdot R \qquad (4.8)$$

where, σ_{pixels} and $\sigma_{degrees}$ represent the spreads in pixels and in degrees of visual angle, respectively; d is the distance in cm between the stimulus and the observer (or the camera in computer vision), and R is the display resolution that is given in pixels/cm.

The profiles of these symmetric filters are shown in figure 4.6, where the filters have been built to simulate the human colour perception of an image of 550 pixels, displayed on a visual field of 20cm and observed from 40cm.

To illustrate how this transformation behaves on a given image we shown in figure 4.7 the results of applying the Spatial–CIELAB transformation on two images presenting an important colour assimilation effect. We can see on the profiles below,
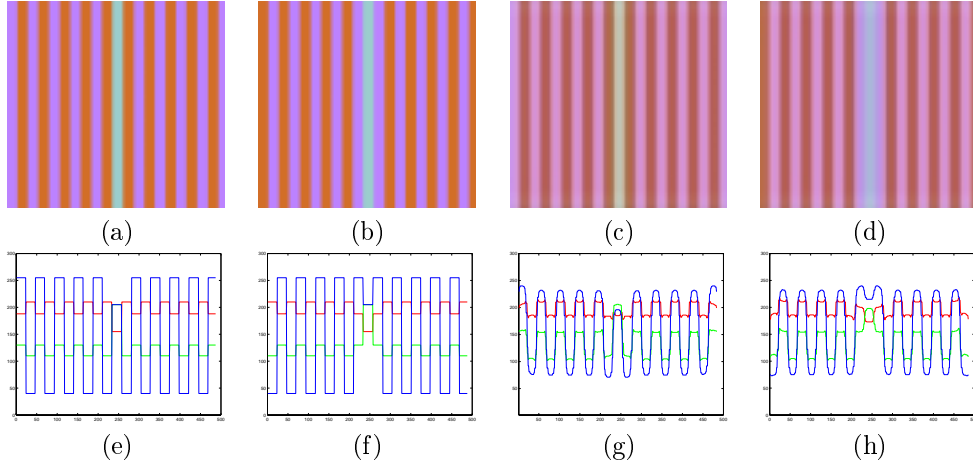
Figure 4.7: (a), (b) Examples of two images presenting important assimilation effects. (c) and (d) Previous images transformed by Spatial–CIELAB. (e), (f), (g) and (h) are the RGB profiles of images (a), (b), (c) and (d), respectively.

how the Spatial–CIELAB transformation makes that the green-blue band is becoming bluish when is surrounded by blue and it becomes reddish when surrounded by red.

4.3 Colour Contrast as a perceptual sharpening

Colour contrast is the complementary mechanism to the assimilation that takes chromaticities of regions with spatial low frequency. Whereas a computational model for colour assimilation has been proposed in computer vision, a computational operator that simulates colour contrast phenomena has not been proposed in the computer vision literature.

In the following sections we will present the main contribution of this work that is devoted to this end, that is, to define an operator that enhances differences in the transitions among colours of regions presenting lower frequencies. While the assimilation effect has been solved by a blurring operator, it seems quite natural that the contrast effect will have to be implemented by a sharpening operator.

The final foal of this operator is to produce a sharpened image that allows a better segmentation of texture blobs in order to be able to compute their local attributes, following human perceptual considerations.

4.3.1 Local perceptual sharpening

In this section and in the subsequents we will progressively define sharpening operators presenting good properties to represent colour contrast. The first and the most common sharpening filter is defined as:

<div align="center">(a)   (b)</div>
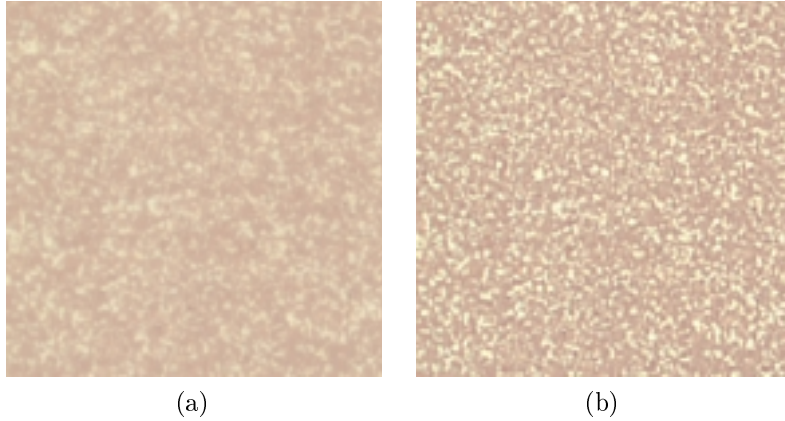
Figure 4.8: Effect of the traditional sharpening operator:(a) original image, (b) sharpening of the image using the usual transform on the RGB space

$$S_c(I, \gamma) = I_c - \gamma \nabla(I_c) \qquad (4.9)$$

where I_c is the c–th channel of a colour image I of dimensions $N \times M$, $\nabla(I_c)$ is the laplacian of the image channel c ($\nabla(I) = \partial^2 I/\partial x^2 + \partial^2 I/\partial y^2$) and γ is a constant that controls the amount of the enhancement. This process is done for each channel separately. Nonetheless, the laplacian operator is very noise sensitive. To avoid this problem, the laplacian of a gaussian (LoG) is used, that is, to smooth the image before the enhancement in order to reduce noise effects. The resulting operator has the following expression:

$$S_c(I, \gamma) = I_c - \gamma LoG(I_c), \qquad (4.10)$$

$$LoG(I) = -\frac{1}{\pi\sigma^4}\left[1 - \frac{x^2 + y^2}{2\sigma^2}\right] e^{\frac{x^2+y^2}{2\sigma^2}} \qquad (4.11)$$

where the $LoG(I)$ expression is centered on zero and with gaussian standard deviation σ. Whichever it is the method used, there is a post–process to clip the output of the responses outside the range of the image (usually $[0 \dots 255]$ in the rgb–space). We will use the notation $S(I, \vec{\gamma})$ to indicate that the operator $S_c(I, \vec{\gamma}[c])$ is applied for each channel of the image and merged together to form a new n–spectral band image.

The first attempt to chromatic contrast perception enhancement is the usual brightness sharpening, but applied to all the bands of the image, that is: $S(I, \vec{\gamma})$. This operator has been applied to the colour texture image of figure 4.8. Apparently, there is a clear enhancement of the texture that form the image. In the original image the transition from one blob to another blob of different colour is very smooth. Even that, some texture is appreciated. Enhancing the image makes the colour blobs more
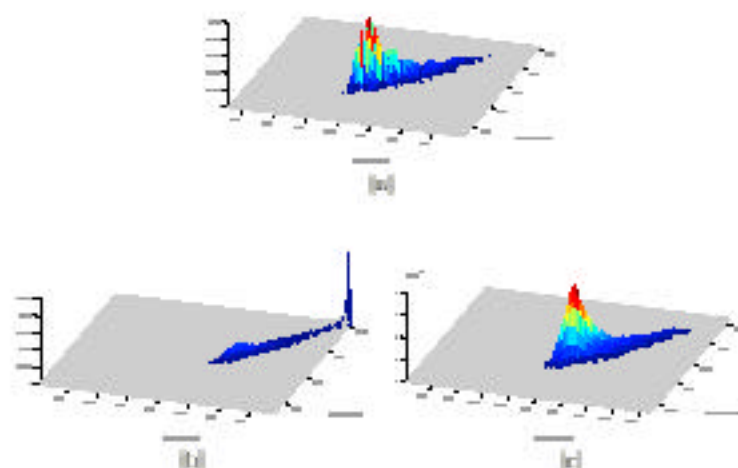
Figure 4.5: (a) Red-green histogram of figure 4.9(a), (b) Red-green histogram of figure 4.9(b), (c) Red-green histogram of image 4.9(c) after a sharpening operator without forcing the dynamic range to be the same than in the original image.

distinguishable, let us see how this effect is projected on the colour image distribution. In this example, the blue channel is the one with more useless information (the colours that appear are mainly formed by red and green). In figure 4.9 we show the histogram of both images using only the red and green channel. In figure 4.9(a) we see the colour distribution of the original image, although we can appreciate two bright colours in the image, the classification presents only one important peak. In figure 4.9(b) we see how that sharpens operator makes to appear that important peaks, one corresponding to pink blobs on the other corresponding to white blobs.

What we will see is a combination of the contrast enhancing sharpening operator. Instead of operating on the RGB space we will operate on the opponent space that give us a more perceptual approach as we have previously introduced.

This change will not be enough to get good contrasted images. The fact that the range of the image is fixed in the maximum and minimum of the dynamic range of the constraints may colour distortion to appear. This is the case of the peak belonging to white in figure 4.9(a), where a maximum is formed. When applying the operator without resmihing the range of the result (fig. 4.9(c)) one of the two main colours disappears. And finally, in both cases the range of the result image is quite different from the range of the input image. In short, neither the limit on

(a)   (b)
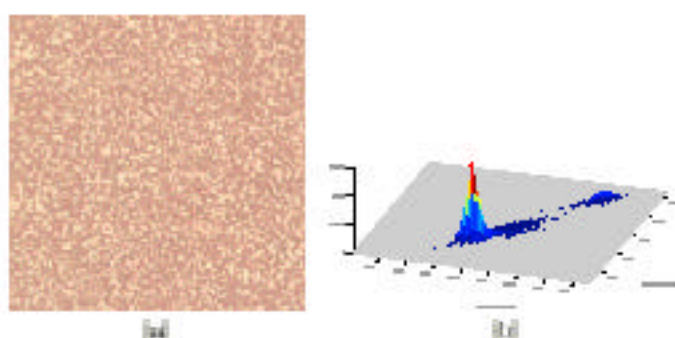
Figure 4.29: Effect of the perceptual space sharpening operator. (a) sharpening of the image using the local sharpening operator T on the figure 4.8(a). (b) the reddish histogram of (a)

dynamic range nor the freedom to surpass this range nor a good choice to improve the underlying texture. What we propose is to mix both points by setting the magnitude the range of the neighbourhood of the input pixel. The neighbourhood limits the output result inside range of the minimum and maximum in a certain neighbourhood as defining the spatial extension to look around. The expression of this new operator, T, is as follows:

$$T(I)_{\min} = RGB(S(C)g(T), \mathfrak{I}[\underline{\quad\quad}])$$ (4.23)

where the magnitudes and subindex $\max(I, \omega)$ and $\min(I, \omega)$ are the maximum and minimum range for each pixel inside a neighbourhood ω. The maximum and minimum threshold can be carried out by means of morphological operators. In this case $\max(I, \omega) = I \ominus \omega$ which is the morphological dilation and $\min(I, \omega) = I \ominus \omega$ being the morphological erosion, in both cases using ω as the structuring element.

The parameter $\mathfrak{I}$ plays an important role for the properties. It allows to control which component of the perceptual stimulus we want to enhance. As an example, figure 4.29 shows the result of the defined operator with a constant $\mathfrak{I}$. In this case the visual appearance of the image is closer to the original image than in the latter image in figure 4.8(a). Although the texture seems not to be as sharper as it is using the traditional sharpening, the histogram shows the opposite. Using T makes the blobs to be very well limited, without odd effects, and the range of the resulting image is in keeping with the original range.

The proficiency of using equation 4.23, whatever the colour space or the range limit that is applied, is that it reinforces the differences in the transition between different colour areas, but not the areas itself. In fact, the behaviour of granularises becomes not to an edge detector. That makes the picture to look like being dramatically enhanced when dealing with small texture images, but not when working with more homogeneous

(a)

(b)

(c)
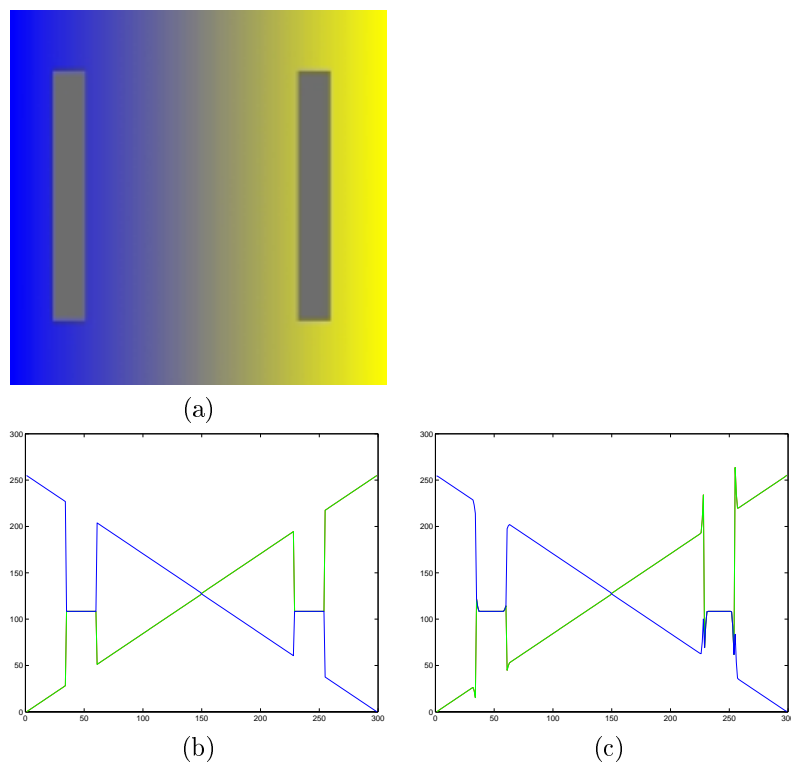
Figure 4.11: Example of the Local perceptual operator for large stimuli: (a) original image, (b) RGB profiles of image (a), (c) RGB profiles of image (a) applying operator T.

ones. Some extra colours are created nearby the edges of the image, and this is what makes the texture emerge substantially. Although the edges are enhanced, and so is the texture, the main area of the colour stimulus is not changed as it should be for a chromatic contrast effect. An example of this situation is figure 4.11. (a) is the image used in explaining colour contrast but enhancing its colours. It can be observed that the effect is on the edges of the stimuli and not on the stimuli themselves. The centres of them are the same when there should be chromatically different. In fact, the operator should spread the edge response all over the areas that form the edge.

4.3.2 Region perceptual sharpening

Based on the work of Grossberg[2] [50] and the operator T, we want to construct a new operator to improve chromatic contrast simulation. The main idea is to recognise inhibited and activated areas, whichever the colour dimension is analysed. When applied, for example, to the red-green opponent colour dimension, the active areas will be the reddish ones and the inhibited areas the greenish ones. Computationally its is equivalent to the intensity of the stimulus. In the preceding example a red area is positive and a green area negative. However, whichever is the sign of the area it can not be considered neither positive nor negative unless it is compared with another area. There will be positive and negative responses when comparing against its surround. A yellow area is a negative area when its surround is red but negative when green. The laplacian operator performs well in such definition because its response is positive in the transition between dark and light, and negative on the contrary. We will use the fact that the laplacian indicates the edge location by a zero cross, i.e: a change between positive an negative response or vice versa. We define an homogeneous area as the points that lie inside the regions surrounded by zero cross points. But as we are working in a discrete domain, the zero–crossings are not well locate. Then, a zero cross are those points where there is a change of sign of the laplacian between it and one of its neighbours. That makes us define two sets of zero–crossings:

$$Z_w(I) = \{\mathbf{p} \in I \mid \exists \mathbf{p_i} \in H(\mathbf{p}) : sgn(LoG(I,\sigma)_{\mathbf{p_i}}) = -1 \ \wedge \ sgn(LoG(I,\sigma)_{\mathbf{p}}) = 1\} \quad (4.13)$$

$$Z_b(I) = \{\mathbf{p} \in I \mid \exists \mathbf{p_i} \in H(\mathbf{p}) : sgn(LoG(I,\sigma)_{\mathbf{p_i}}) = 1 \ \wedge \ sgn(Log(I,\sigma)_{\mathbf{p}}) = -1\} \quad (4.14)$$

where $\mathbf{p}$ stands for the pixels of the image I, $H(p)$ for the pixels belonging to the neighbourhood of $\mathbf{p}$, and $sgn(\mathbf{p})$ is the sign of the intensity value in $\mathbf{p}$, which equals 1 when positive and -1 when negative. $Z_w(I)$ are the zero-crossings taken at the falling edge, and $Z_b(I)$ at the raising edge. $Z_w(I)$ coincide with the limits of the light (or white) areas and $Z_b(I)$ are the limits of dark (or black) areas. I is a one–channel image, being it the responses to one of the opponent channels. Usually $H(\mathbf{p})$ is defined as

$$H(\mathbf{p}) = \{(p_x, p_y - 1), (p_x + 1, p_y), (p_x, p_y + 1), (p_x - 1, p_y)\} \quad (4.15)$$

---

[2]This a psychophysical work on brightness contrast based on on–off lateral geniculate cells, modeling responses in the boundary contour system by a sum of exponential functions that is nearly equivalent to the laplacian of gaussian.

where (p_x, p_y) are the coordinates of $\mathbf{p}$. From these definitions we construct two sets of connected components. The inner points of the connected components are areas where there is no sudden changes, and will be considered homogeneous areas. We will say that a connected component is white when all of its surround is darker than itself. In the same way, black connected components are lighter than all of their surround. From this point, $\neg Z_w(I)$ defines the set of plausible white connected components, $C^w(I) = \{C_i^w(I)\}$, whereas $\neg Z_b(I)$ defines a set, $C^b(I) = \{C_i^b(I)\}$, of plausible black connected components. From now on, we will use the term *region* instead of connected component. The following step is to distinguish those $C_i^w(I)$ and $C_i^b(I)$ that are, actually, white or black regions. We define the white regions as

$$W(I) = \{C_i^w(I) \in C^w(I) \mid \sum_{\mathbf{p} \in C_i^w(I)} -sgn(LoG(I,\sigma)(\mathbf{p})) = |C_i^w(I)|\}, \qquad (4.16)$$

and in a similar way it is defined the set of black regions

$$B(I) = \{C_i^b(I) \in C^b(I) \mid \sum_{\mathbf{p} \in C_i^b(I)} sgn(LoG(I,\sigma)(\mathbf{p})) = |C_i^b(I)|\} \qquad (4.17)$$

Up to this point, not all the pixels are classified as belonging to a black or white region. Those unclassified pixels will be merged in a neutral class, $N(I)$. The pixels in $N(I)$ belong to regions that are surrounded by lighter and darker regions at the same time, and so, can not be classified as black or white regions.

$$N(I) = (C^b(I) \cup C^b(I)) - W(I) - B(I) \qquad (4.18)$$

Thus far, all regions of the image are classified in one of the three types of regions. Moreover, we need to specify how much black or white these regions are. The final image $ELoG$ (*Expanded Laplacian of Gaussian*) will measure how different is a region from its surround assigning at each pixel of the region the maximum difference of all the pixels in this region with its surround (i.e: the laplacian of gaussian).

$$ELoG(I_{\mathbf{p}}, \sigma) = \begin{cases} min_{\mathbf{p_k} \in W_i(I)}(LoG(I,\sigma)_{\mathbf{p_k}}) & : & \mathbf{p} \in W_i(I) \\ max_{\mathbf{p_k} \in B_i(I)}(LoG(I,\sigma)_{\mathbf{p_k}}) & : & \mathbf{p} \in B_i(I) \\ 0 & : & \mathbf{p} \in N_i(I) \end{cases} \qquad (4.19)$$

We use $ELoG(I_{\mathbf{p}}, \sigma)$ when applying the process to the pixel $\mathbf{p}$ of I, and $ELoG(I,\sigma)$ when it is calculated all over the pixels of the image I. Two examples of the $ELoG$ operator applied on monochromatic stimulus are shown in figure 4.12. The upper graphics show the original stimuli in blue lines. Graphics (c) and (d) show the result of laplacian in blue lines and the output of $ELoG$ in red lines. Black regions have positive response whereas white regions are negative. Regions between darker and lighter ones have 0 response. We want to remark how the maximum and minimum inside each region is expanded all around it.

What remains to conclude is to apply the sharpening formula using $ELoG(I)$ instead of $LoG(I)$. We will call the new operator *Expanded Sharpening* $(ES(I,\vec{\gamma})_{\vec{\sigma}})$,

$$ES(I,\vec{\gamma})_{\vec{\sigma}} = I - \vec{\gamma} ELoG(I,\vec{\sigma}) \qquad (4.20)$$
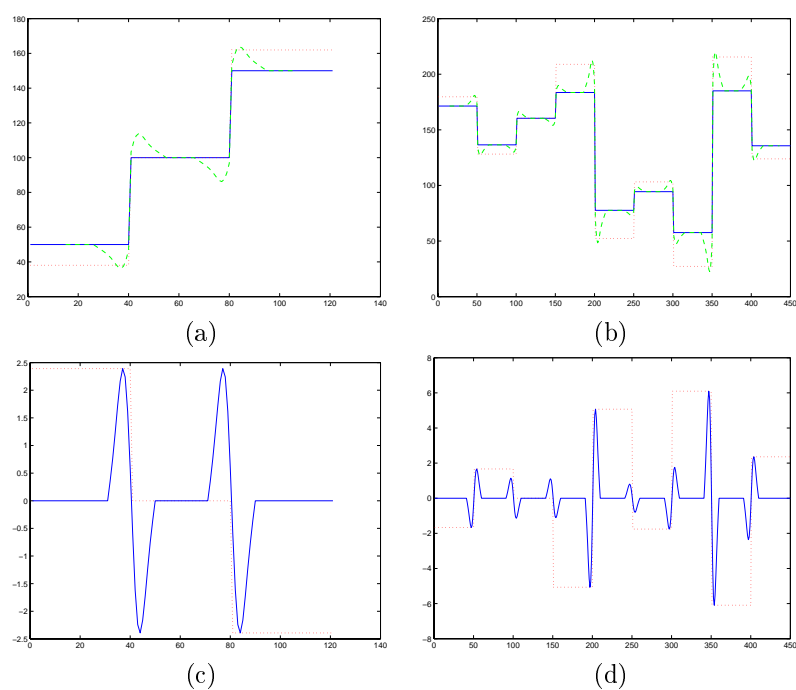
Figure 4.12: Graphic explanation of the effect of operator $ES(I)$, on (a) and (b): In solid lines the original stimulus, in dotted lines the output from the $ES(I)$ operator, compared with the T operator in dashed lines. On (c) and (d): In solid lines the Laplacian of Gaussian response, and in dotted lines the output form $ELoG(I)$.

(a)

(b)

(c)

(d)

(e)

(f)

(g)
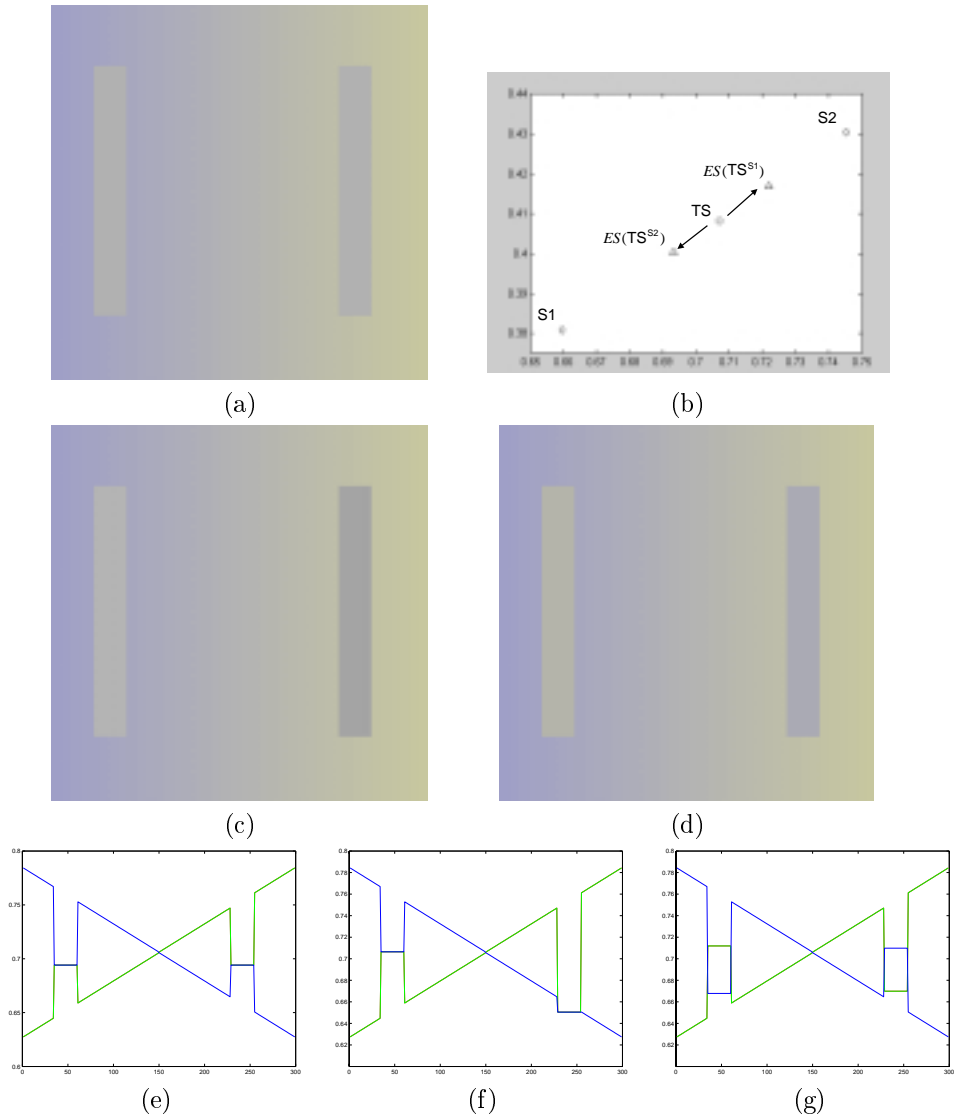
Figure 4.13: Region perceptual sharpening: (a) the input image, (c) and (d) the results with two different parameter configurations. (b) shows the displacement of the chromatic values of the test stimulus in (d) with regard to the original stimulus. And (e),(f),(g) comparison of the profiles of an horizontal line in the RGB space, from left to right: original, (c) and (d).

Returning to figure 4.12 in (a) and (b) the red lines are the response of the new region perceptual operator ES compared to the responses of the local perceptual operator T, green lines. Whereas $T(I)$ only have effect in a short neighbourhood, $ELoG(I)$ works on the whole region. The question is, will it work? and the answer is not so simple. When dealing with perceptual vision the way to validate a model is by means of psychophysical experiments. Some times they are done with a very reduced group of individuals and a short set of test, it is because of the intrinsic complexity of this type of experiments. The kind of tests are a uniform background scene with regular polygons, there are some that are more complex than others. They could be gratings or two simple squares [95, 98, 110, 81]. In any case they should be done by scientists of this field.

In this thesis this problem will not be broached as it is a computational approach to perceptual vision and we are not trying to imitate the human vision but to approximate the images to what humans see. As a matter of fact, this operator has been inspired in the experiments before mentioned. These experiments analyse the reaction to certain isolated stimulus and lead us to look for the different stimuli in the image.

We have introduced the example in figure 4.11 to see the leaks of operator T, in this case the result is what is expected. It is depicted in figure 4.13, where the original colours of the example are used. In (a) there is the input image and in (c) and (d) two examples of the operator changing the parameter $\vec{\gamma}$. (e), (f) and (g) are the horizontal profiles of the central line of the image for the input image and both examples (c) and (d). The red response is occluded by the green one as they are the same. It can be appreciated a shift of the test stimulus against the surround. The chromatic coordinates of the test stimulus for the first example (c) are the same as in the input image as the parameter $\vec{\gamma}$ has been adjusted to work only in the black–white pathway. In the second example $\vec{\gamma}$ is a constant vector, and thus, all pathways are equally weighted. In this case there is a change in the chromatic coordinates of the final stimuli. This situation is plotted in the graphic Fig.4.13(b), where the two surround of the stimuli have chromatic coordinates S1 and S2 (left to right) and the original test stimulus is TS. TS1 and TS2 are the chromatic values of the test in the output image. It is clear that they behave in the same direction that the HVS does.

One such examples of the operator on images of small isotropic texture is illustrated on figure 4.14.

This operator is based on the fact that a region is conceived as inhibited or activated in intensity, red–green or blue–yellow channel. If there is a problem it will be in the definition of the regions and the assumption that a region can be only inhibited or activated for a certain channel. But the reality shows that under certain circumstances it can be inhibited and at the same time activated. This is the case of the example of the bars. What happens when both bars are joined together with a slim bar of the same stimulus? Taking as example the blue–yellow channel, the left grey bar will be an activated region whereas the right bar will inhibit. But as they are connected, they are the same region. When a region is inhibited and activated simultaneously then it belongs to the set of neutral regions that show no reaction and then the result is the same input image. The modified experiment is shown in figure 4.15, where the stimuli are the same colour as before and the results of the operator
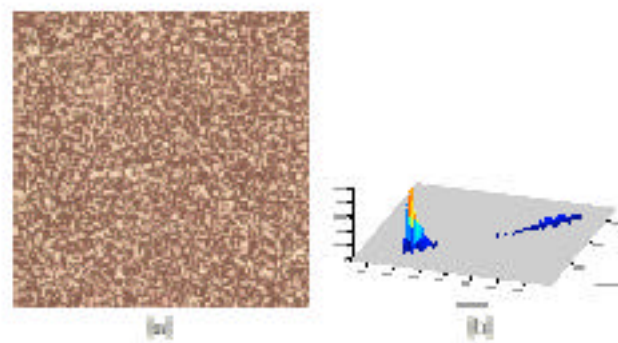
Figure 4.24: The result of applying JSUD on the image in fig. 4.8(a), and a view of its histogram over the red and green bands.

Figure 4.25: Example of the region perceptual sharpening operator in a case where its performance is not good enough. The first stimuli are connected and they transform to connected region being inhibited and activated at the same time.

are not shown because they are exactly the same image.

4.3.3 Spread perceptual sharpening

In short, the local perceptual sharpening operator fails because it does not extend to the centre of the stimulus and the region perceptual sharpening fails because,although it comes from psychophysical ideas, it does not consider one region to have two different behaviours at the same time. To solve this conflict we have designed an alternative operator that combines the good properties of the previous operators. The idea is to use the *LoG* edge enhancing to locate the boundary of regions and to use some of the definitions of the region perceptual sharpening to reduce the number of points used in the operator. The intensity of inhibition/activation in this points will be scattered to the centre of the region no matter which kind of region it is.

Starting form equations 4.13 and 4.14, which define the points that form the borders of the regions of the image, we can take the local inhibition or activation of a region taking the *LoG* in these points. The following step is to construct a surface where its height in a certain point indicates the level of activation of this point, taking into account the intensity on the points of edges that define the region to which it belongs. This surface must have some properties:

1. The points on the boundaries must preserve its energy, i.e: the relationship between adjacent regions must be maintained.

2. The zero crossings between points of the boundaries must remain equal, i.e: there will not be more regions than in the input energy image.

3. Zero crossing can only be added inside a neutral region (defined in Eq. 4.18).

Let us call $\mathcal{S}(\mathcal{X}, \mathcal{Y})$ the operator that constructs this surface from the energies of a set of boundary points, X, giving the activation energy on points Y. An immediate solution is to use some kind of surface interpolation, but not all possible. Some of the possibilities are: nearest neighbourhood interpolation, linear interpolation and cubic Hermite interpolation. Some that are not possible are those based on spline interpolation. The choice of the interpolation method will affect the smoothness of the resultant image. The smoothness is achieved constraining interpolation to certain conditions on the continuity of the first and second derivatives. The complexity of these methods is considerable and it has to be kept in mind when working with large images. In this case, it is reasonable to use linear interpolation, instead. Now we can define the new operator. Since our definition of the operator spreads the energy of the region borders into its inside, we will call it *Spread Sharpening* (*SS*), and similarly the resulting energy surface is a spread modification of the *LoG* surface. Then,

$$SLog(I, \sigma) = \mathcal{S}(LoG(I, \sigma)_{Z_w(i) \bigcup Z_b(I)}, I), \qquad (4.21)$$

is the spread *Log* taking as a control points the energy of the points where there is a change on the inhibition/activation, and evaluated all over the points of image I. Following the same schema than in the local and region perceptual sharpening operators (Eqs. 4.12, 4.20) the final operator will be defined as:

(a)   (b)
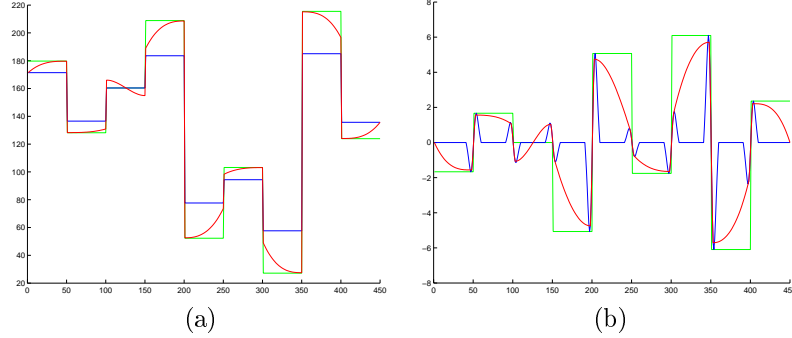
Figure 4.16: Graphic explanation of the functioning of region perceptual sharpening (SS). In (a) the original signal (blue line), the ES output (green) and the SS (red) are plotted. (b) is the response of $SLoG$ in red in front of the LoG in blue.

$$SS(I, \vec{\gamma})_{\vec{\sigma}} = I - \vec{\gamma} ILoG(I, \vec{\sigma}) \tag{4.22}$$

Taking the 1D signal of figure 4.12(b) we will illustrate the effects of the operator. Figure 4.16(a) plots in blue line the original input, in (b) the blue line is the LoG response of the signal, green line is the $ELoG$ response of the previous operator and red line is the $SLoG$ response. It is evident the spread effect of the function $\mathcal{S}$. The function solves de problems of neutral regions and makes the edges influence the inner part of the region. The final output signal is shown in red in (a) compared to the output of operator ES in green.

We noted that the region perceptual sharpening failed when applied to image in figure 4.15. Let us test the performance of this last operator. Figure 4.17(a) depicts the resultant image, whereas in (b) we have shown, as an example, the output from the inhibition/activation function $SLoG$ of the blue–yellow pathway. The profiles shown are from the original image and the output image. The profile from the output of ES applied on the same image is not shown because it is exactly the same as the profile from the input image.

4.3.4 Examples

These operators should be tuned to the contents of the image to adjust the frequencies at which they work better taking into account the distance from which the images are seen. Other parameter to adjust are the ratios of each opponent channel in the contrast response. These adjustments should come from psychophysical measurements. Whereas it seems that there begins to be a consensus on the first set of parameters, contrast begins at least at 1.7 cpd, it is not clear the influence of each channel in the response. Psychophysics agree that the intensity is the most sensitive channel, in second term there is the red–green channel and finally the blue–yellow channel. However we did not find any literature on which their ratios are.
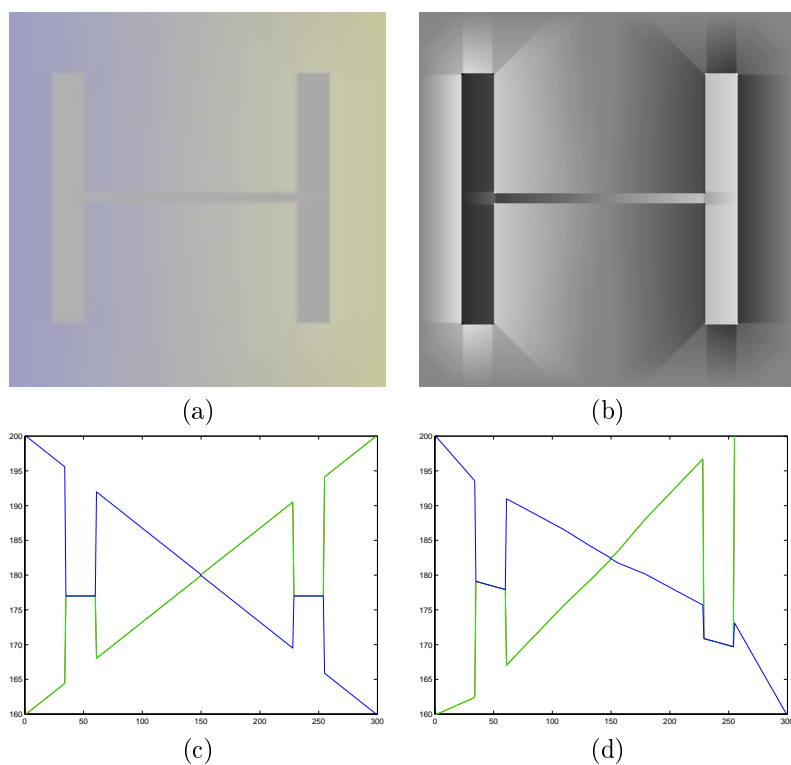
Figure 4.17: Graphic explanation of the spread perceptual sharpening operator: (a) is the spread perceptual operator output applied to the image in figure 4.15, (b) is the $SLoG$ response using linear interpolation of the blue–yellow pathway, (c) is the profile of an horizontal line from the original image in the RGB space and (d) in the case of (a).
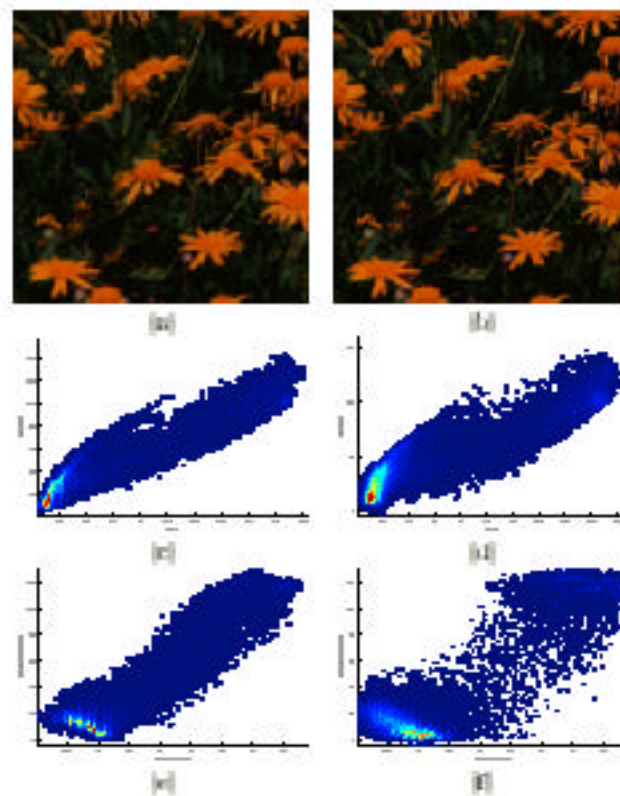
Figure 4.18: Example of the spatial perceptual sharpening. [Vistec]. (a) and (b) original and spatial perceptual sharpened result. (c) and (d) red-green histograms of images (a) and (b). (e) and (f) red/green-blue/yellow histograms of images (a) and (b).

Figure 4.20: Example of the spatial perceptual sharpening [(a)...]. (a) and (b) original and spatial perceptual sharpened result. (c) and (d) red-green histogram of images (a) and (b). (e) and (f) intensity-red/green histogram of images (a) and (b).

Figure 4.26: Example of the spatial perceptual sharpening [Vistex ...]. (a) and (b) original and spatial perceptual sharpened result. (c) and (d) green-blue histogram of image (a) and (b). (e) and (f) red/green-blue/yellow histogram of image (a) and (b).

Figure 4.22 Example of the spetral perceptual sharpening (Vistas _____). (a) and (b) original and spatial perceptual sharpened result. (c) and (d) red-green histogram of images (a) and (b). (e) and (f) intensity-red/green histogram of images (a) and (b). (c) and (d) the 2D-histogram rejecting one of the dimensions in the RGB space. (e) and (f) 2D-histogram in the opponent space.

scale ($\vec{\sigma}$)	$M_1^{ps} > M_1^o$	$M_2^{ps} > M_2^o$	$M_1^{ps} > M_1^o \wedge$ $M_2^{ps} > M_2^o$	M_1^{ps} imp.	M_2^{ps} imp.
high	73.78	67.07	48.17		
medium	99.39	42.68	42.07		
low	63.41	58.54	38.41		
all	99.39	85.98	85.37	159.4	97.1

Table 4.2: Spread percetual sharpening on VisTex image database. Values are in %.

To study the operators we have analysed a set of images from the texture image database VisTex from MIT MediaLab. It is a large database from which we have selected some of them to illustrate the effects of the spread perceptual operator. The figures 4.18, 4.19, 4.20 and 4.21 show four of these images. In all cases: (a) is the original image, (b) is the output image for which the parameters have been chosen empirically, (c) and (d) are the projected histograms of (a) and (b) respectively rejecting one of the dimensions, in each case the dimension rejected was the one that enables to show a better view, and finally, (f) and (g) are the projected histograms on the opponent space, the rejected dimensions have also been chosen to best display the effects of the operator.

The effects are more visible when analysing the opponent space. In the first example the division between green and orange is greater in the perceptual sharpened image than in the original one. Although the printed images do not show a large difference it exists and it is very useful in segmenting colours. The second example show one case where there are some colours but they can not be intuited from neither the RGB nor the opponent RGB 2D–histogram. When the image is perceptually sharpened the opponent histogram show peaks belonging to the colours on the image, that do not appear in the original. The effects on the third example can be seen even in the RGB histogram. Two narrow peaks (those in red and yellow) show the localisation of the red and green leaves. In the last example the effects of the operator are weaker than in the previous images. However this is natural if we consider that the number of colours is large and their spatial location does not produce a colour contrast effect. In this case the operator does not spread colours but concentrates the distributions of colours.

4.4 Validation

While psychophysical test are not done we have validated the operators looking for indexes that show a better discrimination between colours. If the operators perform well the resultant images should be easier to segment in the predominant colours. Following the scheme in [46], we segment the image in two clusters and get a measure of how good this clustering is. A new segmentation is done with three clusters and the measure is calculated. If the previous measure is better than the new one we stop, if not the number of cluster is incremented and the comparison is done again until

scale ($\vec{\sigma}$)	$M_1^{ps} > M_1^o$	$M_2^{ps} > M_2^o$	$M_1^{ps} > M_1^o \wedge$ $M_2^{ps} > M_2^o$	M_1^{ps} imp.	M_2^{ps} imp.
high	82.32	67.68	57.93		
medium	73.78	66.46	52.44		
low	68.29	70.73	48.78		
all	87.80	86.59	76.22	102	87.2

Table 4.3: Region perceptual sharpening on VisTex image database. Values are in %.

scale ($\vec{\sigma}$)	$M_1^{ps} > M_1^o$	$M_2^{ps} > M_2^o$	$M_1^{ps} > M_1^o \wedge$ $M_2^{ps} > M_2^o$	M_1^{ps} imp.	M_2^{ps} imp.
high	73.78	78.66	58.54		
medium	75.61	75.61	59.15		
low	71.34	79.88	59.76		
all	86.59	90.24	79.27	38.06	46.94

Table 4.4: Local perceptual sharpening on VisTex image database. Values are in %.

process	$M_1^{ps} > M_1^o$	$M_2^{ps} > M_2^o$	$M_1^{ps} > M_1^o \wedge$ $M_2^{ps} > M_2^o$	M_1^{ps} imp.	M_2^{ps} imp.
SS vs None	99.39	85.98	85.37	159.4	97.1
ES vs None	87.80	86.59	76.22	102	87.2
T vs None	86.59	90.24	79.27	38.06	46.94

Table 4.5: Summary on perceptual sharpening on VisTex image database. Values are in %.

a maximum on the measure is reached. Instead of using a k-means algorithm as in [46] we used an *Expectation–Maximisation* mixture of gaussians which is more general and fits better the data. Both methods are briefly explained in section 5.2. There are a number of ways of measuring how good a clustering is, Coleman and Andrews enumerate some of them in [21]. In this validation experiment we have selected the following two:

$$M_1 \quad = \quad tr(\mathbf{S}_b)tr(\mathbf{S}_w) \tag{4.23}$$

$$M_2 \quad = \quad \frac{\mathbf{S}_b}{\mathbf{S}_w} \tag{4.24}$$

where $tr(\cdot)$ indicates "trace" or sum of the diagonal elements of a matrix, $\mathbf{S}_w$ is the within groups scatter matrix, a measure of how condensed the cluster is, and $\mathbf{S}_b$ is the between scatter matrix, a measure of the distance between clusters. The scatter matrices are defined in a better context in equations A.2 and A.3 in section A. We will use them here just as a tool.

The measure used should have a maximum when the best clustering is reached. Although M_2 is better when evaluating the dispersion of clusters, it is not upper bounded whereas M_1 is. We have used M_1 to iterate the clustering process and both M_1 and M_2 to measure the behaviour of the operators.

Another problem is that the parameter $\vec{\sigma}$ involved in the operators should be settled specially for each image, however to automatically find the best scale for each image is still an open issue that will derive form this thesis. What we will do is to try three different scales: high, medium and low, keeping the best clustering. The original image is also clustered using the same criterion. When M_i is applied on the original clustered image we will denote it as M_i^o and when done with the sharpened images M_i^{ps}, whichever it is the used operator. The sharpening is done on the two chromatic channels, the intensity is left as it is to show the computational chromatic contrast behaviour. The experiment is done on 164 images of the VisTex image database. Tables 4.2, 4.3 and 4.4 show the results for the SS, ES and T operator. The last two columns of the tables are the percentage of improvement of the measures wit respect to the measure on the original image.

From these results we can conclude that the operators are performing a good separation of colours on the image. And as we suspected, the more complex is the operator the better is its performance. The last row of the tables takes the best clustering in the three scales. Table 4.5 summarises the three operator to show their evolution.

4.5   Discussion

After an introduction to the colour induction phenomena it is concluded that computer vision lacks of an approach to the chromatic contrast effect. While there exists a computational model of colour assimilation, this is not the case for colour contrast.

Our contribution in this subject materialises in three new operators. The first one takes the traditional RGB sharpening operators to a space where colour appearance

is best modeled, adding spatial constraints to the generated responses. It has been illustrated it can work in some circumstances where the stimuli are small, and it can be adequate to very high frequency textures, but there are many situations where it does not fit well. This drives us to psychophysical literature that gives us the trail to search for inhibited and activated regions on the different visual pathways. Once more, there are many cases where the operator is useful but in some others it can not simulate the human visual perception. It was an inflection point to the search of a more general chromatic contrast operator. The result was the *Spread perceptual sharpening* operator that gather the experience in the preceding operators. The core of the operator is the idea of spreading the inhibition and activation of the cells on the transition between regions.

The capacity of the operators to differentiate colours has been tested on a texture image database, performing a segmentation and measuring how good it was compared to the image itself. The results obtained shows a good progression. Although the third operator is the more complete, the knowledge of the scene can advise to choose one of the other operators. This is a matter of complexity. Each operator can cope with more circumstances than the previous one but at the expense of computer resources.

There are open issues that have to be addressed in a near future and they are outside the scope of this thesis. The first one is to find the way to combine both spatial blurring and contrast induction in the same scene. One approach could be to look for different frequency regions in an image and applying the most suitable colour induction. On the other hand, we have presented a method that can be adjusted to viewing distance ($\vec{\sigma}$) and to the weight of each channel in the chromatic contrast effect ($\vec{\gamma}$). Both set of parameters have to be analysed from a psychophysical point of view, and then transferred to the computer vision field.

Chapter 5

Application to surface inspection problems

All the previous chapters define the necessary conditions to broach the problem of surface inspection, basically on industrial problems. This term includes a wide range of problems but we will centre in those where texture and colour simultaneously are the key factors in the problem solution. The work presented here began with an application devoted to the classification of coloured and textured samples of ceramic tiles. Nonetheless, the tools and methods used can be applied for many similar applications.

Considering the perceptual operator introduced in the previous chapter, here we propose a computational colour texture representation based on a multiscale approach. In it, colour measurements are done considering the perceptual blurring that simulates a large distance observer position, and the proposed sharpening to simulate a short distance observer position, from which the blob extraction is derived. From these blobs the features of textons associated to colours are defined.

Finally, we will apply this approach to the classification of ceramic tiles, and to the problem of printing quality classification.

5.1 Building a colour texture representation

In chapter 1 we have briefly introduced previous works on colour, texture and colour-texture computational representations. In chapter 4 we have explained induction phenomena of the human visual system that explains the interaction between these two visual cues when appear jointly on a surface.

Now we want to build a computational colour-texture representation that considers the induction phenomena. To this end, we will follow a common approach in computer vision that is building a feature vector that combines different image properties, but in this case we will make it to take into account the induction phenomena that are involved in the human perception of colour textures.

In figure 4.4 we have plot a model proposal to integrate the most common effects

of colour induction: assimilation and contrast. We have seen that the first can be computationally simulated by a perceptual blurring and the second can be computationally simulated by a perceptual sharpening. Their activity is complementary, the first one is produced when the spatial frequency is high and the second one is produced when the spatial frequency is low. This scale-dependent mechanisms are very common in multiscale approaches in computer vision, and it allows to extract different information of a given image as it is implementing a vision process of looking at the image from different observer positions or looking at different image regions in a more attentive process.

The colour assimilation allows to take a global colour measurement of an image region when observed from a long distance position and when blob details are lost. The colour contrast phenomena allows to take local measurements of the image blobs when observed separately, that is, from a short distance position. We have seen in chapter 1 the need for these two types of measurements, when we propose a colour-texture representation. In this case, and to consider colour and texture interactions we will define the measurement on a set of preprocessed images that will simulate the induction phenomena. These sets of images are constructed in the following steps:

Step 1 A given input image, I, has to be transformed to its opponent colour representation, as it has been seen in chapter 4, we will represent as $Opp(I)_i$ the i channel of the opponent representation of I.

Step 2 An assimilation process is simulated by building the following set of images:

$$\{A(Opp(I)_i, s)\} \tag{5.1}$$

where i represents the image channel and s represents the scale of the perceptual blurring, $A(I, s)$, that represents the convolution of the image I with a kernel defined by the s parameter. In this case, the convolution kernels can be built accordingly for different observer conditions, by considering the psychophysical measurements presented in the Spatial-CIELAB introduced in section 4.2.1. In this case the selected scale will be directed to tune the high spatial frequency image relationships, these set of images will be the basis for global measurements of the image colour.

Step 3 A chromatic contrast process is simulated by building the following set of images:

$$\{C(Opp(I)_i, s)\} \tag{5.2}$$

where i represents the image channel and s represents the scale of the perceptual sharpening, $C(I, s)$, which can be implemented by the spread sharpening defined in section 4.3.3. In this case we do not have the psychophysical measurements that can provide the needed parameters to represent the conditions for a given scale. Therefore we will only use this set of images for the blob segmentation step, but not for colour measurements.

Step 4 The results of the previous steps put the basis for a consistent blob segmentation step based on colour properties. The blob segmentation process will be explained in the following section. The set of k segments of a given sharpened image will be denoted as:

$$\left\{C^{sg}(I_i, s)\right\}_{sg:1...k} \tag{5.3}$$

Once, the colour induction phenomena have been simulated on a set of blurred and sharpened images, now we want to measure colour and texture properties of these images.

In the next sections we will explain how to perform a blob segmentation based on colour properties, afterwards, we will define the global and local measurements that are usually computed for colour and texture representation, and finally we will build the complete computational colour-texture representation we want to propose.

5.2 Perceptual blob segmentation

In the representation of the colour texture that we propose, the blobs play an important role in the description and they should be treated carefully. In the next section we will describe the possible mechanism to obtain the set of images where which content are blobs of similar colour.

The first step to describe colour texture is to separate pixels on the image that share similar colour properties. This step can be seen as a colour-granulometric stage, where we sieve pixels of different colours. Given an image, the output of this segmentation step is a set of images containing the blobs of a specific colour. Although there are many ways to segment images most of them are based on its spatial contents (*region growing, split and merge*, etc) and they are not what we want. Our segmentation has to be independent of the spatial relations between colours. The objective is to return a set of sets of pixels where in each of these set the colour variance is minimum independently of the pixel location. From the segmentation methods we will focus on clustering methods, and from these ones on the bayesian approach.

We can do this as a supervised operation in which approximate colour coordinates are introduced to initialise each colour centre, or in unsupervised mode starting from algorithmically chosen centres. In the second case the clustering algorithm can have non–deterministic solutions, depending on the way these centres are chosen. However most of the times these algorithms tend to the same solution. This point should be kept in mind when working with colourful images.

The basic idea of cluster algorithms is that given an image $\mathbf{I}$ with c different colours, and given a set of initial colours $\{\mathbf{C}_1, \ldots, \mathbf{C}_p\}$ the segmentation step output is a set of images:

$$\{\mathbf{I}_1, \ldots, \mathbf{I}_c\}$$

where each image $\mathbf{I}_i$ contains all the blobs of a specific colour labeled i. Although there are many algorithms to cluster colour data, we have selected two of them for being widely used in computer vision. One of the methods is chosen for its simplicity

and the other one is based on the same idea but a more complex model is behind it. In the last case not only blobs are clustered but also it gives statistical information of its colour content that can be useful for posterior colour classification. These algorithms are the well known *k–means* and *expectation maximisation mixture of gaussians*. In both cases the number of clusters to obtain must be set a priori. To solve this problem a simple method is defined in [46] to analyse how many colour clusters appear in the image. In all of our cases the number of clusters is known since we have information of the production process or we know what we are looking for. We will summarise both clustering methods briefly that are described in detail in [26, 10].

The point of departure in any case should be an image where clustering is helped by its content i.e: the much separate the colours are the easier is the problem of clustering. If we manage colour separation by means of linear transform the problem remains the same as the proportions are maintained. So a non linear transform is needed. In chapter 4 we have introduced the colour contrast perceptual operators. In a normal situation they act to simulate human behaviour when looking at low frequencies patterns. In our examples we have high frequency texture, however we use the spread perceptual sharpening operator (Eq. 4.22) to simulate a human being looking at the sample at very high resolution: In this case colour contrast conditions operates and colour blobs are easier to segment. Figure 5.1 explains this effect where an original sample with their 2D–histogram distribution in the RGB and opponent space are depicted in the left column, and when applying the SS operator (right column) the colours in the image appear more clearly. We are not saying that colour contrast is the effect that helps humans to distinguish colour blobs, in fact if any process is done it is assimilation, but in order to better cluster colour data to describe how colour texture is, humans perform better at high resolution.

5.2.1   K-means clustering

Given a set of n–dimensional data, the goal of the k–means clustering is to find the centres of the colour cluster $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_c$ to be used as their prototypes, where c i the number of clusters that has been set a priori.

Starting form a set of initialised cluster prototypes $\Theta = \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k$ we want to know which is the probability for one pixel $\boldsymbol{x}_i$ to belong to the cluster ω_k, $P(\omega_k \mid \boldsymbol{x}_i, \Theta)$. The algorithm works with the assumption that if the distance between $\boldsymbol{x}_i$ and $\boldsymbol{\mu}_k$ is small the probability is large. Then, we compute the square Eculidean distance between the pixel and all the cluster prototypes, $\| \boldsymbol{x}_i - \boldsymbol{\mu}_p \|^2$. Let $\boldsymbol{\mu}_m$ be the nearest prototype to $\boldsymbol{x}_i$ then, we approximate $\hat{P}(\omega_k \mid \boldsymbol{x}_i, \Theta)$ as

$$\hat{P}(\omega_k \mid \boldsymbol{x}_i, \Theta) = \begin{cases} 1 & : \quad k = m \\ 0 & : \quad \text{otherwise.} \end{cases} \qquad (5.4)$$

Following this schema each pixel of the image is set to belong to one cluster and from the pixels in each cluster the prototypes $\boldsymbol{\mu}_k$ are recalculated. This process is done until some stop condition is reached. This condition could be the number of iterations to do or the stability of prototypes of a minimum threshold to reach of the RMS error.
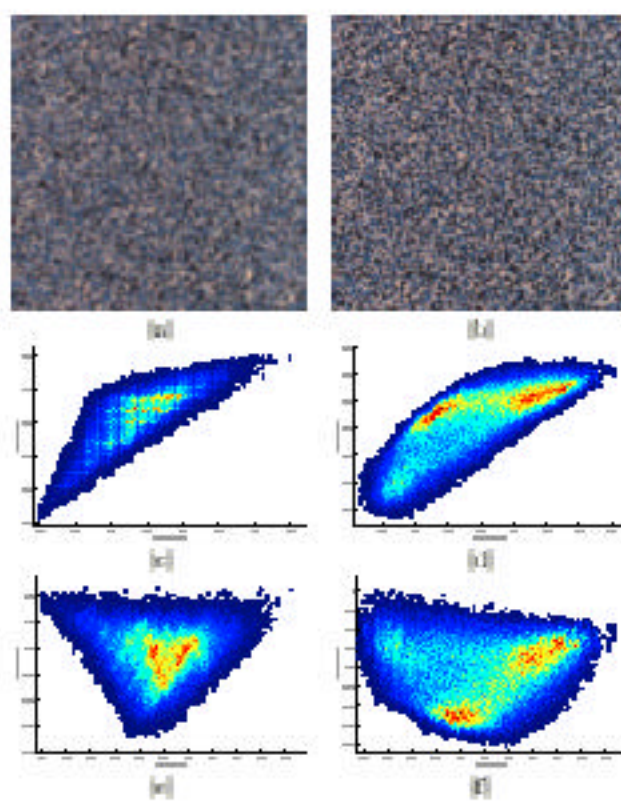
Figure 5.2: Colour distributions of a colour texture image: (a) and (b) original and spread perceptual sharpened result. (c) and (d) the 2D-histogram rejecting one of the dimensions in the RGB space. (e) and (f) 2D-histogram in the opponent space.

Therefore, *k-Means* algorithm is performing a clustering step over the colour space, where clusters are formed considering the euclidean distance between colour coordinates of pixels in this space. Given that we are using an euclidean distance on a colour space we should do it on a perceptually uniform colour space as CIELAB or CIELUV. Some experiments have been done in this sense and they make us to conclude that, for the images of our applications, the segmented blobs are nearly the same, hence, there is no need to introduce a step that needs further calibration and does not improve the clustering result.

5.2.2 Parameter estimation of the colour distribution

In this section we assume the colour distribution as the sum of several gaussian distributions. To get an estimation of the parameters of these gaussian we will use an *Expectation–Maximisation* method.

This method assumes that data are generated from a set of gaussian distributions that when mixed form the final distribution. Its goal is to extract the parameters of the gaussians that best describes the distribution form the data themselves. It is done in a two steps process were in the first and starting from an initial guess an expectation of which are the best parameters is done. The second step takes the set of all possible parameter modifications and chooses the one that maximises the fitting of data with the mixture. With this new guess another iteration is done until some condition is reached (as in the case of k–means procedure). The degree of data fitting is measured in terms of likelihood $\mathcal{L} = \prod_{n=1}^{N} p(\boldsymbol{x}_n)$, where data are N points, $\boldsymbol{x}_n$, and $p(\boldsymbol{x}_n)$ is the a priori probability of a given point to happen. But to reduce the complexity of the problem the log–likelihood is used and then it becomes a minimisation problem of the expression

$$E = -\ln \mathcal{L} = -\sum_{n=1}^{N} \ln p(\boldsymbol{x}_n) \qquad (5.5)$$

$p(\boldsymbol{x})$ is defined in term of the probability of belonging to the gaussian distribution of the mixture given a probability of each particular gaussian to occur, $p(\boldsymbol{x}) = \sum_{j=1}^{k} p((x) \mid j)P(j)$. In the case of mixture of gaussian and for the case of diagonal covariance to simplify the problem, the probability of a given d–dimensional point to belong to a certain gaussian distribution j is

$$p(\boldsymbol{x} \mid j) = \frac{1}{(2\pi)^{d/2}\sqrt{\mid \Sigma_j \mid}} e^{-\dfrac{(\boldsymbol{x} - \boldsymbol{\mu}_j)' \Sigma_j^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_j)}{2}}. \qquad (5.6)$$

Then we can take derivatives on the unknown parameters of the gaussians maximise the likelihood. In this case $\frac{\partial E}{\partial \boldsymbol{\mu}_j}$ and $\frac{\partial E}{\partial \boldsymbol{\sigma}_j}$ for all j he details can be found in [10]. The a priori probability of a certain gaussian $P(j)$ has to be also derived to maximise $\mathcal{L}$. As our intention is jut to pose the method we will not focus in the mathematical description. From this maximisation step we obtain a set of parameters that define the configuration of the mixture, $\hat{\boldsymbol{\mu}}_j$, $\hat{\boldsymbol{\sigma}}_j$ and $\hat{P(j)}$ that are the initial guess for the next iteration.
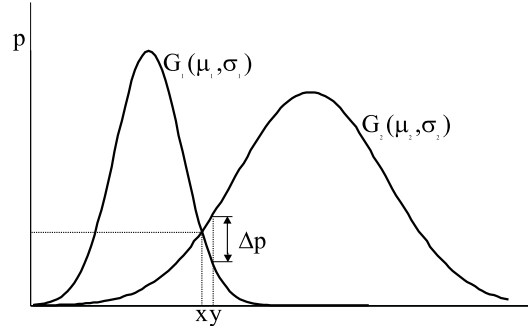
Figure 5.2: Decision criterion: when a point x can belong to more than one gaussian distribution it is rejected. In the example, y will be set as a point in G_2 because its differential to G_1 is large enough.

5.2.3   Decision criterion for clustering

When using mixture of gaussians the pixel $\boldsymbol{x}$ is classified as belonging to the cluster i with maximum posterior probability $p(i \mid \boldsymbol{x}, \Theta_i) = p(\boldsymbol{xx} \mid i, \Theta_i)P(i)/\sum_{j=1}^{c} p(\boldsymbol{x} \mid j, \Theta_j)P(j)$, where Θ_i are the parameters of the i–th gaussian. The original method does not take into account that some times a pixel can belong to two or more different gaussians with nearly equal probability. It is the case of point x in figure 5.2. Those pixels with this ambiguity should not be considered to avoid colours to shift from one cluster to another one. We have introduced a criterion to ensure that pixels are unambiguous. The idea is quite simple, instead of assigning a pixel to the class with the prior criterion, the posterior probabilities for each class are sorted and if the difference between the two maximum posteriors is greater than a certain percentage Δp then the pixel is assigned to the first cluster, in any other case the pixel is keep away in an *ambiguous cluster*. Then, $\mathcal{C}(\boldsymbol{x}, \Theta)$ as the expression to classify a pixel to a cluster i from a set of gaussian parameters $\Theta = \Theta_1 \ldots \Theta_c$ is:

$$\mathcal{C}(\boldsymbol{x}, \Theta) = \begin{cases} i & : \; p(i \mid \boldsymbol{x}, \Theta_i)(1 - \Delta p) \geq p(j \mid \boldsymbol{x}, \Theta_j) \quad \forall j = 1 \ldots c \wedge i \neq j \\ ambiguous & : \; \text{otherwise} \end{cases}$$

$$(5.7)$$

To finish this section we show an example of the two cluster techniques applied both to the input image an the perceptually sharpened image of figure 5.1(a) and (b). If the images are not reproduced with a high quality printing device the colour differences perhaps will not be seen and the attention should focus in the shape of blobs. In figure 5.3 we have applied the k–means method, and a small portion of the central part of the image is shown for better detail. (a) is the detailed input image. (b), (c) and (d) are the three clusters generated using this technique to the input image as it is. The last row are the respective clusters when the input data to the clustering algorithm is the spread sharpened image. Small differences can be appreciated but they can be crucial in describing textures based on blobs
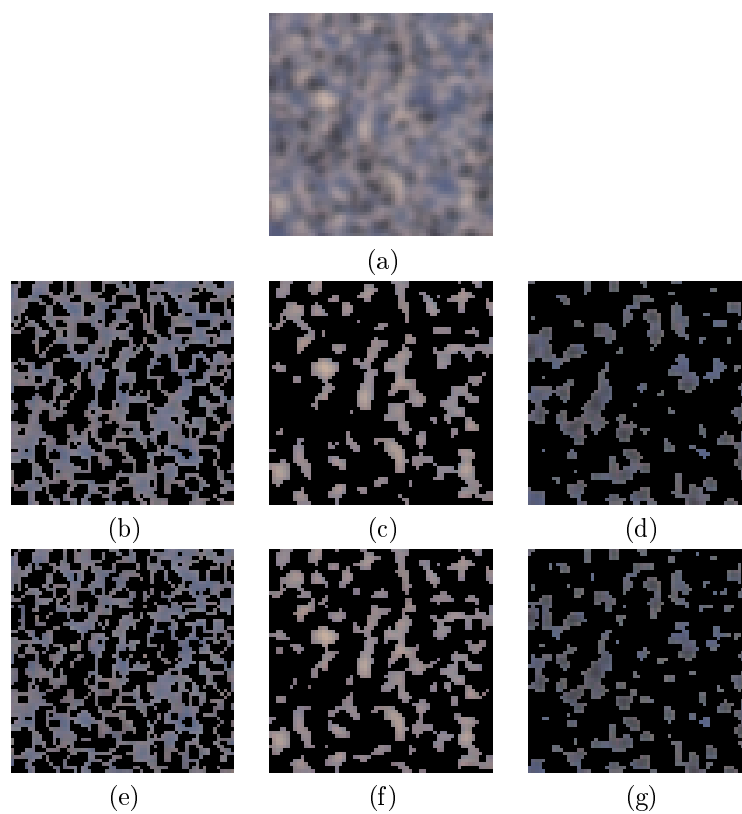
Figure 5.3: K–means clustering example: (a) is the central area from the original image to which clustering is performed. The following row is the result of segmenting in 3 clusters on the original image. Last row when applied to the sharpened image.
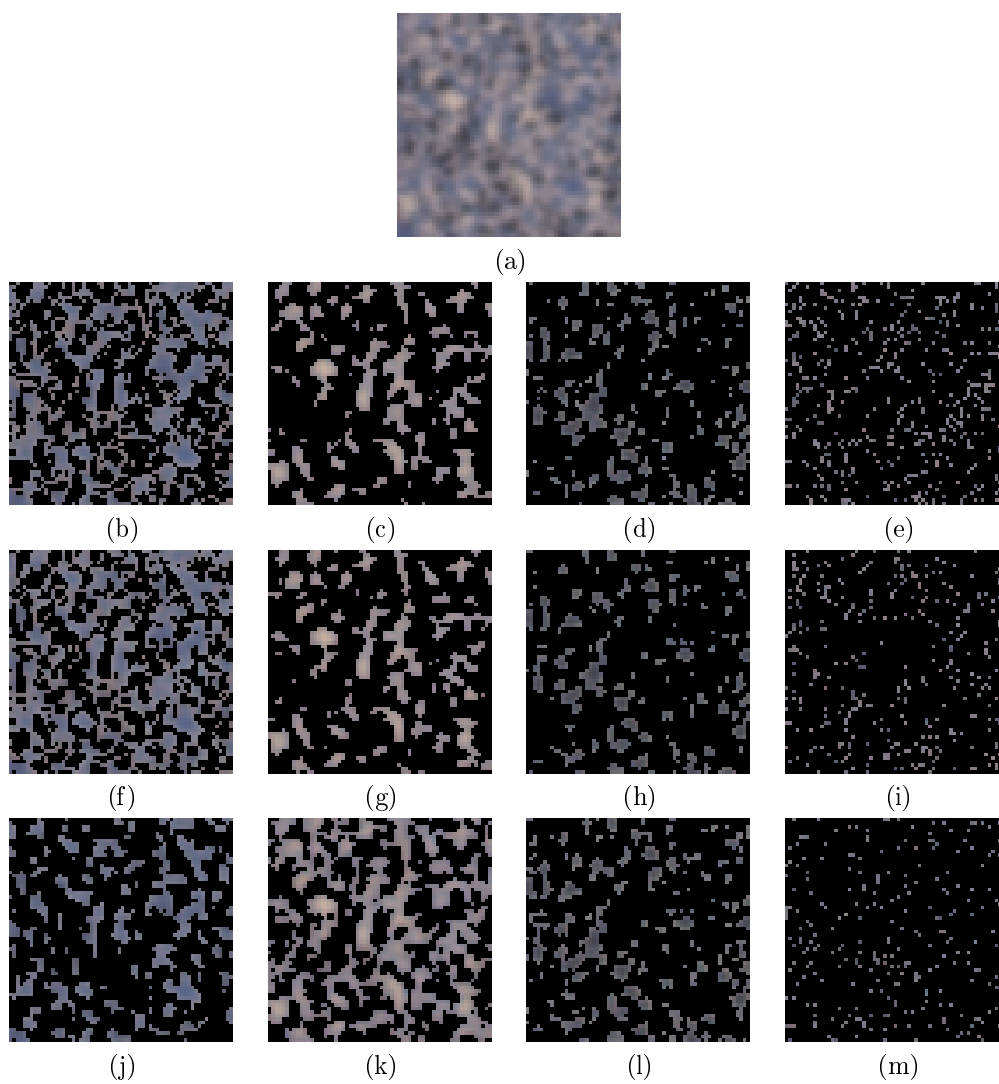
Figure 5.4: Mixture of gaussians clustering example: (a) is the central area from the original image to which clustering is performed. The following row is the result of segmenting in 3 clusters on the original image. The middle row when applied to the sharpened image. The last one when using automatic supervised clustering.

characteristics. In the middle and right cluster, the less populated, some blobs that are merged in the first row are well segmented in the last one. One of the advantages of this method is its simplicity and low computational complexity, however the result are not very good as colours rarely group in spheres. It is more likely to find colour distributions resembling gaussian distributions, and so mixture of gaussian fit well when it is the case. The following picture is from the same image but applying the expectation maximisation mixture of gaussian distributions. Figure 5.4 are the resultant images whit this clustering method. The first row of images contains the image of clusters from the original image and the last image of the row with those points that are classified as *ambiguous cluster* following equation 5.7. It has already improved the clustering without using the perceptual sharpening. The second row of images has the same configuration being the results of the sharpened image. Some characteristics are remarkable, the number of ambiguous pixels reduces in this case and like in the k–means algorithm blobs are better segmented in the case of dark and light blobs. Blue blobs are somewhat messy and is difficult to appreciate improvement. Even in this case when analysed carefully the blob boundaries are better. The last row introduces a new approach to the clustering. As it has been commented these techniques can be supervised initialising the first guess on the centres of the clusters, normally it is done manually. This supervision can be done automatically if some maximum localisation criterion is used. However, in our case the colour data forms a 4–dimensional distribution surface. To construct such surface is very time and space consuming and it is unfeasible. The approach used is to locate the desired number of maximums in a 3D space projecting one of the dimensions and use it to find compatible maximums in the other projections. These maxima are the input centres of the mixture of gaussians. To control the behaviour of the EM algorithm we only iterate on the covariances, fixing the mean of the gaussian distributions. Applying this criterion we obtain the results of the third row of figure 5.4. The density of blobs is more equilibrated and principally, the blue cluster contains more homogeneous colours, and the blobs are more accurate. The light blobs are the ones which receive the colours that are not included in the blue cluster. Nevertheless the inhomogeneity is not very large, in fact when looking at the original image white points are more dispersed. It should be noted that this criterion can not be done in the original image because there not exist clear maxima, whereas in the sharpened image it does.

5.3   Global features

With the following global colour measurements we try to capture a first coarse description of the image using basic statistics. For a given image $\mathbf{I}$ of size $N \times 3$ where each row is a colour triplet and p different types of blobs we will derive from the clustering step the images $\{\mathbf{I}_1, \ldots, \mathbf{I}_p\}$.

Global colour mean: A global colour measurement of the whole image:

$$M_1(\mathbf{I}) = \bar{\boldsymbol{I}} = \frac{1}{N}(\boldsymbol{I}'\mathbf{1}) , \qquad (5.8)$$

where $\mathbf{1}$ is the constant 1 vector of dimension N.

Global colour variance: A global colour measurement of how constant the colour
on the whole image is:

$$M_2(\mathbf{I}) = \frac{1}{N-1} diag((\boldsymbol{I} - \bar{\boldsymbol{I}})'(\boldsymbol{I} - \bar{\boldsymbol{I}})) \tag{5.9}$$

where $diag(A)$ is the diagonal vector of the matrix A.

Obviously these two measures describe very coarsely the colour contents of all the
surface.

5.4 Local features

Now we will give some measurements on local properties of the blobs. We are defining
the measures to get a more detailed description of the geometry and distribution of
colour textons. Most of them are based on the central moments of inertia ([60]),
which are defined as:

$$m_{p,q} = \sum_{(x,y) \in \mathcal{R}} (x - \bar{x})^p (y - \bar{y})^q \tag{5.10}$$

where $\mathcal{R}$ is the region (or blob) of the image where the moment is calculated, and
$\bar{x} = 1/n \sum_x x$ and $\bar{y} = 1/n \sum_y y$ is the centre of gravity with n the total number of
pixels in the region. The moments are calculated for each blob in each segmented
image. Next, we enumerate the different moments proposed to the description of the
form of textons. Although there are many more possible, using high order moments
will capture information on blob geometry, which is usually useless when dealing with
a large amount of blobs. All this measures are over a binary mask of the colour
segmented images.

Blob area It is the moment of order (0,0) that simplifying is the count of the pixels
of a region.

$$a(\mathcal{R}) = m_{0,0} = \sum_{(x,y) \in \mathcal{R}} 1 \tag{5.11}$$

Blob eccentricity A measure of how rounded a region is. When the blob is line–
shaped the value of this measure, ε, is 1 and 0 when it is circular.

$$\varepsilon(\mathcal{R}) = \frac{(m_{2,0} - m_{0,2})^2 + 4m_{1,1}^2}{(m_{2,0} + m_{0,2})^2} \tag{5.12}$$

Elongation A measure proportional to the elongation of the object taken in the di-
rection which maximises the measure. It can be set as the maximum eigenvalue.

$$\lambda_{max}(\mathcal{R}) = 2\sqrt{\frac{m_{2,0} + m_{0,2} + \sqrt{(m_{2,0} + m_{0,2})^2 + 4m_{1,1}^2}}{2m_{0,0}}} \tag{5.13}$$

Orientation The angle, ϕ, between the x axis and the axis around which the blob can be rotated with minimum inertia which is given by the eigenvector to the minimal eigenvalue.

$$\phi(\mathcal{R}) = \arctan \frac{2m_{1,1}}{m_{2,0} - m_{0,2}} \tag{5.14}$$

Instead of recording all the values for each blob in each image the information is reduced to the mean and standard deviation of the measures. For every segmented image $\mathbf{I}_i$ we obtain a set of parameters to define the overall form of blobs in it:

$$
\begin{aligned}
M_1^i &= \frac{1}{N^i} \sum_{b \in Reg(\mathbf{I}_i)} a(b), \\
M_2^i &= \frac{1}{N^i} \sum_{b \in Reg(\mathbf{I}_i)} \varepsilon(b), \\
M_3^i &= \frac{1}{N^i} \sum_{b \in Reg(\mathbf{I}_i)} \lambda_{max}(b), \\
M_4^i &= \frac{1}{N^i} \sum_{b \in Reg(\mathbf{I}_i)} \phi(b)
\end{aligned}
\tag{5.15}
$$

which are the mean values, being $Reg(\mathbf{I}_i)$ the set of blobs in the segmented image $\mathbf{I_i}$, and N^i the number of blobs in the image. And the measures standard deviations:

$$
\begin{aligned}
M_5^i &= \frac{1}{N^i - 1} \sqrt{\sum_{b \in Reg(\mathbf{I}_i)} (a(b) - M_1^i)^2}, \\
M_6^i &= \frac{1}{N^i - 1} \sqrt{\sum_{b \in Reg(\mathbf{I}_i)} (\varepsilon(b) - M_2^i)^2}, \\
M_7^i &= \frac{1}{N^i - 1} \sqrt{\sum_{b \in Reg(\mathbf{I}_i)} (\lambda_{max}(b) - M_3^i)^2}, \\
M_8^i &= \frac{1}{N^i - 1} \sqrt{\sum_{b \in Reg(\mathbf{I}_i)} (\phi(b) - M_4^i)^2}
\end{aligned}
\tag{5.16}
$$

In some cases it can be useful to get a more fine description of the distribution of a certain measure. Then, instead of using the mean and the standard deviation we will use an small dimensional histogram of the parameter in question. For example, we have used in one of the cases that we will present in section 5.6, the histogram of blob areas for each segmented image. As we knew a priori which were the usual size of the blobs we divide them in four bins: very small, small, medium and large blobs. When the texture is thought to be described by a certain parameter it is straightforward to obtain its histogram. The small histogram will be denoted as the feature M_9^i where the number of bins is specific for each problem.

As we want to describe colour textures, a representation of the colour of each cluster is needed. Three more parameters are extracted from each image $\mathbf{I_i}$. We compute the mean colour of all the pixels in each cluster as $M_1(\mathbf{I_i})$

Although strictly speaking only M_1, M_2 and $M_1(\mathbf{I_i})$ are measurements on colour, all the other measures can be considered to extract information of similar colour blobs, defining, separately, how the texture of colours are. The colour space where the colour measurements are done is not specified. Any suitable one can be selected whereas it captures the information that is wanted. As in high frequency patterns colour assimilation is one of the main factors to perceive colour, a perceptual blurring as defined in [12] and introduced in section 4.2 is feasible. With this approach the distance from the scene to the observer can be modeled by a few parameters.

5.5 Proposal for a perceptual colour texture representation

So far we have introduced the methods to isolate the blobs on the image that form the texture based on the colour information. Colour induction has been used to enlarge colour differences and help to a better segmentation of coloured blobs. It is done at different scales to simulate the human process of attentive vision. When done at high frequencies perceptual blurring is performed, and at low frequencies we apply perceptual sharpening.

The second stage has been to define a set of measures to capture global an local features on the colour and texture of the image. At this point there is no connection between both stages.

Our proposal is to merge the two previous points in a single feature vector representing the colour texture descriptor of the image. The schema of this proposal is presented in figure 5.5. We will take the global measures from the set of images obtained from the assimilation process, and the local measures from the blob segmentations using different scales on the contrast process. The number of segments used in the local features will depend on the image content and, when available, on the a priori information of the problem.

From the schema of colour assimilation and contrast a set of images are obtained that have to be recombined to be the input to the feature extraction. Given a scale s_p, where $1 \leq p \leq n$,(i.e.: assimilation conditions) the outputs $A(Opp(I)_c, s_p)$ for $c = 1 \ldots 3$ are combined in a tristimulus image $A(I, s_p)$, that is, the image obtained after an assimilation process at observer conditions defined by s_p. For all the set of images obtained at the different scales for assimilation we compute the global feature vector composed by M_1 and M_2:

$$\{M_f(A(I, s_p))\} \;\; p = 1 \ldots n, \;\; f = 1 \ldots 2 \tag{5.17}$$

A similar process is done with the outputs from the perceptual sharpening, the images $C(Opp(I)_c, s_p)$ for $c = 1 \ldots 3$ and $n+1 \leq p \leq m$, are combined in a tristimulus image $C(I, s_p)$ and then a set of k segmented images is obtained from the clustering process, $\{C^{sg}(I, s_p)\}$ for $sg = 1 \ldots k$. At this point we have $k \times (m - n)$ images as the result of the clustering process on the $m - n$ scales of the perceptual sharpening.

A: *Assimilation Operator*
C: Contrast Operator

$A(Opp(I)_1, s_1)$

$\vdots$

$A(Opp(I)_1, s_n)$

$C(Opp(I)_1, s_{n+1})$

$\vdots$

$C(Opp(I)_1, s_m)$

$A(Opp(I)_2, s_1)$

$\vdots$

$A(Opp(I)_2, s_n)$

$C(Opp(I)_2, s_{n+1})$

$\vdots$

$C(Opp(I)_2, s_m)$

$A(Opp(I)_3, s_1)$

$\vdots$

$A(Opp(I)_3, s_n)$

$C(Opp(I)_3, s_{n+1})$

$\vdots$

$C(Opp(I)_3, s_m)$

$Opp(I)_1$

$Opp(I)_2$

$Opp(I)_3$

I

Global measures

$$\left\{M_f(A(I, s_p))\right\}_{\substack{p:1\ldots n \\ f:1\ldots 2}}$$

Blob segmentation

$$\left\{C^{sg}(I, s_{n+1})\right\}_{sg:1..k}$$

$$\vdots$$

$$\left\{C^{sg}(I, s_m)\right\}_{sg:1..k}$$

Local measures

$$\left\{M_f^{sg}(C^{sg}(I, s_p))\right\}_{\substack{sg:1\ldots k \\ p:n+1\ldots m \\ f:1\ldots 9}}$$
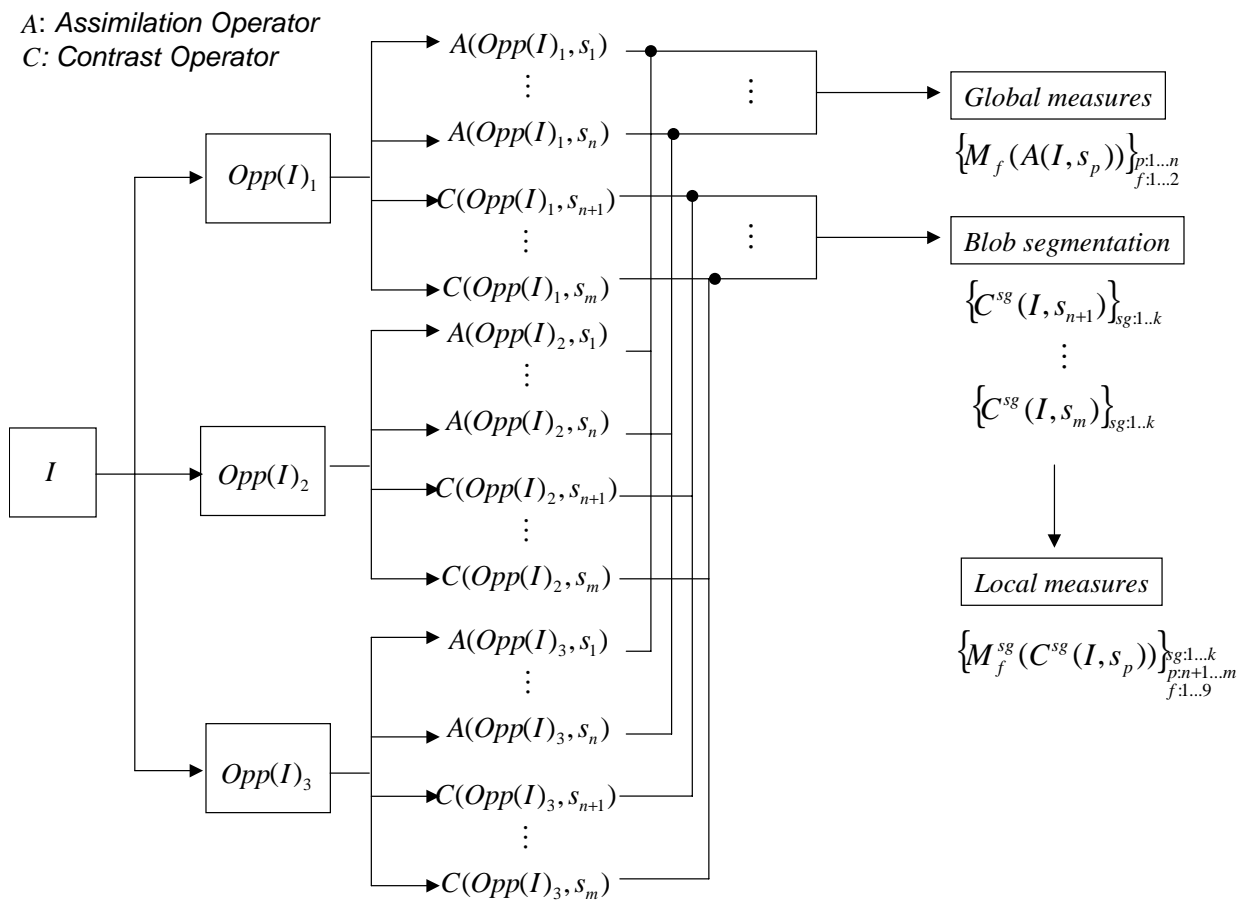
Figure 5.5: Colour–texture representation model.

<div align="center">(a)   (b)</div>
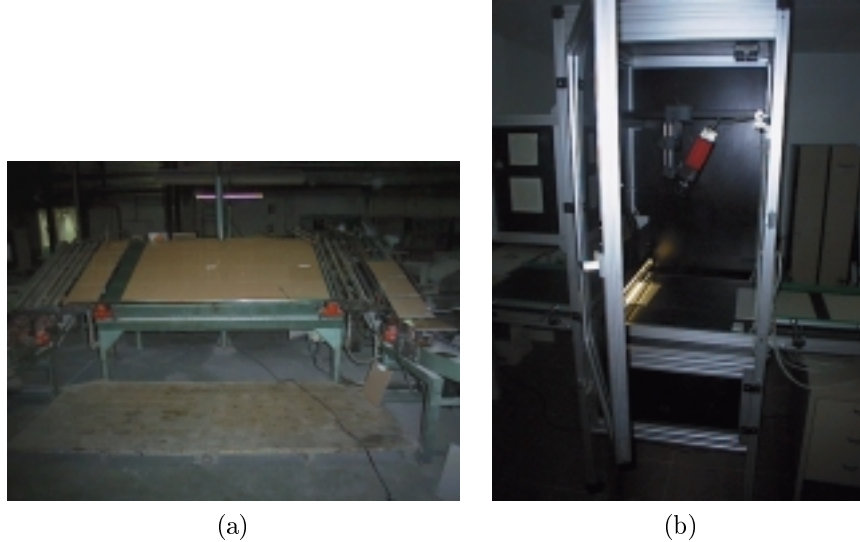
Figure 5.6: Tile classification system: (a) Real conditions in which human operators classify the production. In (b) the off-line classification system designed for this purpose.

From those images the local feature vector is obtained. The features included are those defined in section 5.4, from M_1^i to M_9^i:

$$\{M_f^{sg}(C^{sg}(I, s_p))\} \text{ where } sg = 1 \ldots k, \; p = 1 \ldots n, \; f = 1 \ldots 9 \qquad (5.18)$$

This is a general framework for colour–texture representation, but in each problem the parameters k, m, n, and the desired features f have to be defined to adjust to the knowledge of the problem.

From now on we will describe two applied cases were colour texture is the key factor to analyse the problem. The first case is the classification of ceramic tiles were a representation of the texture is needed to differentiate between classes with small differences in colour and/or texture. The second case is the quantification of printing quality from the appearance of the texture in homogeneous ink patches.

5.6   Case 1: Ceramic tile classification

5.6.1   The problem

In this section we treat a specific problem of classification of polished ceramic tiles. Tile manufacturing needs of pigments and clay which are mixed, melted, sprayed to form the tile substrate, and finally baked. This is a high quality tile whose production can be affected by external factors that are difficult to control, such as humidity, temperature, pressure conditions, origin of clays and colour pigments. Changes in

any of these parameters provoke subtle visual variations of the tile aspect when tiles are placed on the floor, one next to the other. These visual changes are due to small alterations of colour and texture properties of the tiles. It forces an on line classification of the production. At present, the classification is done by human experts, and it always involves subjectivity and loss of repetitiveness. In figure 5.6(a) we can see the place devoted to classify the production. In each production line only one model of tiles is produced. Thus, the classification must be done among classes of each model and not among models. During one day production up to eight classes can be created. A correct and non–subjective ceramic tile classification would allow to avoid returns from customers and to optimise the storage of the production stock reducing stock fragmentation. Previous research in computer vision techniques has contributed with interesting works on this problem [12, 83]. Now we will work on it with our colour texture description approach.

Previous works on the same application only use colour information for the classification task. In one of them [83] the first and second order moments of the RGB histograms are computed as colour and texture measurements, respectively. The second work [12] is based on three-dimensional histograms over the RGB space. The classification process extracts a similarity measure based on the Pearson correlation coefficient between 3D histograms. None of them compute blob measurements.

5.6.2   Human criteria for tile classification

Nowadays, this classification task is performed by specialised workers requiring a training period before to do it. One worker is replaced from the production line every two hours in order to avoid fatigue. It is such a subjective task that two different people can disagree in classifying the same sample. However, they have developed their own jargon to speak about tile differences. With the collaboration of a company of this industrial sector we did an experiment with human operators om order to get maximum information on hoe they do the classification.

Firstly, we asked classification experts to list the vocabulary that has been the basis to develop this work. The following list presents the characteristics they look at:

1. Fine–grained vs. coarse–grained: It is an obvious feature that defines the size of the grains.

2. Opened grain vs. closed grain: it is a measure of the distance among grains of the same size that could be intuitively interpreted as some density factor.

3. Light vs. dark grain colour: The colour properties of a specific type of blobs.

4. Light vs. dark background: The colour properties of the tile background. For some tile models with an important difference between the amount of every colour can provoke a predominant colour and a secondary colour. The first one will be called the background and will the blobs froming this bakground will not be considered.

5. Light vs. dark global colour: The colour properties of the overall colour impression due to the interaction between background and grains or between all grains.

Two tests were done with a group of trained people to search for the motivations they make conclude a certain classification.

- For a set of seven different classes of the same model and for two models they were asked to enumerate and quantify which were the features that, from their point of view, allowed to discriminate between a pair of classes. When describing these differences several tiles of each class were used.

- For each class the expert had to determine which were the three most similar classes.

The conclusion in the first experiment was that they do not focus on the interaction between colours but in the global appearance of tile and in local features of each colour. In the second case it was clear that they admit that some classes can be mixed without a very high inter–class difference, and that when a class can be confused with another one the behaviour is usually symmetric. From the models used in the experiments and from the experience of human operators, it was agreed that human classification is more difficult as the number of colours increases.

5.6.3 Preliminary approach

Although there exist some commercial systems that claim they cope with the ceramic classification problem only the works in [83] and [12] are documented, the others only have a short brochure without any indication of its classification rates and the methods they use. In both cases global colour information has been used, although the influence of the texture has been considered by simulating an assimilation step in the second one.

Our first attempt is to check if using only global colour information is enough or something else is needed.

Similarity measurement

In computer vision the problem of defining similarity measurements has been widely studied for the object recognition task. In [27] it is argued that the recognition is mainly based on distances on a small dimensional space and on the definition of prototypes which can define a class of objects.

In this work, as in [13] we will use the linear correlation coefficient (*Pearson's r*) between image histograms to compare the tiles. Then we define D as a similarity measurement:

$$D(r,s) = 1 - \left| \frac{\sum_i (H(I^r)_i - \overline{H(I^r)})(H(I^s)_i - \overline{H(I^s)})}{\sqrt{\sum_i (H(I^r)_i - \overline{H(I^r)})^2} \sqrt{\sum_i ((H(I^s)_i - \overline{H(I^s)})^2}} \right| \tag{5.19}$$

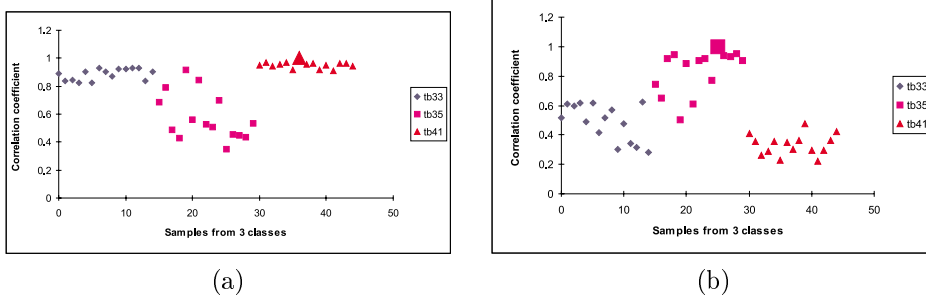<div align="center">(a)   (b)</div>

Figure 5.7: Classification of the same tile set using two different reference tiles: (a) Reference tile is 36. There is no possibility to discriminate between classes tb33 and tb41. (b) Reference tile is 26. Only class tb35 can be discriminated in relation to the other two.

where I^r and I^s are images to be compared, $H(I^j)$ is the histogram of the I^j image and $H(I^j)_i$ denotes de number of pixels of the I^j image having colour i, where i is a triplet of red, green and blue values which range from 0 to 2^b, being b the number of bits used to digitise the image. $\overline{H(I^r)}$ is the histogram mean.

So that, the correlation coefficient helps us to get a similarity value between two different tiles. Every tile is defined by its own histogram. If $D(r, s)$ value is near to 0 it means that images I^r and I^s belong to the same class, on the contrary, a value near to 1 signifies images are from two different classes.

Classification by similarity

The use of similarity measures to classify involves the election of the representative samples for every class of tile. Usually, they are arbitrarily selected among those whose classification is known. This can cause some problems since an algorithm can obtain different results depending on which is the sample used as prototype.

In Boukouvalas's work [13], the classification uses one reference sample. The algorithm classifies based on thresholds on the correlation coefficients between the sample and the reference tiles that have been taken. Applying this method to our samples makes us realise on how important is the choice of the reference sample. Changing the reference sample can vary the capability to discriminate among classes. An example of this problem is shown in figure 5.7. The x axis is the number of sample and the y axis is the distance $D(r, s)$ from each sample to a reference sample, which is different in the first and second plot.

To avoid the above mentioned problem and to get a classification less dependent on the reference chosen, we propose a new method to select the prototypes of each class. N images of each class of tiles were taken to elect the representative sample. From them we obtained a similarity matrix calculated using the measure D between the three–dimensional histograms of all of them.

To select the most representative images for every class, its configuration inside the space of classes has to be analysed. We used *Multidimensional Scaling* as a

method to explore the configuration among them [64]. This method allow us to use a similarity matrix to find the point coordinates of each sample in a $\mathcal{D}$–dimensional space. The distances between pairs of points in the configuration agree with the similarity measurements.

The dimension $\mathcal{D}$ of the space where samples are represented is fixed by the stress measure. This is an error measure between the similarity matrix used and the distances matrix of the points in the new space. In our experiment we generate a 6–dimensional space with an stress value of 0.07, which we consider enough for our purposes. A stress value of 0.05 is considered good to establish the real space dimension.

In this space, we group the samples of the same class. For every group, the first three samples closest to the centroid of the group are taken as class prototypes.

The classification process consists of two steps. Firstly, to calculate the similarity of the input sample to every prototype computed in the above step. Secondly, if the minimum distance does not exceed a certain threshold, $\mathcal{T}$, the input sample is classified as belonging to the same class of the representative sample from which the minimum is obtained. Otherwise a new class is defined and a new space to classify is calculated. The parameter $\mathcal{T}$ determines the stock fragmentation, the smaller $\mathcal{T}$ the higher fragmentation.

Results

The test has been carried out in laboratory conditions, where 90 tiles were used from 3 different models with 3 classes in each model. We used the full histogram to calculate the similarity D implementing the histogram with BTrees in order to save memory space. If the histogram dimension is reduced the method is unfeasible because of the small differences in colour between classes. The classification rate was over 90% using exclusively colour information, but when used in a larger set of samples the classification rates dropped to approximately 75%. This was because colour can not cope with all the cases and both colour and texture has to be considered. There are samples with the same colour distribution that they only differ on how colours are distributed on the sample. These were the first steps in the colour texture inspection problems.

5.6.4   Classification based on proposed perceptual features

From the conclusion of section 5.6.2 we translated the list of visual features to computational features on the image. There is not a one to one equivalence between expert and computer features but we define measures that involve several expert words.

We can associate measurements M_1^i, M_5^i and the small dimensional histogram on $a(\mathcal{R})$, M_9^i, to the first characteristic from the list. M_9^i can also catch information on the second characteristic. Light vs. dark grain colour and Light vs. dark background are represented by measurement $M_1(\boldsymbol{I}^i)$, depending on the segment applied it will be grain or background. Which one it is the segment assigned to background is not relevant because we do not need to explain the reasons because the computational system classify, and for all i, $M_1(\boldsymbol{I}^i)$ will be calculated. $M_1(\mathbf{I})$ and $M_1(\mathbf{I})$ are used

to define the fifth characteristic, Light vs. dark global colour. We also add two more measure to the list of selected ones, M_2^i and M_7^i because when the *grain is open*, in their jargon, blobs are more rounded and have less variation in its elongation. From this set of measures M_1^i and M_5^i are rejected because they are redundant to some extent with M_9^i.

In short and as a summary 2 vectors are obtained as a more global description $(M_1(\mathbf{I})$ and $M_2(\mathbf{I}))$. Two more set of vectors describe globally how are the main statistical properties of each of the colours in the image, $M_1(\boldsymbol{I}^i)$ and are applied to each of the p segmented images. Finally the M_9^i, M_2^i and M_7^i measures geometric properties of the blobs in each of the images $\boldsymbol{I}_i$.

To result these concepts in the general colour-texture representation (section 5.5), parameter f associated to global features is $f = \{1, 2\}$, and f associated to local features is $f = \{2, 7, 9\}$ The last step is to input the feature vector into a classifier. We have used the Linear Discriminant Analysis defined in A.

Results

We have tested the classification method on a variety of feature vectors obtained combining the two clustering methods and applying perceptual blurring and without it, that is: $n = 0$ and $n = 1$ in the colour texture representation model. The segmentation method is done on the original image and in the perceptually sharpened image, $m = 0$ and $m = 2$. We did not combine them, but tested two different configurations $\{n = 0, m = 0\}$ and $\{n = 1, m = 2\}$. As the tiles of the same model have a single main frequency, there is no need to extent the feature vector to a multiscale representation on perceptual blurring and sharpening. For both cases, the frequencies are selected manually for each model.

We have used six different models of tiles with a total set of 514 samples distributed in 47 different classes. Each sample has been divided in three regions which results in 1542 images. One third of the images are randomly selected to be used as learning set and the others are the test set. The classification percentages obtained are those from table 5.1 to 5.6. The first column of the tables is the preprocessing applied to the image which can be the SS operator ($\{n = 1, m = 2\}$) or the image itself ($\{n = 0, m = 0\}$). The second column is the clustering algorithm used where MG stands for Mixture of Gaussians. As the experts recognise that some times the same sample can be included in two different classes, we have obtained the percentage of images classified as the class that they actually belong and the percentage of images that their real class is the first or second best match. Due to the fact that the tiles are previously sorted by humans it is reasonable to take this second percentage as the capability of the system to cope with the problem. The percentages are taken only on the test sample, the learning set is not considered.

Not all the models have the same number of pigments. When it increases the human experts have more difficulties to classify them. The models used in this experiments where Duero, Tiber, Cinca, Orinoco, Ohio and Mijares with 2, 4, 3, 3, 2, and 3 different pigments respectively. Table 5.7 contains the global results for all the models. The conclusions from the tables can be summarised in

- When the number of pigments is greater than two, there is more confusion

Process	Clustering	1st rank	1st or 2nd rank
SS operator	MG	98.75%	100%
	KMeans	97.08%	100%
None	MG	84.16%	92.08%
	KMeans	85.83%	92.91%

Table 5.1: Classification results for Duero model. Number of images in the test set: 240.

Process	Clustering	1st rank	1st or 2nd rank
SS operator	MG	97.5%	100%
	KMeans	97.08%	99.5%
None	MG	92.5%	96.66%
	KMeans	91.25%	96.66%

Table 5.2: Classification results for Tiber model. Number of images in the test set: 240.

Process	Clustering	1st rank	1st or 2nd rank
SS operator	MG	98%	100%
	KMeans	98%	99%
None	MG	89%	92%
	KMeans	87%	93%

Table 5.3: Classification results for Cinca model. Number of images in the test set: 100.

Process	Clustering	1st rank	1st or 2nd rank
SS operator	MG	94.05%	99.01%
	KMeans	91.08%	97.02%
None	MG	88.17%	90.59%
	KMeans	86.13%	89.10%

Table 5.4: Classification results for Orinoco model. Number of images in the test set: 202.

Process	Clustering	1st rank	1st or 2nd rank
SS operator	MG	99.30%	100%
	KMeans	99.30%	100%
None	MG	95.83%	97.22%
	KMeans	94.44%	96.52%

Table 5.5: Classification results for model Ohio. Number of images in the test set: 144.

Process	Clustering	1st rank	1st or 2nd rank
SS operator	MG	100%	100%
	KMeans	100%	100%
None	MG	100%	100%
	KMeans	100%	100%

Table 5.6: Classification results for Mijares model. Number of images in the test set: 102.

Process	Clustering	1st rank	1st or 2nd rank
SS operator	MG	97.47%	99.61%
	KMeans	96.40%	99.03%
None	MG	90.76%	94.36%
	KMeans	89.30%	93.87%

Table 5.7: Average classification results for all models.

between blobs and working on the original image has worst results.

- Although the ratios are not so high than in the previous case, perceptual sharpening operator always gives better results even if the number of different pigments is low.

- The use of the clustering algorithms is not very crucial in the final classification. There are some advantages in the case of mixture of Gaussians, but at the expense of processing time and memory usage. Thus, when the classification ratio is not crucial *kmeans* is a feasible option.

- Due to subjectiveness in the human classification, taking the first and second best classes to fit the sample is an approximation in the uncertainty introduced in the classification process.

Although the number of classes in each model is important, the higher the number of classes is the higher probability of misclassification, but it can happen even with very few classes. In the case of Cinca model there are three different classes but the classification is not perfect, whereas in the Mijares model there is no error with the same number of classes. This is because of the the tiles supplied in the second case are very different between them and any approach will solve it. In regard to Duero, Tiber, Orinoco and Ohio models, the number of grades respectively is 10, 12, 11 and 7.

In spite of these good results, the problem is not completely solved. What we get is an evidence that the colour texture description approach is valid for this problem. However the dynamic creation of new classes inside a model is not obvious, and until it is not done a full on–line system can not be build. To work with this problem we should have all the tiles produced before and after a new class is created, and that needs a more complex logistic when the prototype is done in off–line production.

None of the works found in this field cope with this problem and it is still an open problem to solve in the future.

5.7   Case 2: Printing Quality evaluation

The second industrial problem that we focused on was a particular aspect of the printing quality evaluation. One of the problems of the printing industry is to assure printers can print homogenous colour patches, if they cannot do it very poor final results are obtained. Sometimes the print head produces a vertical degradation that translates to an striped patch which is called a banding effect that diminishes the printing quality. Sometimes it is a very smooth effect and sometimes is very obvious. As many other applications it is done by humans, which implies subjectivity and non–repetitiveness, as in the previous case. To control the printing quality they grade each patch in different levels. The problem is to find a set of parameters to automatically evaluate this grade or, even better, to give a number in a continuous domain of how good the printer is.

Unfortunately we had a very short set of examples. In addition the images were scanned at a medium resolution and we do not know exactly the parameters of the process. Despite these adverse circumstances we have tried the SS operator to this problem.

The perceptual sharpening operator SS can work on any colour space whenever it is tristimulus. Although the opponent space is the best to be used when simulating human behaviour, others can be used to make differences more obvious in specific problems. That is relevant when analysing printed images because the tristimulus used in this process is based on the subtractive primaries *cyan*, *magenta* and *yellow* plus a fourth ink introduced for practical purposes, *black*, and so it is called CMYK model. Then instead of using the opponent space we will use the SS operator on the CMY space as the quality test are usually done with pure inks. With this modification we can operate only on the principal channel of the patch, which is a priori known. For example, if we operate on the opponent space a nearly homogeneous cyan patch will present a colour distribution that will be shared by the two chromatic opponent channels. When performing in the CMY space the representation lays on a single channel and this is the only one that has to be sharpened.

The transform from RGB (the original colour space of the images) and the CMY space can be approximated by:

$$\left[\begin{array}{c} C \\ M \\ Y \end{array} \right] = \left[\begin{array}{c} 1 \\ 1 \\ 1 \end{array} \right] - \left[\begin{array}{c} R \\ G \\ B \end{array} \right] \tag{5.20}$$

When testing the black ink, as it is just luminance, the opponent space is used and the first component is sharpened as there should be no chromatic information.

The images from banding effects are somewhat textured because the white colour of paper sheet becomes more visible as worst is the defect of the printer. That is, the more evident is the defect the more texture appears in the image. A printed image is never perfectly homogeneous but when it is good the colours that appear are very
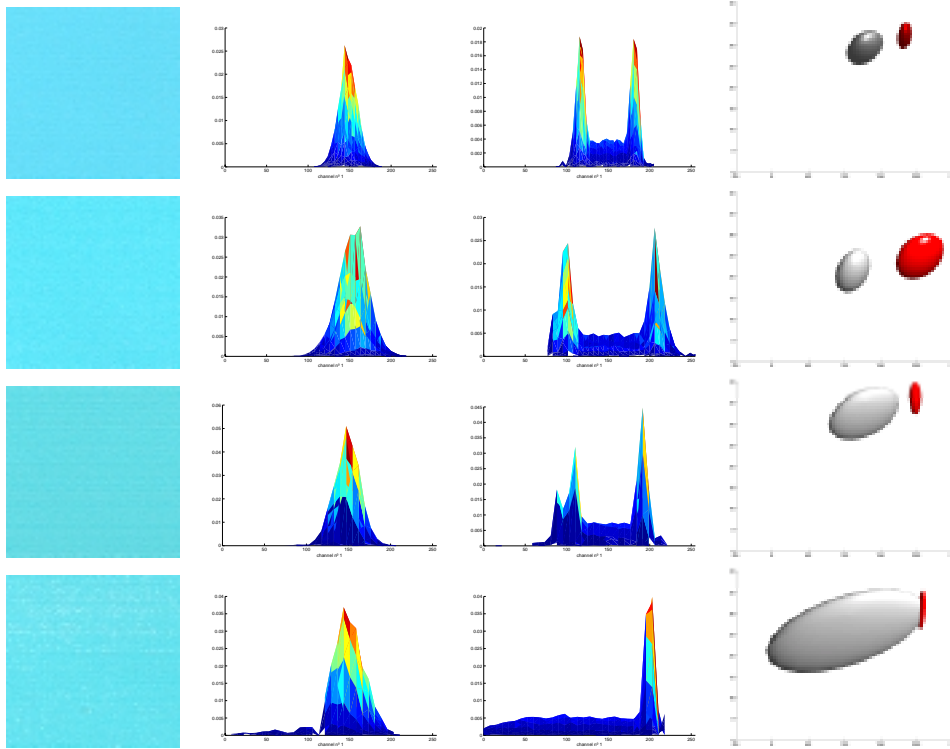
Figure 5.8: Cyan patches from different printers, ordered from top to down by its banding quality. From left to right: original image, colour distribution of the image on the cyan channel, distribution of the image after the SS operator in the CMY colour space, gaussian mixture of the previous distribution.

close one to the other. When there exist non–homogeneity and the colour distribution is analysed the main colours that appear are white (the paper colour) and the hue of the basic ink. Those colours are clearly far away one from the other. We will use this effect to visualise the printer banding error.

What we do is to cluster the colours of the image into two groups using the decision criterion based on modeling de distributions by a mixture of gaussians defined in section 5.2.2. If the region is homogeneous two clusters with their means very close have to be found, and the covariance matrices of both of them have to be compact. One way to measure this compactness is to calculate the ellipsoids that includes 99.5 % of the distribution. The length of the semi–axes is a parameter of how scattered is the colour. If a simpler parameterisation is needed the linear discriminant analysis of gaussian distributions can be calculated and then the mean and covariance will be 2–dimensional instead of 3–dimensional.

It is common that on some inks the banding effect is stronger than on others and this makes each ink to have different grades. For example, gray patches are

Figure 5.9: Magenta patches from different printers, ordered from top to down by its banding quality. From left to right: original image, colour distribution of the image on the cyan channel, distribution of the image after the SS operator in the CMY colour space, gaussian mixture of the previous distribution.
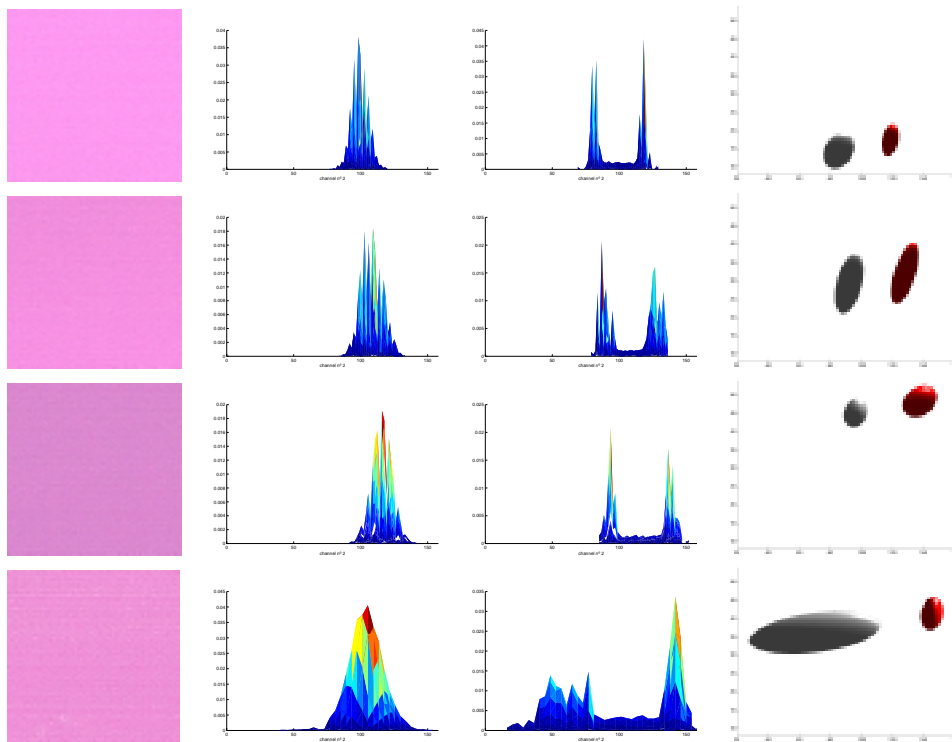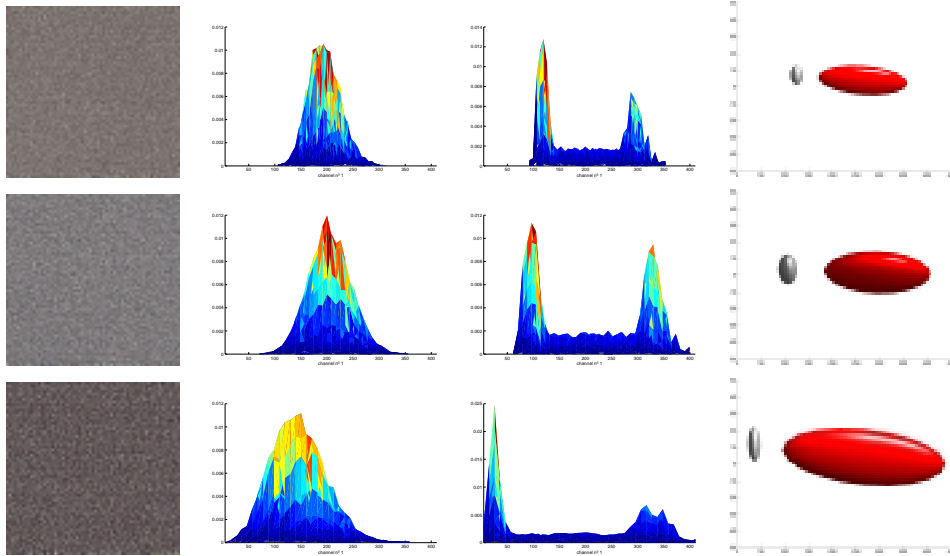
Figure 5.10: Grey patches from different printers, ordered from top to down by its banding quality. From left to right: original image, colour distribution of the image on the intensity channel, distribution of the image after the SS operator in the opponent colour space, gaussian mixture of the previous distribution.

usually better than cyan, and light colours are less sensitive to banding effects than the darker ones of the same hue. It is for that reason that in the examples that we show de number of grades are different and the magnitude of the parameters is specific for each ink.

Figures 5.8, 5.9 and 5.10 shows the results on some samples. The first one, a cyan patch, has four grades shown in the first column of the figure, sorted from best to worst according to the given classification. The first patch is clearly homogeneous whereas the last one is the worst case. In the middle the degradation can be continuous. When plotting the histogram of the cyan channel from the original image, second column, we observe a peak in all cases and in the last one there is a queue that is the white pixels. The other three cases are difficult to evaluate from this distribution. When applying the SS operator in the same view, third column, there appear two peaks when the image is more or less homogeneous, corresponding to the cyan in the image and the last one with one peak of the cyan and a more exaggerated queue of the white. The distance between peaks, if they exist, and their width can characterise the homogeneity of the sample. The results from modeling them with a mixture of gaussians are shown in the fourth column. The first ellipsoid, in red, is linked with the highest cyan in the image and the grey one to or the second cyan in the image or the white pixels. When it captures information of white pixels it is more elongated and its centre is far away from the centre of the red ellipsoid. If it is the coverage of a second cyan then the semi–axes are more similar and the centres are closer.

Figure 5.9 shows the same effect in a magenta sample where the same rules can be applied. And the last image is from a grey patch, where the red ellipsoid holds information of white pixels and the grey one includes the black pixels. In this case there are only three different samples because this colour is not so sensitive to the effect.

No numerical results can be extracted as we would need a number of samples to get a good validation of the parameters and their meaning. Despite this lack the preliminary results show encouraging performance towards its quantification.

5.8 Discussion

In this chapter we have addressed the colour texture representation based on perceptual mechanisms. To define a general framework to describe an image we have taken a multiscale approximation. In this way human attentive processes can be represented computationally. A clustering process on the perceptual sharpened images is done to better adjust the blob segmentation, which can be adjusted to reject ambiguous colours.

Another point of interest is the definition of two sets of features valid for a wide set of different sort of iamges. Global and local features are calculated for every scale needed. The use of feature vectors is not new in representing textures. What it changes from existing works is the use of the features on the images that segregate similar colours.

The general framework is closed defining a general schema that can be adapted to the knowledge of the problem, rejecting features, assimilation scales or contrast scales. Its flexibility allows adapting to a wide variety of problems.

Two of such problems are presented at the end of the chapter. The first one to demonstrate the viability of a classification system of ceramic tiles, which is badly posed as a problem of just colour differences. The second one uses the perceptual segmentation to quantify the banding error of printing industry. In this case, the results are preliminary but they present a good perspective.

Chapter 6

Conclusions and open research directions

Before to come in the details of the main contributions of this work we should review which was the aim of this thesis in its initials. The goal was twofold:

- From an engineering perspective, it was to solve the problems of an automatic inspection system of colour texture surfaces, when the final target is to get quantitative measures of general colour texture properties. We were mainly addressed to the grading of polished ceramic tiles.

- From a scientific perspective, it was to define robust computational colour texture representations that allow to derive similarity judgements which are coherent with human perception and represent enough sound information to allow complex classification tasks.

Taking in mind these two goals, the essential contributions of this work can be summed up in three:

1. A perceptual sharpening operator has been proposed, whose behaviour tries to simulate the chromatic contrast effect demonstrated by the human visual system.

2. A complete computational representation for colour texture has been formulated. It compiles colour representations that consider spatial operations to move chromaticities in a perceptual sense, with a feature vector computed on the modified colour representation that combines global and local features.

3. An on-line colour constancy algorithm for a colour scan line camera that removes dependencies on spatial and temporal variations of lighting conditions, to be used in industrial inspection processes in computer vision.

The main advantages of every contribution will be analysed in the following paragraphs.

The first contribution can be understood as a perceptual pre-processing step that modifies the colour distribution of an image depending on the spatial relations of the image content. The modified distribution presents good properties for being segmented accordingly to the colour appearance of the image blobs.

Even when the number of colours in a scene is not very large, and due to the smoothing effects introduced by the acquisition system, it is common to have unimodal colour distributions for images where a concrete number of colour are perceptually segregated without effort. Unimodality introduced problems at the time to take a colour classification decision within the image. To build sharpened images with a colour distribution where the unimodality has been broken and a clear mixture of different gaussians have appeared accordingly with the number of the perceived image colours can introduced interesting properties for general colour segmentation and for coloured-blob segmentation.

As has been demonstrated in different works on texture perception, as the Julesz texton theory, the attributes of the image blobs have an essential contribution for any texture representation. An efficient blob segmentation is an unavoidable step before computing any local property, as blob size, orientation, chromaticity or contrast.

The proposed computational representation presents the ability to compile the image content properties as a result of a vision process that combines different observations of the same scene or what we could call an integrated colour multi-scale representation. Properties obtained from the images perceptually blurred will represent the image appearance from a distant observer positioning, on the other hand, properties obtained from the images perceptually sharpened will represent the image appearance from an attentive process of the observer across the image and from a close observer positioning to the target. The proposed local features are the same as those used to represent gray-level texture images.

Finally, the on-line colour constancy algorithm developed has more practical than theoretical implications. A linear diagonal model is applied and adapted to the special constraints of a scan line camera. Although its inherent simplicity, it is completely indispensable for any real application where a classification catalog pretends to be stored for a long-term period.

On the basis of the scientific work in this thesis, two practical derivations have been addressed on inspection problems:

1. Ceramic tile classification: We have established the basis for a future system for fully automatic system in this scope. A simplification of the general schema defined for colour texture has been used to grade a large set of samples of ceramic tiles, validating this approximation.

2. Printing quality inspection: We have defined the first steps to quantify the banding effect in commercial printers. This is done using the properties of the defined operators, which improve the colour distribution of the images considering spatial relationships. This takes us to a better interpretation of the degree of banding error.

Once, we have enumerated and commented the most important contributions of this work, let us enumerate other minor contributions and conclusions we want to highlight from all the work:

- Several acquisition architectures has been tested and analysed before to get a final solution. In this work we analyse a list of possibilities and common problems arising when a digital colour acquisition problem is posed. The central conclusions we want to enumerate are the importance of designing criteria that have to take into account when defining such an architecture:

  - The complexity or almost the impossibility of acquiring an homogenously illuminated surface with a matrix camera.

  - The problems on the sensor calibration of some commercial cameras.

  - The high red-sensibility of commercial CCD's that sometimes implies to introduce special filter corrections at the sensor input, resulting in an loss of light intensity.

  - The spatial non-homogeneities introduced by the optic systems: chromatic aberrations, vignetting effects, etc.

  - The non-homogeneous profile provided by light line fiber optics.

  - The influences of spectral distributions of the lamps when they are combined with other problems enumerated before, e.g, the red spectral distribution of a tungsten halogen lamp, added to the special sensibility of CCD's to red inputs increases the needs of filtered inputs.

- Some important corrections are needed when working with CCD cameras with high colour precision, as it is the removing of the dark current. While this is a common matter in astronomy where high precision is also required for photometric measurements, it is not as common in computer vision.

- A diagonal linear model, or Von Kries adaptation model, is the simplest solution to deal with the colour constancy problem. However, most of the commercial CCD cameras does not include completely sharpened sensors. To guarantee the suitability of this diagonal model we have built the sensor transformation that assures sharpened responsivities for a specific line scan camera.

- To compute the sharpening transformation the sensitivities of the camera sensors have been recovered. The method used has been based on a least-square approach. The main problems concluded from this methodology arises from the need of design information of the CCD that should be provided by the camera manufacturer that is not always available, and the need to introduce important constraints due to the generality of the basis functions.

- A wide review of psychophysical bibliography has been done to deal with the perceptual implications of the colour induction phenomena. Although there is an increasing number of works that deal with this issue, there are still some measurements that should be done, in order to be able to have a tabulation on how chromaticities of a large set of basic colours change in front of a large set of inductors. This should be done for a wide set of different spatial frequencies conditions.

Doing research and engineering activities provokes to open new issues to address. In our case some of the future research directions are

- To find a model to combine contrast and assimilation induction in a scene. As we are involved in computer vision it converts to a perceptual sharpening/blurring problem. Our first thought is that a multiresolution approach is needed to select those areas of the image assigned to each effect.

- Sharpening operators have been defined to simulate perceptual human vision but their parameters are left open. Although we have demonstrate their usefulness, psychophysical work is needed to evaluate the validity of this model to match the Human Visual System. In this case the magnitudes and ratios between parameters have to be investigated.

- To go deeply into the field of dynamic clustering to give response to many industrial applications where the clustering has to be done without a priori information. Although with the current results, the on–line surface inspection can be done, the training process of the classifier has to be simplified to be operative.

There also exists open engineering problems that have to be studied

- In the case of ceramic tile inspection, to collect and analyse the prior and posterior samples of the creation of a new grade is created by the production line. Without these data it is not possible to address the problem of a true on–line inspection.

- In the case of printing quality quantification, larger tests have to be done to validate the first results presented, and to evaluate the weight of the proposed parameters in the final quantisation.

- To look for new methods to recover camera sensor sensitivities in a high demanding application. These methods should keep the process as simple as possible for being feasible their use in industrial problems.

- To investigate the effects of the optics distortions in the recovery of the sensor sensitivities, and thus on the final spectral sharpening transform.

Appendix A

Classification method

As our interest is the capacity of the representation of texture and colour information simultaneously, we will not focus on the classification methods. Although there exists many classification approaches [26], considering the nature of the problem we found that a discriminant analysis method was a correct chose. We will need a set of samples to characterise the classes (the learning set) that will be selected randomly from the whole set of each class.

Discriminant analysis can cope with problems where the main characteristics of classes are not a priori known, prototyping classes from the learning sets. The selection criterion of these prototypes has to provide the maximum discrimination ratio overall the learning set. Moreover, distribution of features is unknown, so we have to use a non-parametric discriminant analysis method [63, 74]. One of the methods that fits these constraints is the one based on Fisher discriminant functions. With Fisher's approach there is no need of a priori knowledge of data and it is able to select the best representation maximising the ratio between the inter-class covariance and the intra-class covariance. A linear transform $\mathbf{W}$ is applied over the feature vector $\mathbf{x}$ of a particular image obtaining a new representation, $\mathbf{y} = \mathbf{W}^t \mathbf{x}$, in a new space where discrimination capability has been maximised.

The linear transformation, $\mathbf{W}$, which optimises the discrimination, is obtained by calculating the most significant eigen vectors of the matrix $\mathbf{S}_w^{-1} \mathbf{S}_b$, assuring the maximisation of the following ratio:

$$\frac{\mathbf{W}^t \mathbf{S}_b \mathbf{W}}{\mathbf{W}^t \mathbf{S}_w \mathbf{W}} \ , \tag{A.1}$$

where $\mathbf{W}^t$ stands for the transpose of $\mathbf{W}$, $\mathbf{S}_w$ is the within data sparse matrix defined as:

$$\mathbf{S}_w = \sum_{i=1}^{c} \sum_{\mathbf{x}_k \in C_i} (\mathbf{x}_k - \mu_i)(\mathbf{x}_k - \mu_i)^t \ , \tag{A.2}$$

where c is the number of possible classes and C_i is the set of vectors that are used as learning samples in the i class. The $\mathbf{S}_b$ matrix is the between class sparse matrix,

which is defined as:

$$\mathbf{S}_b = \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^t \ , \tag{A.3}$$

where μ_i is the mean vector of the samples of the i class, N_i is the number of learning samples in the i class and μ is the global mean vector.

We extract the feature vector from an input image, $\mathbf{x}$, and we assign it to the j class if

$$|\mathbf{W}^t \mathbf{x} - \mathbf{W}^t \mu_j| < |\mathbf{W}^t \mathbf{x} - \mathbf{W}^t \mu_i| \qquad \forall \ i \neq j \ . \tag{A.4}$$

A previous step on the classification process is to choose those variables that are significative of the data. Many variables can be used but few of them will be uncorrelated or will be homogenous inside each group. The best way to that is to choose the combinations of variables that best classifies the learning set of samples. This is very time consuming and can not be carried out for many variables. Statistics broach this problem with a stepwise approximation. It begins by selecting the individual variable which provides the greatest univariate discrimination. Then all the remaining variables are paired with the selected one to select the pair that best discriminates the data. This procedure is done until there is no more variables or the contribution to the classification is meaningless. Some test can be done to look for the best discriminating variable, but is has been shown that the results are nearly the same and conclusions vary slightly. We have used SPSS statistical software to do the classification process, which implements Wilks's lambda to test variables:

$$\lambda = \frac{\mid \mathbf{S}_w \mid}{\mid \mathbf{S}_w + \mathbf{S}_b \mid}$$

where $\mid A \mid$ denotes determinant of matrix A. The best discriminatory value is 0 and 1 when variables are useless. Further details on the classifier can be found in [26, 74, 63].

Appendix B

Tile samples

In this appendix we depict some samples of the ceramic tiles models used in section 5.6.

commercial name	number of pigments	number of grades	number of tiles
Duero	2	12	120
Tiber	4	10	120
Cinca	3	3	50
Orinoco	3	11	101
Ohio	2	7	72
Mijares	3	3	51

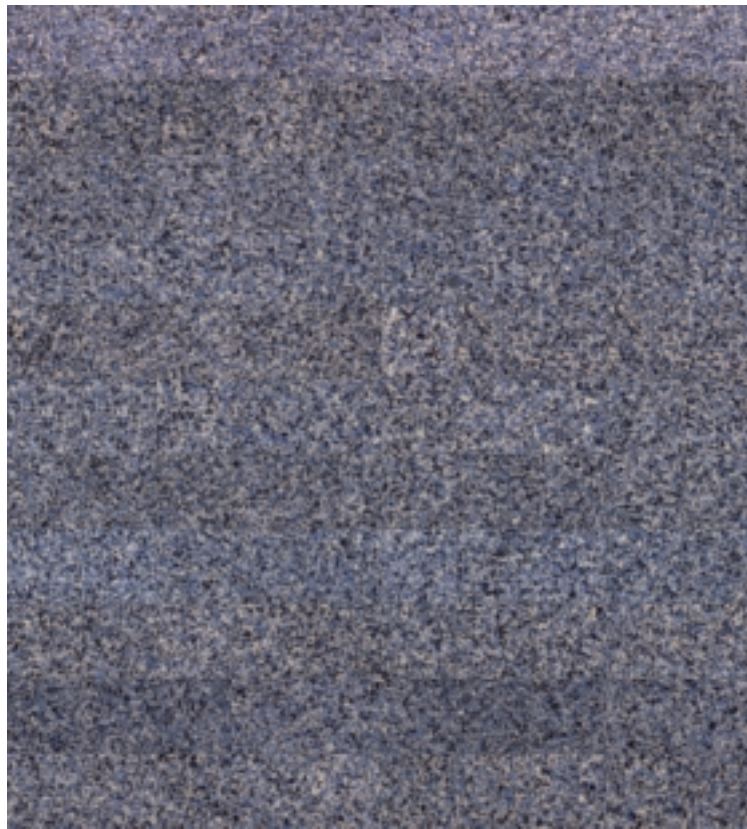**Table B.1:** List of model tiles used in the ceramic tile classification problem.

Figure B.1: Tiber model. Classes are presented on rows.

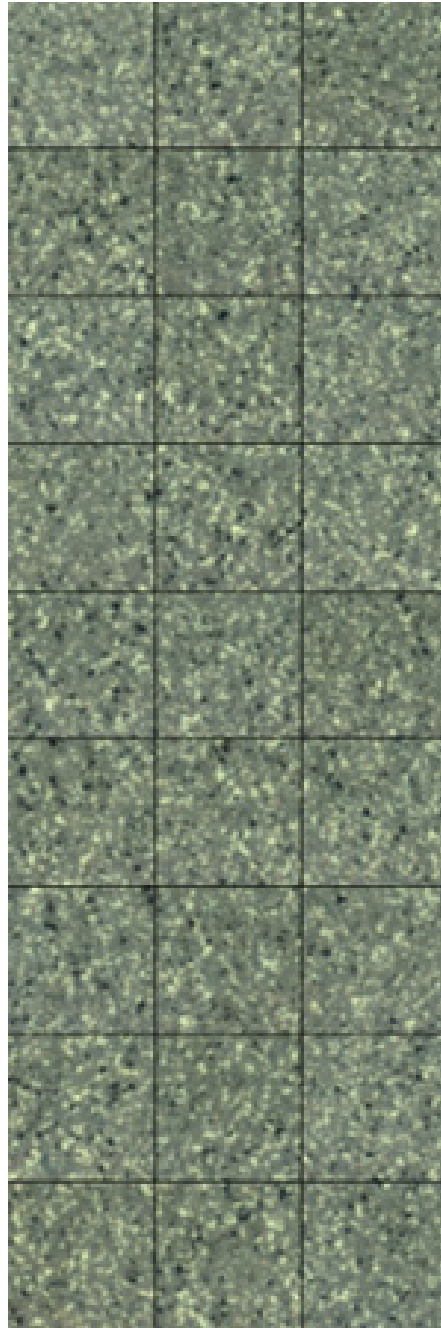**Figure B.2:** Duero model. Classes are presented on rows.

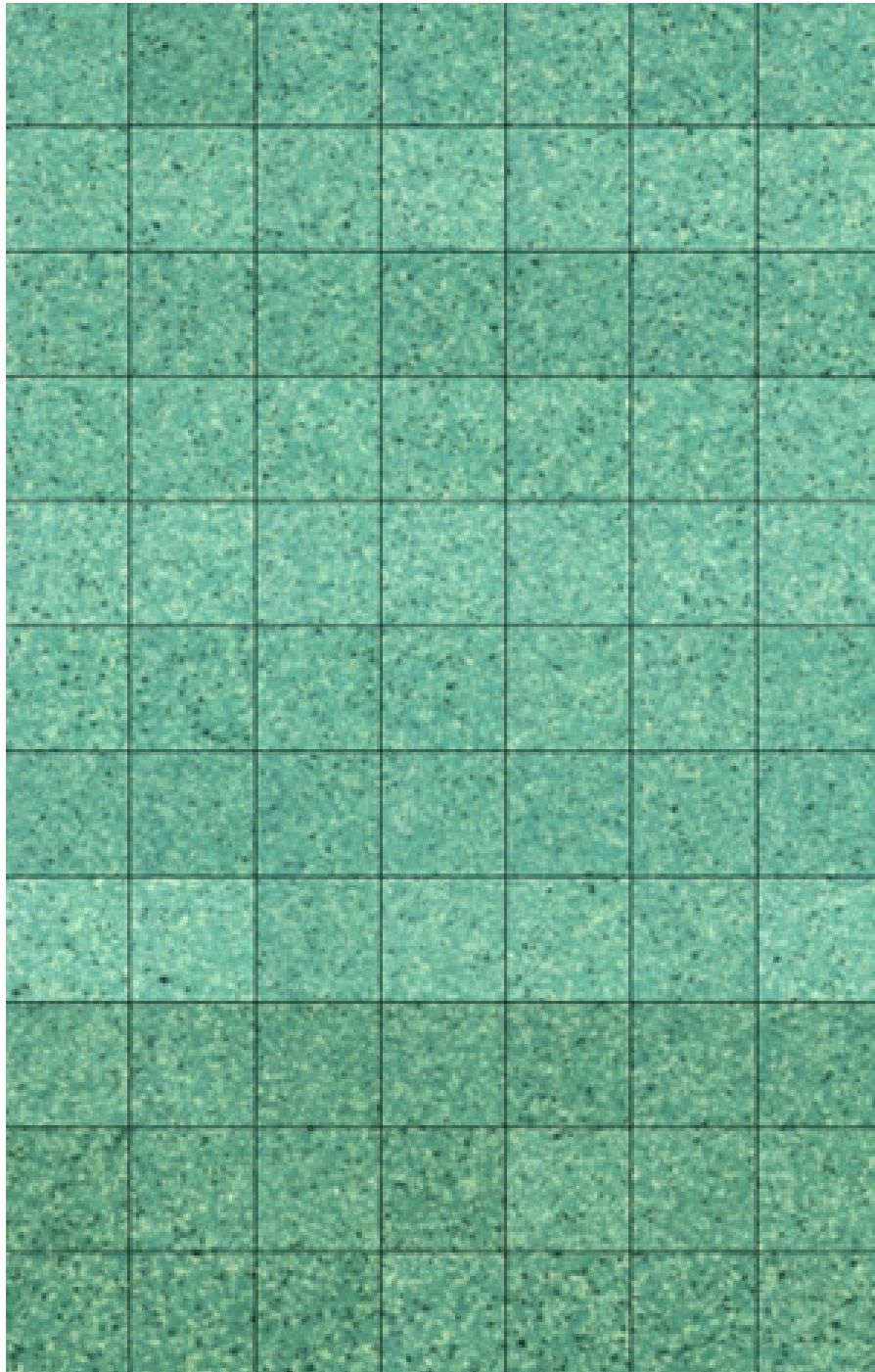**Figure B.3:** Cinca model. Classes are presented on columns.

Figure B.4: Orinoco model. Classes are presented on rows.

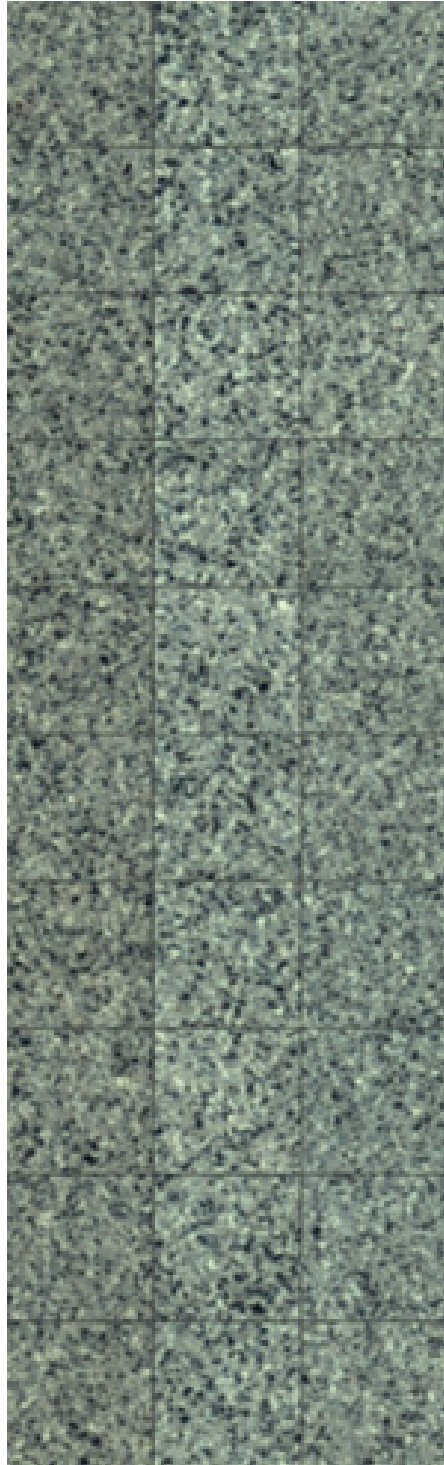**Figure B.5:** Ohio model. Classes are presented on rows.

Figure B.6: Mijares model. Classes are presented on columns.

Bibliography

[1] N. Ahuja and Azriel Rosenfeld. Mosaic models for textures. *IEEE*, PAMI-3:225–235, 1981.

[2] K-H Baeuml and B. Wandell. Color appearance of mixture gratings. *Vision Research*, 36(18):2849–64, 1996.

[3] Ramon Baldrich, Maria vanrell, and Juan José Villanueva. Texture–colour features for tile classification. In *Proceedings of SPIE - The International Society for Optical Engineering*, volume 3826, pages 124–135, 1999.

[4] Ramon Baldrich, Maria vanrell, Juan José Villanueva, and Jaime López-Krahe. Ceramic tile classification based on colour similarity measurements. In *Workshop on European Scientific and Industrial Collaboration on promoting*, 1998.

[5] Kobus Barnard. Computational color constancy: Taking theory into practice, 1995. Simon Fraser University, School of Computing. Available from `http://citeseer.nj.nec.com/279611.html`.

[6] Kobus Barnard. *Practical Colour Constancy*. PhD thesis, Simon Fraser University, School of Computing, 1999.

[7] J. Beck, A. Sutter, and R. Ivry. Spatial frequency channels and perceptual grouping in texture segregation. *Computer Vision, Graphics, and Image Processing*, 37:299–325, 1987.

[8] Robert Benavente, Ramon Baldrich, Maria Cinta Olivé, and Maria Vanrell. Colour naming considering the colour variability problem. *Computación y Sistemas*, pages 30–43, 2000.

[9] Robert Benavente, Gemma Sànchez, Ramon Baldrich, Maria Vanrell, and Josep Lladós. Normalized colour segmentation for human appearance description. In *15th ICPR '2000*, 2000.

[10] Christopher M. Bishop. *Neural Netwrorks for Pattern Recognition*. Oxford University Press, 1997.

[11] C Boukouvalas and M Petrou. Perceptual correction for colour grading of random textures. *Machine Vision and Applications*, 12(3):129–136, Oct 2000.

[12] Constantinos R. Boukouvalas. *Colour shade grading and its applications to visual inspection*. PhD thesis, Surrey University, 1996.

[13] Costas Boukouvalas, Francesco De Natale, Giovanni De Toni, Josef Kittler, Radek Marik, Majid Mirmehdi, Maria Petrou, Philip Le Roy, Roberto Salgari, and Gianni Vernazza. An integrated system for quality inspection of tiles. In *International Conference on Quality Control by Artificial Intelligence*, pages 49–54, 1997.

[14] D. Brainard and W. Freeman. Bayesian color constancy, 1997.

[15] Michael W. Burke. *Image Acquisition*. Chapman & Hall, 1996.

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

[17] T. Caelli and D. Reye. On the classification of image regions by colour, texture and shape. *Pattern Recognition*, 26(4):461–470, 1993.

[18] V. Cardei, B. Funt, and K. Barnard. Modeling color constancy with neural networks. In *Proc. International Conference on Vision Recognition, Action: Neural Models of Mind and Machine*, 1997.

[19] Chad Carson, Serge Belongie, Hayit Greenspan, and Jitendra Malik. Region–based image querying. In *CPR Workshop on Content–Based Access of Image and Video Libraries*, 1997.

[20] E.J. Chichilnisky and B.A. Wandell. Photoreceptor sensitivity changes explain color appearance shifts induced by large uniform backgrounds in dichoptic matching. *Vision Research*, 35:239–254, 1995.

[21] Guy B. Coleman and Harry C. Andrews. Image segmentation by clustering. *Proceedings of the IEEE*, 17(5):773–785, May 1979.

[22] Munsell Color. *Munsell Book of Color, Matte Finish Collection*. Macbeth Division of Kollmorgen Instruments Corporation, 1976.

[23] G.R. Cross and A.K. Jain. Markov random field texture models. *IEEE PAMI*, 5:61–75, 1983.

[24] C.S.Barnes, J. Wei, and S.K. Shevell. Chromatic induction with remote chromatic contrast varied in magnitude, spatial frequency, and chromaticity. *Vision Research*, 39:3561–3574, 1999.

[25] K. Dana, S. Nayar, B. Ginneken, and Koenderink J. Reflectance and texture of real-world surfaces. In *Proceedings of the CVPR97*, pages 151–157, 1997.

[26] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. John Wiley and Sons, Inc, 2001.

[27] S. Duvdevani-Bar and S. Edelman. On similarity to prototypes in 3d object representation. Technical report, Weizmann Institute, 1995.

[28] M. D'Zmura and P. Lennie. Mechanisms of color constancy, 1986.

[29] C. Fach and L.T. Sharpe. Assimilative hue shifts in color gratings depend on bar width. *Perception and Psychophysics*, 40:412–418, 1986.

[30] Mark D. Fairchild. *Color Appearance Models*. Addison Wesley, 1998.

[31] Graham Finalyson, Bernt Schiele, and James Crowley. Comprenhensive colour image normalization. In *5th European Conference on Computer Vision*, pages 475–490, Freiburg, Germany, 1998.

[32] G. Finlayson and S. Hordley. Selection for gamut mapping colour constancy. *Image and Vision Computing*, 17:597–604, 1999.

[33] G. Finlayson, P. Hubel, and S. Hordley. Color by correlation. In *IS&T and SID's 5th Color Imaging Conference: Color Science, Systems and Applications, (Scottsdale, Arizona)*, pages 6–11, 1997.

[34] Graham Finlayson. Color in perspective. *IEEE Transactions on Pattern Analyisis and Machine Intelligence*, 18(10):1034–1038, October 1996.

[35] Graham Finlayson, Mark Drew, and Brian Funt. Spectral sharpening: Sensor transformations for improved color constancy. *Journal of the Optical Society of America*, 11(5):1553–1563, May 1994.

[36] Graham Finlayson, Steven Hordley, and Paul Hubel. Recovering device sensitivities with quadratic programming. In *IS&T and SID's Sixth Color Imaging Conference: Color Science, Systems and Applications*, pages 90–95, 1998.

[37] Graham D. Finlayson, Subho S. Chatterjee, and Brian V. Funt. Color angular indexing. In *European Conference on Computer Vision*, pages 16–27, april 1996.

[38] Graham D. Finlayson, Mark S. Drew, and Brian V. Funt. Color constancy: Generalized diagonal transforms suffice. *Journal of the Optical Society of America*, 11(11):3011–3019, November 1994.

[39] F.Liu and W.Picard. Periodicity, directionality, and randomness: Wold features for image modelling and retrieval. *IEEE Trans. on PAMI*, 18(7):722–733, 1996.

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

[41] K.S. Fu. *Syntactic Pattern Recognition and Applications*. Prentice-Hall, 1982.

[42] A. Gagalowicz, S. De Ma, and C. Tournier-Lasserve. Efficient models for color textures. In *8th ICPR*, pages 412–414, 1986.

[43] David Galadi-Enriquez and Ignasi Ribas. *Manual Practico de Astronomia con CCD*. Ed. Omega, 1998.

[44] J. Garding. Shape from texture and contour by weak isotropy. *Artificial Intelligence*, 64:243–297, 1993.

[45] Michael Gennert, Norman Wittels, and Gary Leatherman. Uniform frontal illumination of planar surfaces: Where to place the lamps. *Optical Engineering*, 32(6):1261–1271, June 1993.

[46] Theo Gevers. A new approach to segmentation of color images. Master's thesis, Universiteit van Amsterdam, 1990.

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

[48] B.B. Ginneken and J.J. Koenderink. Texture histograms as a function of irradiation and viewing direction. *International Journal of Computer Vision*, 31(2/3):169–184, 1999.

[49] L.van Gool, P. Dewaele, and A. Oosterlinck. Survey: Texture analysis anno 1983. *Computer Vision, Graphics and Image Processing*, 29:336–357, 1985.

[50] Stephen Grossberg and Dejan Todorovic. Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. *Perception & Psychophysics*, 43:241–277, 1988.

[51] L. Harvey and M. Gervais. Visual texture perception and Fourier analysis. *Perception and Psychophysics*, 24(6):534–542, 1978.

[52] L. Harvey and M.Gervais. Internal representation of visual texture as the basis for the judgment of similarity. *Journal of Experimental Psychology: Human Perception and Performance*, 7(4):741–753, 1981.

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

[54] F. Heitz, H. Maitre, and Ch.de Couessin. Application of autoregressive models to fine arts painting analysis. *Signal Processing*, 13:1–14, 1987.

[55] Gerald C. Holst. *CCD Arrays Cameras and Displays*. JCD Publishing & SPIE Press, 1996.

[56] Berthold Klaus Paul Horn. *Robot Vision*. McGraw–Hill Book Company, Cambridge, Massachusetts, 1986.

[57] Paul Hubel, Doron Sherman, and Joyce Farrell. A comparison of methods of sensor spectral sensitivity estimation. In *IS&T and SID's 2nd Color Imaging Conference*, pages 45–48, 1994.

[58] L.M. Hurvich and D. Jameson. An opponent-process theory of color vision. *Psychological Review*, 64(6):384–404, 1957.

[59] A.K. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, 24:1167–1186, 1991.

[60] Anil K. Jain. *Fundamentals pf digital image processing*. Prentice Hall, Englewood Cliffs, NJ, 1989.

[61] D. Jeulin. Morphological modeling of images by sequential random functions. *Signal Processing*, 16:403–431, 1989.

[62] B. Julesz and J.R. Bergen. Textons, the fundamental elements in preattentive vision and perception of textures. *Bell Systems Technological Journal*, 62:1619–1645, 1983.

[63] Willima R. Klecka. *Discriminant Analysis*. Sage Publications, Inc., 1987.

[64] J.B. Kruskal and M. Wish. *Multidimensional Scaling*. Sage Publications, Inc., 1978.

[65] Edwin H. Land. The retinex theory of color vision. *Scientific American*, 237(6):108–129, 1977.

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

[67] Raymond L. Lee. Colorimetric calibration of a video digitizing system: Algorithm and applications. *COLOR research and application*, 13(3):180–186, June 1988.

[68] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.

[69] Felipe Lumbreras, Ramon Baldrich, Maria Vanrell, Joan Serrat, and Juan José Villanueva. Multiresolution colour texture representations for tile classification. In *VIII National Symposium on Pattern Recognition and Image Analysis*, pages 227–234, May 1999.

[70] Felipe Lumbreras, Joan Serrat, Ramon Baldrich, Maria Vanrell, and Juan José Villanueva. Colour texture recognition through multiresolution features. In *Quality Control by Artificial Vision*, pages 114–121, May 2001.

[71] J. Malik and P. Perona. Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America*, 7:923–932, 1990.

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

[73] B.S. Manjunath and W.Y. Ma. Texture features for browsing and retrieval of image data. *IEEE-PAMI*, 18(8), 1996.

[74] K.V. Mardia, J.T. Kent, and J.M Bibby. *Multivariate Analysis*. Academic Press, London, 1997.

[75] D. Marr. *Vision*. W. H. Freeman and Company, 1982.

[76] Elzbieta Marszalec and Matti Pietkäinen. On–line color camera calibration. In *Computer Vision and Image Processing*, pages 232–237. Proc. 12th ICPR, 1994.

[77] C.S. McCamy, H. Marcus, and J.G. Davidson. A color–rendition chart. *Journal of Applied Photographic Engineering*, 2(3):95–99, summer 1976.

[78] M.D.Levine. *Vision in Man and Machine*. McGraw Hill, 1985.

[79] M. Mirmehdi and M. Petrou. Perceptual versus gaussian smoothing for pattern-colour separability. In *International Conference on Signal Processing and Communications*, pages 136–140. IASTED/Acta Press, February 1998.

[80] M. Mirmehdi and M. Petrou. Segmentation of color textures. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(2):142–159, 2000.

[81] Eriko Miyahara, Vivianne C. Smith, and Joel Pokorny. The consequences of opponent rectification: the effect of surround size and luminance on color appearance. *Vision Research*, 41:859–871, 2001.

[82] D.K. Panjwani and G. Healey. Markov random field models for unsupervised segmentation of textured color images. *IEEE Trans. Patt. Anal. Mach. Intell.*, 17(10):939–954, October 1995.

[83] Jose A. Peñaranda, Leoncio Briones, and Julian Florez. Colour machine vision system for process control in ceramics industry. In *New Image Processing Techniques and Applications: Algorithms, Methods, and Components II*, pages 182–192. SPIE - The International Society for Optical Engineering, June 1997.

[84] M. Petrou, M. Mirmehdi, and M. Coors. Perceptual smoothing and segmentation of colour textures. In *5th European Conference on Computer Vision*, pages 623–639, Freiburg, Germany, 1998.

[85] Allen Poirson and Brian Wandell. The apprearance of colored patterns: Pattern–color separability. *Journal of the Optical Society of America*, 10(12):2458–2470, 1993.

[86] Allen Poirson and Brian Wandell. Pattern-color separable pathways predict sensitivity to simple colored patterns. *Vision Research*, 36(4):515–526, 1996.

[87] F. Preteux and M. Schmitt. *Image Analysis and Mathematical Morphology*, chapter Boolean Texture Analysis and Synthesis, pages 377–401. Academic Press, 1988.

[88] A.R. Rao and Gerald L. Lohse. Identifying high level features of texture perception. *CVGIP: Graphical Models and Image Processing*, 55:218–233, 1993.

[89] A.R. Rao and Gerald L. Lohse. Towards a texture naming system: Identifying relevant dimensions of texture. *Vision Research*, 36:1649–1669, 1996.

[90] Todd R. Reed and J.M. Hans Du Buf. A review of recent texture segmentation and feature extraction techniques. *CVGIP: Image Understanding*, 57(3):359–372, May 1993.

[91] Leila Shafarenko, Maria Petrou, and Josef Kittler. Automatic watershed segmentation of randomly textured color images. *IEEE Transactions on Image Processing*, 6(11):1530–1543, November 1997.

[92] S. A. Shafer. Using color to separate reflection components. *Color Research and Application*, 10(4):210–218, Winter 1985.

[93] G. Sharma and H. Trussel. Characterization of scanner sensitivity. In *IS&T and SID's Color Imaging Conference: Transforms & Transportability of Color*, pages 103–107, 1993.

[94] S.K. Shevell and J. Wei. Chromatic induction: border contrast or adaptation to surrounding light? *Vision Research*, 38:1561–1566, 1998.

[95] Steven K. Shevell and Jianping Wei. A central mechanism of chromatic contrast. *Vision Research*, 40:3173–3180, 2000.

[96] V.C. Smith, W. Jin, and Joel Pokorny. Color appearance: neutral surrounds and spatial contrast. *Vision Research*, 38:3265–3269, 1998.

[97] V.C. Smith and Joel Pokorny. Color contrast under controlled chromatic adaptation reveals opponent rectification. *Vision Research*, 36(19):3087–3105, 1996.

[98] Vivianne C. Smith, Phil Q. Jin, and Joel Pokorny. The role of spatial frequency in color induction. *Vision Research*, 41:1007–1021, 2001.

[99] K.Y. Song, J. Kittler, and M. Petrou. Defect detection in random colour textures. *Image and Vision Computing*, 14:667–683, 1996.

[100] Michael J. Swain and Dana H. Ballard. Color indexing. *International Journal on Computer Vision*, 7(1):11–32, 1991.

[101] H. Tamura, S. Mori, and T. Yamawaki. Textural features corresponding to visual perception. *IEEE Trans. on System, Man and Cybernetics*, SMC-8:460–473, 1978.

[102] S.C. Tan and J. Kittler. Colour texture classification using features from color histogram. In *Proceedings of the 8th Scandinavian Conference on Image Processing*, 1993.

[103] M. Tuceryan and A.K. Jain. *Handbook of Pattern Recognition and Computer Vision*, chapter Texture Analysis, pages 235–276. World Scientific, 1993.

[104] M. Vanrell, F. Lumbreras, A. Pujol, R. Baldrich, J. Llados, and J.J. Villanueva. Colour normalisation based on background information. In *ICIP 2001, IEEE International Conference on Image Processing*, volume 1, pages 874–877, 2001.

[105] H. Voorhees and T. Poggio. Computing texture boundaries from images. *Nature*, 333:364–367, 1988.

[106] Poorvi L. Vora and H. Joel Trussell. Mathematical methods for the analysis of color scanning filters. *IEEE Transactions on Image Processing*, 6(2):321–327, 1997.

[107] Poorvi L. Vora and H. Joel Trussell. Mathematical methods for the design of color scanning filters. *IEEE Transactions on Image Processing*, 6(2):312–320, 1997.

[108] Brian Wandell. *Foundations of Vision*. Sinauer Associates, Inc, ISBN 0–87893–853–2, 1995.

[109] Brian A. Wandell. Color appearance: The effects of illumination and spatial pattern. Technical report, Psychology Department, Standford University, november, 8 1995.

[110] Thomas Watchler, Thomas D. Albright, and Terrence J. Sejnowski. Nonlocal interactions in color perception: nonlinear processing of chromatic signals from remote inducers. *Vision Research*, 41:1535–1546, 2001.

[111] H. Wechsler. Texture analysis - a survey. *Signal Processing*, 2:271–282, 1980.

[112] C.M.M.de Weert and L. Spillmann. Assimilation: Asymmetry between brightness and darkness. *Vision Research*, 35:1413–1419, 1995.

[113] J. Weickert. Coherence-enhancing diffusion of colour images. *Image and Vision Computing*, 17:201–212, 1999.

[114] G. West and M. Brill. Necessary and sufficient conditions for von kries chromatic adaption to give colour constancy. *J. Math. Biol.*, 15:249–258, 1982.

[115] A.P. Witkin. Recovering surface shape and orientation from texture. *Artificial Intelligence*, 17:17–45, 1981.

[116] G. Wyszecki. *In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.) Handbook of perception and human performance, Volume I: Sensory processes and perception*, chapter Color appearance. John Wiley & Sons, 1986.

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

[118] Xuemei Zhang, J.E. Farrell, and Brian A. Wandell. Applications of a spatial extension to CIELAB. In *SPIE*, 1997.

[119] Xuemei Zhang and Brian A. Wandell. A spatial extension of CEIALB for digital color image reproduction. In *SID*, 1996.