APPENDIX A

HUMAN VISUAL PERCEPTION

A.1 INTRODUCTION

The human visual system (HVS) is the final link in the perception of images and video sequences. A clear understanding of its capabilities and limitations can lead to better image and video processing solutions. In applications whose ultimate goal is to improve the image quality for human consumption, this knowledge allows designers to establish objective performance criteria and quality measures. In machine vision systems (MVS) whose goal is to emulate—and ultimately outperform—their human counterpart, it is absolutely necessary that we know how the human visual system works, which performance limits it imposes, and how this knowledge can be factored into the design of MVS.

In this appendix, we will provide a very brief overview of the human visual system with emphasis on aspects that are relevant—some may say essential—to the researcher and practitioner in the field. This is a long, deep, and fascinating topic for which there are almost as many open questions as there are answers. The interested reader may refer to the "Learn More About It" section at the end of the appendix for suggestions on books and other references that will broaden and deepen their understanding of the field of human vision science.

Practical Image and Video Processing Using MATLAB®. By Oge Marques.

A.2 THE HUMAN EYE

The HVS has two main components, the eye (input sensor) and the brain (information processing unit), connected by the optic nerve (transmission path) (Figure A.1). Image perception consists of capturing the image with the eye, then recognizing it, and finally interpreting its contents in the brain. First, light energy is focused by the lens of the eye onto the sensors on the retina, and then those sensors respond to light energy by an electrochemical reaction that sends an electrical signal down the optic nerve to the brain. The brain uses these nerve signals to create neurological patterns that we perceive as images. In this section, we look at selected anatomical and physiological aspects of the human eye.

Figure A.2 shows a simplified cross section of the human eye. The following are some of the anatomical properties of the eye that are of interest for our discussion:

- The eye contains a *lens* responsible for focusing an image. The movements of the lens are controlled by specialized *ciliary muscles*.
- The anterior portion of the lens contains an *iris diaphragm* responsible for controlling the amount of light that enters the eye. The central opening of the iris is called *pupil* and varies its diameter—approximately from 2 to 8 mm—in a way that is inversely proportional to the amount of incoming light.
- The innermost membrane of the eye is the *retina*, which is coated with photosensitive receptors called *cones* and *rods*. It is on the surface of the retina that an upside-down image of the scene is formed.

The combination of lens, diaphragm, and a back-projection surface is also present in a rudimentary camera, leading to a very popular analogy known as "the eye-camera analogy" (Figure A.3).

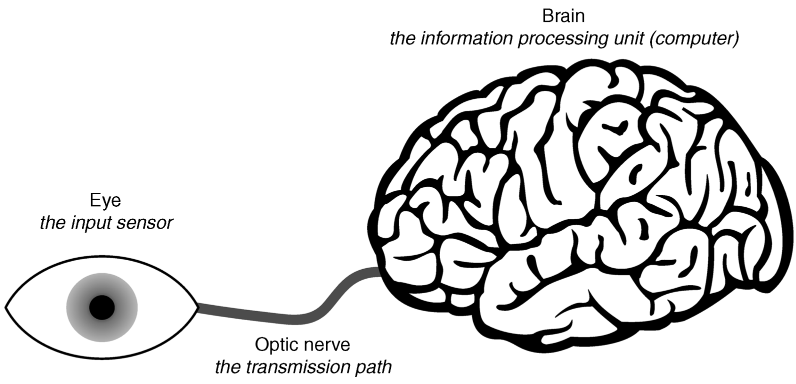

FIGURE A.1 Simplified view of the connection from the eye to the brain via the optic nerve. Adapted and redrawn from [Umb05].

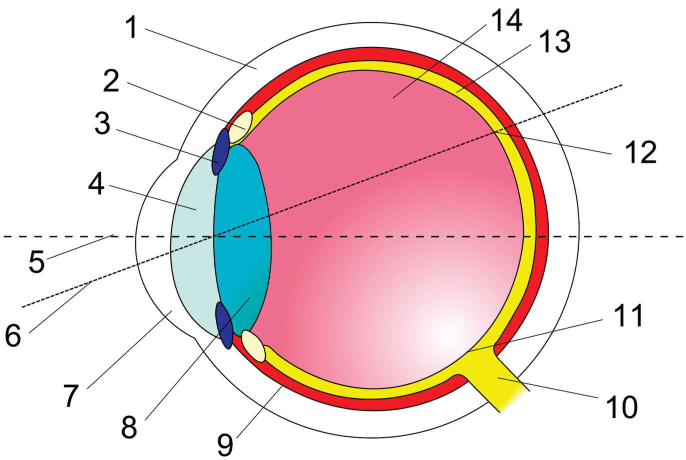

FIGURE A.2 The eye: a cross-sectional view. 1, sclera; 2, ciliary body; 3, iris; 4, pupil and anterior chamber filled with aqueous humor; 5, optical axis; 6, line of sight; 7, cornea; 8, crystalline lens; 9, choroid; 10, optic nerve; 11, optic disk; 12, fovea; 13, retina; 14, vitreous humor. Courtesy of Wikimedia Commons.

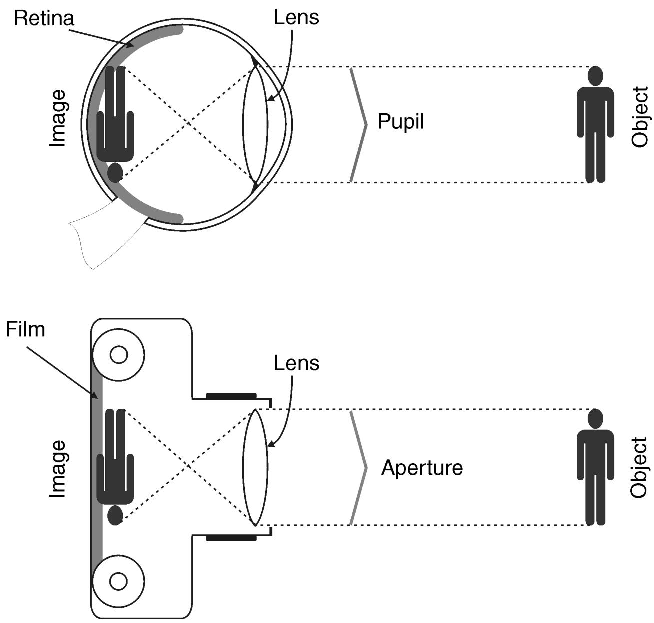

FIGURE A.3 The eye-camera analogy. Adapted and redrawn from [Pal99].

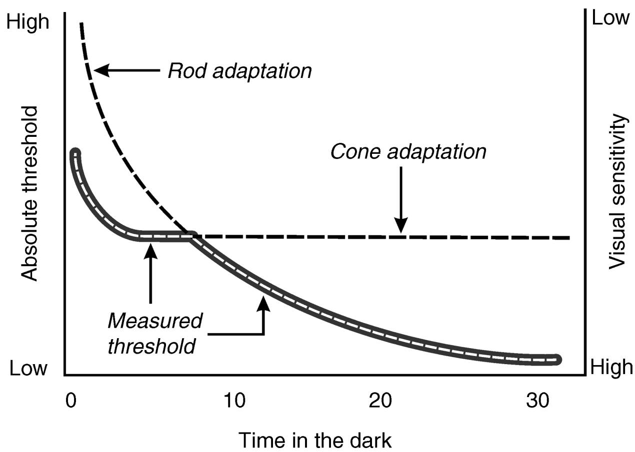

FIGURE A.4 Dark adaptation. Adapted and redrawn from [Pal99].

The surface of the human retina is coated with discrete photoreceptors, capable of converting light into electrochemical reactions that will eventually be transmitted to the brain. There are two types of photoreceptors, *cones* and *rods*, whose names were given based on their overall shape.

Cones (typically 6–8 million in total) are primarily concentrated in the *fovea*—the central part of the retina, aligned with the main visual axis—and are highly sensitive to color. Cones work well only in bright-light (*photopic*) vision scenarios, though. Under low lighting levels, they are not active and our ability to discriminate colors decreases dramatically.[1] Cones come in three varieties—*S*, *M*, and *L*, as in *short*, *medium*, and *long* (wavelengths), roughly meaning light in the red, green, and blue portions of the visible spectrum—each of which is primarily sensitive to certain wavelengths (and the colors associated with them).

The existence of three types of cones provides a physiological basis for the *trichromatic theory of vision*, postulated by Thomas Young in 1802, more than 150 years before it became possible to obtain physiological evidence of the existence of the three types of cones.

Rods outnumber cones (there are 75–150 million rods) and are primarily not concentrated in the fovea; instead, they are distributed over the entire retinal surface (except for the *optic disk*, a region of the retina that corresponds to the perceptual *blind spot*). Rods are not sensitive to color, but are sensitive to low levels of illumination, and therefore responsible for dim-light (scotopic) vision.

Figure A.4 shows how the sensitivity of the retina increases in response to decreasing incoming light in a process known as *brightness adaptation* or *dark adaptation*. The left portion of the curve corresponds to the adaptation experienced by the cones

[1]Anyone who has forgotten where he parked their car and tries to look for it after dark on a poorly lit parking lot knows firsthand how true this is.

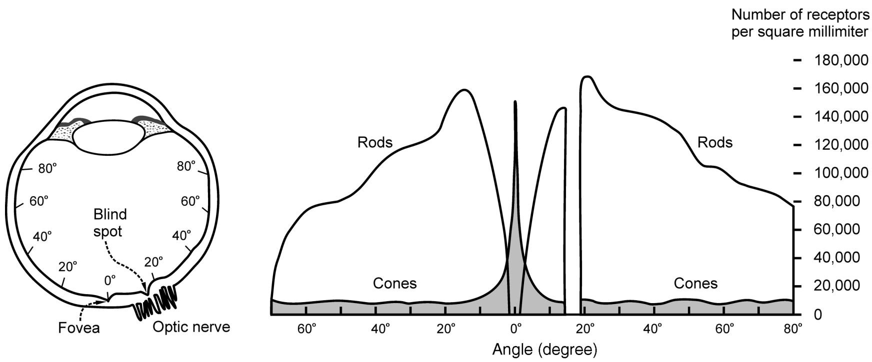

FIGURE A.5 Distribution of rods and cones in the human retina for the right eye (seen from the bottom). Adapted and redrawn from [Ost35].

(photopic vision). The right part indicates the range of time beyond which scotopic vision (primarily peripheral rod vision) becomes prevalent.

Figure A.5 shows the distribution of rods and cones in the retina. The pronounced peak at the center of the fovea (0° in relation to the visual axis) is indicative of the concentration of cones in that region. Note also that there are no cones or rods at a small region about 20° from the optical axis (toward the nasal side of each eye) known as the *blind spot*. This is a constructive limitation, after all the optic nerve must be attached to the retina at some point. It is truly remarkable that the brain "fills in" and allows us to see entire scenes even though some of the light reflected by objects in those scenes falls onto our blind spots (one for each eye).

A significant amount of visual processing takes place in the retina, thanks to a series of specialized (horizontal, bipolar, anacrine, and ganglion) cells. A detailed explanation of these cells and their behavior is beyond the scope of this appendix. It is important to note that although there are more than 100 million light receptors in the retina, the optic nerve contains only a million fibers, which suggests that a significant amount of data processing occurs before the electric impulses ever reach the brain. An additional curiosity is the fact that, since the human eye has evolved as an outgrowth of the brain, the retina appears to have been designed from the inside out: the photoreceptors are not the innermost cells; in fact, surprisingly enough, they point away from the incoming light.

After a retinal image is formed, it is converted into electric signals that traverse the optic nerve, cross the optic chiasm (where the left and right halves of the visual field of each eye cross), and reach the lateral geniculate nucleus (LGN) on its way to a region of the occipital lobe of the brain involved in visual perception: the *visual cortex*. The exact nature of the processing that occurs in the visual cortex, which cells and regions are in charge of what, and the number of possible visual pathways from photoreceptors in the retina to higher order regions of the cortex are among the many aspects over which there has been an enormous amount of research but not much agreement among researchers. Refer to the "Learn More About It" section at the end of the appendix for useful references.

A.3 CHARACTERISTICS OF HUMAN VISION

In this section, we look at a selected subset of characteristics of human vision that are of interest to the image and video processing researcher and system designer. Our goal is to highlight widely accepted facts about the way we perceive properties of scenes, such as brightness, contrast, sharpness (fine detail), color, motion, and flicker. We do not attempt to explain *why this is so* (we leave that to human vision researchers), but instead take an engineering approach and provide qualitative and—whenever possible—quantitative information that can be used in the design of imaging systems.

A.3.1 Resolution, Viewing Distance, and Viewing Angle

Resolution can be defined as the ability to separate two adjacent pixels, that is, resolve the details, in a test grating (such as the EIA test pattern shown in Figure A.6) or any other image. This ability depends on several factors, such as the picture (monitor) height (h) and the viewer's distance from the monitor (d), and the subtended viewing angle (θ) (Figure A.7).

The measure of the number of changes in image intensity for a certain test grating is referred to as its *spatial frequency*. The spatial frequency can be completely characterized by the variation frequencies in two orthogonal directions (e.g., horizontal and vertical). If we call f_x the horizontal frequency (expressed in cycles/horizontal

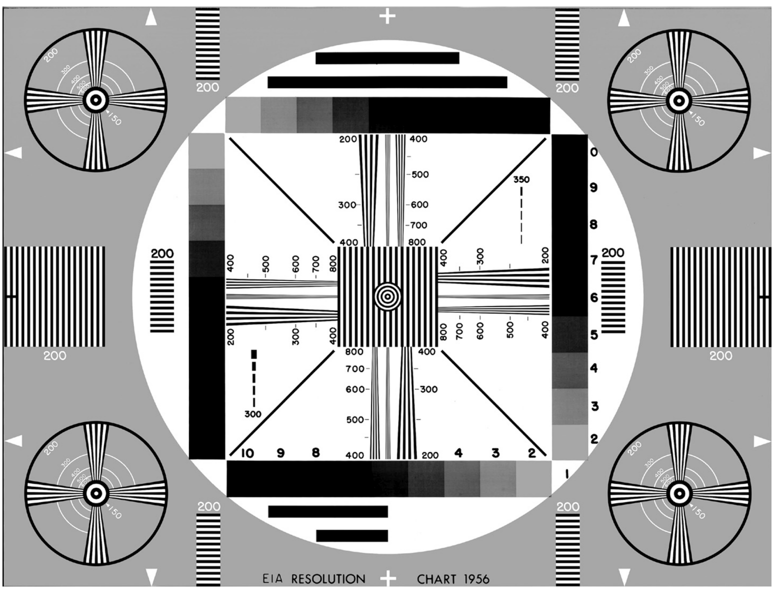

FIGURE A.6 EIA 1956 standard test pattern. Courtesy of http://www.bealecorner.com/trv900/respat/.

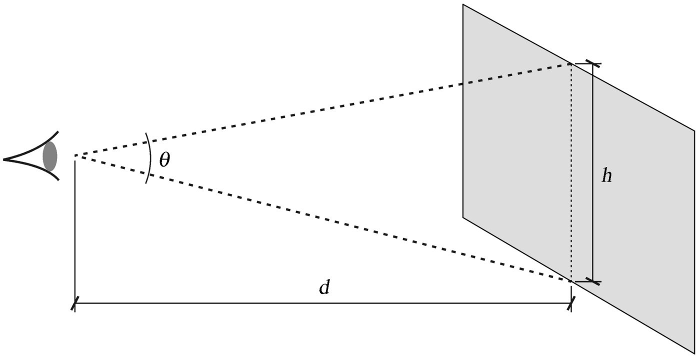

FIGURE A.7 Angular frequency concept.

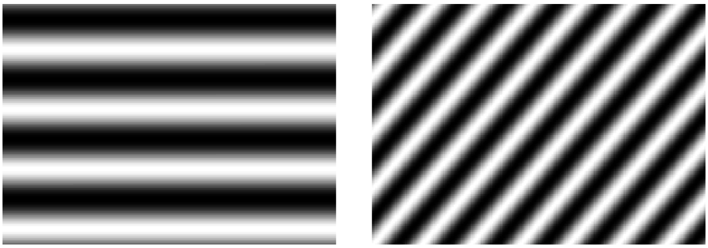

FIGURE A.8 Sinusoidal gratings commonly used for measures of resolution—based on MATLAB code by Alex Petrov: http://alexpetrov.com/softw/utils/.

unit distance) and f_y its vertical counterpart, the pair (f_x, f_y) characterizes the spatial frequency of a 2D image. For example, the test gratings on Figure A.8 have $(f_x, f_y) = (0, 4)$ (Figure A.8a) and $(f_x, f_y) = (7, 4.5)$ (Figure A.8b). These two values can be combined and expressed in terms of magnitude (f_m) and angle (θ):

$$f_\mathrm{m} = \sqrt{\left(f_x^2 + f_y^2\right)} \tag{A.1}$$

$$\theta = \arctan\left(\frac{f_y}{f_x}\right) \tag{A.2}$$

The examples in Figure A.8 have $(f_\mathrm{m}, \theta) = (4, 90°)$ and $(f_\mathrm{m}, \theta) = (8.32, 32.7°)$.

The spatial frequency of a test signal (as defined above) is not very useful in determining the user's perception of those signals because it does not account for the viewing distance. A more useful measure, which not only is a characterization of the signal, but also takes into account the viewing distance and the associated viewing angle, is the angular frequency[2] (f_θ), expressed in cycles per degree (cpd) of viewing angle and defined as follows (Figure A.7):

$$\theta = 2\arctan\left(\frac{h}{2d}\right) \approx \frac{h}{2d}(\text{radian}) = \frac{180h}{\pi d}\ (\text{degrees}) \tag{A.3}$$

$$f_\theta = \frac{f_s}{\theta} = \frac{\pi d}{180h} f_s\ (\text{cpd}) \tag{A.4}$$

A careful look at equation (A.4) reveals that, for the same picture (e.g., grating pattern) and a fixed picture height (PH), the angular frequency increases with the viewing distance; conversely, for a fixed viewing distance, larger display sizes lead to lower angular frequencies. This is consistent with our experience: the same test pattern appears to change more frequently if viewed from farther away and more slowly if displayed on a larger screen.

A common practical application of the concepts of resolution and viewing angle is the determination of the optimal viewing distance for a certain display. Although accurate calculations must take into account other parameters such as display brightness, ambient lighting, and visual acuity of the observer, a typical back of the envelope estimate of the optimal viewing distance can be obtained by assuming a subtended angle of 1 min of arc (1/60 of a degree) between two adjacent TV lines. For conventional standard definition TV (SDTV) displays (480 scan lines and a 4:3 aspect ratio), a viewing distance of about 7 $\times$ PH is usually recommended, whereas typical high-definition TV (HDTV) displays of the same height (with 1080 scan lines and a 16:9 aspect ratio) should be viewed at slightly less than half the distance (3.1 $\times$ PH) to fully appreciate the additional amount of detail available and maintain the same level of spatial frequency eye discrimination (also known as *visual acuity*) (Figure A.9). The horizontal picture angles at those viewing distances are 11° and 33°, respectively (Figure A.10).

A.3.2 Detail and Sharpness Perception

The ability to perceive fine details in an image or video sequence is one of the most important guiding factors in the design of image and video systems since it impacts parameters such as image definition, signal to noise ratio (SNR), and bandwidth. Perception of detail is intimately associated with the concept of *visual acuity*, which can be described as "the smallest angular separation at which individual lines in a grating pattern can be distinguished." The most familiar experience with visual acuity measures you may have had is a visit to the optometrist. In such visit, you are asked to read numbers or letters on a Snellen chart at a standardized test distance of 20 ft (6.1 m). At that distance, the width of the strokes of the letters in the 20/20 row subtends an angle 1 min of arc. Being able to read that row is considered the standard for normal (20/20) vision. Visual acuity varies widely, from $0.5'$ to $5'$ (minutes of arc),

[2] Unfortunately, the expression *angular frequency* is not widely adopted in the literature; readers will often find the same concept expressed as *spatial frequency* instead, which is potentially confusing.

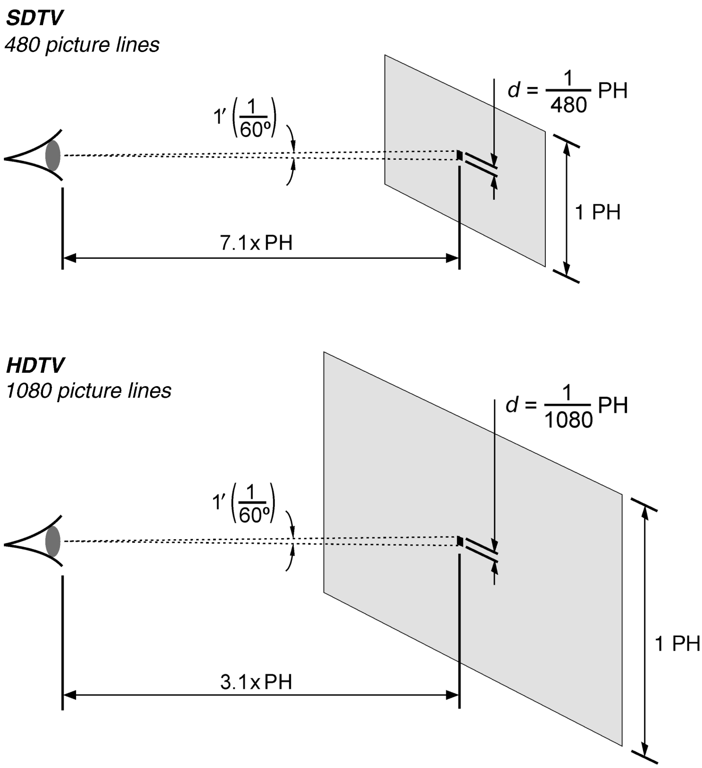

FIGURE A.9 Viewing distance for SDTV and HDTV displays. Adapted and redrawn from [Poy03].

depending on the contrast ratio and the quality of vision of each individual. An acuity of $1.7'$ is usually assumed in the design of image and video systems [LI99].

A.3.3 Optical Transfer Function and Modulation Transfer Function

Most lenses, including the human eye's lens, are not perfect optical systems. Consequently, when visual stimuli with a certain amount of detail are passed through them, they may show a certain degree of degradation. The optical transfer function (OTF) is a way to measure how well spatially varying patterns are observed by an optical system, that is, a way to evaluate the extent of degradation. The OTF is usually expressed as a series of complex numbers—one for each spatial frequency—whose amplitude represents the reduction in signal strength and phase represents the corresponding phase shift. For the sake of simplicity, we will focus primarily on the amplitude component, which is known as *modulation transfer function* (MTF). The MTF is the spatial equivalent of frequency response in electronic circuits.

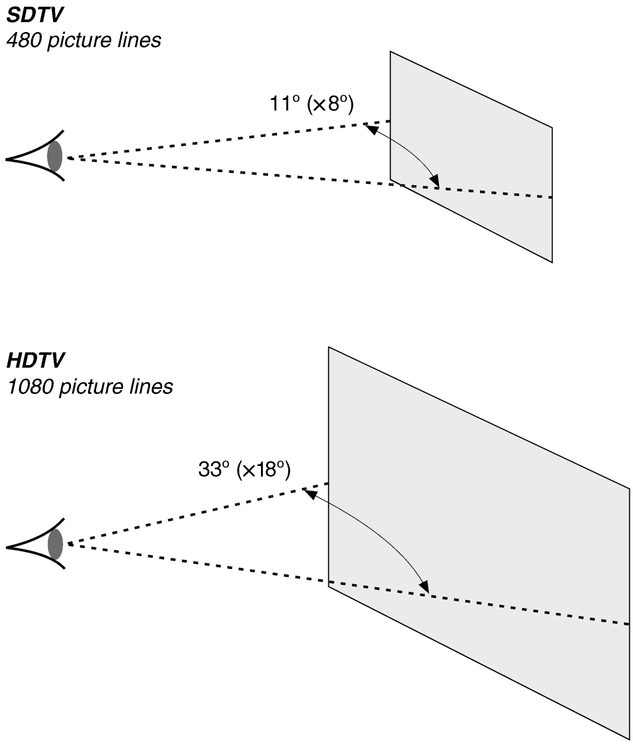

FIGURE A.10 Picture (viewing) angles for SDTV and HDTV displays. Adapted and redrawn from [Poy03].

Figure A.11 illustrates the MTF concept. Part (a) introduces the contrast index (CI), a measure of the amplitude differences between the darkest and brightest portions of the test image. Part (b) shows that a nonideal optical system tested with an input test image with constant CI will exhibit a CI that falls for higher spatial frequencies. Part (c) displays the MTF, which is the ratio between the output CI and the input CI. It is worth noting that while the MTF resolution test illustrated in Figure A.11 provides an objective evaluation of the possible optical degradation experienced by a test signal, the human perception of sharpness is subjective and it is also affected by contrast. As a result, an increase in the image contrast will cause an increased sensation of sharpness even though the MTF is unchanged, an aspect that has been exploited in the design of image display devices.

A.3.4 Brightness Perception

Brightness can be defined as "the attribute of a visual sensation according to which an area appears to emit more or less light" [Poy03].

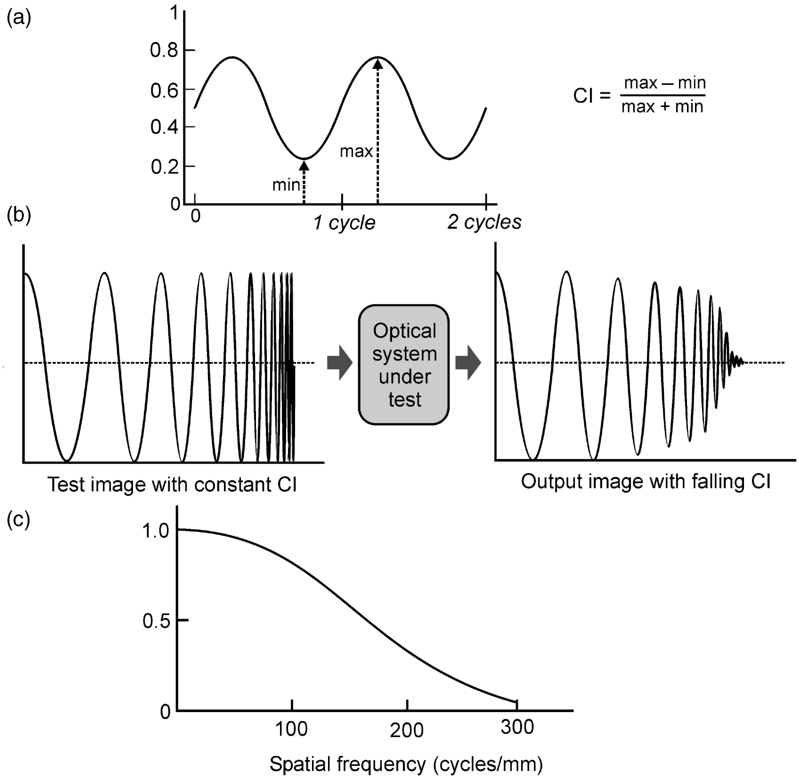

FIGURE A.11 (a) The definition of contrast index; (b) A test image with constant CI results in an output image with falling CI; (c) modulation transfer function: the ratio of output and input CIs. *Note*: When LF response is unity, CI and MTF are interchangeable. Redrawn from [Wat00].

The process of subjective brightness perception in humans is such that the perceived subjective brightness is proportional to the logarithm of the luminous intensity incident on the eye, as shown in Figure A.12. The long solid curve represents the (remarkably high, about eight orders of magnitude) range of luminous intensities that the HVS can adapt to. The plot also shows the transition between scotopic and photopic vision that takes place at low intensity levels. The region where there is overlap between cone-based and rod-based vision is called *mesopic vision*. More important, the small segment at the middle of the curve illustrates the phenomenon of *brightness adaptation*. In this case, after having adapted to a certain brightness level (B_a), the eye is capable of responding to stimuli around that value, provided that they are above another level (B_b in the figure). Any intensities below that will not be perceived. Should the average ambient intensity increase (or decrease), the eyes will adapt to another point in the main solid curve.

The perceived brightness of an area (or object) within an image also depends on the contrast between the object and its surroundings, in what is known as *simultaneous*

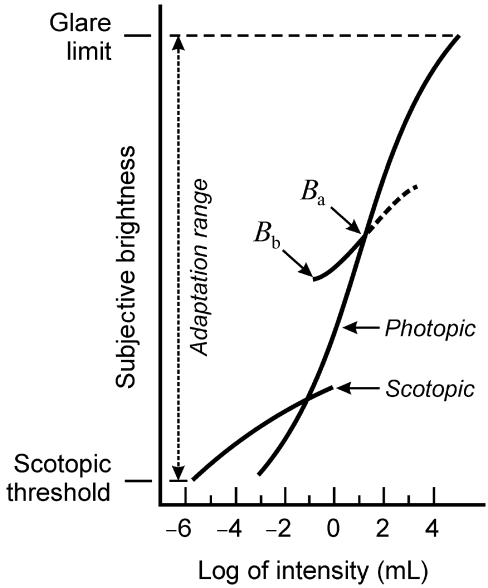

FIGURE A.12 Range of subjective brightness sensations showing a particular adaptation level. Redrawn from [GW08].

contrast (Figure A.13). In other words, we do not perceive gray levels as they are, but in terms of how they differ from their surroundings.

Another well-known way to show that perceived brightness is not a simple function of luminous intensity are the Mach bands (Figure A.14), named after Ernst Mach, who first described the phenomenon in 1865. These bands show that our visual system tends to undershoot or overshoot at the boundaries of regions with different intensities. This is due to the fact that the eye possesses a lower sensitivity to high- and low-spatial frequencies than to intermediate frequencies. It explains our ability to distinguish and separate objects, even in dimly lit scenes, thanks to the accentuated response around the actual edges. A possible implication of this property for designers of imaging

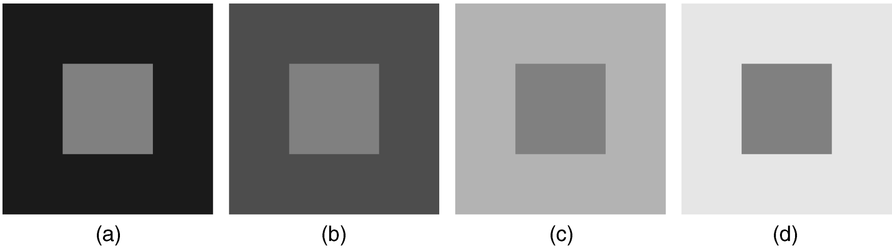

FIGURE A.13 Simultaneous contrast: the center square is perceived as progressively darker as the background becomes brighter (from (a) to (d)) even though it is identical in all four cases.

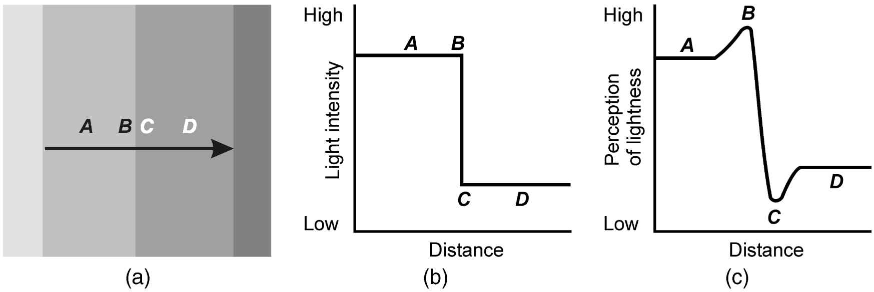

FIGURE A.14 Mach bands.

systems is that the reproduction of perfect edges is not a critical requirement, owing to the eye's imperfect response to high-frequency brightness transitions.

A.3.5 Contrast Ratio and Contrast Sensitivity Function

Contrast ratio is "the ratio of luminances of the lightest and darkest elements of a scene" [Poy03]. Typical contrast ratios are 80:1 (movie theater), 20:1 (TV in a living room), and 5:1 (computer monitor in an office).

The contrast sensitivity of the eye is defined as the smallest brightness difference that can be detected by an observer. Contrast sensitivity is usually measured as the ratio of luminances between two adjacent patches combined in a test pattern that is presented to a human subject (Figure A.15). The observer's field of vision is filled mostly by the surround luminance (Y_0). In the central area, the left portion of the circle has a test luminance value (Y), whereas the right half shows a slightly increased value ($Y + \Delta Y$). Subjects are asked to inform at which point the difference between the two

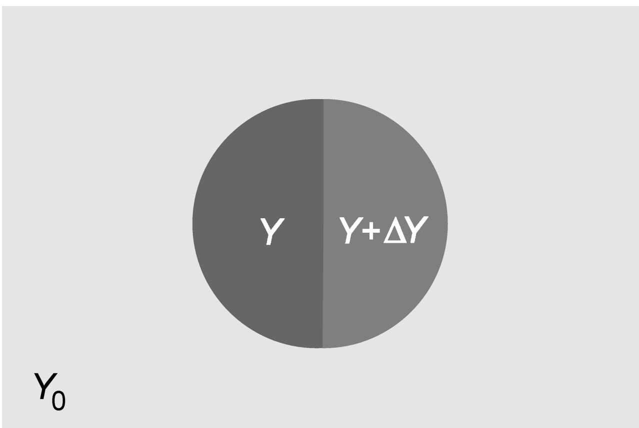

FIGURE A.15 Contrast sensitivity test pattern.

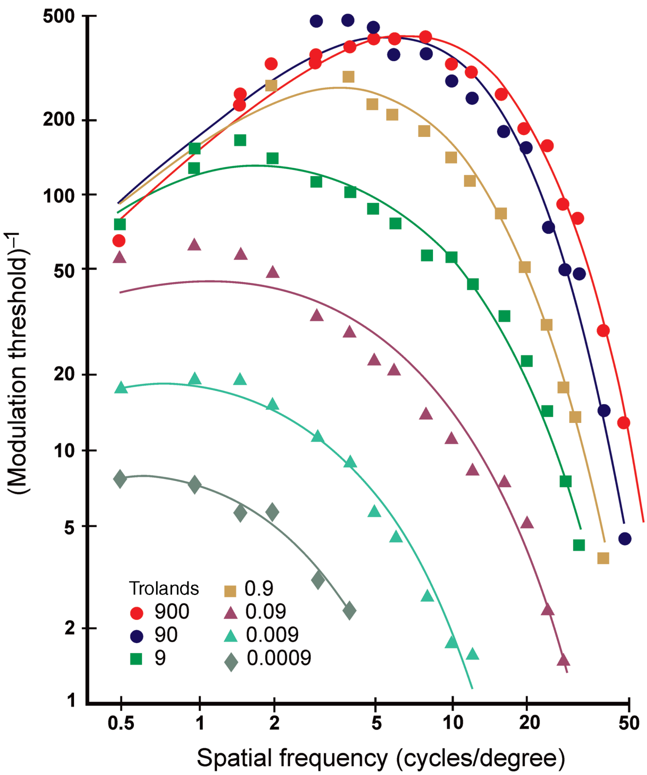

FIGURE A.16 Contrast sensitivity function for various retinal illuminance values (expressed in Td). Redrawn from [VNB67].

halves become noticeable[3] and the corresponding value of Y and ΔY are recorded. The process is repeated for a wide range of luminance values.

Experiments of this type have concluded that over a range of intensities of about 300:1, the discrimination threshold of vision is approximately a constant ratio of luminance. If one plots $\log(\Delta Y/Y)$ as a function of Y, it will show an interval of more than two decades of luminance over which the discrimination capability of vision is about 1% of the test luminance level. In other words, within that range, human vision cannot distinguish two luminance levels if the ratio between them is less than approximately 1.01.

In vision science, contrast sensitivity is also measured using a spatial grating test pattern. The resulting plot is called *contrast sensitivity function* (CSF) and it represents the contrast sensitivity as a function of the spatial frequency (in cycles/degree). Figure A.16 shows a family of curves, representing different adaptation levels, from very dark (0.0009 Td) to very bright (900 Td), where 9 Td is a representative value for

[3]This concept of *just noticeable difference* (JND) is also used in many other psychophysics experiments.

electronic displays.[4] The 9 Td curve peaks at about 4 cycles/degree. Below that spatial frequency, the eye acts as a differentiator; for higher spatial frequencies, the eye act as an integrator. Three important observations can be derived from this graph [Poy03]:

- Beyond a certain spatial frequency (around 50 cycles/degree for the 9 Td curve), the contrast sensitivity falls to very low values (less than 1% of the maximum value), which means our vision cannot perceive spatial frequencies greater than that. The implication for image and video systems designers is that there is no need to provide bandwidth or display resolution for those higher frequency contents, since, ultimately, they will not be noticed.
- Each curve peaks at a contrast sensitivity value that can be used to calculate the number of bits per pixel that must be used to quantize the image. Using more bits per pixel will allow the representation of luminance differences too subtle to be perceived by the human eye.
- The curve falls off for spatial frequencies lower than 1 cycle/degree, which suggests that luminance can be lower (within reasonable limits) in areas closer to the edges of the image without the viewer noticing it.

A.3.6 Perception of Motion

The electrochemical processes associated with the processing of incoming light in the human eye require several milliseconds to be performed and therefore act as a smoothing filter in the temporal domain: what the brain reconstructs is a time-averaged version of the actual input. As a result of this temporal smoothing, there is a *critical flicker*[5] *frequency* (CFF) below which we perceive the individual flashes of blinking light and above which those flashes merge into a continuous, smooth moving image sequence. This is a fundamental property in the design of movie, TV, and video systems. The CFF is directly proportional to the picture luminance and screen size and inversely proportional to the viewing distance.

The temporal frequency response of the HVS depends on several factors, such as viewing distance, display brightness, and ambient lighting. Figure A.17 shows the result of an experiment in which subjects where presented a flat screen whose brightness was modulated by a sinusoidal signal and were instructed to report the lowest modulation level at which the flicker became just noticeable. The reciprocal of that modulation level is referred to as contrast sensitivity and plotted versus the frequency of the modulating signal. Several conclusions can be derived from Figure A.17:

- The temporal response of the HVS is similar to a bandpass filter that peaks at intermediate frequencies and falls off quickly afterward.

[4] A *troland* (Td) is a unit of retinal illuminance equal to object luminance (in cd/m^2) times pupillary aperture area (in mm^2).

[5] The term *flicker* is also used to indicate an image defect usually caused by inadequate frame repetition rate, lower than the eye's temporal cutoff frequency.

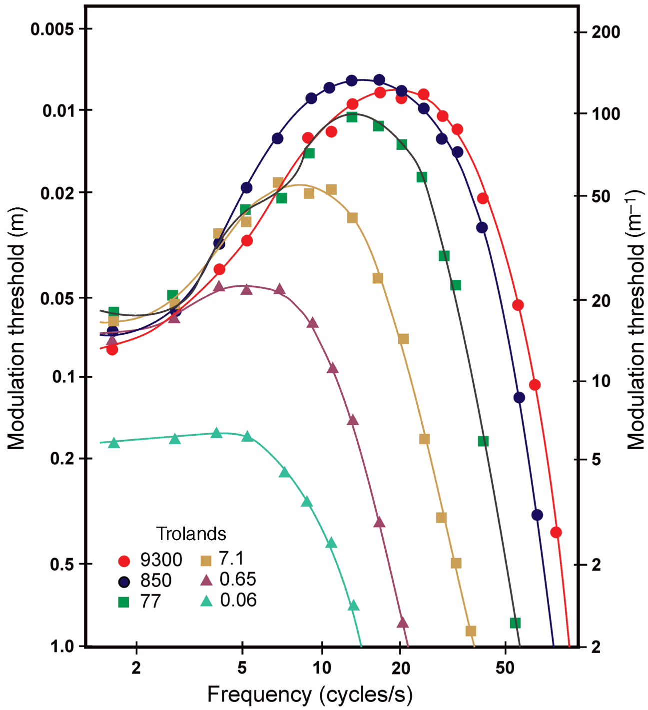

FIGURE A.17 Temporal frequency response of the HVS. Redrawn from [Kel61].

- The peak increases with the mean brightness of the image.
- One reason why the eye has reduced sensitivity at high frequencies is because it can retain the sensation of an image for a short interval after the image has been removed, a phenomenon known as *persistence of vision*.
- The critical flicker frequency is directly proportional to the average brightness of the display. In Figure A.17, the critical flicker frequency varies between 20 and 80 Hz.

A.3.7 Spatiotemporal Resolution and Frequency Response

After having seen the spatial and temporal frequency responses separately, we turn our attention to their combined effect. Figure A.18 shows experimental results by Robson [Rob66]. Figure A.18a shows that at higher temporal frequencies, both the peak and cutoff frequencies in the spatial frequency response shift downward. They also help confirm our intuitive expectation that the eye is not capable of resolving high spatial frequency details when an image moves very fast, compared to its spatial resolution capabilities for static images (Figure A.18b). The key implication of this finding for the design of TV and video systems is that it is possible to trade off spatial

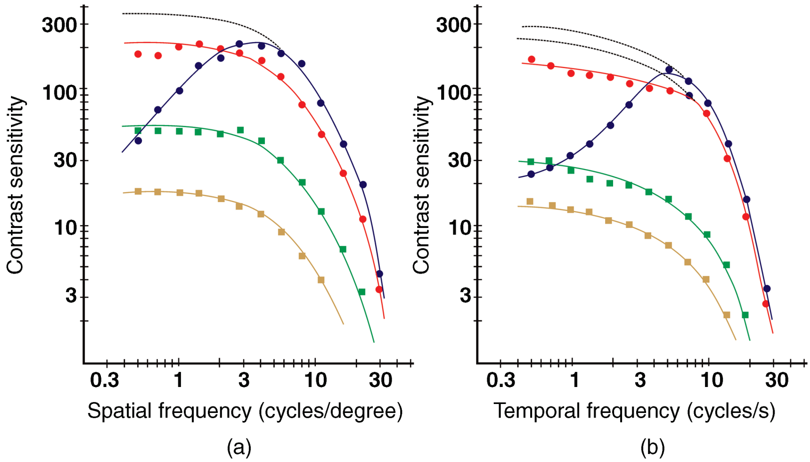

FIGURE A.18 Spatiotemporal frequency response of the HVS: (a) spatial frequency responses for different temporal frequencies (in cpd); (b) temporal frequency responses for different spatial (angular) frequencies (in Hz). Redrawn from [Rob66].

resolution with temporal resolution and vice versa. In Chapter 20, we discussed how this fact is exploited in the design of TV systems using interlaced scans.

The discussion thus far has assumed that the eye is not actually tracking any particular object within the image sequence, which is an obvious simplification that does not correspond to reality. Before we examine the different patterns that arise when tracking takes place, let us define a few important concepts regarding eye movements and their role in human visual perception.

The retina does not respond to the incoming light instantly; it requires between 0.15 and 0.30 s before the brain perceives an image. The early stages are often referred to as *preattentive* vision. After 0.15 s or so have elapsed, attentional mechanisms (some of which are triggered by the presence of salient objects in the scene in a *bottom-up* fashion, while other are dependent on the visual task at hand, the *top-down* component) are factored in and help guide the eyes toward regions of interest within the scene. The result of eye movements around a scene is called a *scanpath*.

Part of the eye movements registered in a scanpath results from involuntary unconscious vibrations known as *saccades*. The eye has a temporal filter mechanism that integrates spatial information from different positions of the retina covered by saccadic eye movements. This temporal filtering is responsible for the phenomenon of "persistence of vision" described earlier in this appendix.

Figure A.19 shows how the perceived temporal frequency changes as a function of eye movements. In both cases, the same object with the same amount of detail moves across the observer's field of view with the same constant speed. The temporal frequency is the product of the amount of detail in the object (usually expressed in lines/mm) and its speed. In Figure A.19a, the eye is fixed and the perceived temporal frequency is high, resulting in motion blur. In Figure A.19b, the eye tracks the moving

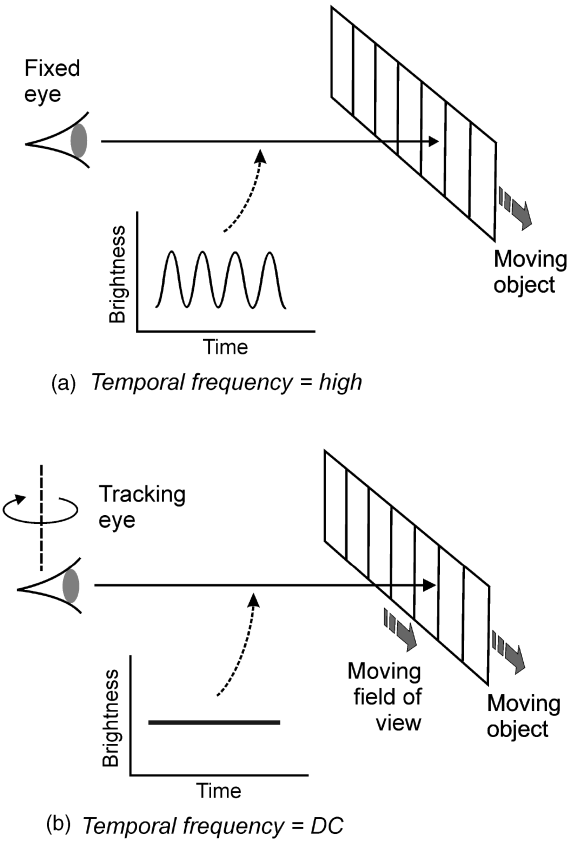

FIGURE A.19 Temporal frequency as a function of eye movements. Redrawn from [Wat00].

object (*smooth pursuit eye movement*), resulting in a temporal frequency of zero and improved ability to resolve spatial detail, besides the absence of motion blur. This is known as *dynamic resolution* and it is how humans judge the ability of reproducing detail in real moving pictures.

A.3.8 Masking

Masking is the reduction in the visibility of one image component (the target) due to the presence of another (the masker). The HVS is subject to several masking phenomena, such as [RP00]:

- *Texture Masking*: Errors in textured regions are usually harder to notice, whereas the HVS is very sensitive to errors in uniform areas.
- *Edge masking*: Errors near the edges are harder to notice.

- *Luminance Masking*: Visual thresholds increase with background luminance, as a result of brightness adaptation. Moreover, higher luminance levels increase the flicker effect.
- *Contrast Masking*: Errors (and noise) in light regions are harder to perceive as a result of the property of the HVS by which the visibility of an image detail is reduced by the presence of another.

A.4 IMPLICATIONS AND APPLICATIONS OF KNOWLEDGE ABOUT THE HUMAN VISUAL SYSTEM

In this section, we summarize some of the most relevant properties of the HVS that have implications for designers of image and video processing systems. They are listed as follows [Ric02]:

- The HVS is more sensitive to high contrast than low contrast regions within an image, which means that regions with large luminance variations (such as edges) are perceived as particularly important and should therefore be detected, preserved, and enhanced.
- The HVS is more sensitive to low spatial frequencies (i.e., luminance changes over a large area) than high spatial frequencies (i.e., rapid changes within small areas), which is an often exploited aspect of most image and video compression techniques. Discarding redundant *low* spatial frequency contents (while preserving edges) leads to computational savings.
- The HVS is more sensitive to image features that persist for a long duration, which means that it is important to employ techniques that minimize temporally persistent disturbances or artifacts in an image.
- HVS responses vary from one individual to the next, which means that subjective evaluations of image and video systems must be conducted with a large number of subjects. This aspect also reinforces the need to find quantitative measures of image and video quality that can be automatically calculated from the pixel data and yet reflect the subjective notion of perceived quality.

LEARN MORE ABOUT IT

The following are some of the books on human visual perception and related fields that may be of interest:

- Goldstein, E. B., *Sensation & Perception*, 7th ed., Belmont, CA: Thomson Wadsworth, 2007.
- Bruce, V., Green, P. R., and Georgeson, M.A., *Visual Perception: Physiology, Psychology and Ecology*, Philadelphia, PA: Psychology Press, 2003.

- Purves, D. and Lotto, R.B., *Why We See What We Do*, Sunderland, MA: Sinauer Associates, 2003.
- Yantis, S. (Ed.), *Visual Perception: Essential Readings*, Philadelphia, PA: Psychology Press, 2000.
- Palmer, S. E., *Vision Science: Photons to Phenomenology*, Cambridge, MA: Bradford Books/MIT Press, 1999.
- Gregory, R. L., *Eye and Brain: The Psychology of Seeing*, Princeton, NJ: Princeton University Press, 1998.
- Rodieck, R. W., *The First Steps in Seeing*, Sunderland, MA: Sinauer Associates, 1998.
- Wandell, B. A., *Foundations of Vision*, Sunderland, MA: Sinauer Associates, 1995.

The following are some of the scientific journals that publish research results in vision science and related areas (in alphabetical order): *Cognition*, *Journal of Vision*, *Nature*, *Nature Neuroscience*, *Perception*, *Perception & Psychophysics*, *Science*, *Spatial Vision*, *Vision Research*, *Visual Cognition*, and *Visual Neuroscience*.

ON THE WEB

- *Vision Science Portal*: an Internet resource for research in human and animal vision, with links to relevant conferences, journals, research groups, software, and much more.
  http://visionscience.com/
- *MATLAB Psychophysics Toolbox*: a widely used (and well-documented) collection of MATLAB functions for psychophysics experiments.
  http://psychtoolbox.org/
- *Project LITE (Boston University)*: a great collection of interactive visual illusions classified by category.
  http://lite.bu.edu/
- *Michael Bach: Optical Illusions & Visual Phenomena*: Another excellent collection of illusions. Many QuickTime movies.
  http://www.michaelbach.de/ot/
- *The Joy of Visual Perception*: a "web book" by Peter K. Kaiser, York University.
  http://www.yorku.ca/eye/