Fundamental Properties of Medical Image Perception

TR91-008
February, 1991

Stephen M. Pizer
Bart M. ter Haar Romeny

Medical Image Display Group
Department of Computer Science
The University of North Carolina
Chapel Hill, NC 27599-3175

The research reported herein was carried out with the partial support of NIH grant number P01 CA47982. Published in Journal of Digital Imaging, February, 1991. UNC is an Equal Opportunity/Affirmative Action Institution.

Fundamental Properties of Medical Image Perception

Stephen M. Pizer*, Bart M. ter Haar Romeny+

*Medical Image Display Research Group
University of North Carolina, Chapel Hill, NC, USA

+Department of Radiology, University Hospital, 3D Computer Vision Group
Utrecht, The Netherlands

Abstract

With a mind toward the effective acquisition, processing, presentation, and reading of radiological images, we present a survey of how the human visual system perceives images. The level is chosen to be suitable for the radiologist, and the relative emphasis on the various visual cues of luminance, color, form, texture, motion, and depth is chosen based on their importance with radiological images. Examples of the radiological relevance of the various visual properties are given. We cover first what the visual system's behavior is and then survey some of the properties of the physiological mechanisms that provide this behavior.

Part I. Visual System Behavior -- What Happens

1. Introduction

The presentation of medical images, whether in 2D and 3D, requires a match between the displayed image and the perceptual capabilities of the human viewer. In 2D at every point in an image a display scale carries, in intensity and/or color, information about the physical parameter, such as x-ray attenuation (for radiography or CT), acoustic impedance change (for ultrasound), or relaxation times (for MRI), that is recorded by the imaging device. The intensity or color changes must allow the detection, localization, and characterization of anatomy or function that is inherent in the imaged data. In 3D display the distribution of the recorded intensity is presented by either intensity levels or surfaces made to appear in 3-space. These intensities or surfaces again need to communicate effectively the corresponding properties in the imaged data. As a result a knowledge of human visual perception summarized in this paper can be helpful to the radiologist. We will largely be referring to the sort of vision that gives information about *what* is in the image rather than metrical information as to *where*, which appears to come from a separate visual subsystem.

In both 2D and 3D the visual system appears to measure a variety of features, more or less independently, and then combine the information from these features to produce an overall percept. As a result we will focus on the properties of the perception of various features: luminance, form, color, motion, flicker, texture, and various three-dimensional cues. A common thread will be the visual system's sensitivity to discontinuities: to edges, corners, and bars in luminance, sharp changes in time, and cliffs in depth. To explain some of the behavior, references to human visual anatomy and neurophysiology will be made from time to time. A summary of this biological structure and operation appears in Part II.

This paper treats the perception of image regions chosen for examination rather than the search process by which such regions are chosen. Thus, we focus more on *perception*, which emphasizes a sort of bottom-up treatment of the visual input, than on *cognition*, which emphasizes a more top-down treatment. A good treatment of the search and cognitive issues in regard to medical imaging can be found in [Kundel, 1987; Nodine, 1990].

2. Perception of Static 2D Images

2.1 Perception of Luminance

Human visual perception is characterized by great adaptability but an inability to make absolute measurements. In particular, we can perceive objects over a wide range of spatial sizes, and we can distinguish intensities over a wide range of luminance. Our vision succeeds very well in perceiving structures except at very high and very low intensities, at very large and very small spatial scale (see figure 5), and at very high and very low rates of temporal change -- elsewhere relative changes are reported.

Figure 1 presents a simple model of human vision, largely ignoring perception of temporal variations. Like most of the explanatory models in this paper, it does not reflect a unanimous view of the vision community, but many of its

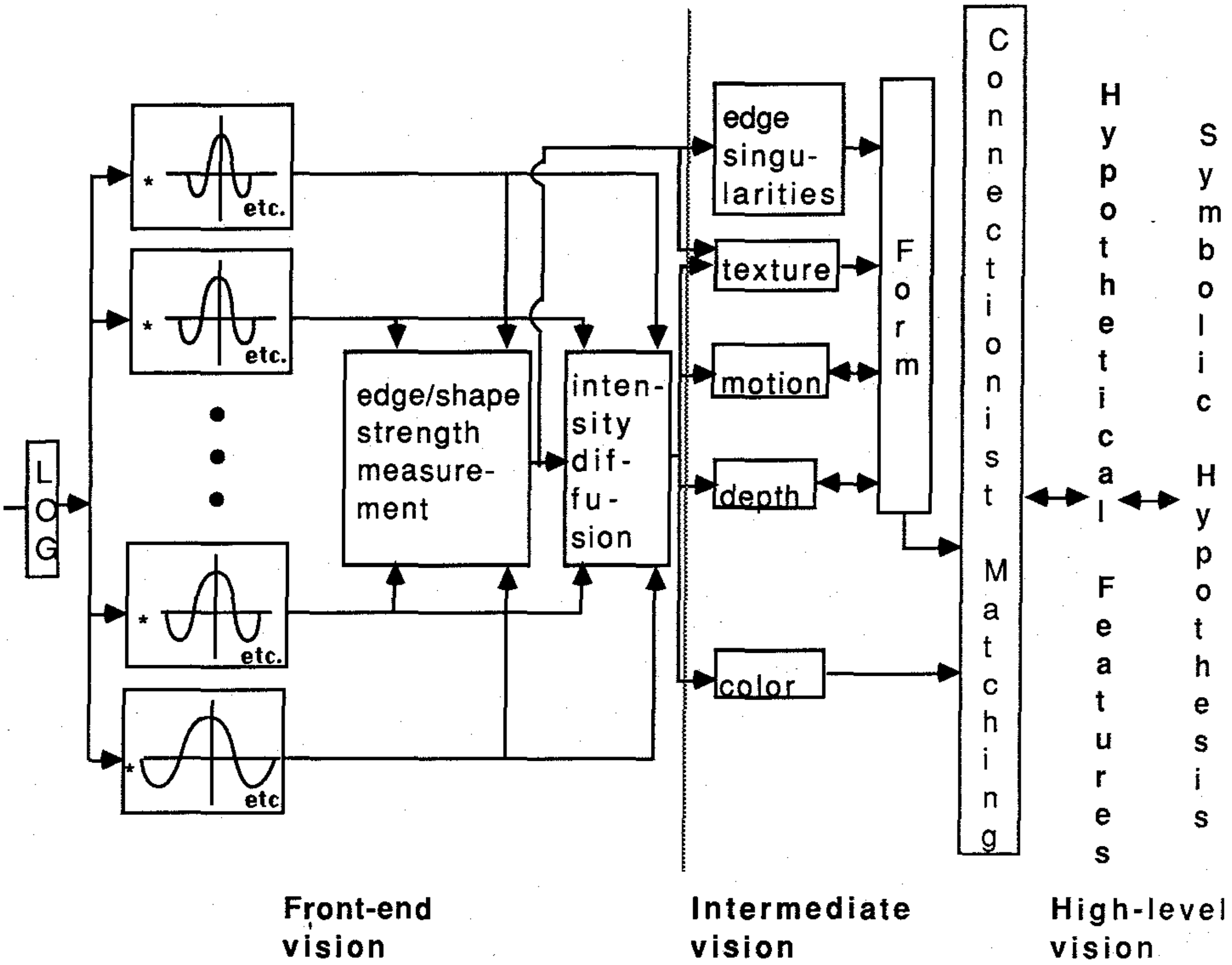


Figure 1. A model of static human vision.
The input, left, is the spatial distribution of luminances sensed by the receptors in the retina with a logarithmic sensitivity. Groups of receptors, in circularly overlapping fields, form receptive fields (RF's) of many sizes, indicated by the first column of boxes. The outside world is sampled via weighted sums of the inputs around each image point (a process indicated by "*"); the weights are indicated by the graphs in the boxes.

To the left of the dotted vertical line is the early visual system, in which all arrows carry images. Each box to the left of that dotted vertical line represents an operation carried out at every image location. The RF's measure edge strength (and other shape features such as curvature) by producing image intensity differences in the input for a range of orientations. Because of the various diameters of RF (indicated by the column of boxes), overlapping at any image point, we see at many 'scales' simultaneously, or with 'multiresolution'.

The features in the columns to the right of the dotted line produce measurements with reference to the locations at which the measurements apply. The rightmost column represents descriptive states rather than analog measurements.

properties are instructive. The stages of this model preceding the dotted line are referred to as 'front-end vision'.

The initial logarithm results in measuring *relative* intensities, and the next column of RF samplings results in measurements giving *relative* information as to sizes. The effect is that pattern properties are judged with details taken relative to the size and intensity of the objects of which they are a part [van der Wildt, 1983]. Often it is not necessary to perceive all fine detail, but the coarser structure suffices, e.g., to see whether an object has moved or to determine the general class of an object. As one moves toward the periphery of the visual field from the central region called the 'fovea', sensitivity to detail decreases because the visual processors for smaller spatial scales successively drop out.

The column of RF samplings produces first or higher order differences relative to nearby locations, so they give strong results at edges, bars, and corners, and in general places where intensity changes or curves sharply either in the intensity dimension or in the spatial dimensions. The output is rich in geometric information, and the visual system is able to do geometric calculations on it. Put differently, if we see the two-dimensional intensity distribution as a hilly landscape, where height corresponds to intensity (see figure 2), the visual system can calculate where steep slopes are, or highly curved height lines, or ridges, cloves, peaks, pits, etc. This corresponds directly to visual features: edges, bars, etc. In fact, the early vision system does not detect features but gives output relative to "featureness", like "edgeness", "cornerness", etc. Additional geometric information is given by comparison of the outputs of two or more neighboring RF's (which are connected by an interneuron giving a delay of transmission between the cells, see section 3 and part II). By this means the visual system can calculate motion via 'optic flow', i.e. for each point how the geometrical features move, or accelerate, relative to the observer and each other. Important features are also the 'catastrophes' when one object moves behind another.

Moreover, each RF box in the diagram and various components of the motion box produce information about changes at a different spatial scale, namely the scale corresponding to the width of the function in them (for a description of the anatomy of the functional units in this layer, and the description of 'scale' see Part II). Thus information about absolute intensity is removed, leaving only information on changes in intensity that occur at the respective spatial scale.

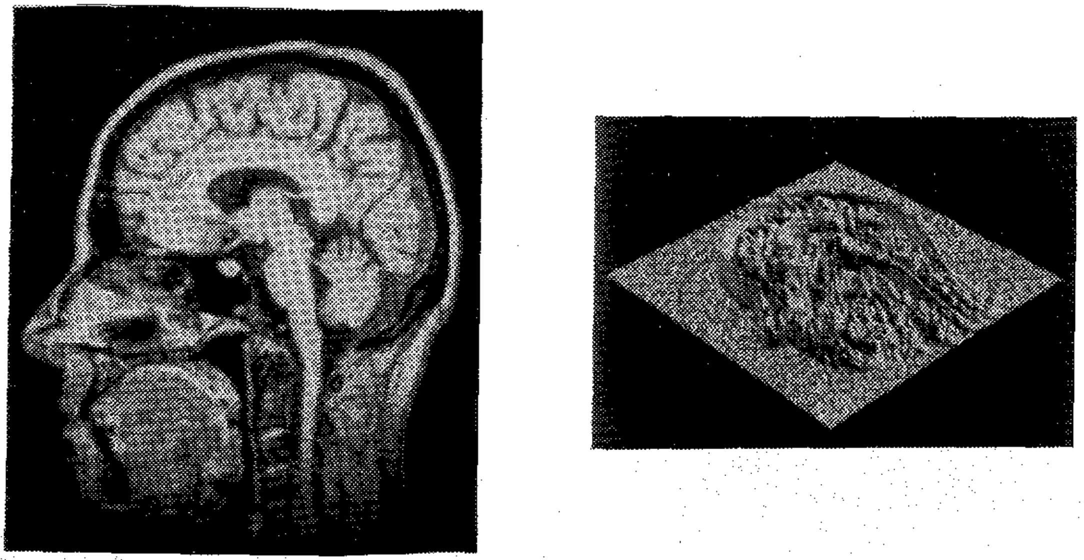


Figure 2. An MRI image and its representation as an intensity surface (landscape).

In addition, each of these change measurement operations is carried out independently by cells that apply the change measurement across a different orientation. Orientation is thus an essential feature of the early visual output. Moreover, the higher order differences give information about curvature as well as orientation.

By a process of co-operation and competition among the local oriented shape feature measurements edge or bar strengths and curvatures are calculated in a way that reflects the continuation of one oriented edge or bar segment into another. The edges produced include not only edges corresponding to sharp luminance changes in the image but also so-called "subjective edges" which are derived from more distant luminance changes or bar ends by a continuation or connection process (see figure 3). The resulting edges, real or subjective, are closed, i.e., they surround regions. The intensity changes initially detected at these edges are then spatially smeared by diffusion, with the perceived contrast of these edges used as an insulation strength that partially blocks the diffusion [Grossberg, 1985]. As a result, intensities perceived within a closed edge are derived from the changes measured at the edge and poorly reflect the actual intensities inside the object.

While these properties of the visual system give it impressive sensitivity in perceiving objects of all sizes and contrasts, they make it a completely untrustworthy measurer of luminance, so conclusions as to the near equality of two separated image locations or even which of the two has greater luminance must be given little credence. The perceived image reflects only relative information about the luminance and the spatial scale and is normally inaccurate about even relative luminance due to the strong dependence on edges and the contrast there. The great effect of edges results in a large collection of "optical illusions" near edges, including 1) the Mach effect (the apparent brightening and darkening on the respective sides of the edge), 2) the appearance of subjective edges completely unsupported by local contrast (see figure 3), and 3) the difficulty in seeing edges well supported by local contrast when a sort of Mach shadow is computationally added to an edge as a result of a method designed to enhance contrast (see figure 4).

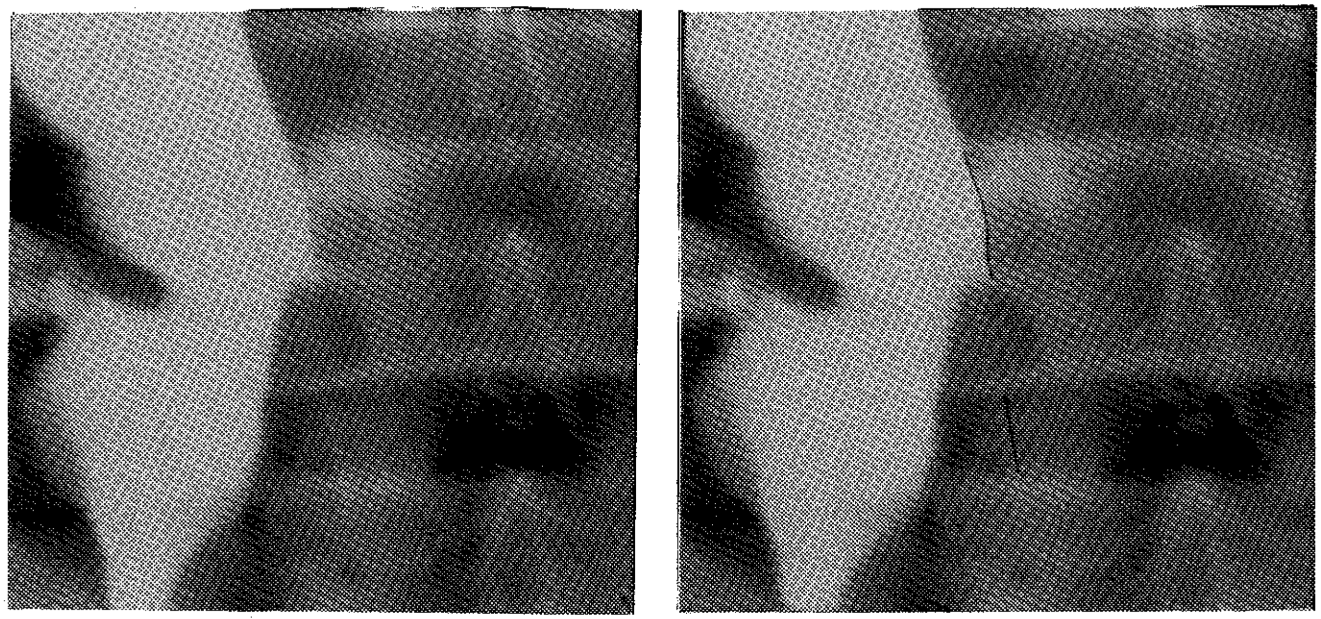


Figure 3: The edges marked by the solid and broken curves in the section of the radiograph on the right are all perceived by the visual system in the image on the left. The one marked by the broken curve is a subjective edge, i.e. it does not exist in the image's intensity variations.

The model properties discussed above emphasize that contrast sensitivity depends strongly on the structure, contrast, and scale of the objects expected or perceived, but very little on the absolute luminance or spatial size except at extremely large or small luminances or sizes. Although human visual abilities are frequently characterized in terms of contrast sensitivity as a function of the spatial frequency of a sinusoidal pattern (see figure 5), this information alone cannot be used to predict behavior when viewing a complex scene. On the contrary, when we speak of just noticeable differences in luminance, a difference that can be seen a fixed fraction of the time (commonly 50 or 75%) when acting at a confidence such that a false call of difference is made at some specified small rate (e.g., 5%), we must realize the heavy dependence on the structure of the target and background. The discrimination capability also depends on the brightness of the background. At low luminances (less than 10 trolands, where the troland is a measure of retinal illumination) a just noticeable difference in

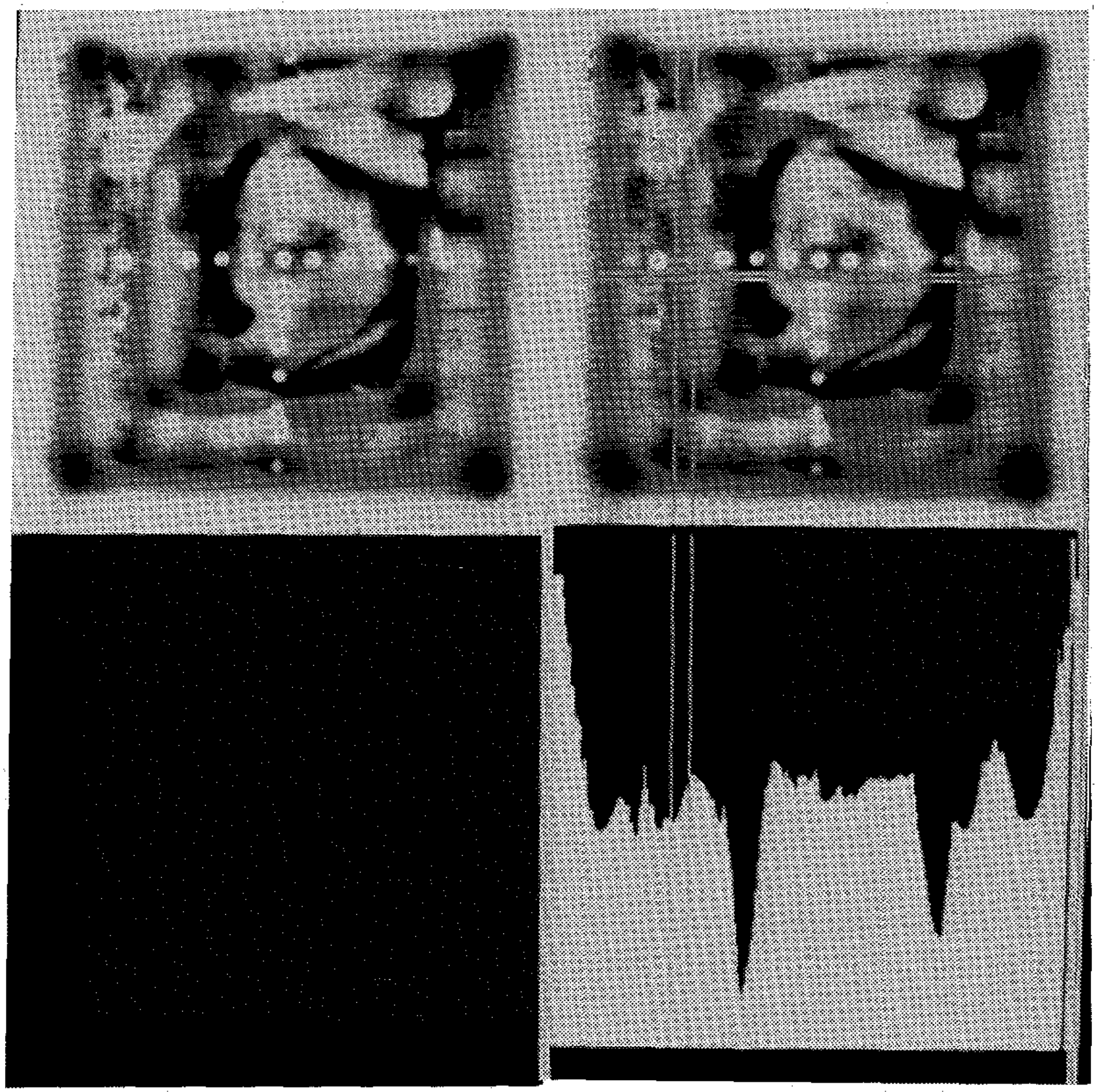


Figure 4. The edge of a radiotherapy treatment field, apparently sharp in the original (not shown) appears smeared (see upper left) when processed by the Adaptive Histogram Equalization (AHE) edge enhancement technique [Pizer, 1986]. A horizontal intensity profile, given at the lower right for the level indicated by the horizontal double line in the upper right image, shows a strong edge in the displayed luminances between the two vertical lines in the right image despite a lack of a perceived edge in the that region.

intensity (see figure 6) is proportional to the square root of the reference intensity (the deVries-Rose law), and at moderate luminances (over 500 trolands) a just noticeable difference is proportional to the reference intensity (the Weber law). However, the constant of proportionality varies strongly with the spatial structure of the pattern in which the contrast is being perceived, with the shape and size of a target, intensities at its edge, and the location of and contrast at nearby edges surrounding the target having the most effect on the perception of the target. It should be realized that while light boxes and CRTs over most of their range follow the Weber law, at their bottom levels one is shading into the deVries-Rose region.

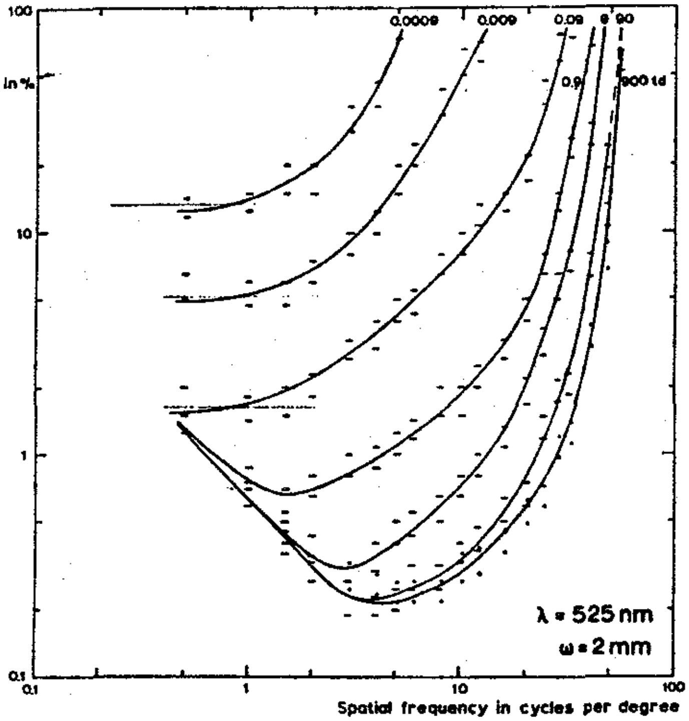


Figure 5. Threshold-modulation curves for green light = 525nm, at seven retinal illuminances (0.0009 - 900 trolands) and a pupil diameter of 2mm. From [van Nes, 1967].

Moreover, while the visual system is capable of adapting over an impressively wide range of luminance (more than 10 orders of magnitude), for any given level of adaptation it can make distinctions over only a relatively narrow range. Even the relative changes in luminance are underestimated at intensities far from that to which the system is presently adapted. Thus, when the viewer is adapted to the middle part of the displayed intensities, sensitivity is lessened at both ends of the scale of luminance. The result is that there is considerable difficulty in appreciating a change in a very dark area of a film when the film is smaller than the illuminated lightbox area and the shutters on the lightbox are not properly set or not available, so the bright borders lead to a strong upward shift of the perceived intensity range. A similar but less pronounced effect affect performance when bright lights are illuminating the environment and thus producing a veiling effect [Gilchrist, 1983]. From these findings we see that the best lighting conditions in a reading room should be not too dark or too bright, but ideally be the same average intensity as the images to be viewed. Many reading room designs lack the possibility of ambient light adjustment.

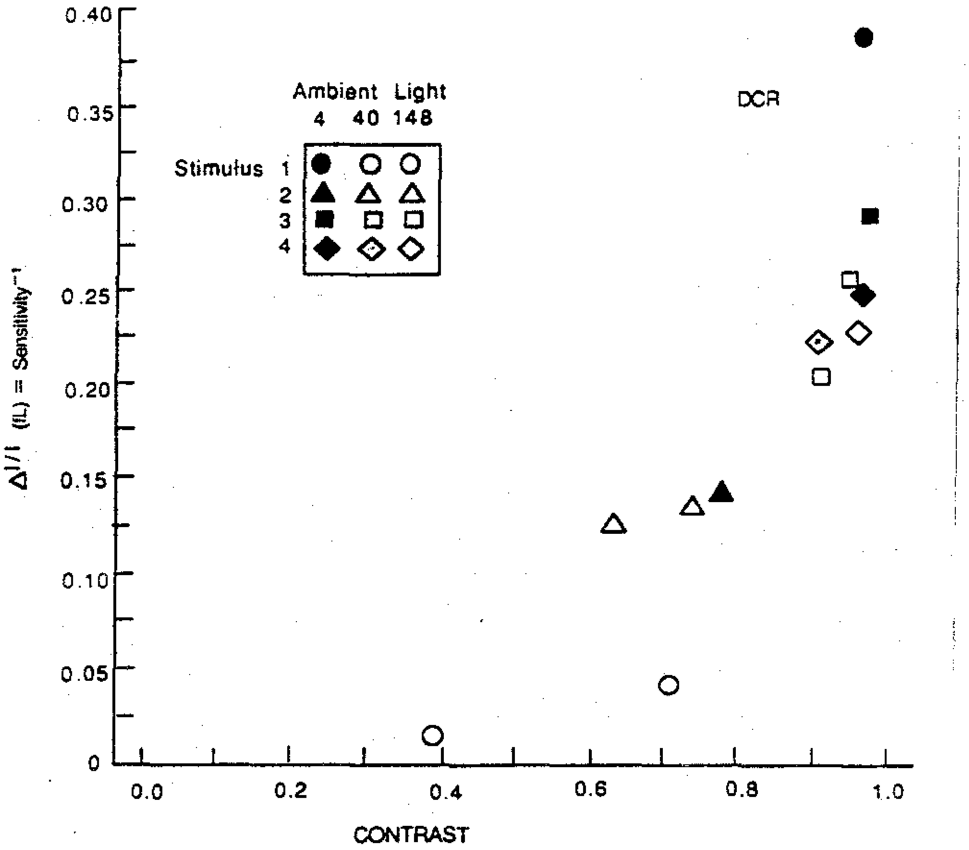


Figure 6. Just noticeable differences in luminance, as a function of background luminance, for a pattern made up of two separated squares on a uniform background [Rogers, 1987].

Noise in the image data changes not only local intensities, but the perception of edges and thus the perceived intensities. Nevertheless, there is good evidence that at low levels of noise the human observer act as a relatively efficient detector compared to the so-called 'ideal observer'. This ideal observer is defined to have an optimal decision signal-to-noise ratio, as long as the noise is uncorrelated from pixel to pixel and the background is relatively uncomplicated

[Burgess, 1988]. Noise correlations, such as those that occur as a result of tomographic reconstruction or image restoration, cannot easily be "undone" by the viewer. The resulting noise "blobs" can easily be misread as structure in the scene.

2.2 Form

The primary result of seeing is not either intensities or edges but objects: regions that carry a spatial structure that is called form. The best evidence is that the edge or bar orientation and curvature information discussed above as well as edge information derived from motion and depth properties discussed below are the relevant features of form [Treisman, 1986]. The edges and corners seen at different levels of spatial scale combine in some way with a hypothesis as to the object generated from expectation or tentative decision to produce the final percept of form.

It has been suggested that the cognitive system hypothesizes a form layout and derives from this hypothesis a set of edge and curvature features at many scales. These features based on expectation or tentative analysis are fed back towards the level at which the sensed feature values arrive, and possibly in addition towards even earlier portion of the visual system, where they can affect what measurements are made. The feature values fed back from the hypothesized image are matched against the features computed from the sensed image (see figure 7). If the match is high enough, the hypothesized form is accepted, and if not (or in parallel) other hypotheses are used as the basis of matching. The matching between hypothesized and sensed image features is probably not carried out by calculating a correlation per se, but by a scheme in which connections among the sensed and expected features mutually reinforce or inhibit to create a pattern of neuron firings corresponding to the acceptance of the hypothesized form that is the perceptual output. Recent theoretical developments in the field of neural networks have stimulated thinking in this direction.

It follows from this explanation that the hypothesis has a strong effect on what is seen -- we see more by accepting or rejecting hypotheses [Gregory, 1970] than by building a view from the visual input. Moreover, the basis for accepting or rejecting the hypothesis is not the full visual input but specific features as to e.g. edges, corners, and bars derived from that input. Radiologists are very familiar with the fact that clinical data can lead to looking for and finding a particular pattern, which would not have been read if no directed vision had taken place. It follows that considerable conservatism is needed with regard to reading faint, expected patterns. At the same time, reading mechanisms that force one to consider a range of expectations can be important to finding the correct diagnosis, since without such mechanisms a finding that jumps out after appropriate direction can be entirely unperceived in the absence of that direction. Without a strong hypothesis, it takes a rather strong corner, edge, or line to cause a weakly held hypothesis to come to the fore.

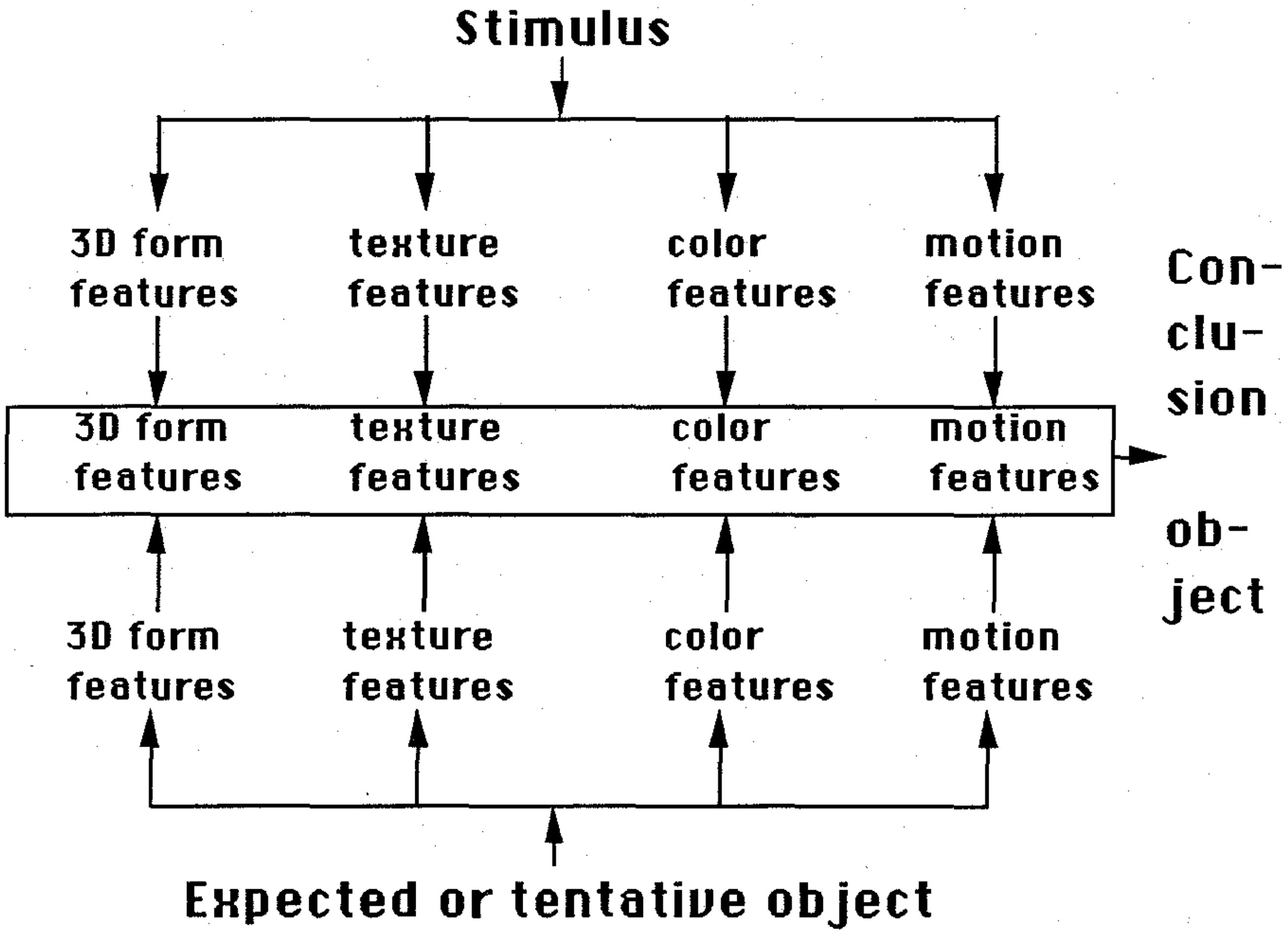


Figure 7. Model of matching of hypothesized and sensed features to produce resultant form.

2.3 Color

The sensory data for color perception comes from the primary receptors called the cones, but the trail leading from them to the percept of color is long and complicated. The three types of cone produce information on light intensity at long wavelengths (redness), medium wavelengths (greenness), and short wavelengths (blueness), but these are not perceived directly but rather are quickly combined into information on luminance and chromanence. Chromanence is given by two values, the position on a scale between red and green and the position on a scale between yellow and blue. Because of this division, we cannot imagine a reddish green or a yellowish blue, but we can imagine a reddish blue (purple) or a yellowish green (chartreuse). Many have attempted to use chromanence to label more than one image intensity value in a single image (e.g., T1 and T2 in a magnetic resonance image) but have failed, partially because their assignments have not recognized the natural color axes. The idea does not work very well even if the natural color axes are used because the human visual system does not separate, for example, purple into red and blue but sees it as reddish blue.

Having separated the sensed image into luminance and chromanence, the visual system deals with the luminance information at high spatial resolution, but it deals with the chromanence information at low spatial resolution. Moreover, just as with grey-scale intensities (luminance), colors are diffused between edges and thus the finally perceived color is derived from relative chromanences across edges rather than absolute chromanences [Land, 1971]. The result is that a color that is measured to have the wavelength combination normally called green can be perceived as red if it is in a certain environment.

The edges affecting the color perception are derived from the luminance information with little or no contribution from chromanence edges. Moreover, features of color patches are derived in a separate part of our visual system that is independent of the subsystem sensing brightness and that records image location only indirectly by reference to topographically represented visual distributions. The resulting color information is later "pasted on" to the forms derived from luminance and motion [Ramachandran, 1987]; as a result a color may appear well incorporated by edges when measurements show it to run across the edges.

We can conclude that color can be very effective to label objects in medical images (as in labeling flow direction in Doppler ultrasound) but that it is not useful for perceiving object boundaries, for defining objects by absolute colors, or even for comparing distant objects in terms of their color. As a result, pseudocolor display, i.e. representation of the intensity of the underlying physical image variable by not only a luminance but also a color, can produce unpredictable and even misleading effects when the chromanence variation between objects is significant compared to the luminance variation. Any use of color is subject to the caveat that a fair fraction of males are somewhat color-blind.

3. Images that Change in Time

3.1. Motion

The use of dynamic images in diagnostic radiology has long been useful for the study of actually moving objects or flows, such as in ultrasound, fluoroscopy, cardiac nuclear medicine, and DSA or MRI of angiographic flow. With the advent of digital techniques dynamic presentation can also be of interest in viewing derived (calculated) 3D objects, such as surface- or volume-rendered anatomy (see figure 12) or physiology from sliced CT, MRI, or ultrasound data, dynamic vector-electrocardiogram presentations, and so on. In the latter case the visualization is enhanced by either moving the object in front of the viewer or the viewer moving around the object.

The human visual system is designed to be used by a moving observer and to respond strongly to objects in motion. It does not simply determine motion from successive time-instances of already perceived form (though it can do that -- as indicated by the box marked "motion" in figure 1). Rather, the system carries out a direct detection of motion via cells that correlate intensity or edge orientation at one position with the same value at a different position at a later time. This means we have neurons forming a separate channel, that are only active when movement is detected. Thus the visual system diagram given in figure 1 shows only the static part of the story. After the original logarithm, receptors are found that sense changes of the image in time as well as the illustrated receptors that sense changes in space. As illustrated in figure 8, each of these motion receptors take input from two receptive fields; after the delay of one of these signals, the two signals are multiplied and then spatially and temporally integrated [Reichardt, 1961]. Thus when the movement of the feature over the bilocal detector is with a speed such that the duration to move from one RF to the other is equal to the delay time between the RF's (see figure 8), then we find an optimum in the output. These motion receptors act on only the luminance portion of the sensed input; color has no effect [Ramachandran, 1977].

As illustrated in figure 9, the subsystem for direct detection of motion consists of many sets of Reichardt detectors. Each set has its own temporal scale (delay) and spatial scale (offset magnitude) and is made of detectors for a range of orientations (offset direction).

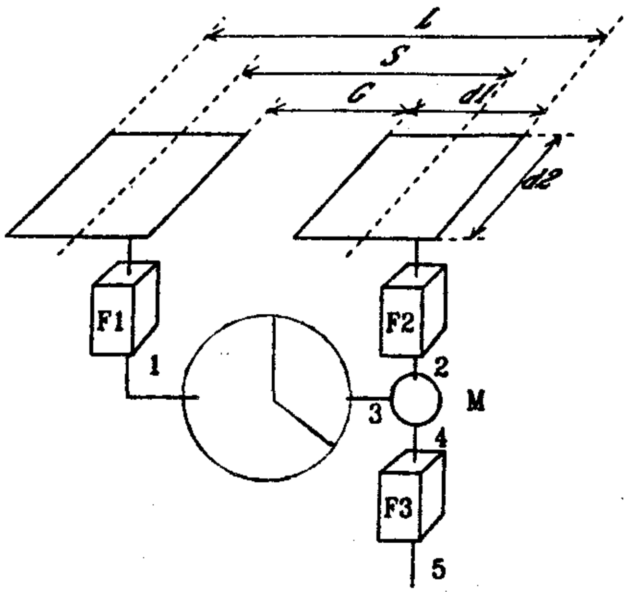


Figure 8. Reichardt motion receptors (compliments of W.A. van de Grind [1986]).

The existence of a separate subsystem for direct detection of motion means that motion is a fundamental visual dimension. Besides providing a percept of velocity, it provides information on depth, helps us to segment the sensed image into objects, and allows us to locate ourselves and our environment relative to each other (for an excellent review see [Nakayama, 1985]). In particular, motion gives

-many 3D cues, not only binocular, but also monocular. The so called 'optic flow field', the direction that the separate points have in a moving image, gives important information as to the orientation of surfaces [Koenderink, 1976, 1986]. In radiology this effect can be appreciated when we compare the display of a rotating 3D reconstruction (quick sequential presentation of reconstructions from many different angles) with a stationary view. Depth perception is substantially increased. In addition to the optic flow field, the time to collision cue provides information regarding the relative distances to points in the environment.

-image segmentation. At the borders of structures moving in front of other structures, sharp changes are found in velocity and directions of movement. Occlusion, i.e., the hiding by one opaque object of another behind it, is one of our strongest depth cues. In the primate receptive fields (RF) are found that are insensitive to uniform motion over the 'center' and 'surround' parts of the RF, but very sensitive to velocity differences between center and surround. Motion therefore helps seeing low spatial frequency information. Very slowly varying contrast structures cannot be seen when stationary, but can be discerned when they are moving.

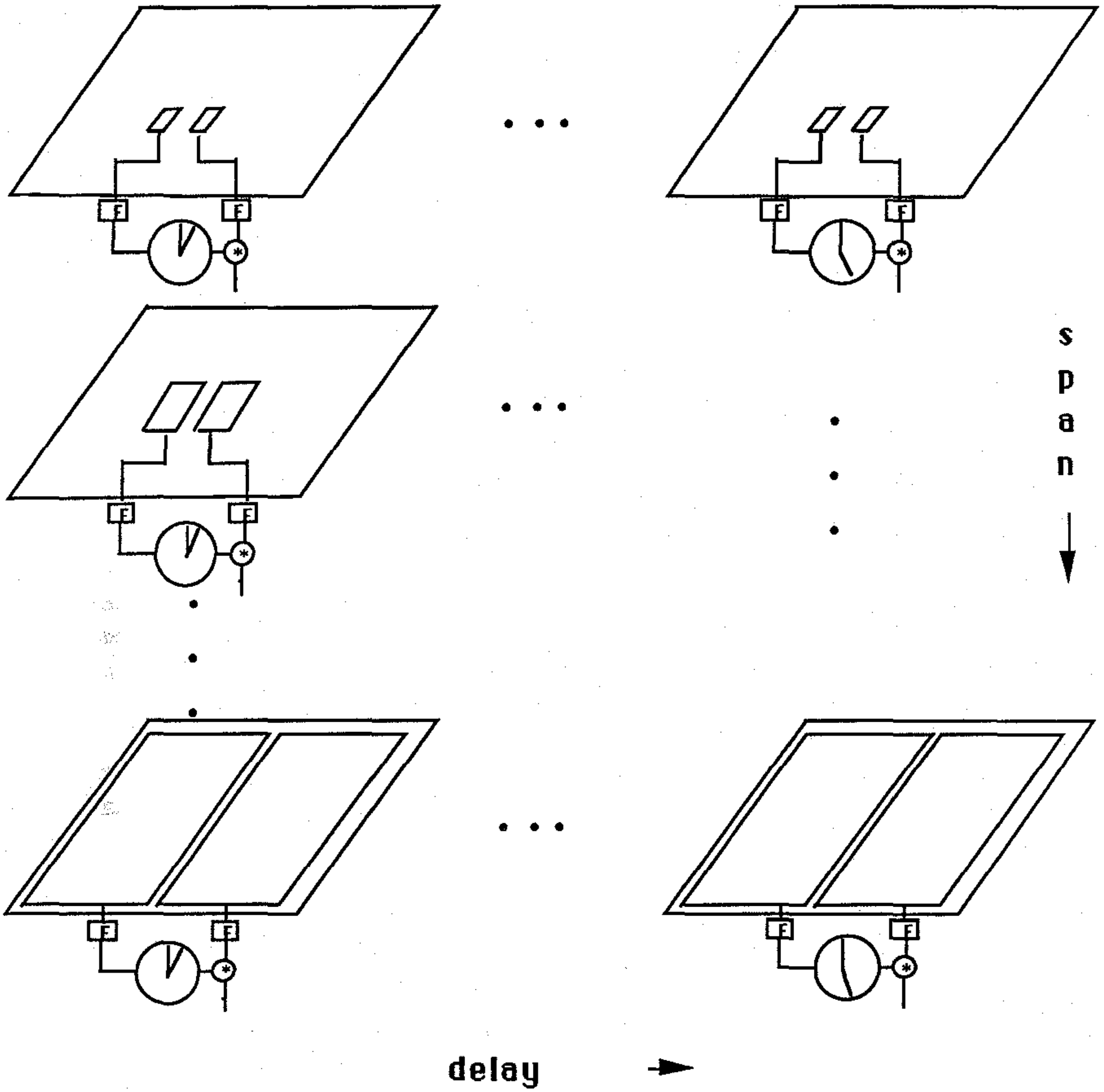


Figure 9. Motion subsystem
Motion is represented in a separate channel by separate neuronal circuitry. There are numerous Reichardt detectors (see Fig. 8) in the visual system for all velocities in all directions, all working simultaneously in parallel process. Each box corresponds to a collection of Reichardt detectors for the range of all possible orientations, but each box corresponds to its own combination of 1) spatial scale of the comparison region (indicated by the size of the box) and the offset between them; and 2) temporal scale and delay (indicated by the clock setting). Both scales vary exponentially.

-updates to our sense of 'what is where in space'. To accomplish this, visual motion information is combined very strongly with our own proprioception (muscular sense of the position of our body), first to locate ourselves, and second to provide the capability of 'active vision,' whereby we can explore the space around us. The subject of active vision will be elucidated in section 5: The 3rd Dimension.

Since motion itself can produce edges that generate form, motion can make an object much more detectable. Note that motion-generated edges also involve the spatial competitions and cooperations that generate subjective edges. The result is that the direct motion subsystem needs not only to be added into the system diagram given by figure 1, but its output needs to be tied into the form and motion feature systems farther to the right in figure 1.

Detection of structure in static images is more difficult than in dynamic studies. It is a well known fact that the frozen photographs from ultrasound studies are by far not as informative as the dynamic study on site, viewing the real time dynamic image. Therefore in many institutions the video signal is also stored on tape. Dynamic viewing also reduces independent noise substantially (see section 3.3: Flicker). Similar remarks can be made for fluoroscopy and heart-catheterization studies. There is a strong trend toward the development of dynamic visualization techniques in all digital modalities: MR dynamic angiography, fast heart studies, ultrafast CT. Not only are these techniques necessary to allow the appreciation of physiological movement and to remove patient motion artifacts, but they also lead to better appreciation of features and structure boundaries as outlined above.

An additional use of motion is suggested by the fact that dazzle-painting for camouflage can hardly or not be seen when it is static but is easily recognized when moving. Thus we might obtain increased sensitivity to changes between a pair of images by using an alternating presentation on a single screen instead of presenting them in adjacent locations. Modern image matching techniques could accommodate for patient movement.

The threshold for detection of motion is about 1 to 2 minutes of arc per second in the presence of a stationary reference, and 10 to 20 minutes of arc per second without. The threshold for peripheral movement is higher (up to a factor of 3) than the threshold for central vision; both increase with decreasing luminance. The precision with which we see differences in velocity is expressed in relative terms, given by the Weber law, that the just noticeable difference in speed is proportional to the the absolute speed. For speed discrimination the constant of proportionality appears to be 5%, except for very slow velocities below 1.5 degrees per second [McKee, 1981].

3.2 After-effects

Images just seen affect our sensitivity in images now to be seen. An exposure to a high luminance in any region of the visual field decreases our sensitivity to luminance in that region and thus causes a lower intensity region to look darker there. This behavior is thought to be related to the need for receptors of luminance to have a period of recovery before they can again 'fire'. The same property holds for other features involving the primary receptors. It is true for each of the colors corresponding to each of the three types of cones, and it is true for motion receptors. One of the oldest demonstrations of the motion after-effect is the 'waterfall illusion'. After observing a steady moving stream for a long time, stationary objects are seen moving in the opposite direction, with a surprising separation of position and motion. This effect again indicates that motion detection is carried out in a channel separate from position detection.

As a result of after-effects radiologists can misread images of one type if they have been exposed for a long period to images with a very different property. After-effects can be blanked by flashing a short bright uniform field. Radiological scientists carrying out observer experiments to measure just noticeable differences or the like need to exercise care in blanking or in waiting an adequate time between stimuli.

3.3 Flicker

Flicker affects our vision of electronic displays or films on light boxes. For viewing medical images, especially for a prolonged time, the observer must not be troubled with flicker of the CRT screen or light box. Flicker sensitivity is significant below a rate of intermittence called the flicker fusion frequency (FFF) (for an overview see [Graham, 1966]). The FFF depends on a large number of parameters: the retinal position, the spatial modulation of the stimulus (fine or coarse structure), the shape of the stimulus, the relative intensity of the stimulus and the surround, whether there is stimulation in the contralateral eye, and the level of dark adaptation. Because all effects work together, it is not easy to study them separately.

At a luminance of the average intensity of viewing radiographs on a lightbox the FFF ranges from 54 Hz in the inferior nasal field to 45 Hz in the central fovea. Indeed it is a well known percept to see flicker in the periphery while in the fovea a flicker-free image is seen. The FFF is highest when the intensities of stimulus and surround are equal. The greatest influence is from the intensity of the stimulus itself; [de Lange, 1952] showed that the FFF increased linearly with the logarithm of the illuminance over more then 7 decades. For very dim light the FFF can be as low as 15 Hz; for the brightest stimulus it goes up to 67 Hz [Brown, 1966]. There is some dependence on stimulus size: larger spots lead to a somewhat higher FFF.

This has implications for viewing medical images. The intensity of CRT screens is some orders of magnitude lower than the intensity of lightboxes. Modern CRT screens have refresh rates from 60 to 100 Hz, and lightboxes are fixed to the power frequency, 60 Hz in the US and 50 Hz in Europe and other areas. The large white area exposed next to the viewed film area on a lightbox may give rise to a substantial perception of flicker. Besides this the existence of the bright area dramatically reduces the visual acuity, i.e. the finest detail that can be discriminated.

4. Texture

Certain patterns of luminance are seen as texture, which affects the parts of a 2D image that are grouped into objects and also can have an effect on the perception of depth in 3D images. Two different categories of texture, geometric and spacing-based, involve somewhat different visual processes. In the geometric form of texture, regions are grouped via elements whose basic features are lines (or elongated regions) with a specified orientation, and corners [Julesz, 1971]. In spacing-based texture, the relative position of regularly or randomly spaced similar elements determines the texture. In this case the texture is defined both by the common or average inter-element spacing and by the properties of the elements. We might, for example, have a coarse texture of small intensity blobs. This latter case obviously involves the multiscale measurements of the early part of the visual system, since the elements are seen at one scale and the spacing at another.

The effect of texture on perception of 2D images comes from the possibility that noise texture can obscure anatomic object patterns (figure 10) and from the possibility that the texture of the anatomic background to an anatomic object of interest can obscure the perception of the object of interest. These form the conspicuity question identified by Revesz and Kundel [1974].

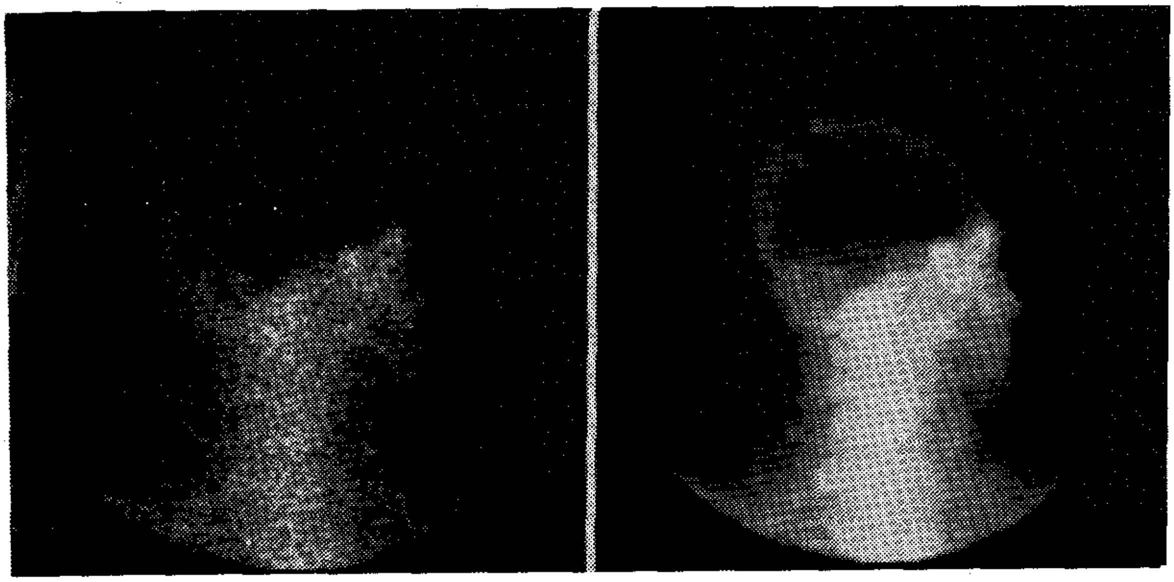
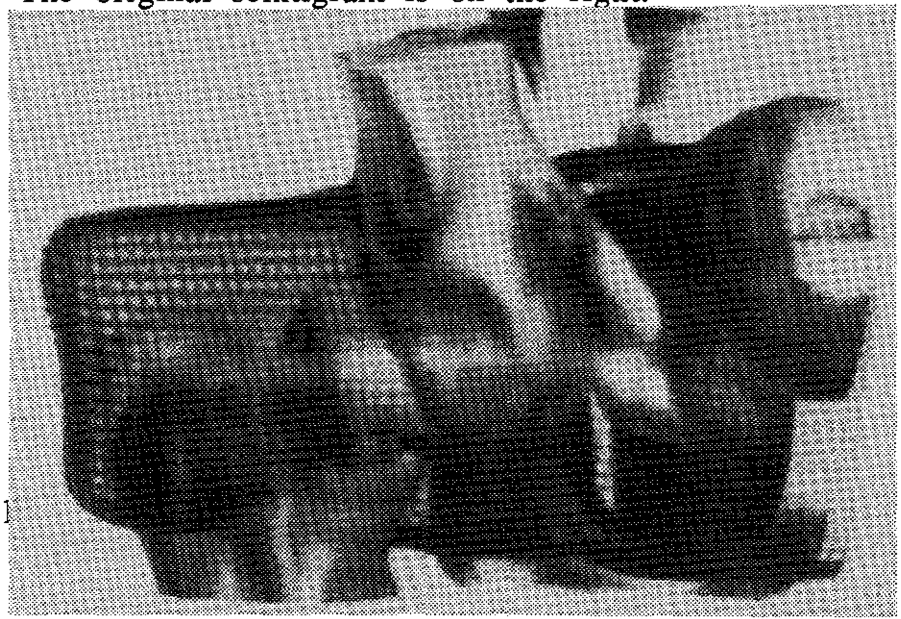


Figure 10. Noise blobs masking real anatomy in a smoothed scintigram(left) of the head. The original scintigram is on the right.



-figure 11

Figure 11. Texture used to convey the shape of a transparent surface. The textured surface show radiation dose, and the opaque surface shows the pelvic bones.

For 3D images texture can be used to help to comprehend shape of surfaces (figure 11). The visual system receives a depth cue based on the fact that spacing of texture elements decreases with the distance of the surface from the observer, and thus with the slant of the surface. By placing texture on a surface according to this rule, the 3D percept can be enhanced. This texture representation has also been used to enhance sensitivity to image intensity changes by so called isometric display, in which intensity is taken as height and the intensity surface resulting from the image is textured.

Another aspect of texture as supporting depth percept is that it gives areas rich with features, which are used for image matching of the two retinal images in binocular stereo. Random dot patterns are much used in research in this area and have been made famous by Julesz (e.g. in his well illustrated book: [Julesz, 1971]).

5. The 3rd Dimension

The human body is three-dimensional. Thus, medical imaging has moved toward image acquisitions in 3D (CT, ECT, MRI, ultrasound) and more recently toward presentations of the image information in three dimensions. These presentations can be divided into two basic categories, those in which image intensity values are placed in visual 3-space, forming a sort of varying intensity, transparent fog. In the other category surfaces, e.g. of organs, are made to appear by simulating the play of light from and through these surfaces (see figure 12). In its primal form, in which the surfaces are explicitly found before the lighting is simulated, this is called 'surface-rendering'. An attractive hybrid of these methods, called 'volume-rendering', involves calculating not only surface shading values everywhere in the 3D space of the image data but also opacities that are made high where surface likelihood is calculated to be high. These shades and opacities are then placed in visual 3-space, forming a sort of varying intensity, variably transparent gel that conveys surfaces where the opacity has been made high.

Most of these presentations allow one to view a 3D structure on a 2D screen. The properties of human 3D vision are of interest to understand how these presentations should be provided.

The percept of depth is derived from many independently computed 'cues'. Especially strongly perceived are strong changes in depth. When such a strong change in depth occurs, it can generate a luminance edge feature, with its subsequent effects on perceived form and brightness.

Perhaps the most powerful depth cue is that of occlusion -- what is hidden or obscured by what. Most of the other cues are dominated by our preattentive conclusion that what is hidden is behind that which is hiding [Nakayama, 1985]. Occlusion generates edges, corners and so-called T-junctions in images (where the silhouette of a faraway object intersects the silhouette of a nearer object), and the visual machinery is well equipped to find just these (see Part II).

Perhaps because of the importance of occlusion and because of the lack of much experience in seeing through fogs, humans are not well equipped to appreciate structure in dense 3D distributions of intensities without much opacity.

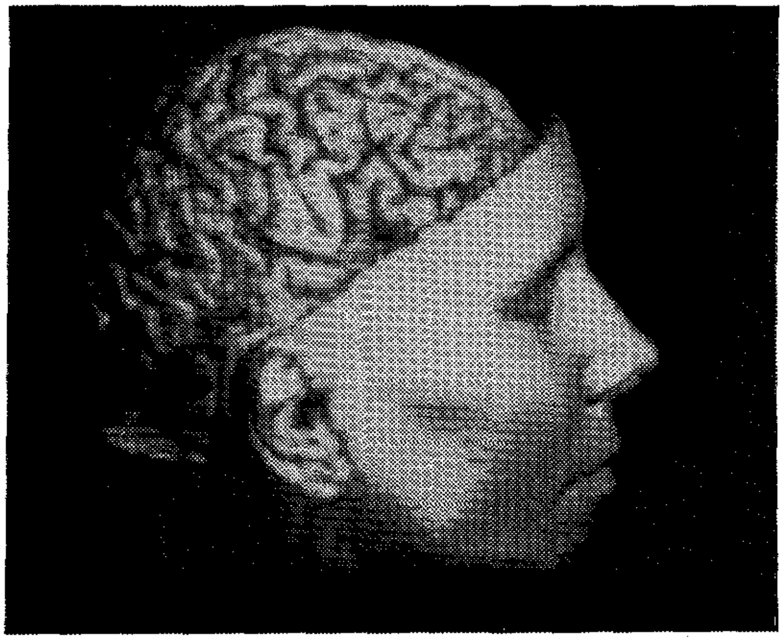


Figure 12. A 3D volume-rendered presentation of the human cortex from 109 MRI slices

Cues related to head-motion are perhaps second in importance only to occlusion. Both the changes in the view of a surface due to parallax, shadings, and reflections, as well as changes in occlusion that result from head motion, have a strong effect. As a result 3D displays that allow the user actually to move his or her head are more effective than those where the object is moved, e.g., by a joystick. Nevertheless, the 3D cue given by motion of the object, called the kinetic depth effect, is a very effective indicator of depth properties of the image. It appears that most of the 3D shape information achieved from motion is derivable from temporally adjacent frames [Todd, 1990] -- in fact, the visual system when shown two frames well-separated in time, unconsciously fills in the intermediate positions and sees the result as a smooth motion. All of these motion cues work best from edges or other sharply defined image features. The importance of the motion cues means that 3D presentations of either surfaces or intensities will be much more effectively seen if motion can be provided, either in terms of precomputed frames or with on-line specifications of the motions desired.

Perceptually, the 3D percept is enhanced if the specification of the motion is tied to a corresponding physical movement of the observer, either by head motion or by hand motion, as if the joystick or trackball were connected to the 3D space being displayed. Then information from the viewer's muscle spindles measuring muscle length, movement, and limb position, as well as a copy of the signal that drives the muscles, updates the internal representation of our percept that the object is at a specific location (a similar mechanism can also be found in eye movements where a continuous update of the changing world is necessary). This finding has led to a whole new class of interactive devices to manipulate computer generated objects [Foley, 1987], like those constructed by surface or volume rendering in radiology. Experiments are being conducted with feedback of head movements to the computer generated CRT display, or even more sophisticatedly, with helmet-mounted displays, where the head rotation and translation is fed back to the computer to interactively (in real time!) update the stereoscopic disparity images viewed by the observer on lightweight LCD television screens on his or her glasses. The viewer has the impression of being in or close to the object, and he or she is able to move around to naturally explore the objects in the image space. With a hand-held 3D mouse, measuring both the rotation and translation of the hand, or the so-called 'data-glove', which has detectors for finger position and orientation, the viewer is able to manipulate anatomic objects or planning objects, such as radiation treatment beams or surgical instruments. Such a system may prove useful for the interactive preparation and exercise for complex operations on patients from 3D image data, before actually starting the real operation. For a realistic percept real-time update of the 3D image is a prerequisite. This is at this moment at the edge of the performance of contemporary graphics devices.

All of the depth cues discussed till now require only one eye. The binocular cues include vergence, i.e., the relative angulation of the eyes toward the surface of attention, and the stereoscopic cue, i.e., the disparity between the views of the two eyes. Except for the fair fraction of persons who are somewhat stereo-blind stereo gives relative depth with good precision, and vergence gives an absolute depth. Both depend on edges or other sharply defined image features. While stereo has been commonly used to view medical and other 3D images and has been deeply studied by visual scientists, its information is strongly dominated by the depth cues mentioned earlier, except perhaps where stereo indicates a sharp change in depth that is not indicated by one of the other cues. An additional 3D cue of moderate effect, but one commonly used in 3D medical image display, is 'shape from shading'. The human visual system makes good use of the variations in surface shading that are produced when the surface is illuminated by directed light sources. In fact, illuminations from above give the strongest cue, presumably because that is the location of the sun. This cue is used by methods of surface and volume rendering that are used in 3D presentation of anatomic surfaces. Yet another, also moderately important cue is 'shape from texture', mentioned in section 5.

6. Variations within and among Observers

While the visual systems of humans with normal vision can be expected to have the basic way of operating in common, the details can be expected to vary from observer to observer. One reason is that the visual system literally grows as a result of the visual experiences of the person (see Part II). Other biological variations cause observers to have shifts in ranges of sensitivity, e.g. to contrast as a function of scale, while having similar range sizes.

A given human will also vary in what he or she sees when the observation is repeated, even if no memory or learning is involved. This inconsistency is frequently modeled by a random internal noise, which may have additive and multiplicative components [Ishida, 1984; Burgess, 1988]. In fact, the neurons that make up our neural circuitry, are relatively noisy and unreliable components, but the networks formed by them have many correcting properties (see Part II).

Part II. How We Do It -- Neuroanatomy of the Visual Pathway

The fundamental properties of visual perception can be understood by studying the behavioral responses of human observers with different visual tasks (psychophysics) or by studying the neurophysiology of the visual system. The last decades have shown a great increase of our knowledge, particularly by the studies of Hubel and Wiesel of the macaque monkey visual cortex, for which work they received the Nobel Prize in 1981 [Hubel, 1982; Wiesel, 1982]. Their work is summarized in the highly recommended book by Hubel [1988].

1. The Retina

The more than 100 million receptors in the retina, the rods and cones, project through an intricate set of retinal layers to about 1 million ganglion cells in the retina whose axons form the optic nerve. So there is a considerable convergence of neuronal connections. It appears that on average 100-150 receptors, projecting via the bipolar, horizontal and amacrine cells to one ganglion cell, form a circular area, called a receptive field (RF). RF's exist in many sizes; the smallest are found in the central region called the fovea, and average diameters increase linearly with eccentricity. Typical diameters of monkey foveal RF's are a few minutes of arc; far out in the periphery they can be 1 degree or more. RF's overlap substantially. This RF structure gives rise to two particularly interesting phenomena: multiresolution and local sign.

1.1. Multiresolution

The sampling of the two-dimensional intensity distribution of the outside world on the retina is done by the RF's, each consisting of many receptors, which means that we see unsharply! Because we see with many different sizes of RF simultaneously, the central nervous system (CNS) has many unsharp representations available simultaneously, each with a different degree of 'defocusing'. This proves to be an important feature, because it effectively averages spatial noise and it generates an automatic hierarchy of the features that we see. An example: when we defocus a picture of a face, we first lose sight of the fine details, like wrinkles and eyelashes; then the nose holes disappear; eventually we are left with only two vague blobs of the eyes; and in the last stage the head itself forms the only remaining blob. This way our central nervous system knows the nesting order and thus the importance of different structures. In contemporary image processing / computer vision techniques this is known as the 'multiresolution' approach, or, because we look at the image on different spatial scales, as 'scale space' processing. The methodology proves to be successful and efficient in both neuronal and computer implementation, because for many operations only the calculation of the coarse structures is necessary. As will be discussed below, receptive fields exhibit extensive specializations, and all of these seem to display multiresolution properties. The abundance of hardware in our central nervous system makes this efficient implementation possible.

1.2. Local sign

The overlap of RF's turns out to be essential. If a certain location on the retina is illuminated, a number of overlapping RF's is stimulated, and the CNS is able to determine the exact location of overlap by determining the correlation between the RF outputs. A neuron is a very good correlation detector: receiving many inputs on its dendrites, it exhibits a threshold, and it acts in a way as a 'coincidence detector': many simultaneous inputs, i.e. high correlation, give rise to a higher probability of firing. It appears that in nerves (like the nervus opticus) it is not the firing patterns along each single fiber that are important, but the correlation that exists between neighboring fibers. Firing of just a single fibre may be an erroneous signal. It appears that such a 'correlation neighborhood' between fibres often has a spatial extent defined by a smoothly decreasing set of weights given by a bell-shape (Gaussian).

This mode of information transfer has many advantages [Koenderink, 1984]. Besides robustness to noise, it provides a means of handling a precise analog signal with many digital lines. You may lose some fibers, through damage or whatever, but the correlation will be kept in order, even if some fibers are lost, so the signal is still there. This very robust solution is a sensible evolution in a creature that has to survive. With this model you may even intermingle nearby fibers, for this has no influence on the local correlation. This is intriguing: it means that it is not too crucial where fibers go exactly when they are growing (in the fetal phase, or when healing after being accidentally cut or damaged), as long as they globally go in the right direction. Here we have a solution for the 'wiring problem': in our computers it is necessary to color or label the wires, for each has a strict destination, which we have to know, especially if we want to repair things. In the CNS this labeling seems to be not necessary. This phenomenon is known as local sign [Koenderink, 1984].

1.3 Receptive field structure

Studying the shape and structure of retinal RF's gives us insight into a number of other phenomena: brightness and contrast perception. The initial response of the rods and cones that are the primary retinal receptors is approximately logarithmically related to local luminance. However, illuminating the retina diffusely over a large area does not produce a response of most retinal ganglion cells because both excitatory and inhibitory connections project on the cell. Most RF's are found to be center-surround, in two classes: 'on-center' (with a ring of inhibitory projecting receptors around it), and 'off-center' (the reverse). For the CNS black and white are equally important. So the messages that the eye sends to the brain have little to do with the absolute intensity of the light illuminating the retina. The cell instead signals the result of a comparison of the amount of light on a certain spot on the retina with the average amount falling on the immediate surround. This implies some counterintuitive results: we see larger spots only by their edges, where the relative intensity changes occur. The absolute intensity does not matter very much. Indeed we are bad estimators of absolute intensity, but we have a dynamic range of luminance as high as 15 orders of magnitude. Our sensitivity to intensity changes also explains the importance of eye movements. If the retinal image is really stabilized on the retina (as is done in some experiments by paralyzing the eye muscles or by mounting an experimental optical system by means of a contact lens on the cornea), all perception of structure is lost after a few seconds. We avoid perceiving the shadow of the blood vessels running in front of the rods and cones by this mechanism. Any slip of the image over the receptors recalls perception immediately, indicating the importance of intensity changes. The eyeball and our head and body movements are continuously taking care of this effect.

1.4 Projections

The ganglions in the retina project to the next stage in the visual system, the lateral geniculate nucleus (LGN), in such a way that neighborhood relations are kept in order. Cells that are neighbors in the retina are also neighbors in the LGN and the next few higher centers. A small number of fibers lead off to take care of pupil size regulation and other instinctual responses. In the LGN again we find grouping in receptive fields, now leading to more complex receptive fields in the outer visual field, i.e. with a more complex structure which can be found by stimulation. From the LGN signals travel further along the geniculo-striate bundle to the visual striate cortex, with its areas V1 and V2, and beyond. Each layer is topographically projected perfectly (a result from local sign?), and it appears that this scheme is found all over in the CNS. There seems to be a strong similarity in our other receptive systems; for example, the pressure sensitive Pucini receptors in our skin exhibit a similar receptive field structure as that found in the retina, as do the hair cells in the organ of Corti in the inner ear, etc. There is no mixing of signals from both retinas in the LGN; this binocular interaction occurs later, in the cortical areas.

Strikingly, there is about three times as much traffic downward from the CNS to the LGN than upward from the LGN to the CNS. While the purpose of this "backwards" traffic is largely unknown, a possible explanation is the control of the selection or summary of data before being sent to higher levels in the cortex, thus reducing the amount of data to be transmitted and processed there.

Already in the retinal ganglions, the cells are divided into three categories, which are maintained through the LGN and the earlier, topographically organized sections of the visual cortex. These categories are named magno, parvo blob, and parvo interblob, for the way they appear under the microscope in histological preparations of the cortex. The existence of these three more or less independent subsystems, each with its own properties, emphasizes the separable feature measurements carried out by the visual system. The parvo interblob system seems concerned with description of scrutinized, static form. It has low contrast sensitivity and temporal resolution, but high acuity. The parvo blob system retains color information and is of low resolution. The magno system seems to be more focused on quick, rough specification of objects. Motion and depth are important features measured by it. It is color-blind and of low spatial resolution, but fast and of high temporal resolution and sensitive to luminance contrast, orientation and form [Livingstone & Hubel, 1988].

2. The Visual Cortex

2.1. Form detection

The human visual cortex is about 2 mm thick, covers a folded area of 30 cm^2 in the back of our head, and contains on the order of 200 million cells. Here 'simple cells', 'complex cells', and 'hypercomplex cells' are found, each with their characteristic RF representation on the retina. These are in increasing order more complex then the center-surround cells found in the retina itself. Most of the RF's do not simply integrate luminance over the field but show a more complex response than RF's in the retina: Simple cells weight local intensities so that there is no sensitivity to orientation. Complex cells are orientation sensitive, i.e. only respond to stimuli oriented in a certain direction, and can be figured as a RF with a number of bands (see figure 13), having the same size or a little larger in the outside visual world than retinal RF's have. They originate from combinations of RF's from the previous layer in the CNS. It is easy to figure that these cells are responsive to lines or edges in a certain direction. They come in all sizes and directions, so sampling of the world can be complete. Like complex cells, hypercomplex cells combine reports from simple cells: they are not only orientation sensitive, but it matters where the end of the line (or edge) stimulus falls on the RF. The effect of the many kinds of cells is to measure sort of oriented, blurry derivatives (differences) of local intensity at various scales, and thus to capture geometrical information.

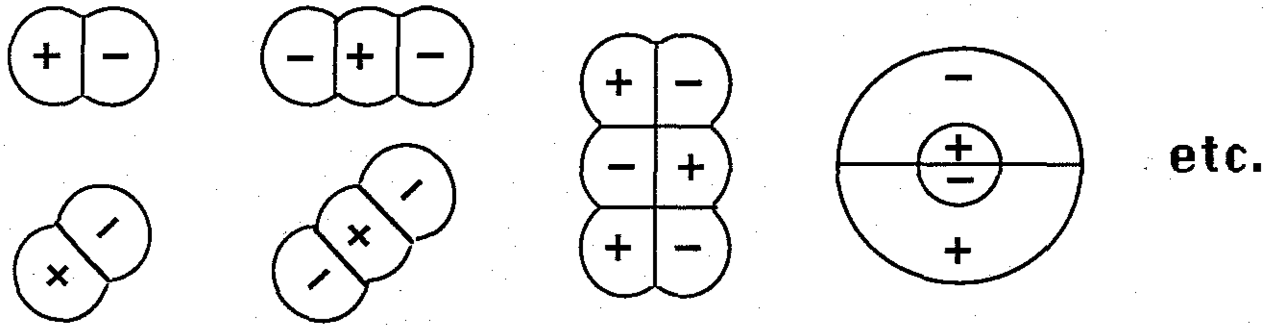


Figure 13. Weighting patterns for cortical cells. The signs "+" and "-" indicate that the marked region is weighted positively and negatively, respectively.

To detect all the basic features in the images that we see, necessary for the perception of shape, recognition, depth, movement etc., the brain has to calculate a large number of certain specific properties from the incoming intensity distribution. These properties are very likely calculated from the outputs of the variety of RF's described above [Koenderink, 1987, 1988; Zucker, 1986]. They do not depend on translation, rotation and scaling, since we recognize objects independent of where they are in our visual field, or how large they are. These features should also be detected while motion occurs, both of the feature and the observer.

For the finding of image shape features such as edges (boundaries), texture, boundary T-junctions (indicating where one object goes behind another object), curvature of isointensity contours, etc., one would like to have available many derived images made from various local spatial differences of the image intensity distribution. The shapes of the different experimentally RF's, as described above, seem very closely to match filters which determine these local difference distributions [Young, 1986]. It has been proposed [Koenderink, 1987, 1988] that in this way the required properties are calculated and so supply a means to form an effective language to perform segmentation, matching, 3D shape comprehension [Koenderink, 1990], etc. Combined with multiresolution, this may also shed some light on why we see so easily the overall form, or 'Gestalt', of structures, discarding the fine details. In a radiograph we recognize a lung immediately as a lung, even under the enormous biological variation.

2.2 *Movement detection*

Motion is coded in separate channels from those measuring spatial intensity variations. Ideas about the possible neurophysiological structure of the mechanism for motion detection come from studies on the visual system of the fly. As mentioned in section 3.1 of part I, detection may be done using two neighboring or slightly overlapping RF's in a bilocal detector: the output of one RF directly projects on an output cell in a next layer, and the other RF projects on this cell via an intermediate neuron that causes a delay in this signal (see figure 8). Then the maximum output is generated in that next layer cell if the stimulus moves with a velocity over the RF's such that the outputs of the RF's arrive simultaneously, because, as we have seen, then the neuron has a greater chance of firing. This happens when the distance (the 'span') between the two RF's is traversed by the stimulus in the same time as given by the delay of the intermediate cell. This mechanism (known as the Reichardt detector) requires an extensive set of many delays over many distances in all orientations (see figure 9), but, as we know, in our CNS there is no lack of hardware/wetware.

This mechanism is confirmed psychophysically in the human visual system [Barlow, 1965], and electrophysiologically in many animals, ranging from insects to mammals. There is also insight in how the sizes and spans of the RF's in the motion detectors are arranged: because our RF's become larger on average as we move toward the periphery of the retina, we can see fine movement differences in the fovea, while thresholds for detection of changes in velocity become greater in the periphery. Faster detectors have a larger span; this span is about linear with the eccentricity. At each eccentricity one can measure [van de Grind, 1986; van Doorn, 1983, 1984] the range of detectable velocities, and about in the center of this range one finds a 'critical velocity' where both spans and delays are minimum. At velocities lower than this critical velocity we use detectors with constant span and delays inversely proportional to the velocity; at higher velocities the delays of the used detectors are constant, and span increases linearly with velocity. This gives a nice model that is able to predict the properties of the human motion system at any eccentricity.

The system is extremely robust to noise: we are able to detect coherent patches in spatiotemporal noise when the noise is 100 times larger than the signal! [van Doorn, 1983]. Again we see the influence of two important factors: we are able to use spatial averaging over a certain area (the areas of the RF's) to reduce the noise, and we are able to use large numbers of the bilocal detectors (up to several thousands) to increase the sensitivity of the detection. Motion detection is rather insensitive to contrast degradation, and it works well at all luminances.

Flicker sensitivity is built up from two components:

1) the temporal characteristics of the receptors in the retina, where the recovery of the cell takes a finite amount of time (this can be measured with the electroretinogram (ERG) in humans), and

2) the delays and feedback of the signal as it travels further into the visual system.

The transmission speed along the neuronal axons and the firing frequencies are relatively slow. The high performance of our visual system as a whole is primarily due to the massively parallel behavior of all cells with their accompanying structure.

2.3 Learning

Contemporary research on perception is being greatly stimulated by the discoveries made in the field of 'neural networks'. The phenomenon of our great capability to learn and to remember has puzzled generations of researchers. A localized structure in the brain, holding our memory, has never been found.

Recently it was realized that the information we remember is stored in the strength of the synaptic connections between the neurons in our brain. If we estimate the total number of cells to be 10^{10} and on average we find 10^4 synapses per neuron, we have on the order of 10^{14} synapses available.

Hebb [1949] was the first to realize that the synaptic strength of a connection is formed by just using it. If a pattern is often seen and the appropriate synapses are often passed, their strengths increase. Also, if synapses are never used, they degenerate and vanish. Thus learning is setting up synaptic structure in our network, which is extensively and even redundantly wired at birth. Seeing the world during the first months of our life shapes our network and thereby the patterns with which we subsequently compare new incoming patterns. This is a continuing process: each new stimulus again reinforces the connections, because they are reused. So by using the system it is also continuously recalibrated on a long term and a rather short term scale (e.g., getting acquainted to new eyeglasses, adjusting the range of grabbing movements as we grow, or getting used to better resolution in medical images, as in CT and MR over the last years).

The plasticity of the visual system after birth was studied thoroughly after the discovery of RF's [Wiesel, 1982]. When one eye of a monkey was closed during the first months after birth, Wiesel found a strong degeneration of just the layers belonging to that eye, both in the LGN as higher cortical areas. It is estimated that 10% of the people have difficulties in the perception of depth, possibly due to, among other causes, early slight strabismus.

Recent 'connectionist' models of human vision and other cognitive processes such as understanding of speech and written language shed light on how we learn. These artificial neural networks with dense diverging and converging connections [Hopfield, 1982] learn by storing information in the strength of the connections. They accomplish direct association of incoming structured information with the information stored during the learning session by recreating the closest prototype pattern where the "fit" (association) to a pattern stored during the interactions is high enough. As a result association with minor variants of learned patterns is possible but recognition of patterns with many unfamiliar features is not easy.

Learning of an extensive reference set is thus a necessary prerequisite for radiological diagnosis. This fact explains some of the difficulties with the acceptance of new modalities with unfamiliar components, such as MR with its many parameters.

This associative form of learning and recognition applies just as well for other sensory inputs, as it does for our internal motor commands to make movements. In general we remember not just one sense, but always the whole context of the simultaneous input that accompanies the stimulus (including even the emotions). An image, familiar in black and white presentation because we have seen it often (e.g. a certain class of radiographs) is not or hardly recognized when a color coding scheme has been applied. This is even so when the absolute intensities of the pixels (when measured) are equal. The relevance of context also explains the importance of reading radiological images using strict protocols.

So we are not perceiving the outside world as a camera, but as a pattern matcher, shaped by this same world that we are moving observers in. Realizing that our perceptual system is formed by our environment, to which we still constantly adapt, and discovering the emerging details of the functioning of general neural structures, learned from early vision processing, are important steps in finding a good match between the presentation of medical images and our perception, and in understanding the extraordinary mechanism of human perception.

Acknowledgements

We are grateful for critical comments on drafts of this paper by Roger Bauman, Edward Chaney, R. Eugene Johnston, Victor Klymenko, and William McCartney. We thank Johannes Blom, James Coggins, Luc Florack, Petra van de Elzen, Wim van de Grind, and Jan Koenderink for helpful discussions and Carolyn Din and Robert Cromartie, Graham Gash, Victoria Interrante and Randy Perry for help with the references and figures, respectively. This paper was prepared under the partial support of Program Project Grant P01 CA47982 from the U.S. National Institutes of Health and SPIN grant #[VS-3DM] 50249/89-01 from the Dutch Ministry of Economic Affairs.

References

Barlow HB, Levick WR: The mechanism of directionally selective units in the rabbit's retina. J. Physiol 178:477-504,1965.

Brown JL, Graham, CH (Ed.): Flicker and intermittent stimulation.
Vision and Visual Perception. John Wiley & Sons, New York, 1966, 251-321.

Burgess A, Colborne B: Visual signal detection IV: observer inconsistency. J. Opt. Soc. Am. A5: 617-627 (1988).

van Doorn AJ, Koenderink JJ: The structure of the human motion detection systemm. IEEE Trans Sys, Man, Cyb, SMC-13(5):916-922,1983.

van Doorn AJ, Koenderink JJ: Spatiotemporal integration in the detection of coherent motion. Vision Res 24(1):47-53, 1984.

Foley, JD: Interfaces for advanced computing. Scient. Amer. 257: 72-81, 1987.

Gilchrist AL, Jacobsen, A: Lightness constancy through a veiling luminance, J. Exp. Psych: Human Percept. and Performance 9(6):936-944, 1983.

Graham, CH (Ed.): Vision and Visual Perception. John Wiley & Sons, New York, 1966.

Gregory RL: The Intelligent Eye, Weidenfeld & Nicolson, London, 1970.

van de Grind WA: The Distribution of Human Motion Detector Properties in the Monocular Visual Field. Vision Res. 26(5):797-810, 1986.

Grossberg S, Mingolla E: Neural Dynamics of Perceptual Grouping: Textures, Boundaries, and Emergent Segmentations. Perception & Psychophysics, 38:2, 1985, 141-171.

Hebb DO: The Organization of Behavior: chapter 4: The first stage of perception: growth of the assembly. 60-78, Wiley, New York,1949.

Hopfield JJ: Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA, 79:2554-2558, 1982.

Hubel DH: Eye, Brain and Vision. Scientific American Library series #22,1988.

Hubel DH: Exploration of the primary visual cortex. 1955-78 (Nobel Lecture), Nature 299, 515-524, 1982.

Ishida M, Doi K, Loo L-N, Metz CE, Lehr JL: Effect of digital image processing on the detectability of simulated low-contrast radiographic patterns: Radiology 150:569-575, 1984.

Julesz B: Foundations of Cyclopean Perception, Univ. of Chicago Press, Chicago,1971.

Koenderink JJ: The concept of local sign. In: A.J. van Doorn, W.A. van de Grind, J.J. Koenderink (Eds.). Limits in perception. VNU Science press:495-547, 1984.

Koenderink, JJ: Optic Flow. Vision Res 26, 1986, 161-180.

Koenderink, JJ: Representation of local geometry in the visual system. Biol. Cybernetics 55, 1987, 367-375.

Koenderink, JJ: Operational significance of receptive field assemblies. Biol. Cybernetics 58, 1988, 163-171.

Koenderink, JJ, Doorn, A.J. van: Local structure of movement parallax of the plane. J.Opt.Soc.Amer. 66, 1976, 717-723.

Koenderink, JJ, Solid Shape, MIT Press, Cambridge, MA, 1990.

Kundel, HL, Nodine, C.F., Thickman, D.T., Toto, L., Searching for lung nodules: a comparison of human performance with random and systematic scanning models, Inv Radiol 22:417-422, 1987.

Land, EH, McCann JJ: Lightness and Retinex Theory, J. Opt. Soc. Am., 61, 1971.

Lange, H de: Experiments on flicker and some calculations on an electrical analogue of the fovea systems. Physica 18, 935-950, 1952.

Livingstone M, Hubel D: Segregation od for, color, movement,, and depth: anatomy, physiology, and perception, Science 240:740-749, 1988.

McKee SP: A local mechanism for differential velocity detection. Vision Res. 21, 491-500, 1981.

Nakayama K: Biological image motion processing: a review. Vision Res. 25:625-660, 1985.

van Nes F L, Bouman MA: Spatial modulation transfer in the human eye. J. Opt. Soc. Am. 57: 401-406, 1967.

Nodine CF, Kundel HL: A visual dwell algorithm can aid search and recognition of missed lung nodules in chest radiographs, Visual Search D. Brogan, ed, London, Taylor & Francis 1990.

Pizer SM, Amburn EP, Austin JD, Cromartie R. , et al: Adaptive histogram equalization and its variations, Comp Vis, Graphics , and Image Proc. 39:355-368, 1986.

Ramachandran VS: Interaction between color and motion in human vision, Nature 328:645-647, 1987.

Ramachandran VS, Gregory RL: Does color provide an input to human motion perception, Nature 275: 55-56, 1978.

Reichardt W: Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. Sensory Communication, Rosenblith, W.A., ed.:303-317, MIT Press, Cambridge MA, 1961.

Revesz G, Kundel HL, Graber MA: The influence of structured noise on the detection of radiologic abnormalities, Invest. Radiol., 9: 479-486, 1974.

Rogers D, Johnston RE, Pizer SM: Effect of ambient light on electronically displayed medical images as measured by luminance-discrimination thresholds, J. Opt Sci Am 4(5):926-83, 1987.

Todd JT: The perception of three-dimensional affine structure from minimal apparent motion sequences, Perception & Psychophysics, in press.

Treisman A: Features and objects in visual processing, Scientific American, 255 (5):114B-125, 1986.

Wiesel TN: Postnatal development of the visual cortex and the influence of environment (Nobel Lecture). Nature 299, 583-591, 1982.

van der Wildt GJ: Contrast Detection and its Dependence on The Presence of Edges and Lines in the Stimulus Field. Vision Res. 23, 1983 , 821-830.

Young RA: Simulation of human retinal function with the Gaussian derivative model. Proc. IEEE CVPR, Miami Fla., 1986, 564-569.

Zucker SW, Hummel RA: Receptive field representation of visual information. Hum. Neurobiol. 5, 1986, 121-128.