Notes for the course of Color Digital Image Processing

Edoardo Provenzi

Preface

Color is a very difficult yet fascinating perceptual feature to analyze. Color is inherently interdisciplinary since it is induced by a very complex chain of mechanisms:

1. A source of visible electromagnetic energy generates a ray of light that reaches our eyes either directly or after the interaction (diffusion, transmission or reflection) with another medium;

2. The light passes through the eyes and it is transduced into electric impulses by the retinal cells (photoreceptors: cones and rods);

3. The optical nerve sends this signal to the brain through complicated (and not yet totally understood) neural mechanisms;

4. The brain interprets the signal carried by the neurons and give us color perception.

So, in order to have a proper comprehension of the models and tools of color image processing, we must first present at least a concise and necessarily simplified overview on the four items listed above. For this reason, these notes will be structured as follows:

- Chapter 1 will deal with the *nature of light*: how it can be produced and it interacts with matter before reaching our eyes;

- Chapter 2 will introduce the *interaction between light and matter*;

- Chapter 3 will present the *eyes physiology and the photometric units*;

- Chapter 4 will concentrate of the *retina*;

- Chapter 5 will deal introduce *CIE colorimetry*, the standard way to represent and measure color;

- Chapter 6 will give a very concise introduction to *neuron dynamics*, higher *brain* features and *contrast perception*;

- Chapter 7 will discuss *histograms* of color images and modern (variational) ways to equalize them, mathematically and perceptually;

- Chapter 8 will introduce *high dynamic range (HDR) images* and the related problem of *tone mapping*.

The author.

Contents

Chapter 1

The nature of light and radiometry

1.1   What is light?

The nature of light has always been a debated topic throughout the history of science because it has the puzzling property of behaving like a wave in some circumstances and like a particle in others. This *duality* can be understood within the framework of quantum field theory.

For the purpose of this course it is not necessary to enter in the complicated models of quantum physics, it is enough to say that the light is a particular electromagnetic wave[1] produced by a non static electric or magnetic field: when an electric or magnetic field varies in time, it induces a non static magnetic or electric field, respectively, so that the process propagates to form a wave in which the electric field, the magnetic field and the direction of propagation of the wave are orthogonal, as depicted in Figure 1.1. The energy of an electromagnetic wave is not continuous, but is distributed in quanta of energy called *photons*.

Very important attributes of waves are:

- The **wavelength** λ, which is the Euclidean distance between two consecutive points of the wave characterized by the same intensity and the same first derivative (e.g. two consecutive maxima or minima), it is measured in *meters* (m) and its prefixes, which are recalled in Figure

---

[1]The electromagnetic theory was developed by James Clerk Maxwell (British physicist 1831-1879).
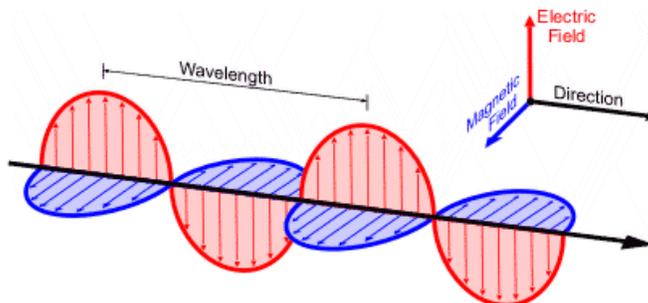
Figure 1.1: An electromagnetic wave.

1.2. A **cycle** is a complete vibration of a wave between two arbitrary points separated by a distance given by the wavelength.

- The **period** T which is the time needed by the electromagnetic wave to perform one complete cycle and it is measured in *seconds* (s);

- The **frequency** ν, which is the inverse of the period: $\nu = \frac{1}{T}$ and it is measured in *hertz* (Hz);

- The **speed of propagation** of the wave, measured in m/s and indicated in general with v. Mechanical waves need a medium to propagate, but electromagnetic wave can also travel in vacuum, in which case the speed is denoted with c and it is a fundamental constant of nature: $c = 299.792, 458\,km/s$.

The relationship among λ, ν and v is the following:

$$\boxed{\lambda = \frac{v}{\nu}}, \; \boxed{\lambda = \frac{v}{\nu}}.$$

Depending on the frequency or, equivalently, on the wavelength, an electromagnetic wave interacts with matter in very different ways, as can be seen in Figure 1.3. We call **light** the electromagnetic radiation that falls in the visible part of the whole electromagnetic spectrum, between 380 and 780 nm, where nm is a nanometer: $10^{-9}m$, or, equivalently, between 400 and 800 THz, where THz is a tera hertz: $10^{12}Hz$ (of course one must use the relation $\nu = c/\lambda$ to pass from the visible wavelength to the visible frequency).

We call **monochromatic light** a visible electromagnetic radiation characterized by the presence of only one wavelength (or frequency) and not by

7

Metric prefixes							
Prefix	Symbol	1000^m	10^n	Decimal	Short scale	Long scale	Since[n 1]
yotta	Y	1000^8	10^{24}	1 000 000 000 000 000 000 000 000	septillion	quadrillion	1991
zetta	Z	1000^7	10^{21}	1 000 000 000 000 000 000 000	sextillion	trilliard	1991
exa	E	1000^6	10^{18}	1 000 000 000 000 000 000	quintillion	trillion	1975
peta	P	1000^5	10^{15}	1 000 000 000 000 000	quadrillion	billiard	1975
tera	T	1000^4	10^{12}	1 000 000 000 000	trillion	billion	1960
giga	G	1000^3	10^9	1 000 000 000	billion	milliard	1960
mega	M	1000^2	10^6	1 000 000	million		1960
kilo	k	1000^1	10^3	1 000	thousand		1795
hecto	h	$1000^{2/3}$	10^2	100	hundred		1795
deca	da	$1000^{1/3}$	10^1	10	ten		1795
		1000^0	10^0	1	one		–
deci	d	$1000^{-1/3}$	10^{-1}	0.1	tenth		1795
centi	c	$1000^{-2/3}$	10^{-2}	0.01	hundredth		1795
milli	m	1000^{-1}	10^{-3}	0.001	thousandth		1795
micro	μ	1000^{-2}	10^{-6}	0.000 001	millionth		1960
nano	n	1000^{-3}	10^{-9}	0.000 000 001	billionth	milliardth	1960
pico	p	1000^{-4}	10^{-12}	0.000 000 000 001	trillionth	billionth	1960
femto	f	1000^{-5}	10^{-15}	0.000 000 000 000 001	quadrillionth	billiardth	1964
atto	a	1000^{-6}	10^{-18}	0.000 000 000 000 000 001	quintillionth	trillionth	1964
zepto	z	1000^{-7}	10^{-21}	0.000 000 000 000 000 000 001	sextillionth	trilliardth	1991
yocto	y	1000^{-8}	10^{-24}	0.000 000 000 000 000 000 000 001	septillionth	quadrillionth	1991

1. ^ The metric system was introduced in 1795 with six prefixes. The other dates relate to recognition by a resolution of the CGPM.

Figure 1.2: The prefixes used in the SI (International System of units).

the superposition of many waves with different wavelengths (or frequencies). If a monochromatic light reaches our eyes, it produces a color sensation that can be distinguished as in Figure 1.4.

By far, the greatest amount of light that arrives on the Earth is produced by the nuclear reactions within the core of our Sun.

The core starts from the center and extends outward to encompass 25 percent of the sun's radius. Its temperature is greater than 15 million degrees kelvin. At the core, gravity pulls all of the mass inward and creates an intense pressure. The pressure is high enough to force atoms of hydrogen to come together in nuclear fusion reactions (something we are trying to emulate here on Earth). Two atoms of hydrogen are combined to create
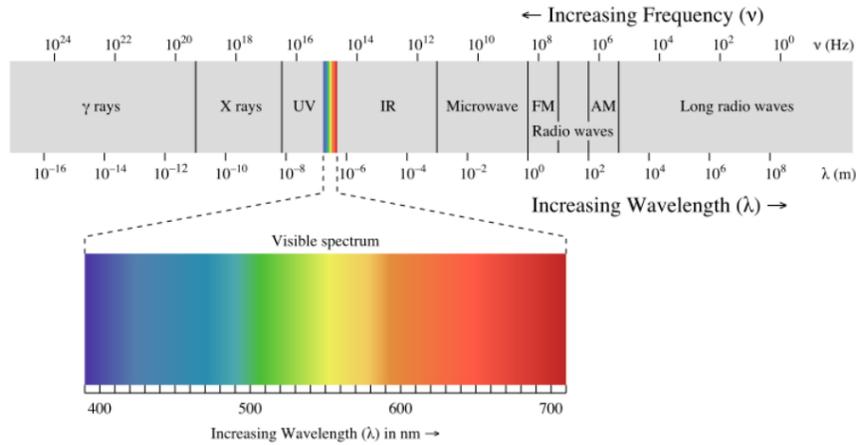
Figure 1.3: The electromagnetic spectrum.
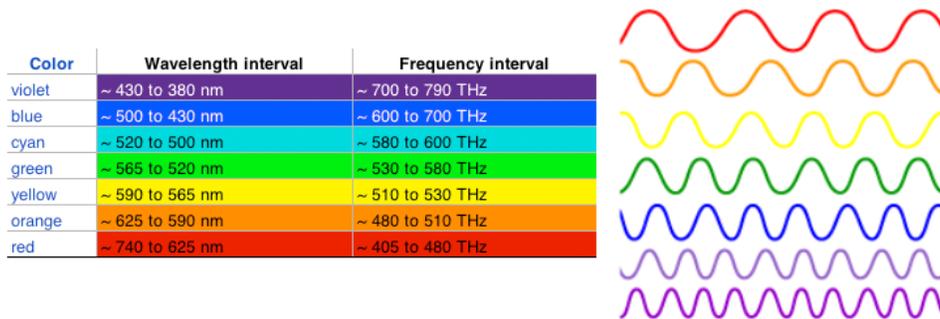
Figure 1.4: Monochromatic colors as a function of wavelength and frequency.

helium-4 and energy in several steps:

- Two protons combine to form a deuterium atom (hydrogen atom with one neutron and one proton), a positron (similar to electron, but with a positive charge) and a neutrino;

- A proton and a deuterium atom combine to form a helium-3 atom (two protons with one neutron) and a gamma ray;

- Two helium-3 atoms combine to form a helium-4 atom (two protons and two neutrons) and two protons.

These reactions account for 85 percent of the sun's energy. The remaining 15 percent comes from the following reactions:

9

- A helium-3 atom and a helium-4 atom combine to form a beryllium-7 (four protons and three neutrons) and a gamma ray;

- A beryllium-7 atom captures an electron to become lithium-7 atom (three protons and four neutrons) and a neutrino;

- The lithium-7 combines with a proton to form two helium-4 atoms.

The *helium-4 atoms are less massive than the two hydrogen atoms that started the process*, so the *difference in mass is converted to energy* as described by Einstein's theory of relativity $E = mc^2$, being m the mass difference, E the energy produced in the reaction and c the speed of light (see later for its value). The energy is emitted in various forms of light: ultraviolet light, X-rays, visible light, infrared, microwaves and radio waves.

The sun also emits energized particles (neutrinos, protons) that make up the solar wind. This energy strikes Earth, where it warms the planet, drives our weather and provides energy for life. We aren't harmed by most of the radiation or solar wind because the Earth's atmosphere protects us.

In fact, the atmosphere strongly reduces the amount of dangerous and highly energetic ultraviolet (UV) and gamma rays, so that the majority of Sun radiation that reaches the surface of our planet penetrating the atmosphere is within the radio and visible part (**sunlight**) of the spectrum, with a small amount of microwave and infrared (IR) waves.

The sunlight is composed with all the wavelengths between 380 and 780 nm, but not in a uniform way: as can be seen in Figure 1.5, the maximum intensity of the sunlight with respect to the wavelength is reached in the 'green' region, around 510 nm. Sources of light like the Sun, i.e. that emit light throughout the visible spectrum are called **heterochromatic** or **equienergetic**, even if this last adjective is quite misleading because it could give the impression of a uniformly distributed spectrum of intensities with respect to frequencies.

To each monochromatic electromagnetic wave of frequency ν we can associate a photon that vibrates with the same frequency. The energy carried by the photon is given by **Planck's formula**:

$$\boxed{E = h\nu}$$

where h is the universal constant of Planck: $h = 6,628 \cdot 10^{-34} Js$, where J is the unit of measure of energy, the *joule*.

The greater the frequency of an electromagnetic wave, the more intense is its energy, so radio waves have very low energy and UV and gamma rays
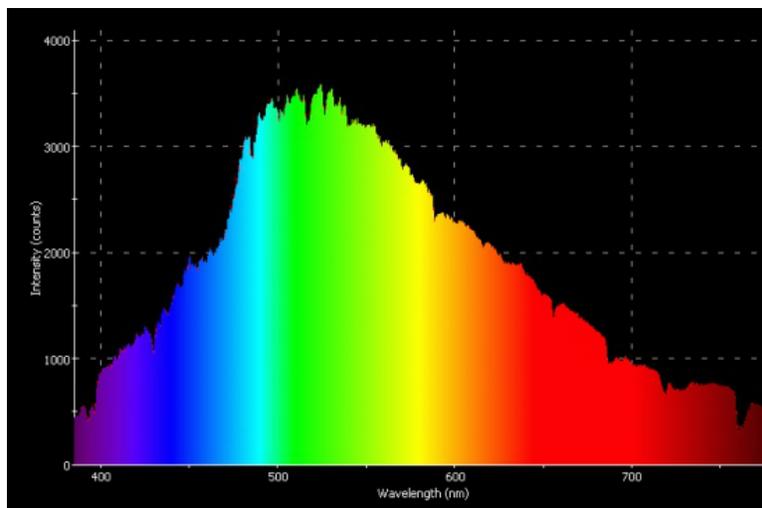
Figure 1.5: The intensity of sunlight relative to the wavelength.

are very energetic, for this reason they are dangerous for human beings and living creatures in general.

Infrared radiation is often improperly called *thermal radiation*. Thermal radiation is electromagnetic radiation produced by everything that has a temperature greater than the absolute zero (0 K (*kelvin*) = -273,15 ^{o}C), it is produced by the movement and interaction of charged particles and molecules. If the body temperature doesn't reach the incandescence level, its thermal radiation falls typically in the infrared region, that is the reason why the two radiations are often identified.

1.2   Radiometric units

The SI, International System of units, has defined a set radiometric units that are very useful when we deal with light. Table 1.6 recaps these quantities and their meaning. The suffix e stays for *energetic*, in contrast to the suffix V, which stays for *visual*, that will be used to define the *photometric units*. Radiometric units are defined without taking into account the interaction with the human eyes, which instead is considered to define the photometric ones, as we will see later in section 4.2.

A particularly important unit is **radiance** L_e, the power of an electromagnetic wave per unit of solid angle and surface orthogonal to the light direction, it is measured in $\frac{W}{sr\,m^2}$, sr being the steradian (dimensionless), the

SI radiometry units

Quantity	Symbol	SI unit	Symbol	Dimension	Notes
Radiant energy	Q_e	joule	J	$M \cdot L^2 \cdot T^{-2}$	energy
Radiant flux	Φ_e	watt	W	$M \cdot L^2 \cdot T^{-3}$	radiant energy per unit time, also called *radiant power*.
Spectral power	$\Phi_{e\lambda}$	watt per metre	$W \cdot m^{-1}$	$M \cdot L \cdot T^{-3}$	radiant power per wavelength.
Radiant intensity	I_e	watt per steradian	$W \cdot sr^{-1}$	$M \cdot L^2 \cdot T^{-3}$	power per unit solid angle.
Spectral intensity	$I_{e\lambda}$	watt per steradian per metre	$W \cdot sr^{-1} \cdot m^{-1}$	$M \cdot L \cdot T^{-3}$	radiant intensity per wavelength.
Radiance	L_e	watt per steradian per square metre	$W \cdot sr^{-1} \cdot m^{-2}$	$M \cdot T^{-3}$	power per unit solid angle per unit *projected* source area. confusingly called "intensity" in some other fields of study.
Spectral radiance	$L_{e\lambda}$ or L_{ev}	watt per steradian per metre³ or watt per steradian per square metre per hertz	$W \cdot sr^{-1} \cdot m^{-3}$ or $W \cdot sr^{-1} \cdot m^{-2} \cdot Hz^{-1}$	$M \cdot L^{-1} \cdot T^{-3}$ or $M \cdot T^{-2}$	commonly measured in $W \cdot sr^{-1} \cdot m^{-2} \cdot nm^{-1}$ with surface area and either wavelength or frequency.
Irradiance	E_e	watt per square metre	$W \cdot m^{-2}$	$M \cdot T^{-3}$	power incident on a surface, also called *radiant flux density*. sometimes confusingly called "intensity" as well.
Spectral irradiance	$E_{e\lambda}$ or E_{ev}	watt per metre³ or watt per square metre per hertz	$W \cdot m^{-3}$ or $W \cdot m^{-2} \cdot Hz^{-1}$	$M \cdot L^{-1} \cdot T^{-3}$ or $M \cdot T^{-2}$	commonly measured in $W \cdot m^{-2} \cdot nm^{-1}$ or $10^{-22} W \cdot m^{-2} \cdot Hz^{-1}$, known as solar flux unit.[nb 5]
Radiant exitance / Radiant emittance	M_e	watt per square metre	$W \cdot m^{-2}$	$M \cdot T^{-3}$	power emitted from a surface.
Spectral radiant exitance / Spectral radiant emittance	$M_{e\lambda}$ or M_{ev}	watt per metre³ or watt per square metre per hertz	$W \cdot m^{-3}$ or $W \cdot m^{-2} \cdot Hz^{-1}$	$M \cdot L^{-1} \cdot T^{-3}$ or $M \cdot T^{-2}$	power emitted from a surface per wavelength or frequency.
Radiosity	J_e or $J_{e\lambda}$	watt per square metre	$W \cdot m^{-2}$	$M \cdot T^{-3}$	emitted plus reflected power leaving a surface.
Radiant exposure	H_e	joule per square metre	$J \cdot m^{-2}$	$M \cdot T^{-2}$	
Radiant energy density	ω_e	joule per metre³	$J \cdot m^{-3}$	$M \cdot L^{-1} \cdot T^{-2}$	

Figure 1.6: SI Radiometric units.

unit of measure of solid angle. Let us recall that a **solid angle** is a portion of the three-dimensional Euclidean space included between two intersecting planes. A solid angle of 1 *sr* is such that the spherical sector defined by the solid angle divided by the square of the sphere radius is 1, as in Figure 1.7. Since the area of the spherical surface is $4\pi r^2$, we have that the total solid angle is $\frac{4\pi r^2}{r^2} = 4\pi$.
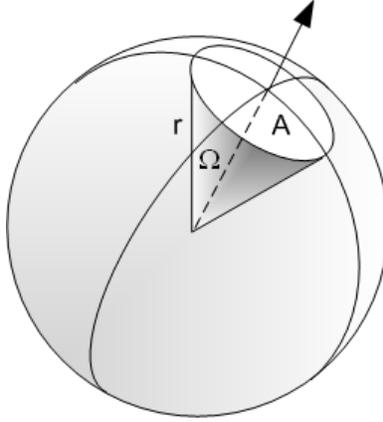
Figure 1.7: Solid angle geometry.

The radiance is usually defined starting from an infinitesimal value:

$$dL_e = \frac{d^2\Phi_e}{dA\cos\theta d\Omega}$$

where the quantities involved in the definition are defined in Figure 1.8. The infinitesimal radiance is defined as the ratio between the power of the light (**optical power**) incident on a surface A through an infinitesimal solid angle $d\Omega$, divided by $d\Omega$ itself and the projected infinitesimal area dA_p (where the projection is meant in the direction orthogonal to the ray of light), i.e. $dA_p = dA\cos\theta$. The use of the second derivative symbol d^2L_e is just a formal consequence of Leibnitz's notation and some authors write $d\Phi_e$ instead of $d^2\Phi_e$.

The total radiance L_e is obtained by integrating dL_e on the projected section A_p and into the solid angle Ω.

The *importance of radiance* is that it is invariant in geometric optics. This means that for an ideal optical system in air, **the radiance at the receiver is the same as the source of the radiation**[2]. This is sometimes called **conservation of radiance** and it depends on the fact that the projected area and the solid angle compensate each other. As an example, if we form a demagnified image with a lens, the optical power is concentrated into a smaller area, so the irradiance E_e is higher at the image. The light

---

[2]For real, passive, optical systems, the receiver radiance is *at most* equal to the source radiance.
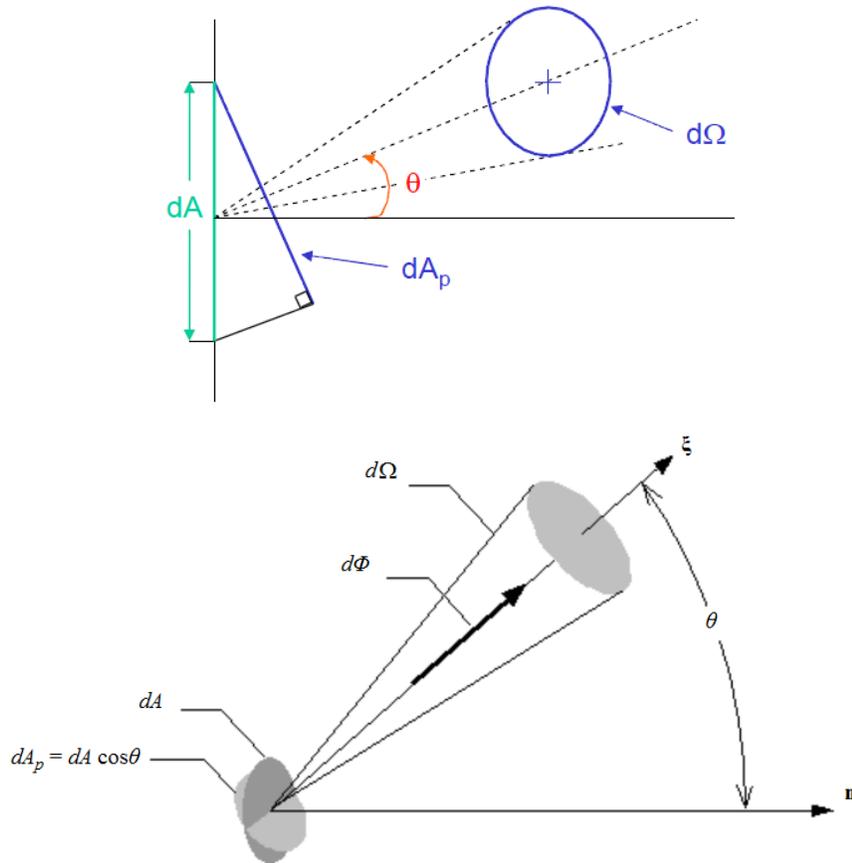
Fig. 1. Geometry of the radiance definition: $d\Phi$ - radiant flux which propagates within an elementary solid angle, $d\Omega$ about the direction of observation, ξ; dA - elementary area of the radiation source with its normal, **n**; $dA_p = dA \cos\theta$ - projection of the elementary area onto a plane perpendicular to the direction of observation, ξ.

Figure 1.8: Geometry of radiance definition 2D (above), 3D (below).

at the image plane, however, fills a larger solid angle so the radiance comes out to be the same assuming there is no loss at the lens.

As we will see later, the acquisition of light through the eyes can be modeled as a geometric optical system, so that the radiance of a light source is the same radiance that hits our eyes.

On the contrary, the irradiance of en electromagnetic radiation[3] E_e **de-**

---

[3]The symbol E_e is used because the irradiance is proportional to $|\vec{E}|^2$, the square of

cays as the square of the distance to the source:

$$E_e = L_e \frac{A}{d^2} \tag{1.1}$$

A being the area of the electromagnetic source and d being the distance to the receiver.

1.3   Black body radiation

For later purposes, namely defining the so-called *correlated color temperature* (CCT), it is necessary to introduce the concept of black body radiation. A black body is an idealized physical body that absorbs all incident electromagnetic radiation, regardless of frequency or angle of incidence[4]. A black body in thermal equilibrium, i.e. at a constant temperature, emits electromagnetic radiation called **black body radiation**. The radiant intensity is emitted according to Planck's law[5]:

$$I(\nu) \propto \frac{1}{e^{\frac{h\nu}{kT}} - 1}$$

where $\propto$ means proportionality up to suitable constants, k is the universal Boltzmann constant $1.38 \cdot 10^{-23} J/K$, ν is the frequency of the radiation and T is the temperature of the black body.

It can be seen that the spectrum of radiant intensity of a black body is determined solely by its temperature and not by the body's shape or composition. Figure 1.9 shows the radiant intensity spectrum emitted by a black body depending on its temperature.

Experimentally, it can be seen that the wavelength corresponding to the maxima of each curves has an inverse proportionality with respect to the temperature, this is called **Wien's law**:

$$\lambda_{\max} = \frac{b}{T},$$

---

the modulus of the electric field.

[4]In opposition to a black body one defines a **white body** as a body whose surface reflects all incident rays completely and uniformly in all directions.

[5]Max Planck (German physicist 1858-1947) received the Nobel prize in Physics in 1918 for his work on quantum physics, notably for postulating that the energy of electromagnetic radiation is packed in quanta of energy $E = h\nu$, thanks to which he was able to predict with great precision the black body radiative spectrum, that is impossible with Maxwell's classical electromagnetic theory.
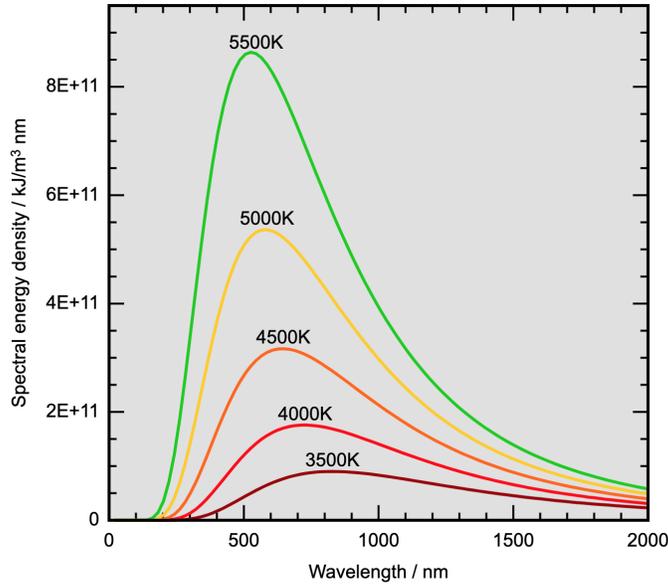
Figure 1.9: The radiant intensity spectrum emitted by a black body depending on its temperature.

b is called Wien's displacement constant: $b = 2.8977685 \cdot 10^{-3} mK$. Of course, if we considered $\nu_{max} = \frac{c}{\lambda_{\max}}$ we would find a linear relationship with the temperature: $\nu_{\max} = aT$, $a = \frac{1}{b}$.

An approximate *realization of a black body* surface is a hole in the wall of a cavity: any light entering the hole is reflected indefinitely or absorbed inside and is unlikely to re-emerge, making the hole a nearly perfect absorber, see Figure 1.10.
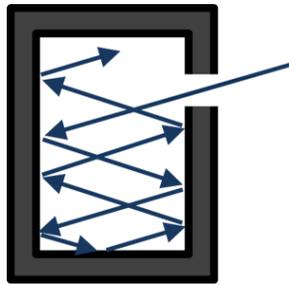
Figure 1.10: Realization of a black body as a cavity with a hole.

16

Stars are approximate black bodies: Figure 1.11 shows a highly schematic cross-section of a star to illustrate the idea. The photosphere of the star where the emitted light is generated is idealized as a layer within which the photons of light interact with the material in the photosphere and achieve a common temperature T that is maintained over a long period of time. Some photons escape and are emitted into space, but the energy they carry away is replaced by energy from within the star, so that the temperature of the photosphere is nearly steady. Changes in the core of the star lead to changes in the supply of energy to the photosphere, but such changes are slow on the time scale of interest here. Assuming these circumstances can be realized, the outer layer of the star is somewhat analogous to the example of a cavity with a small hole in it, with *the hole replaced by the limited transmission into space at the outside of the photosphere*. With all these assumptions in place, the star emits black-body radiation at the temperature T of the photosphere. Using this model the **effective temperature of stars** is estimated, defined as the temperature of a black body that yields the same surface flux of energy as the star. If a star were a black body, the same effective temperature would result from any region of the spectrum. The *Sun has an effective temperature of* **5780 K**, which can be compared to the temperature of the photosphere of the Sun (the region generating the visible light), that is approximatively 5000 K at the outer boundary.
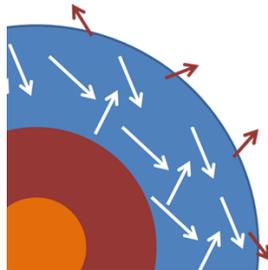
Figure 1.11: Very schematic idealization of the cross-section of a star.

Roughly speaking, nuclear physics predicts that young stars have $\lambda_{\mathrm{m}ax}$ in the region of blue, while old stars have $\lambda_{\mathrm{m}ax}$ in the region of red, this explains why stars appear differently colored in the sky.

Chapter 2

Interaction between light and matter

2.1 Introduction

Non self-luminous objects are those that don't emit autonomously light, so that they can be seen only if they are illuminated. We can distinguish among three types of non self-luminous objects:

1. **Transparent**: those bodies that light can pass right through without being diffused;

2. **Translucent**: semitransparent bodies that can transmit light diffusing it;

3. **Opaque**: they don't let light pass by, they reflect it and/or absorb it.

When opaque materials are hit by light, they partially reflect and partially absorb it. Reflected light permits us to view objects.

Sometimes the photons absorbed allow surface particles to change their energetic state to a higher *unstable* one, so that they decay to a lower energy level and emit a photon that always has a wavelength greater or equal to that of the incoming light (so the emitted photon always has lower or at most equal frequency than the one absorbed, and so is less energetic by virtue of Planck's formula). This phenomenon is generally called **photoluminescence**, in particular we talk about **fluorescence** if this phenomenon ceases immediately after the illumination and about **phosphorescence** if the phenomenon of light emission by the object continues for an appreciable

amount of time after the external excitation of the material. Reflection and fluorescence are, by far, more probable than phosphorescence.

2.2   Interaction between light and surfaces

The starting event that generates the phenomenon of vision is always the same: *a flux of photons entering the eyes.* However, this optical flux can reach our eyes in two different ways:

1. it can come directly from a source of light;

2. it can arrive to the eyes after being reflected by a surface.

In both cases there can be **scattering** (or **diffusion**) by a medium, typically the *air molecules*, that can interact with the photons of the optical flux modifying their direction of propagation.

We say that a light source, a reflecting surface or a scatterer are **Lambertian**[1] or **isotropic** if they emit, reflect or scatter, respectively, the same amount of radiant flux Φ_e in every direction, without privileging a particular one.

If we observe or measure the radiant intensity of a Lambert source, reflecting surface or scatterer at an angle θ with respect to the normal vector to the area of emission, reflection or scatter, respectively, then we find the following expression called **Lambert's cosine law**:

$$\boxed{I_e^\theta = I_e^0 \cos \theta}$$

where I_e^0 is the radiant intensity in the direction of the normal vector, as shown in Figure 2.1.

Lambertian reflection from surfaces is typically accompanied with specular reflection (**gloss**), where the surface luminance is highest when the observer is situated at the perfect reflection direction and falls off sharply. This is simulated in computer graphics with various specular reflection models such as Phong, Cook-Torrance.

A *black body* is an example of a *Lambertian radiator.* We know that black bodies are idealized objects and, in fact, a perfect Lambertian condition is never fulfilled in nature.

---

[1]Johann Heinrich Lambert (1728-1777), swiss mathematician, physicist, philosopher and astronomer. Besides his studies on light, he is best known for being the first one to prove the irrationality of the number π!
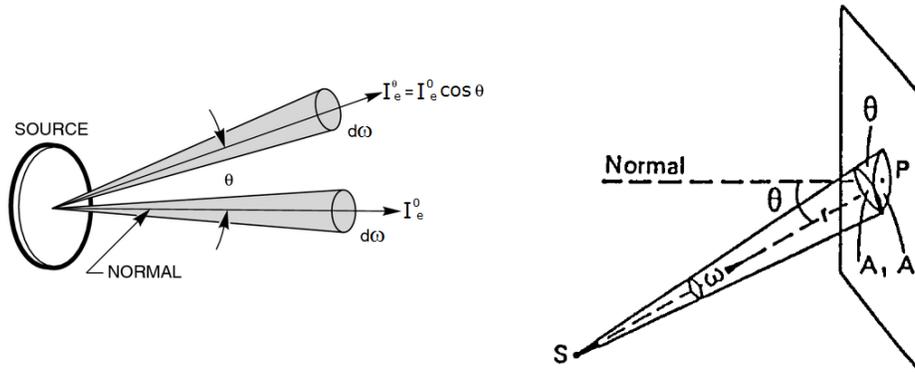
Figure 2.1: Lambert's law.

As an example, the *moon is not a Lambertian scatterer*. If it were so, one would expect to see its scattered brightness appreciably diminish towards its perimeter due to the increased angle at which sunlight hit the surface. The fact that it does not diminish illustrates that the moon is not a Lambertian scatterer, and in fact tends to scatter more light into the *oblique angles* than would a Lambertian scatterer.

The emission of a Lambertian radiator does not depend upon the amount of incident radiation, but rather from radiation originating in the emitting body itself. For example, if *the sun* were a Lambertian radiator, one would expect to see a constant brightness across the entire solar disc. The fact that the sun exhibits limb darkening in the visible region illustrates that it *is not a Lambertian radiator*.

The atmospheric scattering is not Lambertian: the so-called *Rayleigh scattering* between sunlight and the atmosphere particles is such that the spectral radiant intensity $I_{e,\lambda}$ of the sunlight that reaches the surface of the Earth is *proportional to* $1/\lambda^4$, so that the long (reddish) wavelengths are strongly attenuated by the Rayleigh scattering, in favor of the short (bluish) wavelengths and this influences the scattering of the optical flux in the different directions. During sunset and sunrise the angle between the sunlight and the Earth is such that there is only a small fraction of long wavelength in the sunlight, so that the resulting color is more reddish than during the rest of the day.

2.2.1 Reflectance and transmittance

Let us start this section by first recalling Snell's law. When a ray of light propagates from one isotropic transparent medium to another, the direction of propagation of light changes, as shown in Figure 2.2.
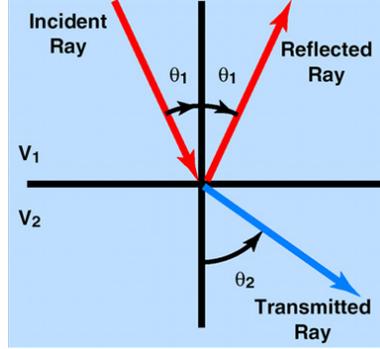
Figure 2.2: Schematization of Snell's law.

With respect to the angles θ_1, θ_2 made with the normal vector to the discontinuity surface we can express Snell's law as follows:

$$\frac{\sin \theta_1}{\sin \theta_2} = \frac{v_1}{v_2},$$

where v_i is the speed of propagation of the light in the i-th medium. If we define the **index of refraction** as $n = \frac{c}{v}$, i.e. the fraction between the the speed of light in vacuum and that in the medium under analysis, we have that, of course $n > 1$ for every material and Snell's law can be reformulated as follows:

$$\boxed{\frac{\sin \theta_1}{\sin \theta_2} = \frac{n_2}{n_1}}.$$

Let us now define two important parameters of a material:

- the **spectral reflectance** $\rho(\xi, \lambda)$ at the point ξ and the wavelength λ is the ratio between the reflected and incident optical flux in ξ and at the fixed wavelength λ:

$$\boxed{\rho(\xi, \lambda) = \frac{\Phi_{e,\lambda}^{\text{reflected}}(\xi)}{\Phi_{e,\lambda}^{\text{incident}}(\xi)}};$$

21

- the **spectral transmittance** $\tau(s, \lambda)$ of a material with thickness s and the wavelength λ is the ratio between the transmitted and incident optical flux at the fixed wavelength λ:

$$\tau(s, \lambda) = \frac{\Phi_{e,\lambda}^{\text{trasmitted}}(s)}{\Phi_{e,\lambda}^{\text{incident}}}.$$

In case of total reflection we have $\rho(\xi, \lambda) = 1$ and $\tau(s, \lambda) = 0$ for every λ.

Chapter 3

Physiology of the eyes and photometric units

The main reference for this chapter is M.D. Fairchild's '*Color appearance models*', Chapter 1 '*Human color vision*', John Wiley and Sons, Second Edition, 2005.

3.1 The human eye

The human eyes are an extraordinary product of evolution. First of all, since sight and sound are the most fundamental senses for humans, nature has placed eyes and ears on the head, so that the stimuli they catch from the outside world can arrive as quickly as possible to the brain.

Our visual perception is initiated and strongly influenced by the anatomical structure of the eye. Figure 3.1 shows a schematic representation of the optical structure of the human eye with some key features labeled. The human eye lies within a bulb of around 2.4 cm and acts as a powerful optical system of around 60 dioptrics. The cornea and lens act together like a camera lens to focus an image of the visual world on the retina at the back of the eye, which acts like the film or other image sensor of a camera. These and other structures have a significant impact on our perception of color.

Our visual field is:

- 90 degrees towards the temple;

- 60 degrees towards the nose;

- 70 degrees above;
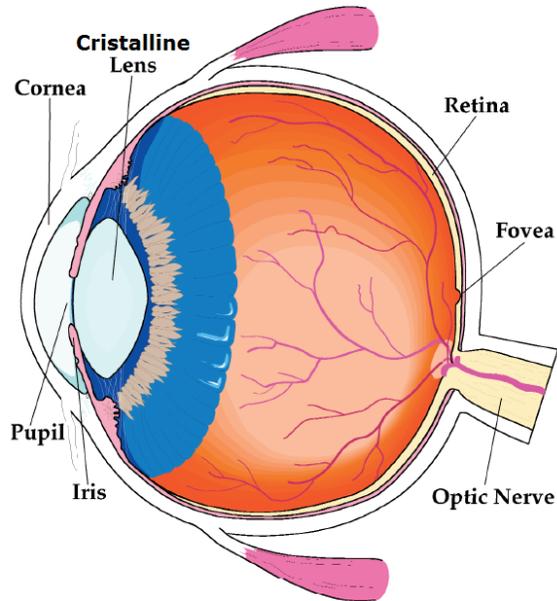
Figure 3.1: A simplified representation of the human eye.

- 80 degrees below.

The light that enters in our eyes, before reaching the photoreceptors of the retina, passes through the cornea, the aqueous humor, the pupil, the crystalline lens and the vitreous humor. In the next subsections we will briefly explain the behavior of all these parts. Finally, in the next chapter, we will analyze in more detail the film of our natural camera: the retina.

3.1.1 The cornea

The cornea is the transparent outer surface of the front of the eye through which light passes. It serves as the most significant image-forming element of the eye since its curved surface at the interface with air represents *the largest change in index of refraction in the eye's optical system*, by virtue of Snell's law this corresponds to a deviation of the rays of light.

The cornea is avascular, receiving its nutrients from marginal blood vessels and the fluids surrounding it. Refractive errors, such as nearsightedness (*myopia*), farsightedness (*hyperopia*), or *astigmatism*, can be attributed to *variations in the shape of the cornea* and are sometimes corrected with laser surgery that reshapes the cornea.

3.1.2 The lens

The crystalline *biconvex lens* serves the function of **accommodation**: it focuses the parallel rays of light coming from relatively distant objects to the retina, in order to have a sharp (*reduced and upsidedown!*) view.

However, as we approach an object to a small distance, the rays of light incoming from this object are not parallel and the lens focuses the rays of light *behind the retina*, so that we see a *blurred image*.

Whenever the image upon the retina becomes blurred, the brain automatically causes the *ciliary muscles* to contract or to relax until the crystalline lens has obtained just the right shape to restore a sharp and defined image upon the retina.

The analogy with a camera here doesn't work, because in a common camera the lens remains fixed, while the focus is moved (manually or automatically) to adapt to difference distance views. When we gaze at a *nearby object*, the lens becomes '*fatter*' and thus has increased optical power to allow us to focus on the near object. When we gaze at a *distant object*, the lens becomes '*flatter*' resulting in the decreased optical power required to bring far away objects into sharp focus.

As we *age*, the internal structure of the lens changes resulting in a *loss of flexibility*. Generally, when an age of about 50 years is reached the lens has completely lost its flexibility and observers can no longer focus on near objects (this is called *presbyopia*, or 'old eye'). It is at this point that most people need reading glasses or bifocals.

Concurrent with the hardening of the lens is an increase in its optical density. The lens absorbs and scatters short-wavelength (blue and violet) energy. As it hardens, the level of this absorption and scattering increases. In other words, *the crystalline lens becomes more and more yellow with age*. Various mechanisms of chromatic adaptation generally make us unaware of these gradual changes.

However, we are all looking at the world through a yellow filter that not only changes with age, but is significantly different from observer to observer. The effects are most noticeable when performing critical color matching or comparing color matches with other observers. The effect is particularly apparent with purple objects.

Since an older lens absorbs most of the blue energy reflected from a purple object, but does not affect the reflected red energy, older observers will tend to report that the object is significantly more red than reported by younger observers.

The crystalline is a layered, flexible structure that *varies in index of*

refraction. It is a naturally occurring gradient index optical element with the index of refraction higher in the center of the lens than at the edges. This feature serves *to reduce some of the aberrations* that might normally be present in a simple optical system.

Finally, let us comment the phenomenon of **chromatic aberration**: as can be seen in Figure 3.2, the long wavelengths are deflected by the lens more closely to the retina than the short wavelengths, this is why the r*ed appears closer to other colors*, this is one of the phenomena used to form a 3D illusion in cinema.
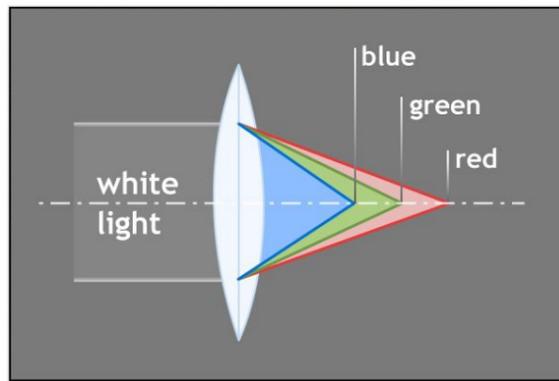
Figure 3.2: Chromatic aberration.

3.1.3   The humors

The volume between the cornea and lens is filled with the *aqueous humor*, which is essentially *water*. The region between the lens and retina is filled with *vitreous humor*, which is also a fluid, but with a higher viscosity, similar to that of *gelatin*.

Both humors exist in a state of slightly elevated pressure (relative to air pressure) to assure that the flexible eyeball retains its shape and dimensions in order to *avoid* the deleterious effects of *wavering retinal images*. The flexibility of the entire eyeball serves to increase its *resistance to injury*. It is much more difficult to break a structure that gives way under impact than one of equal 'strength' that attempts to remain rigid.

Since the indices of refraction of the humors are roughly equal to that of water, and those of the cornea and lens are only slightly higher, the rear surface of the cornea and the entire lens have relatively little optical power.

3.1.4   The iris

The iris is the sphincter *muscle that controls pupil size.* The iris is pigmented, giving each of us our specific eye color. *Eye color* is determined by the concentration and distribution of *melanin within the iris.*

The **pupil**, which is the hole in the middle of the iris through which light passes, defines the *level of illumination on the retina.* Pupil size is largely determined by the overall level of illumination, but it is important to note that it can also vary with non-visual phenomena such as arousal.

This effect can be observed by enticingly shaking a toy in front of a cat and paying attention to its pupils. Thus it is difficult to accurately predict pupil size from the prevailing illumination.

In practical situations, pupil diameter varies from about 3 mm to about 7 mm. This change in pupil diameter results in approximately a *fivefold change in pupil area*, and therefore retinal illuminance.

The change in pupil diameter alone is *not sufficient to explain excellent human visual function over prevailing illuminance levels* that can vary over 10 orders of magnitude.

3.1.5   The retina

The retina is actually *considered a part of the optical nerve* and not of the eye. The optical image formed by the eye is projected onto the retina.

The retina is a thin layer of cells, approximately the thickness of tissue paper, located at the back of the eye and incorporating the visual system's photosensitive cells and initial signal processing and transmission 'circuitry'. These cells are *neurons*, part of the central nervous system, and can appropriately be considered a part of the brain. The photoreceptors, rods and cones, serve *to transduce the information present in the optical image into chemical and electrical signals that can be transmitted to the later stages of the visual system.*

These signals are then processed by a network of cells and transmitted to the brain through the optic nerve. More detail on the retina will be presented in the next section.

Behind the retina is a layer known as the **pigmented epithelium**. This dark pigment layer serves to absorb any light that happens to pass through the retina without being absorbed by the photoreceptors. The function of the pigmented epithelium is to *prevent light from being scattered back* through the retina, thus *reducing sharpness and contrast* of the perceived image.

Nocturnal animals give up this improved image quality in exchange for a highly reflective tapetum that reflects the light back in order to provide a *second chance for the photoreceptors to absorb the energy.* This is why the eyes of a deer, or other nocturnal animal, caught in the headlights of an oncoming automobile, appear to glow!

3.1.6 The fovea

One of the most important structural area on the retina is the fovea. The fovea has a diameter of around 0.1 *mm* and it allows the best spatial and color vision for an angle of about **2 degrees** (called **foveal vision**).

When we look at, or fixate, an object in our visual field, we move our head and rotate our eyes so that the image of the object falls on the fovea.

We call **visual axis** the straight line that connects the center of the fovea with the point that we are looking at. Notice that this, in general, it differs from the eye's optical axis, as show in Figure 3.3.
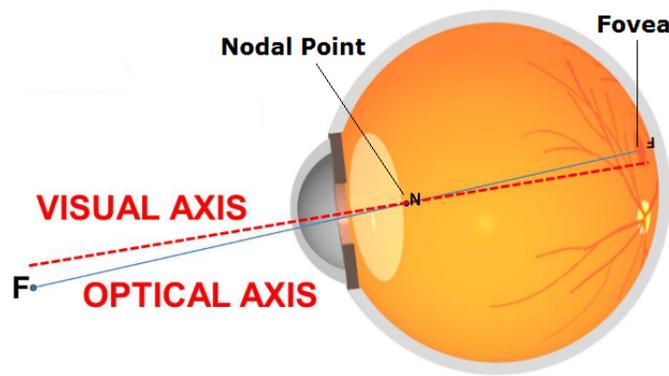
Figure 3.3: Visual vs. optical axis .

As you are reading this text, you are moving your eyes to make the various words fall on your fovea as you read them. To illustrate how drastically spatial acuity falls off as the stimulus moves away from the fovea, try to read preceding text in this paragraph while fixating on the period at the end of this sentence.

It is probably difficult, if not impossible, to read text that is only a few lines away from the point of fixation! The fovea covers an area that subtends about 2 degrees of visual angle in the central field of vision. To visualize 2

degrees of visual angle, a general rule is that the width of your thumbnail, held at arm's length, is approximately 1 degree of visual angle.

Eye saccadic movements

The importance of foveal vision is also stressed by the so-called saccadic eye movements. Saccades are quick, involuntary, simultaneous movements of both eyes in the same direction performed by humans and many other animals (as opposed to most birds).

Our eyes make jerky saccadic movements and stop several times in the so-called **fixations**, moving very quickly between each stop. We process the entire light information acquired during the fixations, while during the saccades we attenuate the importance of low spatial frequencies and only process high spatial frequency, a phenomenon called *saccadic masking.*

The reason for this behavior is twofold:

- It allow foveal (high resolution) view of interesting parts of the scene to build up a mental, three-dimensional map;

- The foveal vision corresponds to the resolution of a digital camera of roughly 1 Mega pixel, if cones had the foveal density throughout the rest of the retina, we would have the resolution of 1 Giga pixel, which is too much information to handle at once for the optical nerve. So saccadic movements are a brilliant way that nature chose to save 'computational space': instead of having an immediate high resolution vision of the world around us, we see sequences of sharp parts of it and then the brain 'links together' these information.

Of course, in order to have the impression of a continuous vision, saccades must be quick, and in fact they are the *fastest movements* produced by the human body: typical saccadic movements while we are reading are of the order of 20-30 *ms*. This is possible thanks to the fact that they are generated by a neuronal mechanism that bypasses time-consuming circuits and activates the eye muscles more directly.

Using *eye trackers* to follow the saccadic movements, researchers have built and studied *saliency maps*, i.e. descriptions of gazes and point where the eyes insist looking at. Even if the interpretation of saliency maps is still debated, there is a general agreement about the fact that strong edges are the first ones to catch the attention, suggesting that first of all our visual system creates a 'rough' depiction of the contours of what we're watching, filling in the finer details in a second part of a multi-scale processing.

3.1.7 The macula

The fovea is also protected by a yellow filter known as the macula. The macula serves to *protect this critical area of the retina* from intense exposures to short-wavelength energy. It might also serve to reduce the effects of chromatic aberration that cause the short-wavelength image to be rather severely out of focus most of the time.

Unlike the lens, the macula does not become more yellow with age. However, there are significant differences in the optical density of the macular pigment from observer to observer and in some cases between a single observer's left and right eyes. The yellow filters of the lens and macula, through which we all view the world, are the major source of variability in color vision between observers with normal color vision.

The humors and the macula contribute to **strongly reduce the contrast of luminance** that hit the retinal cells, **forcing light towards an average gray level**.

The CIE (Commission International de l'Éclairage - International Commission of Illumination) has measured the **Glare Spread Function** (GSF) provoked by the optical system of the eyes as a function of the angle of light acquisition, age and race.

3.1.8 The optical nerve

A last key structure of the eye is the optic nerve. The optic nerve is made up of the axons (outputs) of the ganglion cells, the last level of neural processing in the retina. It is interesting to note that the optic nerve is made up of approximately one million fibers, carrying information generated by approximately 130 million photoreceptors. Thus there is a clear compression of the visual signal prior to transmission to higher levels of the visual system. A one-to-one 'pixel map' of the visual stimulus is never available for processing by the brain's higher visual mechanisms.

Since the optic nerve takes up all of the space that would normally be populated by photoreceptors, *there is a small area in each eye in which no visual stimulation can occur*. This area is known as the **blind spot**. The structures described above have a clear impact in shaping and defining the information available to the visual system that ultimately results in the perception of color appearance.

The action of the pupil serves to define retinal illuminance levels that, in turn, have a dramatic impact on color appearance. The yellow-filtering effects of the lens and macula modulate the spectral response of our visual

system and introduce significant inter-observer variability.

The spatial structure of the retina serves to help define the extent and nature of various visual fields that are critical for defining color appearance. The neural networks in the retina reiterate that *visual perception* in general, and specifically *color appearance*, **cannot be treated as simple point-wise image processing problems**.

Several of these important features are discussed in more detail in the following chapters.

Chapter 4

The retina

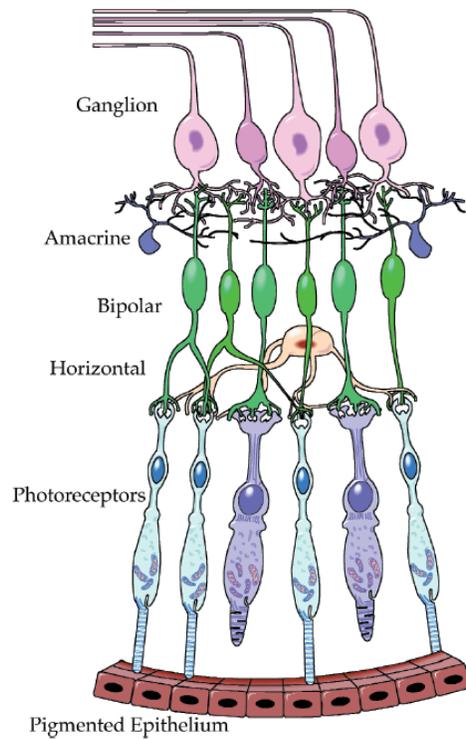Figure 4.1 illustrates a cross-sectional representation of the retina.

Figure 4.1: Composition of retinal layers.

The retina includes several layers of neural cells, beginning with the pho-

toreceptors, the rods and cones. Figure 4.1 provides only a slight indication of the extent of these various interconnections between retinal cells.

The specific processing that occurs in each type of cell is not completely understood and is beyond the scope of this notes. However, it is important to realize that the signals transmitted from the retina to the higher levels of the brain via the ganglion cells are not simple point-wise representations of the receptor signals, but rather consist of sophisticated combinations of the receptor signals.

To envision the complexity of the retinal processing, keep in mind that each synapse between neural cells can effectively perform a mathematical operation (add, subtract, multiply, divide) in addition to the amplification, gain control, and nonlinearities that can occur within the neural cells. Thus the *network of cells within the retina can serve as a sophisticated image computer.* This is how the information from 130 million photoreceptors can be reduced to signals in approximately *1 million ganglion cells* without loss of visually meaningful data.

It is interesting to note that *light passes through all of the neural machinery of the retina **prior** to reaching the photoreceptors.* This has little impact on visual performance since these cells are transparent and in fixed position, thus not perceived. It also allows the significant amounts of nutrients required and waste produced by the photoreceptors to be processed through the back of the eye.

4.1   Rods and cones

Figure provides a representation of the two classes of retinal photoreceptors, rods and cones.

Rods and cones derive their respective names from their prototypical shape. Rods tend to be long and slender while peripheral cones are conical. This distinction is misleading since foveal cones, which are tightly packed due to their high density in the fovea, are long and slender, resembling peripheral rods.

The most important distinction between rods and cones is in visual function. Rods serve vision at low luminance[1] levels (e.g., less than 10^{-3} cd/m^2) while cones serve vision at luminance levels higher than 10 cd/m^2, in the intermediate range both rods and cones work, but with less ability.

---

[1]We will introduce in section 4.2 the photometric units, for now, the only information that matters is that luminance is a sort of 'perceived radiance' and it is measured in cd/m^2, cd being the candela, SI unity of measure of the luminous intensity.
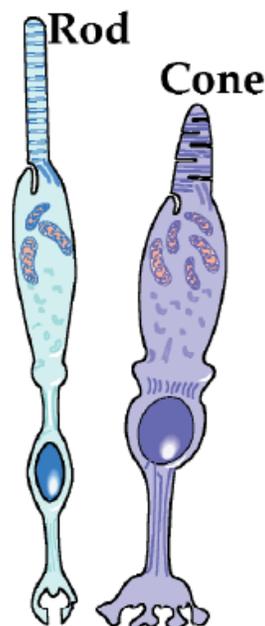
Rod

Cone

Figure 4.2: Prototypical shape of rods (left) and cones (right).

Thus the transition from rod to cone vision is one mechanism that allows our visual system to function over a large range of luminance levels. At high luminance levels (e.g., greater than 10 cd/m^2) the rods are effectively saturated and only the cones function. In the intermediate luminance levels, both rods and cones function and contribute to vision. Vision when only rods are active is referred to as **scotopic vision** (luminance $< 10^{-3}$ cd/m^2). Vision served only by cones is referred to as **photopic vision** and the term **mesopic vision** is used to refer to vision in which both rods and cones are active.

Rods and cones also differ substantially in their spectral sensitivities as illustrated in Figure 4.3, the curves are the CIE **spectral luminous efficiency functions** for scotopic, $V'(\lambda)$, and photopic, $V(\lambda)$ vision wavelengths (see subsection 4.1.1 to know how these curves have been obtained).

There is only one type of rod receptor with a peak spectral responsivity for V' at 510 nm. $V'(\lambda)$ is identical to the spectral responsivity of the rods and depends on the spectral absorption of **rhodopsin**, the photosensitive pigment in rods that bleaches when light is absorbed.

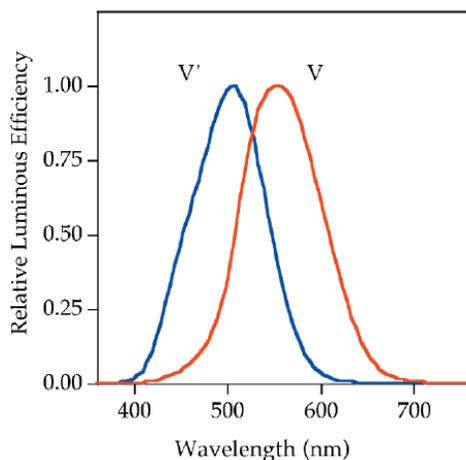Instead, there are three types of cone receptors with peak spectral re-

Figure 4.3: The Purkinje shift.

sponsivities spaced through the visual spectrum induced by three slightly different rhodopsin molecular structures, see Figure 4.4.

The difference in peak spectral sensitivity between scotopic and photopic vision means that with scotopic vision we are more sensitive to shorter wavelengths. This effect, known as the **Purkinje shift**, can be observed by finding two objects, one blue and the other red, that appear the same lightness when viewed in daylight. When the same two objects are viewed under very low luminance levels, the blue object will appear quite bright while the red object will appear nearly black because of the scotopic spectral sensitivity function.

The three types of cones are most properly referred to as L, M, and S cones. These names refer to the long-wavelength, middle-wavelength, and short wavelength sensitive cones, picked at 560 nm, 530 nm and 420 nm, respectively. Sometimes the cones are denoted with other symbols such as RGB, suggestive of red, green, and blue sensitivities. As can be seen in Figure 4.4, this concept is erroneous (in particular because the L cones are picked in the region of monochromatic green-yellow, not red!) and the LMS names are more appropriately descriptive.

The $V(\lambda)$ function represents the overall sensitivity of the cone system with respect to the perceived brightness of the various wavelengths, i.e. a combination of the three types of cone signals rather than the responsivity of any single cone type. $V(\lambda)$ has a pick at 555 nm.

Note that *the spectral responsivities of the three cone types are broadly*
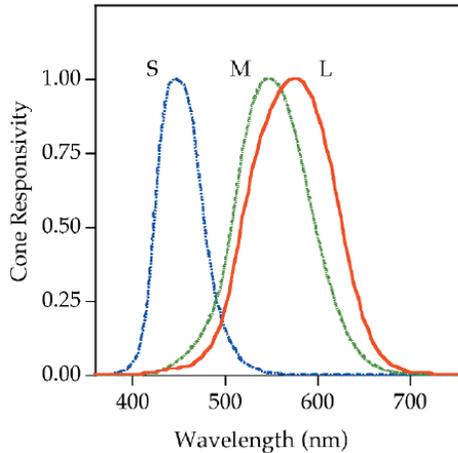
Figure 4.4: Spectral responsivities of L,M and S cones.

overlapping; a design that is significantly different from the 'color separation' responsivities that are often built into physical imaging systems. Such sensitivities, typically incorporated in imaging systems for practical reasons, are the fundamental reason that accurate color reproduction is often difficult, if not impossible to achieve.

The three types of cones clearly serve color vision. Since there is only one type of rod, the rod system is incapable of color vision. This can easily be observed by viewing a normally colorful scene at very low luminance levels.

Another important feature about the three cone types is their relative distribution in the retina.

It turns out that the S cones are relatively sparsely populated throughout the retina and completely absent in the most central area of the fovea. There are far more L and M cones than S cones and there are approximately twice as many L cones as M cones.

The relative populations of the L:M:S cones are approximately 12:6:1. *These relative populations must be considered when combining the cone responses.* (plotted with individual normalizations in Figure 4.4) to predict higher level visual responses. Figure 4.5 provides a schematic representation of the foveal photoreceptor mosaic with false coloring to represent a hypothetical distribution with the L cones in red, M cones in green, and S cones in blue. Figure 4.5 is presented simply as a convenient visual representation of the cone populations and should not be taken literally.
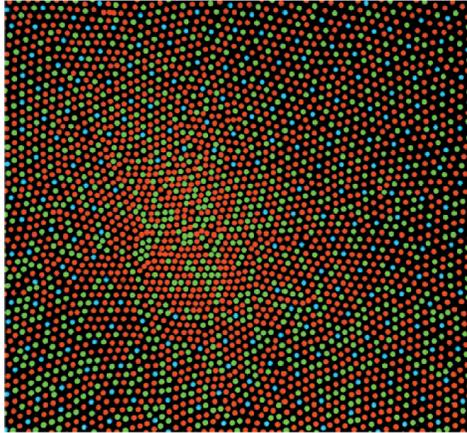
Figure 4.5: Each colored dot represents a particular type of cone (L,M or S) in the retina. Notice that there are no S cones or rods in the fovea.

There are *no rods present in the fovea*. This feature of the visual system can also be observed when trying to look directly at a small dimly illuminated object, such as a faint star at night. It disappears since its image falls on the foveal area where there are no rods to detect the dim stimulus. Figure 4.6 shows the distribution of rods and cones across the retina.
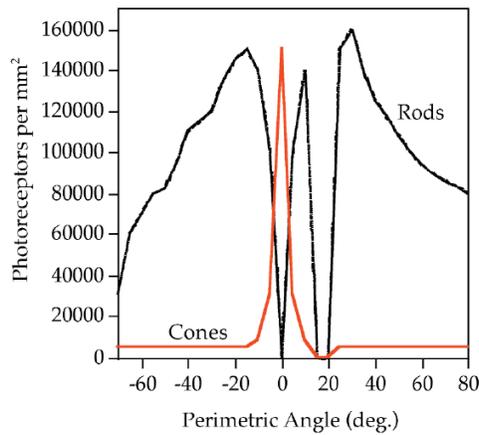
Figure 4.6: Distribution of rods an cones across the retina.

Several important features of the retina can be observed in Figure 4.6. First, notice the extremely large numbers of photoreceptors. In some reti-

nal regions there are about 150 000 photoreceptors per square millimeter of retina! Also notice *that there are far more rods* (around 130 million per retina) *than cones* (around 7 million per retina). This might seem somewhat counterintuitive since cones function at high luminance levels and produce high visual acuity while rods function at low luminance levels and produce significantly reduced visual acuity (analogous to low-speed fine-grain photographic film vs high-speed coarse-grain film).

The solution to this apparent mystery lies in the fact that *single cones feed into ganglion cell signals while rods pool their responses over hundreds of receptors* (feeding into a single ganglion cell) *in order to produce increased sensitivity at the expense of acuity.* This also partially explains how the information from so many receptors can be transmitted through one million ganglion cells. Figure 4.6 also illustrates that cone receptors are highly concentrated in the fovea and more sparsely populated throughout the peripheral retina while there are no rods in the central fovea.

Rods become denser and denser as we move far away from the fovea. The fact that they are distributed on the lateral parts of the retina and respond quicker than cones to visual stimuli is proven, for example, by looking at a TV monitor in dim light conditions with the back of our eyes and noticing the flickering given by the fact that the TV signal is not continuous but pulsed: this can deceive our cones, but not our rods! This anisotropic distribution is, again, a remainder of our evolution: primitive men had to defend themselves from the attack of fierce creatures during the night, that typically attack from the side, and this favored the selection of human beings with a prevailing number of rods in the lateral part of the retina.

A final feature to be noted in Figure 4.6 is the blind spot. This is the area, $12-15^0$ from the fovea, where the optic nerve is formed and there is no room for photoreceptors. *One reason the blind spot generally goes unnoticed is that it is located on opposite sides of the visual field in each of the two eyes.* However, even when one eye is closed, the blind spot is not generally noticed.

Figure 4.6 provides some stimuli that can be used to demonstrate the existence of the blind spot and the filling in phenomena.

To observe your blind spot, close your left eye and fixate the cross in Figure 4.7(a) with your right eye. Then adjust the viewing distance of the figure until the spot to the right of the cross disappears when it falls on the blind spot. Note that what you see when the spot disappears is not a black region, but rather it appears to be an area of blank paper. This is an example of a phenomenon known as **filling in**. Since your brain no longer has any signal indicating a change in the visual stimulus at that location, it

Figure 4.7: Experiment to prove the existence of the blind spot and the filling in phenomena.

simply *fills in the most probable stimulus*, in this case a uniform white piece of paper. The strength of this filling in can be illustrated by using 4.7(b) to probe your blind spot. In this case, with your left eye closed, fixate the cross with your right eye and adjust the viewing distance until the gap in the line disappears when it falls on your blind spot. Amazingly the perception is that of a continuous line since that is now the most probable visual stimulus. If you prefer to perform these exercises using your left eye, simply turn the book upside down to find the blind spot on the other side of your visual field.

The filling in phenomenon goes a long way to explain the function of the visual system. The *signals* present in the *ganglion cells* represent *only local changes in the visual stimulus*. Effectively, only information about spatial or temporal transitions, i.e. *edges*, is *transmitted to the brain*. Perceptually this code is sorted out by *examining the nature of the changes and filling in the appropriate uniform perception until a new transition is signaled*. This coding provides *tremendous savings in bandwidth to transmit the signal* and can be thought of as somewhat similar to run-length encoding that is sometimes used in digital imaging.

4.1.1 Experiments to measure spectral luminous efficiency functions

There are various techniques to measure the spectral luminous efficiency functions $V(\lambda)$ and $V'(\lambda)$. The CIE curves for $V(\lambda)$ and $V'(\lambda)$ shown in Figures 4.3 and 4.4 are obtained by averaging the results of different laboratories that used the technique of mono and heterochromatic adjustment.

This technique consists in separating the visual field in two parts illuminated by two light of different intensity. If we use two monochromatic lights, then, by changing the light intensity we will arrive at a point in which the

separation between the two fields is imperceptible. The sensitivity to intensity variation is different for each wavelength selected in the monochromatic illuminant. This is called **monochromatic adjustment**.

In **heterochromatic adjustment** instead we take as reference the luminous efficiency of the monochromatic adjustment for a light of 555 nm and we measure the *relative efficiency* by fixing this light in one part of the field and using lights with other wavelengths in the other part. Of course in this case there won't be a perfect intensity match, so that the edge between the two parts will still be visible and match will correspond to the minimal contour visibility.

4.1.2 Photo-electrical response of cones and rods

We have already said that the humors and macula strongly reduce the range of luminance hitting the retina. When the photo-chemical transduction from light to electric impulses performed by the photoreceptors occurs, we have another range reduction. In fact, it has been measured that when a photoreceptor absorbs a luminance L, the electric potential of its membrane changes accordingly to the empirical law known as **Michaelis-Menten's equation** or **Naka-Rushton's equation** when $n = 1$:

$$r(L) = \frac{\Delta V}{\Delta V_{\max}} = \frac{L^n}{L^n + L_S^n},\qquad(4.1)$$

where $\Delta V_{\max}$ is the highest difference of potential that can be generated, n is a constant (measured as 0.74 for the rhesus monkey) and L_S is the luminance at which the photoreceptor response is half maximal, called the *semisaturation level*, and which is usually associated with the *level of adaptation*. Each type of cone is most sensitive over a particular waveband and the semisaturation constant depends on the amount of light in the particular waveband that reaches it, not on the global luminance of the light source.

In Figure 4.8 we show the compressive effect of this behavior in arbitrary units for rods and cones.

It is interesting to note that these functions show characteristics similar to those found in all imaging systems. At the low end of the receptor responses there is a *threshold*, below which the receptors do not respond. There is then a *fairly linear portion of the curves*, followed by response *saturation* at the high end. Such curves are representations of the photocurrent at the receptors and represent the very first stage of visual processing. These signals are then processed through the retinal neurons and synapses until a
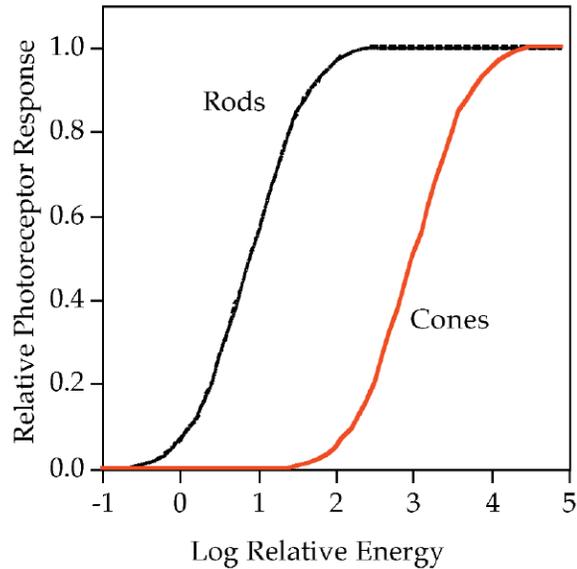
Figure 4.8: Compressive effect of Michaelis-Menten's equation in arbitrary units.

transformed representation is generated in the ganglion cells for transmission through the optic nerve.

We will consider again eq. (4.1) in the light adaptation process.

4.2 Radiometric vs Photometric units: Luminous intensity and luminance

We conclude this chapter by relating radiometric and photometric units. Radiometric units describe physical attributes of light, while *photometric unities take into account the eyes' response to light stimuli.*

As we have seen, the response of photoreceptors to light stimuli is different in each wavelength, the spectral luminous efficiency functions $V(\lambda)$ and $V'(\lambda)$ of cones and rods, respectively, quantify this response.

Given a generic radiometric unit X_e, we define the corresponding photometric unit X_V (the suffix V stays for 'visual') multiplying X_e by the adimensional weight given by $V(\lambda)$ or $V'(\lambda)$ (depending if the visual condition is photopic or scotopic) and integrating over the visual spectrum,

i.e.

$$X_V = K_m \int_{380\,nm}^{780\,nm} X_e(\lambda)V(\lambda)\,d\lambda \qquad \text{Photopic conditions;}$$

$$X_V = K'_m \int_{380\,nm}^{780\,nm} X_e(\lambda)V'(\lambda)\,d\lambda \qquad \text{Scotopic conditions,}$$

where K_m and K'_m are two constants whose value is $K_m = 683\,lm/W$ and $K'_m = 1700\,lm/W$, they are called **luminous efficiency** in photopic and scotopic condition, respectively. Mesopic conditions are still a matter of research and no commonly accepted definition is available. In practice, the integration is never performed and one considers a finite sum with a small step in wavelength, typically of the order of units of nanometers. Of course, for a monochromatic radiation of wavelength λ, there is no need to consider any integration.

The table in Figure 4.9 provides the list of photometric units.

Quantity	Symbol	SI unit	Symbol	Dimension	Notes
Luminous energy	Q_v	lumen second	lm·s	$T·J$ [nb 3]	units are sometimes called *talbots*
Luminous flux	Φ_v	lumen (= cd·sr)	lm	J	also called *luminous power*
Luminous intensity	I_v	candela (= lm/sr)	cd	J	an SI base unit, luminous flux per unit solid angle
Luminance	L_v	candela per square metre	cd/m^2	$L^{-2}·J$	units are sometimes called *nits*
Illuminance	E_v	lux (= lm/m^2)	lx	$L^{-2}·J$	used for light incident on a surface
Luminous emittance	M_v	lux (= lm/m^2)	lx	$L^{-2}·J$	used for light emitted from a surface
Luminous exposure	H_v	lux second	lx·s	$L^{-2}·T·J$	
Luminous efficacy	η	lumen per watt	lm/W	$M^{-1}·L^{-2}·T^3·J$	ratio of luminous flux to radiant flux
Luminous efficiency	V			1	also called *luminous coefficient*

Figure 4.9: SI photometric units.

Notice that

- The luminous energy Q_V is the photometric equivalent of the radiant energy Q_e;

- The luminous flux Φ_V, measured in **lumen**, $lm = cd \cdot sr$, is the photometric equivalent of the radiant flux Φ_e;

- The luminous intensity I_V is the photometric equivalent of the radiant intensity I_e;

- The luminance L_V, measured in cd/m^2, is the photometric equivalent of the radiance L_e and so it is **conserved in geometric optical systems**;

- The illuminance E_V, measured in **lux**, $lx = lum/m^2$, is the photometric equivalent of the irradiance E_e;

- The luminous emittance M_V is the photometric equivalent of the radiant emittance M_e.

To understand why the K_m and K'_m constants are introduced we need to talk about the definition of one of the 7 fundamental units of the SI, the luminous intensity, and its unit of measure, the **candela**. In 1979, the SI defined 1 cd as *the luminous intensity I_V, in a given direction, of a source which emits a monochromatic radiation of wavelength $540 \cdot 10^{12} Hz$ with a radiant intensity I_e in that direction of* $1/683$ W/sr. The name refers to the fact that the luminous intensity of 1 cd is approximatively that produced by a real candle.

Three parameters of the definition must be explained:

1. The SI has chosen a frequency ν of $540 \cdot 10^{12} Hz$ because the corresponding wavelength $\lambda = v/\nu$, being v the speed of light in standard air, is 555 nm, which corresponds to the peak of $V(\lambda)$, so such a monochromatic radiation is chosen to minimize the radiant energy needed to produce the same luminous intensity;

2. To define 1 cd a radiant intensity of $1/683$ W/sr is needed because, being $V(555\,nm) = 1$, we have

$$I_V = 683\,lm/W\frac{1}{683}W/sr = 1\,lm/sr = 1\,cd.$$

The value $K_m = 683\,lm/W$ has been chosen simply to maintain coherence with the value of the candela defined in 1948 by stating that 'the luminous intensity of a black body radiator at the solidification temperature of the platinum is 60 cd/cm^2'. The definition has been changed because of the difficulty in creating a perfect black body radiator and thanks to the advances in laser so that a monochromatic radiation of 555 nm that propagates in a well specified direction can be produced without major efforts.

3. Similarly, the value $K'_m = 1700\,lm/W$ allows equivalence with the former definition of candela in scotopic conditions.

Examples of luminous intensities:

- Fission bomb right after the explosion: 10^{12} cd;

- Lightning: 10^{10} cd;

- Sun from the surface of the Earth: 10^9 cd;

- Full moon: 10^3 cd;

- $25W$ isotropic fluorescent bulb: 135 cd;

- Standard candle: 1 cd;

- LEDs can range from 10^{-3} to 10^1 cd (ultra bright leds).

The candela is the only SI unit in which the interaction between a physical quantity (radiant intensity) and humans is explicitly taken into account.

4.2.1 Radiometers and Photometers and Spectrophotometers

Radiometers are instruments that can measure radiant units (as radiance, irradiance and so on), the most precise radiometer available nowadays is the so-called **cryogenic radiometer** which capitalizes the fact that at temperatures close to the absolute zero (0 K = -273,15 ^0C) it is possible to establish a very precise correspondence between optical power and electric power, that can be measured with great precision (0.01%). Less precise, but most frequently used, radiometers use photoresistors, photodiodes, photomultipliers or photon counters, whose description is beyond the scope of these notes.

Photometers are instruments that can measure photometric units (as luminance, illuminance and so on) because the are able to perform the operations to pass from radiometric to photometric units. An example of photometer is shown in Figure 4.10.

Spectrophotometers can analyze the photometric propertied of the different wavelengths of light by using a **monochromator**, which is an optical device that transmits a mechanically selectable narrow band of wavelengths of light chosen from a wider range of wavelengths available at the input. A monochromator can use either the phenomenon of *optical dispersion in a prism*, or that of diffraction using a *diffraction grating*, to spatially separate the wavelengths of light. It usually has a mechanism for directing the selected wavelength to an exit slit.

Figure 4.10: A photometer.

Once luminous intensity, luminance or illuminance are measured, one can pass from one to another by using their definitions, e.g. the relation between luminance L_e and illuminance E_e is

$$L_e = E_e \frac{d^2}{A},$$

where A is the area of the emitter and d is the distance between emitter and receiver.

Chapter 5

CIE colorimetry and illuminants

As we have already said, the CIE is the Commission International de l'Éclairage (International Commission on Illumination) is the international authority on light, illumination, color, and color spaces. It was established in 1913 as a successor to the Commission Internationale de Photométrie and is today based in Vienna, Austria.

CIE colorimetry is the **metrology of the psychophysical color stimulus**. In its very fundamental form CIE colorimetry is based on so-called **Class-A observations** or color matching experiments (invented by the German polymath Hermann **Grassmann** (1809-1877) around 1853): if two stimuli with unequal physical characteristics produce, under identical exterior circumstances, the same sensation, then we regard them to be equivalent.

Color metrology (the science of defining quantitative measures and metrics) is possible because, despite the fact that color is a subjective and non conveyable perception and it strongly depends on the context, some fundamental rules can be elaborated, in particular there are experimental evidences that a vast majority of people agree about **color differences**. Since a metric is nothing but a meaningful 'distance' between two values of a quantity, this fact turns color into a measurable entity.

More advanced descriptions of psychophysical phenomena, for example, the determination of the **brightness** of a stimulus in a complex context composed with different colored lights, is beyond the realms of basic CIE colorimetry, and the determination of its physical correlates is part of **advanced colorimetry**. Items belonging to this later group are, for example,

color appearance models. Such phenomena are called **Class B observations** and will be dealt later.

5.1 Viewing conditions

Various aspects of the viewing field impact on the color appearance of a light stimulus. colorimetry needs conventions. Such conventions are necessary because two colour stimuli looking similar under one viewing condition might look different when seen under other conditions (field of view, adaptation, direction of viewing, etc.). Hence accurate definitions and descriptions of the components of the viewing field as shown in Figure 5.1 are necessary for the development and correct use of a color experiment. Here we follow the definitions given by two of the most important researchers in color science: Hunt and Fairchild.

- **Stimulus** is a color element for which a measure of color appearance is required. Typically, the stimulus is taken to be a **uniform patch** of about 2^0 **angular subtense** to be consistent with foveal vision;

- A **proximal field** is the immediate environment of the color element considered, extending typically for about 2^0 from the edge of that color element in all or most directions;

- The **background** is defined as the environment of the color element considered, extending typically for about 10^0 from the edge of the proximal field in all, or most directions. When the proximal field is the same color as the background, the latter is regarded as extending from the edge of the color element considered;

- A **surround** is a field outside the background. In practical situations, the surround can be considered to be the entire room or the environment in which the image is viewed. For example, printed images are usually viewed in an illuminated (average) surround, projected slides in a dark surround, and domestic television displays in a dim surround;

- An **adapting field** is the total environment of the color element considered, including the proximal field, the background, and the surround, and extending to the limit of vision in all directions.
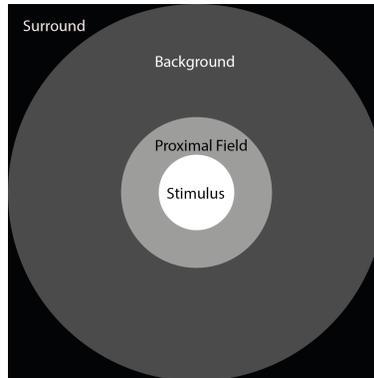
Figure 5.1: Components of the adaptive field.

5.2 Additive and subtractive color mixing. Color matching experiments and standard CIE observers

There are two fundamental methods of producing color stimuli: **additive** and **subtractive** color **mixing**:

- In additive color mixing lights are mixed in a so-called photometric sphere (often called Ulbricht sphere) and viewed via the exit port of the sphere, or superimposed on the same spot of a reflective screen or emitted separately but closely and fused in our eyes, as in color TV displays, where the color sensation in our eye is produced by the additive mixture of tiny red, green, and blue lights, where the single spots are so near to each other that our eye is unable to resolve them spatially and we see the mixture of the lights. By changing the intensity of the single spots, different mixed colors can be produced;

- In subtractive color mixing, pigments remove some part of the visible spectrum. Superposing several colorants of different concentrations on each other will change the color of the reflected light.

Basic colorimetry is the description of the results of color matching experiments, is built on additive color mixing because the laws of additive color mixing (called Grassmann's laws) are simpler than those of subtractive color mixing.

A **color matching experiment** consists typically in a human tester, embedded in a dark room and adapted to dim light conditions, placed in

48

front of a so-called bipartite field: one part of the field is characterized by the light of a reference test stimulus, the other by three lights that can be varied in order to match the reference stimulus.

In Figure 5.2 we can see the apparatus typically used to create a bipartite field: the three matching stimuli are mixed in a photometer sphere (often called **Ulbricht sphere**) and viewed via the exit port of the sphere. A mirror blends the light coming from the test stimulus into one half of *the viewing angle field needed for the experiment.* A black light trap absorbs the light of the test lamp not hitting the white reflecting plate. Two light baffles in the sphere (not shown in the figure) secure that no direct light from the sources or the mixture field reaches the observer.
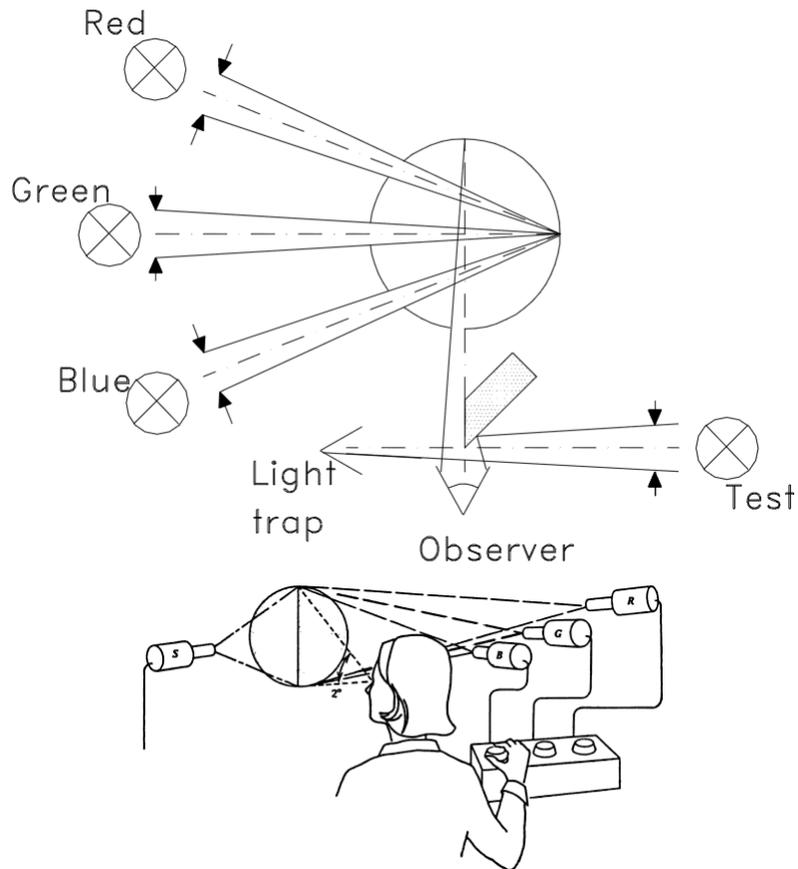

Figure 5.2: Above: schematic drawing of the principles a visual colorimeter. Below: superposition of projected lights

One may ask why exactly three light sources are used. The answer is that it has been proven experimentally that three independent lights (i.e. none of them can be obtained by mixing the other twos) are necessary and enough to obtain the match, this fact suggested to the polymaths Thomas **Young** (English 1773-1829) and Hermann **Helmholtz** (German, 1821-1894) the idea of the existence of three independent photoreceptors in our eyes a century before their physiological discovery! Their theories are nowadays part of the so-called **trichromatic theory** of color vision. This theory only describes in a successful way the very first part of the visual cascade of events and must be complemented with models of higher vision mechanisms to explain many phenomena, as we will see later.

Let us now mathematically model the color matching experiments: let $i = 1, 2, 3$, then

- $S_i(\lambda)$ denotes the spectral sensitivity distributions of the three types of retinal **cones** (Figure 4.4);

- $C(\lambda)$ is the spectral optical energy of a **test stimulus**;

- $\Lambda = [380, 780]nm$ is the interval of visible wavelengths.

As showed in Figure 5.3, after the photons that compose $C(\lambda)$ have been absorbed by the cones, the cascade of events that induces the color sensation will be started by these three **analysis coefficients** or **cone activation coefficients**:

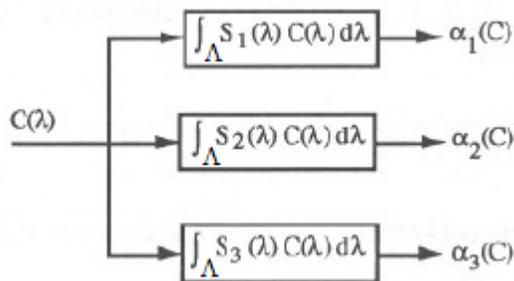$$\alpha_i(C) = \int_\Lambda S_i(\lambda)C(\lambda)\, d\lambda. \tag{5.1}$$

Figure 5.3: Schematic drawing of the principles a visual colorimeter.

A fundamental property of human vision is **metamerism**, which can be explained as follows: suppose we have two color stimuli $C_1(\lambda)$ and $C_2(\lambda)$

such that
$$\alpha_i(C_1) = \alpha_i(C_2) \qquad \forall i = 1, 2, 3,$$

then the analysis coefficients of the two color stimuli are exactly the same and they produce the same color sensation even if the spectral distribution of the two stimuli is different.

Metamerism is a direct consequence of the fact that our eyes perform an additive mixing of light instead of responding to each monochromatic radiation separately (as, e.g. ears do with sound of different frequency).

Metamerism is fundamental in colorimetry because it allows **color reproduction**, without metamerism TV's and projectors would be useless and, more in general, we wouldn't be able to reproduce color.

Since, instead, experiments shows that we can reproduce the same color sensation of a given stimulus by adding three synthesize lights, let us consider three light sources, that we will call **primaries**, with optical energy distribution given by $P_i(\lambda)$, $i = 1, 2, 3$ that must be independent, in the sense that none of the primaries can be matched by the additive mixture of the other two primaries. The inventor of modern linear algebra, Grassmann, used the color primaries as one of the first practical *examples of basis of a linear space*.

The synthesis of the three primaries will be obtained by a generic linear combination, in particular we are interested in the linear combination with coefficients β_k, $k = 1, 2, 3$ such that:

$$\alpha_i(C) = \int_\Lambda \left[\sum_{k=1}^3 \beta_k P_k(\lambda) \right] S_i(\lambda) \, d\lambda, \qquad i = 1, 2, 3,$$

because in this case the mix of the three primaries will be metameric to the test color stimulus C. We can rearrange the previous formula in this way

$$\alpha_i(C) = \sum_{k=1}^3 \beta_k \int_\Lambda P_k(\lambda) S_i(\lambda) \, d\lambda, \qquad i = 1, 2, 3,$$

let us write
$$a_{ik} = \int_\Lambda P_k(\lambda) S_i(\lambda) \, d\lambda, \qquad i = 1, 2, 3,$$

so that

$$\alpha_i(C) = \sum_{k=1}^3 a_{ik} \beta_k = \int_\Lambda S_i(\lambda) C(\lambda) \, d\lambda, \qquad i = 1, 2, 3, \tag{5.2}$$

which are three equations called 'color matching equations'.

In practice, for reasons that will be explained in section 5.4, it is convenient to calibrate the primaries P_k against a reference source of light considered 'white' (**reference white** from now on) and with known optical energy distribution $W(\lambda)$.

Let us denote with w_k the matching coefficients needed to match the primaries to W, i.e. those numbers such that

$$\alpha_i(W) = \sum_{k=1}^{3} w_k \int_\Lambda P_k(\lambda) S_i(\lambda) \, d\lambda \qquad i = 1, 2, 3.$$

Given a test color stimulus $C(\lambda)$, a white reference $W(\lambda)$, the three primaries $P_i(\lambda)$ and their matching coefficients β_k to $C(\lambda)$ and w_k to $W(\lambda)$, we define **the tristimulus values of C with respect to the primaries P_k and the reference white W** as follows:

$$\boxed{T_k(C) = \frac{\beta_k}{w_k}}, \qquad k = 1, 2, 3.$$

We can see that the tristimulus values of a color stimulus can be interpreted as the components of C with respect to the basis given by the three primaries P_k, normalized with respect to the reference white W, which can be considered as a sort of 'unit color' since its tristimulus values are obviously 1:

$$T_k(W) = \frac{w_k}{w_k} = 1, \qquad k = 1, 2, 3.$$

So the tristimulus values of a color stimulus C are the quantities of each primary (relative to the reference white) that one must use to metamerically match C with a superposition of the primaries P_k.

Among all color stimuli there are monochromatic optical radiations. A monochromatic stimuli of spectral optical energy 1 can be expressed by the Dirac delta $C(\lambda) = \delta(\bar{\lambda} - \lambda)$ and its tristimulus values are usually denoted by $T_k(\bar{\lambda})$.

For a monochromatic stimulus eq.(5.1) becomes

$$\alpha_i(\delta(\bar{\lambda} - \lambda)) = \int_\Lambda S_i(\bar{\lambda}) \delta(\bar{\lambda} - \lambda) \, d\bar{\lambda} = S_i(\lambda),$$

where S_i is the sensitivity of the i-th cone. Eq. (5.2) becomes

$$\sum_{k=1}^{3} a_{ik} \beta_k = S_i(\lambda), \qquad k = 1, 2, 3.$$

but $T_k(\lambda) = \beta_k / w_k$, so that $\beta_k = w_k T_k(\lambda)$ and thus

$$\boxed{\sum_{k=1}^{3} a_{ik} w_k T_k(\lambda) = S_i(\lambda)}, \qquad \forall \lambda \in \Lambda, \ k = 1, 2, 3.$$

Since $S_i(\lambda)$, a_{ik} and w_k can be computed, this equation allows us to calculate the value of $T_k(\lambda)$ for every $\lambda \in \Lambda$ and so to construct the so called

Color matching functions

$$\begin{aligned} T_k : \ \Lambda &\longrightarrow \ \mathbb{R} \\ \lambda &\longmapsto \ T_k(\lambda). \end{aligned}$$

Since the values $T_k(\lambda)$ are the tristimulus values of a monochromatic light $\delta(\bar{\lambda} - \lambda)$, the tristimulus values of any other color stimulus $C(\lambda)$ will be obtained by integrating the color matching functions multiplied by $C(\lambda)$:

Tristimulus of C in terms of color matching functions

$$\boxed{T_k(C) = \int_{\Lambda} C(\lambda) T_k(\lambda) \, d\lambda}.$$

Experimentally, it can be seen that the luminance of $C(\lambda)$ can be obtained by summing the luminances of its components:

$$L_V(C) = K_m \sum_{k=1}^{3} T_k(C) w_k \int_{\Lambda} P_k(\lambda) V(\lambda) \, d\lambda,$$

where K_m is the luminous efficiency coefficient and $V(\lambda)$ is the photopic luminous efficiency function.

The luminance of a color contributes, together with its saturation (which measures the distance between pure colors and white) and hue (the dominant wavelength that defines the overall chroma of the color) define three alternative coordinates for describing color.

In many situations, it is useful to fix the luminance value and consider the so-called **chromaticity coordinates**, given by:

$$t_k = \frac{T_k}{T_1 + T_2 + T_3}, \qquad k = 1, 2, 3,$$

where, of course $t_1 + t_2 + t_3 = 1$, so that it is necessary to specify only two chromaticity coordinates, since the third one is just 1 minus the sum of the other two. As we will see, the chromaticity coordinates are directly related to saturation and hue of a color.

5.2.1 The CIE 1931 standard colorimetric observer

To be able to repeat an additive color match precisely the observation conditions have to be standardized. For the CIE 1931 standard colorimetric observer a 2^0 foveal field of observation and a dark surround was chosen.

The CIE 1931 standard colorimetric observer was derived from the results of two experimental investigations, conducted by W. D. **Wright** and J. **Guild**. The two investigations used different primaries, but when transforming the results to a common system the agreement was surprisingly good, despite the fact that *the number of observers was only 7 in Guild's work and only 10 in Wright's!*

To be able to define a standard observer the spectral compositions and the luminances of the primaries have to be specified. Single wavelengths were used: 700 nm for the red, 546.1 nm for the green, and 435.8 nm for the blue primary. The red stimulus was selected using an interference filter in order to be in the part of the spectrum where the sensation of red changes very slowly with respect to wavelength. The green and blue stimuli were chosen to coincide with the emission lines of of a mercury discharge lamp as it facilitated the wavelength calibration.

The 'unit intensity' of the primaries was defined by stating their luminances. The requirement the matching between the three primaries specified above and a test stimulus with **equienergetic spectrum**, i.e. a light with the same radiant energy for each wavelength. The results have shown that if 1 cd/m^2 of red light was used, then 4.5907 cd/m^2 of green and 0.0601 cd/m^2 blue light was needed to match the color of a stimulus with equienergetic spectrum, see Figure 5.4. The mixture of them matches the equienergetic white stimulus of 5.6508 cd/m^2.
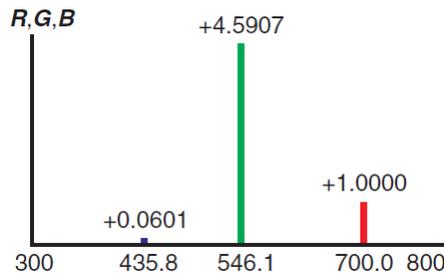

Figure 5.4: CIE 1931 primaries.

As we have seen, we can obtain the color matching functions by fixing a monochromatic light from 380 nm to 780 nm and then modifying the

intensities of the primaries chosen above to obtain a match. The color matching functions in the CIE 1931 system are written $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, $\bar{b}(\lambda)$ and they depicted in Figure 5.5.

The negative lobes of the red curve need an explanation. The observations have shown that at 520 nm green stimulus couldn't be matched by any combination of the three matching stimuli. Exact match could only be obtained if some red light was mixed to the test stimulus, i.e. the color mixture of given amounts of the green and blue matching stimuli will match the mixture of the test and red stimuli. Mathematically, the fact that from one of the primaries light is mixed with the light of the test stimulus is written with a negative sign for that stimulus (which, of course, doesn't mean a 'negative light', which does not make any sense).
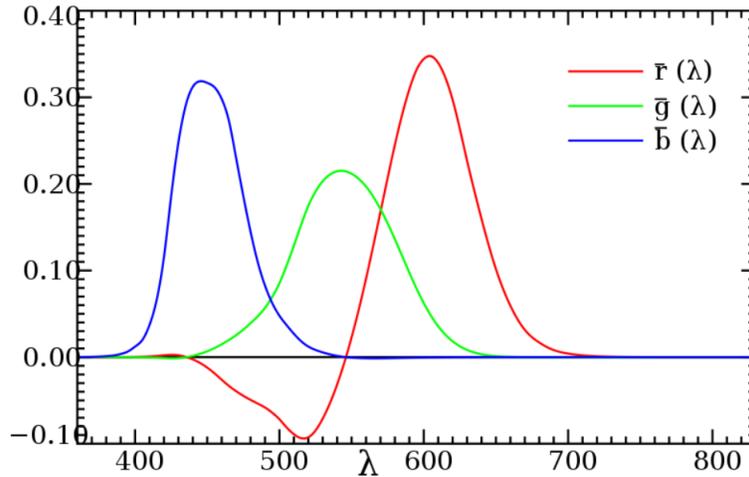

Figure 5.5: CIE 1931 $(\bar{r}, \bar{g}, \bar{b})$ color matching functions.

5.2.2 The CIE 1964 supplementary standard colorimetric observer

The CIE 1931 trichromatic system is recommended only for small, 1^0-4^0 size, stimuli. We need, however, the description of larger stimuli as well, where the stimulus falls on a larger area of the retina than the one covered by the macula lutea, or where we see the stimulus partly out of fovea. For that purpose the CIE standardized a large field colorimetric system, based on the visual observations conducted on a 10^0 visual field. A 10^0 visual field represents a diameter of about 9cm at a viewing distance of 0.5 m.

The researches that developed the experiments were **Stiles**, **Burch** and **Speranskaya**. The primaries used are R (645.2 nm), G (526.3 nm), and B (444.4 nm). The new color matching functions are similar to those of the 2^0 visual fields, but not identical, and they are represented in Figure 5.6.
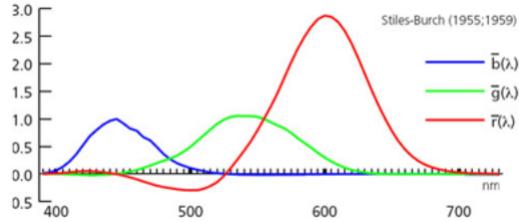

Figure 5.6: CIE 1964 $(\bar{r}, \bar{g}, \bar{b})$ color matching functions.

5.3   Color spaces

Color spaces allow a numerical specification of color stimuli, which is very useful for practical purposes.

5.3.1   The RGB color space

In the RGB color space the tristimulus values of a color stimulus of radiance $C(\lambda)$ are written as R,G and B and are obtained with the $\bar{r}$, $\bar{g}$ and $\bar{b}$ CIE color matching functions:

$$R = \int_\Lambda C(\lambda)\bar{r}(\lambda)\, d\lambda,$$

$$G = \int_\Lambda C(\lambda)\bar{g}(\lambda)\, d\lambda,$$

$$B = \int_\Lambda C(\lambda)\bar{b}(\lambda)\, d\lambda,$$

with

$$\int_\Lambda \bar{r}(\lambda)\, d\lambda = \int_\Lambda \bar{g}(\lambda)\, d\lambda = \int_\Lambda \bar{b}(\lambda)\, d\lambda,$$

because the equienergetic stimulus has identical components.

The tristimulus values of the reference white are $R = G = B = 1$.

The primaries of the RGB color space are the monochromatic radiations chosen to define the standard colorimetric observer: 700 nm for the red, 546.1 nm for the green, and 435.8 nm for the blue primary.

Mixing these fixed primaries we can obtain only a portion of all visible colors, called **gamut**, which is the triangular region that can be seen in Figure 5.7. This doesn't mean that the other colors can never be represented, it just means that they cannot be represented with this particular choice of primaries. Other primaries can represent a different gamut. No practical set of three primaries has been found that can reproduce the whole perceivable color gamut.
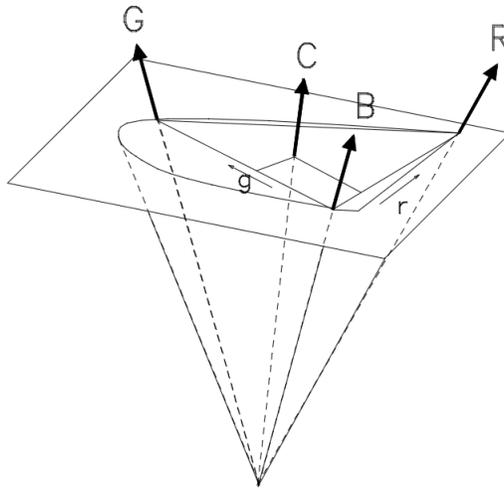

Figure 5.7: Gamut of visible colors versus RGB gamut.

A plane has been drawn where the primaries have unit values, this is called **Maxwell plane**. On the Maxwell plane the chromaticity coordinates $r = R/(R + G + B)$ and $g = G/(R + G + B)$ are drawn.

Many times, in particular for computer applications, the RGB color space is represented as a cube, as in Figure 5.8.

5.3.2 The XYZ color space

When the CIE trichromatic system was built it was felt that the manipulation with negative values in the color matching functions may lead to calculation errors, so a set of fictitious (non-physical) non-negative color matching functions $\bar{x}$, $\bar{y}$ and $\bar{z}$ has been built (both for the 2^0 and for the 10^0 observer).
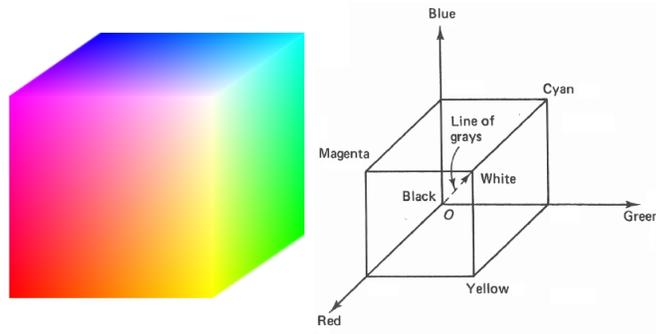
Figure 5.8: RGB cube.

The tristimulus values with respect these new set of color matching functions are

$$X = \int_\Lambda C(\lambda)\bar{x}(\lambda)\,d\lambda,$$

$$Y = \int_\Lambda C(\lambda)\bar{y}(\lambda)\,d\lambda = \text{Luminance},$$

$$Z = \int_\Lambda C(\lambda)\bar{z}(\lambda)\,d\lambda,$$

and $\bar{x}$, $\bar{y}$, $\bar{z}$ were built in such a way that the new Y tristimulus value would give the **luminance** of the color stimulus, chosen due to the already remarked importance of the luminance in optical systems.

The $\bar{x}$, $\bar{y}$ and $\bar{z}$ color matching functions are depicted in Figure 5.9.

Also here, the tristimulus values of the reference white are $X = Y = Z = 1$.

The chromaticity coordinates in this system are x, y, z:

$$x = \frac{X}{X + Y + Z}$$

$$y = \frac{Y}{X + Y + Z}$$

$$z = \frac{Z}{X + Y + Z}$$

$x + y + z = 1$.

We stress that the XYZ tristimulus coordinates don't have a direct physical interpretation, since they have been built for mathematical convenience,
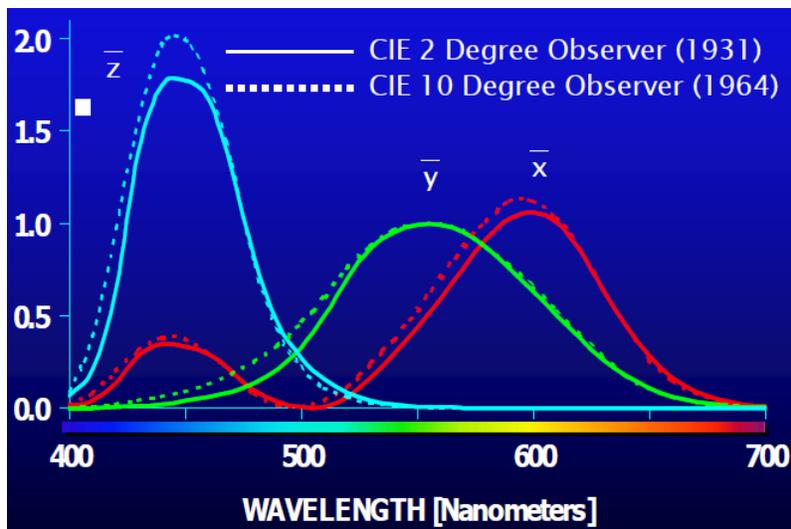
Figure 5.9: The 2^0 and 10^0 fictitious CIE XYZ color matching functions.

however they are related to the RGB tristimulus coordinates through the following linear transformation:

$$
\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.490 & 0.310 & 0.200 \\ 0.177 & 0.813 & 0.011 \\ 0.000 & 0.010 & 0.990 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}
$$

5.3.3 CIE chromaticity diagram

The CIE considered, as in Figure 5.10 a projection of the XYZ color reference to form the so called XYZ chromaticity diagram, or 'CIE flag', shown in Figure 5.11.

The CIE chromaticity diagram must be interpreted as follows:

- The curved portion of perimeter represents monochromatic colors (the wavelengths are indicated) with highest saturation;

- The straight part of the perimeter is called the 'purple line';

- The horizontal chromaticity coordinate is proportional to the amount of red;

- The vertical chromaticity coordinate is proportional to the amount of green;
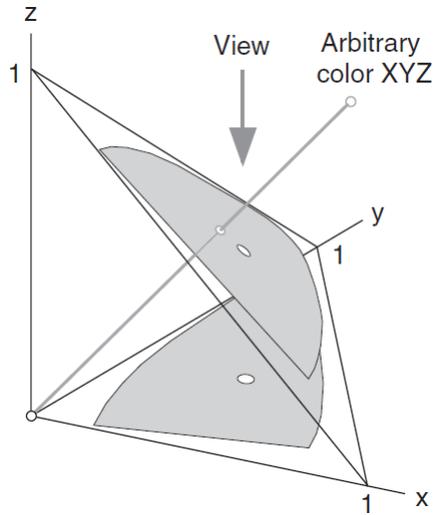
59

Figure 5.10: Projection of the CIE XYZ chromaticity space.

- The saturation decreases as we approach the barycenter of the diagram.

-

In Figure 5.12 we show the typical gamuts of monitors and printers inside the chromaticity diagram.

There are many other color spaces, that will be discussed in the lab classes.

5.3.4 Non uniformity of the CIE chromaticity diagram

The CIE chromaticity diagram has been built only with information given by the first stages of human vision taken in controlled conditions. Moreover, no interaction from the context of a scene is taken into account.

It is not surprising that the CIE chromaticity diagram, even if very useful for industrial purposes, doesn't give a faithful color representation. One of the first signals of this drawbacks was singled out by the experiments of Mac Adam (early 40's of the twentieth century), which have shown the so-called not uniformity of the CIE diagram, as shown in Figure 5.13.

The so-called MacAdam ellipses (10-times magnified in Figure 5.13) define areas in which colors are visually identical. The CIE chromaticity diagram is not uniform: the ellipses are much bigger in the green region than
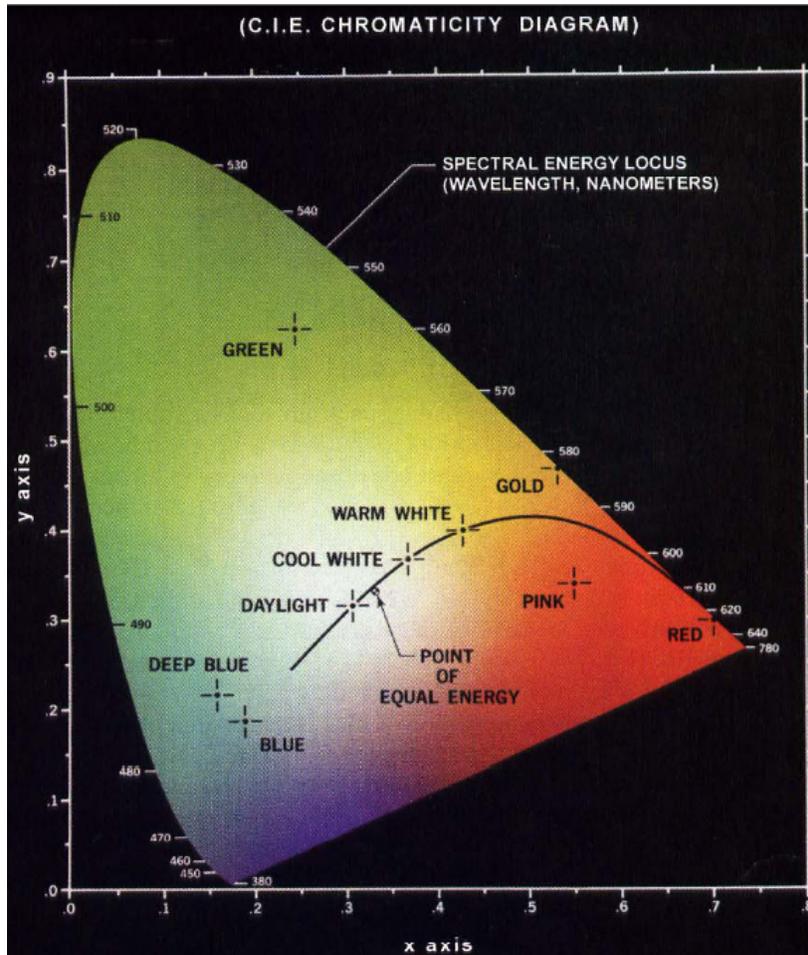
Figure 5.11: The CIE chromaticity diagram.

in the red region. To remedy this problem the CIE tried to 'stretch' the chromaticity diagram in such a way that Mac Adam ellipses are evenly distributed, creating a plethora of almost uniform color spaces called CIELAB, CIELUV, etc.

5.4   Light sources and illuminants

As we have seen, the tristimulus coordinates of the white reference are $(1, 1, 1)$. Changing reference white produces a color shift, so the reference
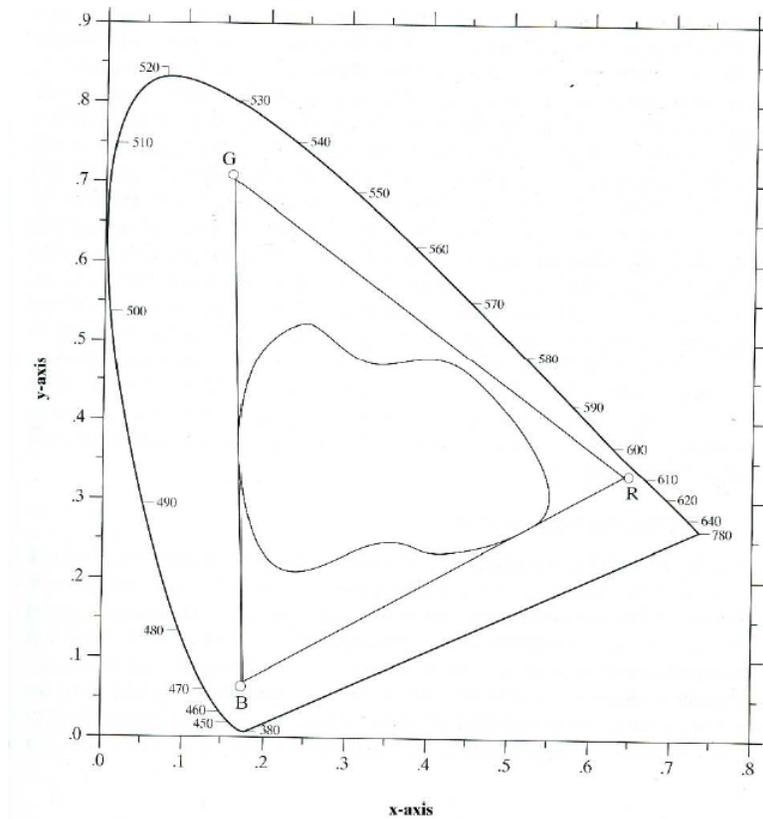
Figure 5.12: Triangular area: gamut of a standard R,G,B monitor. Irregular area: typical gamut of a printer.

white must be chosen with care. In order to know how to properly select the reference white it is essential to introduce some basic information about light sources and illuminants.

5.4.1 Artificial light sources

The most common artificial light sources are:

- Incandescent lamps (traditional, halogen, dichroic);

- Gas discharge (or florescent) lamps;

- Solid state lighting (LEDs).
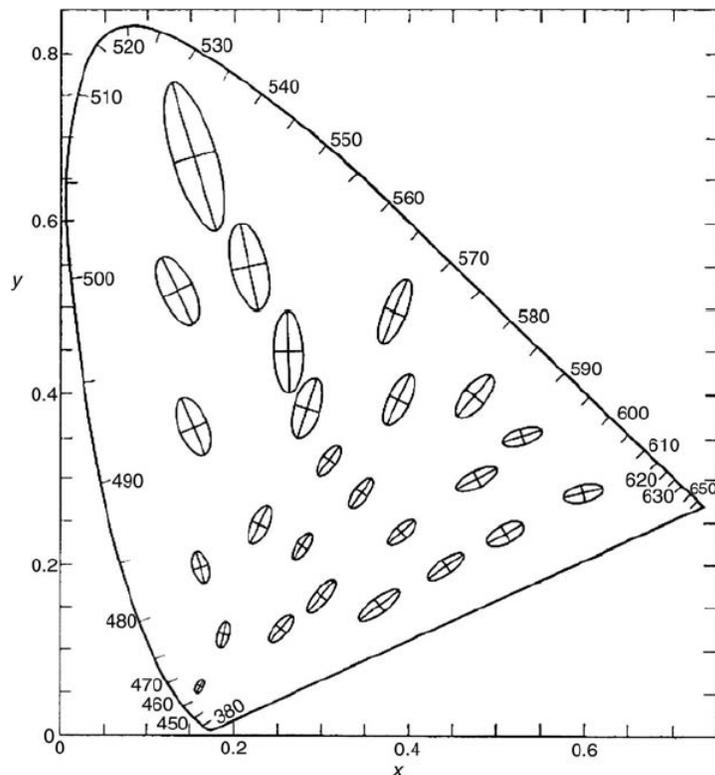
Figure 5.13: MacAdam ellipses 10-times enlarged.

Incandescent lamps

Incandescent lamps emit electromagnetic radiation by the **Joule effect**, which is the heating of a conductive material induced by the flow of electric current. The heating (typically infrared radiation) is accompanied with visible light and a small component of ultraviolet radiation. In Figure 5.14, we can see the typical incandescence light bulb.

The conductive filament generally used is given by **tungsten** and it was developed by George **Coolidge** of the General Electric Company (USA) in 1908 (before that time a very inefficient carbon filament was used).

During its functioning, the atoms of the filament sublimate (passing from solid to vapor state) and so the filament tails off becoming thinner and thinner until it breaks. For this reason, in the bulb it is introduced an inert gas, typically **argon** or **krypton**, to delay the filament evaporation

63

Figure 5.14: Traditional incandescent bulb.
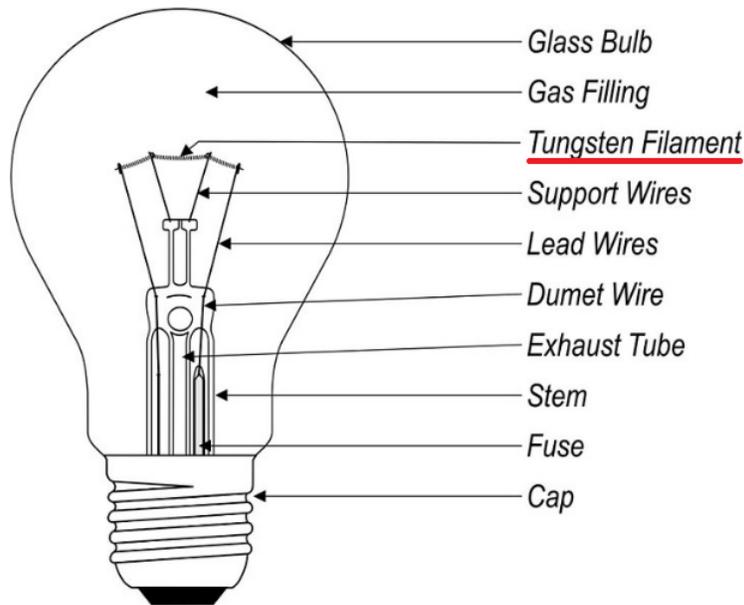
and reduce the blackening of the bulb due to the deposit of tungsten vapors. The spectrum of the traditional incandescent lamp is depicted in Figure 5.15. In can be seen that the long wavelengths are prevailing with respect to the middle and short ones, this is the reason why traditional incandescent lamps produce a light that is considered 'warm' (reddish).

These traditional incandescent lamps are disappearing from the market because of their high energetic cost in favor of halogen incandescent lamps, see Figure 5.16. Their functioning principle is the same as the traditional, what changes is the fact that the **double spiral tungsten filament** here is introduced in a **quartz bulb** filled with halogens (**iodine** and **bromine**), chemical elements that help the regeneration of the filament. In fact, after the sublimation, the tungsten atoms react with the halogen atoms to create halogen tungsten which then dissociates and deposits back to the filament, thus regenerating it, even if not perfectly.

To have this kind of chemical reactions we need the temperature of the filament to remain above 2000^0C, this is the reason why the transparent bulb of halogen incandescent light sources is made of quartz and not simple glass (which would be fused at that temperature).

The advantages of halogen incandescent lamps with respect to the tra-
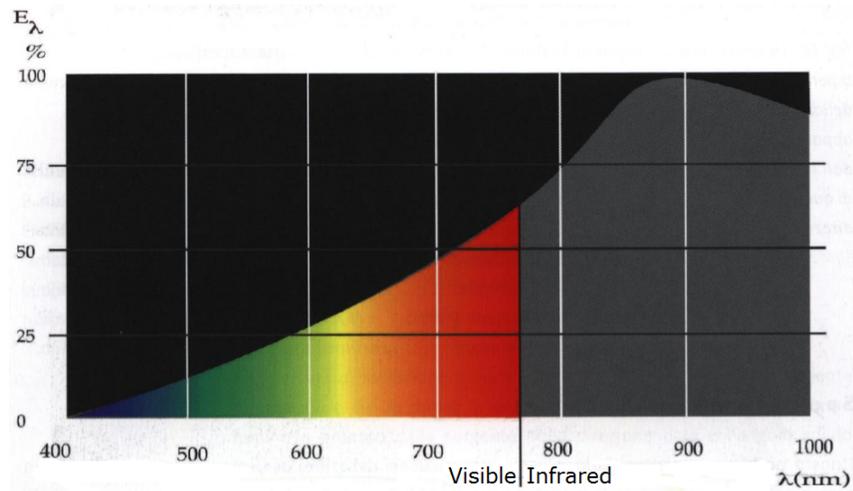
Figure 5.15: Traditional incandescent bulb spectrum.
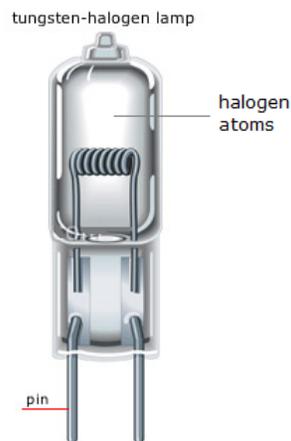
Figure 5.16: Halogen incandescent bulb.

ditional ones is that they last for longer time, the bulb doesn't blackens and the light produced has a more uniform spectrum, as can be seen in Figure 5.17.

The disadvantages consist in the fact that, due to the fact that they work at higher temperatures, they produce more energetic UV radiation, which, however, can be reduced by suitably treating the quartz bulb.

The infrared radiations of halogen lamps are also significant and this can
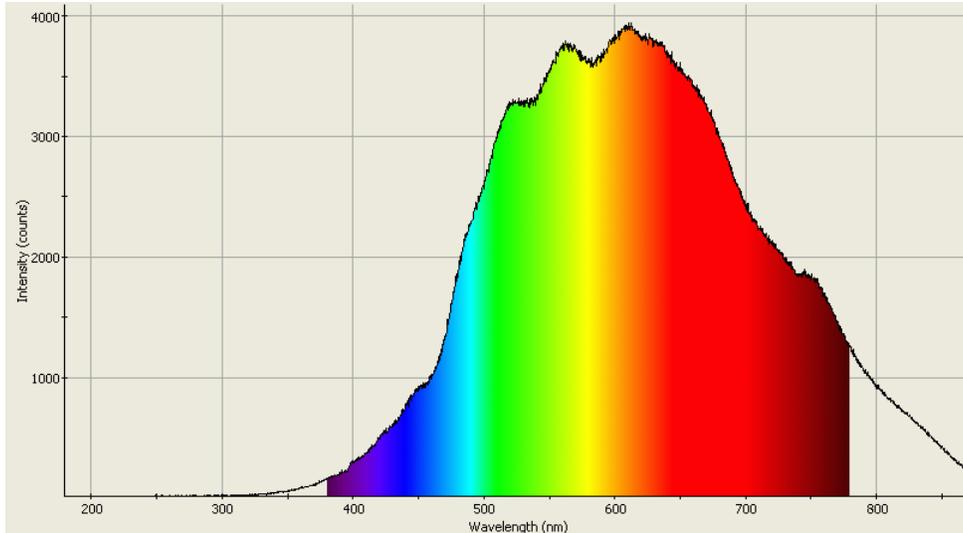
Figure 5.17: Halogen incandescent bulb spectrum.

create problems as, e.g. drying of materials. To reduce this problem, the so-called dichroic halogen lamps have been developed, see Figure 5.18.
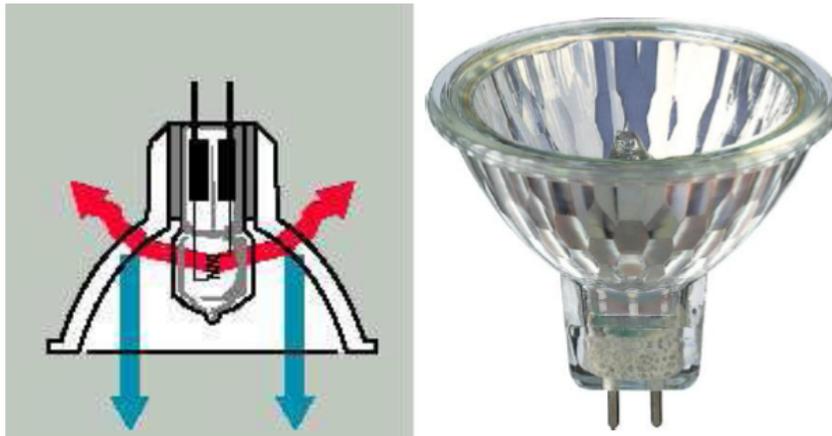
Figure 5.18: Halogen dichroic bulb.

The particular mirror configuration is able to scatter up to the 65% of the infrared radiation to the back of the lamp. Some more advanced dichroic lamps instead of scattering the infrared radiation to the back of the lamp, scatter it directly in the direction of the filament, which increases its

66

temperature and so it needs less electrical power to function.

Fluorescent lamps

A fluorescent lamp or fluorescent tube is a gas-discharge lamp that uses electricity to excite **mercury vapor** and argon, xenon, **neon**, or krypton. The excited mercury atoms produce short-wave *ultraviolet light that then causes a phosphor to fluoresce*, producing visible light. A fluorescent lamp converts electrical power into useful light much more efficiently than incandescent lamps. The scheme of a fluorescent lamp can be seen in Figure 5.19
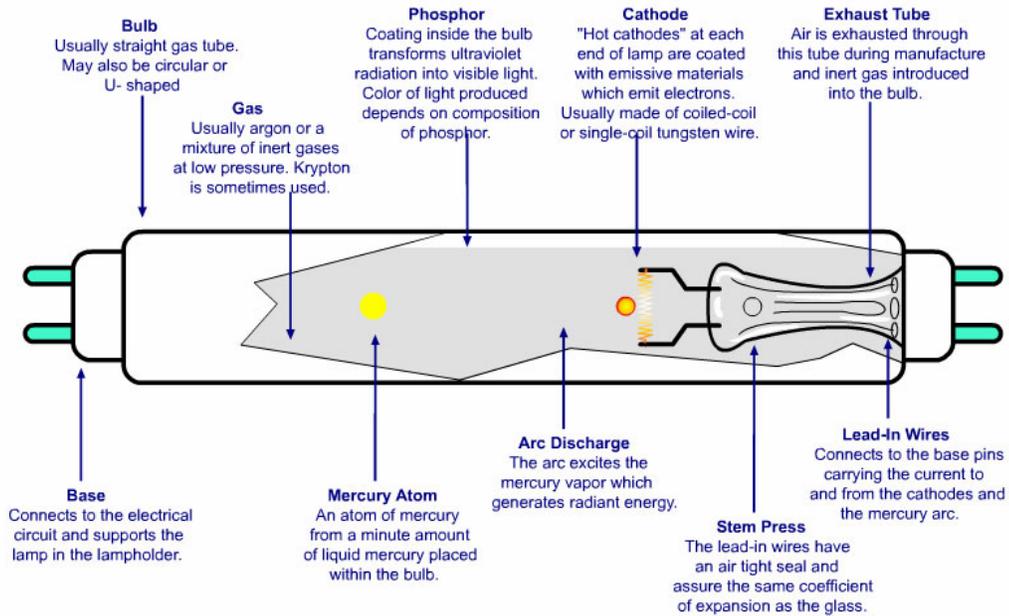
Figure 5.19: Fluorescent tube.

The spectrum of a fluorescent light source is not continuous, as for the incandescent light bulb, but is composed with spikes of radiations that cluster around well defined wavelengths, as can be seen in Figure 5.20. This is given by the fact that the photons produced have an energy that belongs to well defined ranges (the level gaps of the gas and the fluorescent material).

The principal merit of fluorescent light sources is that they are less energy-consuming than the incandescent sources and they do not produce UV rays. The major drawback is the lack of a continuous spectrum and the fact that they need more time than the incandescent sources to start

generating light.

Spectrum of Compact Fluorescent Light
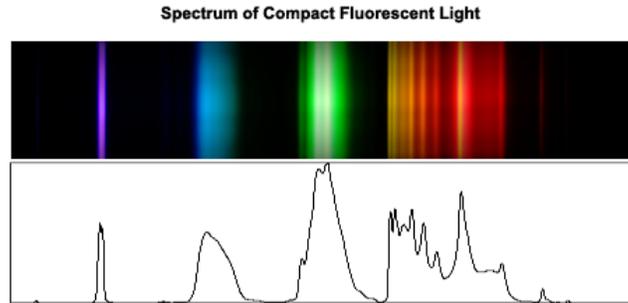
Figure 5.20: Fluorescent spectrum.

LEDs

LED is the acronym of Light Emitting Diode and it is based on the phenomenon of **electroluminescence**. To explain it we must first explain what is a diode, which requires the concept of semiconductor.

The most energetic electrons in solids occupy the so called valence and conduction bands, as in Figure 5.21. The valence electrons are bound to individual atoms, as opposed to conduction electrons (found in conductors and semiconductors), which can move freely within the atomic lattice of the material.
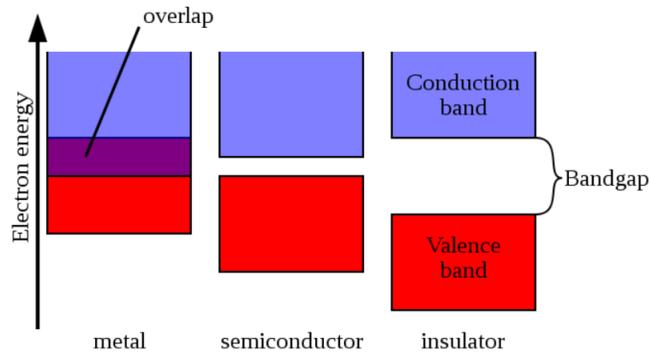
Figure 5.21: Valence and conduction bands in conductors, semiconductor and insulators.

On a graph of the electronic band structure of a material, the valence band is located below the conduction band, separated from it in insulators

and semiconductors by a band gap. In metals, the conduction band has no energy gap separating it from the valence band.

When a difference of potential is applied to a semiconductor, the valence electrons can 'win' the energy gap and pass to the conduction band. This passage is accompanied with the emission of a photon.

A diode is given by a p-n junction, i.e. an interface between two types of semiconductor material, **p-type** and **n-type**, inside a single crystal of semiconductor. It is created by **doping** a semiconductor creating an excess (n-type) or a lack (hole) of electrons (p-type).

When a light emitting diode is switched-on, **the electrons are able to recombine with electron holes** within the device, releasing energy in the form of photons. The **color of the light** (corresponding to the energy of the photon) is **determined by the energy gap of the semiconductor**.

The first diodes were able to produce only red and green/yellow light, but nowadays technology is able to produce also blue light, so that they can be combined, as in Figure 5.22, to produce different shades of white light, whose typical spectrum is depicted in Figure 5.23

Figure 5.22: Joining a red, green and blue led we can nowadays produce 'white' light.

The major advantages of LEDs is that they require much less electrical power to produce light than incandescent sources and that they can be built in order to avoid the generation of ultraviolet and infrared electromagnetic radiation. Their major drawback is their cost, which is still high.

5.4.2   Features of artificial light sources

The principal features of an artificial light source are:
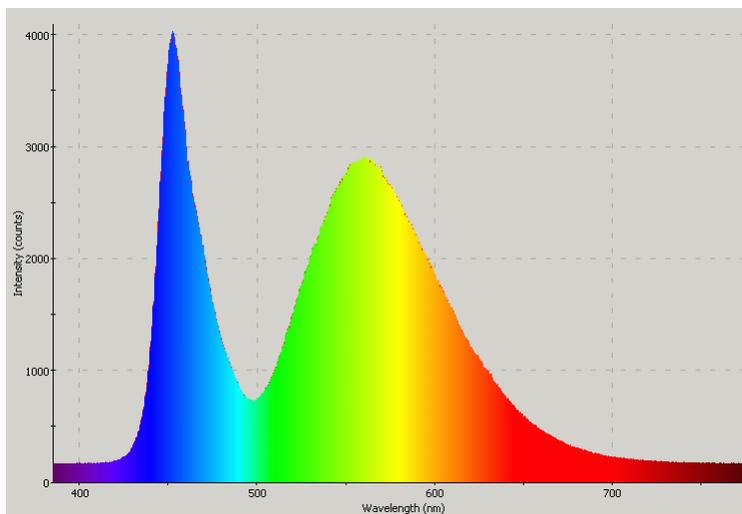
Figure 5.23: The typical spectrum of a 'white' light generated by a configuration of 3 LEDs.

- Color rendering index (CRI);

- Correlated color temperature (CCT);

- Luminous efficiency and average lifetime.

The **color rendering index** (CRI), sometimes called color rendition index, is a quantitative measure of the ability of a light source to reproduce the colors of various objects faithfully in comparison with an ideal or natural light source. Light sources with a high CRI are desirable in color-critical applications such as photography, cinematography and in museum lighting. Its CIE definition is the following: '*effect of an illuminant on the color appearance of objects by conscious or subconscious comparison with their color appearance under a reference illuminant*'.

To have a satisfying CRI the light source **must have in its spectrum all the visible wavelengths and they must be fairly homogeneously distributed**. The CRI of traditional incandescent and fluorescent lamps is quite poor because they tend to privilege the red and the green/blue part of the visual spectrum, respectively. The CRI of high temperature halogen lamps and LEDs instead are more satisfying.

The **correlated color temperature** (CCT) of an artificial light source is the temperature of an ideal black body radiator that radiates light of

70

comparable hue to that of the light source. Color temperatures over 5000K are called **cool colors** (blueish white), while lower color temperatures (2700-3000 K) are called **warm colors** (yellowish white through red).

It must be stressed, however, that this is a psychological relation and it is in contrast to the physical relation implied by Wien's displacement law seen in chapter 1, according to which the spectral peak is shifted towards shorter wavelengths (resulting in a more blueish white) for higher temperatures.

In Figure 5.24 we can see the so-called **Planckian locus**, i.e. the *color that an incandescent black body would take in the CIE chromaticity diagram as the black body temperature changes*. It goes from deep red at low temperatures through orange, yellowish white, white, and finally bluish white at very high temperatures. The straight segments represent points of constant CCT.
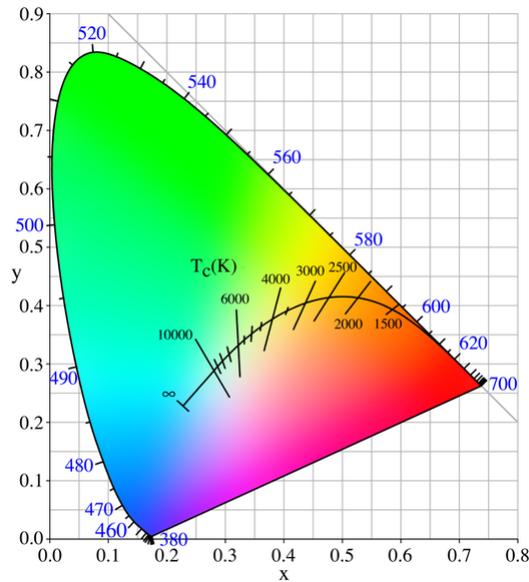
Figure 5.24: The Planckian locus.

Finally, the **luminous efficiency** and **average lifetime** of an artificial light source are economical features. The average lifetime is the number of continuous hours of functioning after which the 50% of a high number (the specification of this number may change) of light sources stop functioning. LEDs and fluorescent lamps have a high average lifetime, halogen incandescent lamps have increased a lot their average lifetime, while traditional incandescent lamps have the lowest lifetime.

The luminous efficiency is the ratio between the total luminous flux emitted by a device and the total amount of input electrical power it consumes. It is measured in lm/W. The highest luminous efficiency is that of LEDs and fluorescent lamps, then we have halogens and traditional lamps.

5.4.3 CIE illuminants

A standard illuminant is a theoretical source of visible light with a profile given by its spectral power distribution. Standard illuminants provide a basis for comparing images or colors recorded under different lighting and they are published by the CIE.

A list of the most commonly used standard illuminants follows:

- Illuminant **A**: CIE standard illuminant A is intended to represent typical, domestic, *tungsten-filament lighting*. Its relative spectral power distribution is that of a Planckian radiator at a temperature of approximately 2856 K. CIE standard illuminant A is used in applications of colorimetry involving incandescent lighting;

- Illuminants **B** and **C** are *daylight simulators*. They are derived from Illuminant A by using a liquid filters. B served as a representative of noon sunlight, with a correlated color temperature (CCT) of 4874 K, while C represented average day light with a CCT of 6774 K. They are poor approximations of any common light source and deprecated in favor of the D series. The profile of spectral power distribution of illuminants A,B and C is represented in Figure $\mathbb{R}$fig:abc;

- Illuminants **D** have been derived by Judd, MacAdam, and Wyszecki. The D series of illuminants are constructed to portray *natural daylight* illumination at open-air in different parts of the world. They represent a substantial improvement of the B and C series, in spite of being difficult to physically produce artificially, but are easy to characterize mathematically. In particular, the D65 (or D_{65}corresponds roughly to a *midday sun in Western/Northern Europe*, as any standard illuminant is represented as a table of averaged spectrophotometric data, so that **any light source which statistically has the same relative spectral power distribution can be considered a D65 light source**. There are no actual D65 light sources, only simulators. The number 65 refers to the fact that the CCT of the D65 illuminant is 6500 K (as the CCT of the D50 y 5000 K, and so on). The spectral power distribution of the D65 illuminant is depicted in Figure 5.26.

In Figure 5.27 we can see the position of white points of the standard illuminants in the CIE chromaticity diagram. Finally, in Figure 5.28 there's the list of chromaticity coordinates and 'color' of the white points of the standard illuminants.
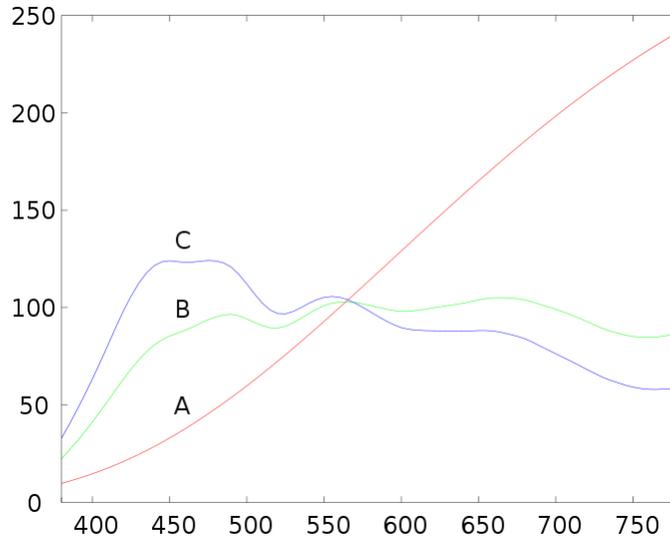
Figure 5.25: The A,B,C CIE standard illuminant spectral power distributions.

5.4.4 White balance in digital images: the von Kries trasformation

The source of light (natural or artificial) that we use to illuminate a scene directly affect the analogical or digital pictures that we take, as can be seen in Figure 5.29.

The overall red and green hue on the second and the third picture of Figure 5.29, respectively, is called **color cast** and it is due to the fact that the incandescent and fluorescent light sources have a scarce color rendering index. The first daylight picture instead doesn't have any cast because natural daylight has the highest color rendering index.

To (at least partially) eliminate the color cast, the easiest transformation that can be implemented on a RGB image is the so-called von Kries transformation, which simply consists in normalizing the tristimulus values over
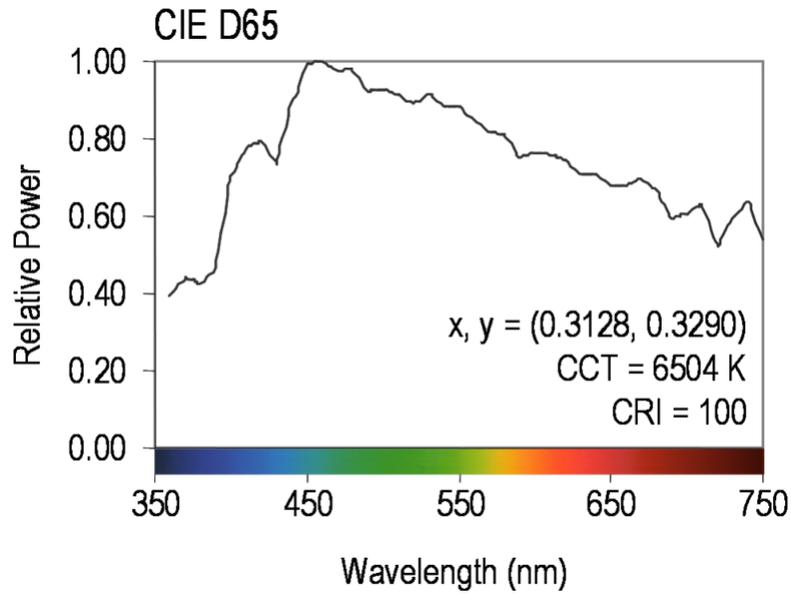
Figure 5.26: The D65 spectral power distributions.
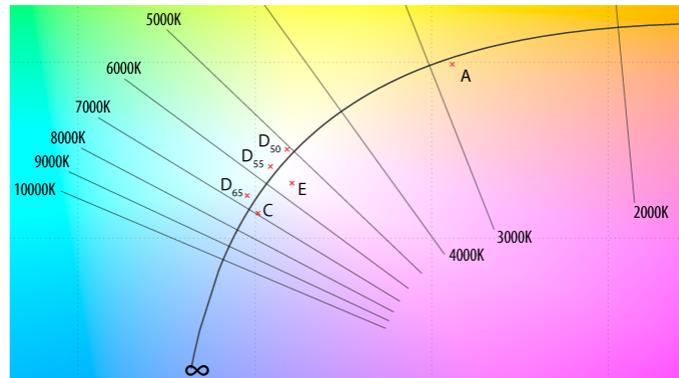
Figure 5.27: White points of standard CIE illuminants in the chromaticity diagram.

the maximum value found in the whole image, for each chromatic channel separately:

$$R_{\text{von Kries}} = \frac{R_{\text{Original}}}{R_{\text{max}}}, \ G_{\text{von Kries}} = \frac{G_{\text{Original}}}{G_{\text{max}}}, \ B_{\text{von Kries}} = \frac{B_{\text{Original}}}{B_{\text{max}}},$$

so that the highest tristimulus values become unitary: $R_{\text{max}} = G_{\text{max}} =$

74

<div align="center">**White points**</div>

Name	CIE 1931 2°		CIE 1964 10°		CCT (K)	Hue	Note
	x_2	y_2	x_{10}	y_{10}			
A	0.44757	0.40745	0.45117	0.40594	2856		Incandescent / Tungsten
B	0.34842	0.35161	0.34980	0.35270	4874		{obsolete} Direct sunlight at noon
C	0.31006	0.31616	0.31039	0.31905	6774		{obsolete} Average / North sky Daylight
D50	0.34567	0.35850	0.34773	0.35952	5003		Horizon Light. ICC profile PCS
D55	0.33242	0.34743	0.33411	0.34877	5503		Mid-morning / Mid-afternoon Daylight
D65	0.31271	0.32902	0.31382	0.33100	6504		Noon Daylight: Television, sRGB color space
D75	0.29902	0.31485	0.29968	0.31740	7504		North sky Daylight
E	1/3	1/3	1/3	1/3	5454		Equal energy
F1	0.31310	0.33727	0.31811	0.33559	6430		Daylight Fluorescent
F2	0.37208	0.37529	0.37925	0.36733	4230		Cool White Fluorescent
F3	0.40910	0.39430	0.41761	0.38324	3450		White Fluorescent
F4	0.44018	0.40329	0.44920	0.39074	2940		Warm White Fluorescent
F5	0.31379	0.34531	0.31975	0.34246	6350		Daylight Fluorescent
F6	0.37790	0.38835	0.38660	0.37847	4150		Lite White Fluorescent
F7	0.31292	0.32933	0.31569	0.32960	6500		D65 simulator, Daylight simulator
F8	0.34588	0.35875	0.34902	0.35939	5000		D50 simulator, Sylvania F40 Design 50
F9	0.37417	0.37281	0.37829	0.37045	4150		Cool White Deluxe Fluorescent
F10	0.34609	0.35986	0.35090	0.35444	5000		Philips TL85, Ultralume 50
F11	0.38052	0.37713	0.38541	0.37123	4000		Philips TL84, Ultralume 40
F12	0.43695	0.40441	0.44256	0.39717	3000		Philips TL83, Ultralume 30

Figure 5.28: Table of white points of standard illuminants.

$B_{\max} = 1$. The result of this trasformation can be seen in Figure 5.30.

Coherently with what just said, professional photographers use to shoot the first picture of an indoor scene to a perfectly reflecting white paper: if the sheet of paper is not white in the picture, then a suitable white balance must be performed.

The von Kries transformation must be used carefully: in Figure 5.31 it can be seen a negative effect of it (typical for low key images, i.e. pictures taken in dim light conditions).
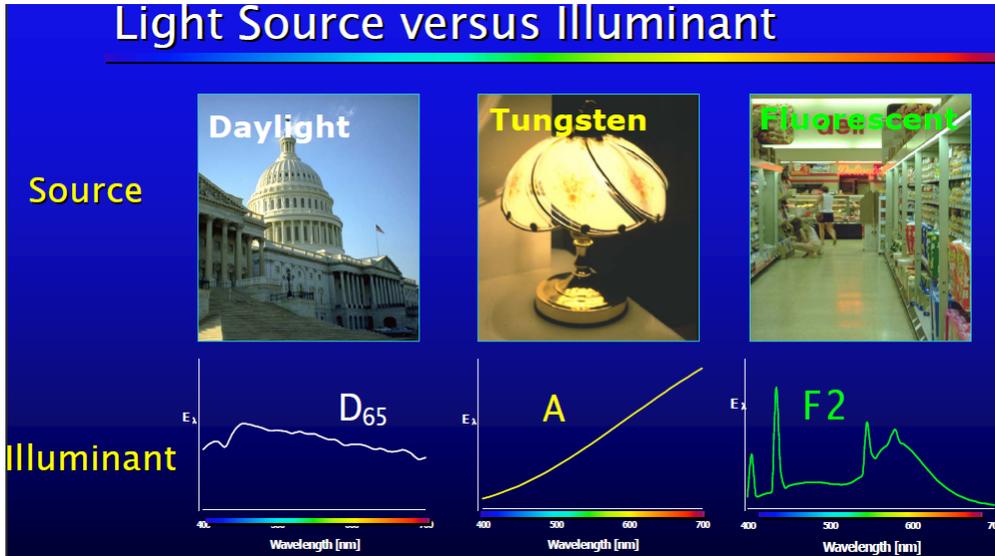
Figure 5.29: Different sources of light and how they affect digital pictures.

Figure 5.30: Effect of von Kries trasformations.
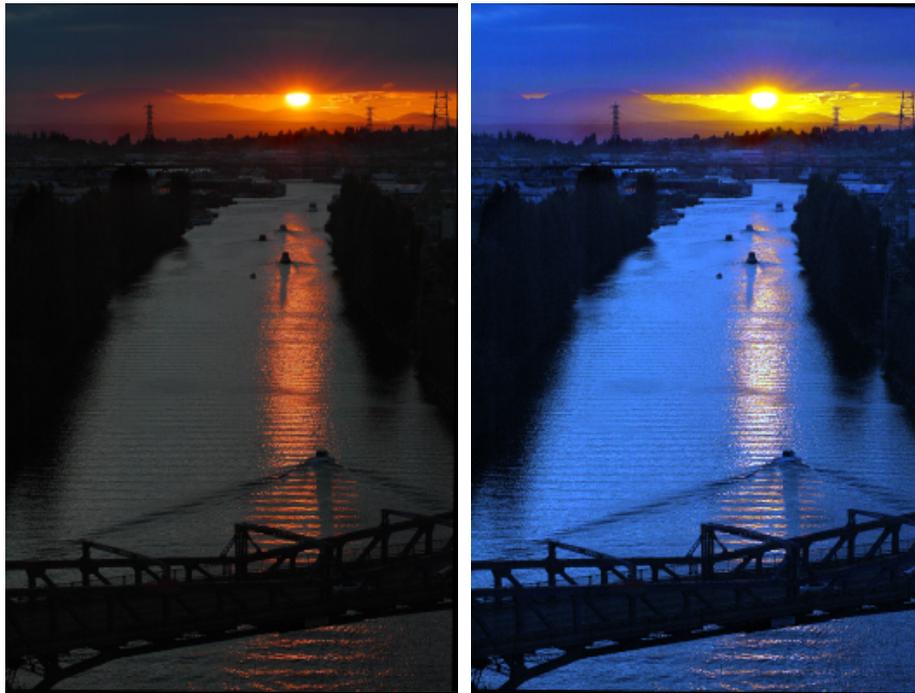
Figure 5.31: Nagative effect of von Kries trasformations. *Left*: original, *Right*: von Kries transformed.

Chapter 6

Beyond light acquisition: neurophysiological and phenomenological properties of the human visual system

The neural processing of visual information is quite complex within the retina and becomes enormously more complex at later stages. The aim of this chapter is to provide a brief overview of the paths that some of this information takes. It is helpful to begin with a general map of the steps along the way. The optical image on the retina is first transduced into chemical and electrical signals in the photoreceptors. These signals are then processed through the network of retinal neurons: horizontal, bipolar, amacrine, and ganglion cells.

The ganglion cell axons gather to form the optic nerve, which projects to the **lateral geniculate nucleus** (LGN) in the thalamus. The LGN cells, after gathering input from the ganglion cells, project to visual area one called V1 in the **occipital lobe** of the cortex.

At this point, the information processing begins to become amazingly complex. Approximately 30 visual areas have been defined in the cortex with names such as V2, V3, V4, MT, etc. Signals from these areas project to several other areas and vice versa. The cortical processing includes many instances of feed-forward, feed-back, and lateral processing. Somewhere in this network of information our ultimate perceptions are formed. There is still much work to do in order to have a clear understanding of our visual system.

6.1 The other retinal cells, receptive fields and brain vision areas

The electrical signal generated by rods and cones passes to horizontal, bipolar, amacrine and ganglion cells before arriving to the optical nerve.

Horizontal and amacrine cells laterally interconnect in a very complicated, and still not completely understood, way the output of photoreceptors and bipolar cells, respectively.

Ganglion cells are among the most studied retinal cells and what we know for sure up to now is that they spontaneously *fire* action potentials[1] at a base rate while at rest. **Excitation** of retinal ganglion cells results in an *increased firing rate* while **inhibition** results in a *depressed rate of firing*. For this reason the *magnitude of the signal* is represented in terms of the *number of spikes of voltage per second fired by the cell* rather than by the voltage difference across the cell wall.

To represent the physiological properties of these cells, the concept of **receptive fields** becomes useful. A receptive field is a graphical representation of the area in the visual field to which a given cell responds. In addition, the nature of the response (positive or negative) is typically indicated for various regions in the receptive field. As a simple example, the receptive field of a photoreceptor is a small circular area representing the size and location of that particular receptor's sensitivity in the visual field. Figure 6.1 represents some prototypical receptive fields for ganglion cells.
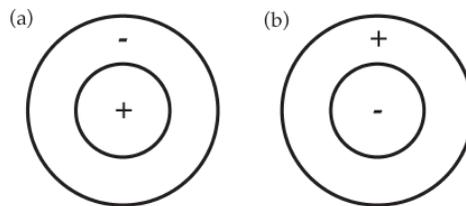
Figure 6.1: Typical center-surround antagonistic receptive fields: (a) on-center, (b) off-center.

They illustrate **center-surround antagonism**, which is characteristic at this level of visual processing. The receptive field in Figure 6.1(a) illustrates a positive central response, typically generated by a positive input

---

[1]In physiology, an *action potential* is a short-lasting event in which the electrical membrane potential of a cell rapidly rises and falls, following a consistent trajectory.

from a single cone, surrounded by a negative surround response, typically driven by negative inputs from several neighboring cones. Thus the response of this ganglion cell is made up of inputs from a number of cones with both positive and negative signs. The result is that the *ganglion cell does not simply respond to points of light, but serves as an edge detector* (actually a 'spot' detector). Using a digital image processing analogy, we can think of the ganglion cell responses as similar to the output of a convolution kernel designed for edge detection.

Figure 6.1(b) illustrates that a ganglion cell response of opposite polarity is equally likely. The response in Figure 6.1(a) is considered an **on-center** ganglion cell while that in Figure 6.1(b) is called an **off-center** ganglion cell.

Often on-center and off-center cells will occur at the same spatial location, fed by the same photoreceptors, resulting in an enhancement of the system's dynamic range. This is a crucial step that works in opposition to the glare and luminance contrast reduction provoked by the eye bulb.

Note that the ganglion cells represented in Figure 6.1 will have **no response to uniform fields** given that the positive and negative areas, in this case, would balance each other. This illustrates one aspect of the image compression carried out in the retina: *the brain is not bothered with redundant visual information*, **only information about changes in the visual world is transmitted**!

These changes are magnified in a local way (**local contrast enhancement**), as can be seen in the Mach bands effect shown in Figure 6.2 (left). As we approach the vertical band on the right the gray level appears lighter, as we approach the vertical band on the left the gray level appears darker, in spite of that the luminance value in each vertical gray band is constant, as can be seen in Figure 6.2 (right).

The excitation-inhibition processing can explain the Mach bands effect, consider Figure 6.3: in the situation 1 excitation and inhibition fire signals equally, so a uniform patch is perceived; in the situation 2 a part of the inhibition component of the receptive field is activated by a region of highest luminance and so it prevails, generating a sensation of darker gray; on the contrary, in the situation 3, a part of the inhibition component of the receptive field is activated by a region of lowest luminance and so it is dominated by the excitation component, which produces a sensation of lighter gray.

This spatial information processing in the visual system is the fundamental basis of the important impact of the background on color appearance.

Figure 6.4 shows that in addition to spatial opponency, there is often **spectral opponency in ganglion cell responses**. Figure 6.4(a) shows a
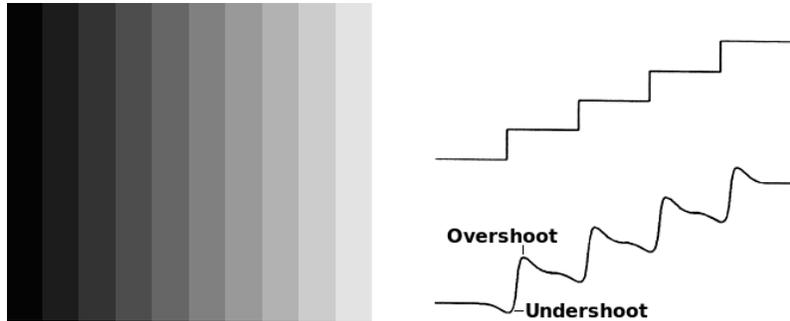
Figure 6.2: *Left*: Mach bands effect. *Right*: real and appearent luminance pattern.
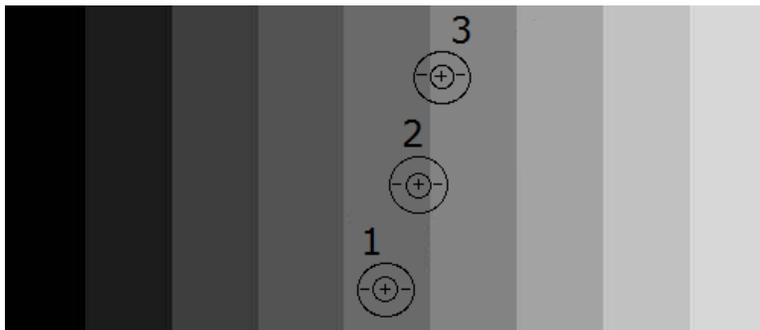
Figure 6.3: Excitation/inhibition explaination of the Mach band effect.

red-green opponent response with the center fed by positive input from an L cone and the surround fed by negative input from M cones. Figure 6.4(b) illustrates the off-center version of this cell.
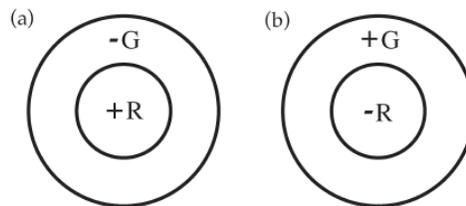
Figure 6.4: Examples of (a) red-green and (b) green-red spectrally and spatially antagonistic receptive fields.

Thus, *before the visual information has even left the retina, processing*

has occurred with a profound affect on color appearance, as we will see in the next sections.

As we said, when the visual signal reaches the lateral geniculate nucleus, it is projected to the primary visual area in the brain called V1 and there the visual path becomes enormously intricate. We just mention that the areas involved in color processing are mainly the V1 and V4 (also used for orientation and spatial frequency). A large part of these areas is dedicated to the process of the foveal signal, proving again the significance of foveal vision over extra-foveal one. Figure 6.5 shows the area dedicated to processing of the visual signal in the brain.
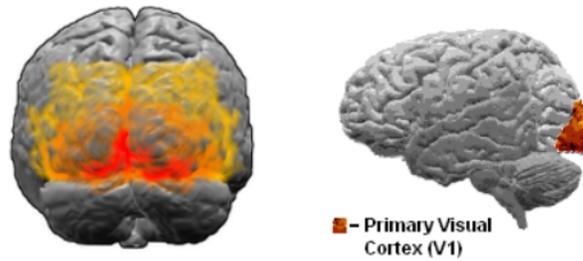
Figure 6.5: The visual cortex.

We conclude this section with the very schematic description of the visual path presented in Figure 6.6.
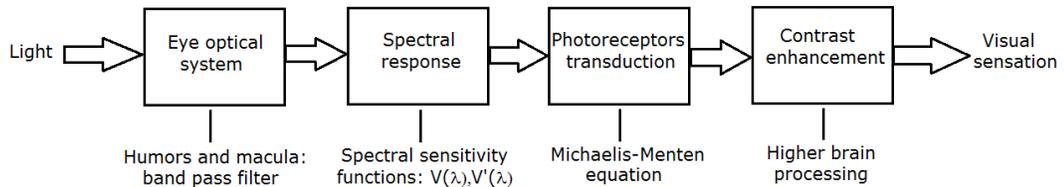
Figure 6.6: Simplified Human Visual System model.

6.2 Phenomenological mechanisms of color vision

Historically, there have been many theories that attempt to explain the function of color vision and its phenomenological features. In this chapter we will present only the most important insights about color vision, some of these will be developed in more detail in the next chapters.

6.2.1 Trichromatic theory

In the later half of the 19th century, the trichromatic theory of color vision was developed, based on the work of **Maxwell**, **Young**, and **Helmholtz**. They recognized that there must be three types of receptors, approximately sensitive to the red, green, and blue regions of the spectrum, respectively. The trichromatic theory simply assumed that three images of the world were formed by these three sets of receptors and then transmitted to the brain where the ratios of the signals in each of the images was compared in order to sort out color appearances. The trichromatic (three-receptor) nature of color vision has been 'reinforced' by the so-called **Retinex theory** of Land and McCann (1964).

6.2.2 Hering's opponent-colors theory

At around the same time, Hering proposed an opponent-colors theory of color vision based on many subjective observations of color appearance. Hering noted that certain hues were never perceived to occur together. For example, a color perception is never described as **reddish-green** or **yellowish-blue**, while combinations of red and yellow, red and blue, green and yellow, and green and blue are readily perceived.

This suggested to Hering that there was *something fundamental about the red-green and yellow-blue pairs* causing them to oppose one another. Similar observations were made of **colored simultaneous contrast** (Figure 6.7) in which objects placed on a red background appear greener, on a green background appear redder, on a yellow background appear bluer, and on a blue background appear yellower.

Hering proposed that there were three types of receptors, but Hering's receptors had bipolar responses to **light-dark**, **red-green**, and **yellow-blue**. At the time, this was thought to be physiologically implausible and Hering's opponent theory did not receive appropriate acceptance.

In the middle of the 20th century, Hering's opponent theory enjoyed a revival when quantitative data supporting it began to appear. Figure 6.8 illustrates that the first stage of color vision, the receptors, is indeed trichromatic as hypothesized by Maxwell, Young, and Helmholtz. However, contrary to simple trichromatic theory, the three 'color-separation' images are not transmitted directly to the brain. Instead the neurons of the retina (and perhaps higher levels) encode the color into opponent signals. The outputs of all three cone types are summed $(L + M + S)$ to produce an achromatic response that matches the CIE $V(\lambda)$ curve as long as *the sum-*
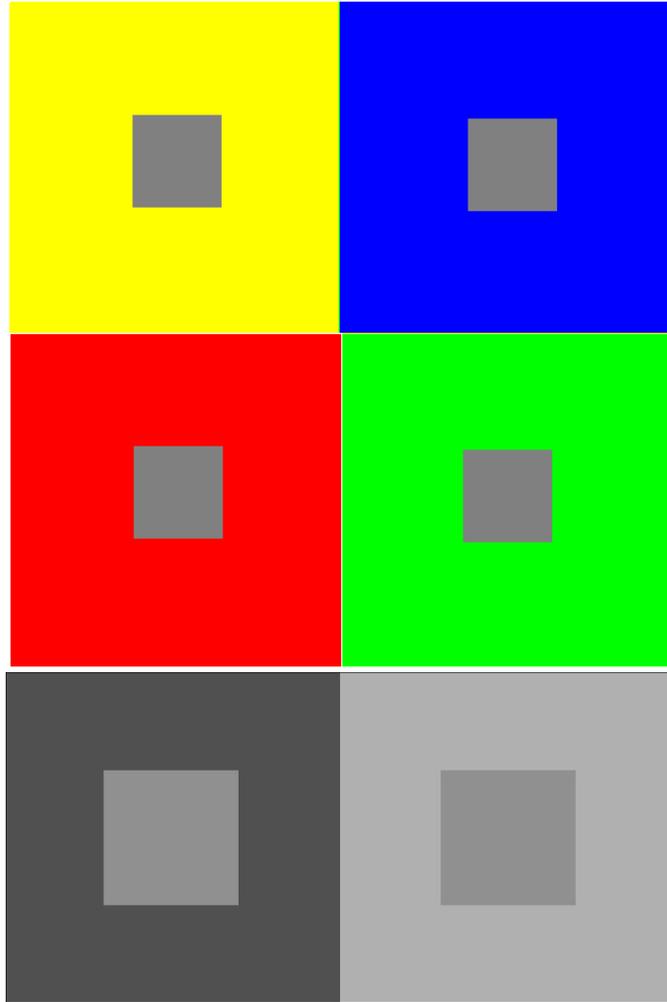
Figure 6.7: *Top*: Colored simultaneous contrast. *Bottom*: Grayscale simultaneous contrast. In both pictures, the inner gray squares have exactly the same physical luminance, however, their perceived luminance is very different.

mation is weighted with the relative populations of the three cone types.

Differentiating the cone signals allows construction of red-green ($L - M + S$) and yellow-blue ($L + M - S$) opponent signals. The transformation from LMS signals to the opponent signals serves to decorrelate the color information carried in the three channels, thus allowing **more effi-**

cient signal transmission and reducing difficulties with noise. The three opponent pathways also have distinct spatial and temporal characteristics that are important for predicting color appearance. The importance of the transformation from trichromatic to opponent signals for color appearance is reflected in the prominent place that it finds within the formulation of many color appearance models. Figure 6.8 includes a schematic diagram of the neural 'wiring' that produces opponent responses.
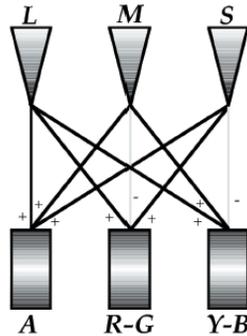
Figure 6.8: Schematic illustration of the encoding of cone signals into opponent colors signals in the human visual system.

6.2.3 Color vision deficiencies

Some color vision deficiencies are caused by the lack of a particular type of cone photopigment. Since there are three types of cone photopigments, there are three general classes of these color vision deficiencies, namely protanopia, deuteranopia, and tritanopia.

- An **protanope**, **is missing the L-cone photopigment** and therefore is unable to discriminate reddish and greenish hues since the red-green opponent mechanism cannot be constructed;

- A **deuteranope** is **missing the M-cone photopigment** and therefore also cannot distinguish reddish and greenish hues due to the lack of a viable red-green opponent mechanism. Protanopes and deuteranopes can be distinguished by their relative luminous sensitivity since it is constructed from the summation of different cone types. The protanopic luminous sensitivity function is shifted toward shorter wavelengths;

- A **tritanope** is **missing the S-cone photopigment** and therefore **cannot discriminate yellowish and bluish hues** due to the lack of a yellow-blue opponent mechanism.

Why the disparity between the occurrence of color vision deficiencies in males and females? This can be traced back to the genetic basis of color vision deficiencies. It turns out that the most common forms of color vision deficiencies are sex-linked genetic traits.

The genes for photopigments are present on the X chromosome. Females inherit one X chromosome from their mother and one from their father. Only one of these need have the genes for the normal photopigments in order to produce normal color vision. On the other hand, males inherit an X chromosome from their mother and a Y chromosome from their father.

If the single X chromosome does not include the genes for the photopigments, the son will have a color vision deficiency. If a female is color deficient, it means she has two deficient X chromosomes and all male children are destined to have a color vision deficiency. It is clear that the genetic 'deck of cards' is stacked *against males* when it comes to inheriting deficient color vision. Knowledge regarding the genetic basis of color vision has grown tremendously in recent years.

6.2.4 Adaptation mechanisms

It is not enough to consider the processing of color signals in the human visual system as a static 'wiring diagram'. The dynamic mechanisms of adaptation that serve to optimize the visual response to the particular viewing environment at hand must also be considered. Thus an overview of the various types of adaptation is in order. Of particular relevance to the study of color appearance are the mechanisms of dark, light, and chromatic adaptation.

Dark adaptation

Dark adaptation refers to the change in visual sensitivity that occurs when the prevailing level of illumination is decreased, such as when walking into a darkened theater on a sunny afternoon. At first the entire theater appears completely dark, but after a few minutes one is able to clearly see objects in the theater such as the aisles, seats, and other people. This happens because the visual system is responding to the lack of illumination by becoming more sensitive and therefore capable of producing a meaningful visual response at the lower illumination level.

Figure 6.9: Images illustrating the color discrimination capabilities that are missing from observers with various color vision deficiencies: (a) original images, (b) protanope, (c) deuteranope, (d) tritanope.

Figure 6.10 shows the recovery of visual sensitivity (decrease in threshold) after transition from an extremely high illumination level to complete darkness. At first, the cones gradually become more sensitive until the curve levels off after a couple of minutes. Then, until about 10 minutes have passed, visual sensitivity is roughly constant. At that point, the rod system, with a longer recovery time, has recovered enough sensitivity to outperform the cones and thus the rods begin controlling overall sensitivity. The rod sensitivity continues to improve until it becomes asymptotic after about 30 minutes.

The five-fold change in pupil diameter is not sufficient to serve vision over the large range of illumination levels typically encountered. Therefore, **neural mechanisms must produce some adaptation**.
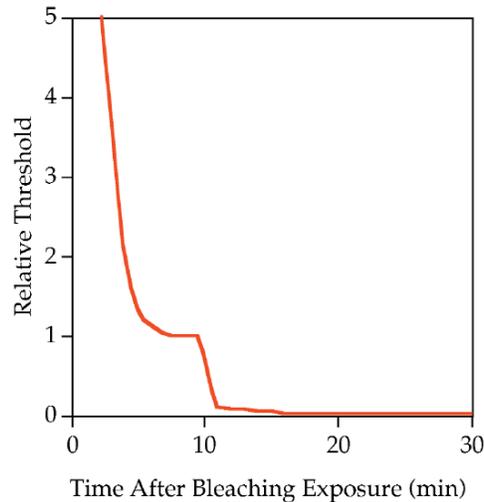
Figure 6.10: Dark-adaptation curve showing the recovery of threshold after a bleaching exposure. **The break in the curve illustrates the point at which the rods become more sensitive than the cones**.

Light adaptation

Light adaptation is essentially the inverse process of dark adaptation. However, it is important to consider it separately since its visual properties differ. Light adaptation occurs when leaving the darkened theater and returning outdoors on a sunny afternoon. In this case, the visual system must become less sensitive in order to produce useful perceptions since there is significantly more visible energy available.

The same physiological mechanisms serve light adaptation, but there is an asymmetry in the forward and reverse kinetics resulting in the time course of light adaptation being on the order of 5 minutes rather than 30 minutes. Figure 1.15 illustrates the utility of light adaptation.

As we have seen in Section 4.1.2, the **retinal system** has a limited **output** dynamic range of about **2 orders of magnitude**, say 100:1, available for the signals that produce our perceptions, as we have seen discussing the Michaelis-Menten's equation (4.1). The **world** in which we function, however, includes illumination levels covering at least **10 orders of magnitude** from a starlit night to a sunny afternoon. Fortunately, it is almost never important to view the entire range of illumination levels at the same time.

If a single response function were used to map the large range of stimulus intensities into the visual system's output, then only a small range of the

88

available output would be used for any given scene. Such a response is shown by the dashed line in Figure 6.11. Clearly, with such a response function, the perceived contrast of any given scene would be limited and visual sensitivity to changes would be severely degraded due to signal-to-noise issues.
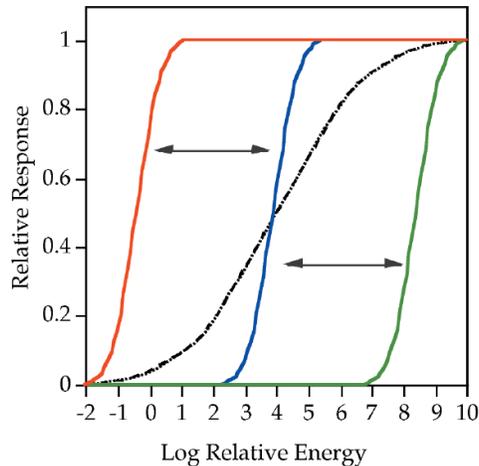
Figure 6.11: Illustration of the process of light adaptation whereby a very large range of stimulus intensity levels can be mapped into a relatively limited response dynamic range. Solid curves show a family of adapted responses. Dashed curve shows a hypothetical response with no adaptation.

On the other hand, light adaptation serves to produce a family of visual response curves which represent the Michaelis-Menten retinal response to light stimuli as illustrated by the solid lines in Figure 6.11. These curves map the useful illumination range in any given scene into the full dynamic range of the visual output, thus resulting in the best possible visual perception for each situation. **Light adaptation can be thought of as the process of sliding the visual response curve along the illumination level axis** in Figure 6.11 **until the optimum level for the given viewing conditions is reached**.

Light and dark adaptation can be thought of as **analogous to an automatic exposure control in a photographic system**.

Chromatic adaptation

Chromatic adaptation can be observed by examining a white object, such as a piece of paper, under various types of illumination (e.g., daylight, flu-

orescent, and incandescent). Daylight contains relatively far more short-wavelength energy than fluorescent light, and incandescent illumination contains relatively far more long-wavelength energy than fluorescent light. However, the paper approximately retains its white appearance under all three light sources. This is because the S-cone system becomes relatively less sensitive under daylight to compensate for the additional short-wavelength energy and the L-cone system becomes relatively less sensitive under incandescent illumination to compensate for the additional long-wavelength energy.

Chromatic adaptation is the largely independent sensitivity control of the three mechanisms of color vision. This is illustrated schematically in Figure 6.12, which shows that the overall height of the three cone spectral responsivity curves can vary independently. While chromatic adaptation is often discussed and modeled as independent sensitivity control in the cones, there is no reason to believe that it does not occur in opponent and other color mechanisms as well.
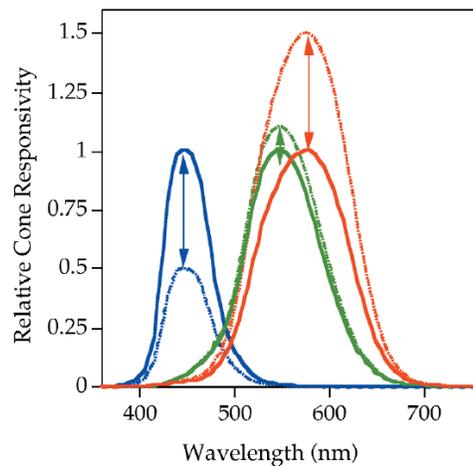
Figure 6.12: Conceptual illustration of the process of chromatic adaptation as the **independent sensitivity regulation of the three cone responsivities**.

Chromatic adaptation can be thought of as **analogous to an automatic white-balance in photo or video cameras**. Figure 6.13 provides a visual demonstration of chromatic adaptation in which the two halves of the visual field are conditioned to produce disparate levels of chromatic adaptation.
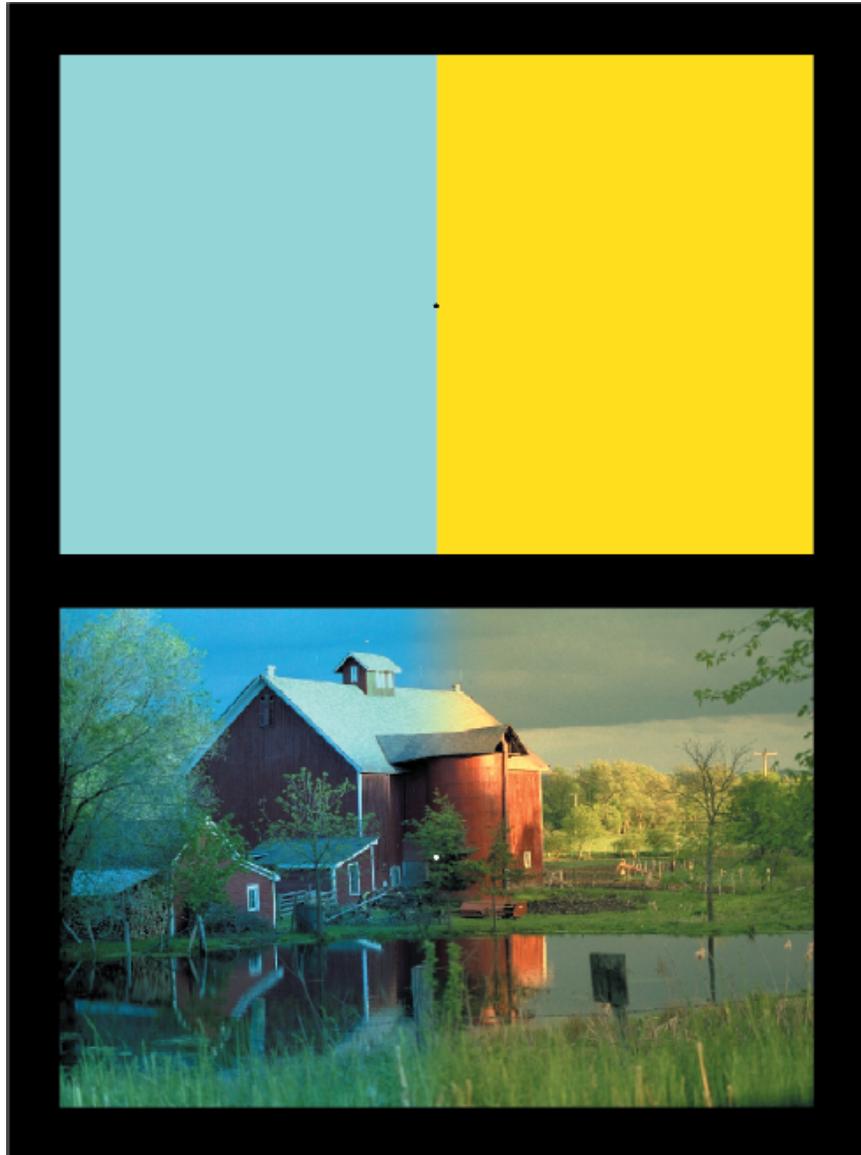
Figure 6.13: A demonstration of retinally localized chromatic adaptation. Fixate the black spot in between the uniform blue and yellow areas for about 30 seconds then shift your gaze to the white spot in the center of the barn image. Note that the barn image appears approximately uniform because we have adapted to these new illumination conditions.

6.3 Visual Mechanisms Impacting Color Appearance

There are many important cognitive visual mechanisms that impact color appearance. They include memory color, color constancy, discounting the illuminant, and object recognition:

- **Color memory** refers to the phenomenon that recognizable objects often have a prototypical color that is associated with them. For example, most people have a memory for the typical color of green grass and can produce a stimulus of this color if requested to do so in an experiment. Interestingly, the memory color often is not found in the actual objects. For example, green grass and blue sky are typically remembered as being more saturated than the actual stimuli;

- **Color constancy** refers to the everyday perception that the colors of objects remain unchanged across significant changes in illumination color and luminance level. Color constancy is served by the mechanisms of chromatic adaptation and memory color;

- **Discounting the illuminant** refers to an observer's ability to automatically interpret the illumination conditions and perceive the colors of objects after discounting the influences of illumination color;

- **Object recognition** is generally driven by the spatial, temporal, and light-dark properties of the objects rather than by chromatic properties.

There is no better phenomenon to show how color memory and object recognition influence our perception as the famous (shocking) Adelson's chessboard optical illusion (Figure 6.14).

6.3.1 Physical and perceived contrast: Weber-Fechner's and Stevens' laws

As we have seen, the eye's optical system and the response of photoreceptors strongly reduces the range of light intensity that can be processed. To compensate this reduction the HVS has developed a system to enhance contrast perception already in the retina with lateral inhibition and further in the brain with higher perceptual features. For this reason it is necessary to **distinguish between physical and perceived contrast**.
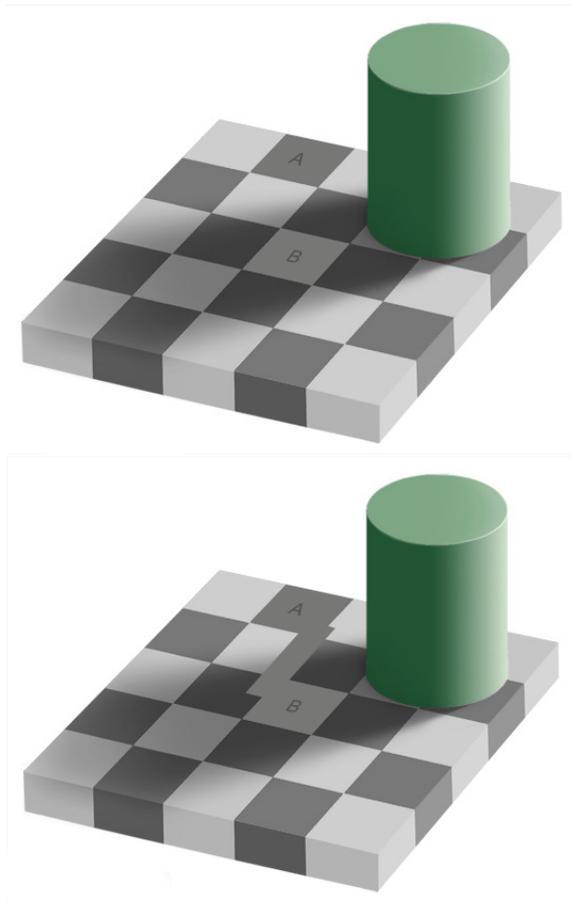
Figure 6.14: Adelson's chessboard: the squares A and B have exactly the same luminance, but color memory, object recognition and spatial processing make us perceive them as very different.

There are plenty of definitions of physical contrast, perhaps the most used and simple is the **Michelson definition of physical contrast**: let I_1 and I_2 be the luminous intensity coming from two adjacent areas, then Michelson's contrast is defined as the ratio between the absolute difference among the two intensities and their arithmetic mean, i.e.

$$C_M(I_1, I_2) = \frac{|I_1 - I_2|}{I_1 + I_2},$$

sometimes a factor of 2 is introduced to restore a proper arithmetic mean at

the denominator. Michelson's definition is interesting because it defines the *contrast relative to the average intensity*, which, as we will see in what follows, is much more important than the absolute difference among luminous intensities. Also notice that we can define Michelson's contrast as

$$C_M(I_1, I_2) = \frac{\max(I_1, I_2) - \min(I_1, I_2)}{\max(I_1, I_2) + \min(I_1, I_2)}.$$

Psychophysics is the science that aims at **modeling** in a mathematically rigorous way **the magnitude of human perception in response of external stimuli**. Typically, there is a high non-linear relationship between external stimuli and human response. Contrast perception of light stimuli is a very important example of such relationships.

The German physicist Ernst **Weber** (1795-1878) in the second half of the 19th century was one of the first scientists in history to develop some psychophysical experiments to test contrast perception in a very constrained setting: a dark-adapted human observer was put in a dim room in front of a white screen on which a narrow beam of light was thrown in the center of the visual field. The luminous intensity I of the beam was increased very slowly and the observer was asked to tell whether he/she could perceive an intensity change. Nowadays we call the least perceptible intensity change ΔI the **JND** for *Just Noticeable Difference*.

Weber found out that the JND increased proportionally with the luminous intensity[2], i.e.

$$\textbf{Weber's law} \qquad \boxed{\frac{\Delta I}{I} = K},$$

i.e. $\Delta I = K \cdot I$, and $K \simeq 0.08$ is called **Weber's constant**. Weber's law says that, as we increase the background light I, the difference ΔI must increase proportionally in order to be able to appreciate $I + \Delta I$ as different from I. This partially explains why we are more sensitive to noise in dark areas of a visual scene and thus why it is more important to perform a good denoising in dark areas of digital images rather than in bright ones. This last consideration is a practical application of a psychophysical phenomenon.

The founder of psychophysics, the German experimental psychologic Gustav Fechner (1801-1887), gave the following interpretation of Weber's law: he introduced the adimensional quantity $s(I)$ called light **sensation**

---

[2]Weber's law is approximately valid not only for the visual sense, but also for all the other senses, with different values of Weber's constant.

and stated that the difference of sensation $\Delta s(I)$ is proportional to a slightly modified Weber's ratio, i.e.

$$\Delta s(I) = k\frac{\Delta I}{n + I}$$

where $k > 0$ is a constant and $n > 0$ is a quantity often interpreted as *internal noise* in the visual mechanism. Fechner transformed this finite difference equation into a differential equation and integrated it to obtain an analytical law for s:

$$ds(I) = k\frac{dI}{n + I}$$

integrating both sides from I_0, the threshold above which luminous intensity is perceivable, i.e. such that $s(I_0) = 0$ and $s(I_0 + \varepsilon) > 0$ for all $\varepsilon > 0$, to a generic value of I, we obtain

$$\int_{I_0}^{I} ds(I) = k\int_{I_0}^{I} \frac{dI}{n + I} \iff s(I) - s(I_0) = k[\log(n + I) - \log(n + I_0)],$$

by using the properties of the logarithm and using the fact that $s(I_0) = 0$ we get the so-called **Weber-Fechner's law**:

$$\boxed{s(I) = k\log\left(\frac{n + I}{n + I_0}\right).}$$

Weber-Fechner's law says that **the sensation of luminous differences**, in the very constrained context of Weber's experiment, **grows at the logarithm of the luminous intensity**.

We must stress the **limitations of Weber-Fechner's law**:

1. Firstly it is valid only for very simple visual scenes, as that considered by Weber in his experiment, because, as we will see soon, **the presence of a non-trivial context changes completely contrast perception**;

2. Secondly, even for very simple visual scenes, we must stress that Fechner's assumption that we can maintain the validity of Weber's law **passing from finite to infinitesimal light intensity differences** is correct only for luminous intensities intermediate between the minimum and the maximum perceivable light. As we approach these extreme situations, this assumption fails dramatically due to strong non-linearities in the visual mechanism, so that Weber-Fechner's law doesn't hold anymore.

In 1957, the American psychophysicist Stanley Smith Stevens (1906-1973), proposed a power-law to express sensation to external stimuli (not only light). He collected magnitude estimation data from multiple observers, averaged the data across subjects, and then fitted a power function to the data. Because the fit was generally reasonable, he concluded a power law such as the following was correct:

$$\boxed{s(I) = k_S I^{\gamma}} \qquad \text{Stevens' law,}$$

$k_S > 0$ being a suitable constant and γ being a suitable exponent that changes for every sense. For example, $\gamma = 0.33$ for a 5 degrees light stimuli shown in a dark room and it changes if these conditions are changed.

Stevens' methodology is different than Weber's one. The principal methods used by Stevens to measure the perceived intensity of a stimulus were *magnitude estimation and magnitude production.* In magnitude estimation with a standard, the experimenter presents a stimulus called a standard and assigns it a number called the modulus. For subsequent stimuli, subjects report numerically their perceived intensity relative to the standard so as to preserve the ratio between the sensations and the numerical estimates (e.g., a light perceived twice as bright as the standard should be given a number twice the modulus). In magnitude estimation without a standard (usually just magnitude estimation), subjects are free to choose their own standard, assigning any number to the first stimulus and all subsequent ones with the only requirement being that the ratio between sensations and numbers is preserved. This procedure is criticized by many psychophysicists because it ignores any individual differences that may obtain and indeed it has been reported that the power relationship does not always hold as well when data are considered separately for individual respondents.

It is worthwhile finishing this section by stating that both Weber-Fechner's and Stevens' laws just give a rough average idea of how we perceive changes in luminous intensity. In Weber's experimental condition it is certain that sensation is a concave monotonically increasing function of light stimulus, such as logarithm or a power law with an exponent between 0 and 1.

However, it is **very rare to find Weber's experimental conditions in real life**, where the importance of the context, plays a fundamental role.

Chapter 7

Histograms of color images: variational equalization and perceptually inspired contrast enhancement

In this chapter we will recall the classical histogram equalization processing and then introduce a variational version of this transformation. This will give us the opportunity to discuss recent advances in perceptual contrast enhancement of color images.

Before this discussion it is worthwhile introducing the notation that we are going to use from now on to manipulate an RGB color image, i.e. an image whose intensity in the three chromatic channels are related to the tristimulus values RGB.

We will denote by $\Omega \subset \mathbb{Z}^2$ its **spatial domain**, $|\Omega|$ the number of pixels and by $x \equiv (x_1, x_2)$ and $y \equiv (y_1, y_2)$ the coordinates of two arbitrary pixels in $\mathfrak{I}$. We will always consider a **normalized dynamic range** in $[0, 1]$ (simply by dividing and multiplying for 255 before and after the computation, respectively), so that a **color image function** will be denoted with

$$\vec{I}: \quad \Omega \quad \longrightarrow \quad [0, 1] \times [0, 1] \times [0, 1]$$
$$x \quad \mapsto \quad (I_R(x), I_G(x), I_B(x))$$

where each scalar component $I_c(x)$ defines the intensity level of the pixel $x \in \Omega$ in the Red, Green and Blue channel, respectively.

We stress that we will perform every computation on the scalar components of the image, thus **treating each chromatic component sepa-**

rately. Therefore, we will avoid the subscript c and write simply $I(x)$ to denote the intensity of the pixel x in a given chromatic channel.

7.1 Classical histogram equalization processing

Let us recall what is the histogram of a 'continuous' digital image: let $\lambda \in [0, 1]$ be a generic intensity level, then the normalized **histogram** of I computed in λ, $h(\lambda)$, is:

$$h(\lambda) = \frac{1}{|\Omega|} \, \text{Area}\{x \in \Omega \mid I(x) = \lambda\} \qquad \lambda \in [0, 1],$$

i.e. the **occurrence probability** of the level λ in the image, that is, how many times the level λ appears in the image.

In Figure 7.1 we can see two examples of histograms.

Figure 7.1: Histograms of a low contrast (first column) high contrast (second column) image.

The normalized **cumulative histogram** of I computed in λ, $H(\lambda)$, is:

$$H(\lambda) = \frac{1}{|\Omega|} \, \text{Area}\{y \in \Omega \mid I(x) \leq \lambda\} \qquad \lambda \in [0, 1],$$

i.e. the **probability to find a pixel with intensity less than λ**.

Of course, the relationship between h and H is:

$$\boxed{H(\lambda) = \int_0^\lambda h(t)\, dt}, \quad \boxed{H'(\lambda) = h(\lambda)},$$

i.e. H **is the integral function of** h in the interval $[0, 1]$ and the first derivative of H in each level gives the histogram of that level.

It will be useful for later purposes to notice that the relationship $H(\lambda) = \int_0^\lambda h(t)\, dt$ can be re-written as follows

$$H(\lambda) = \int_0^\lambda h(t)\, dt = \int_0^1 \operatorname{sign}^+(\lambda - I(t))\, dt$$

where

$$\operatorname{sign}^+(\xi) = \begin{cases} 1 & \text{if } \xi \geq 0; \\ 0 & \text{if } \xi < 0. \end{cases}$$

and its 'spatial version'

$$H(I(x)) = \int_\Omega \operatorname{sign}^+(I(x) - I(y))\, dy \tag{7.1}$$

An image is said to be **equalized** if it has the *same occurrence probability for all levels*, i.e. if $h(\lambda) \equiv k$, the value of k can be determined by integrating h:

$$1 \underset{h \text{ is normalized!}}{=} \int_0^1 h(\lambda)\, d\lambda = \int_0^1 k\, ds = k \int_0^1 ds = k,$$

which implies $k = 1$ (1 must not be interpreted as a probability of 100%, but $\frac{100\%}{|\Omega|}$ since the histogram is normalized!).

The equalization condition on the histogram $h(\lambda) \equiv 1$ can be traduced to the condition $H(\lambda) = \int_0^\lambda d\lambda = \lambda$ on the cumulative histogram, as visualized in Figure 7.2
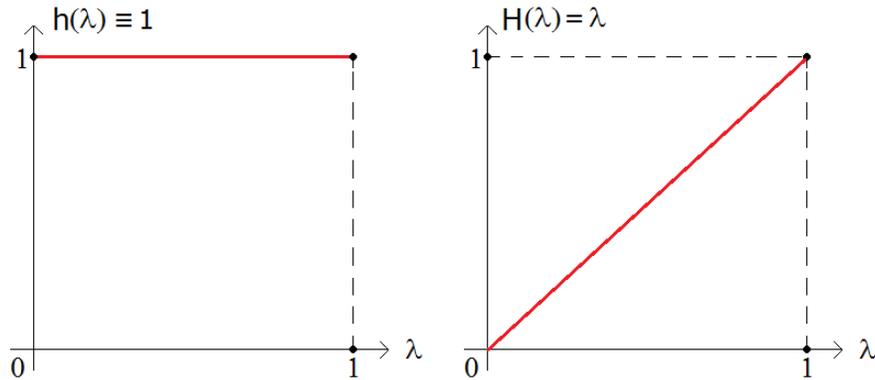

Figure 7.2: Histogram and cumulative histogram of an equalized image.

Considering the explicit discrete nature of a digital image, we can express h and H as follows

$$h(\lambda) = \frac{1}{|\Omega|} \sum_{x \in \Omega} \delta(\lambda - I(x)) \qquad \lambda \in [0, 1],$$

$$H(\lambda) = \frac{1}{|\Omega|} \sum_{x \in \Omega} \operatorname{sign}^+(\lambda - I(x)) \qquad \lambda \in [0, 1],$$

where

$$\delta(t) = \begin{cases} 1 & \text{if } t = 0; \\ 0 & \text{if } t \neq 0. \end{cases}$$

Many times, an image does not have a balanced number of intensities over the range $[0, 1]$: some values appear many times (where the histogram has a **peak**), others less frequently and some level may never appear, see Figure 7.3 (left column). However, if an image has equalized histogram, then all the levels appear with the same frequency of occurrence, see Figure 7.3 (right column). It is evident that an image with equalized histogram carries a larger amount of (quantitative, not necessarily qualitative) information.
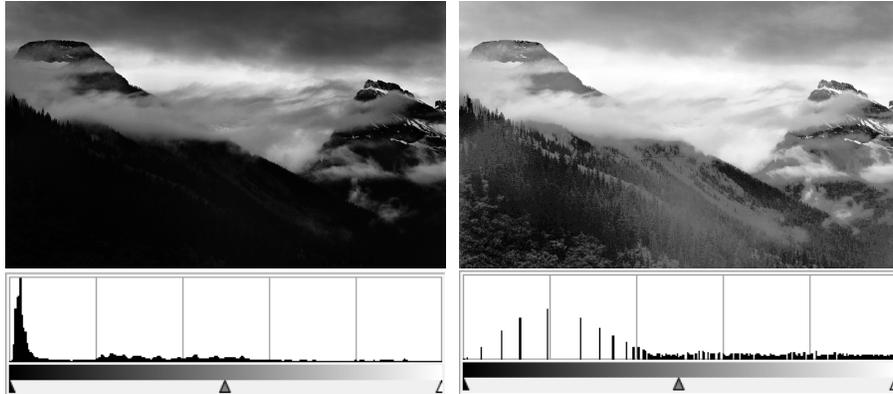
Figure 7.3: A famous picture of Ansel Adams before (left) and after (right) histogram equalization. Of course in the digital domain a perfect equalization is almost never impossible to achieve, so that approximations must be considered.

Classical **histogram equalization** is the transformation $\varphi : [0, 1] \to [0, 1]$, $r \longmapsto \varphi(r) = s$ which modifies the level distribution of an image in such a way that its histogram is as homogeneous as possible. To avoid the

inversion among the order of level lines, one usually assumes $\varphi'(r) \geq 0 \; \forall r$, i.e. that φ is a non decreasing function.

To determine analytically φ it is useful to observe that, of course, the total number of pixels must be conserved after the transformation. Translating this obvious consideration at an infinitesimal level we can say that the number of pixels with levels that fall in the intensity range $[r, r + dr]$ of the original image must be identical to the number of pixels with levels that fall in the intensity range $[s, s + ds]$ of the final (equalized) image.

If we denote with h_i (i=input) the normalized histogram of the original image and with h_o (o=output) the normalized histogram of the transformed image, then what we have just said can be written analytically through the equation: $h_i(r)dr = h_o(s)ds$. If we require the output image to be equalized, i.e. $h_o(s) = 1 \; \forall s$, then the previous infinitesimal equation can be re-written as $h_i(r)dr = ds$, but then, if we integrate both sides we find (recall that $s = \varphi(r)$, so the theorem of substitution of variable in integrals must be used)

$$\int_0^r h_i(t)dt \underset{s=\varphi(r)}{=} \int_{\varphi(0)}^{s=\varphi(r)} ds = s - \varphi(0).$$

In classical histogram equalization one usually sets $\varphi(0) = 0$, so that

$$\int_0^r h_i(t)dt = s,$$

but $\int_0^r h_i(t)dt = H(r)$, the normalized cumulative histogram of the original image, and $s = \varphi(r)$, hence the equalization transformation φ is nothing but the integral function H:

$$\boxed{r \underset{\varphi}{\longmapsto} s = H(r) : \text{Histogram equalization}}.$$

Thus, *the intensity level transformation $s = \varphi(r)$ realizes histogram equalization when changes each level r of the original image into the value of the normalized cumulative histogram corresponding to that same level.*

When we try to implement this transformation on a digital image of course we face the problem that, in the digital domain, it makes no sense to consider infinitesimal ranges of intensity levels, so that we must not expect equalization to be perfect (as can be seen in the histogram shown in the second row to the right in Figure 7.3).

Notice that histogram equalization can produce nice results, at least in terms of information content of an image, but can also destroy images. This typically happens for low key images, in fact, in these case the cumulative

histogram of dark levels is already close to 1, so that contrast of bright levels can even be decreased by histogram equalization, as can be seen in Figure 7.4
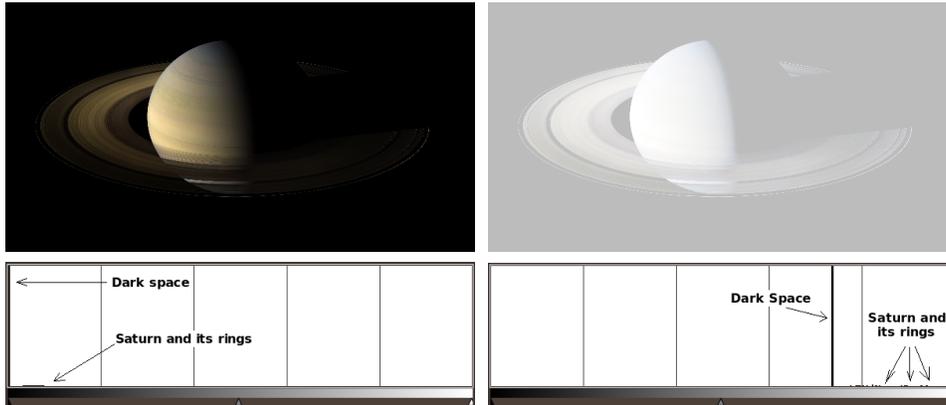

Figure 7.4: Effect of histogram equalization (right) on a low key image (left). Notice that the histogram of the 'equalized' image on the right starts exactly at the level defined by the normalized cumulated histogram in the level 0 of the original image.

What about histogram equalization in color images? In general, equalization in the three independent chromatic channels can be dangerous, since unnatural color can be generated by the unrelated stretching of the three histograms, as can be seen in Figure 7.5 (right column), the equalization of only the luminance channel in general avoids this problem Figure 7.5 (center column).

As Figure 7.5, a 'correct' histogram equalization of color images is not a trivial task to perform. The HVS automatically performs an equalization of light information, so we could take advantages of the HVS properties to implement a more sound histogram equalization. This can be performed in many different ways, in these notes we will show how to do it through the so-called variational techniques, but to do that we need to introduce the basic definitions and results about variational principles and show how histogram equalization can be interpreted in the variational framework.

7.2   Variational principles in digital imaging

Variational principles belong to one of the most important and useful fields of pure and applied mathematics: **Optimization**. In this section we will
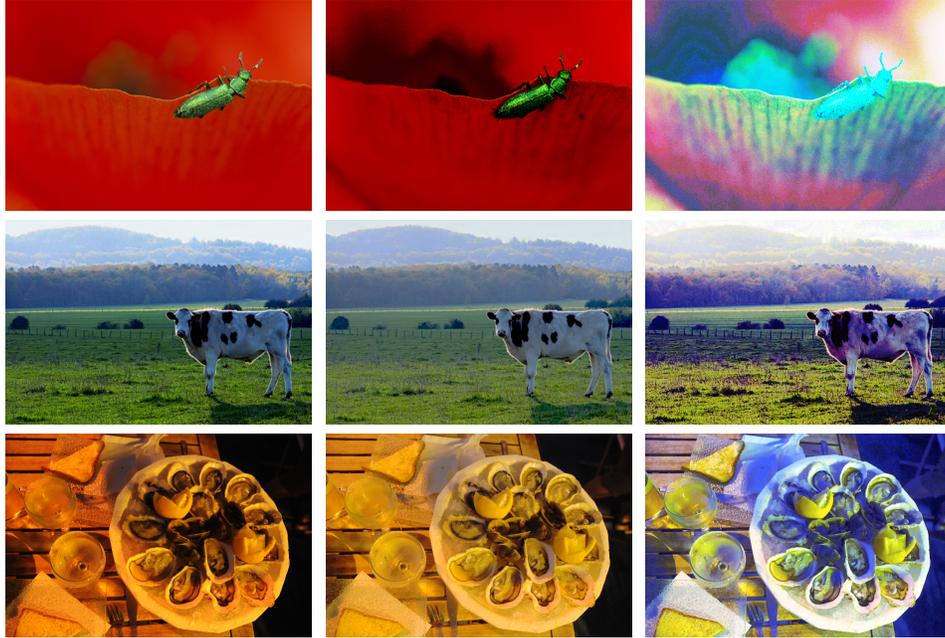
Figure 7.5: *Left*: original color image. *Center*: histogram equalization of the luminance. *Right*: histogram equalization of the three independent chromatic channels.

introduce just the basic information about variational principles applied to imaging.

One of the classical problems of ordinary calculus is finding the extremal points of a function $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$. Fermat's theorem states that, if f is differentiable in D, then each extremal point $x_0 \in D$ of f must satisfy

$$\boxed{\nabla f(x_0) = 0}, \textbf{ Stationarity} \text{ in terms of the gradient.}$$

If f has a global minimum (maximum) in $\vec{x}_0$, then we say that $\vec{x}_0$ is the **argmin** (**argmax**) of f in D:

$$\vec{x}_0 = \operatorname{argmin}_{x \in D}\{f(x)\} \quad \Longleftrightarrow \quad f(x_0) = \min_{x \in D}\{f(x)\},$$

$$\vec{x}_0 = \operatorname{argmax}_{x \in D}\{f(x)\} \quad \Longleftrightarrow \quad f(x_0) = \max_{x \in D}\{f(x)\}.$$

Let us recall that the gradient $\nabla f(x_0)$ has a profound relationship with $D_v f(x_0)$ directional derivatives of f in x_0 along a given direction described

103

by a unitary vector v, $\|v\| = 1$, expressed by

$$\langle \nabla f(x_0), v \rangle = D_v f(x_0),$$

where the directional derivative $D_v f(x_0)$ is defined as

$$D_v f(x_0) = \lim_{h \to 0} \frac{f(x + hv) - f(x)}{h} \equiv \left. \frac{d}{dh} \right|_{h=0} f(x + hv),$$

and it's the speed of variation of f along the direction defined by v when we move out of x_0. The previous formula implies that $D_v f(x_0)$ is given by the scalar product of the gradient $\nabla f(x_0)$ with v, i.e. by the projection of the gradient in the v direction. Thanks to this, we can express the stationarity condition for an extremal point in terms of directional derivatives:

$$\boxed{D_v f(x_0) = 0, \ \forall v \in \mathbb{R}^n}, \quad \textbf{Stationarity} \text{ in terms of directional derivatives.}$$

These results can be generalized from real-valued functions defined on $\mathbb{R}^n$ to real-valued functions defined on **functional spaces**. We recall that a functional space is a vector space whose elements are functions, e.g.

$$\mathcal{F} = \{f : D \subseteq \mathbb{R} \to \mathbb{R}\}$$

$$\mathcal{F}_{\mathrm{I}} \subset \mathcal{F} = \{f : D \subseteq \mathbb{R} \to \mathbb{R}, \ f \text{ injective}\}$$

$$\mathcal{F}_{\mathrm{E}} \subset \mathcal{F} = \{f : D \subseteq \mathbb{R} \to \mathbb{R}, \ f \text{ exhaustive}\}$$

$$\mathcal{C}^1(\Omega) = \{f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}, \ f \ n \text{ times differentiable with continuity in } D\}$$

$$L^2(\Omega) = \{f : \Omega \subset \mathbb{R}^n \to \mathbb{R}, \ \int_\Omega |f(x)|^2 dx < \infty\}$$

$L^2(\Omega)$ is called the Hilbert space of finite-energy functions on Ω, because the integral that appears in its definition is called 'energy' of the function f and its finite by definition.

Each continuous (normalized) $I : \Omega \subset \mathbb{R}^2 \to [0, 1]$ or digital discrete $I : \Omega \subset \mathbb{Z}^2 \to \{0, \ldots, 255\}$ image function can be seen as an element of $L^2(\Omega)$, since the image values are bounded and the image support Ω is finite. Image functions can be seen as elements of more complicated functional spaces, but for the purposes of this notes it is not necessary to go further into this discussion.

Variational principles applied to imaging interpret image functions I as variables of functions defined on $L^2(\Omega)$ called **energy functionals**:

$$\begin{aligned} E : \quad L^2(\Omega) &\longrightarrow \mathbb{R} \\ I &\longmapsto E(I) \equiv \langle E, I \rangle, \end{aligned}$$

where the notation $\langle E, I \rangle$ comes from the fact that *if the energy functional E is linear and continuous, then it can always be interpreted as a scalar product*, which in turn can be written as a suitable *integral*. This important result is called **Riesz representation theorem** and explains why practically all the functionals that are used in imaging are written as integrals (or sums, in the discrete domain).

A variational principle in imaging starts with the proper selection of an energy functional E, whose extremal points (maxima or minima, depending on the problem) must satisfy some *optimal conditions* to solve (completely or partially) a problem. The solution, i.e. the argmin or argmax of E is **the optimal image function** $I : \Omega \to \mathbb{R}$ that maximizes the property selected to define the functional E. If we are dealing with *RGB color images*, then we can apply the same reasoning on each scalar chromatic component I_c of the image, $c \in \{R, G, B\}$.

Without enter in detail, we can say that a typical example is given by the so-called *denoising functional*, whose argmin will be the image function with the least possible noise with respect to certain parameters of the functional.

Variational principles are used throughout imaging and often give the possibility to have a '*view from above*' which allows to more profoundly understand a problem and link together different algorithms designed to solve it.

A natural question that arises when dealing with variational principles is the following: is there any stationary condition that a functional must satisfy in order to be able to find its extremal points? The answer is yes and the stationary condition for a functional is the direct generalization of the stationary condition of a function in terms of directional derivatives.

Let us call the **Gateaux derivative** or **first variation** of a functional E in the image function I along another image function J the real number defined as

$$\delta E(I, J) = \lim_{h \to 0} \frac{E(I + hJ) - E(I)}{h} \equiv \frac{d}{dh}\bigg|_{h=0} E(I + hJ).$$

The stationary condition for a functional E can be stated by saying that, if I is an extremal point of E, then

$$\boxed{\delta E(I, J) = 0, \ \forall J}, \ \textbf{`Euler-Lagrange equations'}.$$

Typically the Euler Lagrange equations are implicit or too difficult to be solved, for this reason many numerical schemes have been invented to approximate the solution I of the Euler-Lagrange equations. Probably the

most famous and used consists in the **gradient-descent** and **gradient-ascent** methods:

$$\partial_t I = \pm \delta E,$$

where, of course, the minus defines the descent and the plus the ascent, and t is the step parameter of the scheme.

We summarize in the following table the similarities between the classical and functional optimization.

	Optimization in $\mathbb{R}^n$	Functional Optimization
To be optimized	Function $I : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$	Functional $E : L^2(\Omega) \to \mathbb{R}$
Variable	$x \in \Omega$	$I \in L^2(\Omega)$
Stationary condition	$\nabla I = 0$	$\delta I = 0$

Let us now show the first variations of two important functionals.

Theorem: Given the two functional

$$E_1(I) = \int_\Omega \psi(I(x)) \, dx;$$

$$E_2(I) = \iint_{\Omega^2} \phi(I(x), I(y)) \, dxdy$$

where ψ is a differentiable function defined on the codomain of I and ϕ is a differentiable function defined on the 2-th Cartesian power of the codomain of I, then their first variations are:

$$\delta E_1(I, J) = \int_\Omega \left. \frac{\partial \psi}{\partial I} \right|_{I(x)} J(x) dx \equiv \int_\Omega \psi'(I(x)) \, J(x) \, dx \qquad (7.2)$$

and

$$\delta E_2(I, J) = \iint_{\Omega^2} \left(\left. \frac{\partial \phi}{\partial I} \right|_{I(x)} J(x) + \left. \frac{\partial \phi}{\partial I} \right|_{I(y)} J(y) \right) \, dxdy. \qquad (7.3)$$

In the next sections we will use these results to show how histogram equalization can be written as a variational principle.

7.3 Histogram equalization through variational techniques

In the paper

G. Sapiro, V. Caselles: *Histogram modification via differential equations*,
    Journal of Differential Equations, Vol. 135(2), pp. 238–266, 1997

the authors has proven that histogram equalization can be interpreted as the minimization of an energy functional, as proven in the next theorem.

Theorem (Caselles-Sapiro): The image that minimizes the functional

$$E_{\text{Hist.Eq}}(I) = 2 \int_\Omega \left(I(x) - \frac{1}{2} \right)^2 dx - \frac{1}{|\Omega|} \iint_{\Omega^2} |I(x) - I(y)| \, dx dy \, ,$$

i.e. the image

$$I^* = \underset{I}{\operatorname{argmin}} \left(E_{\text{Hist.Eq}}(I) \right)$$

has equalized histogram.

Proof. : by linearity, we can compute the first variation of the two terms of the energy functional separately and then add the results. For that, it is convenient to introduce these names:

$$D_{\frac{1}{2}}(I) = 2 \int_\Omega \left(I(x) - \frac{1}{2} \right)^2 dx \, ;$$

$$C(I) = \frac{1}{|\Omega|} \iint_{\Omega^2} |I(x) - I(y)| \, dx dy \, .$$

By virtue of formula (7.2), we have

$$\delta D_{\frac{1}{2}}(I, J) = \int_\Omega 4 \left(I(x) - \frac{1}{2} \right) J(x) \, dx, \qquad (7.4)$$

and by virtue of formula (7.3), we have

$$\delta C(I, J) = \frac{1}{|\Omega|} \iint_{\Omega^2} [\operatorname{sign}(I(x) - I(y)) J(x) - \operatorname{sign}(I(x) - I(y)) J(y)] \, dx dy$$

$$= \frac{1}{|\Omega|} \iint_{\Omega^2} \operatorname{sign}(I(x) - I(y)) J(x) \, dx dy +$$

$$- \frac{1}{|\Omega|} \iint_{\Omega^2} \operatorname{sign}(I(x) - I(y)) J(y) \, dx dy.$$

Now, interchanging the role of the 'mute' variables x and y in the second integral of the last step, we have that

$$\frac{1}{|\Omega|} \iint_{\Omega^2} \operatorname{sign}(I(x) - I(y)) J(y) \, dx dy = \frac{1}{|\Omega|} \iint_{\Omega^2} \operatorname{sign}(I(y) - I(x)) J(x) \, dy dx$$

but then, using the oddness of the sign function,

$$\frac{1}{|\Omega|} \iint_{\Omega^2} \text{sign}(I(x)-I(y))J(y)\, dxdy = -\frac{1}{|\Omega|} \iint_{\Omega^2} \text{sign}(I(x)-I(y))J(x)\, dydx.$$

Hence, we can write

$$\delta C(I, J) = \frac{1}{|\Omega|} \iint_{\Omega^2} \text{sign}(I(x) - I(y))J(x)\, dydx +$$
$$+ \frac{1}{|\Omega|} \iint_{\Omega^2} \text{sign}(I(x) - I(y))J(x)\, dydx =$$
$$= \frac{2}{|\Omega|} \iint_{\Omega^2} \text{sign}(I(x) - I(y))J(x)\, dydx$$

that can be conveniently rearranged as follows

$$\delta C(I, J) = \int_{\Omega} \left(\frac{2}{|\Omega|} \int_{\Omega} \text{sign}(I(x) - I(y))\, dy \right) J(x)\, dx. \qquad (7.5)$$

Now, since $\delta E_{\text{Hist.Eq}}(I, J) = \delta D_{\frac{1}{2}}(I, J) - \delta C(I, J)$, by using formulas (7.4) and (7.5) we have

$$\delta E_{\text{Hist.Eq}}(I, J) = \int_{\Omega} 4\left(I(x) - \frac{1}{2} \right) J(x)\, dx - \int_{\Omega} \left(\frac{2}{|\Omega|} \int_{\Omega} \text{sign}(I(x) - I(y))\, dy \right) J(x)\, dx$$

i.e.

$$\delta E_{\text{Hist.Eq}}(I, J) = \int_{\Omega} \left[4\left(I(x) - \frac{1}{2} \right) - \frac{2}{|\Omega|} \int_{\Omega} \text{sign}(I(x) - I(y))\, dy \right] J(x)\, dx.$$

The stationary condition $\delta E_{\text{Hist.Eq}}(I, J) = 0$, $\forall J$, implies that the expression in the square bracket must be zero, i.e.

$$\delta E_{\text{Hist.Eq}}(I, J) = 0 \quad \Longleftrightarrow \quad 4\left(I(x) - \frac{1}{2} \right) - \frac{2}{|\Omega|} \int_{\Omega} \text{sign}(I(x)-I(y))\, dy = 0,$$

so that the Euler-Lagrange equation relative to the energy functional $E_{\text{Hist.Eq}}$ is the following implicit equation

$$2\left(I(x) - \frac{1}{2} \right) - \frac{1}{|\Omega|} \int_{\Omega} \text{sign}(I(x) - I(y))\, dy = 0,$$

that can be suitably re-written as

$$\frac{1}{|\Omega|} \int_{\Omega} \text{sign}(I(x) - I(y))\, dy = 2I(x) - 1. \qquad (7.6)$$

Now, using the identity $\text{sign}(t) = 2\text{sign}^+(t) - 1$, we can express the left-hand side of the Euler-Lagrange equation as

$$\frac{1}{|\Omega|} \int_\Omega (2\text{sign}^+(I(x) - I(y)) - 1)\, dy = \frac{2}{|\Omega|} \int_\Omega \text{sign}^+(I(x) - I(y))\, dy - \frac{\int_\Omega dy}{|\Omega|}$$
$$= 2H(I(x)) - 1,$$

where we have used the fact that $\frac{1}{|\Omega|} \int_\Omega \text{sign}^+(I(x) - I(y))\, dy$ is the spatial version of the cumulative histogram $H(I(x))$, as noticed in eq. (7.1).

Thus, the Euler-Lagrange eq. (7.6) is equivalent to $2H(I(x)) - 1 = 2I(x) - 1$, i.e. to $H(I(x)) = I(x)$, but then

$$\delta E_{\text{Hist.Eq}}(I, J) = 0 \quad \Longleftrightarrow \quad H(I(x)) = I(x),\ \forall x \in \Omega,$$

which means that the image function I which satisfies the Euler-Lagrange equations of the functional $E_{\text{Hist.Eq}}(I)$ has an equalizes histogram.

In the quoted paper, G.Sapiro and V.Caselles proof that the functional $E_{\text{Hist.Eq}}(I)$ has a global minimum when I is the solution of the Euler-Lagrange equation, and that completes the proof of the fact that the image $I^* = \underset{I}{\text{argmin}}\, (E_{\text{Hist.Eq.}}(I))$ has an equalized histogram. $\qquad\square$

7.3.1 Interpretation of the variational histogram equalization

The interpretation of the energy functional $E_{\text{Hist.Eq}}(I)$, whose argmin is an image with equalized histogram, gives a first example of the power and importance of variational principles.

To understand why, let us start the histogram equalization energy functional is $E_{\text{Hist.Eq}}(I) = D(I) - C(I)$, where the two functional terms $D(I)$ and $C(I)$ are

$$D_{\frac{1}{2}}(I) = 2 \int_\Omega \left(I(x) - \frac{1}{2} \right)^2 dx$$

and

$$C(I) = \frac{1}{|\Omega|} \iint_{\Omega^2} |I(x) - I(y)|\, dx dy,$$

so the minimization of $E_{\text{Hist.Eq}}(I) = D(I) - C(I)$ is achieved through the **minimization of** $D(I)$ and the **maximization of** $C(I)$, because $C(I)$ has a minus sign in front of it, so it becomes more negative as we increase it!

Let us discuss the meaning of the two functional terms:

- $D_{\frac{1}{2}}(I)$ is called **global quadratic dispersion** term around the middle gray level $1/2$ and it is minimized when $I(x) = 1/2$ for all $x \in \Omega$, i.e. the minimization of this term tends to turn the image I into a constant gray image.

- $C(I) = \frac{1}{|\Omega|} \iint_{\Omega^2} |I(x) - I(y)| \, dxdy$ is called **global contrast** term because is built by integrating the absolute value of the intensity differences between two pixels, and we know that $|I(x) - I(y)|$ is proportional to the Michelson contrast among the pixels x and y. As said before, to minimize $E(I)$ we must maximize $C(I)$, and this corresponds to maximize the global Michelson contrast of the image I.

Thus, the image I^* that minimizes $E(I)$ is the image that realizes the optimal balance between **two opposite effects**: from one side the minimization of $E(I)$ tries to **set all the levels to the average gray** $1/2$ and, from the other side, it tries to **spread the intensity levels apart, as far as possible from each other**. *The equilibrium among these two conflicting actions, dispersion control and contrast enhancement, realizes histogram equalization.* This is a *highly non intuitive result* that is very difficult, if not impossible, to achieve without variational principles.

One practical consequence of this result is that, applying for example the gradient descent technique (or another numerical scheme to minimize $E_{\text{Hist.Eq}}(I)$, we can stop the minimization process before reaching the complete equalization, thus realizing a *partial equalization* that can nonetheless be useful for someone's purposes and avoiding the problems reported in section 7.1.

However, for the purposes of color image processing, the most important consequences of the theorem just proven are theoretical, in fact, as we will see in the next section, we can modify the functional $E_{\text{Hist.Eq}}(I)$ in such a way that the basic principle of histogram equalization, i.e. the balance between dispersion control and contrast enhancement, is preserved but we can change the analytical form of the terms $D(I)$ and $C(I)$ on the basis of human visual perceptual features, so that the argmin image I^* of the modified functional will optimally represent color perception in a suitable sense.

7.4 Perceptually-inspired color enhancement of digital imaging

In chapter 6 we have examined some HVS properties related to adaptation and contrast perception. Let us very briefly recall the basic facts:

1. **Adaptation**: the Michaelis-Menten equation (4.1) describes the response of photoreceptors to light stimuli and it shows that the HVS is able to adapt to the average radiance level of any visual scene;

2. **Weber-Fechner's law of contrast perception**: in Section 6.3.1 we have seen that the response of the HVS to light changes is roughly logarithmic because, for non-extreme light intensities, the ratio between the just noticeable difference JND among light intensities and the average intensity is approximately constant;

3. **Local contrast enhancement**: phenomena as Mach bands of simultaneous contrast show that the HVS performs a contrast enhancement that depends on the local information of light intensity around each point of a visual scene;

4. **Color constancy**: is the ability to adapt to different color temperatures of light sources in order to perceive colors almost constantly, a feature that is vital for object recognition.

These four properties are considered **the most fundamental phenomenological features of color vision**.

It must be noticed that adaptation and local contrast enhancement are opposite features: adaptation happens mostly in the retina and can be related to the control of dispersion around an average radiance level; local contrast enhancement is induced by neuron dynamics and interactions and tends to magnify intensity differences on the basis of local information.

We have seen that this is exactly the same behavior underlying histogram equalization, for this reason it can be inferred that **the HVS implements a (local and highly non linear) histogram equalization of the radiance** signal corresponding to each visual scene.

This means that one can search for a possible modification of the energy functional related to histogram equalization to embed the fours fundamental phenomenological properties of the HVS just commented into a variational framework.

This has been achieved in the paper

where the adaptation mechanism has been modeled with a dispersion term written as

$$D(I) = \int_\Omega d(I)\, dx$$

d being a differentiable real-valued function of I and local contrast enhancement has been modeled as the functional term

$$C_w(I) = \iint_{\Omega^2} w(x,y) c(I(x), I(y))\, dx dy,$$

$w : \Omega \times \Omega \to \mathbb{R}^+$ being a spatial weight and c being a differentiable real-valued function of two pixel intensities.

The global energy functional is the sum of these two terms:

$$E(I) = D(I) + C_w(I).$$

Of course there are infinitely many functionals that can be written like that, however, in the paper the authors have proven that when the other two properties, color constancy and Weber-Fechner's law, are taken into account, then there can be only one class of functionals that fulfill all the four properties at once. In other words, the selection of $D(I)$ and $C_w(I)$ is driven by the HVS properties and it turned out to be unique.

Without entering in the quite difficult details of this selection, we report here the dispersion and contrast terms.

The dispersion term $D(I)$ is given by the entropy around the average value

$$\mu = \frac{1}{|\Omega|} \int_\Omega I(x) dx$$

of the image I, i.e.

$$D(I) = \int_\Omega \left[\mu \log \frac{\mu}{I(x)} - (\mu - I(x)) \right] dx,$$

since entropy is a measure of disorder, minimizing $D(I)$ means minimizing the possibility to find intensity value other than μ, which is exactly what we expect from the dispersion term.

The contrast term instead is

$$C_w(I) = \iint_{\Omega^2} w(x,y)\, \varphi\left(\frac{\min\{I(x), I(y)\}}{\max\{I(x), I(y)\}}\right) dx\, dy, \qquad \varphi'(t) > 0$$

to understand its action let us concentrate on the so-called perceptual contrast function

$$c^\varphi(I(x), I(y)) = \varphi\left(\frac{\min\{I(x), I(y)\}}{\max\{I(x), I(y)\}}\right), \qquad \varphi'(t) > 0$$

which is minimized when

- $\min\{I(x), I(y)\}$ decreases;

- $\max\{I(x), I(y)\}$ increases,

hence, the minimization of c^φ indeed induces contrast enhancement. The local nature of this enhancement is guaranteed by the weight factor w. Typically w is a Gaussian kernel with center in x, its standard deviation σ can be decided by a user to increase or decrease the locality of contrast enhancement. Very small values of σ push towards *sharpening*.

For the purposes of perceptually-inspired color enhancement one can also add a coefficient α to have a degree of freedom and balance the action of the two terms, thus considering this energy functional:

$$E^\varphi_{\alpha,w}(I) = \alpha D(I) + C_w(I), \quad \alpha \in \mathbb{R}^+.$$

The Euler-Lagrange equations of $E^\varphi_{\alpha,w}(I)$ are implicit equations that cannot be solved directly. For this reason the functional is minimized through a *gradient descent* w.r.t. $\log I$. The continuous gradient descent equation is:

$$\partial_t \log I = -\delta E^\varphi_{w,\alpha}(I)$$

By discretizing the equation above one can generate an iterative algorithm which converges to a fixed point image I^*, i.e. $I^* = \mathrm{argmin}(E^\varphi_{w,\alpha}(I))$.

It is interesting to notice that this framework gives a sort of general house for perceptually-inspired models of color correction: in fact, by varying φ, it is possible to obtain variational formulations of two well known algorithms of this type, namely Retinex and ACE, respectively:

$$C^{\mathrm{id}}_w(I) = \iint_{\Omega^2} w(x,y)\, \frac{\min(I(x), I(y))}{\max(I(x), I(y))}\, dx\, dy,$$

$$C_w^{\log}(I) = \iint_{\Omega^2} w(x,y) \log \left(\frac{\min(I(x), I(y))}{\max(I(x), I(y))} \right) dx dy.$$

The iterative equations in these cases are: Iterative equation from *gradient descent* technique ($\varphi \equiv \mathrm{id}$):

$$I^{k+1}(x) = \frac{I^k(x) + \Delta t \left(\alpha \mu + R_{I^k}^{\mathrm{id}}(x) \right)}{1 + \Delta t(\alpha)},$$

where

$$R_{I^k}^{\mathrm{id}}(x) := \int_{\mathfrak{I}} w(x,y) \frac{I^k(y)}{I^k(x)} \mathrm{sign}^+(I^k(x) - I^k(y)) \, dy$$

$$- \int_{\mathfrak{I}} w(x,y) \frac{I^k(x)}{I^k(y)} \mathrm{sign}^+(I^k(y) - I^k(x)) \, dy,$$

I^k: image at the k-th step, I^0: original image, and, when $\varphi \equiv \log$:

$$I^{k+1}(x) = \frac{I^k(x) + \Delta t \left(\alpha \mu + R_{I^k}^{\log}(x) \right)}{1 + \Delta t(\alpha)},$$

where

$$R_{I^k}^{\log}(x) := \int_{\mathfrak{I}} w(x,y) \, \mathrm{sign}(I^k(x) - I^k(y)) dy.$$

In Figure 7.6 we compare the action of these variational algorithms to that of classical histogram equalization: it can be seen that the problems related to the generation of unnatural colors and excessive contrast enhancement are minimized by the perceptually inspired modification of histogram equalization (the pictures correspond to the minimization of the functional $E_{\alpha,w}^{\mathrm{id}}(I)$.
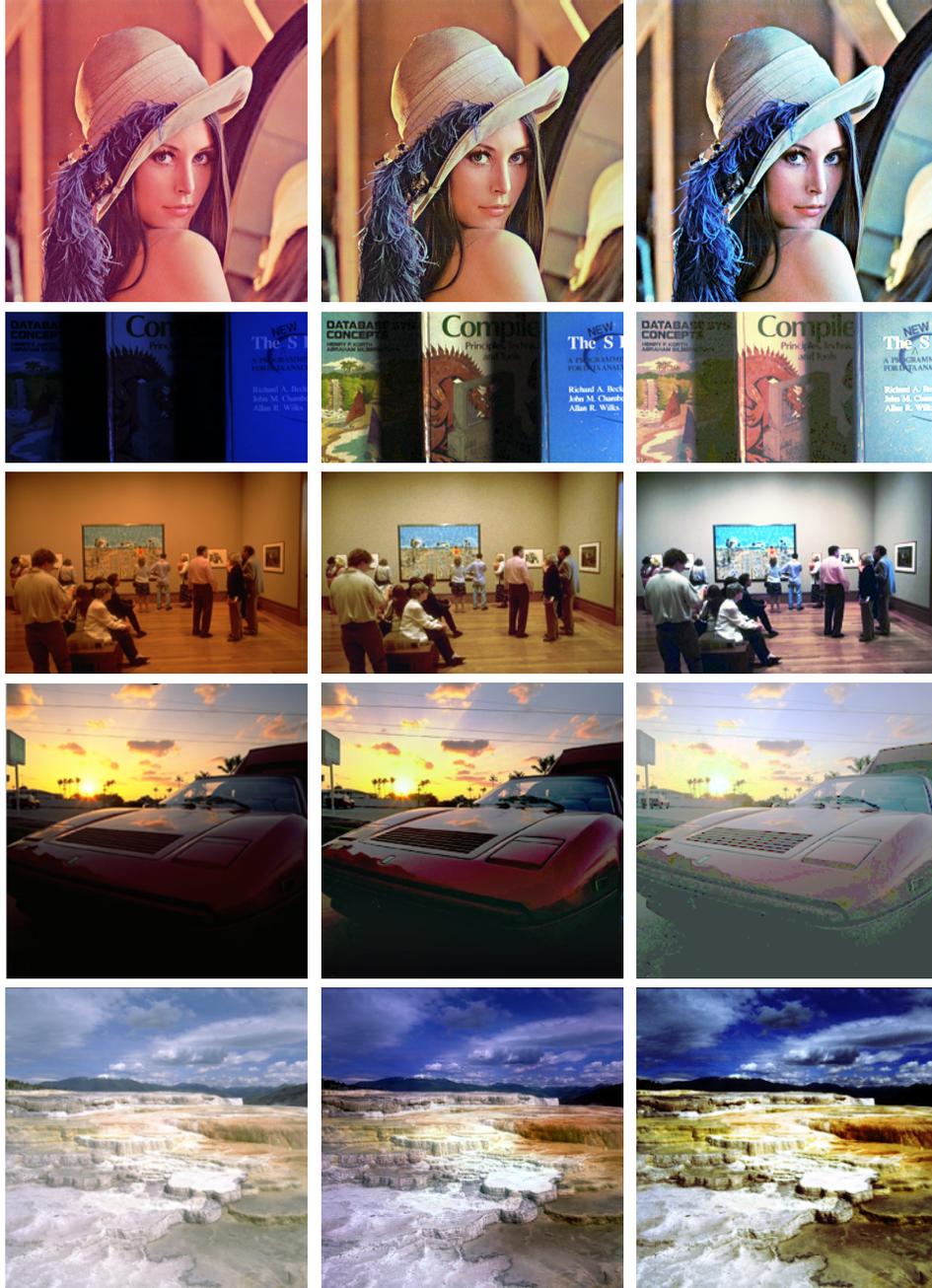
Figure 7.6: *Left*: original image. *Center*: result of the perceptually-inspired variational algorithm with $\varphi =$id. *Right*: histogram equalization of the three independent chromatic channels.

Chapter 8

High Dynamic Range (HDR) imaging and the tone mapping problem

High Dynamic Range images are a relatively recent way to store data proportional to the radiance of a visual scene. In this chapter we will show how to build these images and the so-called tone mapping problem, i.e. how to compress them in order to be shown on a standard screen.

8.1   Radiance maps generation

In the SIGGRAPH conference of 1997, P. Debevec and J. Malik presented the following paper

'Recovering High Dynamic Range Radiance Maps from Photographs'
Paul E. **Debevec** and Jitendra **Malik**
Proceedings of the 24th annual conference SIGGRAPH conference, 1997,
Pages 369-378

downloadable at: http://debevec.org/Research/HDR/debevec-siggraph97.pdf, presented the first fast, simple and robust method of recovering radiance maps from a set of ordinary photographs taken with different exposures.

These radiance maps are somewhat improperly called High Dynamic Range (HDR) Images, we say 'improper' because an HDR image is not a digital image in the classical sense: it cannot be shown on a standard screen since its range typically span more than the 2 orders of magnitude

available in standard screens, it is more an array of pixels whose intensity is proportional to the radiance of a visual scene.

In this section we will revise the work of Debevec and Malick. Let us begin by recalling that, as seen in Section 1.2, the *radiance* L_e is the optical power per unit of solid angle and surface orthogonal to the light direction and it is measured in $\frac{W}{sr\,m^2}$, while the *irradiance* E_e is the optical power incident on a surface and measured in $\frac{W}{m^2}$. We also recall that the radiance information is important because, if we multiply it by the human spectral sensitivity function $V(\lambda)$ and integrate in the visible spectrum, we obtain the luminance L_V, which is the photometric quantity that starts the visual process.

Figure 8.1 shows why it is so difficult to recover the radiance with a single (analogical or digital) photo.
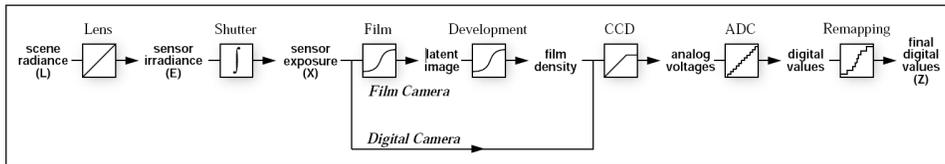

Figure 8.1: Image Acquisition Pipeline shows how scene radiance becomes pixel values for both film and digital cameras. Unknown nonlinear mappings can occur during exposure, development, scanning, digitization, and remapping.

Non-linear mechanisms as the typical sigmoidal response of a film or a CCD, see Figure 8.2, makes it impossible to achieve a linear mapping of a general scene radiance. However, whenever we fix an exposure, i.e. a time in which the shutter of a camera remains open, there is an intermediate range of radiance which is linearly transformed by a camera. If we change the exposure, the linear response region will slide over the entire radiance range. In Figure 8.3 we can see the effect of different exposure.

The basic idea of Debevec and Malick was to suitably 'fuse' the information obtained by the linear response with different exposures in order to have a global linear mapping of the scene radiance. They achieved that through a constrained least-square model, let us see how.

They considered a camera fixed on a tripod, that takes pictures of a static scene from a static vantage point with different exposure durations Δt_j, $j = 1, \ldots, P$ (P as pictures). They also assume that the camera resolution is high enough to consider the irradiance values $E_e(x)$ for each pixel $x \in \Omega$
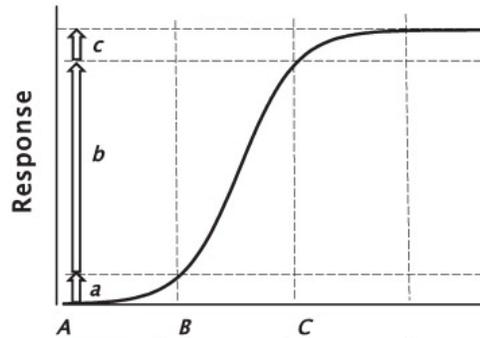
Figure 8.2: Typical sigmoidal response of a camera to light stimuli, notice that the behavior in the region bounded between B and C is approximately linear.

Figure 8.3: From left to right, images taken with decreasing time exposure. If Δt is *big*, one can recover radiance *details in dark areas* of the scene, if Δt is *small*, one can recover radiance *details in bright areas* of the scene.

as constant, Ω being the spatial support of the image taken, as usual. They denoted pixel values by $Z_j(x)$, $x \in \Omega$, $j = 1, \ldots, P$. Finally, they define the exposure X as the product of the irradiance $E_e(x)$ at the camera sensor in the pixel x and the exposure time, Δt, $X_j(x) = E_e(x)\Delta t_j$, measured in J/m^2.

The **reciprocity assumption** states that only the product $E_e(x)\Delta t$ is

important, i.e. halving $E_e(x)$ and doubling Δt will not change the resulting optical density. Under extreme conditions (very large or very low Δt), the reciprocity assumption can break down, a situation described as **reciprocity failure**. In typical analogical films, reciprocity holds within the range of 10^{-4}-10 seconds. In the case of a CCD, reciprocity holds under the assumption that each site of the CCD measures the total number of photons it absorbs during the integration time.

Assuming that reciprocity holds, we can write the **reciprocity equation** as:

$$Z_j(x) = f(E_e(x)\Delta t_j)$$

where f is an yet unknown function given by the composition of the characteristic sigmoidal curve of the CCD as well as all the nonlinearities introduced by the later processing steps shown in Figure 8.1.

Debevec and Malick made the 'reasonable assumption' that the function f is smooth and monotonically increasing, so its inverse f^{-1} is well defined. Knowing the exposure $X_j(x)$ and the exposure time Δt_j, the irradiance $E_e(x)$ is recovered as $E_e(x) = X_j(x)/\Delta t_j$. Knowing the relation between radiance and irradiance expressed by eq. 1.1, one can reconstruct the radiance L_e in the scene. The objective function of Debevec and Malick is thus f^{-1}.

Observe that applying f^{-1} to both sides of the reciprocity equation we obtain

$$f^{-1}(Z_j(x)) = f^{-1}(f(E_e(x)\Delta t_j)) = E_e(x)\Delta t_j$$

taking the logarithm of both sides

$$\log f^{-1}(Z_j(x)) = \log(E_e(x)\Delta t_j) = \log E_e(x) + \log \Delta t_j$$

Denoting $g \equiv \log f^{-1}$, $g : \{Z_{\min}, \ldots, Z_{\max}\} \subseteq \{0, \ldots, 255\} \to \mathbb{R}$, we obtain the following system of $|\Omega| \cdot P$ equations ($|\Omega|$ being, as usual, the number of pixels of each image):

$$g(Z_j(x)) = \log E_e(x) + \log \Delta t_j, \tag{8.1}$$

the exposure times Δt_j and the digital values $Z_j(x)$ are known, while the $|\Omega|$ irradiance values $E_e(x)$ and the $Z_{\max} - Z_{\min} + 1$ values taken by the function g and are unknown, with the only assumption that g is monotonic and smooth.

Debevec and Malick solved the previous equation in the function g and the irradiances $E_e(x)$ in a **least-squared sense**, i.e. they searched for the

$|\Omega|$ irradiance values $E_e(x)$ and the $Z_{\max} - Z_{\min} + 1$ values taken by the function g that **minimize the following constrained quadratic objective function**:

$$
\begin{aligned}
\mathcal{O} \;=\;& \sum_{x \in \Omega} \sum_{j=1}^{P} \{w(Z_j(x)) \left[g(Z_j(x)) - \log E_e(x) - \log \Delta t_j\right]\}^2 \\
& +\; \lambda \sum_{z=Z_{\min}+1}^{Z_{\max}+1} \left[w(z)g''(z)\right]^2 ;
\end{aligned}
$$

where:

- the *first term* ensures that the solution satisfies the set of equations 8.1 in a least squares sense;

- the *second term* is a **smoothness term** on the sum of squared values of the second derivative of g to ensure that the function g is smooth, in fact this term is minimized when g'' is minimum and the second derivative measures the acceleration of variation of g, so that low values of g'' correspond to the absence of singularities in the behavior of g. In the discrete setting $g''(z) = g(z-1) - 2g(z) + g(z+1)$;

- The scalar λ weights the smoothness term relative to the data fitting term and should be chosen appropriately for the amount of noise expected in the measurements of the values $Z_j(x)$;

- Due to saturation, at the extreme values $Z_{\min}$ and $Z_{\max}$ of the dynamic range, $g(z)$ is typically less smooth. A weighting function $w(z)$ is added to emphasize the smoothness and fitting terms toward the middle of the curve:
$$
w(z) = \begin{cases} z - Z_{\min} & \text{if } z \leq Z_{\mathrm{mid}} \\ Z_{\max} - z & \text{if } z > Z_{\mathrm{mid}} \end{cases}
$$
  as represented in Figure 8.4

The least-square problem above can be solved in a fast and robust way by using the *singular value decomposition* (SVD) method. Given measurements of $|\Omega|$ pixels in P photographs, we have to solve for $|\Omega|$ values of $\log E(x)$ and $Z_{\max} - Z_{\min}$ samples of g. Thus, the size of the system of equations is of the order $|\Omega|P + Z_{\max} - Z_{\min}$, which, even for small images, requires very expensive calculations. To reduce the computation, we can consider a random distribution of $n << |\Omega|$ pixel values that ensure a sufficiently
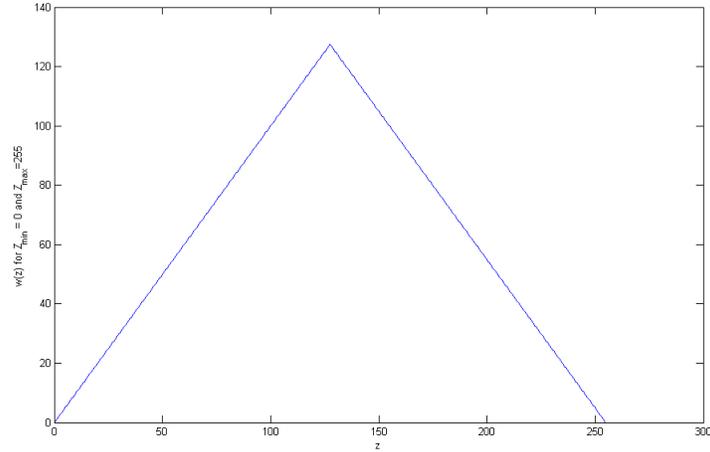
Figure 8.4: Shape of the weighting functions used by Debevec and Malick.

overdetermined system, i.e. $n(P-1) > Z_{\max} - Z_{\min}$. For the pixel value range $Z_{\max} - Z_{\min} = 255$ and $P = 11$ photographs, Debevec and Malick declare that a choice of n on the order of 50 pixels is adequate. The running time is on the order of a few seconds on an ordinary PC.

The **number of photographs needed** to best recover the radiance map, of course, *strongly depends on the extension of the dynamic range of the photographed scene*: for a not too contrastated scene 3 pictures can be enough, for a high contrastated scene 5 or 7 pictures are recommended.

Once found the values of $g(z)$, the logarithmic radiance values can be expressed as:

$$\log E_e(x) = g(Z_j(x)) - \log \Delta t_j,$$

however, for robustness, all the available exposures for a particular pixel are used to compute its radiance, for this it is convenient to reuse the weighting function $w(z)$ to give higher weight to exposures in which the pixel's value is closer to the middle of the response function, where nonlinearities are less probable, and perform the weighted sum:

$$\log E_e(x) = \frac{\sum_{j=1}^{P} w(Z_j(x))(g(Z_j(x)) - \log \Delta t_j)}{\sum_{j=1}^{P} w(Z_j(x))}.$$

Finally, an exponential is performed: $E_e(x) = \exp(\log E_e(x))$ for each $x \in \Omega$.

8.2 Tone mapping

Once we have obtained the radiance map with the Debevec-Malick method, it arises the problem of showing this image on an ordinary Low Dynamic Range (LDR) screen or printer. To do that, we must process the HDR image in order to reduce back its dynamic range to the 2 orders of magnitude of LDR screens and printers.

This might seem a contradiction: we struggle to build an HDR image from multiple exposures and then we turn back to a LDR image, why doing that? The answer relies in the fact that the dynamic reduction must be implemented in a clever way in order to preserve as much as possible the details contained in the HDR image and the appearance of colors of the visual scene photographed. Such a transformation is called 'tone mapping'.

Of course, tone mapping cannot be performed simply by a linear normalization, i.e. the transformation

$$I(x) \longmapsto \frac{I(x) - I_{\min}}{I_{\max} - I_{\min}},$$

$I_{\max}$ and $I_{\min}$ being, respectively, the highest and lowest intensity of the HDR image in a given chromatic channel, scales the HDR image intensities to the range $[0, 1]$, but the radiance ranges in an HDR are so different that such a transformation would typically generate an output image that is black all over, except for the pixels whose intensity has the same magnitude of the brightest pixel, as can be seen in Figure 8.5

How can one perform a sound tone mapping? Once again, the HVS properties can help us. In fact, the human visual system performs a very efficient tone mapping in two opposed stages:

- firstly, intraocular **veiling glare** (caused by scattered light in the eye bulb) reduces the luminance range on the retina, and retinal photoreceptor response expressed by the Michaelis-Menten formula (4.1) further reduces the range of the electric signal passed to the optical nerve;

- secondly, neurophysiology of the visual system produces a physiological simultaneous contrast which increases the differences in appearances by comparing the responses from different part of the retinal luminance image. Vision's simultaneous contrast mechanism changes dramatically any correlation of scene luminance and appearance.

The world that humans see is almost always one with nonuniform illumination and surround, and with dynamic ranges greater than the information
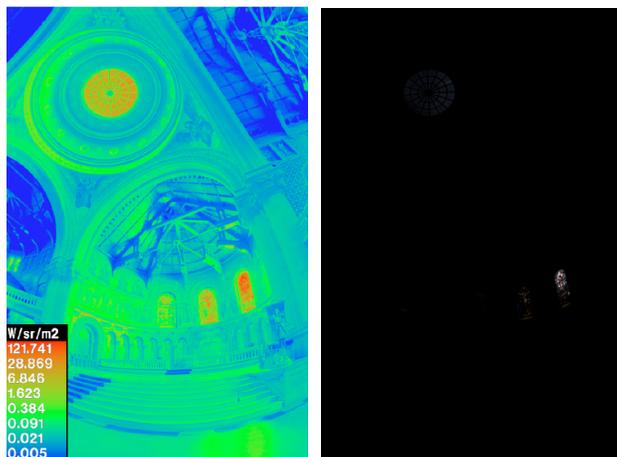
Figure 8.5: *Left*: false color image showing the radiance levels of the Memorial church image, the total radiance range is 10^6. *Right*: result of a linear scaling on the HDR Memorial church image.

capacity of the optic nerve. It is highly probable that for this reason humans have evolved in order to produce the two opposite mechanisms quoted above.

A tone mapping algorithm that implements these two stages has been proposed in: S. Ferradans, M. Bertalmío, E. Provenzi, V. Caselles: '*An analysis of visual adaptation and contrast perception for tone mapping*', IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 33(10), 2002-2012, October 2011. The first stage of that tone mapping algorithm implements a version of the Michaelis-Menten equation (4.1) and the second stage uses the variational contrast enhancement discussed in the previous chapter. In Figure 8.6 we can see the results of this method compared to those of a global logarithmic tone mapping transformation.
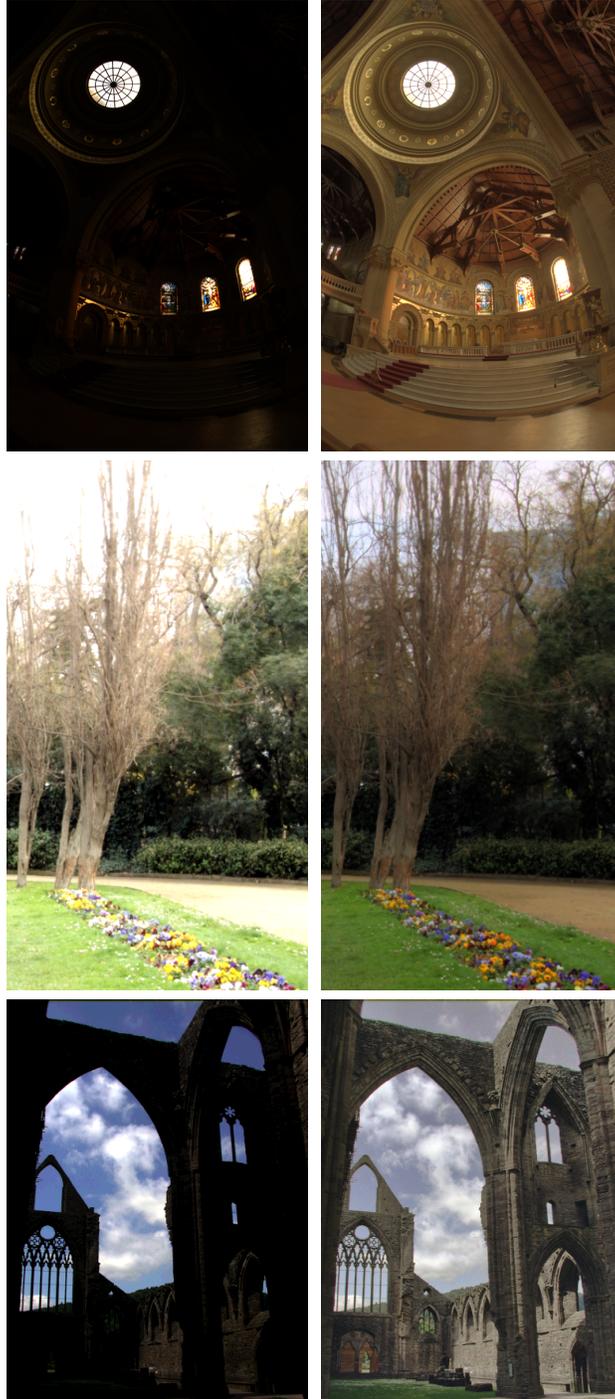
Figure 8.6: *Left*: results of a global logarithmic tone mapping. *Right*: results of the tone mapping proposed in the paper quoted above.