VISION

Bart M. ter Haar Romeny
Faculty of Biomedical Engineering, Eindhoven Technical University, the Netherlands

Keywords: visual pathway, visual system, geometry, receptive field, vision

Contents

Summary

Most of the visual system is located in the visual cortex in the back of our head. This is one of the best-studied brain areas. Much is known about the wiring diagram of the first layers, the cell types, and the feedback loops. This chapter describes the 'wetware' structure of the first stages of the visual system, and attempts to model the mathematical operations carried out there. This bio-mimicking may give useful input for robust image analysis and computer vision routines. The first stages are found to measure a multi-resolution set of images at the retinal level, and to take derivatives of the image at many orientations at the cortical level. All different physical parameters of the visual scene seem to have their own 'channel' with dedicated receptive fields, such as for motion, depth, shape and color. The chapter gives an overview of the current insights into the neurophysiology of the visual system, and describes some computer vision models for feature detection, adaptive control and for multi-scale imageanalysis.

1. Introduction

Vision is our most important sense. The visual system is a signal processing system with spectacular performance. First of all, the mere quantity of information processed in the visual system is enormous. But most importantly, the system is capable to extract *structural* information from the visual picture with astonishing capabilities and with real-time speed. The visual system is one of the best-studied systems in the brain. We

are, however, still a very long way from understanding its most intricate details. In this chapter we focus on the first stages of vision, where data from neuro-physiology, anatomy and computational models give rise to a reasonable insight in the processing that happens here.

After the measurement through the eye, the *looking*, our *seeing* is done in the visual cortex, a number of specialized cortical areas in the back of our brain. It has been estimated that 25% of all our estimated 10^{10} brain cells are in some way involved with the visual process.

A study of the visual system can be done from many different viewpoints:

- Eye physics: the study of the physical limitations due to the optics and retinal structure.
- Psychophysics: how well does the visual system perform? What are the limits of perception? Can it be tricked?
- Neurophysiology and -anatomy: a study of the system's wetware organization, e.g. by measurement of the electrical activity of single or small groups of cells, following neural pathways and connections.
- Functional imaging: measure the functional activity of arrays or large clusters of cells.
- Computational models: the field of computer vision mimicking the neural substrate to understand and predict its behavior, and to inspire artificial vision routines. Despite many efforts and high mathematical sophistication, today we still see a very limited performance of computer vision algorithms in general.

Traditionally, vision is coarsely divided into three levels: front-end, intermediate and high level vision. The visual front-end can be defined as the measurement and first geometric analysis stage, where associative memory and recognition do not yet play a role. The outputs of the front-end go to all further stages, the intermediate and high levels. In the front-end the first processing is done for shape, motion, disparity and color analysis (in more or less separate parallel channels). The intermediate level is concerned with perceptual grouping, more complex shape, depth and motion analysis and first associations with stored information. The high level stages are concerned with cognition, recognition and conscious perception. This chapter focuses on front-end vision, as this is by far the best understood, and its principles may guide progress in the research of higher levels. High-level vision, where the cognitive processes take place, is a huge research area, and the most difficult one. It is the domain of many scientific disciplines, the cognitive sciences.

The visual system turns out to be extremely well organized. The retinal grid of receptors maps perfectly to the next layers in the brain in a retinotopic fashion: neighborhood relations are preserved. Two cells, next to each other in the retina, map to cells next to each other in higher stages, i.e. the lateral geniculate nucleus and the primary visual cortex. We recognize a cascade of steps, many (or all?) of the stages equipped with extensive feedback to earlier stages. The mapping is not one-to-one: there is neighborhood convergence (many nearby cells to one) and neighborhood divergence (one cell to many nearby).

Many models can be proposed for the assumed working of the visual front-end in its visual processing, and extensive literature can be found. A successful recent realization is that the front-end can be regarded as a *multiscale geometry engine*. What does this mean? Multiscale, or multiresolution, is the property of the visual system to measure and process the information at many simultaneous levels of resolution. It is a direct consequence of the physics paradigm that the world consists of objects and structure at many sizes, and they should be measured with apertures at these sizes, i.e. at these different resolutions.

Our retina is not sending down the measurements of single rods and cones, but of *groups* of rods or cones. Such a (typically circular) group is called a *receptive field*. They come in a large range of sizes (minutes of arc to many degrees), and measure the world consequently from sharp to very blurred. For very fine details we employ the smallest receptive fields, for the larger structures we employ the larger receptive fields. Moreover: it is an extra dimension of measurement: we sample not only the spatial and temporal axes, but also along the scale axis.

The notion of geometry engine is reflected in the important contemporary model that the front-end visual system may extract *derivatives* in space and time to high order from the visual input pattern on the retina.

It turns out that in the visual front-end we have separate parallel channels for shape, motion, color and disparity processing. In our visual system we generate an extremely redundant set of measurements: for every possible value of a parameter we seem to have a specialized receptive field, e.g. for every velocity, for every direction, for all sizes, necessary differential order and all orientations. We will study this framework in greater detail in the next sections, while following the *visual pathway*, i.e. the neuronal path of the information from the retina into the brain.

2. The eye

The eyes are the two moveable stereo cameras to measure the light distribution reflected and emitted from the objects in the world around us. The image is formed upside down on the retina, which is the layer of light sensitive receptors in the back of our eye. The breaking power of the eye optics (60 diopters; A diopter is defined as the measure for the breaking power of a lens. It is equal to the inverse of the focal distance of the lens, $1/f$, and is expressed in m^{-1}) is due to both the cornea (43 diopters) and the lens (17 diopters), and can be varied over 8 diopters due to accommodation of the lens.

The eye has a diameter of about 17 mm. The processing of information starts already in the retina, as this really is extended brain tissue: similar neurotransmitters are found in the retina as in the cortical brain tissues, and we recognize the same strictly layered structure in the retina as we will meet later in the visual cortex.

3. The retina

The retina consists coarsely of three layers of cells (figure 1). The light sensitive receptors, i.e. the rods and the cones, are located in the back of the eye behind the other

layers. The reason for this is the close neighborhood to the nursing vessel bed in the back of the eye, the pigmented cells or chromatoid. The rhodopsine molecules in the receptors that are bleached by the light can in this way be easily replenished. The middle layer contains horizontal, bipolar and amacrine cells (figure 1). The front layer contains the about one million ganglion cells, whose axons form together the optic nerve, the output of the retina.

The rods and cones are packed tightly together in a more or less hexagonal array. We have much more rods then cones. It is estimated that we have about 110,000,000 to 125,000,000 rods in our retina, and about 6,400,000 cones. The diameter of a rod in the periphery is about 2.5 μm, or about 0.5 minute of visual angle. In the fovea, the central area of the retina, the receptors become smaller, about half this size (figure 2).
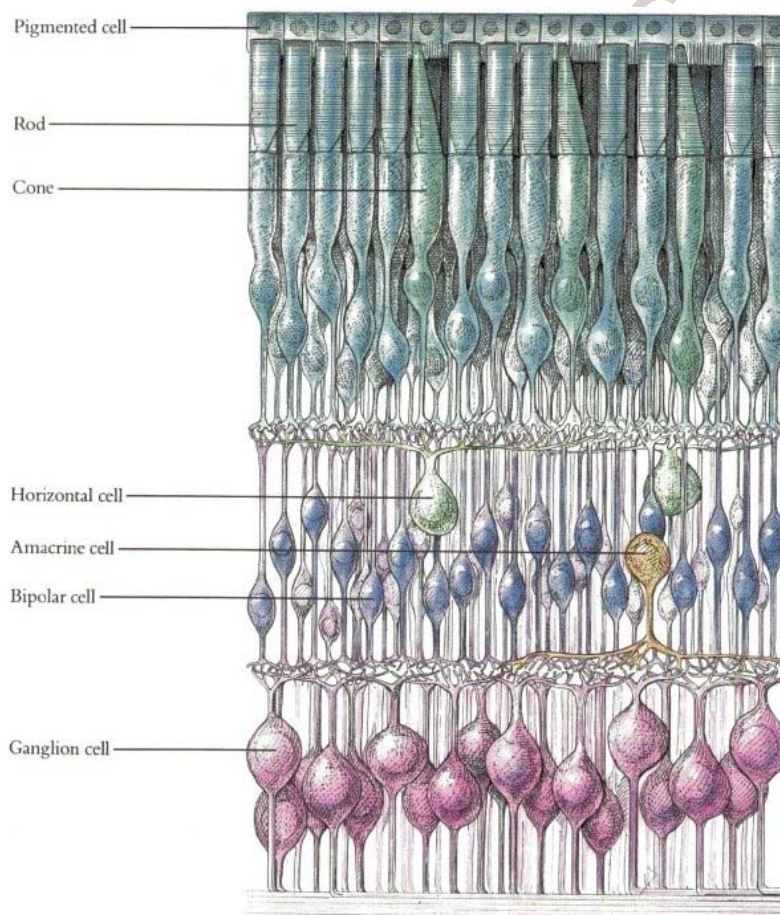

Figure 1: The cell layers of the retina. Light is coming from below. The receptors touch the chromatoid layer, from which the bleached rhodopsine is replenished. The bipolar cells connect the receptors with the ganglion cells. The horizontal cells enable lateral interaction. The function of the many types of amacrine cells is unclear. They may have a role in motion processing. The ganglion cells at the bottom form the output and are the only cells that generate action potentials. From [Hubel 1988].
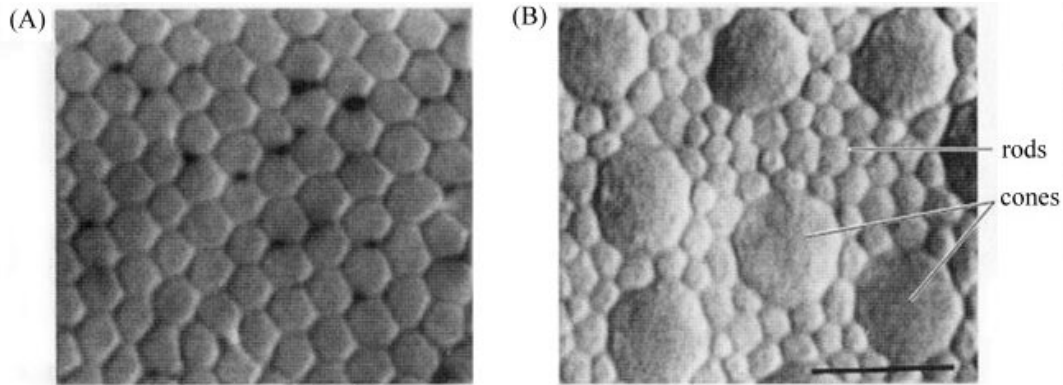
Figure 2: Hexagonal packing of the rods in the fovea (left) and rods/cones in the periphery (right) in the human retina. Scale bar = 10 μm.

Rods are only used at very dim light levels. This is rod vision, or *scotopic vision*. In normal daylight lighting conditions, they are in a completely saturated state, and have no role in perception. Rods are single color, so in the dim light we see no colors. The optimal sensitivity is in the green-yellow. Cones are used at normal light levels, i.e. *photopic vision*. Cones come in three types, for long (red), medium (green) and short (blue) wavelength sensitivity. Therefore these types are called L, M and S cones.
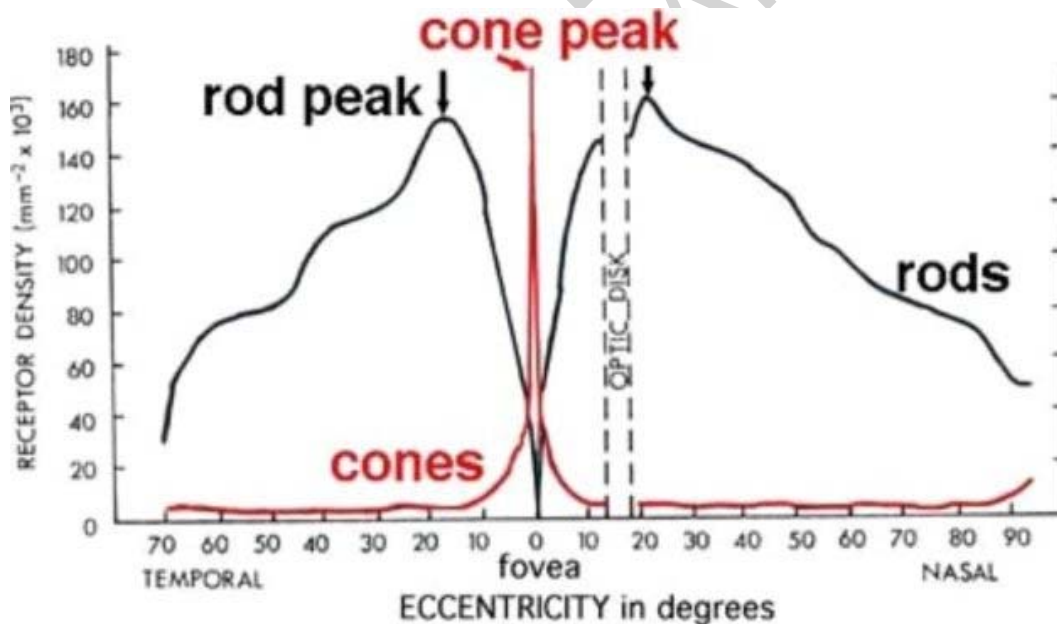

Figure 3: Density of retinal rods and cones as a function of eccentricity. The central area of the fovea is rod-free

The receptors are not evenly distributed over the retina: in the fovea (about 1.5 mm or 5.2 degrees in diameter; one degree of visual angle is equal to 288 μm on the retina), we find an area free of rods of a roughly 250 - 750 μm. The number of cones in the fovea is approximately 200,000 at a density of 17,500 cones/degree^2. The rod free area is about

1°, thus there are about 17,500 cones in the central rod-free fovea. The density distribution of rods and cones as a function of eccentricity is given in figure 3.
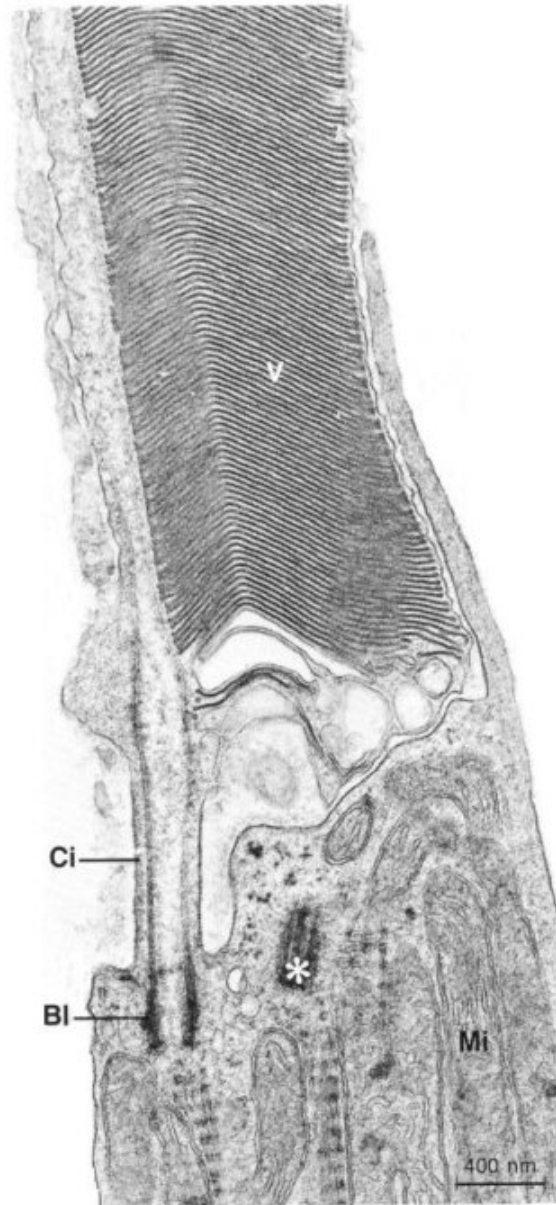

Figure 4: Electron-microscopic cross-section of a rod. In the upper section are the disks with the membrane vesicles (V), which contain the visual pigment, i.e. the rhodopsine molecules. Ci = cilium, the small connecting tube between the top (outer segment) and bottom part (inner segment) of the cell. Mi = mitochondrion. Bl = basal body. Scale bar: 400 μm.

Figure 4 shows an electron-microscopic cross-section of a single rod. On top the hundreds of disks can be seen, which contain the light sensitive rhodopsine molecules. Even a single photon is capable to evoke a measurable chemical reaction. We need

about 3-4 photons to see them consciously. The sensitivity range of the retina is impressive: the ratio between the dimmest and the brightest light we can see (without damage) is 10^{14} !

The caught photon changes the structure of the rhodopsine molecule, after which a whole chain of events generates a very small voltage change at the base of the cell, where it is transmitted to the next layer of cells, to the horizontal and bipolar cells. A rod or cone does *not* generate an action potential, just a small so-called receptor-potential of a few millivolts. Much research has been done on this process; it is beyond the scope of this chapter to go into more detail here.

-
-
-

TO ACCESS ALL THE **28 PAGES** OF THIS CHAPTER,
Visit: http://www.eolss.net/Eolss-sampleAllChapter.aspx

Bibliography

Excellent and all easy-to-read books for an introduction to the physiology of human vision are the following:

Hubel, D. (1988). *Eye, Brain and Vision*, 240 pp. New York: Scientific American Library [written by the Nobel laureate as an overview of his total oevre]

Rodieck, R. W. (1998). *The First Steps in Seeing*, 562 pp. Sunderland MA, USA: Sinauer Associates Inc. [handbook on retinal physiology and modeling]

Zeki, S. (1993). *A Vision of the Brain*, 366 pp. Oxford UK: Blackwell Scientific Publications [full with historical details of the development of vision science]

Kandel, E. R., Schwartz, J. H., Jessell, T. H. (2000). *Principles of Neuroscience, 4rd ed.*, 1137 pp. London: Prentice Hall International Inc. [the classical textbook for medical students by Nobel laureate Kandel]

Wandell, B. A. (1995). *Foundations of Vision*, 476 pp. Sunderland MA, USA: Sinauer Associates Inc. [engineering approach to vision science]

Reference books:

Kessel, R. G., Kardon, R. H. (1979). *Tissues and Organs, a Text-Atlas of Scanning Electron Microscopy*, 319 pp. San Francisco: W.H. Freeman and Company [one of the best photobooks available on biological tissues and organs]

Website: *Visionary*, a dictionary for Vision: http://cns-web.bu.edu/pub/laliden/WWW/Visionary/Visionary.html

Website: *Webvision*, an overview of the structure and function of the vertebrate retina by Kolb, Fernandez and Nelson: http://webvision.med.utah.edu/

Good tutorial books for the multiresolution modeling and theory of vision and the applications to medical imaging:

Ter Haar Romeny, B. M. (2001). *Front-End Vision and Multiscale Image Analysis*. Dordrecht, the Netherlands: Kluwer Academic Publishers [a tutorial book in Mathematica 4 with complete sourcecode to

do image processing experiments].

Lindeberg, T. (1994). *Scale-Space Theory in Computer Vision*. Dordrecht, the Netherlands: Kluwer Academic Publishers [classical primer for the field of computer modeling of multiscale image processing]

Ter Haar Romeny, B. M. (1994). *Geometry-Driven Diffusion in Computer Vision*, 439 pp. Dordrecht, the Netherlands: Kluwer Academic Publishers [a summary of nonlinear filter theories in computer vision].

Florack, L. M. J. (1997). *Image Structure*, 269 pp. Dordrecht, the Netherlands: Kluwer Academic Publishers [a textbook for mathematicians on multiscale image analysis]

Biographical Sketch

Bart M. ter Haar Romeny is full professor in Biomedical Image Analysis at the Department of Biomedical Engineering at Eindhoven University of Technology. He received a M.Sc. in Applied Physics from Delft University of Technology in 1978, and a Ph.D. on neuromuscular nonlinearities from Utrecht University in 1983. After being the principal physicist of the Utrecht University Hospital Radiology Department he joined in 1989 the Image Sciences Institute at Utrecht University as associate professor. His research interests are mathematical aspects of visual perception, in particular to apply 'biomimicking' in (bio-)medical imaging. His monograph "Front-End Vision and Multi-Scale Image Analysis" is a popular textbook. He initiated a number of international collaborations on those subjects. He is an active teacher in (inter-)national courses, a senior member of IEEE.. He has been chairman of the Dutch Society for Clinical Physics, and the Dutch Biophysical Society.