Spatiochromatic information content of natural scenes.

by Carlos Alejandro Párraga

A thesis submitted to the University of Bristol in accordance with the requirements of the degree of Master of Science in the Faculty of Science, Department of Psychology.

July, 1995.

Abstract

The spatiochromatic characteristics of a dataset of 29 natural scenes (representative of the natural terrain) were explored. The image capture was done using a video camera capable of recording images through a set of 31 narrowband interference filters spanning the visible spectrum. The system allows the measurement of spectral radiance and reflectance for every point within a given scene. The Fourier amplitude spectrum characteristics of the dataset were explored across the visible spectrum. Our results show that (a) the mean distribution of spectral reflectance and radiance in our dataset is consistent with the shape of the h.v.s. spectral sensitivity curves and (b) the characteristics of the Fourier amplitude spectrum show no significant variation with wavelength. We explored the consequences of having visual receptors different from ours. Our results show that (c) there is no significant variation of the Fourier amplitude spectrum with the bandwidth of the receptor and (d) luminance images produced from these hypothetical receptors show a Fourier amplitude spectrum whose characteristics depend on those of the receptors. Luminance and chrominance images, based on assumptions about human cone response and signal coding were derived from the dataset. Our results show that (e) the different spatial transfer functions of colour and luminance in human vision are not reflected in the spatial-frequency characteristics of our scenes which appear to be rich in high-spatial-frequency chrominance information (f) the Fourier amplitude spectrum of our luminance and chrominance images, when plotted in log-log co-ordinates is a straight line with slope similar to that found for achromatic images in other studies.

Para Claudia, la dueña de mis horas
(por haberme prestado algunas para completar esta Tesis).

For Mum, Dad and all those who supported me at home.

Acknowledgements:

I wish to acknowledge and express my gratitude to all those who made it possible for this thesis to be written.

To my supervisor Dr Tom Troscianko for his patience and advice, for trusting me from the beginning and for providing an excellent environment to work in.

I have benefited with discussions with members of the PSRC especially Dr Gavin Brelstaff who gave me his advice and criticisms. He also gave me enormous help to understand and learn how to use the software tools he skilfully wrote for processing images, and to understand the basics of the DRA-Camera. Other members of the PSRC also helped me with the job of recording the scenes I used in this work. They are Ms Kay Nelson and Mr Derek Carr.

Thanks to Mr Tim Wood who gave me advice on English grammar.

I also wish to acknowledge the staff of the University of Bristol Botanical Gardens for kindly providing help in the recording of part of the scenes I use in this work.

Author's declaration:

Part of this work was performed with the collaboration of members of the PSRC of the University of Bristol. I wish to state that I have not participated in either the construction or the calibration of the DRA-Camera as described in Chapter 2 (sections 2.1 to 2.3). The development of the software tools used for image processing in this work (Chapter 2, section 2.8) was done by Dr G. Brelstaff. My contribution to the software tools employed within this work is reduced to the original idea for the *"reilluminate"* algorithm (as described in Chapter 2, section 2.6 and Appendix A, section A6) kindly implemented by Dr G. Brelstaff and all validity tests described in Chapter 2 (section 2.9). The scene grabbing on the field was done with the assistance of Ms K. Nelson, Mr D. Carr and Dr G. Brelstaff. Part of the contents of Appendix A (sections A1 to A5) was summarised from "The DRA-Camera Theory" a report on the physics of the DRA-Camera submitted to the DRA by Gavin Brelstaff in January 1994.

All views expressed in this dissertation are those of the author and not of the University.

Table of contents

List of Tables and illustrative material

Tables

Figures

Abbreviations

3D	Three-dimensional
a/d	Analogue/Digital
CCD	Charged-coupled device
CCU	Camera control unit
Chrom	Chrominance
DRA	Defence Research Agency
e-o	Electro-Optical
h.v.s	Human visual system.
IR	Infra red
L, M, S	Low, Middle, Short
l.g.n	Lateral geniculate nucleous
Lum	Luminance
LUT	Look up table
ND	Neutral density
PC	Personal computer
PSRC	Perceptual Systems Research Centre
lab	Laboratory
sf	Spatial frequency
UV	Ultra violet

Definitions

Chapter 1

Background

1.1 *The visual pathway*

The pathway of light perceived by the h.v.s. can be sketched as follows: light enters the eye and passing through its optics reaches the retina (see Figure 1.1). Here a group of light receptors (cones and rods) converts the light into neural activity. This neural activity (visual information) is then transmitted to the lateral geniculate nucleus (l.g.n.) through the optic nerve (see Figure 1.2). The left l.g.n. receives fibers from retinal areas stimulated by light from the right visual field in both eyes and the opposite happens with the right l.g.n. From here, l.g.n. fibers project to the striate cortex (in the back of the brain) on the same side.

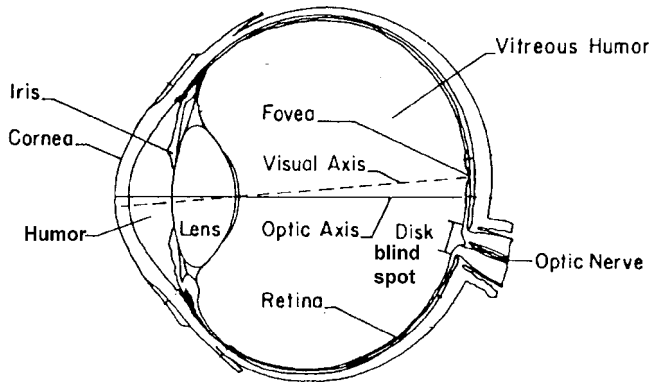

Figure 1.1: Schematic cross section of the eye
(from De Valois and De Valois 1990)

The neurophysiological approach to vision has been relatively successful in understanding the precortical levels of the visual system: the retina and the lateral geniculate nucleus. This may be because the relationship between neural responses and the physical properties of the optical image are more direct at these precortical levels. At higher levels of the visual system some less understood phenomena such as consciousness, alertness and attention play significant roles in perception. There is a

large body of information available about the visual system of primates at the precortical stages. The neural organisation of this system reveals general principles that can be applied to man.

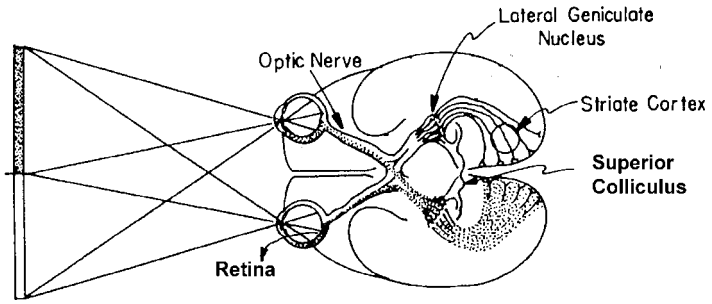

Figure 1.2: Schematic diagram of the main neural pathways involved in vision (from De Valois and De Valois 1990).

1.2 *The receptors*

The h.v.s. has three types of cone photoreceptor (labelled L, M, and S) characterised by different spectral sensitivities -low, middle and short wavelength sensitivity respectively. The spectral sensitivity curves of these cones were derived by Smith and Pokorny (1972, 1975) using a combination of psychophysical measures and some reasonable assumptions. They obtained their data using dichromats (red or green colour-blind observers) and flicker photometry to ensure that only one chromatic mechanism contributed to the spectral sensitivity curve each time. Figure 1.3 shows the L, M, and S spectral sensitivity curves. The other receptor type (rod) plays no role in colour vision. They are about 10 times as sensitive as cones and thus their main contribution is at very low (scotopic) levels of light. Although there are 20 times more rods than cones in the human retina, cones make the largest contribution to the information going to subsequent brain centres (De Valois and De Valois 1990). The largest concentration of cone photoreceptors in the retina occurs in a centre called the fovea. Although the fovea subtends a visual angle of only about 1 to 2 degrees, it plays a disproportionately large role in vision. Figure 1.4 shows a diagram of the density of receptors as a function of radial eccentricity in the human retina. Most of the studies are in fact dedicated to investigating foveal vision.

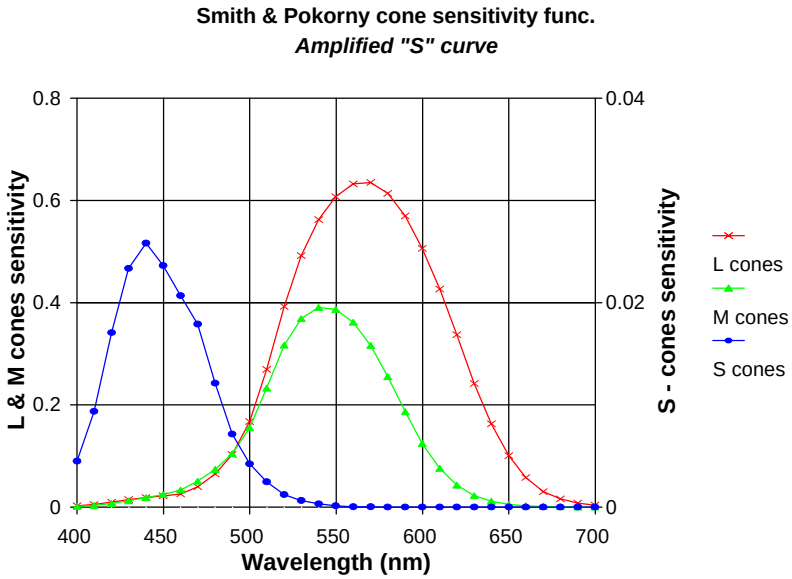

Figure 1.3: Spectral sensitivity of L, M and S mechanisms. The scale for the S-cones spectral sensitivity is amplified (right). From Smith and Pokorny's (1975) data.

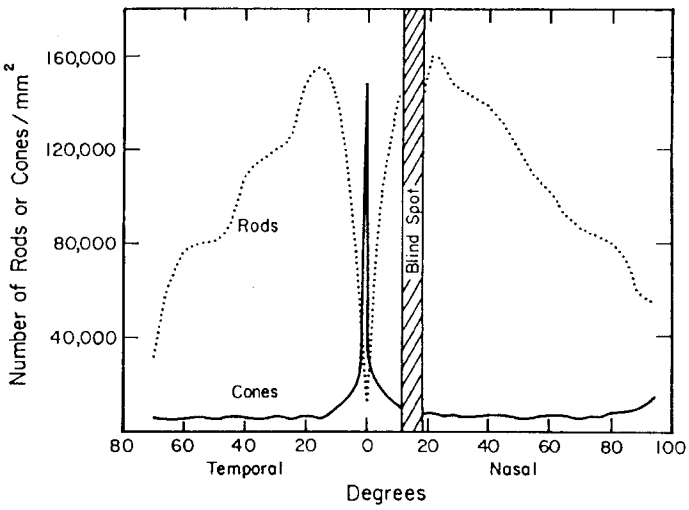

Figure 1.4: Density of receptors vs. eccentricity in the human retina (from De Valois and De Valois 1990).

1.3 Spatial and chromatic antagonisms

In the primate fovea the output of a single cone is transmitted to a single ganglion cell by a bipolar cell system. This signal is ultimately delivered to the visual cortex by relay neurones in the l.g.n. Horizontal and amacrine cells receive inputs from groups of cones creating cone opponent interactions. These kinds of interactions (see Figure 1.5) have been detected in cones, horizontal cells, bipolars, ganglion cells and in neurones of the retino-geniculate pathway to the primate striate cortex. However, the precise effect is

not yet understood. Cone opponency allows the h.v.s. to extract information from chromatic contrast due to local spatial differences in the spectral composition of the light reaching it (see Figure 1.6). The comparison between the response of a single cone to those of its neighbours provides information about achromatic contrast due to the local spatial distribution of the overall energy of the light reaching the cone (see Figure 1.7). These two forms of visual contrast, chromatic and achromatic, may vary independently of one another across an edge in the viewed scene. This leads Gouras (1991a) to suggest that they require different neural networks for their detection.

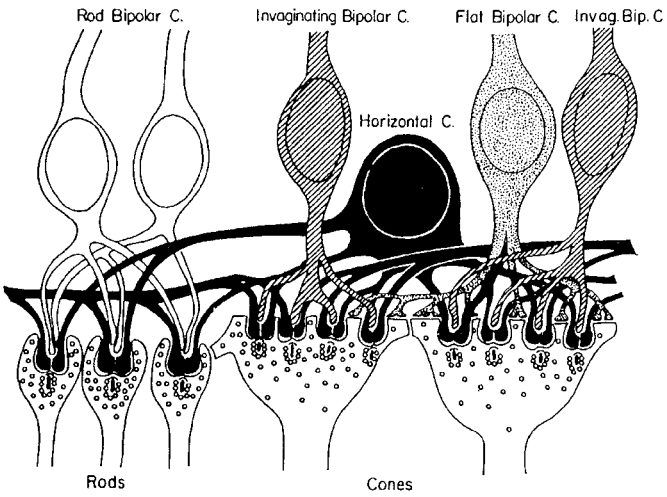

Figure 1.5: Opponent interactions in the retina (from De Valois and De Valois 1990)

Two different pathways become evident at the lateral geniculate nucleus (l.g.n.). The l.g.n. is the main route by which the retina communicates with the cerebral cortex in primates. All retinal axons that reach the l.g.n. connect to geniculate neurones which in turn transmit signals onward to the striate cortex. Thus retinal axons are distributed into specific layers of the l.g.n. to facilitate the transmission of their information to appropriate layers in the striate cortex. These l.g.n. layers are called *magnocellular* and *parvocellular* and it is well established that they are two functionally different visual pathways that operate in parallel on each local area of the retinal image. The parvocellular system is composed of smaller cells with slower response to the stimuli and is thought to provide the visual cortex with the bulk of the information for all chromatic vision and also at least achromatic vision at high spatial resolution. The magnocellular system is composed of larger cells, is thought to be sensitive to

achromatic but not to chromatic contrast, and responds faster than the parvocellular system to the stimuli (P. Gouras, 1991a).

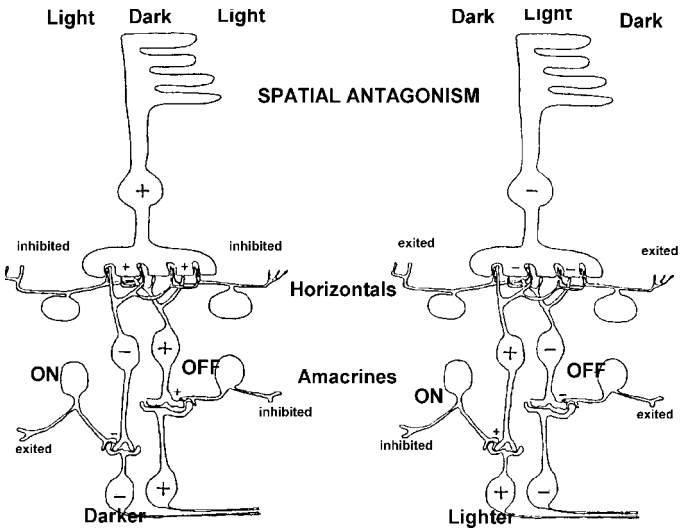

Figure 1.6: Spatial antagonism occurring between different areas of the retina (from Gouras 1991a).

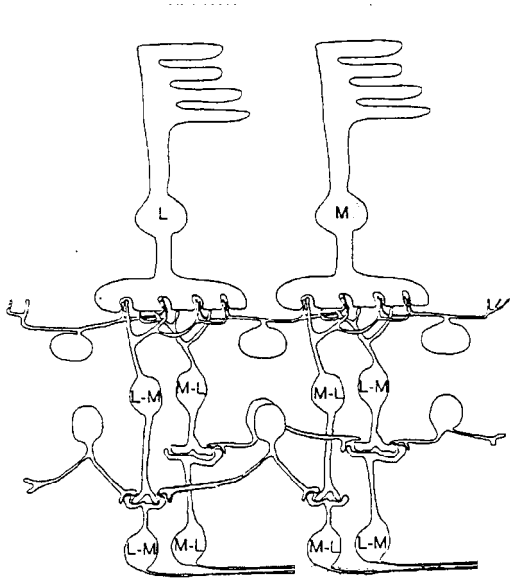

Figure 1.7: Spectral antagonism in the retina (from Gouras 1991a).

Derrington *et al.* (1984a) distinguished two groups of cells in the parvocellular layer of the macaque's l.g.n. One of these groups of cells receives opposed, but not equally balanced, inputs from only L and M cones (Derrington *et al.* assigned the label L-M to these l.g.n. cells). The other group of cells receive inputs from S cones almost equally opposed to a combined input from L and M cones (these they labelled S-(L + M)). A

second study (Derrington and Lennie 1984) reports relatively high sensitivity for achromatic contrast and higher temporal frequency tuning in magnocellular units than in parvocellular units of the l.g.n.

At this stage (l.g.n.) all visual information has been arranged into pools of photoreceptors sampled on the retina. An estimation of the size of the pool of photoreceptors sampled can be provided by psychophysical measurements of the spatial resolution for pure chromatic contrast. This can be done using (pure) colour gratings in which only the wavelength varies sinusoidally between two extremes and the whole grating is equated for luminance. These measurements (Mullen 1985) show that both S-(L+M) (also called "blue- yellow") and L-M (also called "red-green") contrast sensitivity functions have similar low-pass characteristics. Figures 1.8 and 1.9 show the L-M and S-(L+M) contrast sensitivity functions respectively to chromatic and achromatic gratings.

Comparison between contrast sensitivity functions for chromatic and achromatic gratings show that, at low spatial frequencies, contrast sensitivity is greater for the chromatic grating than for the achromatic, and at high spatial frequencies the opposite occurs.

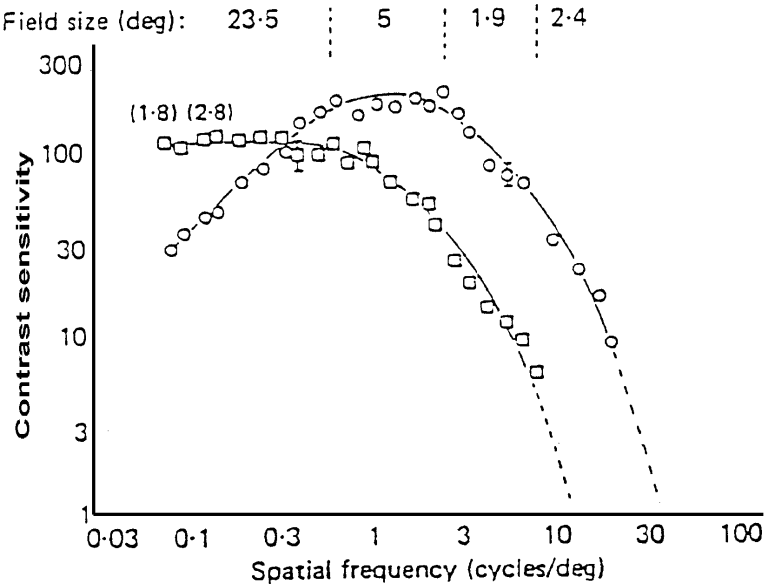

Figure 1.8: L-M contrast sensitivity function for chromatic (squares) and monochromatic (circles) gratings. Taken from Mullen (1985).

At this point we should introduce some of the fundamental questions to be discussed

later in this work: Why are the chromatic and achromatic contrast sensitivity thresholds tuned to favour different portions of the spatial frequency spectrum? Is this physiological imbalance reflected in the physical spatial frequency content of natural images? If so, we could conclude that the visual system has evolved to code the information existing in the visual environment.

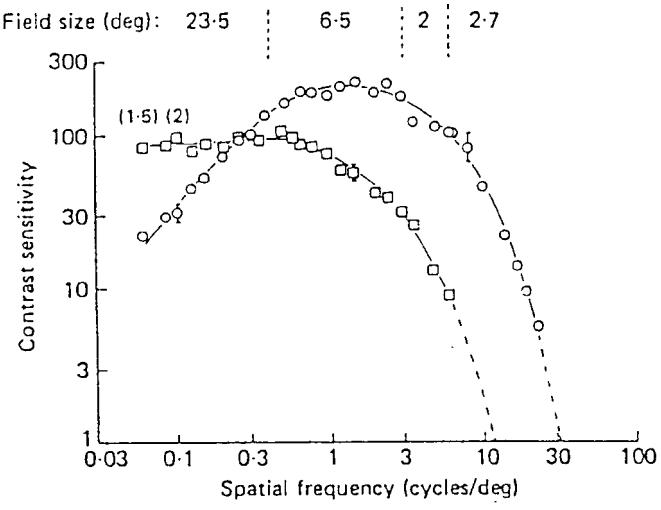

Figure 1.9: S-(L+M) contrast sensitivity function for chromatic (squares) and monochromatic (circles) gratings. Taken from Mullen (1985).

1.4 *Spatial frequency channels in the achromatic domain*

There is a considerable body of evidence supporting the suggestion that the contrast sensitivity function represents the envelope of many more narrowly tuned filter mechanisms (channels) that pass part of the information that may impinge upon them. Figure 1.10 (De Valois and De Valois 1990) illustrates the above idea. Each of these narrowly tuned spatial frequency filters is responsive to only a fraction of the total range covered by the contrast sensitivity function. They simultaneously analyse a given pattern, each responding as the pattern contains energy within its particular spatial frequency band (parallel processing). The most convincing phychophysical demonstration in support of the existence of multiple channels is spatial frequency specific adaptation (Blakemore and Champbell 1969). Here the contrast sensitivity function is measured before and after the subject is adapted to a monochromatic grating of single spatial frequency and the results show a band-pass loss in contrast sensitivity around the adaptation frequency. It is assumed that these psychophysically measured

spatial frequency channels are cortical in origin. One reason is because of the orientation selectivity of the channels, a property shown by striate cortex and later levels. A second reason is the interocular transfer of spatial frequency adaptation effects.

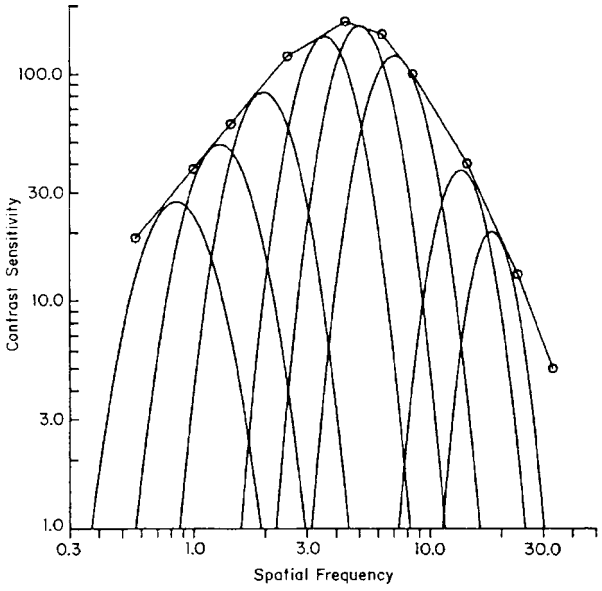

Figure 1.10: Contrast sensitivity function viewed as the envelope of more narrowly tuned spatial freq. selective channels (De Valois and De Valois 1990)

An important attempt to relate the statistical structure of the environment to the coding derived from these narrowly tuned filter mechanisms of the visual system was made by Field (1987). He considers the statistics of six achromatic scenes from the natural environment and compares various coding schemes on how they represent information in such natural scenes. Plotting the Fourier amplitude spectra of these images, Field shows that the greatest value is at low spatial frequencies and it decreases as spatial frequency increases. The amplitude spectra fall off quickly by a factor of roughly f^{-1} (where f is spatial frequency). This means that the total energy (see Appendix B) between for example, 2 and 4 cycles/deg will equal the energy between 4 and 8 cycles/deg. Such a spectrum is implied by a natural (fractal-like) environment that has s.f. energy invariant with scale (viewing distance). For a scene to be scale invariant, its statistics must remain constant as one magnifies any local region of the scene. Field also examines a model of the behaviour of cortical cells derived from principles from information theory (Shannon and Weaver 1949) and Gabor's (1946) theory of communication. The function first proposed by Gabor to analyse time-varying signals

was:

$$f(x) = \sin(2 \cdot \pi \cdot k \cdot x + \theta) \cdot e^{\frac{-x^2}{2\sigma^2}}$$

This function can describe the profile of the receptive fields of cortical neurones (similar to a Gaussian modulated by a sinusoid). The two-dimensional variation of this function is described as the product of a two dimensional Gaussian and a sinusoid. Field labels as *"sensors"* individual Gabor functions located at a point within the scene and representing a single hypothetical cortex cell. He organised these sensors into *"channels"* (spatial arrays of sensors tuned to a common orientation and s.f.). He concludes that such a collection of channels with sf-bandwidths constant in octaves and orientation bandwidths constant in degrees produces an even distribution of the information (from the set of six scenes) across the array. Figure 1.11 shows the relations between the size of a channel in the frequency domain and the size and spacing in the space domain.

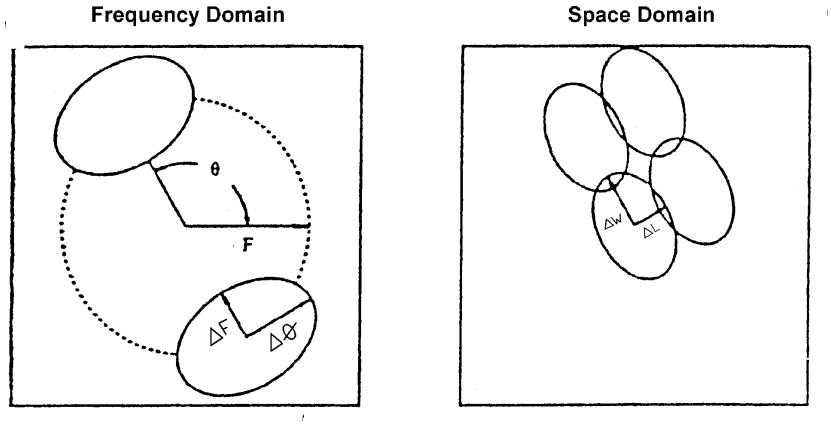

Figure 1.11: Fields channels represented in the frequency domain and in the space domain. A channel with a bandwidth ΔF in the frequency domain consists of an array of sensors with a width ΔW in the space domain. Taken from Field (1987).

Figure 1.12 similarly shows the relations between the spacing of the channels in the frequency domain and the spacing of the sensors in the space domain. Field found the optimal sf-bandwidth to be in the range 0.5 to 1.5 octaves. Also the optimal constant ratio between sf and orientation bandwidth (measured in the two-dimensional Fourier space) was found to lie between 0.5 and 1.0. Field also applies log-Gabor codes (a function that is a Gaussian on a logarithmic frequency axis) to the images and compares them to standard Gabor codes. One relevant property of these log-Gabor channels is that

they are always at zero at the zeroth sf and so a combination of them need not "over-represent the low frequencies". Field claims that his findings are in good agreement with the neurophysiological literature for simple cortical cells although his evenly distributed channels over a rigid sensor grid do not model the variability in the spatial frequency tuning of different cortical cells (De Valois *et al* 1982). Again, attention should be drawn to the extremely small sample of scenes (six) used to reach Field's conclusions.

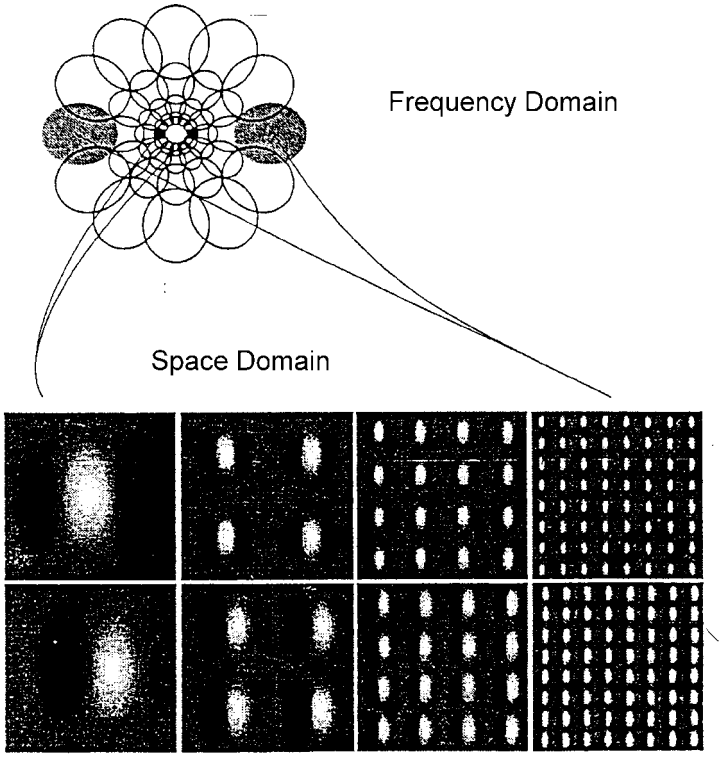

Figure 1.12: A possible coding scheme of Field's channels represented in the frequency domain
and in the space domain. Channels have constant spatial frequency bandwidths in octaves
and constant orientation tuning in degrees. Taken from Field (1987).

A further attempt to explore the characteristics of natural scenes using a larger sample of images was made by Tolhurst *et al* (1992). They analysed the amplitude spectra of 135 digitised photographs of natural scenes and found that relatively few conform exactly to Field's suggestion explained above (*amplitude* $\propto f^{-1}$). He plotted the amplitude spectra of his images in double-logarithmic co-ordinates and found that instead of being a straight line with slope equal to -1, about 25% show significant curvature and the average slope is -1.2. He also reports considerably varied slopes from

-0.8 to -1.5. These findings imply that it would be an over-simplification to consider the amplitude spectra of natural scenes as essentially the same and represented by: $amplitude \propto f^{-1}$.

The previous analyses have been performed on achromatic images. In this work we extend this study to the colour domain, examining how these properties vary, both across the visible spectrum and for chromatic images.

1.5 *Luminance and chrominance.*

The L-M opponent system can be thought of as signalling a push-pull between absorption rates of the L and M cone types. The zero output from this signal corresponds to a white, grey or black surface. It can also be achieved using a pure yellow surface (Hurvich and Jameson 1957). Light reflected from pigments with a relatively strong reflectance at short wavelengths would produce a chromatic output of one sign whereas light reflected from pigments with relative strong reflectance at long wavelength will cause a response with an opposite sign. It is possible to determine a spectral sensitivity curve for this opponent colour system using a hue cancellation technique. This technique consists of finding the relative proportions of a mixture of red (L wavelength) and green (M wavelength) light until the mixture looks neither reddish nor greenish to a normal observer. The spectral sensitivity curve is produced by repeating the measurements over different wavelengths (King-Smith 1991). Figure 1.13 shows the resultant spectral sensitivity of the L-M (chromatic) and L+M (achromatic) system calculated from Smith and Pokorny (1975) spectral sensitivity functions. These spectral sensitivity curves correspond to a weighted sum and a difference between spectral sensitivities of M and L cones, that is:

$$Achrom.signal = \boldsymbol{a_1}.L + \boldsymbol{a_2}.M;$$

$$Chrom.signal = \boldsymbol{a_3}.L - \boldsymbol{a_4}.M.$$

where a_1, a_2, a_3, and a_4 are weights to be determined (in Figure 1.13 they were considered as equal to 1 for simplicity). The achrom.signal (or *luminance*) carries information about the spatial distribution of light intensity. The chrom.signal (or *chrominance*) carries information about the spatial distribution of wavelength differences. Chrominance therefore refers to the purely *chromatic* content of the visual scene, and should be independent of luminance. Different ways of defining

chrominance satisfy this requirement to various extents. This will be examined in detail in Chapter 4.

The presence of "+" and "-" signs in the above equations reflect the assumptions that the responses of L and M receptors are linear. If we assume, for example, logarithmic responses, the equations above would reflect this fact by changing to:

$$Achrom.signal = L \cdot M \quad \text{and} \quad Chrom.signal = L/M.$$

Several workers support the view that the assumption often made that

$$a_1 = a_2 = a_3 = a_4 = 1$$

is only qualitatively correct. Estimations of the ratio between the weights have been made based on flicker and acuity criteria (Ingling Jr. and Tsou 1988) and on information theory based criteria (Buchsbaum and Gottschalk 1983). For simplicity in this chapter, all the weights will be considered as equal to 1.

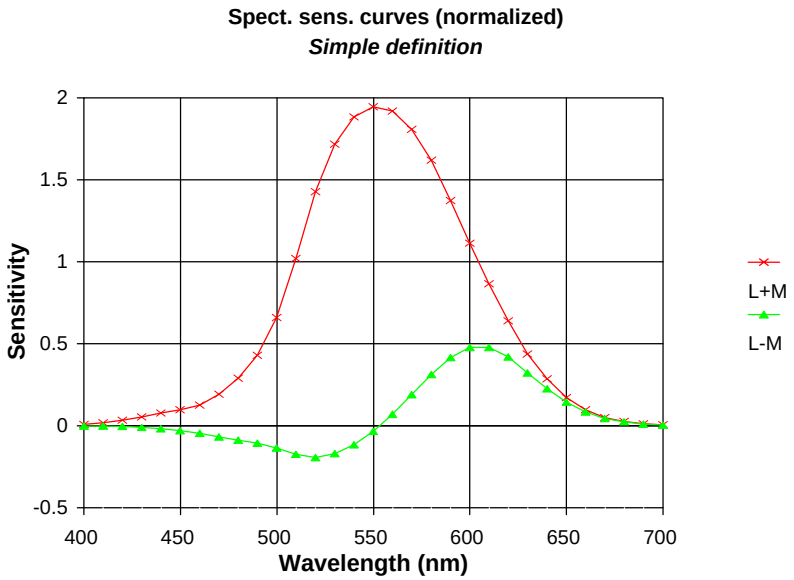

Figure 1.13: Spectral sensitivity of L-M and L+M systems. Calculated from Smith and Pokorny's (1975) data.

It is possible to plot the cone absorption (number of photons absorbed) of one cone type versus another for different natural objects when illuminated by direct sunlight. When the L and M cone's absorption responses for five common coloured mineral pigments are plotted in this way they tend to cluster fairy close to a line called "white line" or "achromatic line" (King-Smith 1991). Burton and Moorhead (1987) analysed nineteen natural scenes in terms of the relative absorption levels in the L, M and S sensitive

cones. Their analysis led to a similar conclusion. These observations have implications for the transmission of visual signals to the brain. If independent signals from the L and M cone types were sent individually along the optic nerve, the information content of the scene would tend to be duplicated in the two channels, leading to inefficiency in transmission. Using only information theory based criteria (Buchsbaum and Gottschalk 1983) it is possible to show that efficient information transmission is achieved by the transformation of the initial three colour mechanisms (L, M and S) into an achromatic (L+M) and two opponent chromatic channels (L-M) and (S-(L+M)).

The transmission of achromatic (L+M) and chromatic (L-M) information through the *parvocellular* optic nerve neurons has been modelled using a multiplexing scheme (Ingling Jr. and Martinez 1983). Such a scheme consists of transmitting both signals within different non-overlapping spatial frequency bands (a similar method to the one used in AM radio transmission where different radio signals are transmitted in different non-overlapping frequencies). For this analysis the simple-opponent L-M units that predominate in primate fovea (Derrington *et al* 1984a) were assumed to form a single channel signalling both chromatic and achromatic information. The problem here is that the chromatic and achromatic signals do, however, overlap (Lennie and D'Zmura 1988). Looking for an efficient decoding strategy, Ingling and Martinez postulate that it will be advantageous to know what kind of visual environment the system to be designated must operate in and to know what the signals are used for. They propose a demultiplexing strategy that interprets high s.f.'s modulation as luminance and low s.f.'s modulation as chrominance and the boundary is decided to be somewhere in the middle. This demultiplexing is clearly incomplete (e.g. it predicts that low s.f. luminance would be seen as colour modulation), but is attractive as it is consistent with the attenuation of high s.f. chrominance perception found by psychophysical experiments (Mullen 1985).

1.6 *"Ecological" explanations for the properties of the luminance and chrominance coding by the h.v.s.*

As seen before, there are three major factors (hardware constraints, priorities for survival and physical content of the environment) that could justify the broad characteristics of the h.v.s. If we expect the constraints imposed by the environment to be the most relevant, the facts already known about the h.v.s should correspond with the statistics of natural images. In particular, the psychophysical results obtained by Mullen

(see Figure 1.8), should have a correspondence in the statistics of the visual environment. Specifically, the luminance to chrominance signals ratio of natural images should be greater than 1 for high spatial frequencies and smaller than 1 for low spatial frequencies (a more extensive analysis of this is provided in Chapter 4). Several workers have been exploring to what extent the characteristics of the visual environment are consistent with facts already known about the h.v.s. and trying to find a sort of "ecological" justification of these facts. Following this line, Derrico and Buchsbaum (1991) proposed a computational model that decomposes a scene into an achromatic component and a chromatic component from the L and M initial components using an eigenvector (Karhunen-Loeve) transformation. They claim that their model has psychophysical correlates in the visual system and can be implemented physiologically using simple combinations of retinal receptive fields. In their analysis of a set of three scenes they found that the achromatic images contain the high spatial frequencies and the chromatic images have little or no energy in this range. However, the small set of scenes used as input for this analysis and the fact that not all of them are natural, limit the generality of their conclusion.

An analysis of the Fourier energy distribution of the chromatic and achromatic components was later performed using a larger set of scenes (Brelstaff and Troscianko 1992). The results only weakly support the idea that physiological imbalance between colour and luminance transfer functions is reflected in the spatial frequency content of natural scenes. The authors used the same set of scenes analysed by Burton and Moorhead (1987) in which, unfortunately, saturated objects like flowers were deliberately avoided and thus the range of chromaticity values is very small. The input set of scenes was larger (twenty scenes) but arguably still not a representative sampling of the natural environment due to its very limited gamut of colours. To give an approximate idea of what we might expect for the *lum/chrom* energy distribution to be - assuming that the contrast sensitivity functions (Mullen 1985) reflect somehow the statistics of the environment- we proceed as follows:

Figure 1.14 shows another plot of psychophysical measurements of the contrast sensitivity functions for luminance modulation and for isoluminant-colour modulation (taken from Mullen 1985). In this case the *x*-axis have been divided into 8 SF-bands (similar to those produced by Brelstaff and Troscianko (1992) and to those produced by

"FBANDS" -see Chapter 2, Table 2.2). Vertical lines represent the boundaries of these bands in the Fourier space in the plot. The *band-pass* and *low-pass* nature of luminance and chrominance signals are manifest in these contrast sensitivity functions. The area (integral) under each curve in each interval is calculated and plotted on Figure 1.15.

Notice that the x-axis in Figure 1.14 is logarithmic and because of this the areas determined by intervals on the right side are much bigger than the areas determined by intervals on the left side of the graph.

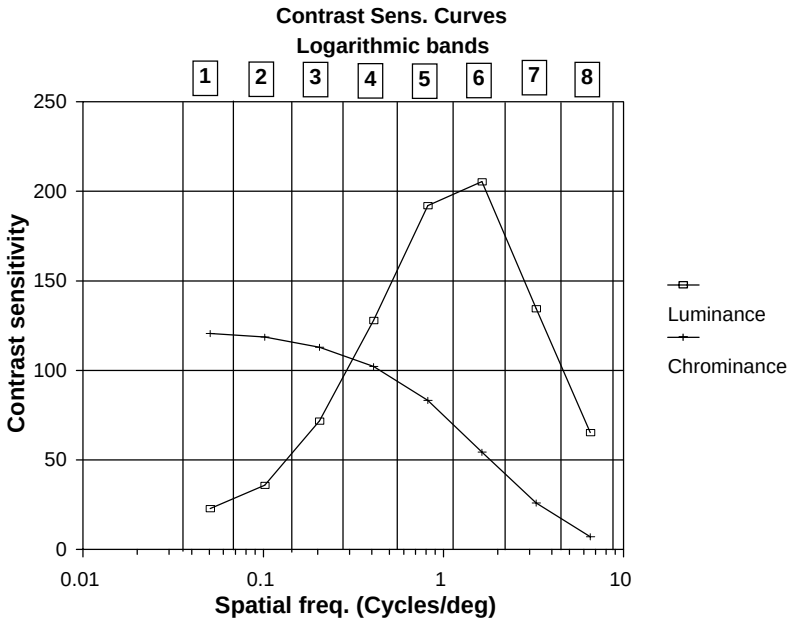

Figure 1.14: Psychophysical measurement of the contrast sensitivity functions (Mullen 1985). Sf-bands are represented on the plot by vertical lines.

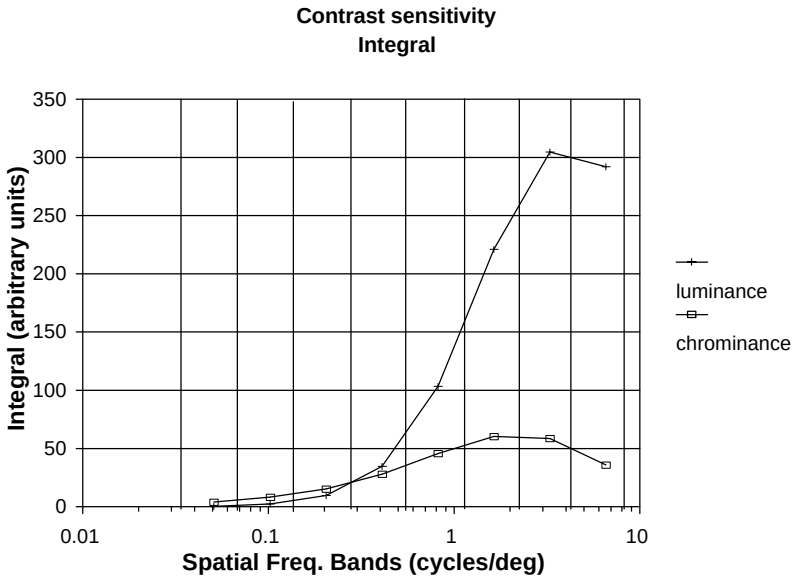

Figure 1.15: Value of the integral under the curves of the previous Figure.

Following this, curves on Figure 1.15 are scaled so that the total area under each of the contrast sensitivity curves for all intervals equals 1.Figure 1.16 shows the *lum/chrom* area ratio for each band. It also gives an approximate idea of what we might expect to find in the spatial frequency analysis of the natural scenes if the psychophysical measurements match the physical properties of the environment.

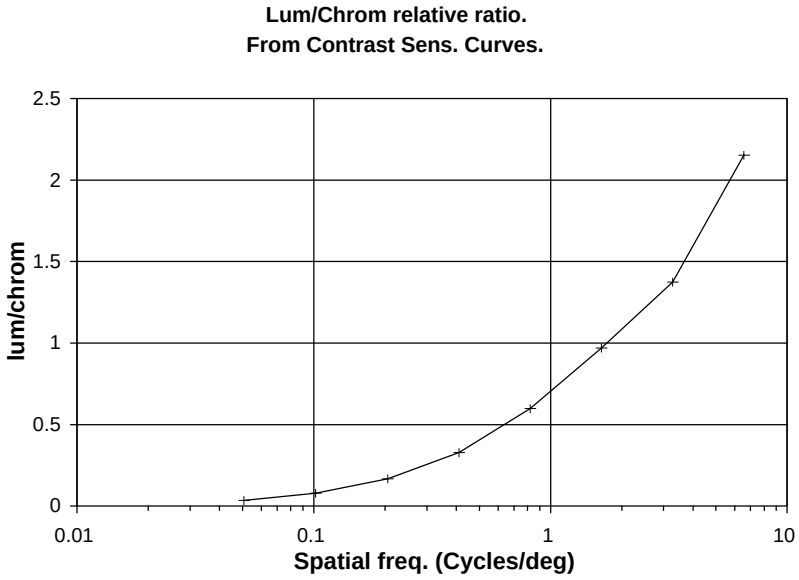

Figure 1.16: *Luminance to chrominance* ratio obtained from the contrast sensitivity curves (Mullen 1985).

When it comes to model the spatial and chromatic properties of the h.v.s., the “ecological” approach requires a consistent knowledge of the statistical properties of the visual environment. Efficiency in representing information in the visual system and its potential evolutionary advantages plays also a major role in this approach (Brelstaff and Troscianko 1992; Derrico and Buchsbaum 1991; Buchsbaum and Gottschalk 1983; Field 1987, 1989; Burton and Moorhead 1987; Atick 1992). If one starts from this "efficiency principle" along with biological constraints it is possible to predict neural processing. Thus some of the cortical computational strategies have been proposed from physical properties of the stimulus environment (Derrico and Buchsbaum 1991; Buchsbaum and Gottschalk 1983; Field 1987, 1989).

1.7 Aim of this work - The questions.

The aim of this work is to report some of the statistical properties of the visual environment using a larger sample size than has been used before and to relate the properties to observed characteristics of neural processing (such as the number of initial colour mechanisms -cone types-, shape and location along the wavelength axis). Specifically, the questions addressed are the following:

Q1) What is the mean distribution of spectral reflectance and radiance of natural scenes? Are the Smith and Pokorny (1975) cone sensitivity functions consistent with an efficient sampling of this mean distribution?

Q2) Is there any significant variation of the average distribution of spatial frequency (Fourier content) of natural scenes with wavelength and with spectral bandwidth of the receptor? Could these variations have determined some of the spatial properties of the colour receptive fields of the h.v.s.?

Q3) The models of spatiochromatic coding represent the chromatic signal as a linear combination of L and M cone output. Is there any significant variation of the average distribution of spatial frequency in terms of chrominance (L-M) related to (a) the separation in the wavelength axis between L and M receptors and (b) their bandwidth?

Q4) Is the physiological imbalance between colour and luminance (Mullen 1985) reflected in the physical spatial frequency content of natural images? Does this depend on the way we model the opponent interaction between cones?

Q5) Are the characteristics of the amplitude spectra previously found for achromatic images similar to those of *chrominance (L-M) images*? Does this depend on which definition of *chrominance* we use to model this opponent interaction?

Chapter 2

Methods

Technical background

For this exploration of the statistical properties of the natural environment we require a representative sample of natural scenes. This implies also a system for image acquisition (camera), image processing (computers), storage (disks, tapes, etc.) and display (monitors). We require for the system to provide a measure of radiance as a function of wavelength for each pixel of our images. Using this information we calculate the statistics of the environment and contrast them with some of the known features of the h.v.s. An ordinary colour (CCD) TV camera seems to be the most obvious choice except for the fact that it makes use of its own colour model. Colour models are 3D co-ordinate systems in where each colour is represented by a single point which facilitates the specification of colours in some generally accepted (standard) way. Most colour models in use today are either hardware-oriented or applications-oriented and their assumptions do not represent the world directly. Furthermore, any trichromatic sampling of this kind does not allow subsequent recoding into, for example, more than three channels or into channels of varying bandwidths. We wished to have the freedom to be able to model any possible *n-channel* sampling where *n* is less than 31. Another possible option is the use of photographic film and a digitizer. Here the process of calibration is more complicated because the output has to be compensated for the non-linearities of the photographic process. The main reason, however, for the choice of video rather than photographic technology was the greater ease of digitising the images of a video representation.

In order to provide a more direct (and free of erroneous assumptions) representation of the natural environment we employ a multispectral digital camera as described in the following sections.

2.1 The Defence Research Agency-funded multispectral camera

The gathering of all statistical properties of a large sample of natural visual images used in this work was made by a video camera capable of recording images through narrowband interference filters and storing the resultant images on a computer disk. The development of the camera was funded by the Defence Research Agency (DRA).

The DRA-camera system and controller device were partially constructed by Custom Camera Designs Ltd. (of Wells, Somerset) and completed, assembled, calibrated and operated by members of the Perceptual Systems Research Centre (PSRC) of the University of Bristol between 1990 and 1993. Originally, it was constructed to achieve the following aims:

a) To construct an electro-optical (e-o) system capable of representing the spectral distribution of light in terrain scenes.

b) To validate the e-o system with calibrated images.

2.2 Characteristics of the DRA-camera

The DRA-camera consists of an e-o mechanism built around a "Pasecon" tube, a camera control unit (CCU), a carousel slide changing filter mechanism, a portable 386-PC, a real-time video monitor and a battery power supply. The scheme of the DRA-camera is shown in Figure 2.1. The Pasecon tube was chosen because it has good linearity throughout the full spectral range and a low-noise output at low light inputs. Figure 2.2 illustrates the camera tube sensitivity between 400 and 800 nm (from manufacturer's data). The original Camera used to convert the video signal to digital form (analogue/digital conversion) for storage in the internal frame-store card and then reconvert to analogue form (digital/analogue conversion) for both display and as input to the Data Translation frame card. Given that the final digital/analogue device proved to be very prone to drift with temperature and since the required images were to be stored and processed in digital form, digital images were transferred directly from the camera after the first analogue/digital conversion. A card was designed and built to enable clean and noise-free images to be transferred digitally directly from the camera's first a/d converter into the frame-store on the Data Translation card.

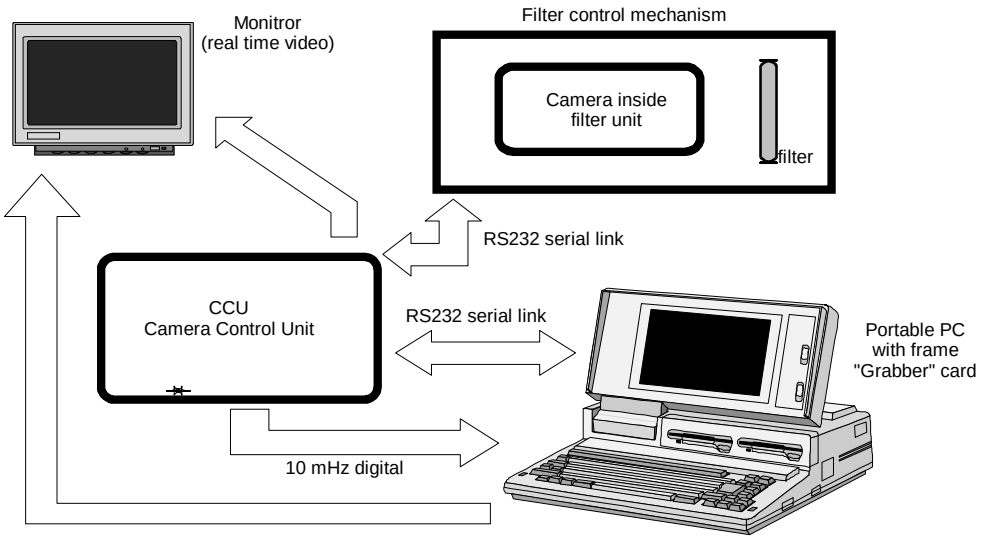

Figure 2.1: Scheme of the DRA-Camera.

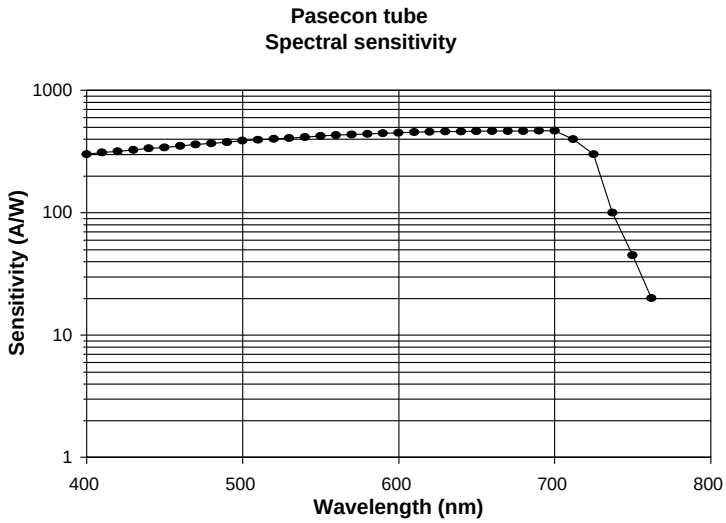

Figure 2.2: Camera tube sensitivity between 400 and 800 nm (from manufacturer's data).

The slide changing filter mechanism was added to the e-o mechanism to allow the DRA-camera to sequentially grab images through a set of 31 optical filters. These are chromatically narrowband in the range of 400 to 700 nm and with 10 nm spacing between their peaks. Figure 2.3 shows a plot of transmittance versus wavelength for the whole set of filters employed. The entire system is controlled by the portable 386-PC and mounted on a trolley along with a 12 Volts inverter to supply 240 Volts mains in the field. Manual fixed focal length lenses (Fuji CF25B, f/1.4, 25 mm) with a field angle of 28.71 x 21.73 deg were used. The marked aperture settings on the lens are: f1.4, f2, f2.8, f4, f5.6, f8, f11 and f16.

A PC program (written in 'C' under MS-DOS 3.0) sends a value for the integration time (see below) to the DRA-camera control unit which grabs the image and transfers it from its frame card to the PC frame card. The whole system allows a sequence of 31 chromatically narrowband filtered 8-bit images (256 x 256 pixels x 256 grey-levels) to be grabbed. These images correspond only to the central part of the visual field supplied by the lenses. The field angle of this picture is thus equal to 14.35 x 14.35 deg. Given this, each pixel subtends an angle of 0.056 x 0.056 deg (3.36 arc minutes) approximately. This value is of the order of the size of a foveal cone diameter (1 arc minute approximately). Once recorded, the image set is transferred to an IBM (RS/6000) workstation for processing.

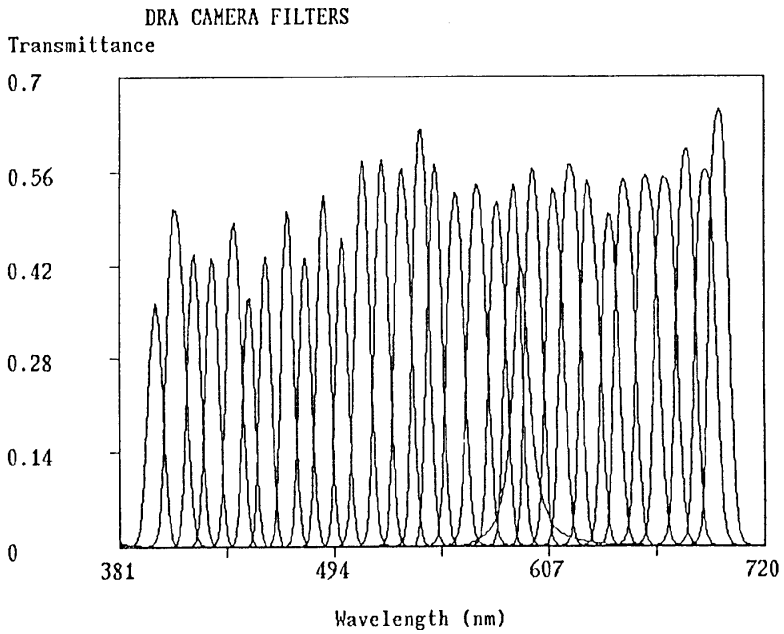

Figure 2.3: Transmittance of the filters employed.

A graphical user interface (written in 'C' under AIX on the RS/6000 accessing the OSF/Motif X-Windows System) allows convenient access to the data as grabbed image sets. Using this it is possible to display each image on screen and to display a spectrograph of the light at the point in the image where the mouse cursor has been pressed in the display window. Either grey level, radiance or reflectance can be shown on the graph. Figure 2.4 illustrates this facility. This requires the images to be calibrated for reflectance. This was achieved in the manner described below.

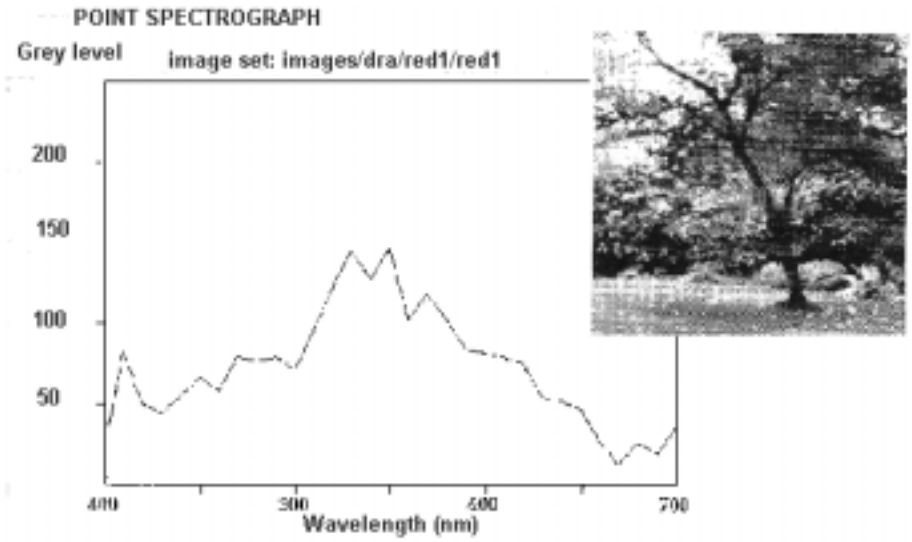

Figure 2.4: Display window of the IBM-RS/6000 employed to process the dataset.

2.3 *The DRA-camera calibration*

The DRA-camera system was calibrated to write out image files along with relevant header information such as the integration time and lens aperture from which multi-spectral light-measurements can be reconstructed. Figure 2.5 shows a measurement of the system output versus time since system boot. Notice that the minimum time necessary to ensure a stable output was around 6000 seconds (1hour 36 min). All measurements described below were obtained several hours after the system boot.

The calibration strategy consisted basically of:

a) Finding for any given filter and aperture setting the number of integration frames required to achieve a reasonably large dynamic range.

b) Correcting the non-linear characteristics of the DRA-camera and the off-axis variation across the camera target.

c) Converting the grey level output of the DRA-camera into measurements of the spectral radiance (and spectral reflectance) across the target.

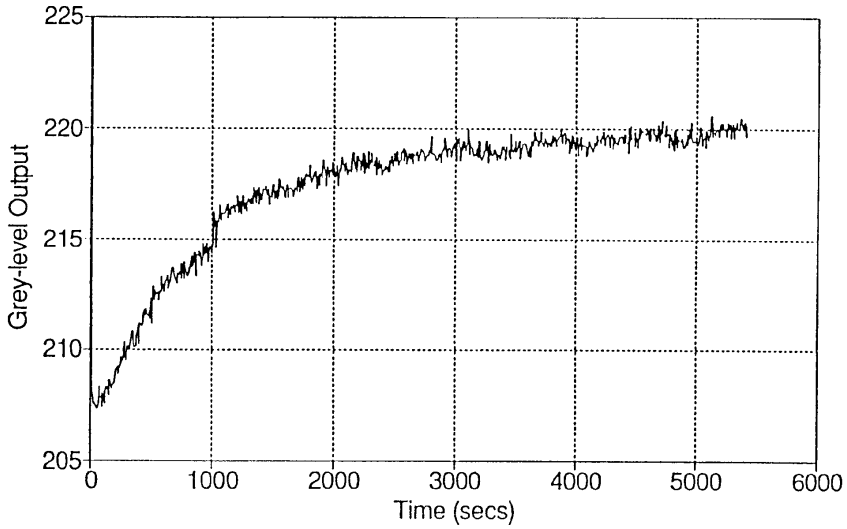

Figure 2.5: System output versus time since system boot.

(a) Choosing the right dynamic range for each image

Very bright regions like sky and specularities are not of direct interest, and would severely compress the dynamic range (0, 255) if properly represented. No sophisticated strategy was used to identify these bright regions. Instead, the system was created to allow the adjustment of the integration time until bright regions, other than sky and specularities, produce responses near to (but not above) the maximum. The real-time video monitor provided an effective way to identify these bright regions on the field. For example, it is possible to choose a ceiling value of 254 (in grey-levels) so that 90% of the recorded scene will not exceed it. Alternatively, it is possible to choose among nine small square regions regularly distributed on the scene and manipulate its statistics (e.g. median, mean, maximum). For example, one might select the median value within the small square region in the centre not exceeding the ceiling.

Once both the statistics (S) and ceiling (C) are defined the algorithm tries to automatically find the value for the integration time (IT) so that S(IT) is near to (but not above) C. The accepted difference between C and S can be specified by a tolerance parameter. Given that this calibration is repeated for each filter on the carousel, an algorithm has been developed to accelerate the process. This algorithm can predict a suitable value for IT using the results of previous attempts, and check if the light levels change between individual trials. If that happens, the calibration process is halted and the problem reported.

(b) Corrections of non-linear characteristics of the DRA-camera

To correct the non-linear characteristics of the DRA-camera, a look up table (LUT) was produced as follows:

The DRA-camera was pointed to a blank piece of white paper (test card) the radiance of which was 0.002 WSr^{-1} nm^{-2} m^{-1} as measured with a TopCon Spectro-radiometer (model SR1). The test card was lit using a steady illuminant, a tungsten bulb with constant current through it. To avoid possible contributions from light with wavelengths outside the visible range the 570 nm interference filter was always used and all measurements were taken in a dark room. As Figure 2.2 shows, the camera tube sensitivity decays drastically in the infra red (IR) part of the spectra ensuring that no IR radiation will affect the measurements in spite of any secondary peak the filter could have in this region. Figure 2.6 shows the apparatus set up.

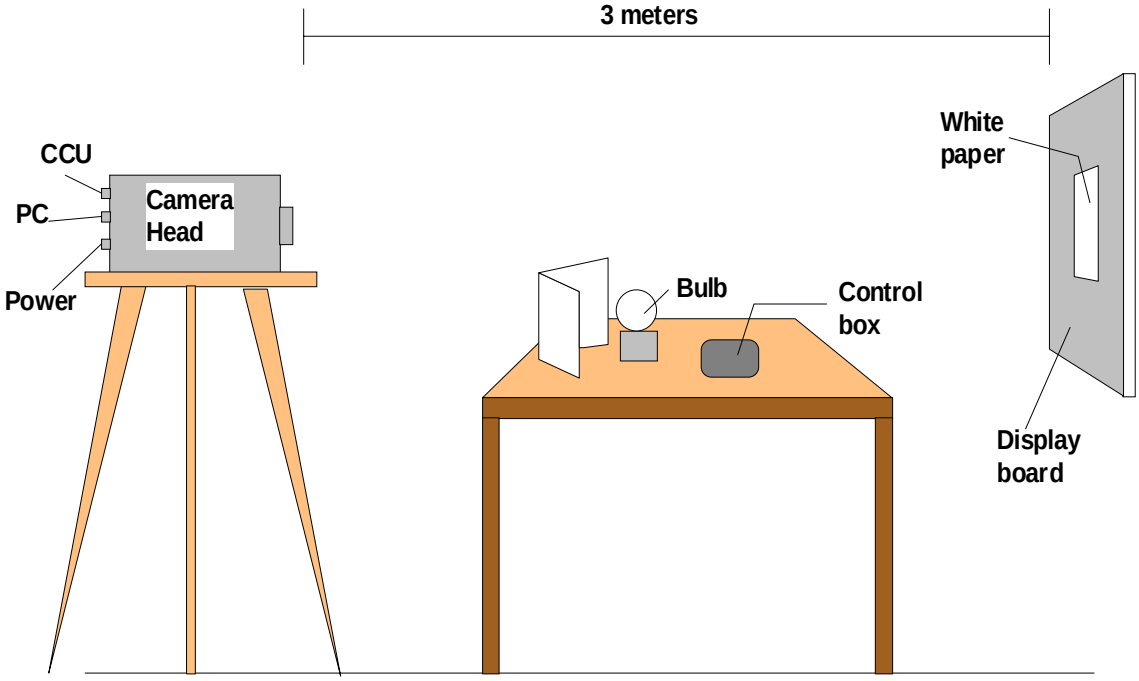

Figure 2.6: apparatus set up for calibration.

The system output within the central square region was estimated by varying the integration time from 1 to 200 frames for all 8 lens aperture settings. This provides a series of eight curves. To obtain a further eight curves a transmission filter with neutral density 0.5 was slotted into the camera behind the 570 nm filter. The procedure described above was repeated 10 times to obtain variability data and 16 exemplary curves (one for each of the above conditions) were selected and used to create the look-

up table.

The same procedure used for calibrating the centre of the scene was repeated for all nine regions of the scene. An additional set of 8 further regions lying on the perimeter of the scene was used to provide extra information about the off-axis output. As before, this data was stored in an off-axis correction file that was simultaneously used with the LUT to convert the DRA-camera output into light measurement.

(c) Conversions into spectral radiance

This was obtained by matching the output of the DRA-camera through each of its 31 filters to that of the TopCon SR1 when both systems were pointing to a standard Kodak grey card. Further measurements of the spectral radiance by the DRA-camera were compared to that obtained with the TopCon SR1 using a collage of coloured papers illuminated with a constant current tungsten lamp. The relative error of the matching between both systems was less than 5% in the range 400-570 nm. In the range 580-700 nm the relative error was bigger but always less than 10%. For one red sheet the match for some unknown reason was out by as much as 20% in the spectral range 650-700 nm. Figures 2.7(a) and 2.7(b) provide an example of this matching.

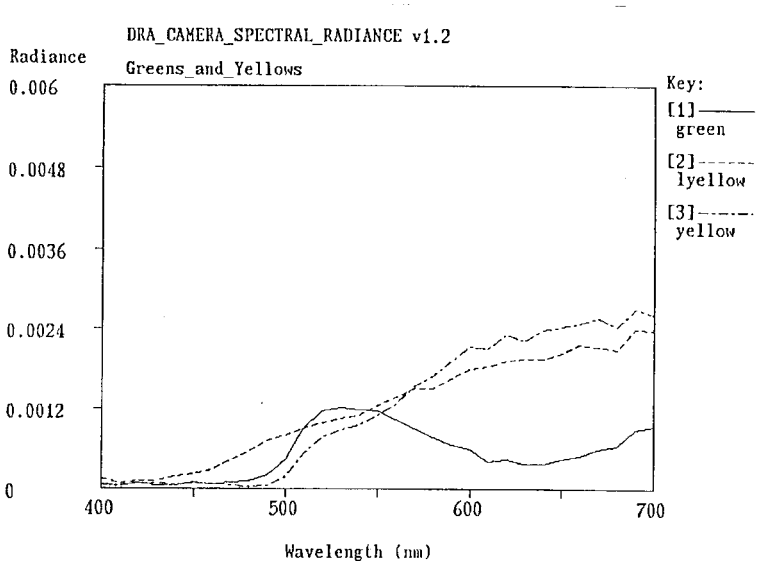

Figure 2.7(a): Spectral radiance obtained with the DRA-camera

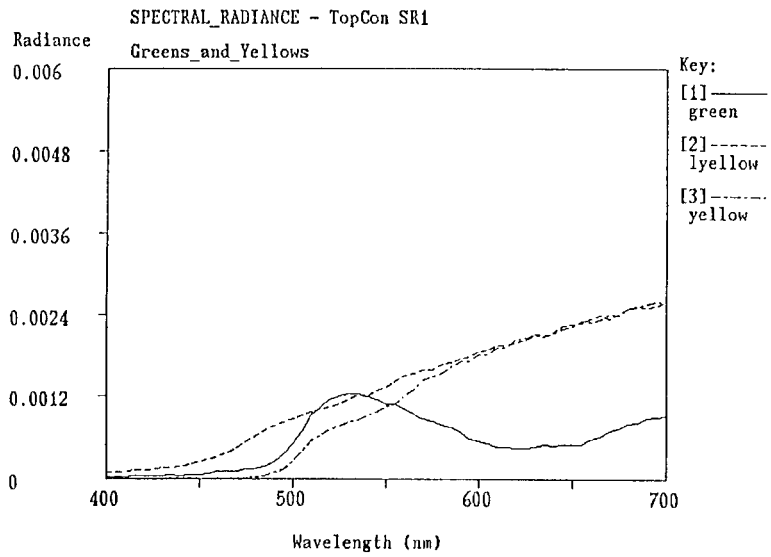

Figure 2.7(b): Spectral radiance obtained with the TopCon SR1.

(d) Conversions into spectral reflectance

To convert the DRA-camera output into measurements of spectral reflectance, it is necessary to know the spectral characteristics of the light falling over the scene.

$$\textit{Scene Radiance} = \textit{Illumination} \times \textit{Reflectance}$$

This information can be obtained by placing an object of known reflectance (like the Kodak standard grey card) into the scene and measuring its radiance. The algorithm allows the introduction of the XY position of the grey card on the screen and performs the conversion automatically. This assumes that light falling over the grey card is an accurate sample of the light falling over the whole scene and that the reflections are approximately independent of the subtended angle (Rayleghian or diffuse reflections).

2.4 *Scene grabbing*

A dataset of 31 scenes was obtained for this analysis between September 1993 and January 1994. Each of these scenes contains 31 images taken through a different spectral filter. The normal recording procedure was as follows:

a) Warming up of the DRA-camera systems and battery checking. This was usually done the day before to ensure that the system would be fully warmed during the recording. As shown on Figure 2.5 these temperature changes can alter the system

output.

b) Selection of the place and time of the day for the recording. This was done the day before (after checking the forecast information). To avoid changes in light because of changes in sun position, all pictures (except the ones taken inside the lab) were generally taken around noon. Several places were chosen depending on their facilities for moving the trolley with the equipment and the drift of people around. Several scenes were recorded near the Royal Fort House of the University of Bristol, others in the University of Bristol Botanical Gardens. Some practical problems that influenced our selection of the place and daytime will be discussed in the next section.

c) Positioning of the grey card in a visible place in the scene. Most of the pictures (29) contained no artificial objects with the exception of the Kodak grey card, and (in two cases) a tripod. The card was placed in all scenes for calibration and light checking purposes.

d) Recording of the spectral radiance of the grey card using the TopCon SR1. This was made to detect changes in the spectral characteristics of the light before and after every picture recording. The measurements were taken from approximately the same place as that of the DRA-camera and just before starting the proper recording of the scene.

e) Recording of the scene. Considering the relative long time (about 5 minutes) needed to grab each scene it was often necessary to wait until weather conditions were acceptable (see next section about practical problems) before starting.

f) Second spectral measurement of the grey card radiance. The procedure was similar to (d) and was made immediately after finishing the scene recording.

g) A photograph was taken of approximately the same scene to be used as a colour image reference.

h) Back in the laboratory the grabbed images were transferred into the IBM/6000 workstation for further analysis.

2.5 *Common practical problems and their solutions*

Most of the practical problems we faced were related to changes in the scene during the recording. A different solution was formulated for each one.

a) Linear light changes. These were generally caused by changes in the sun position during the recording of the scene. These alterations never affected the spectral properties of the light and were detected by comparing the SR1 measurements before and after the recording. To avoid them, most of the recording was done around noon and a special algorithm "*reilluminate*" was developed to minimise its effects (see below).

b) Light fluctuation, mainly due to short-term variations in cloud cover. This generally affected one or two of the images in the scene. To detect them, it was necessary to compare the SR1 spectral measurements on the grey card with those of the DRA-camera in the same place. The most effective way to avoid these fluctuations was to wait until the sky was completely clear or completely overcast before grabbing the scene. Small errors were corrected with the "*reilluminate*" program (see below).

c) Small movements of objects (such as tree branches) in the scene during imaging. These were due to wind and their negative effects were more intense for objects at a short distance. Negative effects were minimised by grabbing medium and long distance pictures but often it was necessary to wait for these movements to cease. All the naturally lit short distance scenes were taken in the glass houses of the University of Bristol Botanical Gardens in order to avoid this problem.

2.6 *Corrections made by software*

The "*reilluminate*" program was created to correct small light fluctuation and linear light changes of particular scenes of the dataset. It uses the spectral reflectance image of the scene along with one of the SR1 measurements (the program allows one to choose between any of them) taken on the grey card. It "reilluminates" the scene with the light spectrum falling over the grey card so that the radiance measured with the DRA-camera on the grey card matches the SR1 measurement. Appendix A gives a detailed explanation of its fundamentals along with the DRA-camera mathematical theory.

Modifications of the data were avoided as much as possible and the use of this kind of correction was limited to five particular cases of the dataset. To decide whether it was necessary to use this program or not, the following points were considered:

a) The fit between SR1 measurements taken on the grey card before and after the

recording of each scene. A special algorithm (Press *et al.* 1991) was applied to compare both spectral measurements.

b) Comparison between the spectral radiance of the grey card as measured by the SR1 and the spectrum obtained by the DRA-camera in the same place.

c) Light and wind conditions on the day when the scene was recorded. This information was registered in a table for each scene of the dataset.

After correcting a scene, the comparison described in (b) was performed again in order to evaluate the effects of the "*reilluminate*" program. However, all further analyses in terms of Fourier amplitude were performed separately on the corrected dataset and on the original one, and there does not appear to be a marked difference between both sets of results.

2.7 *Some statistics of the dataset*

Although there is no formal agreement about what is considered a representative sampling of the visual environment, we ensure that (within our limitations) some of the most common natural objects are represented in our dataset. These include plants with different shapes, textures and colours, flowers (mostly in bright red colours), trunks, branches, grass, yellow leaves, trees, bushes, rocks and sky. Some of these objects were artificially arranged in the laboratory trying to make this arrangement to look as casual as possible. Others are just images of the British countryside (in which buildings or other human artifacts were avoided) or of gardens (either taken in the Botanical Gardens or in the University of Bristol Royal Fort Gardens). Our general aim was to generate a set of images which could conceivably be representative of the environment in which primate vision evolved. Of course, we cannot be sure that we have achieved this aim. However, we felt that the inclusion of vegetation seen from different distances might satisfy this requirement in part. Later on we shall discuss the extent to which our images (obtained in typical diffuse Northern Temperate illumination) may differ from images obtained under more directional illumination. Practical constraints limited our choice of possible images. Bright parts such as sky result in large saturated portions of the image and were deliberately avoided. Only four of our scenes contain regions of sky. Moving objects such as clouds are also avoided. Very strong shadows produced by

solar illumination move during the period of image acquisition and are consequently avoided. Water (i.e. lakes, ponds, etc.) is also avoided because both of the bright reflections and the movement. All the scenes were recorded in Autumn and Winter (between October 1993 and January 1994) and this can be considered as a limitation in the range of possible natural colours and textures.

Of the 31 scenes recorded, 29 contained no other artificial object than the grey card, and in two cases a tripod to attach it. These were the ones used in our analysis. The remaining two included artificial objects and were used to test the quality of the lenses (as is described in Chapter 5).

Different classifications of the dataset can be performed depending on the characteristic considered:

a) In terms of lighting, nineteen scenes were illuminated by the sun (these are called *sun-illuminated*), directly or through cloud cover, nine were illuminated by incandescent lighting and one by fluorescent lighting (these are called *artificially-illuminated*).

b) In terms of distance to the main objects in the scene, four *long distance* scenes were taken. This means that the main objects were in the range 0.1-4.0 km. The rest of the dataset includes objects in the range 0.5-50.0 m (called *short-distance* scenes).

c) In terms of the corrections performed, the dataset can be classified in three different categories:

> Corrected scenes (5 in total). "*Reilluminate*" was used to correct light fluctuation on several filtered images.
> Slightly corrected (14 in total). "*Reilluminate*" was used to produce the adjustment of the DRA-camera measurement to that of the SR1 on the grey card.
> Without corrections (10 in total) These were also called *selected scenes* in later chapters.

2.8 *Software tools*

Several software tools were developed in 'C' to run on scenes on the RS/6000. The most frequently used in this analysis were:

"DO_LMS": This uses as input the 31 narrowband images that constitutes a given scene. The program first converts the value of each pixel into radiance or reflectance (as described in section 2.3) according to a selection by the user. A special option allows the user to implement the corrections described in section 2.6 (i.e. the “reilluminate” algorithm). Following this, the program weights each of the 31 monochromatic images and adds them together to generate each of the L, M, S cone sensitivity images. The weighting is produced according to the Smith and Pokorny (1975) cone sensitivity functions (see Figure 1.3) which are normalised to have a maximum sensitivity value equal to one. The output of this program is a set of three floating-point images, each of them corresponding to the L, M and S cone sensitivity (see Figure 2.8).

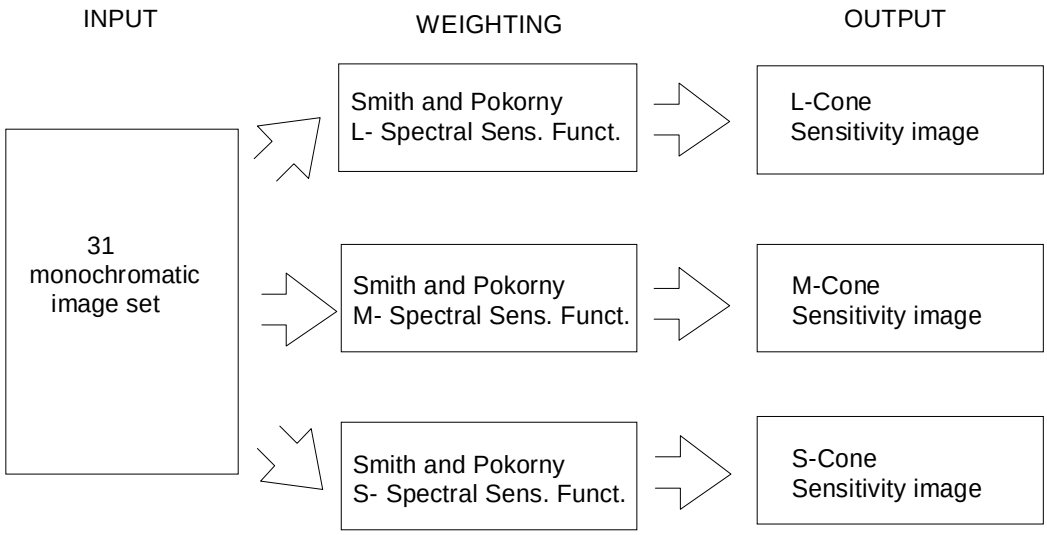

Figure 2.8: Scheme of the “DO-LMS” algorithm.

"FBANDS": This produces the Fourier amplitude spectra of an image and displays it as a 2-dimensional intensity function. For example, if "FBANDS" is applied to a floating-point image A, the output will be another floating-point image where brightness is proportional to the amplitude spectrum of A. In this representation, the distance from the centre of the image is proportional to the spatial frequency value and the angle from the horizontal represents the orientation (see Appendix B). Given that the dynamic range of Fourier spectra is much higher than the display device is able to faithfully reproduce (255 grey levels), the centre of the plot has been removed. This corresponds to the highest value and is analysed separately in our analysis (it also represents the average luminance of the image A). Figure 2.9 shows a 3-dimensional diagram of the Fourier amplitude spectra of an image.

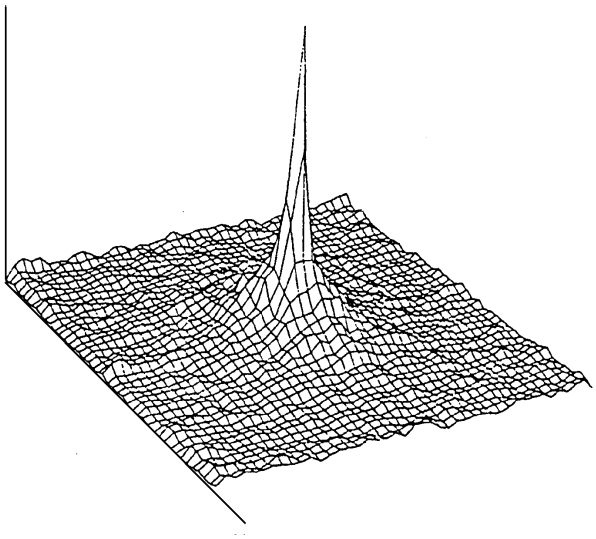

Figure 2.9: 3-dimensional diagram of the Fourier amplitude spectra of an image. Taken from Field (1987).

The program also divides the Fourier space into 8 concentric annuli (bands), each of them representing a given range of the spatial frequency spectrum. The total Fourier amplitude and its average value within the band (averaged across orientation) are measured into these bands and the results presented in a table. It is possible to choose between two different sets of concentric bands, the *"linear"* and the "*logarithmic"*. Tables 2.1 and 2.2 show the range (in pixels measured from the centre of the screen and their equivalent in cycles/deg) covered by each band in both cases. Figure 2.10 shows a scheme of this annuli-shaped division. These *linear* and *logarithmic* band spacing produce different results when applied to our dataset (see later in Chapter 5 for a more extensive discussion of this).

We choose the number of logarithmic bands to be equal to 8 because this is a convenient number for dividing the 256 x 256 pixels screen. In this case the minimum bandwidth corresponds to one pixel (see Table 2.2). To keep the symmetry between both versions, the number of bands is equal to 8 in the linear case too.

All the software tools were developed by the PSRC between 1991-1993.

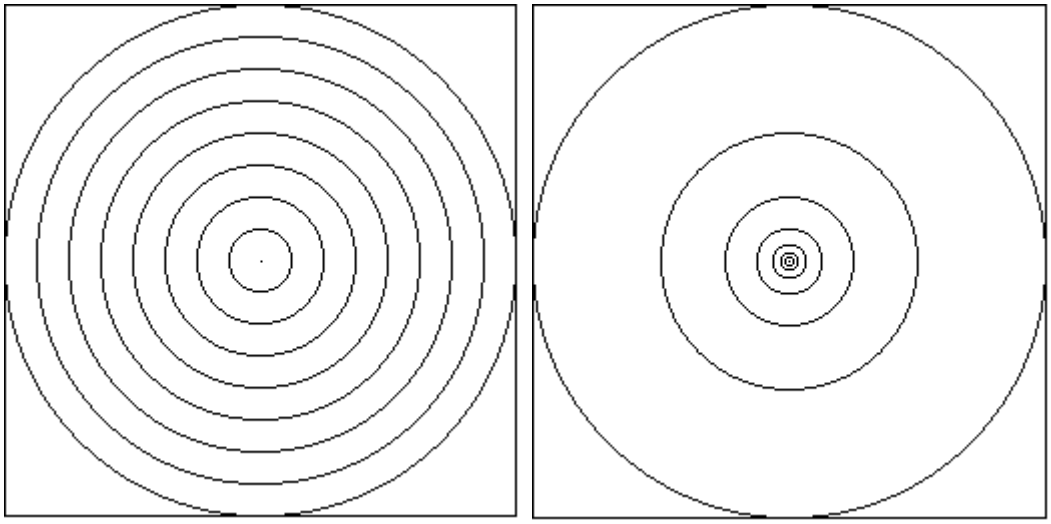

Figure 2.10: Annuli-shaped division of the Fourier space. *Linear* (left) and *logarithmic* (right).

2.9 *Validity tests*

Several validity and repeatability tests were performed to check whether the system (DRA-camera plus software tools) was working properly.

The DRA-camera output was compared to that of the TopCon Spectro-radiometer for several coloured objects to ensure that radiance/reflectance measurements were reliable.

Synthetic images of sine gratings with spatial frequencies of 0.008, 0.1 and 0.5 cycles/pixel were constructed using a special algorithm designed for this purpose. Following this, they were processed using FBANDS. The amplitude spectrum image showed the peaks corresponding to the spatial frequencies mentioned above. Logarithmic bands 2, 6 and 8 respectively showed amplitude values 1000 times greater than the rest.

A natural image was filtered using a spatial frequency low-pass filter and then processed using FBANDS. The results showed the corresponding shift in the Fourier amplitude spectra towards the low spatial frequencies. A similar analysis was performed using high-pass filters. The data obtained using the "*reilluminate*" algorithm was compared with the results obtained without it and the difference proved to be non significant for this analysis.

Further lab testing of the camera lenses were performed to see whether there is any

defocus on the blue side of the spectrum that could affect our analysis. These results are more intensively discussed in Chapter 5.

Table 2.1: Linear Bands				
Band	**Range**		**Centre**	**Nb**
	pixels	***cycles/deg***	***cycles/deg***	***Pixels***
1	1 - 16	0.03 - 1.10	0.57	888
2	17 - 32	1.10 - 2.20	1.65	2516
3	33 - 48	2.20 - 3.30	2.75	4116
4	49 - 64	3.30 - 4.40	3.85	5716
5	65 - 80	4.40 - 5.50	4.95	7352
6	81 - 96	5.50 - 6.60	6.05	8924
7	97 - 112	6.60 - 7.70	7.15	10564
8	113 - 128	7.70 - 8.80	8.25	12110

Table 2.2: Logarithmic Bands				
Band	**Range**		**Centre**	***Nb***
	pixels	***cycles/deg***	***cycles/deg***	***Pixels***
1	1 - 1	0.03 - 0.07	0.05	8
2	2 - 2	0.07 - 0.14	0.10	16
3	3 - 4	0.14 - 0.28	0.21	44
4	5 - 8	0.28 - 0.55	0.41	180
5	9 - 16	0.55 - 1.10	0.83	640
6	17 - 32	1.10 - 2.20	1.65	2516
7	33 - 64	2.20 - 4.40	3.30	9832
8	65 - 128	4.40 - 8.80	6.60	38950

The tables above show the range covered by each band of the FBANDS algorithm in the 8 bit representation of the Fourier space and their equivalent in cycles/degree. The band centre is the value used to represent a given band in all the plots of this work. Nb represents the total number of pixels included by each band. All conversions were done using the field of view of the DRA-camera.

Chapter 3

Physical quantification of spatio-chromatic properties of our dataset

As mentioned in Chapter 1, the quantitative knowledge of statistical properties of natural signals is required in order to relate "efficiency principles" and biological constraints to predict neural processing. The material in this chapter deals primarily with the exploration of further statistical properties of our dataset. Given the potential evolutionary advantages of an efficient representation of information by the visual system, in this Chapter we explore the spatial frequency structure with respect to wavelength of the dataset and report its properties. This analysis can be separated into two broad sections:

a) Report on the average variation over wavelength of the magnitude (centre of the Fourier spectrum diagram) of the signal.

b) Report on the average variation over wavelength of the modulation (Fourier spectrum not including the centre) of the signal.

Here the term “signal” refers to spectral radiance and reflectance. Given the diversity of our scenes, it is necessary to normalise the outcome of the statistical analysis to represent and compare them. A mathematical description of the normalisation is provided.

The dataset was also processed to produce new images by adjusting the spectral bandwidth and the location of a hypothetical receptor along the wavelength axis through the visible region of the spectrum. This was done in order to investigate the consequences of having photoreceptors different from our own.

Further consequences of our findings for spatio-chromatic coding in h.v.s. will be discussed in Chapter 5.

The human visual system is only sensitive in the spectral region of greatest solar energy. The sun emits a spectrum that is quite similar to that of a blackbody at a temperature of about 5777 K. Figure 3.1 (Siegel and Howell 1972) shows the hemispherical spectral emissive power of a blackbody for several different temperatures (see top curve in the figure). The coincidence between the peak of the solar energy and the visible region (between 400-700 nm) confirm that it is an evolutionary advantage for humans to gather visual information in the spectral region where the solar energy is maximised.

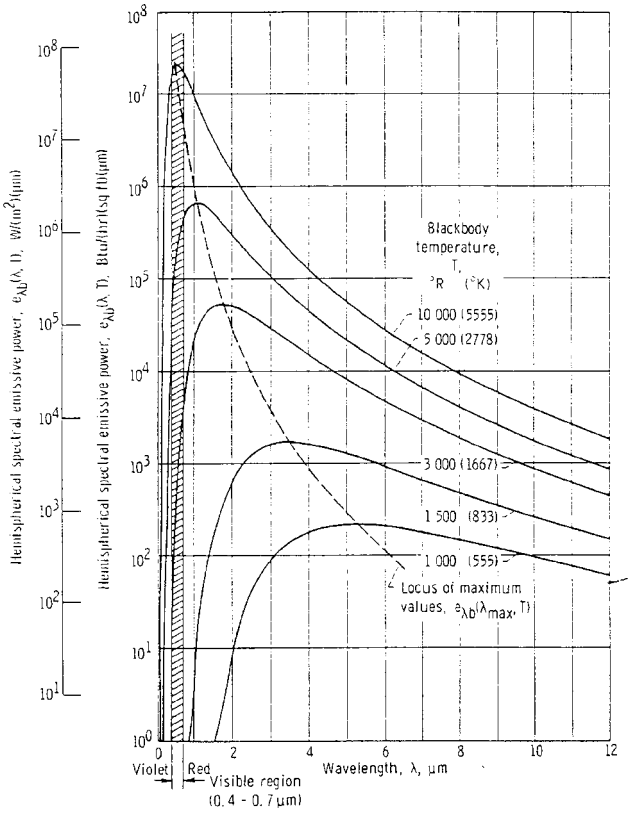

Figure 3.1: Hemispherical emissive power of blackbody for several different temperatures (Siegel and Howell 1972)

Curves B and C in Figure 3.2 (Le Grand 1968) correspond to standard illuminants (energy distributions) of direct sunlight and overcast sky used in colorimetry. Curve A corresponds to a tungsten (incandescent) source. They also represent spectral distributions of light reflected from a perfect white surface under these illuminants. Curves B and C in Figure 3.2 are examples of the typical energy distribution of the light

that illuminates our outdoor scenes and curve A is like that of the light illuminating our indoor scenes (with the only exception of one scene illuminated by fluorescent tubes).

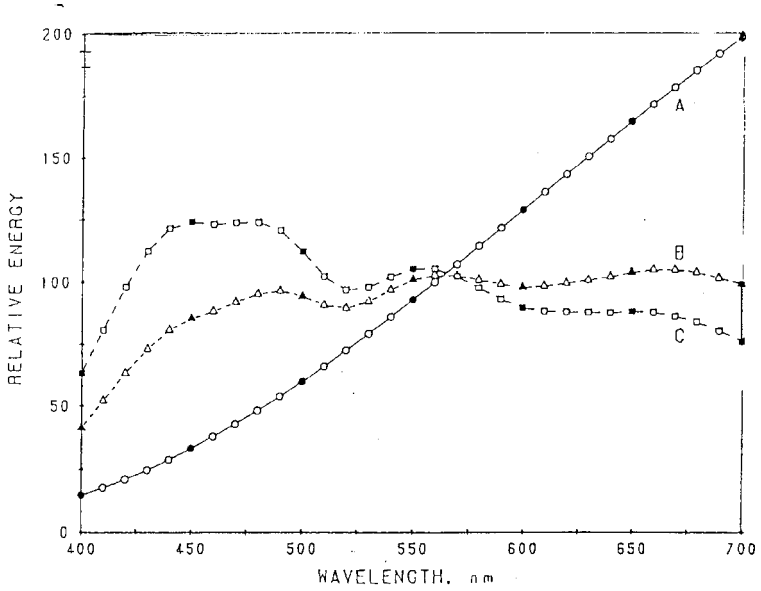

Figure 3.2: Typical energy distributions of light that illuminates our scenes: A- tungsten (incandescent) source, B- direct sunlight C- overcast sky (Le Grand 1968).

Most of the light that reaches the eye comes from objects that do not themselves emit light but rather reflect some portion of the light falling on them. The amount of light transmitted by a surface (i.e. its radiance) depends on both the quantity of illumination and its reflectance:

Scene Radiance = Illumination x Reflectance.

The h.v.s. extracts information from the radiance of the environment using transformations of the initial three colour mechanisms (L, M and S -see Chapter 1). Part of these transformations is reflected in the spectral sensitivity curves for the aforementioned *luminance (L+M)* and *chrominance (L-M)* channels. Using the DRA-camera and the TopCon Spectroradiometer, as described in Chapter 2, we were able to measure both spectral radiance and spectral reflectance in every point of our scenes.

The following section consists of a comparison between the human spectral sensitivity curves and the mean spectral radiance and reflectance extracted from our scenes.

To obtain the mean distribution of spectral radiance and reflectance across all scenes of our dataset, the data was processed using the following definitions:

a) **Image** (256 x 256 x 8 bit): chromatically narrowband filtered file grabbed using the DRA-camera. The index λ is used to indicate centre wavelength in nm: λ = 400, 410, 420, ..., 700nm.

b) **Scene**: each is a collection of *31 filtered images* grabbed with the DRA-camera. The index i is used to indicate each particular scene: ***i***= 1,2,3,...,29

c) **Dataset**: entire collection of 29 *scenes* recorded as described in Chapter 2.

Given:

$$S_i = (I_{i,400}\ ,\ I_{i,410}\ ,...,\ I_{i,\lambda}\ ,...,\ I_{i,700}.)$$

where S_i represents a particular scene of our dataset and $I_{i,\lambda}$ represents a particular colour filtered image (31 of which form the scene). The mean radiance within every image was obtained for all scenes

$$R_{i,400}\ ,\ R_{i,410}\ ,\ ...\ ,\ R_{i,\lambda}\ ,\ ...\ ,\ R_{i,700}$$

where $R_{i,\lambda}$ = mean radiance of $I_{i,\lambda}$

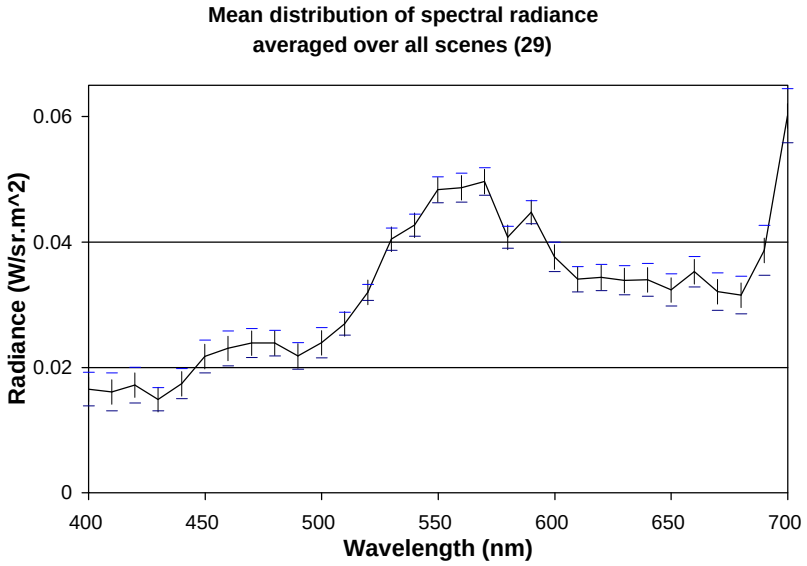

Figure 3.3: Distribution of mean spectral radiance for all scenes.

Figure 3.3 shows the distribution of mean spectral radiance for all 29 scenes and their standard error. Every point on the plot (one for each spectral filter λ) corresponds to:

$$R_\lambda = \frac{1}{29}\sum_{i=1}^{29} R_{i,\lambda}$$

Figure 3.4 shows the mean spectral radiance for selected scenes: *long-distance*, *short-distance* and *artificially-illuminated* sets (as described in Chapter 2 section 2.7) separately. Notice the peaks in the 700 nm filtered image (far red) for all the different sets. This will be discussed later.

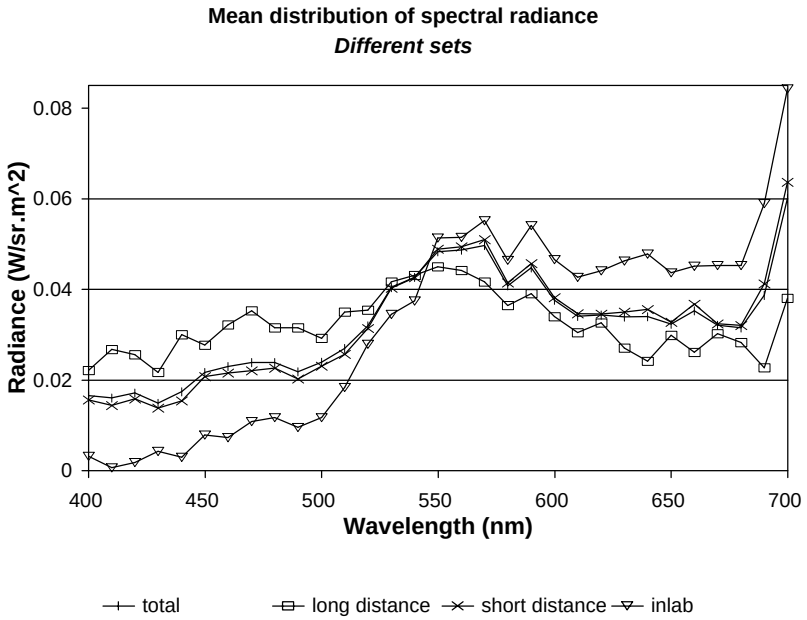

Figure 3.4: Distribution of mean spectral radiance for selected scenes: *long viewing distance*, *short distance*, *artificially illuminated* and *total* (all).

A similar analysis was performed using the derived mean spectral *reflectance* of the scenes. This magnitude is designed to be independent of the illumination, and is shown in Figure 3.5.

Figure 3.6 (Kirk 1984) shows the reported absorption spectrum of the two most common types of chlorophyll (denoted *a* and *b*). In both cases the absorption has two maxima, one around 450 nm and a second one at 650 nm. These are also manifested on the mean spectral reflectance of our dataset (Figure 3.5) which shows a corresponding depression around 450 nm and another near 650 nm.

As seen earlier (Figure 3.1) the h.v.s. is optimised to gather information in the regions of the electromagnetic spectrum where the solar energy is maximised.

If there is an advantage of being more sensitive to the spectral range in which solar

energy is maximised, it seems sensible for this to apply within the visual spectrum. Here we ask: is the spectral sensitivity of the human visual system to some further extent optimised to gather information from the visual environment?

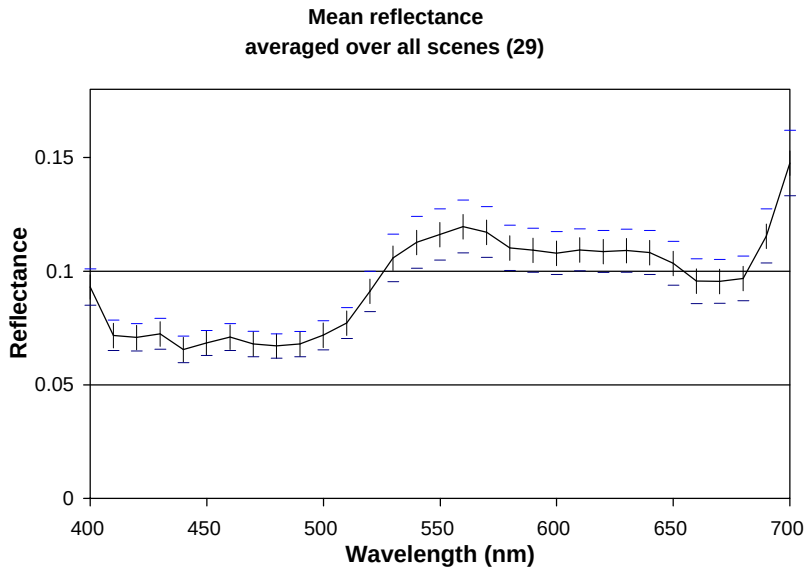

Figure 3.5: Distribution of mean spectral reflectance for all scenes.

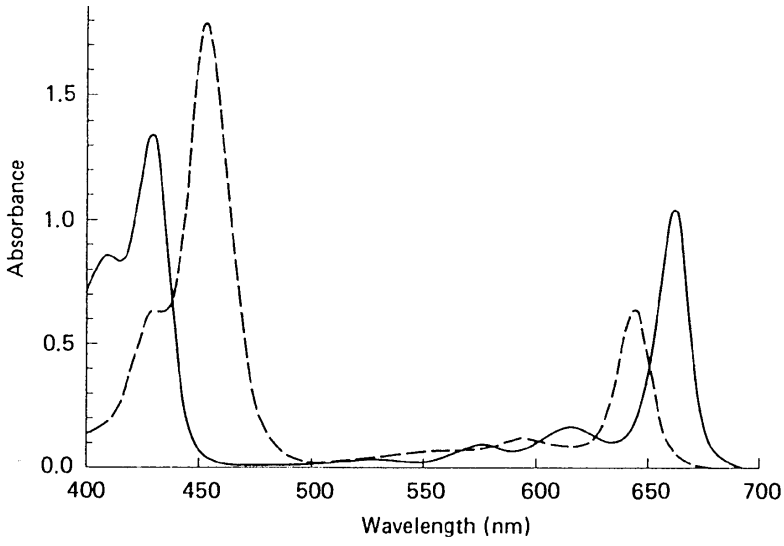

Figure 3.6: Reported absorption spectrum of chlorophyll *a* and *b* (Kirk 1984).

Figure 3.7 (a) shows a comparison between the spectral sensitivity curves for the *achromatic (L+M)* and *chromatic (L-M)* pathways and the mean spectral radiance averaged across all the dataset. The spectral sensitivity curves were produced from the Smith and Pokorny (1975) cone sensitivity curves using Ingling and Tsou (1988)

definition of chrominance and luminance. In this definition (see Chapter 1) the chrominance curve crosses zero at the neutral yellow reference point (580 nm). Notice the coincidence between the achromatic (L+M) and radiance curves' peaks at 550 nm. It is also noticeable that the highest value of radiance is in the far-red (and considering the absorption spectrum of the chlorophyll it is possible in the infra-red) and it does not correspond to the spectral sensitivity curves. This lack of sensitivity to the highest value of spectral radiance in natural scenes may be explained in terms of the "hardware constraints" discussed in the Introduction, like the increase of photoreceptors' noise with temperature (Aho *et al*. 1988; Barlow 1988).

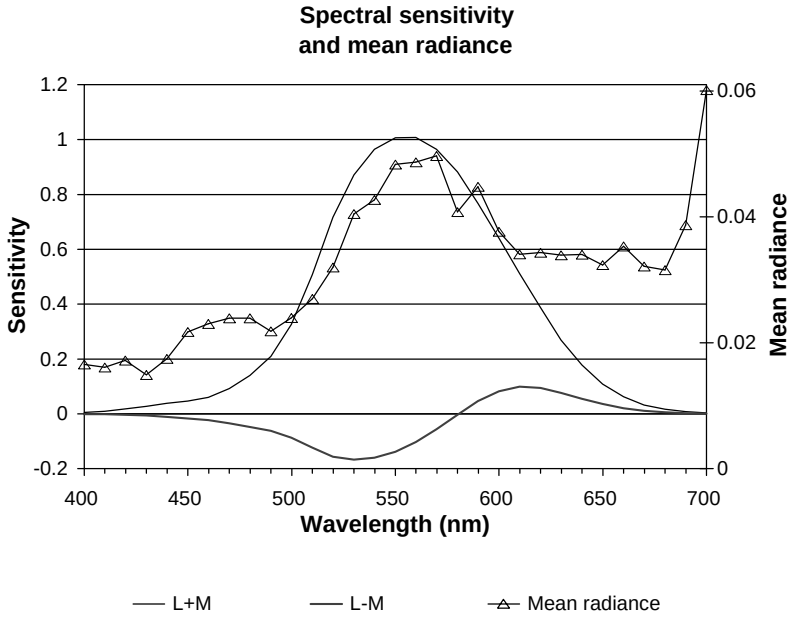

Figure 3.7 (a): Comparison between the spectral sensitivity curves and the mean distribution of spectral radiance of the dataset. The curves were obtained from Ingling and Tsou's (1988) definition of luminance and chrominance (chrominance crosses zero at the neutral yellow reference wavelength -about 580 nm).

Figure 3.7 (b) shows a similar comparison between the spectral sensitivity curves and the mean spectral reflectance averaged across all scenes. The achromatic and reflectance peaks are again coincident except for the dark red peak of the spectral reflectance. Comparison of derived reflectance is designed to be independent of any kind of light used to illuminate the scenes.

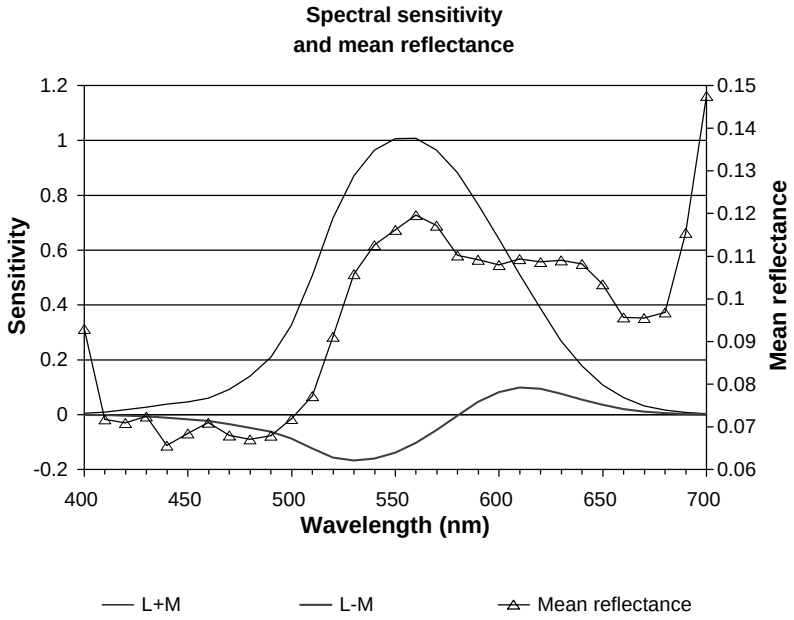

Figure 3.7 (b): Similar comparison between spectral sensitivity curves and the mean distribution of spectral reflectance of the dataset.

The positive and negative peaks of the chromatic (L-M) curve in both Figures 3.7 (a) and 3.7 (b) represent wavelengths where the output of the parvocellular L-M channel (see Chapter 1) has largest magnitude. Noticeable is the correspondence between those peaks and the shape of the mean radiance and reflectance curves. This may possibly represent the selective value of having an optimum coincident with the maximum environmental radiance or reflectance values.

Although the above ideas may seem reasonable, there is a risk of a misleading conclusion:

a) The spectral radiance and reflectance environmental average are clearly not the only factors determining our spectral sensitivity.

b) The shape of the L-M curve in Figures 3.7 (a) and 3.7 (b) could have its biological significance in optimising the discrimination of colours just different from that of green leaves (Neumeyer 1991). Best discrimination abilities are represented by the maximum slope of the L-M curve. In the analysis above no "discrimination task" was considered (such as for example to find a red cherry on a green leafy background).

3.2 Mean spectral distribution of amplitude spectra (Fourier content) of the radiance and reflectance scenes - Analysis of the modulation of the signal

In Chapter 1 we discussed how the amplitude spectra of natural images (averaged across all orientations) was reported to be proportional to f^{α} , where *f* is the spatial frequency and slope α assumes mean values of about -1.0. This was reported by Field (1987) for luminance (in fact achromatic) images. Tolhurst *et al.* (1992) found a mean value of
α= -1.2 for a different and larger set of achromatic images. If we assume that this kind of spatial structure of achromatic images is consistent with the coding scheme of the cortex our question is: is this spatial structure a peculiarity of the visual environment independent of colour? And if not, has the human visual system evolved to cope with a peculiarity of the visual environment restricted to certain regions of the visual spectrum? And finally: does the value of α depend on the region of the visible spectrum considered?

Here we analyse the consistency of the f^{α} relationship across the visible spectrum. "FBANDS" was applied to each filtered image of every scene of the dataset.

To make it easier to compare the results across the diversity of our scenes we normalised the results in two different ways:

a) **Image-based normalisation (*e'*),** in which the Fourier amplitude in each *image* of every scene is given equal weight when the results are averaged.

b) **Scene-based normalisation (*e"*),** in which the Fourier amplitude in each *scene* of the entire dataset is given equal weight when the results are averaged.

A formal description of these normalisations referred to as *e'*, *e"* respectively, is given in the next section. In the section after that, the results are described.

3.2.1 Mathematical description of the normalisation notation

In this section we use the same mathematical notation as in the previous one. To maintain the consistency, it is necessary to add two new definitions:

a) **SF band:** each of the spatial frequency bands in which the program "FBANDS" divides the Fourier space as described in Chapter 2. To indicate them we used the index

b: *b*= 1, 2, 3, ... , 8; where 1 corresponds the lowest SF band and 8 to the highest SF band in the Fourier space. See Tables 2.1 and 2.2 for the corresponding values in cycles/deg.

b) **Number of pixels** in each SF band: this can vary depending on the alternative version of "FBANDS" used (*logarithmic* or *linear*) and the band "*b*" considered. The index used is N_b. For the *linear* version: N_b= (888, 2516, 4116, 5716, 7352, 8924, 10564, 12110) and for the *logarithmic* version: N_b= (8, 16, 44, 180, 640, 2516, 9832, 38950). See Figure 2.10 and Tables 2.1 and 2.2 on Chapter 2 for more information.

Using the above notation, the total Fourier amplitude spectra in band *b* of image λ in scene *i* is denoted: $E_{b\,\lambda\,i}$

Absolute values of Fourier amplitude spectrum vary notably from one scene to other and even from one image to another in the same scene. The two different normalisations used to compare Fourier amplitude across wavelength are:

Image-based normalisation: the value of $E_{b\lambda,i}$ is normalised so that the total Fourier amplitude spectra corresponding to each *image* is the same (unit).

$$E'_{b\lambda i} = \frac{E_{b\lambda i}}{\sum_{b=1}^{8} E_{b\lambda i}}$$

Scene-based normalisation: similarly, in this second normalisation the total Fourier amplitude spectra corresponding to each *scene* is the same (unit).

$$E''_{b\lambda i} = \frac{E_{b\lambda i}}{\sum_{\lambda=400}^{700}\sum_{b=1}^{8} E_{b\lambda i}}$$

To make the resulting amplitude spectrum per band easier to compare across bands and to relate with previous findings for achromatic images (Field 1987, Tolhurst *et al.* 1992) the *average amplitude* within each band is compiled. This is equivalent to the amplitude spectra averaged across orientations as mentioned in the previous studies. To do this we divide the normalised amplitude spectra of each band by the number of pixels within the band.

$$e'_{b\lambda i} = \frac{E'_{b\lambda i}}{N_b}; \qquad e''_{b\lambda i} = \frac{E''_{b\lambda i}}{N_b}$$

These definitions of normalised amplitude spectra (*e'* and *e"*) were applied to the radiance and reflectance scenes and averaged over the dataset, i.e.

$$\overline{e'}_{b\lambda} = \frac{1}{N}\sum_i e'_{b\lambda i}$$

$$\overline{e''}_{b\lambda} = \frac{1}{N}\sum_i e''_{b\lambda i}$$

where N = total number of scenes (29). $\overline{e'}$ and $\overline{e''}$ are the magnitudes we use to show average spatial structure across the dataset.

3.2.2 *Average distribution of $\overline{e'}$ and $\overline{e''}$. Radiance scenes - Results*

Figures 3.8 (a) and 3.9 (a) show for radiance, the distribution of averaged Fourier amplitude over wavelength for each of the eight *SF bands*. In Figure 3.8 (a) the average is computed according to *image-based normalisation* so the quantity plotted is $\overline{e'}$ (as defined in the section above). In Figure 3.9 (a) the average is computed according to *scene-based normalisation* so the quantity plotted is $\overline{e''}$ (also defined above).

In order to obtain values evenly distributed along the spatial frequency axis (not bunched at the high frequency end) the *logarithmic* version of "FBANDS" is applied in all cases.

To show the dimension of the Fourier space, a logarithmic axis is added in Figures 3.8 (b) and 3.9 (b). The central spatial frequency in cycles/degree of each SF band is outlined to identify it in the spatial frequency axis.

In Figures 3.9, the values corresponding to the centre of the Fourier space (mean value of radiance of each filtered image) were included in the graph as **band 0**. These values are identical to those in Figure 3.3 but are here plotted logarithmically.

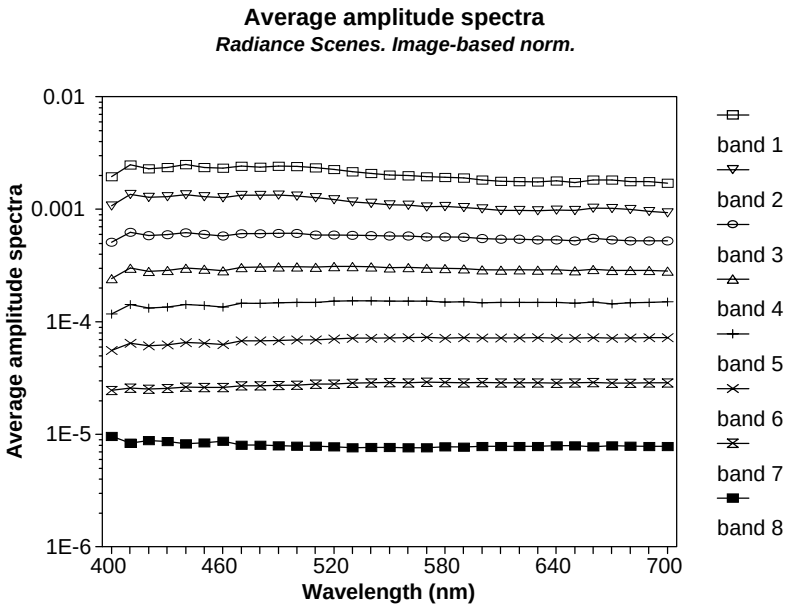

Figure 3.8 (a): Variation of the average amplitude spectra ($\overline{e'}$) across the visible spectrum for all SF bands. Image-based normalisation and radiance scenes.

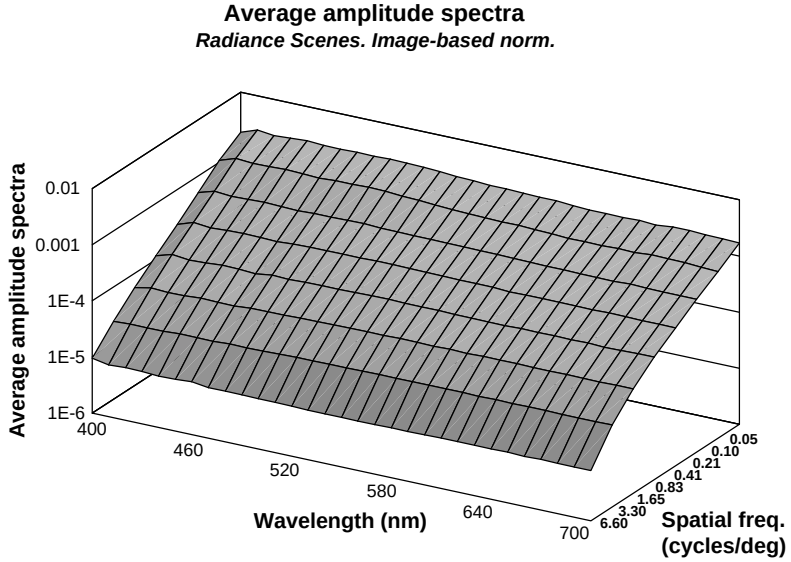

Figure 3.8 (b): 3-dimensional plot of the Figure above. One of the axes displays the central value of spatial frequency (in cycles/deg) corresponding to the SF bands.

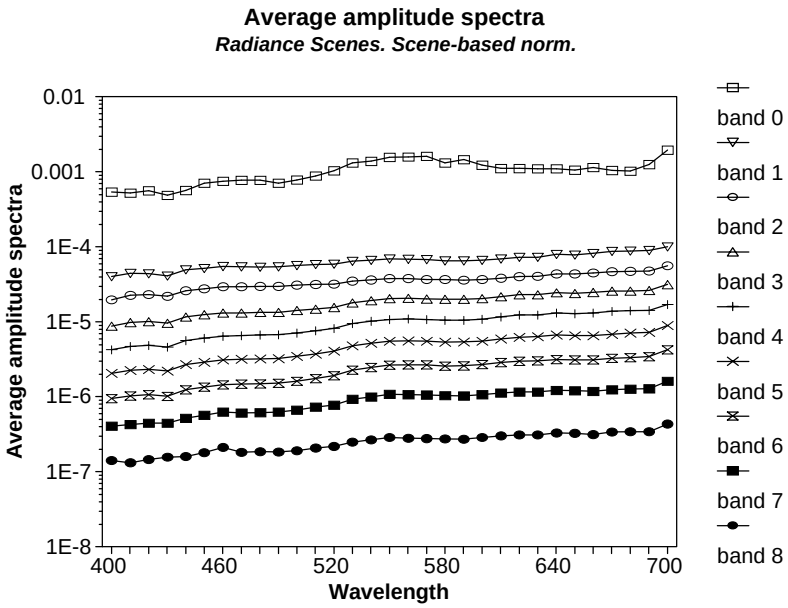

Figure 3.9 (a): Variation of the average amplitude spectra ($\overline{e''}$) across the visible spectrum for each SF band. Scene-based normalisation and radiance scenes.

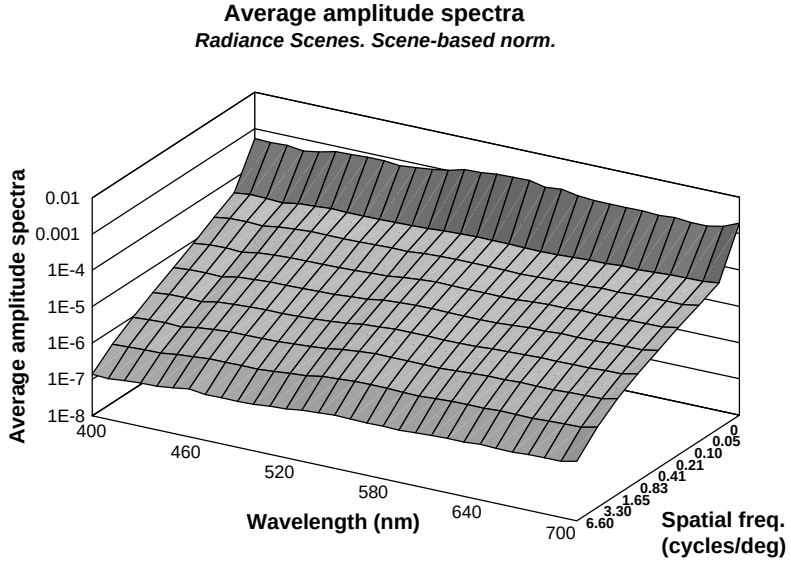

Figure 3.9 (b): 3-dimensional plot of the previous Figure. One of the axes displays the central value of spatial frequency (in cycles/deg) corresponding to the SF bands

From Figures 3.8 and 3.9 notice the following features:

a) The averaged Fourier amplitude always decreases as the spatial frequency increases. This is independent of the wavelength considered (the horizontal lines in Figures 3.8 (a) and 3.9 (a) never cross each other).

b) This decrease can be described by the relation $\overline{e'} = k.f^{\alpha}$, where α takes values between -1.0 and -1.2 (mean α= -1.10, std= 0.03). The 3-dimensional plot in Figures 3.8 (b) and 3.9 (b) show a surface composed of roughly straight lines lying along the spatial frequency axis.

c) The average amplitude spectra corresponding to the highest SF band (band 8) is slightly smaller than expected from the relation $\overline{e'} = k.f^{\alpha}$ (the plotted surface drops).

Scene-based normalisation (Figures 3.9) constrains fewer degrees of freedom, so the plot is less flat than in Figures 3.8.

3.2.3 *Average distribution of $\overline{e'}$ and $\overline{e''}$. Reflectance scenes.*

All the previous measurements were repeated for the reflectance scenes. Figures 3.10 and 3.11 show the variation of average amplitude spectra in *image-based normalisation (e')* and *scene-based normalisation (e")* respectively for them. The *logarithmic* version of "FBANDS" was again used. The results are similar to those described before (approximately constant distance between the curves and no crossings) for both normalisations. The value of slope α is again calculated and the mean is α= -1.10, std= 0.04.

The most significant difference between average Fourier spectrum for radiance and reflectance scenes is denoted in Figures 3.9 and 3.11 (scene-based normalisations). The latter one varies less across wavelength than the former one. This is because *radiance scenes* include scenes illuminated using tungsten lighting whose spectrum is described in Figure 3.2. The influence of tungsten is to raise the values of average Fourier amplitude when the wavelength increases. The reflectance scenes are designed to be independent of the kind of illumination used.

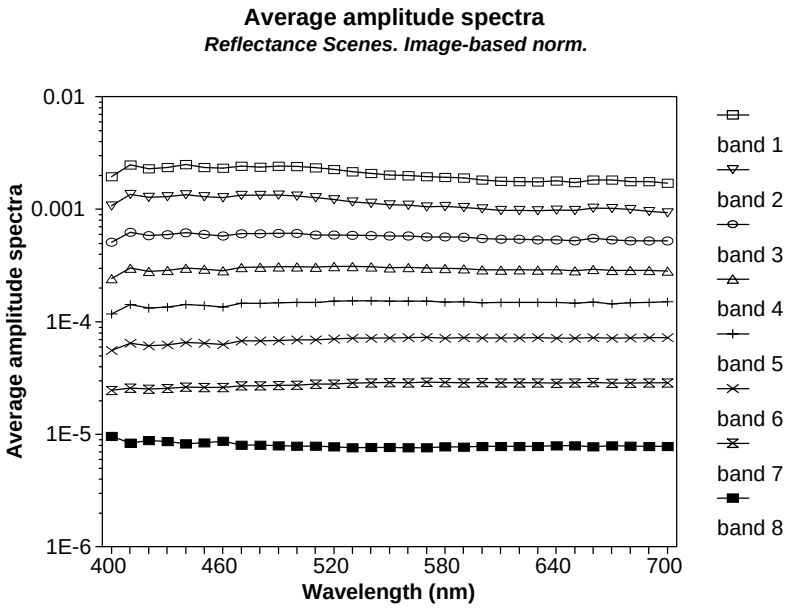

Figure 3.10 (a): Variation of the average amplitude spectra ($\overline{e'}$) across the visible spectrum for all SF bands. Image-based normalisation and reflectance scenes.

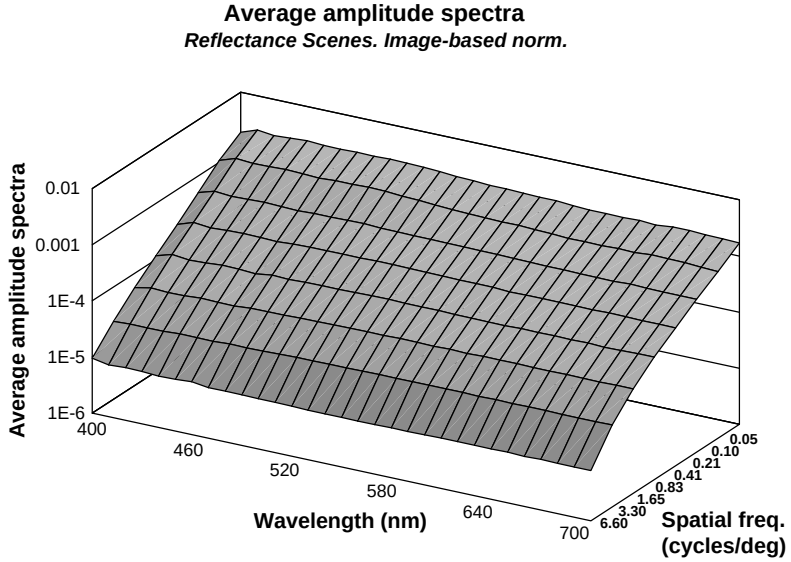

Figure 3.10 (b): 3-dimensional plot of the Figure above. One of the axes displays the central value of spatial frequency (in cycles/deg) corresponding to the SF bands.

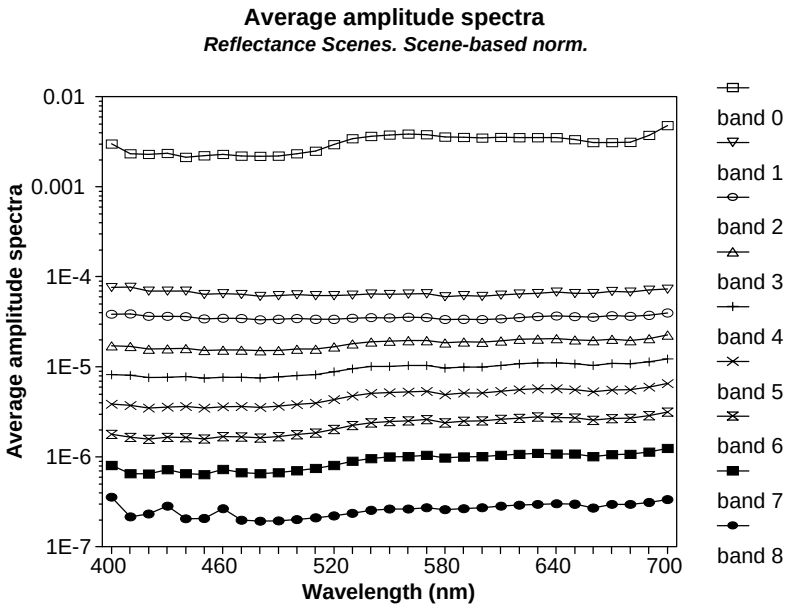

Figure 3.11 (a): Variation of the average amplitude spectra ($\overline{e''}$) across the visible spectrum for each SF band. Scene-based normalisation and radiance scenes.

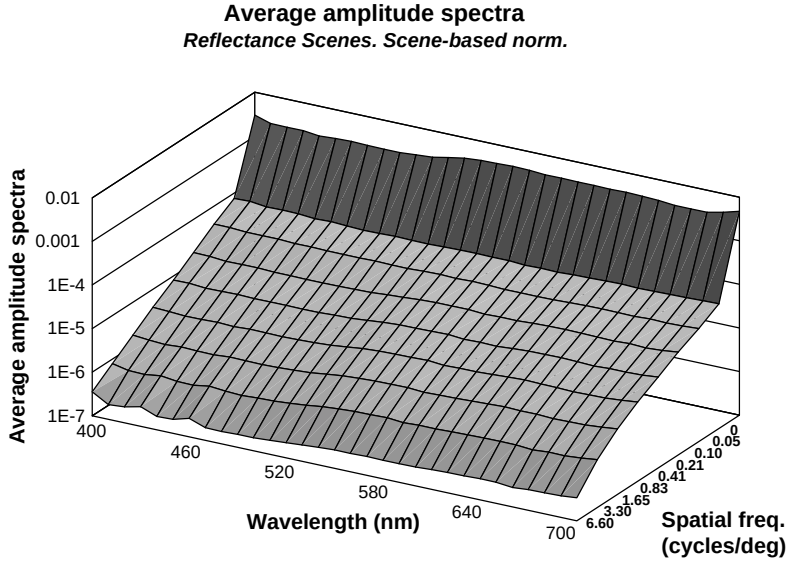

Figure 3.11 (b): 3-dimensional plot of the Figure above. One of the axes displays the central value of spatial frequency (in cycles/deg) corresponding to the SF bands.

3.2.4 Quantification of the slope (α) across wavelength

For radiance and reflectance scenes, the slope (α) varies with wavelength. Figure 3.12 shows how the mean slope varies for the curves displayed in Figure 3.9 (the values of band 0 were not included). Notice how the slope α is closer to the value α= -1 for

wavelengths between 580nm and 640 nm. Standard error of the mean α was included in the plot. A similar shape was obtained for *radiance* and *reflectance scenes* using both *scene and image-based normalisation.*

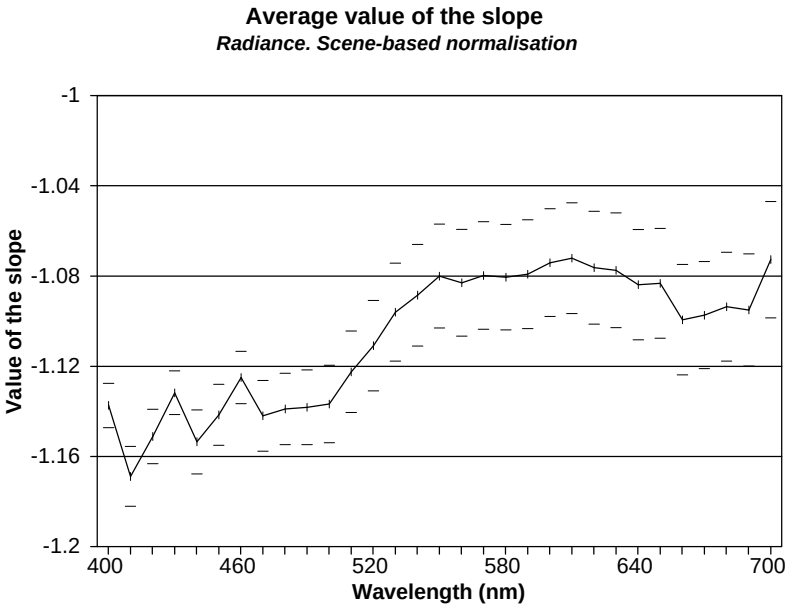

Figure 3.12: Mean value of the slope (α) across the wavelength. *Radiance scenes* and *scene-based normalisation* were employed. The standard error is also shown in the plot.

From Figure 3.12 notice the following:

a) The average slope across the visual spectrum is α= -1.10, std= 0.03.

b) Slopes tend to be less steep than the average for values of wavelength greater than 530 nm. For wavelengths smaller than 530 nm, the values of α are steeper than the average.

c) The values of α closest to -1 are between 570-640 nm.

The trend of the plot in Figure 3.12 is strongly influenced by the presence of sky in four of the scenes included in the average. These scenes (labelled as *long-distance* scenes in Chapter 2) show steeper slopes (α is close to -1.3) in the range 400 to 520 nm and values of α close to -1.1 in the range 640 to 700 nm. Although originated in the presence of sky, this lack of Fourier energy in the high frequency range of the blue side of the spectrum might provide an explanation for the lower number of blue cones in the retina, their wider spacing and their absence in the fovea.

The former sections provide a rough description of the spatial frequency structure of our dataset across the visible spectrum. Some of the characteristics of the structure found for filtered images within the range 570-640 nm are similar to those reported for achromatic images (Field 1987). However, from Figure 3.12 it is clear that Field's assumption of absolute scale invariance is only a first approximation -an over-simplification. In Chapter 5 we will explore the consequences of these findings.

3.3 *Variation of $\overline{e}'$ and $\overline{e}''$ with spectral bandwidth of the receptor*

Results given in the previous section suggested that the statistics across the visible spectrum even if not completely scale-invariant, are close to it, with the amplitude spectra in general decreasing monotonically with the spatial frequency. The data explored above is for chromatic bandwidth of about 10 nm. The human visual system gathers information through three receptors (L, M, S cones) which have bandwidths of the order of 100 nm. Some of the specifications of the shape and disposition of these L, M, S fundamentals can be understood in terms of the constraints imposed by the environment. For example, receptors must be spectrally wide and must overlap in order to best discriminate spectral colours. Our question at this point is: could the scale-invariance structure of the environment play some role in these specifications, such as for example limiting the spectral bandwidth of the receptors to provide optimal information coding? Are the average amplitude spectra of natural scenes still proportional to f^{α} when the spectral bandwidth of the receptors increases? And is this true across the whole visible spectrum? The L, M and S cone fundamentals are not equally spaced along the visible spectrum. However, for a general analysis, we process the data from our scenes in order to produce an output similar to three hypothetical receptors equally spaced along the visible spectrum. Following this, we explore how the Fourier amplitude spectra of the images produced by these receptors vary depending on the bandwidth of the receptors. The range explored goes from a very narrow band (10 nm) to the order of an L,M,S receptor bandwidth.

Here we use the same notation as in the former sections. More definition is required:

Receptor: combination of a given group of images belonging to the same scene, is denoted:

$$R_{c,b} = \frac{\Delta\lambda}{b} \cdot \sum_{\lambda=c-(b-\Delta\lambda)/2}^{c+(b-\Delta\lambda)/2} I_{\lambda} ,$$

where $\Delta\lambda$ = *10 nm*, ***c*** is the *centre* of the receptor and ***b*** is the receptor *bandwidth*.

For example, the combination of the images I_{640} ,I_{650} and I_{660} is represented by the receptor centred in *c=650* with a bandwidth *b = 30 nm* and denoted:

$$R_{650,30} = \frac{1}{3} \sum_{\lambda=640}^{660} I_{\lambda}$$

Given that we are only interested in variations of the Fourier structure with bandwidth of the receptor, the region of the visible spectrum in which this is placed is not of primary interest. In the previous sections all the images had 10 nm bandwidth and are evenly distributed along the visible spectrum. Here the centre *c* of such images (receptors) is selected to have the following equidistant values:

c = 450, 550 and 650 nm

and the following spectral bandwidths (*b*):

b = 10, 30, 70 and 110 nm.

To reduce any influence of the illumination, we used *reflectance* scenes.

The *logarithmic* version of "FBANDS" was applied to these receptors. The results were averaged as described in the previous section to obtain $\overline{e'}$ (*image-based normalisation*) and $\overline{e''}$ (*scene-based normalisation*). Figure 3.13 shows a plot of the average amplitude spectra using *image-based normalisation* ($\overline{e'}$) of the $R_{450,30}$, $R_{550,30.}$, $R_{650,30}$, receptors in each of the SF bands. The features observed in the spectrally narrowband images on Figures 3.8 to 3.11 are present again, when plotted on double-logarithmic co-ordinates. $\overline{e'}$ decreases approximately linearly as the spatial frequency increases. The slope (α) is again consistent with the previous findings (α=-1.12, α= -1.10 and α= -1.07 for $R_{450,30}$, $R_{550,30.}$ and $R_{650,30}$ respectively). The same characteristics were found for *scene-based normalisation* ($\overline{e''}$).

The plot in Figure 3.13 shows that $\overline{e'}$ can be described using $k.f^{\alpha}$ for all the receptors examined. A more precise account of small differences in the spatial frequency structure

when the spectral bandwidth of the image increases can be obtained from Figure 3.14. The curves represent different receptors centred in *c= 650* with spectral bandwidth *b =10 nm, 30 nm, 70 nm and 110 nm.* Figures 3.15 and 3.16 correspond to a similar analysis of the receptors centred in *c= 550 nm* and *440 nm* respectively.

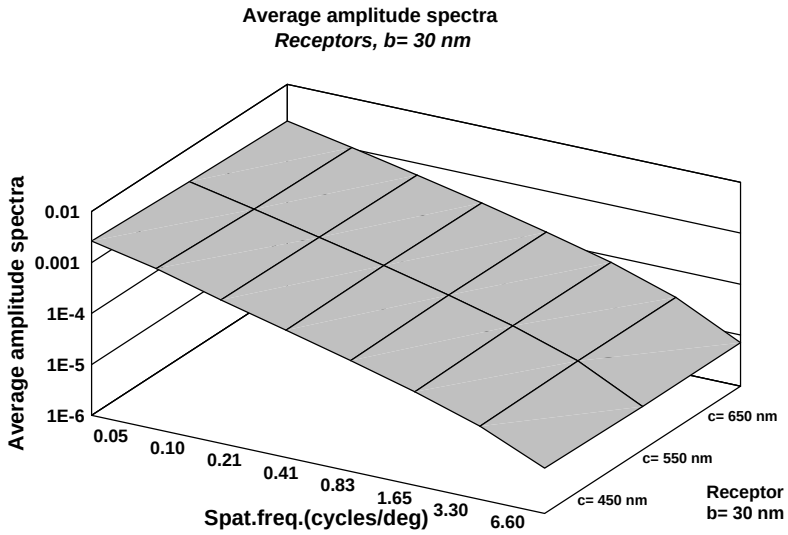

Figure 3.13: Variation of $\overline{e'}$ for the blue, green and red receptors. Bandwidth of all receptors b= 30 nm. The centres were c= 450, 550, and 650 nm for the "blue", "green" and "red" receptors respectively.

Table 3.1 shows how the values of the slope (α) vary according to the bandwidth of the receptor considered (with α taking values between -1.07 and -1.15). The standard deviation corresponding to the average over the dataset was also included.

Value of the slope α	**c = 650 nm**	**c = 550 nm**	**c = 450 nm**
b = 10 nm	-1.06 ± 0.06	-1.10 ± 0.06	-1.14 ± 0.04
b = 30 nm	-1.07 ± 0.12	-1.10 ± 0.13	-1.13 ± 0.17
b = 70 nm	-1.08 ± 0.13	-1.11 ± 0.13	-1.14 ± 0.17
b = 110 nm	-1.09 ± 0.12	-1.12 ± 0.14	-1.15 ± 0.17

Table 3.1: Value of the slope (α) and standard deviation for receptors centred in *c= 650, 550* and *450* nm and bandwidth *b= 10, 30, 70* and *110* nm

In general, Table 3.1 shows a slight tendency for the average value of α to decrease when the bandwidth increases. Again the nearest to unit slopes (α) correspond to the c= 650 nm and c= 550 nm receptors.

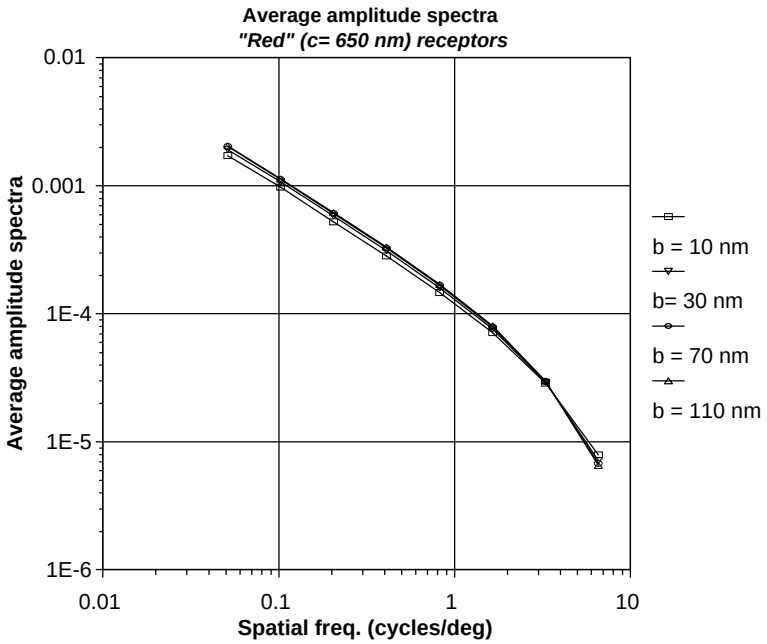

Figure 3.14: Variation of the average amplitude spectra ($\overline{e'}$) with the SF bands. Each plot represents a different spectral bandwidth (b) of the "red" receptors (b= 10, 30, 70 and 110 nm).

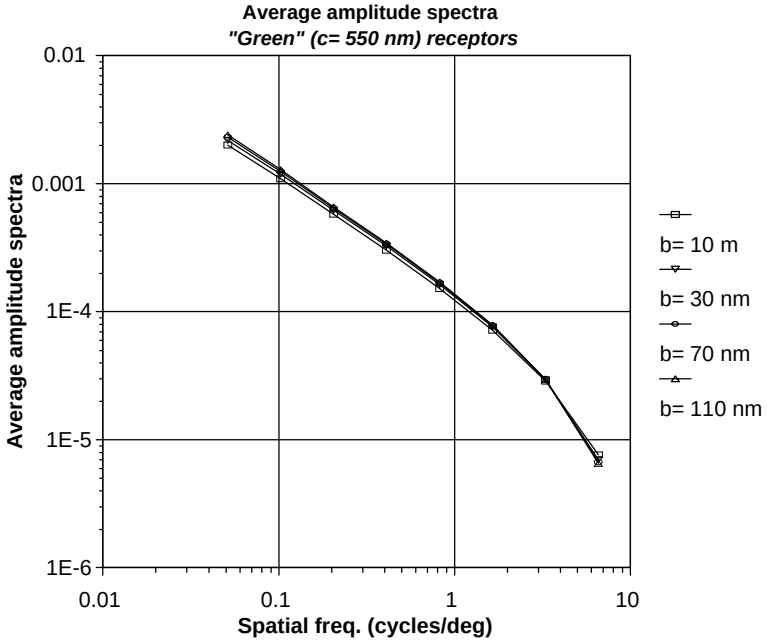

Figure 3.15: Similar plot as the above but considering only "green" (c = 550 nm) receptors.

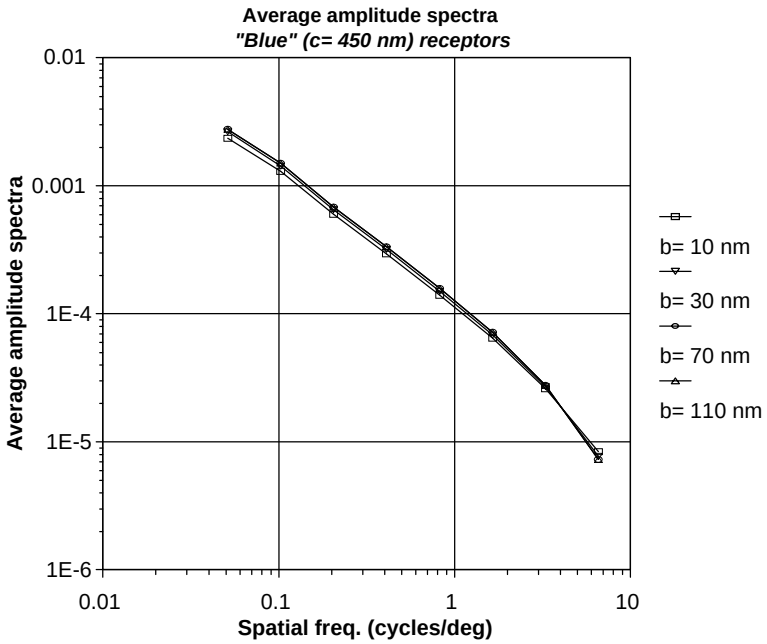

Figure 3.16: Variation of the average amplitude spectra ($\overline{e'}$) with the SF bands. Each plot represents a different spectral bandwidth (b) of the "blue" receptors (b= 10, 30, 70 and 110 nm).

Apparently there is no evidence of remarkable changes in the structure of the Fourier amplitude spectrum with the increase of the bandwidth of the receptor considered. To see whether the changes of the value of α in Table 3.1 have any statistical significance, we tested the equality of the values in each column. Student's t-test (Dunn and Clarke 1987) was applied to the largest difference in each column using a 95% confidence interval and the results showed that there is no significant variation between the mean values of α.

The consequences of these findings for the human visual information coding will be discussed in Chapter 5.

3.4 *Variation of the average amplitude spectra ($\overline{e'}$) in terms of chrominance (L-M) with the separation between L and M in the wavelength axis*

In Chapter 1 we mentioned the concept of *chrominance* as a linear combination of the Smith and Pokorny L and M cone responses. The information extracted from the visual environment using this particular combination of spectral receptors depends on the spectral bandwidth of the receptors and on their location along the visual spectrum. Changes in the Fourier content of the environment produced by any combination of receptors could lead to inefficiency in the coding scheme of the visual cortex. Here we

explore to what extent the bandwidth and the location of the actual L and M cone fundamentals in the visible spectrum may have been determined by variations in the regularity of that structure.

We analyse the Fourier amplitude spectrum of our dataset using an analogous combination of the “red”(c= 650~ 570 nm) and “green”(c= 550 nm) receptors. The central wavelength of the “green” receptor is fixed at c= 550 nm whilst that of the “red” one varies, taking values of c= 650, 630, 600 and 570 nm. The bandwidth of both receptors was also modified, having values of b= 10, 30, 70 and 110 nm.

Precisely, the combinations were:

$R_{550,10} - R_{650,10}$	$R_{550,30} - R_{650,30}$	$R_{550,70} - R_{650,70}$	$R_{550,110} - R_{650,110}$
$R_{550,10} - R_{630,10}$	$R_{550,30} - R_{630,30}$	$R_{550,70} - R_{630,70}$	$R_{550,110} - R_{630,110}$
$R_{550,10} - R_{600,10}$	$R_{550,30} - R_{600,30}$	$R_{550,70} - R_{600,70}$	$R_{550,110} - R_{600,110}$
$R_{550,10} - R_{570,10}$	$R_{550,30} - R_{570,30}$	$R_{550,70} - R_{570,70}$	$R_{550,110} - R_{570,110}$

All methods employed in this section are analogous to those used in the previous section.

The value of the average amplitude spectra ($\overline{e'}$) for *image-based normalisation* is plotted in Figures 3.17 to 3.20 for all *"red-green"* images above. Figure 3.17 shows how the slope of $\overline{e'}$ vs. spatial frequency varies when the centre (c) of the "red" receptor takes the values 650, 630, 600 and 570 nm. Receptors of b = 10 nm width were used. Figures 3.18, 3.19 and 3.20 show the same for receptors of b = 30, 70 and 110 nm width respectively. All these figures exhibit a similar structure of the Fourier amplitude spectrum to that of the previous sections. However the slope varies depending of the distance between the receptors considered.

All the plots seem again to be described by the function *amplitude* = $k.f^{\alpha}$ with the only exception being the value corresponding to the highest spatial frequency band examined. As in the previous section, all the values of the coefficient α and its standard deviation for all the curves plotted on Figures 3.17 to 3.20 are displayed on Table 3.2. The first row shows the values for the spectral bandwidth (b) of the receptors involved in the combination (10 nm, 30 nm, 70 nm and 110 nm) and the first column shows the

values for the centre (c) of the “red” receptor (570 nm, 600 nm, 630 nm and 650 nm).

Again we performed Student’s t-test to see whether the difference between the values in Table 3.2 are statistically significant. The test was performed between the largest difference of the value of α in each row and in each column on Table 3.2 using a 95% confidence interval. All differences proved to be statistically significant under these conditions. Notice that α is closer to the value $\alpha = -1$ when the distance between the "red" and "green" receptor increases and when the bandwidth of the receptors considered increases. This might have some implications for the h.v.s. since the optimum value ($\alpha = -1$) is achieved by spectrally separating the receptors and broadening their bandwidths. Given that L and M receptors are not, in fact, spectrally separated, we deduce that there might be a more compelling reason for doing this which is not considered here. The tendency shown in Table 3.2 might not be significantive enough to determine the evolution of the h.v.s. This is discussed later in Chapter 5.

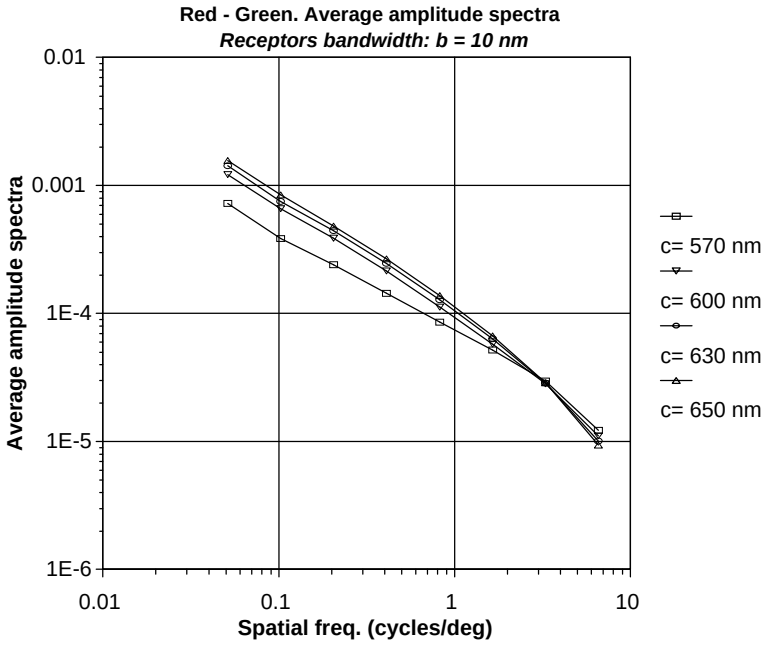

Figure 3.17: Variation of the average amplitude spectra ($\overline{e'}$) with the SF bands. Each plot represents a different centre (c) of the “red” receptor (c= 650, 630, 600 and 570 nm). Bandwidth of the receptors considered b= 10 nm.

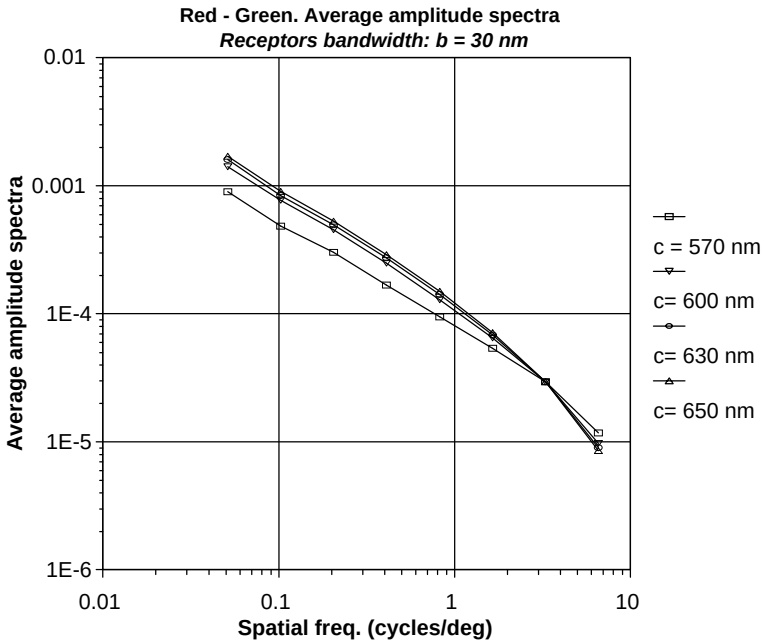

Figure 3.18: Variation of the average amplitude spectra ($\overline{e'}$) with the SF bands. Each plot represents a different centre of the "red" receptor (c= 650, 630, 600 and 570 nm). Bandwidth of the receptors considered b = 30 nm.

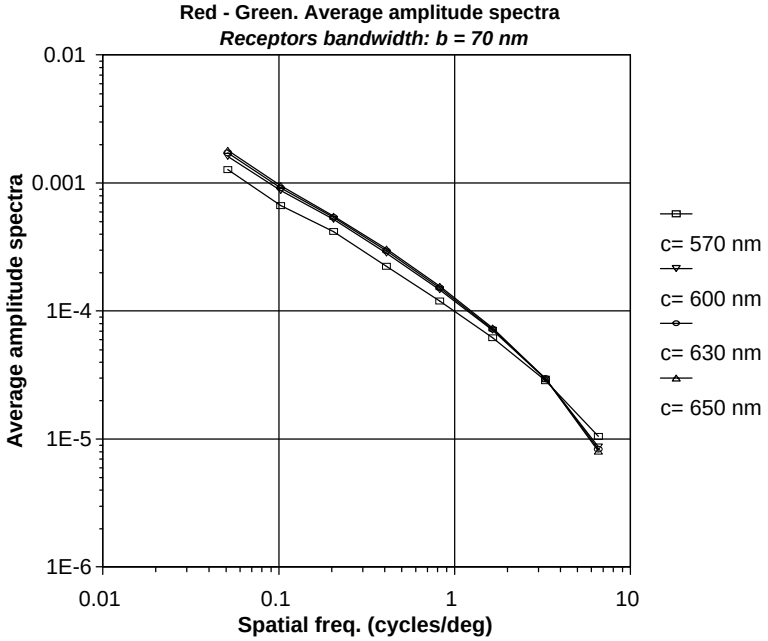

Figure 3.19: Similar plot as the above but considering receptors b= 70 nm.

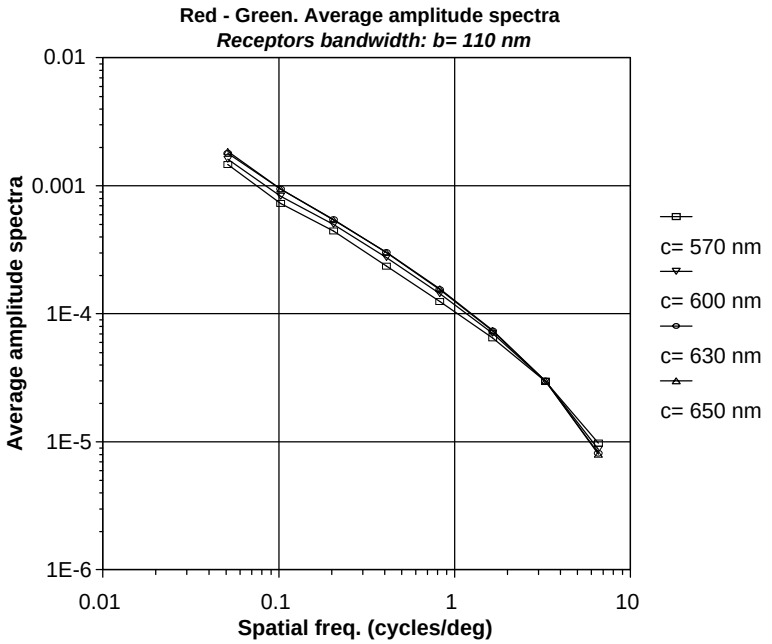

Figure 3.20: Variation of the average amplitude spectra ($\overline{e'}$) with the SF bands. Each plot represents a different centre (c) of the "red" receptor on the visible spectrum (c= 650, 630, 600 and 570 nm). Bandwidth of the receptors considered b = 110 nm.

Value of the slope α	**b: receptor's bandwidth**			
c: centre of the "red" receptor	**10 nm**	**30 nm**	**70 nm**	**110 nm**
650 nm	-0.98 ± 0.15	-1.01 ± 0.13	-1.02 ± 0.15	-1.02 ± 0.14
630 nm	-0.95 ± 0.14	-0.98 ± 0.149	-1.01 ± 0.14	-1.02 ± 0.15
600 nm	-0.90 ± 0.14	-0.95 ± 0.14	-0.99 ± 0.14	-0.98 ± 0.15
570 nm	-0.76 ± 0.12	-0.83 ± 0.12	-0.91 ± 0.15	-0.94 ± 0.14

Table 3.2: Slope (α) and standard deviation for combinations of "red" and "green" receptors of different bandwidth (b) and different centres (c).

Chapter 4

Applying existing analysis to the new dataset

Chapter 1 outlined some conclusions about the statistics of natural scenes based on a relatively small sample of the natural visual environment. In some cases (Burton and Moorhead 1987; Brelstaff and Troscianko 1992) the gamut of colours was very limited and in others (Derrico and Buchsbaum 1991; Field 1987) the number of scenes was arguably small. Even in one of these cases, (Brelstaff and Troscianko 1992) the results were clearly inconclusive. The acquisition of a new dataset of 29 natural scenes generates a new opportunity for extracting information that may further test these analyses.

4.1 *Analysis of colour and luminance information in natural scenes*

The question arises as to which aspects of human vision we wish to compare with the statistics of our images. We have three receptor types (L, M and S) and our vision differs in the fovea and the periphery. However, S-cones are absent from the fovea, which is arguably the most important area of the retina. Furthermore, Mullen (1985) investigated foveal vision and the main aim of our work was to compare image statistics with her results. Therefore, we limit our analysis to the responses of the L and M cones.

In a spatial-frequency analysis of a set of 20 colour slide scenes, Brelstaff and Troscianko addressed the question: is the physiological imbalance between colour and luminance transfer functions reflected in the s.f. content of natural scenes? To this end *luminance* and *chrominance* images were defined as linear combinations of the Smith and Pokorny **L** and **M** cone response images (their analysis was restricted only to L and M cone responses):

$Lum = a_1 \mathbf{L} + a_2 \mathbf{M}$; $Chrom = a_3 \mathbf{L} - a_4 \mathbf{M}$

using the following three definitions:

(a) Simple definition:

$a_1 = a_2 = a_3 = a_4 = 1.$

Lum = **L** + **M**; *Chrom* = **L** - **M**

(b) Ingling and Tsou (1988): estimations of the ratio between the coefficients were derived using a flicker and acuity based criteria.

$a_1/a_2 = 1.02$; $a_3/a_4 = 0.41$.

Lum = 1.02 **L** + **M**; *Chrom* = 0.41 **L** - **M**

(c) Buchsbaum and Gottschalk (1983): coefficients were derived using an information theory based criterion.

$a_1 = 0.887$; $a_2 = 0.461$;

$a_3 = 0.46$; $a_4 = 0.88$.

Lum = 0.887 **L** + 0.461 **M**; *Chrom* = 0.46 **L** - 0.88 **M**

The same definitions are computed for our dataset of 29 multispectral scenes by the following five stages:

a) "DO_LMS" (see Chapter 2) is applied to each of the radiance scenes. Three images (L, M and S) are obtained for each of the scenes.

b) Two of the three output images (L and M) are linearly combined as described in (a), (b) and (c) to obtain the luminance *(lum)* and chrominance *(chrom)* images.

c) For every *lum* and *chrom* image, a *Fourier amplitude spectrum image* is derived using "FBANDS" (see Chapter 2). The centre (mean luminance level) of this image is removed. Both Fourier spectrum images are scaled so that each contains total amplitude equal to unit in the spatial frequency spectrum. The proportion of the total Fourier amplitude present in each band within the *"logarithmic"* set of concentric bands is measured.

d) The ratio of the *lum* image amplitude to that of the *chrom* image amplitude is plotted

for each spatial frequency band.

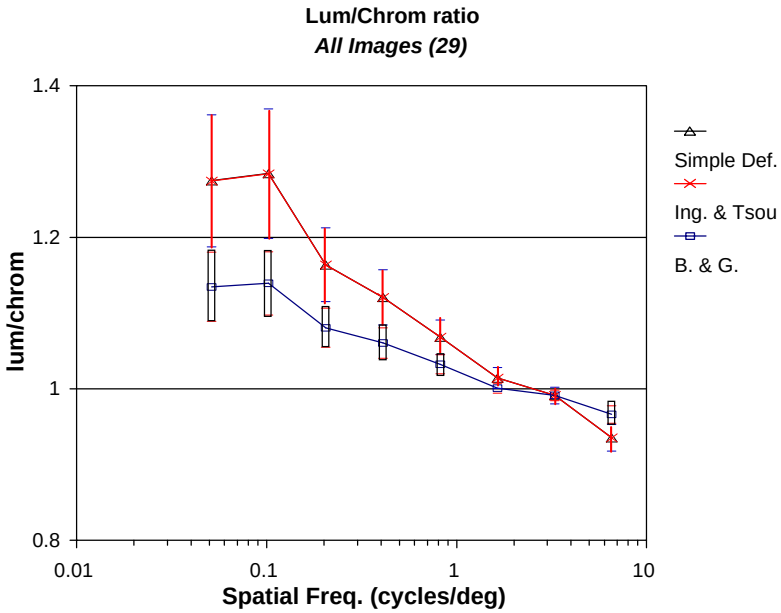

Figure 4.1: Average ratio of the *lum* image amplitude to that of the *chrom* image amplitude for all the dataset. The three definitions are used. Standard error is shown on the plot.

Figure 4.1 shows the average for all 29 scenes using the three definitions of *lum* and *chrom.* The numbers on the *x*-axis are converted from the Fourier amplitude spectrum image (cycles/pixel) to cycles/deg. The spatial frequency limit of 128 cycles/image width corresponds to the Nyquist frequency of the frame store sampling process (i.e. 0.5 cycles/pixel). Considering that the total image width corresponds to 14.58 deg of visual angle (data obtained from the DRA-camera lens), the highest spatial frequency that our system is capable of resolving is 8.78 cycles/deg. The standard error is also shown in all plots.

To check whether the results have something to do with the original quality of our scenes, we plot similar amplitude ratio *(lum/chrom)* corresponding to scenes without any correction *(selected set).* The *simple definition* (a) is used. Figure 4.2 shows that the trend of the plots is basically similar. The *selected* set is slightly shifted down in the range of low spatial frequency.

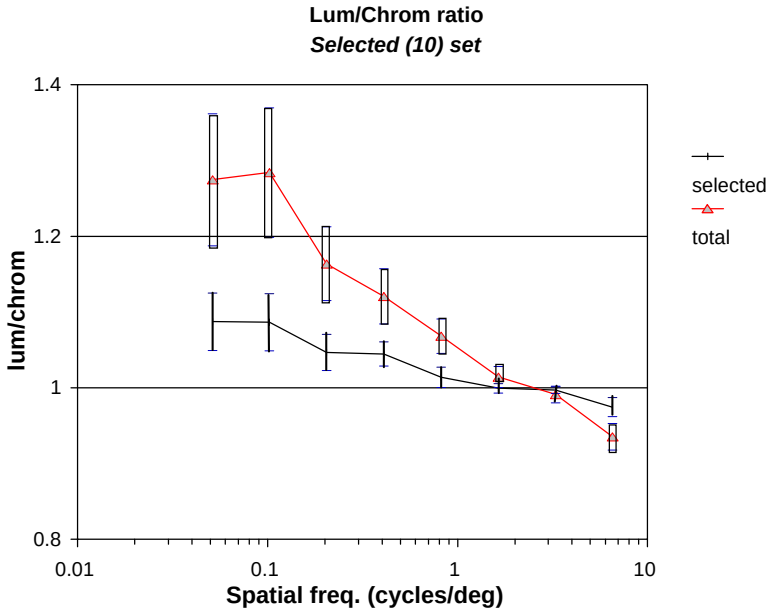

Figure 4.2: Average ratio of the *lum* image amplitude to that of the *chrom* image amplitude for scenes without any correction (*selected* set) and for the rest of the scenes (*total*). The simple definition of *luminance* and *chrominance* is used.

The values of the plot change considerably when the average for the *long distance* (see Chapter 2) set of scenes is plotted (see Figure 4.3). In this set most of the objects (trees) are in the range 0.1~4.0 Km and there is a considerable proportion (50% approx.) of grass and sky in the scene. Figure 4.3 shows that the proportion of Fourier amplitude in *lum* to that of *chrom* for low spatial frequencies is much higher in those scenes. The absence of any *long distance* scenes justifies why the *selected* set is less steep than the rest in Figure 4.2. This *selected* set of scenes only includes scenes taken in ideal conditions (mainly inside the lab) which is not the case for the *long distance* scenes.

Considering that the results obtained for the *long distance* set (4 scenes) are considerably different from the rest of the scenes, we decide to analyse both groups separately. The following analyses are performed using only the *short distance* set.

The average for the *outdoor* (sun-illuminated) and *indoor* (artificially-illuminated) image set is plotted on Figure 4.4. No noticeable difference was found between these sets.

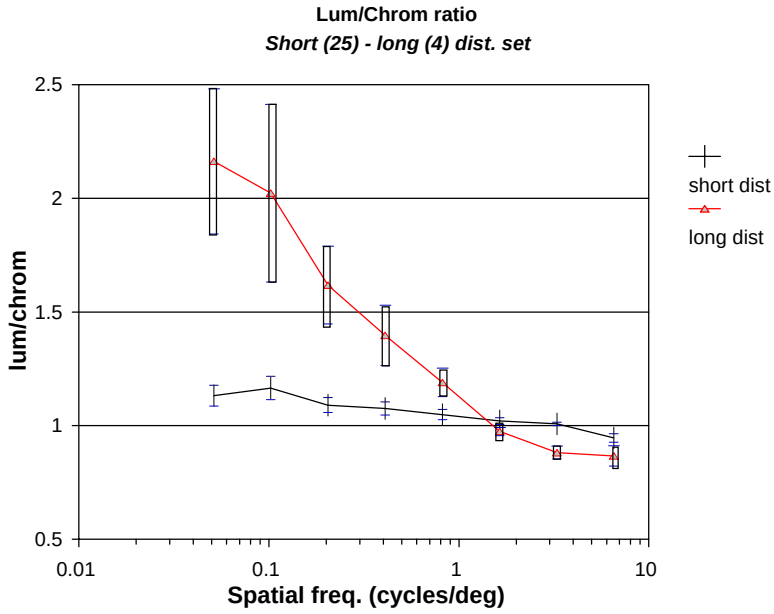

Figure 4.3: Average ratio of the *lum* image amplitude to that of the *chrom* image amplitude for the *long-distance* viewing set (objects in the range 10 m to 4 Km and including grass and sky) and the *short-distance* viewing set (objects up to 10 m far away from the camera).

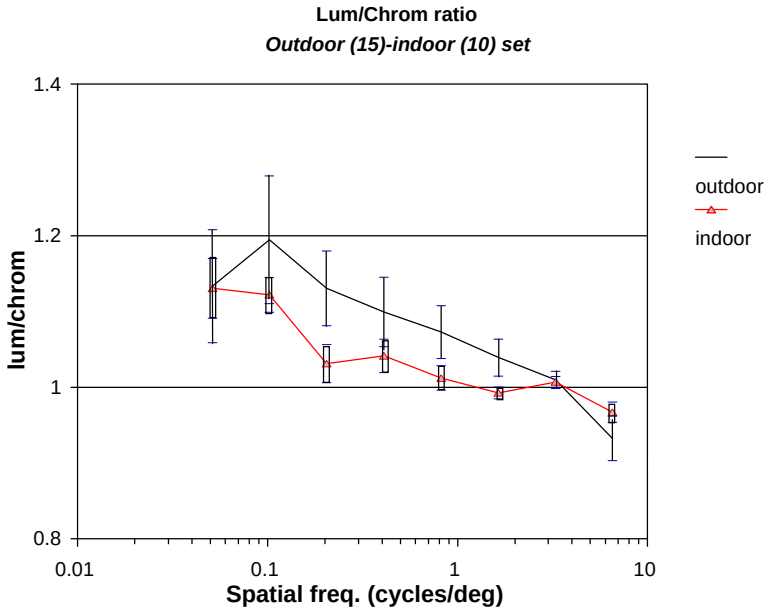

Figure 4.4: Average ratio of the *lum* image amplitude to that of the *chrom* image amplitude for the *outdoor* (sun-illuminated) and the *indoor* (artificially illuminated) sets of scenes.

In Figures 4.1, 4.2, 4.3 and 4.4 the trend is always the same: Fourier amplitude in *lum*

images is proportionally higher than in *chrom* images for low spatial frequencies (0.03 - 2 cycles/deg). This ratio changes for high spatial frequencies (2 - 8.7 cycles/deg).

In the case of the *long distance* set, this tendency is considerably increased. To clarify this view, all short distance scenes were again divided into different groups (outdoor, indoor and selected) and plotted together with the total average (all groups) in Figure 4.5. There does not seem to be any difference between them.

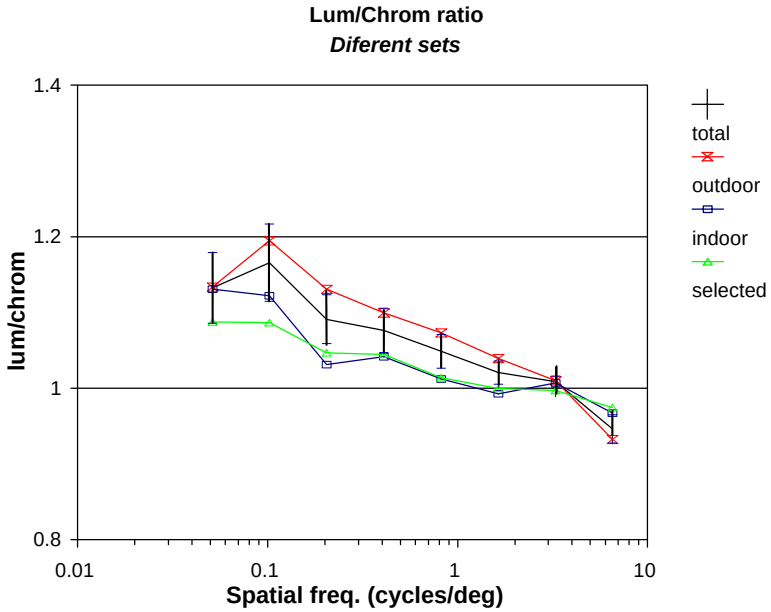

Figure 4.5: Average ratio of the *lum* image amplitude to that of the *chrom* image amplitude for short-distance scenes. These scenes are divided again into different groups (*outdoor*, *indoor* and *selected*) and plotted along with the average for all groups (*total*). The standard error is also shown in the plot for the *total* group

All our analyses (Figures 4.1 to 4.5) show an entirely different relationship from that expected if psychophysical measurements of contrast sensitivity functions (see Figure 1.16 in section 1.6 of the background Chapter) had an “ecological” explanation. The expectations from Mullen’s (1985) results especially disagree with the results for the *long distance* image set.

Our results clearly show that there is no rationale for the differences in human contrast sensitivity thresholds for *lum* and *chrom* in the statistics of our dataset. A more general discussion about these findings is found in Chapters 5 and 6.

4.2 *Analysis of colour and luminance information in natural scenes using a shadow-removing definition.*

Most of the time, natural scenes vary across their extent in both colour and luminance. Sometimes, they covary spatially, but most of the time their distributions are only partially correlated. For example, consider a uniformly red and bright object upon an uniformly green and dark background. In this case both colour and luminance share the same spatial distribution, but normally objects reflect light from a distant source (like the sun) and thus the illumination casts shadows (luminance discontinuities) in the world.

In general this shadowing only produces small changes in the colour of the light reflected by an object. For example, objects in the shade under sunlight illumination reflect higher proportion of light scattered by the atmosphere that is more skewed toward the short wavelengths (bluish light) than objects under direct sunlight. This change in colour is trivial compared with the massive change between the intensity of the light reflected by objects directly illuminated by the sun and by scattered light.

In a situation like the above, the spatial distribution of wavelength differences (chrominance) would give a more realistic representation of the real world than the spatial distribution of intensity differences (luminance).

Imagine a shadow over a large coloured surface. It contains high spatial frequencies Fourier energy in both luminance and chrominance. If we remove the shadow on the *chrom* image using a special definition of chrominance, the final *chrom* image will contain less energy in the high spatial frequency range. From this it follows that the world may be richer in high sf Fourier energy in luminance and in low sf Fourier energy in chrominance. This is what we expect from Mullen's measurements as mentioned in Chapter 1.

To account for the fact that none of the definitions of colour used in this work so far is independent of the effect of shadowing, here we include an extra definition of chrominance which removes the shadowing over our scenes. The *shadow-removing definitions* of *luminance* and *chrominance* are:

Lum = **L+M**; *Chrom* =**(L - M) / (L + M)**

The definitions above were tested on a special scene consisting of some vertical coloured cards (red, green, blue, grey and brown). The scene was illuminated by a projector which had a slide in its gate such that the top of the scene was clear and the bottom had a ND filter. The corresponding L and M images were obtained and combined using the *simple* definition and the above definition of *chrom*. Following this, the values of *chrom* between the directly illuminated (top) half of the image and the shadowed (bottom) half were measured in all coloured cards for both definitions:

Simple definition: *chrom = L-M*					
	brown	red	blue	green	grey
bright half	0.00065	0.00047	0.0037	0.00036	0.00064
shadowed half	0.00040	0.00031	0.00023	0.00024	0.00043
Shadow-removing definition: *chrom = (L-M)/(L+M)*					
	brown	red	blue	green	grey
bright half	0.329	0.429	0.260	0.302	0.317
shadowed half	0.330	0.429	0.256	0.304	0.309

These measurements show an effective removal of the shadow in the second image.

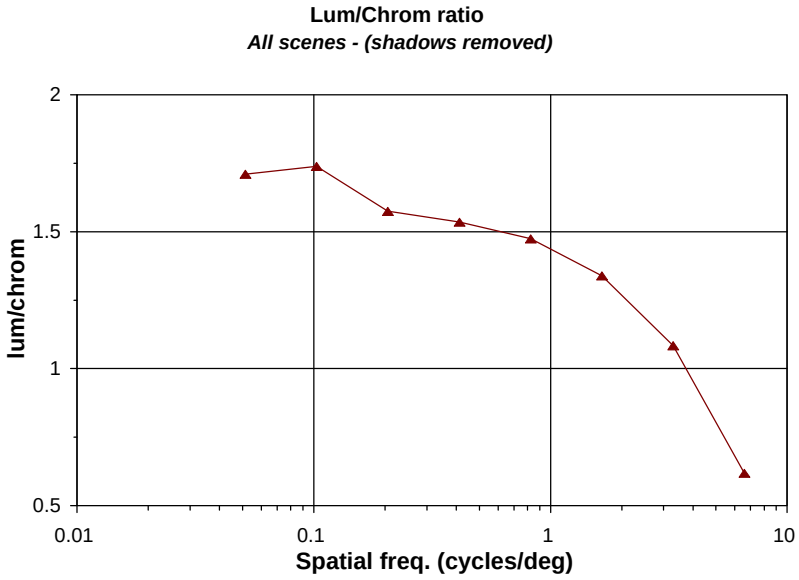

Figure 4.6: Average ratio of the *lum* image amplitude to that of the *chrom* image amplitude for the *shadow-removed* scenes. This average is done across all the dataset.

Working with the *shadow-removing* definition of *chrom*, we repeat our measurements as described in the previous section. The *lum* to *chrom* average is similarly obtained for all scenes and Figure 4.6 shows the results. In this Figure we see that the basic features obtained for the other definitions still remains here: for low spatial frequencies there is relatively more *luminance* than *chrominance* information. This is reversed only for the highest spatial frequency considered in the plot (the value of lum/chrom is lower than 1). The only notable new feature is the sagging of the plot. Using the previous definitions this was slightly pronounced in the opposite direction.

From this we conclude that shadowing does not affect the broad conclusions obtained in the previous sections.

4.3 *Analysis of the Fourier amplitude spectra of lum and chrom images*

Another analysis of the spatial frequency content of natural scenes has been carried out by Field (1987). He relates the statistical structure of the environment to the coding properties of the visual system (see Chapter 1). Field derived the amplitude spectra for six achromatic images (equivalent to the "FBANDS" analysis mentioned on Chapter 2). He found that the amplitude spectrum of these images is greatest at 0 spatial frequency (the mean intensity of the image) and falls off quickly by a factor of roughly f^{-1} (where f is spatial frequency). This is a straight line with slope $\alpha = -1$ when plotted on double logarithmic co-ordinates. This leads to an invariance in frequency content when magnifying any part of an image (See Chapter 1 and Appendix B).

Other workers (Tolhurst *et al*. 1992) examined the amplitude spectra of 135 digitised photographs of natural scenes and found that relatively few of them follow the previous statement. They suggest that the average slope of the Fourier amplitude spectra when plotted on double logarithmic co-ordinates is $\alpha = -1.2$ (steeper than the one suggested by Field).

Following the above, the *lum* and *chrom* images were obtained from every scene of our dataset and subjected to similar analysis. The goal here is to find whether they show similar statistics and to estimate the average value of the slope (α). Notice that both Field and Tolhurst *et al.* applied their analysis only to an achromatic image set. Our analysis is as follows:

a) "DO_LMS" (see Chapter 2) is applied to each of the *reflectance* scenes again and two of its output images are combined using the *simple* definition *(lum = L + M, chrom = L - M).*

b) The Fourier amplitude spectra image is derived for every *lum* and *chrom* image using "FBANDS" (see Chapter 2). The centre is again removed. Both Fourier images are scaled so that each contains unit amplitude in the total spatial frequency spectrum. The average amplitude spectrum (averaged across orientation) within each band of the "logarithmic" set of concentric bands is measured. The logarithmic set of bands was chosen because it provides an even distribution of the sampled data when plotted along the logarithmic Fourier space axis.

c) Figure 4.7 shows the average amplitude spectra for the *lum* and *chrom* images for all the dataset (29 natural scenes). When plotted in log-log axis, these average amplitude spectra (averaged across orientation and across the dataset) are roughly straight lines. The only possible exceptions are the values corresponding to the high spatial frequency section of the Fourier space.

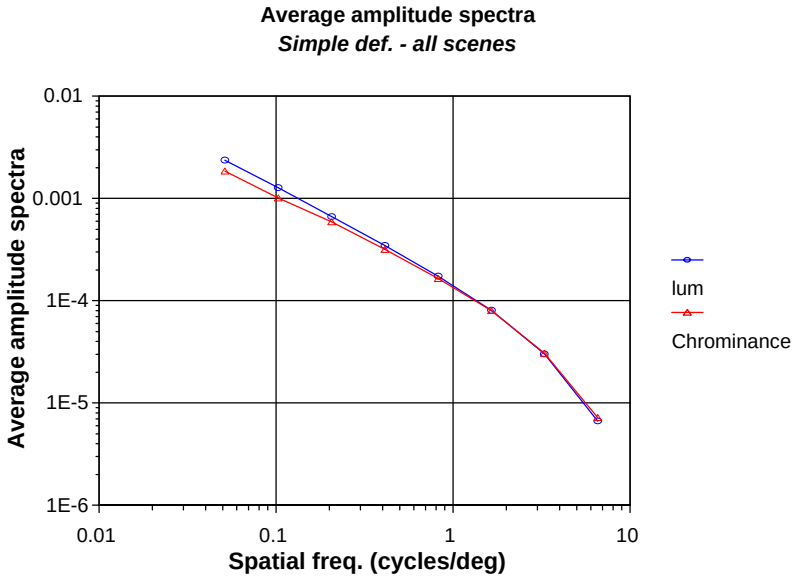

Figure 4.7: Amplitude spectrum for *lum* to *chrom* averaged across orientation for all the dataset (29 scenes)

The slope (α) of this line is measured for each scene individually using linear regression. The average values and their standard deviation are:

	slope (α)
Lum	-1.11 ± 0.13
Chrom	-1.06 ± 0.11

Similar analysis is performed using the *shadow-removing* definition of *lum* and *chrom* as described in section 4.2. Figure 4.8 shows the average amplitude spectra for these *lum* and *chrom* images. The resulting values of the slope (α) are:

	slope (α)
Lum	-1.11 ± 0.13
Chrom	-0.94 ± 0.12

The values of the slope for chrom images are only slightly different from the previous ones. This means that *shadow-removing* definition produces no important variation in terms of average amplitude spectra for the *chrom* scenes.

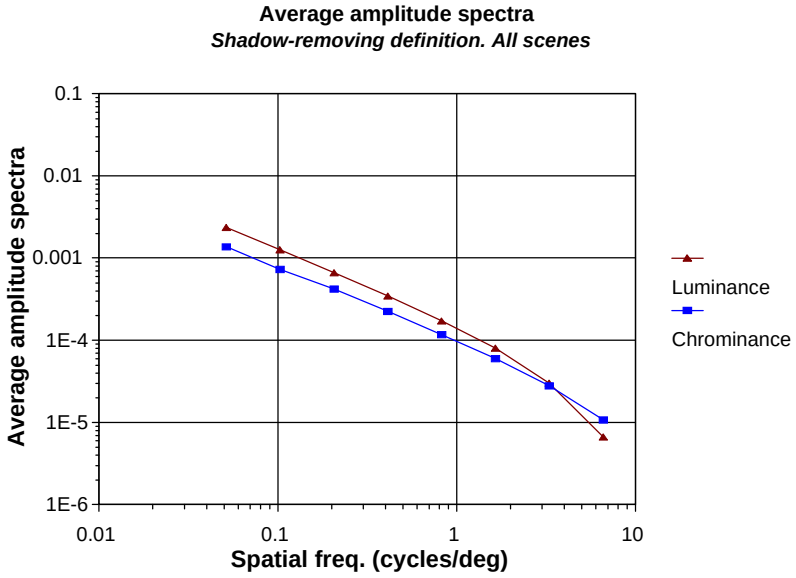

Figure 4.8: Average amplitude spectra for *lum* and *chrom* images when the shadow-removing definition is used.

The results in Figure 4.7 and in the previous section show that the amplitude spectra averaged across orientation follows:

$$\text{amplitude} = k \cdot f^{\alpha}; \qquad (k = \text{constant})$$

for our dataset. As seen above, the average value for the slope (α) of the line

corresponding to the *lum* scenes is bigger than the corresponding to the *chrom* scenes. This is consistent with the previous findings in terms of the *lum/chrom* ratio for the dataset (Figures 4.1 to 4.5).

The above result approximately confirms Field's findings in terms of the amplitude spectra and power spectra of his set of six achromatic images and extends the implications to the colour image set. If the statistics of natural scenes are stationary (i.e. they remain constant if any local region of the scene is magnified), the array of sensors described by Field (see Chapter 1 and Appendix B) produces an even distribution of the information. Following this, Field considers that the human visual system is efficient because it is well matched to the statistical redundancy of the visual environment. It is possible to further criticise his work by considering the effects of cortical neural noise, the optics of the eye, and the spacing between cortical neurons which does not correspond to the rigid grid proposed by him. For a more extensive discussion and review of the implications of these findings for human colour vision see Chapters 5 and 6.

Chapter 5

Discussion

Consequences for human visual system coding

In this Chapter we discuss the findings exposed in Chapters 3 and 4 in terms of their consequences for the h.v.s. information coding. Several facts are already known about *how* the early stages of the h.v.s. process information. For example, we have models of the distribution and spectral sensitivities of the photoreceptors in the retina, the interconnections among receptors, horizontal cells, bipolar cells, etc. in the retina, the retinogeniculate colour and luminance organisation, and the spectral sensitivity function. We also have evidence of the band pass and low pass filter properties of the contrast sensitivity functions (which is thought to be the envelope of many more narrowly tuned spatial frequency selective channels) and the arrays of neurons selective to local regions of space, spatial frequency and orientation in the primary visual cortex. In this work, we focused on *why* some of these strategies of representing the visual environment might evolve. Our procedure is to analyse the statistical structure of the visual environment, trying to find the relation between it and the visual system's representation of this structure.

5.1 *The spectral sensitivity functions of the human visual system and the spectral radiance and reflectance of natural scenes.*

The h.v.s. "hardware" is optimised to extract information in the spectral region where the solar radiation is maximised. As mentioned before, the IR and UV spectral limits represent a compromise between the hardware constraints of the h.v.s. and this requirement. Extending the sensitivity into the UV would have required the capacity to repair damage to cells and to further correct chromatic aberrations within the eye. To extend this sensitivity into the IR would have lowered the signal-to-noise ratios because the human body continuously radiates IR radiation according to body temperature. This body heat is also present within the eyeball so if we were sensitive to the IR we would

probably have higher noise interfering with the information from the real world. The fact that even within these hardware limitations, the h.v.s. extracts information in the spectral region where the solar energy is maximised leads us to the assumption that there is an evolutionary advantage in doing this and the advantage may be relevant within the visible spectrum itself.

In Chapter 3 we compared the spectral sensitivity functions for the luminance and chrominance channels of the h.v.s. and the mean spectral radiance and reflectance from our scenes. The curves on Figures 3.7 (a) and 3.7 (b) show a correspondence between these functions within the range 400-680 nm. For wavelengths greater than 680 nm the spectral radiance and reflectance of our scenes have higher values and little correspondence to the contrast sensitivity curves. The explanation for this peak in the far red may come from the spectral reflectance of chlorophyll. As mentioned before, there is a compromise between the spectral distribution of the energy available in the natural environment and the hardware constraints of the h.v.s. It is likely that the advantages of having a receptor tuned to IR may not compensate for the disadvantages of having to deal with the noise produced by human heat.

The shape of the spectral sensitivity function for the chrominance channel may have another explanation. The slope of the curve also represents the discrimination abilities of the h.v.s. The steeper the slope, the better the h.v.s. can discriminate between opponent colours and the curve may have its significance in optimising the discrimination of some colours from the background (in this analysis we did not include discrimination tasks as a goal of the h.v.s. and we only considered the energy distribution across the visible spectrum).

5.2 *The spectral distribution of spatial frequencies (Fourier content) of the natural scenes.*

In Chapter 1 we discussed the *luminance* and *chrominance* amplitude spectra of natural scenes and the coding properties of the cortical cells. There it was manifest the advantage of processing images with stationary statistics according the model proposed by Field (1987). These images produce an even distribution of the information among the array of scale-invariant sensors modelled by the theory. The consequences of processing images with non-stationary statistics are less clear. Nevertheless, we can

assume that according to the model, it is an advantage for the h.v.s. to process images with statistics as close as possible to the "optimal" stationary statistics.

Figures 3.8 to 3.11 show the spatial frequency structure (Fourier content) along the visible spectrum. The structure found for achromatic *lum* and *chrom* images (Field 1987; Tolhurst *et al.* 1992) is repeated along the visible spectrum for all the wavelengths examined. From this we can infer that the h.v.s. is not constrained by the spectral distribution of the Fourier content to extract information from any particular spectral region. The only modification of the spectral distribution of the Fourier content is shown in Figure 3.12. What varies is the value of α (slope of the amplitude spectra when plotted on double-logarithmic co-ordinates) across the wavelength axis.

As stated earlier, the consequences for the h.v.s. of processing images with nearly stationary statistics are not clear. Figure 3.12 shows that the most "optimised" (close to -1) value of α occurs in the region of the visible spectrum where the h.v.s. is optimised for collecting information. From this we could infer that if there is an advantage of coding images where α is as close as possible to the optimal value of α= -1.0, this is represented by the distribution of the receptors across the visible spectrum.

Subsequently, both the spectral distribution of the energy and the spectral distribution of the Fourier content of natural scenes, along with hardware constraints, might suffice as a qualitative explanation of why the h.v.s. samples information within the visible spectrum in the way it does.

5.3 *The spectral distribution of spatial frequencies (Fourier content) of natural scenes and its relation to the bandwidth of the receptor.*

It has been argued that the broad nature of L, M, and S receptor bandwidth is because of a strategy of avoiding "blind spots" for wavelength. In addition, having many receptors would ensure that all the visible spectrum is scanned but would impair our acuity for colour. Here we have a trade-off between having many receptors or improving the acuity for colour. The answer was clearly to have a few spectrally broad receptors. In the case of the fovea, where receptors are more densely packed to gain acuity, we have only two types (L and M cones).

On the other hand, the fact that most natural objects are not spectrally narrowband

implies that spectral bandwidth of the receptors should not necessarily be broad. Given that the L, M, and S spectral sensitivities are indeed spectrally broad this suggests that it is necessary to explore some other characteristics of the environment in search of an explanation. In Chapter 3 we explored how the spatial characteristics of natural scenes changed when three hypothetical receptors become increasingly broader across the wavelength spectrum.

This increase in the bandwidth of the receptors produces no further changes in the Fourier structure of the spectral region examined. As Table 3.1 shows, there is also no statistically significant change in the slope of the Fourier amplitude function when the bandwidth of the receptor increases. The slope measured in all cases is coincident with the average slope within each spectral region (see Figure 3.12). This leads us to conclude that the spatial structure of the environment has not played any key role in the design of the spectral bandwidth of the receptors. Other features of nature must be relevant for the human visual system to adopt the strategy of using spectrally broad L, M and S receptors. To account for these it is necessary to include survival strategies like recognising predators or efficient search for food, which were not considered here.

5.4 *Influence of the separation between the L and M receptors on the spectral distribution of spatial frequencies (Fourier content) of chrominance information extracted from natural scenes.*

The fact that the L and M cone sensitivity functions overlap accounts for efficient colour discrimination. Less understood is why this overlap involves a broad region of the visible spectrum and why the peaks of the L and M sensitivity functions are spectrally close to each other. In the same manner as in the previous section, we explored the effects on the spatial frequency structure of the chrominance (L-M) information channel of modifying the *spectral distance* between those receptors. The results are summarised in Table 3.2 and Figures 3.17 to 3.20. Table 3.2 shows the values of the slope (α) for the chrominance images when the distance between L and M receptors and the bandwidth of the receptors are modified. Notice how the slope moves away from the optimum value (α= -1) when the centre of the L receptor approximates to the centre of the M receptor. On the other hand, the value of α seems to approximate to the optimum when the spectral bandwidth of the receptors increases.

Overlapping the spectral sensitivity functions of the L and M cone receptors leads to a steeper value of α (thus, less optimised) when the chrominance (L-M) image is considered. On the other hand, increasing the L and M receptors' bandwidth produces the opposite effect on the chrominance image. These two effects compensate each other, so when the bandwidth is 110 nm and the centre of the L cluster is in 570 nm the average value of α is -0.94 (not far from the optimum). Consequently, the worst value of α
(-0.76) is for receptors of 10 nm width and 10 nm spacing along the wavelength axis.

Again we face the problem of the significance of the small variations in the value of α for the h.v.s. coding scheme. If we assume that these variations are important to optimise the coding, and the optimisation of the coding is of some value for survival, then the data on Table 3.2 could help to explain why the h.v.s. extracts colour information mainly using two largely overlapping receptors.

There are some important features not explained by our results. For example there seems to be an important reason for the h.v.s to have the peaks of the L and M cone receptors separated by a small spectral distance (see Figure 1.3 -spectral sensitivity of L, M and S mechanisms) despite the fact that this leads to non-optimised values of α. This could be related to an optimum colour discrimination of targets lying over a green background, which is not included in our analysis. The change in the value of α caused by the proximity of the L and M cone receptors is compensated for by broadening their spectral bandwidth.

From the above we conclude that the h.v.s. has adopted an adequate strategy to extract colour information from the environment but that many of the chosen parameters are somewhat uncritical. This strategy consists of collecting information from the spectral region where the statistics are nearly scale-invariant and combining this information in a way that compensates for digressions from this invariance.

5.5 *The lum/chrom ratio of natural scenes and the contrast sensitivity function in humans*

As seen in Chapter 4 (Figures 4.1 to 4.5) the shape obtained for the *lum/chrom* Fourier energy ratio suggests an entirely different relationship from that expected from

Mullen's (1985) measurements for the contrast sensitivity functions (Figures 1.14 to 1.16). This is especially marked for the long-distance set of scenes and also true for the *shadow-removing definition* tested. There may be different reasons for such a discrepancy:

(a) The high spatial frequency chrominance information may be irrelevant for primate vision task.

Consider as an early task of the visual system the segregation of different parts of a visual scene in order to identify discrete objects. The spatial distribution of wavelength differences (colour) would contribute to the differentiation of some objects from their surrounds and make them more identifiable despite the irrelevant contours produced by the shadowing. Consider the problem of finding a cherry in a cherry tree. Typically, the cherry will have a different luminance and a different colour compared to its background. Our luminace vision gives us a good representation of the boundaries of the cherry. However, other luminace properties of this scene will also produce luminance boundaries (e.g. shadows, textures, light gradients). What is needed to aid the perceptual segregation of the cherry and its background is an independent representation of its properties. This is conveniently provided by the low-pass characteristics of colour vision which enhance the global properties of the cherry (its redness) and largely ignore the boundary representations already provided by luminance. Such independence provides greater robustness in the segregation task.

(b) The neural machinery has evolved to compensate for the chromatic aberrations of the optical system of the eye.

These aberrations, called *longitudinal (or axial)* and *transversal (or radial)* occur in all lenses and produce different effects on the image displayed on the retina.

Longitudinal (axial) aberration: for example, if the eye is accommodated (focused) on a distant red target of 700 nm wavelength, a distant violet one of 400 nm will be seriously blurred. Le Grand (1967) estimated that a subject emmetroptic for yellow-orange (589.3 nm) would become myopic by 1 diopter for violet (430.8 nm) and hyperopic by nearly 0.6 diopter for deep red (768.1 nm). There is no possibility that the mixed rays from a white target can all be optimally focused in the retina.

Transversal (radial) aberration: longitudinal aberration is an aberration just in position

(different chromatic images are located in different planes). Transversal aberration is an aberration in dimension (the size of the retinal image varies with colour). These effects are usually combined and the final result is that only one monochromatic image is in focus in the retina and the others are surrounded by a blurred border (Le Grand 1967).

These aberrations affect principally the chromatic information that reaches the retina, producing variations in luminance from them. For example, consider a red-green pure colour grating where the dominant wavelength varies sinusoidally across the stripes of the grating and the luminance is equal at every point of the display. This grating can, for example, be produced by interleaving two sinusoidal luminance gratings of 630 and 520 nm. When the eye focuses on the 630 nm component, the longitudinal aberration will cause the other component to be defocused (with no change in its spatial frequency or its phase, but with a reduction in its amplitude). This effect will produce a grating that varies in luminance as well as in chrominance.

On the other hand, transversal aberration causes patterns produced by light of two different wavelengths to be slightly differently magnified at the retina. If the same chromatic grating as mentioned above is used, the retinal images of the 630 and 520 nm gratings would have slightly different spatial frequencies producing luminance artifacts.

An example artifacts produced by chromatic aberrations on the retina are the chromatic fringes that appears near a black and white edge (Walker 1977).

Artifacts are present on the retina and distort the real information about the outside world. The filtering of the high spatial frequency colour information existing in the contrast sensitivity functions (Mullen 1985) may be the way the visual system eliminates that part of the message that carries wrong information about the real world. One example of this is described by Berry and Wilson (1993). They reported that interference fringes decorating caustics appear black and white with high contrast instead of highly coloured as expected by the theory.

(c) Physiological constraints:

The optic nerve is a bottleneck in the information path from the retina to subsequent brain centres. There is therefore strong pressure to maximise the amount of information transmitted per optic nerve neuron, which implies strong post-receptoral re-coding of the stimulus information. Ingling and Martinez (1983) suggested that a single opponent

cell can produce both summing and differencing signals in order to send both luminance and chrominance information down a single neuron (multiplexing technique). The differencing signal (chrominance) is produced at low spatial frequencies and the summing signal (luminance) at high spatial frequencies. “Spatial frequency” is used here to signify the degree of change compared to a neighbouring neuron in the retinotopic array. This scheme predicts (and critically depends on) different lum and chrom contrast sensitivity functions. A combination of this and the considerations in *(b)* may be a powerful set of reasons for the development of our contrast sensitivity profile.

(d) Absence of shadows in our dataset.

As mentioned before, strong shadows were avoided as much as possible in our dataset. In our set of indoor scenes taken inside the laboratory, diffuse lighting was employed to illuminate them. A shadow implies high spatial frequency luminance but very little high spatial frequency chrominance since colour remains relatively constant across a shadow boundary. Figure 5.1 shows the lum/chrom amplitude ratio measured for the special scene described in Chapter 4, section 4.2.

This particular scene consisted of vertical coloured cards with a strong shadow produced using a projector and a ND filter. Both lum and chrom were obtained using the simple definition mentioned in Chapter 4. Notice how the ratio between lum and chrom has a completely different trend from that obtained for our dataset (Figure 4.1). When the *shadow-removing* definition of chrominance (as defined in Chapter 4, section 4.2) is applied to this scene, one obtains a result in which there is little high s.f. information in the chrominance domain leading to a positive slope in Figure 5.1.

Notice in Figure 5.1 that the ratio lum/chrom is closer to what we expect from Mullen’s (1985) contrast sensitivity measurements (see Chapter 3). The absence of shadows in our dataset perhaps could explain our failure to find an agreement between these psychophysical measurements and our lum/chrom figures.

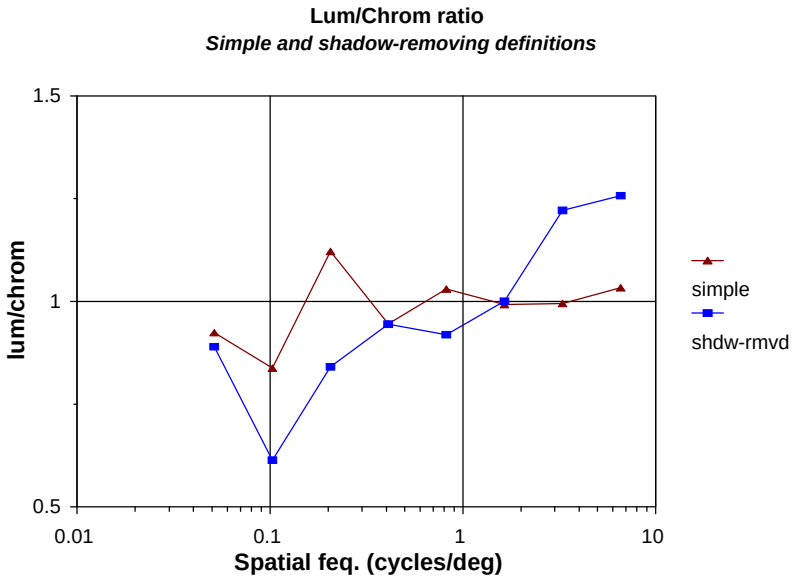

Figure 5.1: Lum/chrom ratio for a special scene consisting of a collage of coloured papers with a strong shadow boundary. *Simple* and *Shadow-removing* definition were used.

On the basis of these results, it may be possible to argue that the characteristics of the h.v.s. could be tuned to a certain type of scene which does not readily exists in the (northern temperate) environment in which we analysed scenes. If the world consisted of relatively large, uniform coloured areas and strong shadows were cast across these, then the Fourier amplitude ratio of such a set of scenes could be expected to show the characteristic upward slope predicted from Mullen's data. It can be established in future research whether there are real visual environments for which these characteristics can hold.

(e) Artifacts in the acquisition of our dataset:

To be discussed later in this chapter.

5.6 *The luminance and chrominance amplitude spectra of natural scenes and the coding properties of the cortical cells.*

As Tolhurst *et al.* (1992) suggested, the assumption that the amplitude spectra of natural scenes follow:

$$e(f) = k \cdot f^{\alpha} \qquad ;(k\text{= constant})$$

with α= -1.0 made possible the development of computational studies of the efficiency of the visual cortex's coding scheme (Field 1987, 1989) and psychophysical paradigms for studying the processing of natural images by the visual system (Tadmor and Tolhurst 1990). This "efficient" coding was modelled from neuro-physiological studies of the visual cortex (Blakemore and Campbell 1969) using Gabor's (1946) theory of communication (see Chapter 1). As discussed in Chapter 1 and Appendix B, the spatial frequency selective channels proposed by Field have constant bandwidth in octaves. If the statistics of the natural scenes are also stationary (α= -1.0), then all different sensors will carry equivalent amounts of information. From here, he concludes that coding a scale invariant image into an array of scale-invariant sensors produces an even distribution of the information (i.e. well distributed capacity within each channel on average).

Our results in Chapter 4 show that for the natural scenes of our dataset, the amplitude spectra still follow the same relationship but the mean values of the slope (α) are different: α= -1.11 ± 0.13 for *lum* images and α= -1.06 ± 0.11 for *chrom* images. The mean values of the slope using the *shadow-removing* definition of *chrom* also change little compared to the previous figures: α= -0.94 ± 0.12. To see whether this change in the slope (α) is statistically significant we applied Student's t-test using a 95% confidence interval to the value of α obtained for chrom images in both definitions. The results indicate that they are statistically different.

The fact that amplitude spectra of the luminance images of our dataset have a steeper slope when plotted in log-log co-ordinates implies that the efficiency of the coding scheme suggested by Field will be slightly diminished: how important is this change in the slopes?

Consider an image where the amplitude spectrum follows $e(f)= k f^{\alpha}$ with $\alpha \neq -1$. The amount of energy between frequency f_0 and frequency nf_0 when viewed at a distance d is:

$$\int_{f_0}^{nf_0} k \cdot f^{2\alpha} \cdot 2\pi f \cdot df = \frac{2k\pi}{2\alpha+2} \cdot \left[\left(nf_0\right)^{2\alpha+2} - f_0^{2\alpha+2} \right]$$

$$= k_0 \cdot \left(n^{2\alpha+2} - 1\right) \cdot f_0^{2\alpha+2} \qquad \text{(Eq. 5.1)}$$

where k and k_0 are constants.

If the viewing distance is increased by a factor a the frequency range will be shifted to the range between af_0 to anf_0. The corresponding value of the energy within those bands will be:

$$k_0 \cdot \left(n^{2\alpha+2} - 1\right) \cdot a^{2\alpha+2} \cdot f_0^{2\alpha+2}$$

For example, if the value of α is -1.1, and the viewing distance is increased by a factor of a= 2 the corresponding change in the energy inside the band will be

$$a^{2\alpha+2} = 0.87$$

Thus the energy on increasing the viewing distance by a factor of 2 will be 13 per cent less.

Even when the viewing distance is the same, the values of the energy vary between two adjacent octaves. Figure 5.2 shows the values for the energy calculated using Eq. 5.1 (see above) for slopes α= -1.2, α= -1.15 and α= -1.1. k_0 is an arbitrary constant. The Figure shows how different from a horizontal line (expected from scale invariant scenes where α= -1) are the values of the energy according to the variation of the value of α.

Although the previous section shows the difference in terms of Fourier energy of having values of α different from -1, it is a long way from saying that the model of the h.v.s. coding scheme constructed by using Gabor's theory of the information is inefficient. These changes in the average slope might be assimilated within the limitations of the model, which treats the visual system as having constant bandwidths measured in octaves although the spatial frequency bandwidths of cortical cells show a fair degree of scatter. Another fact which is not considered in the model is the evidence that high spatial frequency cells have narrower bandwidths (in octaves) at higher spatial frequencies. The average bandwidth considered by Field is 1.4 octaves.

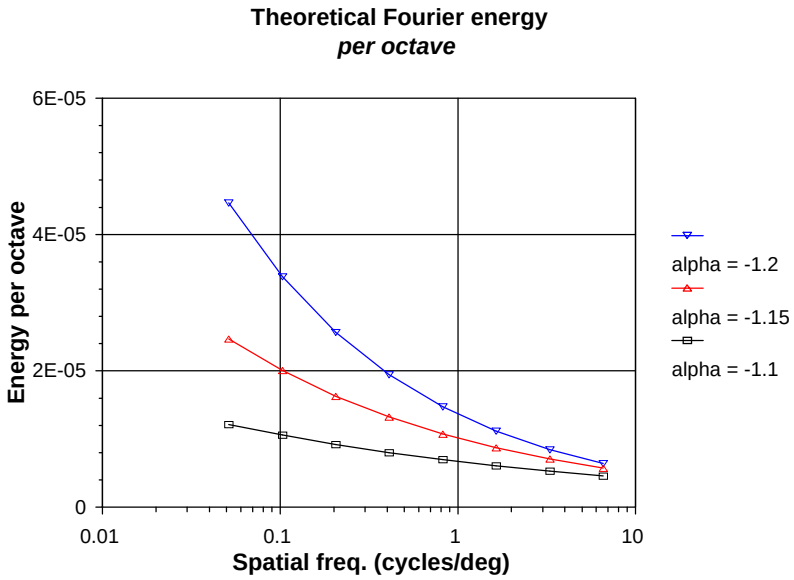

Figure 5.2: Estimated values of the Fourier energy per octave for three different values of α.

The fact that the statistics corresponding to the *chrominance* channel are roughly the same as that of the *luminance* channel may have implications for the h.v.s. chromatic system. There is both physiological and psychophysical evidence for multiple spatial frequency channels in the colour domain, i.e. there is a similarity between the spatial filtering characteristics of colour and luminance striate cortex cells, (Thorell *et al.* 1984) and there is selective colour adaptation in the h.v.s. (Bradley *et al.* 1985). This evidence is more limited than in the case of the luminance channels (see Chapter 1) and it suggests that multiple colour sensitive spatial frequency channels are less finely tuned for spatial frequency than the equivalent luminance channels.

Our findings of similar statistics for the *lum* and *chrom* images in Chapter 4 support the idea that colour-based spatial vision also operates very much like luminance-based spatial vision.

5.7 Possible artifacts that might have led us to the previous results

The most important artifacts could be summarised as:

a) Calibration artifacts.

b) Effects of light changes and wind.

c) Poor focus - chromatic aberrations of lens.

d) Unrepresentative visual environment in our dataset.

(a) Calibration artifacts: As stressed in Chapter 2, there was a mismatch between the measurements of the TopCon SR1 spectro-radiometer and the results obtained with our DRA Camera. This mismatch was marked especially for red sheets of paper in the spectral range 580-700 nm. There appears to be no obvious relationship between this mismatch and our results in Chapters 3 and 4.

(b) Effects of light changes and wind: These are relevant when the integration time is high or when we combine one or more images to produce L, M, S cones output, *receptors*, etc. In these cases a test group of scenes was considered separately. This test group was composed of scenes that obviously could not have been affected by light changes or wind, as in the scenes taken inside the lab, or in particularly good conditions. The results corresponding to the test group were compared with those of the rest of the dataset to confirm the independence of the findings from this possible artefact.

(c) Poor focus - chromatic aberrations of lens: For the *lum* images, this would have led to less high spatial frequency Fourier energy and thus steeper slopes than those previously reported for achromatic images (Tolhurst, *et al.* 1992), which is not the case here.

Poor focus might also have affected some images in a given scene leading to lesser values of the Fourier amplitude spectra in the high spatial frequency range and thus steeper slopes. For example, chromatic aberrations in our DRA camera lens might be responsible for defocusing images taken in the blue range of the spectrum and this might have induced the values shown in Figure 3.12. Here we analyse this problem.

In our analysis of the slope of the Fourier amplitude spectra we explored several images along the visible spectrum. These images were either chromatically narrowband (filtered) images or linear combinations of them (*receptors*, simulations of the L and M cones output, etc.). In all cases the algorithm employed to obtain the slope α of the amplitude spectra was the *logarithmic* version of FBANDS as described in Chapter 2. We choose this algorithm for two important reasons:

(a) The results show less dependence on high spatial frequency than, for example, using

the *linear* version of the same algorithm. This is because of the annulus-shaped division of the Fourier space performed (see Figure 2.10). In the logarithmic arrangement, there is only one measurement of the average amplitude spectra corresponding to high spatial frequencies (most external ring in Figure 2.10). The influence of this region upon the rest of the Fourier space is thus minimised. In the other cases, the same area is normally covered by several annuli which are more influential at the time of evaluating the amplitude spectra's slope. Other researchers (Field 1987; Tolhurst *et al.* 1992) evaluate the average amplitude spectra of this region in multiple annuli which are still more critically dependent on defocus.

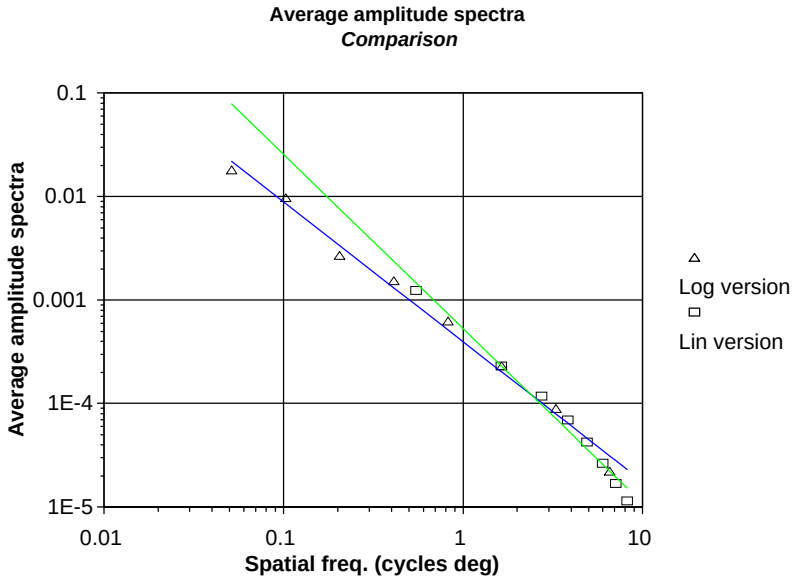

Figure 5.3: Comparison between the slope obtained with the *logarithmic* and the *linear* version of FBANDS. The former produces a distribution of values more evenly spaced along the spatial frequency axis than the later, resulting in a more precise value of the slope.

(b) The data obtained using the logarithmic version of FBANDS are more evenly spaced when plotted in log-log co-ordinates. As shown in Chapters 3 and 4, FBANDS in its *logarithmic* version produces an output consistent with 8 estimations of *E(f)* (average amplitude spectra) equally spaced along the logarithmic spatial frequency axis. To use the *linear* version to produce the same output would have led to an over-representation of the high spatial frequency energy when estimating the slope of the plot on double-logarithmic co-ordinates. Figure 5.3 illustrates how a different version of FBANDS could lead to different results when estimating the slope α.

The same image was analysed using the *linear* version of FBANDS and the *logarithmic* version and the slope α obtained strongly depends on how evenly spaced is the data along the spatial frequency axis. The same might be true for other studies. Figure 5.4 shows the averaged amplitude spectra of four images from Tolhurst's dataset (Tolhurst *et al.* 1992). Notice the concentration of the data on the right side of the plot. This could have lead to an over-representation of that sector when estimating the slope of any straight line in the log-log plot. The situation is more complex considering that the sector over-represented is specially influenced by defocus noise.

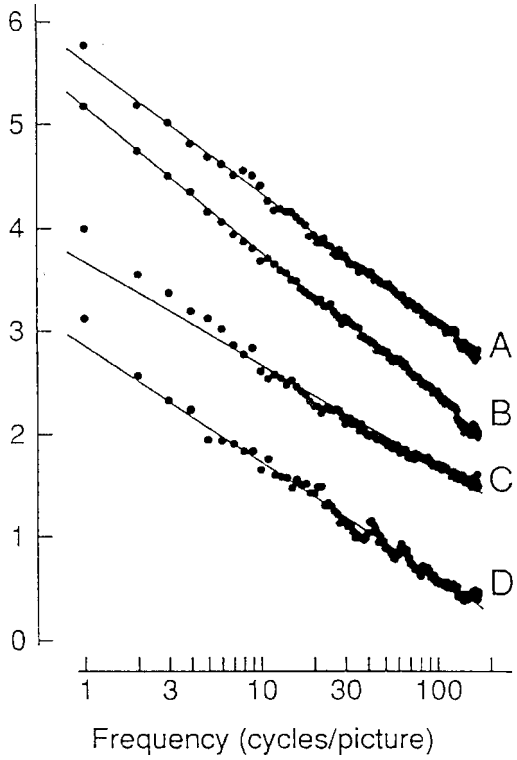

Figure 5.4: Amplitude spectra averaged across orientation for natural images. From Tolhurst *et al.* (1992).

Despite the fact that the computer algorithm we chose is not strongly affected by defocus, we checked whether or not the defocus produced by our lenses could influence the measurements of α across the visible spectrum. For this we measured the amplitude spectra slope across the visible spectrum of a pair of non-natural scenes. The processing was similar to that described in Chapter 3 (logarithmic version of FBANDS, *image-based normalisation*). Objects in these scenes were chosen to be completely artificial and the scene was recorded under controlled conditions inside the lab. Both scenes

consisted of a set of coloured papers on a grey background and were illuminated with tungsten lamps. Figures 5.5 and 5.6 show the amplitude spectra across the visible spectrum for scene 1 and scene 2 respectively. Figure 5.7 shows the values of α for these non-natural scenes across the visible spectrum. From Figure 5.7 we see that the value of α for scene 1 does not show a decrement for the blue side of the visible spectrum. In contrast, its value is surprisingly high. The values of slope for scene 2 are approximately constant across the spectrum. This is contrary to what is expected in case of a defocus in the blue side of the spectrum. We therefore rule out an explanation based on chromatic aberration of the camera lens.

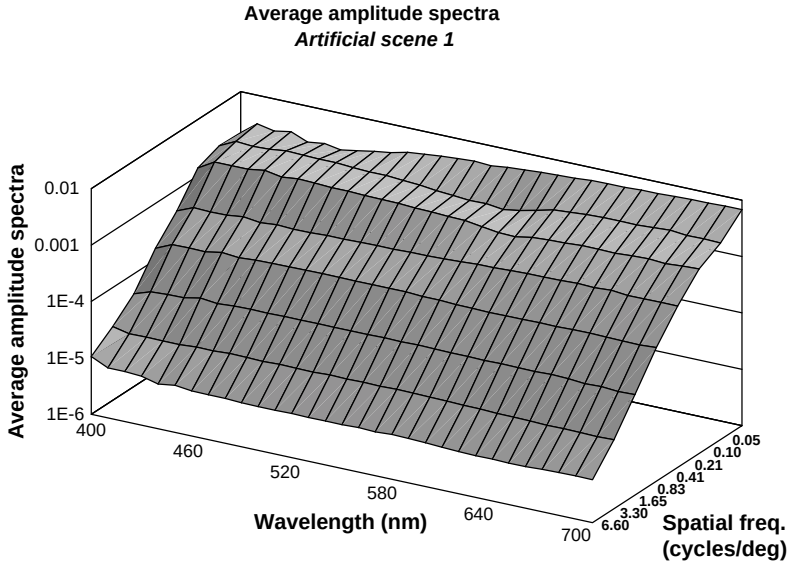

Figure 5.5: Amplitude spectra across the visible spectrum for scene 1.

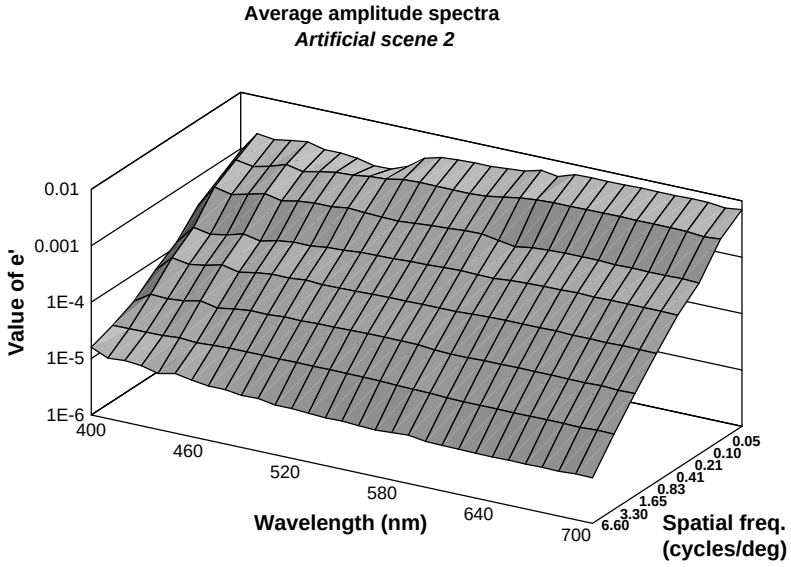

Figure 5.6: Amplitude spectra across the visible spectrum for scene 2.

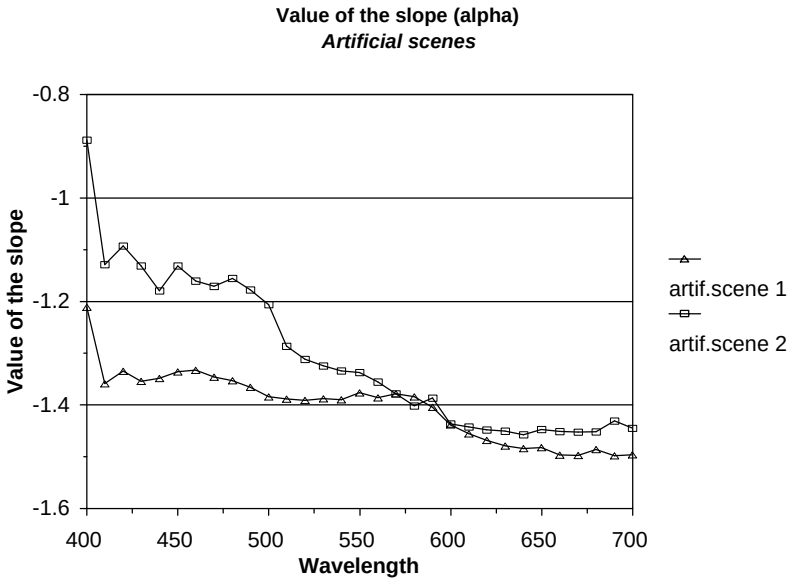

Figure 5.7: values of α for artificial scenes 1 and 2 across the visible spectrum.

(d) Unrepresentative visual environment in our dataset: Section 5.5 pointed out that the characteristics of the h.v.s. could be tuned to a certain type of visual environment which is not fairly represented by our dataset. In that environment, strong shadowing might play a greater role and consequently, the statistics in terms of the slopes of the amplitude spectra of lum and chrom images would be different. Within the constraints imposed by the time of the year, the geographical location, the lack of strong shadowing, etc., our dataset intends to be as representative as possible of the natural environment. Further work is required to establish whether different environments (e.g. strong shadows and directional illumination) would give rise to different image statistics, possibly in closer agreement with the psychophysical data.

Conclusions

Our analysis of the dataset of natural scenes recorded using the DRA-Camera leads us to the following conclusions:

1) The Smith and Pokorny (1975) cone sensitivity functions are consistent with the mean distribution of spectral reflectance and radiance of our natural scenes. Our analysis only considered the coincidence between the two maxima as an efficiency criterion. Other factors may need to be considered here to reach a more general conclusion.

2) The shape of the average distribution of spatial frequency (Fourier content) of natural scenes shows no significant variation with wavelength. The average value of the slope (α) tends to be closer to -1 in the range 570 - 640 nm. It is not clear yet to what extent this provides support to the theories about the function of cortical cells posited in Chapter 1. No conclusive results about the influence of the spectral bandwidth of the receptor were obtained.

3) The amplitude spectrum of L-M images was explored for different bandwidths and separations between the L and M receptors. The average distributions of amplitude spectra have similar characteristics to those found in the previous measurements. Our results in Table 3.2 show a tendency of the slope (α) to take steeper values when the bandwidth of the receptors is increased and when the separation is increased.

4) The physiological imbalance between colour and luminance (Mullen 1985) is not reflected in the spatial content of our dataset. This is unlikely to arise from artefacts of our dataset and it does not depend of the definition of luminance and chrominance used in this work.

5) The amplitude spectrum of *luminance and chrominance* images exhibits approximately the same characteristics as the amplitude spectra of *achromatic* images investigated by other workers (Field 1987; Tolhurst *et al.* 1992). The average value of the slope (α) measured here is $\alpha = -1.11 \pm 0.13$ for *lum* and $\alpha = -1.06 \pm 0.11$ for *chrom.*). The value of α obtained using a *shadow-removing* definition of chrom. is

$\alpha = 0.94 \pm 0.12$.

In this work we considered the optimisation to extract colour and spatial information from static scenes as the most important factor in our analysis of the h.v.s. Despite that there should always be an advantage in being optimised to extract that information from the environment, in some cases this might not be the most important selective reason. As our results show, in some situations the optimisation of this information extraction is apparently an important goal (1), sometimes its importance is not clear (2), and in other situations it is secondary. A complete analysis examining other factors such as motion, stereo vision, priorities for survival, etc., must be considered in order to better understand their influence on the operation of the human visual system.

Appendix A

The DRA Camera

In order to better understand the work of the DRA Camera, the following sections (A1 to A5) were summarised from the draft "DRA Camera Theory" by Gavin Brelstaff, Perceptual Systems Research Centre, University of Bristol, UK, October, 1993.

A1 DRA-Camera theory

The DRA Camera was constructed to work as a simple image forming device. Its optics are described in Figure A1. Image irradiance is converted into signal by the sensor using the following assumptions:

a) The system is properly focused.

b) There is no vignetting

From those assumptions we can write the following equation:

$$E(x,y) = L(x,y) \cdot a \cdot k(x,y) \ ,$$

where *E(x,y)* is the flux per unit area at the arbitrary point *(x,y)* or *scene irradiance* in $W.m^{-2}$. *L(x,y)* is the flux per unit foreshortened surface area per unit solid angle or *scene radiance* that projects on *(x,y)* in $W.m^{-2}.sr^{-1}$. *a* is the *transmittance of the aperture* and *k* is the *optical sensitivity*. The last one is usually dependent on the position *(x,y)*.

To complete the previous equation we need to add an interference filter of *spectral transmittance F(λ)* and to consider the *spectral sensitivity* s(λ) of the camera's sensors. Then the above equation becomes wavelength *(λ)* dependent as follows:

$$E(x,y,\lambda) = s(\lambda) \cdot F(\lambda) \cdot L(x,y,\lambda) \cdot a \cdot k(x,y)$$

Here we assume that chromatic aberrations in the DRA Camera optics are negligible and then *a* and *k(x,y)* do not vary with wavelength.

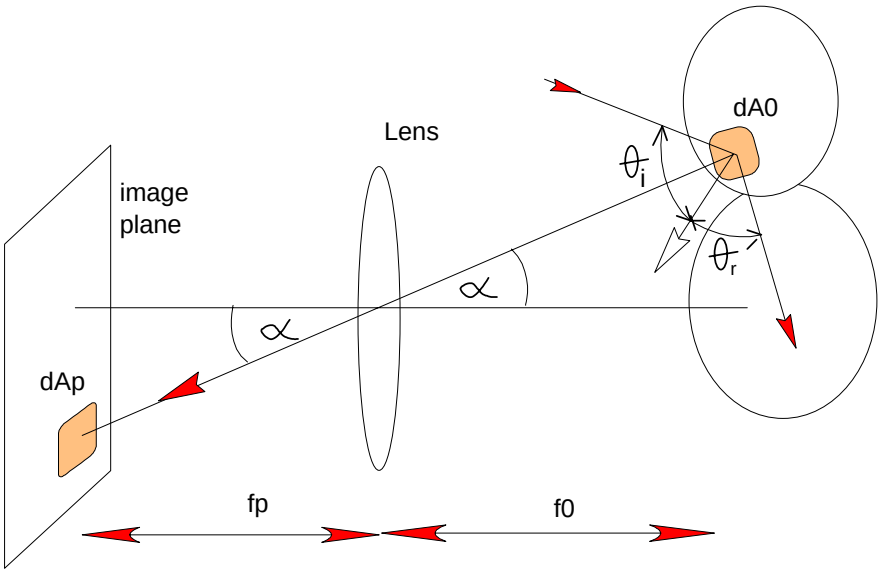

Figure A1: Scheme of the DRA-Camera.

***A2** Measurements of the image radiance*

The value of image irradiance on position *(x,y)* corresponds to a sample over a chosen wave-band depending on the particular interference filter used (Figure 2.3 shows a graph of the transmittance of the filters employed). This can be determined by integrating all contributions across that wave-band:

$$E_f(x,y) = a \cdot k(x,y) \int_{-\infty}^{\infty} s(\lambda) \cdot F_f(\lambda) \cdot L(x,y,\lambda) \cdot d\lambda$$

The suffix *f* above denotes the particular filter in use.

Given that we want to estimate the value of the image radiance L_f from the measured image irradiance E_f, we need to consider the way L_f is being sampled. This can be approximated as the mean value of $L(\lambda)$ within the wave-band considered. Considering $L(\lambda)$ as constant over the wave-band simplifies the above equation:

$$E_f(x,y) = a \cdot k(x,y) \cdot L_f(x,y) \cdot T_f; \qquad \text{(Eq. A1)}$$

where $T_f = \int_{-\infty}^{\infty} s(\lambda) \cdot F_f(\lambda) \cdot d\lambda$ is called *wave-band transmittance* and is regarded as constant although it may vary with temperature. The values of T_f vary from filter to filter. These can be estimated from both $s(\lambda)$ and $F_f(\lambda)$. The former one is taken from supplier's information and the second one is measured using the TopCon Spectro-

radiometer (as referred in Chapter 2).

A3 *Discrete sampling of* E_f

The equations above were assumed to hold for an infinitesimal point *(x,y)* on the image. In practice, the image is tessellated by a rectangular grid of rectangular pixels. The signal is integrated over the area of the pixel. All the assumptions that lead us to the previous equation are still valid, except that the position *(x,y)* is now referred to as a discrete point on the pixel grid and $E_f(x,y)$ is an spatially averaged signal.

A4 *Temporal sampling of* E_f

The sampling referred above also occurs over a given integer number of video frames (this period is also called *integration time t*). As discussed in Chapter 2, the integration time is decided for each narrowband image in order to allow a good dynamic range to be obtained.

The *grey level* (g_f) output of the camera was found not to be directly proportional to the integration time *t*. Experiments were performed in order to establish the relationship between the grey-level output, the incident irradiance and the integration time. This is referred as *time averaged value of* E_f and denoted $\overline{E}_f(g_f,t)$.

The calibration strategy to obtain the form of the function $\overline{E}_f(g_f,t)$ is described in Chapter 2. The value of g_f was measured for given values of $\overline{E}_f$ and *t.* Constant values of $\overline{E}_f$ were achieved using a constant current illuminant and a variety of aperture settings. From Eq. A1 we can derive absolute values of $\overline{E}_f$ by considering each term in turn:

-L_f: *scene radiance* is provided by a white card illuminated by constant current illuminant. It value was measured using a TopCon Spectro-radiometer.

-*a: aperture transmittance* is controlled to take on a range of known values. To obtain values that fall between the aperture f-stops is by achieved using neutral density filters.

-*k(x,y)* the optical sensitivity is assumed to be unit considering a fixed point near the optic axis. Spatial variations were mapped separately.

-T_f the transmittance is known as described above.

In practice our interest is limited to investigating the correlation between g_f, t and a as everything else is kept constant.

A5 *Obtaining spectral reflectance from $E_f(x,y)$*

Spectral reflectance is related to spectral radiance and illumination as follows:

$$I_f(x,y).R_f(x,y) = E_f(x,y)$$

in which $I_f(x,y)$ is the *illumination* of the image measured in W.m^{-2}, $R_f(x,y)$ is the *spectral reflectance* in a given direction and $E_f(x,y)$ is the *spectral radiance* in that direction. To simplify our problem we make the assumption that all reflection in the scene is approximately independent of the angle considered (Rayleghian). For our purposes, this is true if we avoid reflections from plants in short-distance images, etc. Another assumption that we make in order to obtain the value of reflectance is that $I_f(x,y)$ is fundamentally constant (independent of the position x,y) across all the image:

$$I_f \cdot R_f(x,y) = E_f(x,y)$$

This is not true in the case of strong shadowing over our scenes. This was avoided as much as possible, and diffuse lighting was used for scenes taken in the lab (indoor scenes). With these assumptions in mind and measuring the radiance of an object of known reflectance, we were able to obtain the illumination I_f of the scene:

$$I_f = \frac{E_f(x_0,y_0)}{R_f(x_0,y_0)}$$

The above was done using the standard Kodak grey card which was placed into the scene to that effect. Using this value of known illumination we were able to obtain the spectral reflectance for all points within the image using:

$$R_f(x,y) = \frac{E_f(x,y)}{I_f}$$

Changes in the illumination during the recording of the scene could introduce undesirable variations in our estimations of spectral radiance. For example, suppose that during the recording of the scene X there was a substantial change of the lighting when filter f_0 was added in front of the lenses. Then the value of $E_{f_0}(x,y)$ is affected by a factor *c*:

$$E'_{f_0}(x,y) = c \cdot E_{f_0}(x,y)$$

The value of I_{f_0} (as calculated in the above section) is also affected by the same factor:

$$I_{f_0} = \frac{E'_{f_0}(x_0,y_0)}{R_{f_0}(x_0,y_0)} = \frac{c \cdot E_{f_0}(x_0,y_0)}{R_{f_0}(x_0,y_0)}$$

but the reflectance obtained from there remains unaffected:

$$R_{f_0}(x,y) = \frac{E'_{f_0}(x,y)}{I_{f_0}}$$

As mentioned in Chapter 2, there were external measurements of spectral radiance over the Kodak grey card for each scene of the dataset. They were taken before and after the recording of each scene in order to check whether there was a substantial change in the illumination (*linear light changes*, see Chapter 2) during it. From these measurements we were able to compare the spectral radiance of the grey card as measured with the DRA-camera and the TopCon SR1 and see whether or not there was any variation (*light fluctuations*, see Chapter 2) in the illumination during the recording. The measurements also provided a reliable sample of the spectral characteristics of the illumination of our scenes. "*Reilluminate*" is designed to

a) compute the spectral illumination $\tilde{I}_f$ of the scene from external measurements using the TopCon SR1.

b) obtain the spectral reflectance $R_f(x,y)$ of the scenes (unaffected by changes in the illumination) as described above.

c) "reilluminate" the scene using these values. The radiance is then corrected to compensate changes in the illumination: $E_f(x,y) = \tilde{I}_f \cdot R_f(x,y)$

The effects of *reilluminate* were tested in our image analysis. Images processed with this algorithm proved to be indistinguishable from the rest in terms of Fourier spectra.

Appendix B

Fourier amplitude spectra of natural scenes

Image analysis

Two-dimensional Fourier transform theory has played a key role in this work. These transformations had been applied by the image processing algorithm "FBANDS" (in either its *linear* or *logarithmic* versions). Here we emphasise the understanding of the "FBANDS" output and how it is related with previous research methods by Field (1987) and others.

B1 *Fourier transforms*

Consider a function *f(x,y)* where *(x,y)* are the co-ordinates of any point. If *f(x,y)* is continuous and integrable and *F(u,v)* is integrable, the following Fourier transforms pair exists:

$$F(u,v) = \iint_{\infty} f(x,y) \cdot e^{-j \cdot 2\pi (ux+vy) \cdot dx \cdot dy}$$

$$f(x,y) = \iint_{\infty} F(u,v) \cdot e^{j \cdot 2\pi (ux+vy) \cdot du \cdot dv}$$

where *u* and *v* are the frequency variables

The Fourier spectrum, phase and power spectrum, respectively are:

$$\left|F(u,v)\right| = \sqrt{R^2(u,v) + I^2(u,v)}$$

$$\phi(u,v) = \tan^{-1}\left[\frac{I(u,v)}{R(u,v)}\right]$$

$$P(u,v) = \left|F(u,v)\right|^2 = R^2(u,v) + I^2(u,v)$$

Having the above notation in mind, we can now introduce the *discrete* Fourier transform pair for the two-variable case:

$$F(u,v) = \frac{1}{MN} \cdot \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \cdot e^{-j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)}$$

for u = 0, 1, 2, 3, ..., M - 1, v = 0, 1, 2, 3, ..., N - 1

$$f(x,y) = \cdot \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u,v) \cdot e^{j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)}$$

for x = 0, 1, 2, 3, ..., M - 1 and y = 0, 1, 2, 3, ..., N - 1.

This is applied to sampled two-dimensional functions. For example, *f(x,y)* now represents samples of the function $f(x_0 + x\Delta x, y_0 + y\Delta y)$ for x = 0, 1, 2, 3,..., M - 1 and y = 0, 1, 2, 3,..., N - 1. The sampling is now in a two-dimensional grid with divisions of width Δx and Δy in the x and y axis. Similar comments apply to *F(u,v)* and the sampling increments in the spatial and frequency domains are related by:

$$\Delta u = \frac{1}{M\Delta x}, \text{ and } \Delta v = \frac{1}{N\Delta y}$$

The Fourier spectrum, phase, and power spectrum are also given by the same equations as in the continuous case. Unlike the continuous case, existence of the discrete Fourier transforms is of no concern, because *F(u,v)* always exists.

As seen in Chapter 2, each image of our dataset can be considered as a discrete two-dimensional function (in where *f(x,y)* can be either radiance or reflectance, depending of the case) of the spatial co-ordinates *x,y*. FBANDS was designed to produce a display in which luminance is proportional to the amplitude of $|F(u,v)|$. In this display (see the polar diagram on Figure B1) the distance from the centre $(u^2+v^2)^{1/2}$ typically corresponds to the spatial frequency considered and the orientation ϕ to the phase angle.

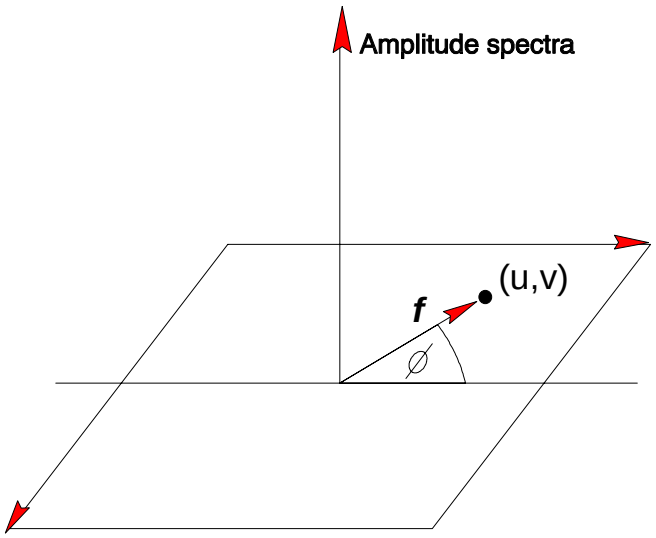

Figure B1: Polar display representing the function *E(f)* in the Fourier space. Typically it is a 256 x 256 digital display where the distance from the centre corresponds to the spatial frequency *f* and ϕ to the orientation.

***B2** Scale invariant images*

These special images (also called fractal-like images) have the particularity that their statistics do not change when viewing distance is increased or decreased. For example, if we measure the contrast within these images as the variance in pixel intensities, it should remain constant if we magnify any particular region of them. On average, the variance should be the same on magnifying any region.

This invariance in contrast can be related to the amplitude power by Parseval's theorem:

$$\frac{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left(f(x,y)\right)^2}{MN} = \sum_{u=0}^{M-1}\sum_{v=0}^{N-1}\left|F(u,v)\right|^2 ;$$

The left side of the above equation represents the variance (from mean =0) of the image and $|F(u,v)|^2$ is the power spectrum. The right side of the equation represents the total energy. If we consider that the variance remains constant independently of the scale, the total energy must also be independent of the scale. For consistency with the existing literature in the vision field, we use the expression *Fourier energy density* denoted as *g(f)* for the term $|F(u,v)|^2$.

For example, if we expect an image to be scale invariant, the total amount of energy

between frequency f_0 and frequency nf_0 must remain constant when we shift the range of frequencies to the range af_0 to anf_0 (this is equivalent to a magnification).

$$\int_{f_0}^{nf_0} g(f) \cdot 2\pi f \cdot df = \int_{af_0}^{anf_0} g(f) \cdot 2\pi f \cdot df = C$$

For this to occur, *g(f)* must be proportional to f^{-2}. Then the *amplitude spectra* must be proportional to f^{-1}.

From the above we conclude that if the power spectrum falls off as f^{-2} then the image will have constant variance at all scales.

***B3** Image transforms*

Despite the complexity of natural images, they share some statistical features that distinguish them from random-dot patterns. Some of these features were explored by Field (1987, 1989) who analysed a set of 6 images in terms of their *Fourier amplitude spectra* and *power spectra*.

In his work Field obtained the value of amplitude spectra for his set of six digitalysed natural images and found some consistent statistics as follows:

a) When plotted in a two-dimensional diagram (as the one shown on Figure B1) *the amplitude spectra* show its greatest value at low spatial frequencies.

b) If amplitude spectra are averaged across all orientations the falloff is remarkably similar for all scenes. Field suggests that this falloff is roughly proportional to f^{α} ,where $\alpha = -1$. From this follows that the power spectrum $g(f) \propto f^{-2}$.

c) When the *amplitude spectra* are plotted in double logarithmic co-ordinates the graph is a straight line with slope -1. Figure B2 shows this value averaged across all orientations for the six images analysed by Field.

These kind of images are *scale invariant.* This means that we would expect the relative contrast energy to be independent of the viewing distance.

Another study in the same field by Tolhurst *et al.* (1992) revealed different results. They processed 135 digitised pictures and found that only a very restricted sample of them conform to the characteristics listed above. When plotted in double logarithmic co-

ordinates they are closer to a straight line with an average slope of -1.2, instead of the 1.0 previously suggested. Figure 5.3 in Chapter 5 shows the averaged amplitude spectra of four images from Tolhurst's dataset. Notice the concentration of the data on the right side of the plot. As explained before this could have led to an overrepresentation of that sector when estimating the slope of any straight line in the log-log plot.

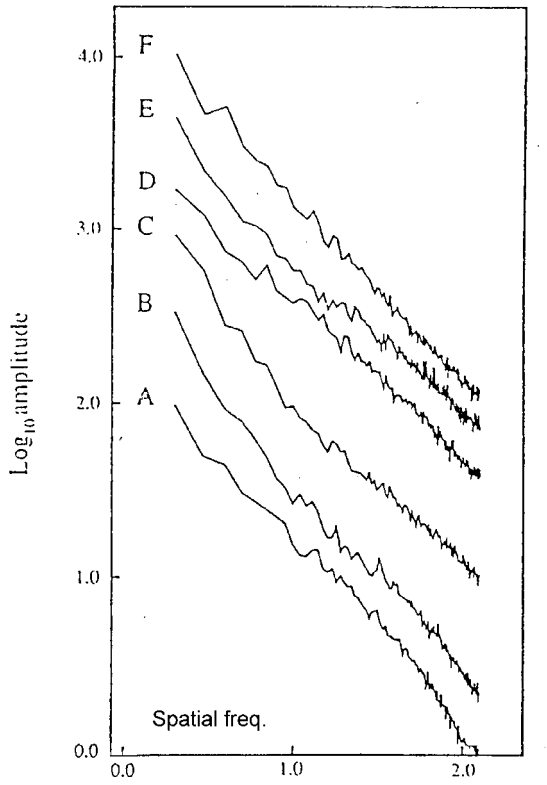

Figure B2: Fourier energy for six images averaged across orientations (from Field 1987)

References

Aho A. C., Donner K., Hyden C., Larsen L.O. and Reuter T. 1988. *Low retinal noise in animals with low body temperature allows high visual sensitivity*. Nature Lond. 334, pp. 348-350.

Atick J. J. 1992. *Could information theory provide an ecological theory of sensory processing.* Network 3, pp. 213-251.

Barlow H. B. 1957. *Purkinje shift and retinal noise.* Nature Lond. 179, pp. 225-256.

Barlow H. B. 1988. *The thermal limit to seeing.* Nature Lond. 334, p. 296.

Bennet A. T. D. and Cuthill I. C. 1994. *Ultra violet vision in birds: what is its function?*. Vision Research, Vol. 34, N° 11, pp. 1471-1478.

Berry M. V. and Wilson A. N. 1994. *Black-and-white fringes and the colours of caustics* Applied Optics. Vol. 33, N° 21, pp. 4714-4964.

Blakemore C. and Campbell F.W. 1969. *On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images.* Journal Physiol. Lond. 203, pp. 237-270.

Bradley A., Switkes E, De Valois K.K. 1985. *Orientation and spatial frequency selectivity of adaptation to isoluminant color patterns.* Invest. Ophthalmol. Vis. Sci. (Suppl.). 26, pp. 182.

Brelstaff G. and Troscianko T. 1992. *Information content of natural scenes: implications for neural coding of colour and luminance.* S.P.I.E. Vol.1666, Human Visual Processing and Digital Display, pp. 302 - 309.

Buchsbaum G. and Gottschalk A. 1983. *Trichromacy, opponent colour coding and optimum colour information transmission in the retina.* Proc. R. Soc. Lond. B220, pp. 80-113.

Burton G. J. and Moorhead I. R. 1987. *Color and spatial structure in natural scenes,* Applied Optics Vol. 26, pp. 157-170.

Campenhausen C.. *Photoreceptors, lightness constancy and colour vision.* 1986. Naturwissen. 73, pp. 674-675.

De Valois R. L., Thorell L. G. and Albretch D. G. 1982. *Spatial frequency selectivity of cells in macaque visual cortex.* Vis. Res., 22, pp. 545-599.

De Valois R. S., De Valois K. K. 1990. *Spatial Vision.* Oxford Psychology Series N° 14, Oxford Science Publications.

Derrico J. B. and Buchsbaum G. 1991. *A computational model of spatiochromatic image coding in early vision.* Journal of Visual Communication and Image Representation, 2, pp. 31-37.

Derrington A. M. and Lennie P. 1984. *Spatial and temporal contrast sensitivities of neurons in lateral geniculate nucleous of macaque.* J. Physiol. 357, pp. 219-240.

Derrington A. M., Krauskopf J. and Lennie P. 1984a. *Chromatic Mechanisms in lateral geniculate nucleous of macaque.* J. Physiol. 357, pp. 241-265.

Dunn, O.J. and Clark V.A. 1987. *Applied statistics: analysis of variance and regression*. 2nd edition- Wiley series in probability and mathematical statistics. pp. 54-60.

Field D. J. 1987. *Relation between the statistics of natural images and the response properties of cortical cells.* J. Opt. Soc. Am., Vol.4, N°. 12, pp. 2379-2394.

Field D. J. 1989. *What the statistics of natural images tell us about visual coding.* S.P.I.E. 1077, pp. 269-276.

Gabor D. 1946. *Theory of communication.* J. Inst. Electr. Eng., 93, pp. 429-457.

Gouras P. 1991. *Cortical mechanisms of colour vision.* Vision and visual disfunction (Vol. 6). ed. by P. Gouras, General Editor: J. R. Cronly-Dillon. The Macmillan Press Ltd. pp. 179-197.

Gouras P. 1991a. *Precortical physiology of colour vision.* Vision and visual disfunction (Vol. 6). ed. by P. Gouras, General Editor: J. R. Cronly-Dillon. The Macmillan Press Ltd. pp. 163-178.

Hurvich L. M. and Jameson D. 1957. *An opponent process theory of colour vision.* Phychol. Rev., 64, pp. 384-404.

Ingling Jr. C. R. and Martinez E. 1983. *The spatiochromatic signal of the r-g channel.* Colour Vision: Physiology and Psychophysics. ed. by J. D. Mollon & L. T. Sharpe, pp. 433-444.

Ingling Jr. C. R. and Tsou B.H. 1988. *Spectral sensitivity for flicker and acuity criteria.* J. Opt. Soc. Am. A, 5, 1374-1378.

King-Smith P. E. 1991. *Chromatic and achromatic visual systems.* Vision and visual disfunction (Vol. 6). ed. by P. Gouras, General Editor: J. R. Cronly-Dillon. The Macmillan Press Ltd. pp. 23-42.

Kirk J. T. O. 1983. *Light and photosynthesis in aquatic ecosystems.* Cambridge. Cambridge University Press. Fig 8.7.

Le Grand Y. 1967. *Form and space vision.* Translated by Michel Millodot and Gordon G. Heat. Indiana Univ. Press, Bloomington and London.

Le Grand Y. 1968. *Light, colour and vision.* (2nd edition). Translated by R.W.G. Hunt, J. W. T. Walsh and F. R. W. Hunt Somerset. N. J. Halsted Press.

Lennie P. and D'Zmura M. 1988. *Mechanisms of colour vision.* CRC Critical Reviews in Neurobiology, Vol. 3, 4, pp. 333-400.

Lythgoe J. N. 1991. *Evolution of visual behaviour.* Vision and visual disfunction (Vol. 2). ed. by J. R. Cronly-Dillon and R. Gregory. The Macmillan Press Ltd. pp. 3-14.

Mullen K. T. 1985. *Contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings.* J. Physiol. 359, 381-400.

Neumeyer C. 1991. *Evolution of colour vision.* Vision and visual disfunction (Vol. 2). ed. by J. R. Cronly-Dillon and R. Gregory. The Macmillan Press Ltd. pp. 284-305.

Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P. 1992. *Numerical recipes in C.* Second Edition, Cambridge University Press, pp. 656-681.

Shannon C. E. and Weaver W. 1949. *The mathematical theory of communication,* (U. Illinois Press, Champaign, Ill.).

Smith V. C. and Pokorny J. 1972. *Spectral sensitivity of color-blind observers and the cone photopigments* Vision Res. Vol. 12, pp. 2059-2071.

Smith V. C. and Pokorny J. 1975. *Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm.* Vision Res. Vol. 15, pp. 161-171.

Snyder A. W. 1977. *Acuity of compound eyes, physical limitations and design.* J. Comp. Physiol. 116, pp. 161-182.

Tadmor Y. and Tolhurst D. J. 1990. *Is the human visual system optimised for processing natural images.* Perception 19, p. 398 A.

Thorell L. G., De Valois R.L., Albretch D. G. 1984. *Spatial mapping of monkey VI cells with pure colour and luminance stimuli.* Vision. Res., Vol. 24, pp. 751-769.

Tolhurst D. J. Tadmor Y. and Tang Chao. 1992. *Amplitude spectra of natural images.* Ophthal. Physiol. Opt.,Vol.12, pp. 229-232.

Walker J. 1977. *The flying circus of physics with answers.* Wiley, New York, 1977, Article 5.129.