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

(a) (b) (c)

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

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

$L_x(\lambda, \Theta)$ 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,

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

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

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

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

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

where $\boldsymbol{\mathcal{T}}$ 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

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

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

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

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 $\boldsymbol{\mathcal{T}}^i = \mathbf{U}^{-1}$.

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

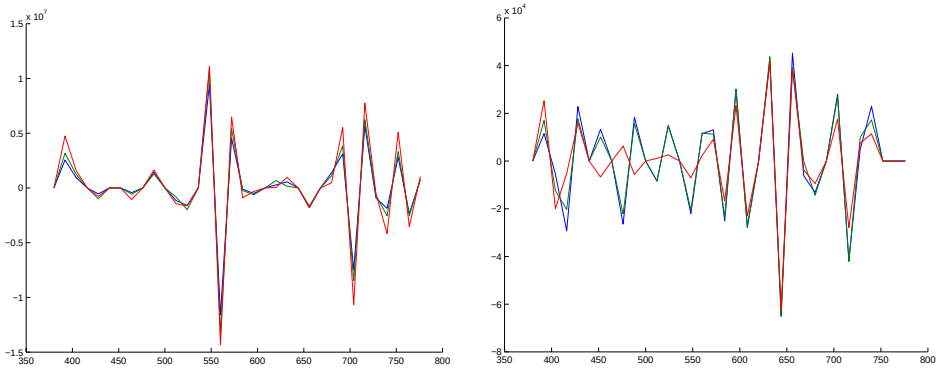
$$\rho_x^k = (\vec{L}_x)' \vec{R}^k \qquad k = 1, 2, 3 \tag{3.10}$$

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



(a) (b

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 $\mathbf{L}$ 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:

$$\begin{pmatrix} \rho_1^k \\ \rho_2^k \\ \vdots \\ \rho_{24}^k \end{pmatrix} = \begin{pmatrix} I(\lambda_1)S_1(\lambda_1) & I(\lambda_2)S_1(\lambda_2)\cdots I(\lambda_W)S_1(\lambda_W) \\ I(\lambda_1)S_2(\lambda_1) & I(\lambda_2)S_2(\lambda_2)\cdots I(\lambda_W)S_2(\lambda_W) \\ \vdots & \vdots \quad \cdots \quad \vdots \\ I(\lambda_1)S_{24}(\lambda_1) & I(\lambda_2)S_{24}(\lambda_2)\cdots I(\lambda_W)S_{24}(\lambda_W) \end{pmatrix} \begin{pmatrix} \hat{R}^k(\lambda_1) \\ \hat{R}^k(\lambda_2) \\ \vdots \\ \hat{R}^k(\lambda_K) \end{pmatrix} \tag{3.12}$$

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:

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

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:

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

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

$$R^k(\lambda_{v+1}) \geq R^k(\lambda_v) \qquad v = 1, \ldots, w-1 \tag{3.14}$$

$$R^k(\lambda_{v+1}) \leq R^k(\lambda_v) \qquad v = w, \ldots, W-1 \tag{3.15}$$

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:

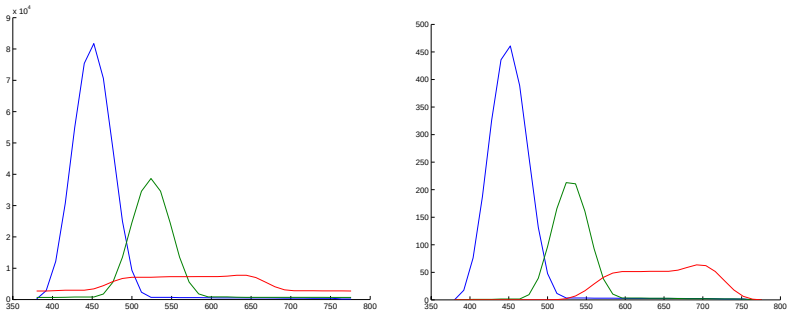
$$\mathbf{B}_1 = c, \quad \mathbf{B}_2 = sin(x), \quad \mathbf{B}_3 = cos(x), \quad \mathbf{B}4 = sin(2x), \quad \mathbf{B}_5 = cos(2x), \cdots$$

and combining them following the equation:

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

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

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



(a) (b)

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:

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

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:

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

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	$\begin{pmatrix} 1.014 & 0.117 & 0.015 \\ 0.088 & 1.028 & 0.042 \\ -0.027 & 0.112 & 1.005 \end{pmatrix}$		

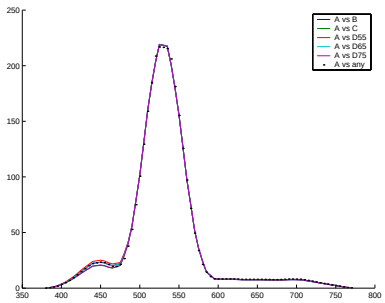


Figure 3.5: Modified camera sensor sensitivity to improve DTM. All transform are similar but not identical. The dotted line is the best fit to all transforms..

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 $\boldsymbol{\mathcal{R}}^{\boldsymbol{X}}$ 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

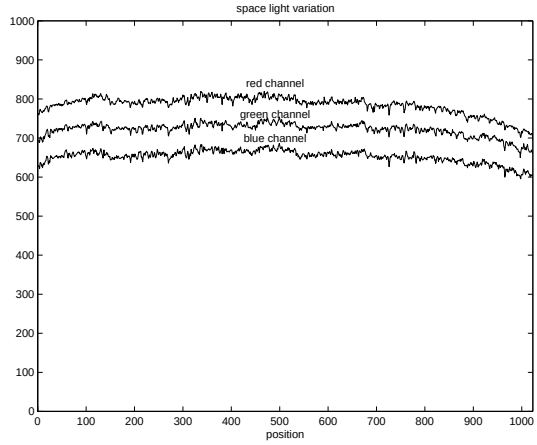


Figure 3.6: Light spatial variation: it shows the fall–off at the edges of the image due to the cosine–4th law.

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:

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

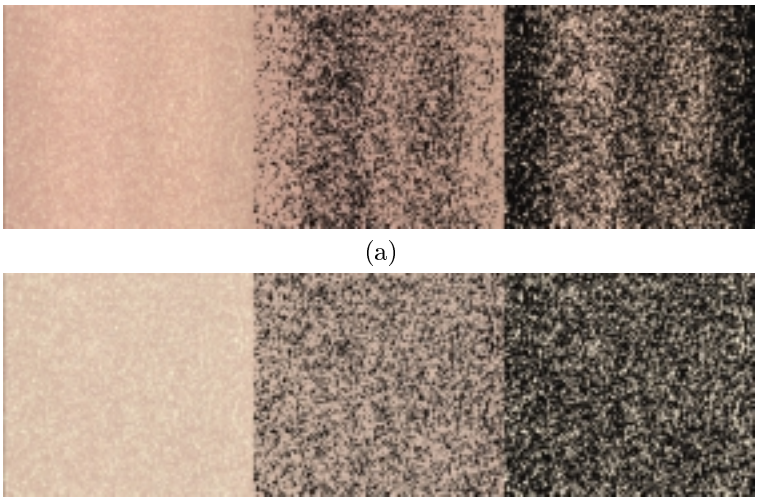
and for each channel, k, of the image we have

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

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.

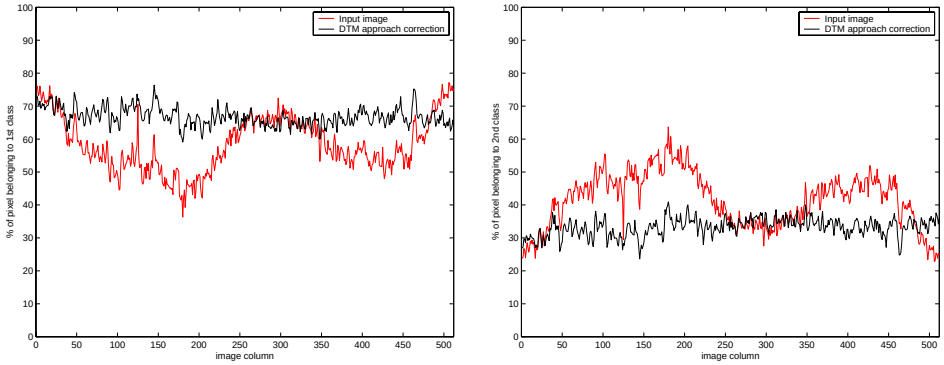
$t_i < t < t_j$, being t_i and t_j the times when two consecutive white reference images have been taken, we can express the canonical descriptor of $\mathbf{p}_x$ as

$$\mathbf{p}_x^c = \mathbf{S}_x \mathbf{T}_x^{t_i} \mathbf{p}_x \,, \tag{3.20}$$

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

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

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.

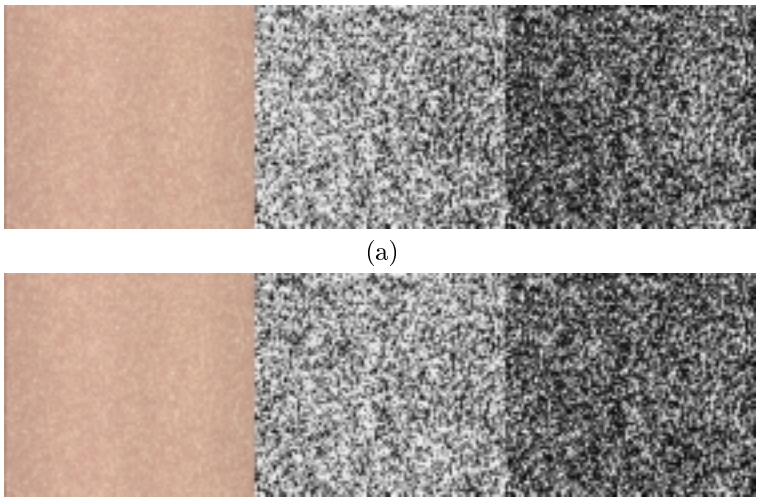
We have measured the difference between the colour correction with and without this sensor sharpening for a line scan camera. The error between both transforms is

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

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.

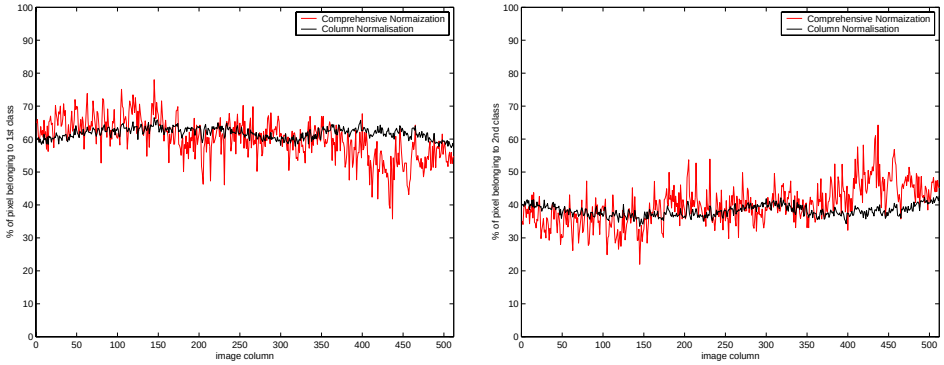
work based on the *Comprehensive Normalisation* of Finlayson presented in [31]. The comprehensive colour normalisation is an iterative algorithm, which tries to remove shading and light colour, successively. One of the assumption of this normalisation is that light colour is constant all over the scene, 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.