Chapter 3

II - Perception

Knowledge about the human visual system (HVS) is useful to the computer vision scientist in a number of respects. First, the HVS sets the norm for several products. If images are displayed, e.g. in an interactive system for carthographic updates or medical diagnosis, image quality will have to be satisfactory. Secondly, limitations in human vision or performance will open market opportunities if a computer can be made to do the job. As an example, high precision delineation of structures in images is something humans are not particularly good at (it being boring for one thing), but a computer could do it rather easily once receiving interactive assistance in finding the locus of attention. Another example is quantitative mensuration as part of visual inspection. Third, the HVS has inspired – and will even more so in the future – the problems to be tackled and the ways to do it. Responses of neurons in the visual cortex have inspired people to use certain filters. Stereo vision has been implemented for robot vision. In fact, the very idea of using the visible part of the spectrum for the extraction of so rich a host of information would probably have never been taken seriously if its feasibility would not be illustrated so convincingly by the HVS. As will become clear, the HVS is not a kind of radiometric devise. What the HVS exactly does nobody can tell, but *human vision is an active process*, not a matter of pure registration.

3.1 II - Brightness (Luminance)

Consider a light with spectral radiant flux $c(\lambda)$ that is zero everywhere except at $\lambda = \lambda_r$, where λ_r denotes a fixed reference wavelength. A light that consists of only one spectral component (or wavelength) is called a *monochromatic light*. Suppose we ask a human observer to compare the brightness of a monochromatic reference light with $c(\lambda_r)$ with that of another monochromatic light with $c'(\lambda_t)$ where λ_t is a test wavelength. By brightness is meant the observed intensity of the light. The power required for equal brightness can be obtained by showing two patches of light with a fixed $c(\lambda_r)$ and to let the observer decrease or increase the amplitude of $c'(\lambda_t)$ until he feels they match in brightness. The ratio $c(\lambda_r)/c'(\lambda_t)$ where $c(\lambda_r)$ and $c'(\lambda_t)$ match in brighness, is called the *relative luminous*

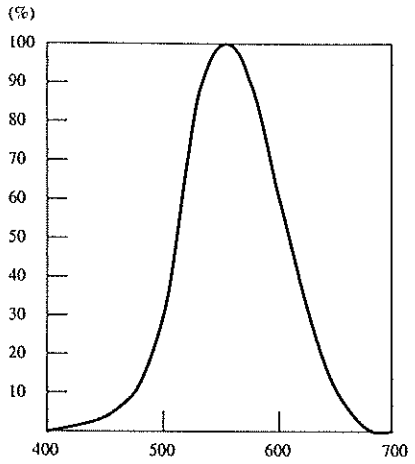

Figure 3.1: *C.I.E. relative luminous efficiency function for daylight vision ("photopic vision").*

efficiency of a monochromatic light λ_t relative to λ_r, and is approximately independent of the amplitude of $c(\lambda_r)$ under normal viewing conditions. The wavelength λ_r used is 555 nm (yellow-green light), at which a typical observer has maximum brightness sensitivity. For this choice of λ_r, the relative luminous efficiency is always less than or equal to 1, since $c(\lambda_r)$ is not greater than $c'(\lambda_t)$; that is, it takes less energy at λ_r to produce the same brightness. The relative luminous efficiency as a function of λ is called the *relative luminous efficiency function* and is denoted by $v(\lambda)$. Two monochromatic lights with $c_1(\lambda_1)$ and $c_2(\lambda_2)$ appear equally bright to an observer when

$$c_1(\lambda_1)v(\lambda_1) = c_2(\lambda_2)v(\lambda_2).$$

The relative luminous efficiency function $v(\lambda)$ depends on the observer. Even with a single observer, a slightly different $v(\lambda)$ is obtained when measured at different times. To eliminate this variation, the *C.I.E. standard observer* was defined in 1929, based on experimental results obtained from a number of different observers. C.I.E. stands for Commission Internationale de l'Eclairage, an international body concerned with standards for light and colour. The resulting function $v(\lambda)$ is called the *C.I.E. relative luminous efficiency function* and is shown in fig. 3.1 for vision under daylight conditions ("photopic vision"). The luminous efficiency function is roughly bell-shaped with a maximum at $\lambda = 555$ nm. A similar curve can be drawn for night vision ("scotopic vision"). It looks very similar but should be shifted about 50 nm to the left.

We conclude that the brightness impression created with the human observer is not a measure for the actual power in the incoming light. Radiometric units are therefore not apt for specifying observed brightness. Since the word "brightness" has some connotation of colour vividness the term *luminance* has been introduced instead. Luminance is the photometric equivalent of the radiometric radiance. A complete overview of photometric terminology will not be given here. In fact, once acquainted with radiometric quantities,

the reader will find that each radiometric quantity has a photometric quantity associated to it. Normally it will suffice to interchange the words radiant and *luminous*. Examples of photometric quantities are luminous flux, luminous emittance, and luminous intensity. *Illuminance* is the photometric counterpart of irradiance.

Although in view of the latest standards the *lumen* (unit of luminous flux) would make the best candidate to be the basic unit (SI unit) of photometry, this role is so far still played by the *candela* (unit of luminous intensity). Anyway, the relation between both is very direct and simple, since one candela corresponds (by definition) to one lumen per steradian.

Before 1979 the candela was defined on the basis of 1/60 of the luminous intensity radiated within one steradian in a perpendicular direction by 1 cm^2 of a blackbody (thorium oxide) at 2042 K, the temperature of melting platinum, under a pressure of 101 325 newtons per square meter. Because this standard was difficult to reproduce, a new one was adopted that links up photometry and radiometry directly. As such, the definition relies on that of the Watt, which is precisely known. The candela is now defined as follows:

> The candela is the luminous intensity in a given direction of a source that emits monochromatic radiation of frequency 540×10^{12} Hertz and that has a radiant intensity in that direction of (1/683) Watt per steradian.

In fact this definition has been formulated in a way consistent with the older one within the limits of uncertainty. The frequency in the definition corresponds to 555 nm, the peak in the luminous efficiency function. At this frequency one lumen stands for 1/683 Watt = 1.46 mWatt. The connection between photometric and radiometric units for other frequencies is given through the luminous efficiency curve of fig. 3.1. The inverse yields a scale factor that tells how much more power is in a lumen of light at another wavelength. The luminous flux of a light with spectral radiant flux $c(\lambda)$ can be calculated as

$$l = k \int_{\lambda=0}^{\infty} c(\lambda) v(\lambda) d\lambda$$

and is measured in lumens, k is 683 lumens/watt, $c(\lambda)$ is in watts, $v(\lambda)$ is unitless, and λ has a unit of m.

Until about 1940 photometric measurements were primarily based on the visual matching of brightness. Although by nature very close to the actual goal, i.e. quantifying visual impressions, such *visual photometry* methods were rather inconvenient, but also of low precision and poor reproducibility due to the variability between observers and over time. Modern methods use physical detectors almost exclusively. The radiometric results are transformed by incorporating factors that take account of visual response characteristics for the standard observer. The match of the transducer with the C.I.E. luminous efficiency function is therefore crucial. This *physical photometry* is more convenient, precise, and reproducible than visual photometry.

Much of the confusion in photometry arises from the parallel use of several units. We confine our discussion to a simple enumeration of some of these:

- 1 stilb = 10^4 candela / m^2
- 1 lux = 1 lumen / m^2
- 1 phot = 1 lumen / cm^2
- 1 footcandle = 1 lumen / ft^2
- 1 nit = 1 candela / m^2

3.2 II - Colour

Human observers can distinguish many more colours than they can distinguish intensity levels (shades of grey). The study of colour is therefore also important for computer vision for other reasons than aesthetic judgment. Transforming different intensity levels to different colours (so-called "pseudo-colours") facilitates discerning details. Colour plays an important role in biological as well as computer vision as a rich feature that helps detecting (segmenting) and recognizing objects in cluttered scenes.

The representation of colour

Before we discuss ways to model the perception of colour, is is useful to first consider the different aspects that seem relevant. The perceptual "dimensions" of colour are *luminance (brightness), hue,* and *saturation* or *chroma.* The hue of a colour refers to that aspect that we usually refer to as "the colour": the overall colour type such as red, green, etc. Saturation specifies the vividness of the colour. Starting from a monochromatic light (e.g. red light) we have maximum saturation. Then we can start adding white, evolving to pink. The more white we add, the lower the saturation. In the limit we obtain pure white and the saturation becomes zero. Figure 3.2 shows the baseball-shaped representation of colour space organised according to these concepts. Along the vertical axis we find brightness, evolving from black (no power) to ever brighter light. We see that with very dark or bright light colours can hardly be distinguished. Hue specifies a meridian, i.e. a vertical section (on the drawing) through the axis of the baseball. Each meridian corresponds to a monochromatic colour (e.g. in order of increasing wavelength, cfr. table 2.1), or a purple, as will be explained later. Saturation tells how far from the axis the colour is to be situated. Our "baseball" is normally referred to as the *C.I.E. Lab colour space.* Without going into the details, L is correlated with brightness, a with the relative amounts of red versus green, and b with the relative amounts of yellow versus blue.

After this rather qualitative discussion, a more quantitative approach is outlined. Colour representation methods are based on the classical theory of Thomas Young. [1], He introduced and demonstrated the idea that an arbitrary colour can be generated by

[1]Thomas Young (1773-1829) was a British physician and physicist, who had been a student at London, Edinburgh, Göttingen, and Cambridge before he became a staff member of St. George's Hospital in London. He was involved in the study of optics from both a medical (*The mechanism of the eye* (1800))

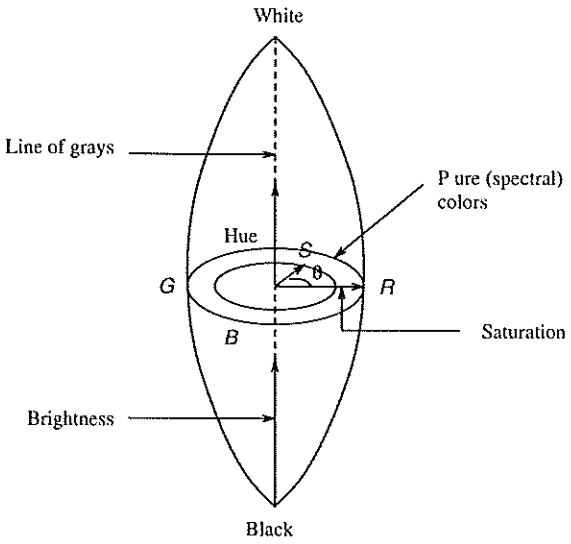

Figure 3.2: *Perceptual representation of the colour space. The brightness varies along the vertical axis, hue θ varies along the circumference, and saturation varies along the radius.*

an appropriate mixture of three primary colours. Don't misunderstand this. This is *not* to say that e.g. the light corresponding to any single wavelength could be made identical to a linear combination of just three primary colours. Of course, by investigating the spectrum, radiometry would be able to tell the difference. The *percept* created with the human observer can be made indistinguishable, however, by suitably mixing the three primary colours. Later on, this method obtained its physiological underpinning, as three different types of cones in the human retina where found with distinct spectral sensitivity. The corresponding human response curves, $H_1(\lambda), H_2(\lambda), H_3(\lambda)$ for short, are shown in fig. 3.3. As can be seen, the curves have their maximum in the yellow-green, green, and blue regions, respectively. Notice the substantial overlap and the low sensitivity of the "blue cone". A source with spectral radiant flux $C(\lambda)$ will produce responses R_i, $i = 1, 2, 3$, for the three cones as the integrated response over the spectral radiant flux:

$$R_i(C) = \int H_i(\lambda)C(\lambda)d\lambda, \qquad i = 1, 2, 3$$

If $C_1(\lambda)$ and $C_2(\lambda)$ are two spectral radiant flux characteristics that produce identical responses (i.e. $R_i(C_1) = R_i(C_2)$ for $i = 1, 2, 3$) then these two sources are perceived as having the same colour.

In the qualitative baseball sketch, luminance information was literally considered orthogonal to the colour or *chrominance* information. As comes out of experiments, this is justified. The relative amounts needed for three sources to match a colour is quite independent of the overall luminance of the colour. If the luminance of the colour to be matched

and a physics point of view (he reintroduced Huygens' wave theory). He also played an important role in the decyphring of hieroglyphs.

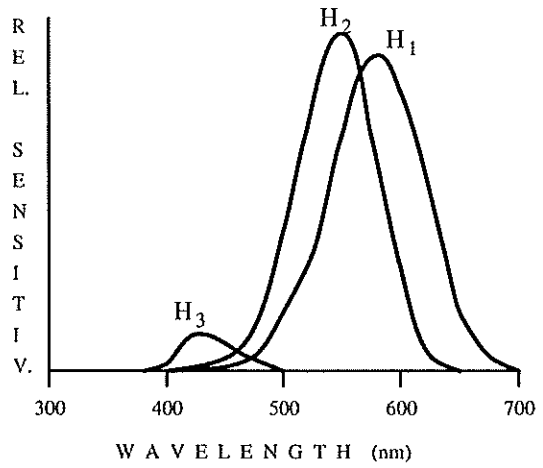

Figure 3.3: *Spectral sensitivity curves for the three types of retina cones.*

is doubled, so are the appropriate amounts of the three sources [2].

The above describes normal trichromatic vision. Colour blindness corresponds to three forms of abnormal colour vision: monochromatism, dichromatism, and anomalous trichromatism. The monochromat only discerns different luminances, chrominance information is lost completely (only 0.003 % of the population). The dichromat's vision can be modeled by using two cone types (between 2 and 3 % of males), and anomolous trichromatism requires three cone types, but different ones than for normal colour vision (between 5 and 6 % of males). When designing visual colour outputs as a source of information, it is important to keep these rather large groups in mind.

As a further remark, the perception of colour cannot be reduced to the outputs of the three cone types. There is much more to it, as the phenomenon of 'colour constancy' shows. Colour constancy is discussed in chapter 13.

3.3 * II - Contrast

From the section on colour constancy (??) it already became clear that the perception of the world is more than a collection of the sensations at the front-end of the visual system. Basically, the HVS introduces context information. This section in this sense is the sequel of the section on brightness. Indeed, when the impression of brightness is tested with respect to a background, *relative luminance* comes out to be crucial.

Suppose a homogeneously bright central pattern (e.g. square) is shown on top of a homogeneous background, as in fig. 3.4. Apparent brightness is seen to heavily depend on the background luminance. In fig. 3.4, the central squares are all of identical luminance, but their perceived brightness varies as a function of background luminance. It can be considered a general rule that the HVS is much better in making relative than absolute

[2]This is one of Grassman's laws on the mixing of colours.

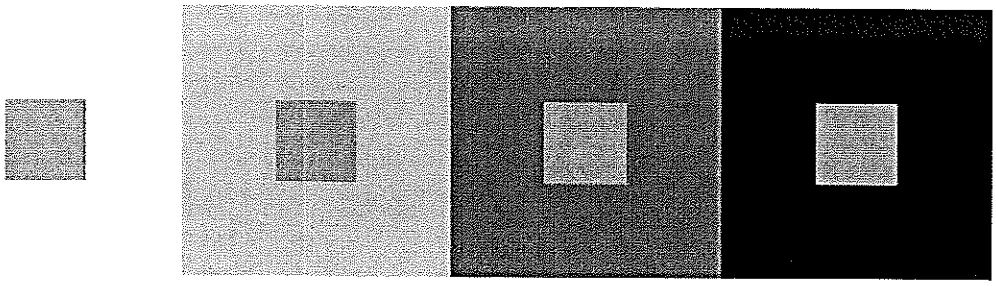

Figure 3.4: *Apparent brightness is not simply related to region luminance. The inner squares have equal luminance in all 4 cases, but brightness appears different.*

judgments. If contrast is sufficiently high, the task of telling whether the central square or the background is the brightest is easy (i.e. specifying contrast sign), but not to give an absolute estimate of brightness for either of them.

Of course, contrasts that are too small cannot be detected either. There is a contrast detectability threshold, a *just noticeable difference* or *jnd* that should be surpassed. The detection of contrast is described by *Weber's law*, one of the many findings on logarithmic responses in biological perception. It states that the *jnd* Δl is proportional to the background luminance l_{back}:

$$\frac{\Delta l}{l_{back}} \sim c_W.$$

Typically the *Weber fraction* c_W is a few percent. Also above this threshold remains the ratio of contrast over background luminance a perceptually important measure. Apparent brightness will be the same for equal ratios for a wide range of background luminances. This phenomenon is known as *brightness constancy*. It should be noted that *jnd*s are much lower for scotopic vision. They also depend on the eccentricity of presentation on the retina (see section 3.4.1), being generally lower at the periphery.

Also the sizes of the patterns are important. If cosine patterns of different spatial frequency are shown, then the Weber fraction can be seen to be a function of that frequency. This differential sensitivity (gain factor of the visual system) is demonstrated by fig. 3.5, where the spatial frequency decreases to the right while the contrast increases upward; the curve along which the pattern is just visible represents a kind of *modulation transfer function* (see chapter 5) of the reader's visual system. This interpretation from linear system theory should not be used blindly, however, since the visual system is highly non-linear in some respects. This differential sensitivity is drawn in the back curve of figure 3.6. From this figure we notice that contrast sensitivity is largest for some range of spatial frequencies, around 5 cycles per degree of visual angle, and has a bandpass characteristic. A similar phenomenon can be observed for temporal frequencies, i.e. contrast reversals are applied at a certain pace. The effect is shown in the same figure 3.6 as additional equal temporal frequency sections of what is actually a sensitivity surface. The figure illustrates

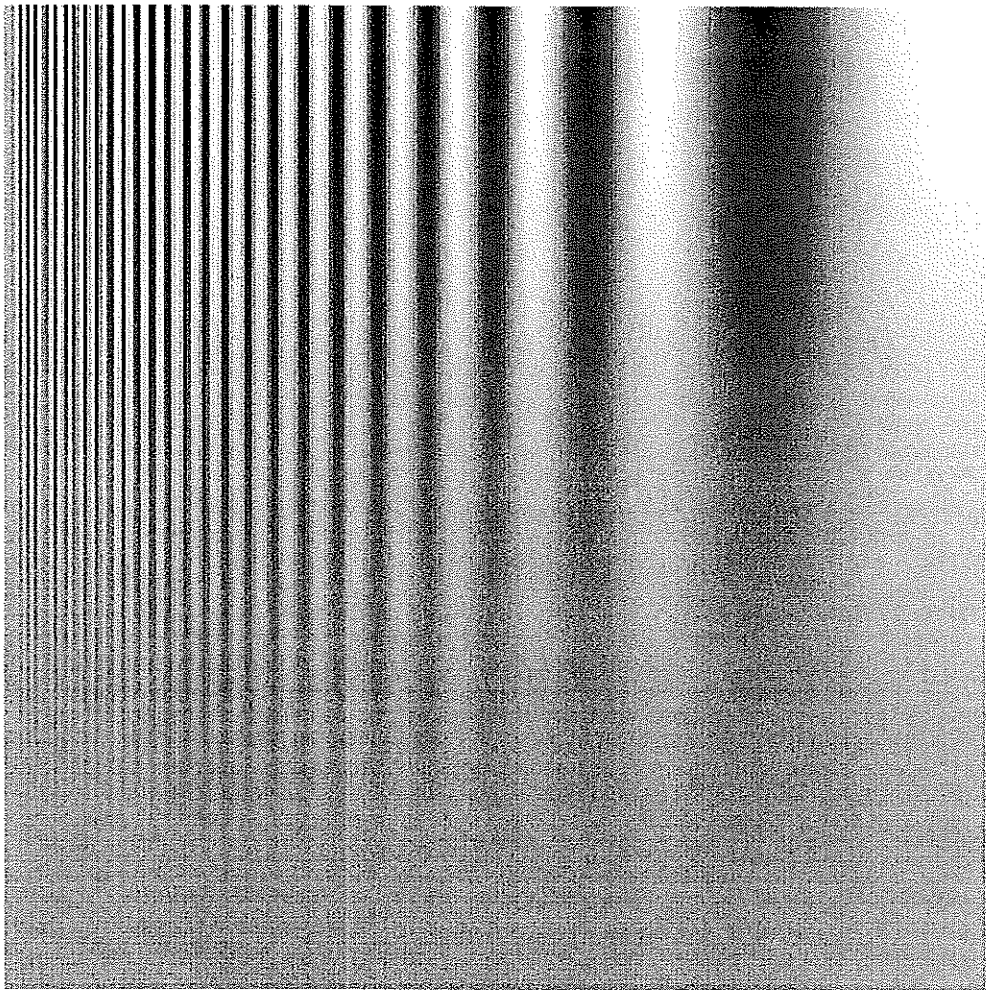

Figure 3.5: *Sinusoidally modulated pattern.*

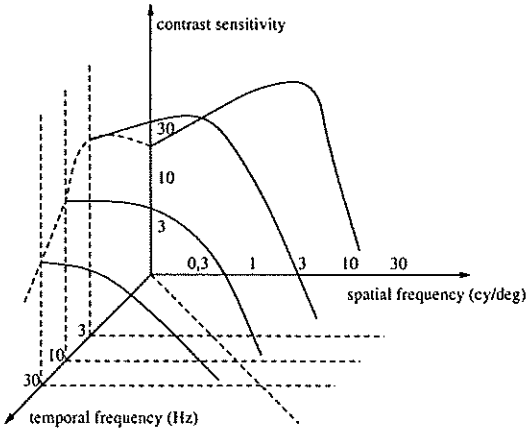

Figure 3.6: *Spatio-temporal contrast sensitivity.*

how the spatial cut-off frequency decreases with increasing temporal frequency. This means the spatial resolution moving objects are perceived with is lower. The temporal resolution is determined by what is perceived as flickering. The temporal frequency at which the pattern appears steady to the viewer is the *critical fusion frequency* and is about 60 to $70Hz$. Such information is crucial in the design of displays (see chapter 4).

Contrasts of single-colour patterns are perceived in a similar way as luminance contrast discussed above, but the peak and cut-off frequencies lie at lower spatial and temporal frequencies.

As will be discussed later (see 3.4.1), the spatial resolution is also dependent on the position with respect to the center of gaze. This is due to the inhomogeneous distribution of light sensitive elements over the retina. This is not so relevant here, since people will normally keep the scenery of interest in the central, i.e. most dense, part.

The differences in sensitivity to different spatial frequencies explains another phenomenon, the so-called *Mach bands*: overshoots observed at abrupt intensity transitions. Looking at fig. 3.7, the constant intensity rectangles appear to have varying intensity, being brighter on the side of the darker neighbouring rectangle, and darker along the brighter neighbouring rectangle. This percept is explained by the fact that step edges contain a broad gamut of frequencies, including very high ones beyond the cut-off point. The frequency dependent sensitivity of the visual system will not preserve all frequencies to the same degree, thereby creating the Mach effect. This effect might be useful, since it tends to increase edge contrast. Computer vision algorithms with a similar effect are frequently used in e.g. document processing (in particular unsharp masking, see chapter ??).

As has been amply illustrated now, apparent brightness is not a pure matter of local luminance, but all kinds of context effects have to be taken into account. Again, processing by the HVS is seen to be more involved than might be assumed from basic psychophysical experiments. We could literally say there is more than meets the eye, i.c. a lot of massively parallel postprocessing by several millions of neurons.

The impossibility of predicting perception from simple measurements on the input is corroborated even more by the so-called Benussi-Koffka ring of fig. 3.8. Although the intensity of the gray ring is constant, its two halfs are seen as having different brightness by the simple addition of the horizontal black lines (right drawing). The HVS seems to smooth out local brightness impressions at the edges of a region over that region, as long as no borders have to be crossed. This would explain the uniform impression for the left, and the non-uniform impression for the right ring.

Conversely, edge induced contrasts seem to propagate the edges into uniform regions, as shown in fig. 3.9. Although the corresponding regions are not separated from the background by any kind of edge, we clearly see the characters R,E,A,D, a triangle with its corners occluding three circles, and a triangle overlaying another triangle and 4 squares. It seems as if the edges of the figure were extended. The brightness of these purely imaginary shapes even seems to be higher than that of the "background"!

This *fill-in* capacity of the HVS is remarkable and might well have developed as a weapon against camouflage so abundant in nature. It is interesting to note that fill-in is not operational without intensity contrast, e.g. isoluminant colour patterns do not show the

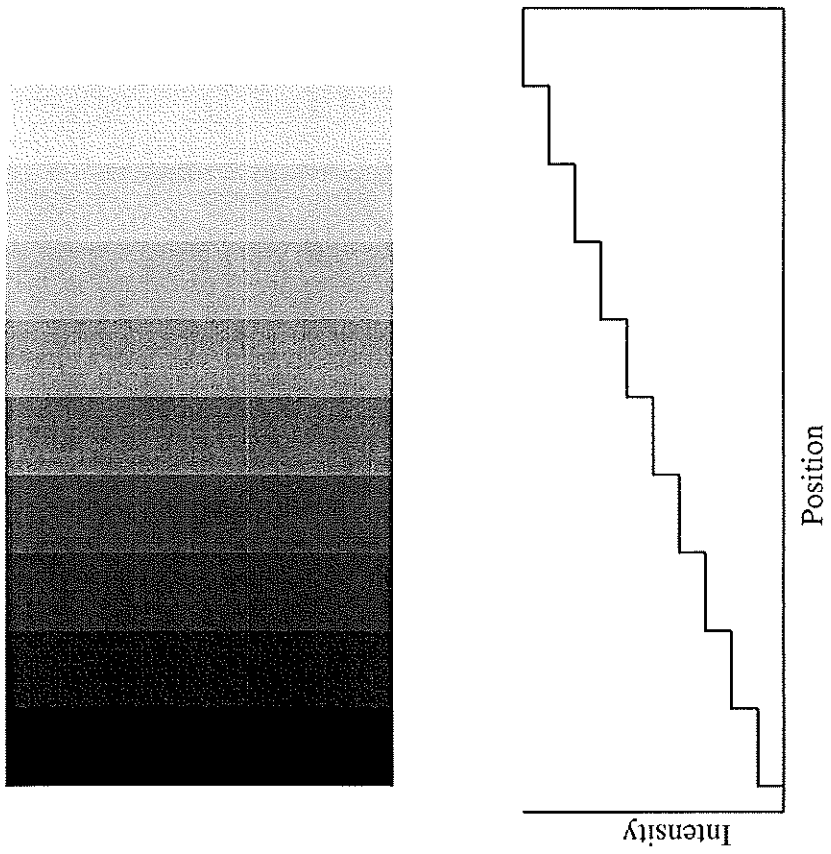

Figure 3.7: *Illustration of Mach bands.*

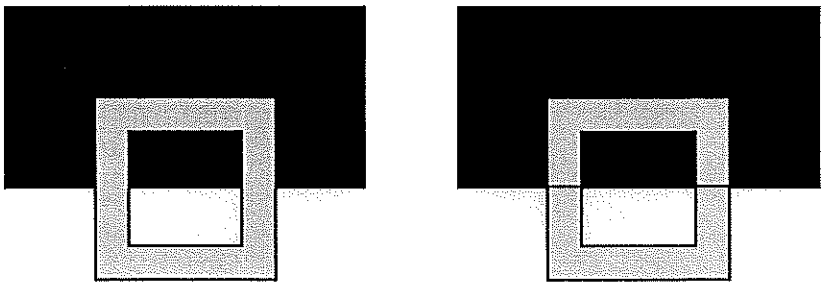

Figure 3.8: *The Benussi-Koffka ring.*

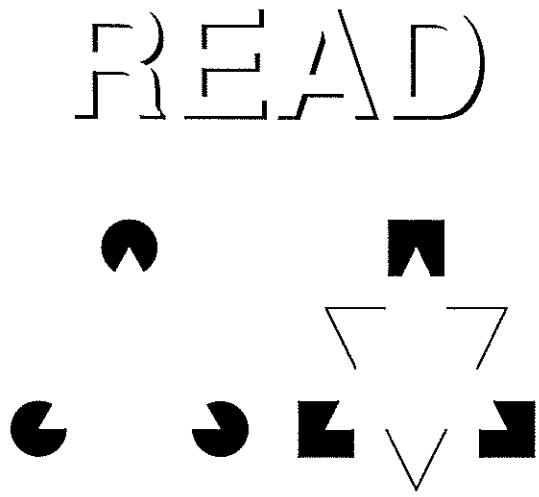

Figure 3.9: *"Fill-in" processes.*

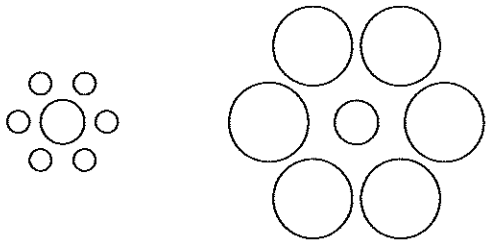

Figure 3.10: *Ebbinghaus illusion.*

fill-in effect.

3.4 * II - Shape

As noted before for brightness, the HVS is far more strongly oriented towards making relative than absolute judgments. The same holds for size. It is not so difficult to tell which of a pair of yardsticks is longer, but one has a hard time to tell how long they actually are. This is illustrated in fig. 3.4, showing the *Ebbinghaus illusion.* The central circle in both figures has the same diameter, yet it seems larger in the left figure. Again we see the direct influence of context on very basic estimates. Along the same lines, consider figure 3.11. The *Müller-Leyer illusion* is very strong: both horizontal lines have equal length, but this seems almost unbelievable.
It is possible to mislead the HVS when it comes to the judgment of several other features of shapes. For most of these illusions no final or satisfactory explanation has been found yet.
Fig. 3.12 shows segments of the same circle, yet their curvature seems different. It has

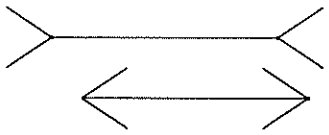

Figure 3.11: *Müller-Leyer illusion.*

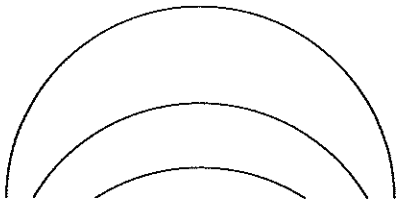

Figure 3.12: *Segments of circles having the same curvature.*

been suggested the false impression arises by the fact that the visual system somehow takes the orientational differences at the segment end points. The lower segment clearly has a lower such difference and would therefore be judged to have a lower curvature.
Collinearity isn't safe either. Fig. 3.13 shows an oblique straight line, interrupted by a vertical bar. Having to choose one of the two oblique lines at the right to collinearly extend the line on the left, a human observer (who is not in the middle of reading a section on optic illusions) will pick the wrong one.
How about parallelism? As fig. 3.14 shows, perfectly parallel lines seem to intersect under the influence of an interfering, superimposed pattern.
Even the straightness of a line should be judged with care after being confronted with fig. 3.15.

Each of the foregoing illusions has something to do with the interference of differently oriented patterns. It seems that the presence of a particular orientation raises the threshold for that orientation, a form of *adaptation*, and that similar orientations are therefore perceived as tilted more away from them. This is in line with suggestions that the visual system operates with banks of filters sensitive to a gamut of orientations and a kind of weighted output of such filters is used to obtain a high orientation precision (so-called *acuity*). If adaptation lowers the output of some filters, the weighted output will give a false result.

One of the remarkable aspects of the HVS is the speed at which is is able to *segment* (i.e. decompose, see chapter 14) a scene into the entities it comprises, e.g. objects. One school of perception psychology – the *Gestaltists* [3] – introduced the idea that the HVS uses a set of rules of thumb, the *Gestalt laws of organization*. These describe how to go about finding entities in the scene, i.e. how to *group* the multitude of visual inputs. Research into these laws has been very intensive in perception psychology for many years, but has

[3] After the German *Gestalt*, a word difficult to translate (probably the reason why it survived so many language barriers), meaning something like percept, appearance, stature.

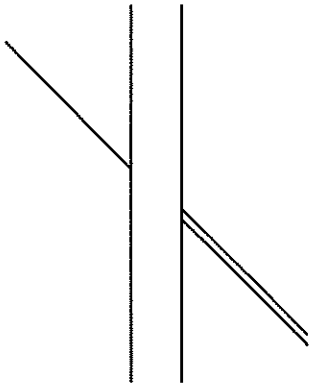

Figure 3.13: *Poggendorf illusion.*

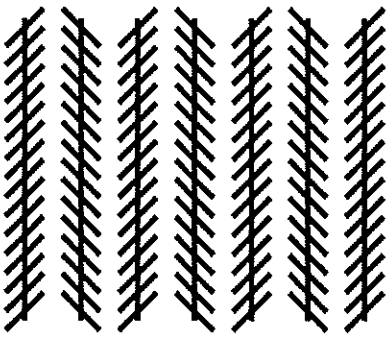

Figure 3.14: *Zollner illusion.*

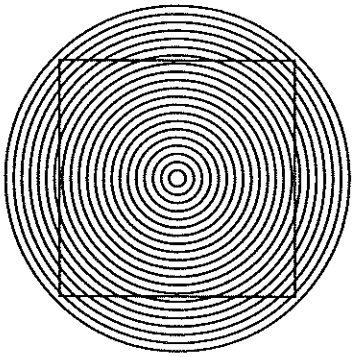

Figure 3.15: *Square in distress.*



Figure 3.16: *Gestalt law of proximity: oblique grouping is preferred over horizontal grouping, due to smaller distances.*



Figure 3.17: *Rubin vase.*

unfortunately never been able to come up with quantitative statements on the relative importance of these laws.
The *law of proximity* states that grouping is preferred over entities that are close (fig. 3.16). The *law of similarity* states that similar entities tend to end up in the same grouping. The *law of good continuation* states that curves will be seen to extend over crossings via the branch that yields the highest smoothness. The *law of closure* states that closed figures are seen as units. The *law of symmetry* prescribes that symmetric groupings be preferred.

An intriguing aspect of the segmentation problem is how the HVS can so easily tell objects and background apart: the *figure-ground problem.* One could argue that this is not a crucial issue, since upon exchanging figure and background the outlines of objects would not change, except for a sign reversal of the curvature. A simple illustration will suffice to show the opposite. Consider fig. 3.17, a drawing of the *Rubin vase.* One can see a vase or two faces, but not both simultaneously, and what one sees depends on which part of the figure is interpreted as background. The HVS "switches" between both interpretations.

3.4.1 * The eye

The eye, a masterpiece of biological engineering, was seen by Darwin [4] as a very important test case, where gradual increases in complexity had to be found in order for his evolutionary theory to hold. None of our other senses is able to bombard the brain with so rich and so plentiful an amount of information. It has everything our most sophisticated cameras

[4]cfr. the following quote from the Origin of Species: "To suppose that the eye with all its inimitable contrivances for adjusting the focus to different distances, for admitting different amounts of light, and for the correction of spherical and chromatic aberration, could have been formed by natural selection, seems, I freely confess, absurd in the highest degree." In later editions of his *Origin of Species* he introduced Lamarckist ideas.

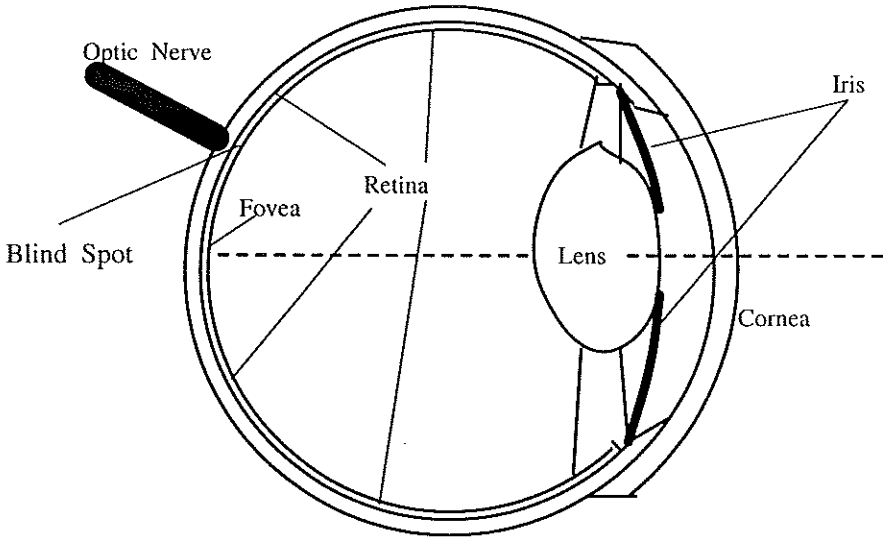

Figure 3.18: *Architecture of the human eye.*

have to offer, and more.

The basis for its operation are pigment molecules that absorb light and thereby change their chemical structure and consequently cell membrane potentials. In many-cellular animals, these electric signals would have to travel over considerable distances compared to the cell size. As with the members of a technologically advanced society, the organisation of such organism is based on specialisation. Some cells have developed in special photodetectors. Other cells, nerve cells or *neurons*, transport and process the information.

For some animals, like earthworms, the photoreceptors are spread over the body surface. Other animals have developed eyespots, that are clusters of such cells. There can be more than two eyes. Bees, for instance, have two large facetted eyes, and three smaller ones. The latter seem to be used for judging precisely the rise and decline in brightness, in order to determine optimal times for departure to and return from food gathering places. The facetted eyes are one way of discerning spatial patterns. Indeed, in order to achieve this, some form of directional selectivity had to be implemented. Each receptor has its own orientation and is shielded off from other directions because it is placed in an opaque concavity. In the case of the bee, one of the specific advantages is that the precise position of the sun can be estimated, as well as the translational velocity of patterns, both useful in determining flight routes. Another solution for directional selectivity is to use a lens type of mechanism that directs light towards the different receptors that are placed together in a single cusp, depending on the angle of the incoming light. This is the basic architecture of the vertebrate eye, discussed next.

The vertebrate's eye uses a convex lens and a layer of photoreceptors at the back of a spherically shaped cavity. Figure 3.18 shows the architecture of the human eye. The cornea, the iris, the lens, and the retina are the major components.

The *cornea* is a protective transparent tissue and is the portion of the eye that is in contact with the outside world.

The *iris* is nature's diaphragm. Its opening (pupil) is regulated to control the amount of light that enters the eye.
The *lens* absorbs ultraviolet and infrared light. It has remarkable focusing capabilities. Depending on the distance of objects, muscles change the lens: flatter for distant objects and thicker for objects nearby.
The *retina* is the layer of photoreceptors that covers the surface of the eye cavity opposite to the lens. They come in two classes: *rods* and *cones*. A curious fact is that the light sensitive part of the cells lies oriented away from the lens. Although the retinal surface is obscured by blood vessels at places, we don't have the impression of looking through mazes of such network. This is a first role the fill-in mechanism might play (see 3.4). Biological vision fills in the different blind spots.

There are about 6 to 7 million *cones* in each eye, most of them in the central portion of the retina, the so-called *fovea*. It is a relatively small region that extends to about 1 degree off the visual axis. The cones are only responsive to sufficiently high levels of irradiance. They underly *photopic* (day-light) vision and yield colour and sharp edge impressions. When luminance levels surpass 1 to 3 cd/m^2 the cones dominate perception and photopic vision conditions hold. The connection from cones to optic nerve is one-to-one, thereby preserving spatial detail (see section 3.5). The observer tends to keep the fovea fixated on the scene parts of interest, via a combination of gradual pursuit and sudden motions (so-called *saccades*).
There are even more *rods*, approximately 100 million. They are spread out more evenly over the whole retina and the outputs of several rods converges on a single nerve of the optic nerve. Hence, away from the fovea (where the cones are), visual resolution drops off heavily. Their irradiance thresholds are much lower. They are responsible for night (*scotopic*) vision. In a moonlit scene we do not observe colours since rods are insensitive to colour. Rods dominate perception up to a luminance of about $10^{-3}cd/m^2$.
Vision in the intermediate range between $10^{-3}cd/m^2$ and $1cd/m^2$ is based on the response of cones and rods. It is referred to as *mesopic* vision.

3.5 * II - The visual pathway

A sketchy overview of the *visual pathway*, i.e. the sequence of parts in the brain that are involved in visual processing, is shown in fig. 3.19.

The "gates" to the visual pathway are the eyes, already described in the previous section. It would be a mistake to compare the eye to nothing more than a pure input device though. In fact, filtering operations on the input are already carried out in the retina itself. A layer of *ganglion cells* that get their input from the photoreceptors apply a kind of bandpass filtering, both in space and time. In fact, many intermediary neurons are involved, such as bipolar and amachrine cells. The *receptive field* of a cell is the area on the retina to which a cell responds, and often the term is meant to include information on the spatial change in response within that area. Receptive fields come in different sizes. At the fovea, the cells with the smallest *receptive fields* are found. For a ganglion cell, the

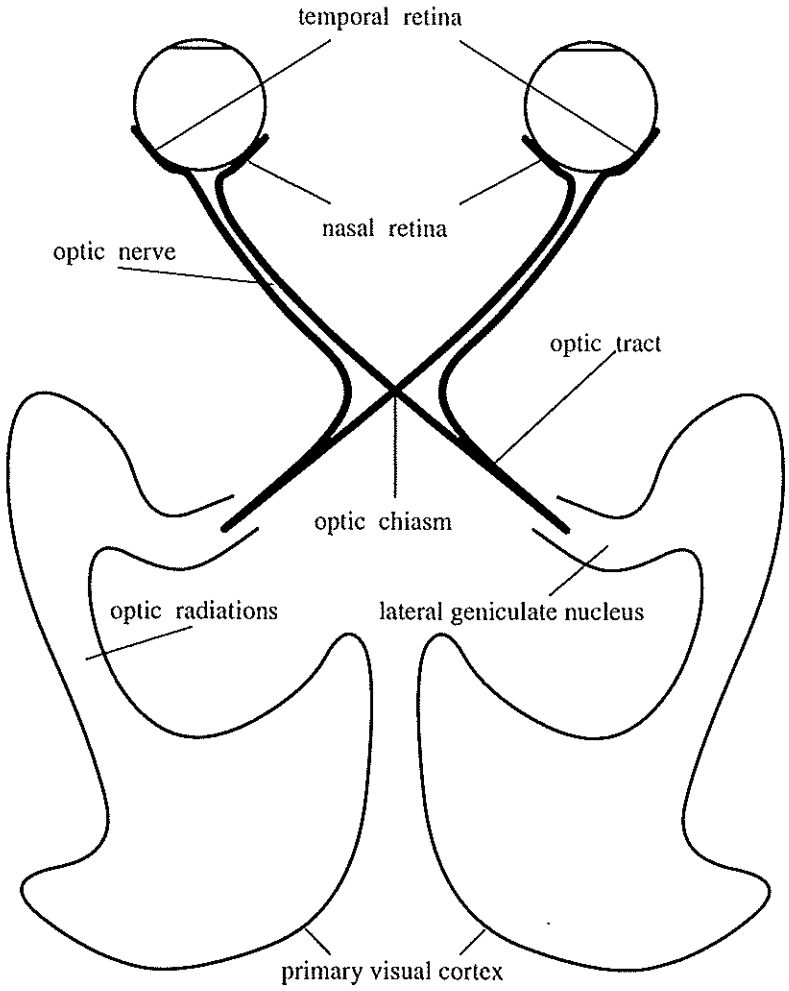

Figure 3.19: *Sketch of the visual pathway.*

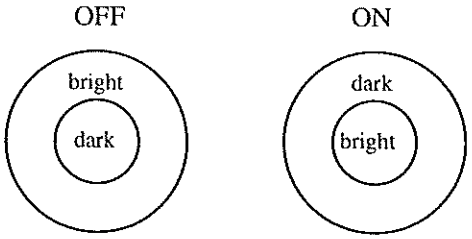

Figure 3.20: *On and off center ganglion cells.*

receptive field has a *center-surround* structure in that the cell responds to e.g. a dark spot on a bright background, or conversely, to a light spot on a dark background. These two subclasses of responses correspond to two families of ganglion cells, the former called *off cells*, and the latter *on cells*. A schematic drawing of the receptive fields for these classes is given in figure 3.20. In reality the transition in response between the two regions of the receptive field is gradual.

From the eye, the visual stimuli are transported via the *optic nerve*. Notice how the left and right parts of both visual fields are conveyed to the same brain half at the *optic chiasm*. Due to the reversal by the lens, the right half of the visual field is projected onto the *temporal retina* of the left eye and the *nasal part* of the right eye. The nasal parts of the retinas are the halfs on the side of the nose. The temporal parts are the opposite halfs. It is a widespread but erroneous belief that all information from the left eye is passed on to the right brain half and v.v. Rather, the light pattern from the *left part of the visual field* is processed by the right hemisphere of the brain and v.v.

From the optic chiasm, most of the fibers lead to the *lateral geniculate nuclei* (LGN). These thalamic bodies have often been compared with a relay station, but this is an oversimplification, since they receive extensive feedback from the parts further along the visual pathway (the cortex) also. Receptive fields for these LGN-cells are comparable to those of the retinal ganglion cells. The same center-surround structure is observed. At this level already some specialisation seems to occur and two main information streams have been hypothesized, the so-called parvo-cellular and magno-cellular systems. A clear insight in their role is lacking.

From the LGN, signals are transmitted to the primary visual cortex, via the *optic radiations* to the *calcarine sulcus*. This is the cortical concavity in which these neurons are found. From the primary visual cortex, connections fan out to several cortical regions. Processing in the primary visual cortex has been studied carefully. It is here that Nobel price winners Hubel & Wiesel found their orientation-specific cells, the so-called *simple cells*. These cells were found to act more or less like linear convolution filters (see chapters 5 and ??). Three typical receptive field types are shown in figure 3.21. The precise role of the other cortical regions is less clear. Some seem specialised in motion detection, and further away from the primary visual cortex, cells sensitive to specific shape icons were found, yet more abstract than replica images of your grandmother [5].

[5]It has been suggested for some time that individual persons and objects would have their own neuron

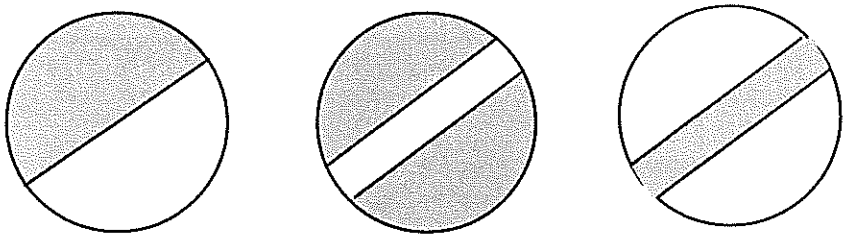

Figure 3.21: *Three receptive field structures of simple cells. Shading indicates the pattern to which the cells are maximally sensitive.*

Along a large part of the pathway, the information is *retinotopically* organised, meaning that the signals arising from nearby places in the retina are propagated and processed by nearby neurons. This is the case at least up to the first stages of cortical filtering in the primary visual cortex. Another typical aspect is that neurons are stacked in layers. Although many different types of receptive fields are found, usually neurons in the same layer belong to the same type. Processing in layers is found in the different structures along the pathway, including the retina, the lateral geniculate nucleus and the visual cortex. The pattern of processing is therefore one of projecting one retinotopic layer on the next, exploiting spatial organisation and massive parallelism. Forward connections are almost without exception accompanied by feedback connections. Simultaneously, there is a tendency for several neurons in one layer to map on the same neuron in the next. This systematic convergence explains that receptive fields get larger.

As far as the neurons themselves are concerned, the encoding of information has a kind of hybrid digital/analog nature. The *dendrites* of the neuron are branches that reach out to other neurons, from where they collect incoming information. The signals are transported to the neuron cell body, the *soma*. If the resulting change in membrane potential is sufficiently large, then the cell is activated. The information is carried as a *train of action potentials* (spikes), generated near the neuron cell body, and propagated along the *axon* to the dendrites of other neurons. Connections between neurons are called *synapses*. This brief description is rather caricature-like, but gives an idea.

3.6 * II - Primary visual cues

Not available yet, the text I mean...

for recognition somewhere in the brain. These hypothesized cells were called *grandmother cells*. Ideas of single-neuron representations have largely been abandonned and distributed representation of items by groups of neurons became the predominant view held in the community. The latter triggered renewed and intensive research into artificial neural networks. Nevertheless, face sensitive cells have been found.

3.7 * II - Grouping

Not available yet

3.8 * II - Further reading

Several excellent books on the perceptual and neurophysiological aspects of vision have been published recently. Of course, the reader will find many subjects in these references that haven't even been touched upon in this book. Hubel [1], who together with Wiesel won a Nobel-prize for their work on the primary visual cortex, gives a clear overview of visual functions in the early stages of the brain's visual pathway. Zeki's book [5] contains one of the best neurophysiological accounts ever, certainly for those who are interested in the global, architectural working principles of the visual cortex.
Both [2] and [4] contain several excellent tutorial texts on different aspects of both the perception and neurophysiology of vision.

Bibliography

[1] D. Hubel, *Eye, Brain, and Vision*, Scientific American Library, 1988

[2] K. Leibovic, ed., *Science of Vision*, Springer, NY, 1990

[3] R. McCluney, *Introduction to Radiometry and Photometry*, Artech House, Boston, 1994

[4] L. Spillmann and J. Werner, eds., *Visual Perception, the Neurophysiological Foundations*, Academic Press, San Diego, 1990

[5] S. Zeki, *A Vision of the Brain*, Blackwell Scientific Publications, Oxford, 1993