

Implementations and Implications of Foveated Vision

Cornelius Weber and Jochen Triesch

Frankfurt Institute for Advanced Studies, Goethe University Frankfurt, Germany

Abstract: Humans are equipped with "space-variant" vision, i.e. a concentration of photoreceptors, retinal ganglion cells and other visual resources at the central fovea, and a sparser coverage of other regions within a wide 180 degree field of view. If the entire visual field was equipped with foveal ganglion cell resolution, then the brain would have to cope with approximately 350 times more visual information.

We will review the very few existing hardware implementations and patents involving space variant vision. Still, the majority of the computer vision community traditionally turns a blind eye on space-variant vision, because it comes along with distorted image representations, complicating standard geometry-based processing. Recent learning algorithms for feature detection and transformations are more flexible and may cope with foveated images.

Foveated vision requires an active vision system: ballistic eye-movements termed "saccades" frequently move the fovea to points of interest in the visual field. The metric of saccades is adjustable, and the resolution increase at the fovea may play a role in supplying the feedback to the system. Furthermore, saccades are related to visual space perception and embodied vision.

Keywords: Retina, space variant vision, camera, machine learning.

1. INTRODUCTION

The retina of the human eye is often compared to the CCD or CMOS photosensor of a camera. However, in addition to photoreceptors, the retina has several neuronal layers that perform pre-processing of the incoming image stream before the retinal ganglion cells (RGCs) project to recipient visual areas of the brain. Pre-processing involves the following elements:

- compression of the 6.4 million cone- and 110 million rod photoreceptors to one million RGC fibers that connect to the brain [34],
- transformation of the sustained photoreceptor light responses to transient RGC responses [43] and
- spatial filtering for contrast enhancement and noise removal [33].

All these processing steps happen in a space-variant resolution that assigns most resources to the center of the visual field, the fovea.

We will review biological findings and biologically-inspired computer implementations and patents. Our emphasis is here on space-variant vision. A high foveal- and a low peripheral resolution are custom-tailored for active scene analysis with a dynamic focus on a current point of interest. Such active vision systems are common in biology and may be part of future robots.

2. BIOLOGY

Foveated vision is largely biological inspired. In Section "Retinal Processing", we will take a more detailed look at the abovementioned three elements of pre-processing. The retinal circuitry compresses and alters the visual signal, independently of foveation that considers only the spatial layout of its cells. A reader who wishes to focus on foveation may continue directly with Section "2.2. Space-Variant Vision".

2.1. Retinal Processing

Circuitry in the Retina. The processing steps in the retina are done on five layers of neurons in the following order:

1. Photoreceptors (rods and cones) respond to light in a sustained fashion.
2. Horizontal cells provide inhibitory feedback to surrounding photoreceptors via hemi gap junctions [33].
3. Bipolar cells receive their input mainly from the photoreceptors. Some are tuned to faster and some to slower fluctuations in the visual signal [33]. There are separate ON-transient, ON-sustained, OFF-transient and OFF-sustained bipolar cells [39].

4. Amacrine cells, which are mostly inhibitory (GABA), seem to lack an axon and are with 29 identified types the most diverse cell class in the retina [39]. There is inhibitory feedback from the amacrine cells to the bipolar cells.
5. Ganglion cells receive their input from amacrine or bipolar cells, and some may feed back via gap junctions at least to the amacrine cells [32]. Bistratified RGCs (they have dendrites on both sides of the amacrine layer) include an ON-OFF direction selective cell [39] that fires at onset as well as at offset of a visual signal.

One function of the retinal circuitry is to convert a sustained light response of the photoreceptors into a phasic response that reflects light changes: only what changes is of interest. While bipolar cells respond slowly and in a sustained fashion, most RGCs respond rapidly and transiently [79]. Transient RGC responses can be generated by the inhibitory amacrine cells truncating a more sustained excitatory input [43].

Another function can be described by a spatial filter operation endowing the RGCs with center-surround type receptive fields. The negative feedback from horizontal cells to the cones is critical for the formation of this antagonistic surround [16].

Receptive Fields. Many ganglion cell types have an ON-center OFF-surround receptive field, meaning that they respond maximally to a center light-onset surrounded by light-offset. Conversely, there are OFF-center ON-surround types. An intuition about these "Mexican hat" shaped receptive fields is that "the center Gaussian sums locally correlated signals to improve the signal-to-noise ratio, and the surround Gaussian subtracts broadly correlated signals to reduce redundancy" [10]. Since the ganglion cells' receptive fields overlap in the retina, it was shown that for natural images with statistical correlations between nearby positions, such an overlap maximizes the information in the ganglion cells' population responses by increasing the signal-to-noise ratio at the cost of greater redundancy [10].

Parallel Visual Streams. The retina is covered respectively with at least 17 distinct ganglion cell types [20,50]. Each of these has different function and resolution.

The most prominent classification is into Magno (M) and parvocellular (P) streams: M ganglion cells are large and respond transiently to illumination changes. P ganglion cells are small and respond in a sustained fashion; furthermore, many of them are color selective [54].

Both M and P RGCs are sub-divided into ON-center and OFF-center RGCs. ON-center cells are activated at light onset in the center of their receptive fields (RFs), whereas OFF-center cells fire in response to light offset in the RF center [33]. Both have an antagonistic surround.

The P cells are sub-divided into different color channels (the numbers give their percentages from [81]):

- red-center-ON -- green-surround-OFF (21%),
- green-center-ON -- red-surround-OFF (11%),
- green-center-OFF -- red-surround-ON (9%),
- red-center-OFF -- green-surround-ON (5%),
- blue-center-ON -- yellow-surround-OFF (5.7%),
- yellow-center-ON -- blue-surround-OFF (0.3%; possibly not existent).

The color-antagonism makes these cells spectral filters, narrowing the spectral band to which they respond best [34]. But, why is the opponent response collected from the spatial surround (if we would not distinguish center and surround, we would have just two red-green channels instead of four!)?

As an effect of the surround opponency, large spots are processed by the opponent channel, but small spots that fall only on the center part of a receptive field are not. Thus, small spots are more likely to be perceived without color amplification; there is no opponency. This forms a basis for a higher resolution of the monochromatic pathway, involving high spatial frequency edge detectors, compared to the color-sensitive pathway in the visual cortex.

It is still debated whether the color surround antagonism results from a random placement of RGC receptive fields, or whether RGCs actively shape their receptive fields by means of learning to select their input [11]. For example, red- and green- (long and medium wavelength) selective cones are locally randomly dispersed in the retina. Since the RGC centers are connected only with very few cones, random fluctuations would almost inevitably produce a center selectivity for a specific wavelength. Small anisotropies, such as an elliptical RGC input field, can dramatically alter the color input [65].

The abovementioned RGCs project to the lateral geniculate nucleus (LGN), in which one can distinguish ON- and OFF-layers, as well as M- and P-layers [70]. The M stream then leads through dorsal visual cortical areas, and gives rise

to motion analysis and space judgements. The P stream leads through ventral visual cortical areas, and gives rise to high-resolution and color vision with implications for object analysis. Together, these cortical pathways are associated with conscious visual perception.

Further examples of RGC types that project to further pathways are:

- ON-OFF-center RGCs. They respond to both light onset and offset. They are believed to project to the superior colliculus (SC) where many neurons in the superficial, optical layer have similar response properties [22]. Furthermore, long-latency and short-latency ON-OFF cells can be distinguished [12].
- Large and fast RGCs that are deemed to belong to the M-stream provide input to both, the LGN and the SC [50].
- Photosensitive ganglion cells innervate the suprachiasmatic nucleus of the hypothalamus, which is the circadian pacemaker of the human brain [8]. Neither rods nor cones are required to activate these RGCs which are small in number and large in size.

OFF center cells have been found to be more numerous and to branch more densely and thus collect more synapses per visual angle than ON center cells [56]. This has been related to the statistics of natural scenes, i.e. dark fine structure on light background (e.g. black text on paper) occurs more often than the reverse.

2.2. Space-Variant Vision

Compression Ratio of 350. We estimate how much the visual system is compressed w.r.t. an alternative implementation in which the entire visual field was represented with foveal resolution. We use data [34] that has been collected from several studies [47,53,15,69].

The highest density of cones at the center of the fovea is 161,900/mm². In the fovea, there are 2 ganglion cells per cone, hence leading to a RGC density of $2 \cdot 161,900/\text{mm}^2 = 323,800/\text{mm}^2$. If the total area of each human retina, which is 1094mm², was covered by this RGC density, then the number of RGCs would amount to $1094 \cdot 323,800 = 354,237,200$. Since every RGC sends its axon through the optic nerve, it would contain as many axons. However, the number of axons in the optic nerve of each eye is only 1,000,000, which is roughly 350 times less. This is the factor by which foveated information is less than a hypothetical full-resolution information.

Other researchers have estimated that the brain would weigh 60kg if it processed full-resolution vision in the entire visual field [5].

Photoreceptors. The density Y of cone photoreceptors in 1000 cells/mm² as a function of visual field eccentricity E in degrees can be described by a combination of three exponentials in the form [24]

$$Y = c_1 \exp(\lambda_1 E) + c_2 \exp(\lambda_2 E) + c_3 \exp(\lambda_3 E) \quad (1)$$

with six constants which are for the human: $c_1 = 166.9517$, $c_2 = 21.3213$, $c_3 = 4.5576$, $\lambda_1 = -1.35$, $\lambda_2 = -0.1455$, $\lambda_3 = -0.0041$. These values were taken from least-square error minimization of human cone density data by [47,63]. Since cone photoreceptor density is measured in distance from fovea, a separate calculation had been performed to convert this to visual angle for the above data.

The rod photoreceptors permit peripheral grey-level vision in darkness. There are no rods in the center of the fovea, and rod responses are believed to saturate at normal light levels. Therefore, we pay little attention to them, here.

Retinal Ganglion Cells. The RGC density is more extremely foveated in the center and it falls off more strongly toward the periphery, compared to the cones. Only the RGCs send their axons to the brain, hence their distribution matters to all following information processing. The reason why there is a relatively more homogeneous distribution of photoreceptors may be to increase the signal-to-noise ratio in the periphery at negligible cost.

A mathematical formula for RGC densities as a function of visual field eccentricity E in degrees is given in [64] based on data of [47,52] as

$$D^{1/2} \sim M = M_0 / (1 + a E + b E^3) \quad (2)$$

with, e.g., parameters $a = 0.29$, $b = 0.000012$ for the temporal retinal quadrant.

The cortical magnification factor M describes how many mm of cortex are devoted to 1° of visual angle; $M_0 = 8\text{mm/degree}$ is the value for the fovea. The density D of RGC receptive fields (RFs) in RFs/mm² is proportional to M^2

[64] (i.e., the area measure D relates to the square of the length measure M). The RGC RF density equals the RGC cell density almost everywhere, except that the RGC cell bodies are displaced away from the foveola (center of fovea); hence, the RF density is the functionally relevant entity.

Equation 2 is plotted in Fig. (1) using the individual parameters of the four quadrants of the visual field. The field of view extends furthest in the temporal quadrant where the fit is good up to 80° . For the nasal and inferior quadrants, the fit was given up to 60° ; for the superior quadrant up to 45° .

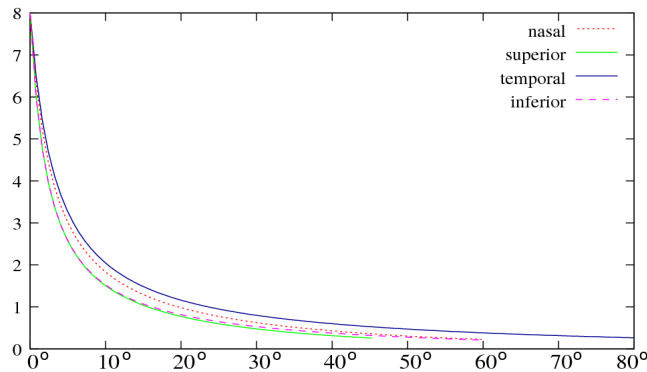


Fig. (1). RGC distributions in the four quadrants of the visual field as function of eccentricity in degrees, from [64]. The y-axis is approximately the square root of the RGC density with arbitrary scaling constant (see text for details).

Visual Target Areas. The topographic mapping of the visual field to the surface of the striate cortex (visual area V1) can be obtained directly by recording cortical responses to light spots at varying locations in the visual field. This circumvents measuring photoreceptor and RGC densities. From such data, a cortical mapping function is characterized as a complex logarithmic (“log-polar”) mapping [71] which maps the radius r and angle φ in the visual plane to a cortical cartesian position (x,y) according to

$$\begin{aligned} x &= \ln r \\ y &= \varphi . \end{aligned}$$

Hence, the round retina is projected to a rectangular cortical model layout with one cortical axis for retinal radius and the other for retinal angle.

The logarithmic mapping is an idealization and has an aesthetical flaw of a singularity at the foveal center $r = 0$. This can be fixed by adding a small constant to the radius before taking the log, or by assigning a uniform resolution to a small center disk.

It is anatomically plausible that the lower visual areas devote their resources proportionally to the amount of incoming fibers. However, in monkeys, the central visual field seems to have an even larger cortical representation [64]. It is unknown whether this increased foveal representation in cortex w.r.t. retina is innate (e.g. via chemical markers), or the result of activity-driven learning (e.g. micro-saccades which increase the activity in visual area V1 [38] might be more effective near the fovea, where receptive fields are small, than in the far periphery, where receptive fields are large).

Further Anisotropies. So far we have described one general RGC density function. When looking at specific RGC subtypes, we find that these deviate from the average density distribution.

The percentage of retinal ganglion cells that project to the SC is 6% in the fovea, increasing slightly with eccentricity, as measured in macaque [49]. (The relatively larger percentage of SC innervation from the peripheral visual field is likely to be paired with a higher percentage of the ON-OFF type of RGC that projects to SC in the periphery.) Hence the relatively weaker foveation in the SC pathway coincides with the role of the SC in controlling eye-movements, which are always directed away from the foveal fixation point.

Bearing in mind that dorsal cortical areas (“where”-pathway), like the SC, are relatively more peripherally connected, while ventral cortical areas (“what”-pathway) are more foveally connected. Attention even increases effective foveation in ventral area neurons: a study finds large receptive fields when measured with plain background,

which shrink to a small region around the fovea when the scene is complex [3].

Another aspect is that early differential innervation of "what"- and "where"-pathways may be causal to the differentiation of these two visual streams during development.

RGCs projecting to the pretectum that controls reflexes of pupil and lense, are primarily found in the inferior and nasal retinal quadrants [80]. These quadrants represent the upper and lateral visual field, which may be more relevant in determining the light conditions than the lower visual field, and hence may be preferred for adjusting the pupils. There is also an asymmetry in cone density: in mouse retina, short wavelength sensitive cones (blue light) are higher concentrated in ventral (inferior) retina, while medium wavelength sensitive cones (green light) are more in the dorsal half of the retina [74].

There is a greater than threefold variability in cone density between human individuals at the fovea [14]. These strong variations level out almost completely at a retinal eccentricity of 0.3mm, which corresponds to roughly 1° of visual angle. The size of V1 and other visual cortical areas show roughly twofold individual variability in the macaque [91], whether or not this is a consequence of retinal variations.

Development. The number and density of RGCs are the result of a process of organization. It was found in chick embryo that an initial overproduction of RGC number is down-regulated by RGC signals involving neural growth factor (NGF), inhibiting the generation of new RGCs and killing incoming migratory RGCs [23]. A second phase of RGC death coincides with the time of their axons reaching the respective target areas. It is likely that those RGCs that cannot establish proper connections die off, possibly via a lack of neurotrophic input from the target neurons, such as brain-derived neurotrophic factor (BDNF).

The different density distributions of cones and ganglion cells seems to develop independently of each other, since the bipolar cells that connect them are "born" only several days later in the mouse [28].

The center five degrees of the newborn retina are underdeveloped and may not be fully functional in the newborn infant [1]. Photoreceptors in the periphery mature earlier than in the fovea [29]. Furthermore, the motion-sensitive magnocellular (M) pathway matures earlier than the color-sensitive parvocellular (P) pathway [27]. This might reflect that peripheral- and motion-sensitive viewing, which determines where to look at, needs to be in place before foveal viewing, which identifies objects, can develop meaningfully.

Topographic Mapping. Foveated vision has implications for the development of topographic neural projections, such as to the SC, because it leads to strong distortions. The lower-vertebrate analogue of the SC is the tectum, and models of the retino-tectal projection assume a space-invariant visual representation, since fish, amphibia and many birds do not have much of a fovea. The correspondences between retinal and tectal positions that are to be neurally connected, are determined in these models by chemical gradients that vary linearly over retinal and tectal space [25].

These models may poorly generalize to foveated vision where most of the retinal cells are near one point at the fovea: since a linear gradient barely shows any change near this point, many retinal cells would become chemically indistinguishable at their targets. A solution might be a chemical concentration that changes steeply at the fovea. A molecule that is concentrated on the temporal but not nasal half of the retina could meet this demand, even though this was first found in chicken which do not have a pronounced fovea [40,41]. Animals that have foveated vision rather need to be examined.

Summary. Beyond foveation, a striking aspect of biological vision is a division of functionalities into different streams. The two most prominent are (i) a fast stream that is less foveated and that localizes objects of interest in the dorsal visual pathway and the SC, and leads to eye movements, and (ii) a slow stream that is highly foveated and that analyzes a focused object in the ventral visual pathway. Both together are required for scene analysis.

3. IMPLEMENTATIONS

3.1. Space-Variant Vision

Implementations of space-variant vision can vary considerably depending on the purpose and the specific processing steps to be modeled.

Photoreceptor/RGC Densities. An implementation of the RGC density distribution of Eq. 2 is shown in Fig. 2. The RGC center points have been placed in a ring-like fashion, where the thickness of each ring and the number of cells within it is given by Eq. 2. Disadvantages of this approach is that the total number of cells assigned is not known beforehand and has to be adjusted iteratively; yet it may not be possible to assign an exact given number of cells, due to symmetry reasons.

Furthermore, the RGC positions will have to be placed onto a square grid, if pixels from a digital photo taken with a conventional CCD/CMOS photosensor need to be sampled. Near the fovea, the positions fit badly into the pixel grid, and the density of the RGC cells may go beyond the resolution of the digital photo. Together, many considerations have to be made to obtain a practical cell distribution.

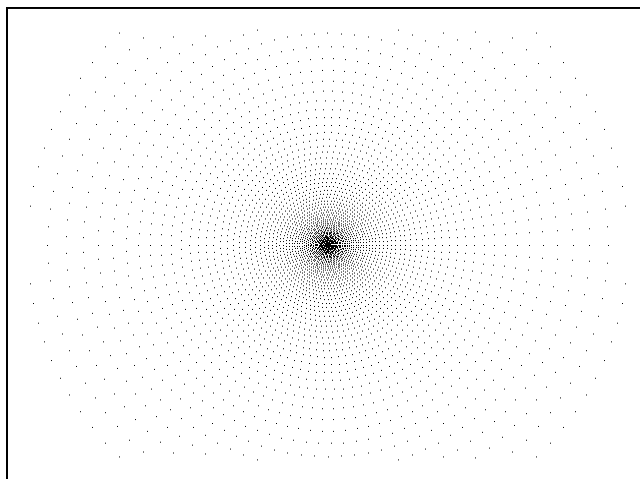


Fig. (2). Foveated vision. Sample points shown in black are arranged according to RGC density.

As an alternative to analytic density and mapping functions, retinal cell densities can be modeled by self-organization [5]. The positions of retinal cells can "grow" by first defining an approximate desired density function and then organizing the cell positions by a Kohonen Self-Organized Map, which smoothes this density function.

As part of a detailed eye model, Deering (2005) [17] uses a model in which the cones rearrange themselves by repetitively applying forces between the cones. This dynamic process yields an observed hexagonal arrangement of cones that cannot be achieved by placing cones in rings around the fovea.

Foveated Vision Hardware. Projects between 1981 and 1997 have developed space variant CCD, CMOS and color CMOS photosensor chips (e.g. [68]; see also Section "3.3. Patents"). Their resolution is highest at the center and falls off at the periphery. These research projects have been completed successfully with the construction of some demonstrator chips, but the production of these special-purpose photosensor chips was not commenced on a commercial basis; instead, space-variant vision is today generally performed by software on the basis of high-resolution commercial photosensor chips.

Direct Mapping to Target Areas. The densities of the retinal photoreceptors and ganglion cells are often bypassed by considering directly the mapping of the visual field to the visual target areas in the brain, such as in [6,76], also for the mapping from the retina to the SC. A simple geometrical description, such as a log-polar mapping [71], also lends itself to some image processing algorithms, such as spatial frequency analysis, which are generally difficult to perform on space variant representations [9].

3.2. Receptive Fields

In the following, we will review *generative* models that explain receptive field properties of RGCs by a learning algorithm. Generative models have hidden layer neurons (the RGCs) which capture in their activations all information of the input layer neurons (the photoreceptors), such that this information may be reconstructed by feedback. It may seem that each of the abovementioned "parallel" visual streams is concerned with only one specific "information channel".

Generative Models. The center-surround type filter properties of the RGCs can be learned using a generative model [75]. Such a model has two layers, an input- and a hidden layer (e.g. bipolar cells and ganglion cells) with bi-directional connections between them. The hidden layer units' activations are used to reconstruct ("generate") the activations of the input units. This architecture resembles that of an autoencoder. Over the course of showing a lot of input data, the connection strengths (weights) between both layers are adjusted so to make the reconstruction optimal;

the reconstruction error is a cost term that is minimized during learning.

To avoid degenerate solutions, generative models are mostly given additional constraints or cost terms:

- A cost term on the connection strengths (synapses) leads to center-surround RGC-like receptive fields [75], if natural images are given as input training data.
- The hidden layer may have less units than the input layer, forcing data compression by the hidden layer. An example is principle component analysis (PCA).
- Only a small number of hidden units may be allowed to be active at any given time. Such "sparse coding" leads to the extraction of independent features from natural images by the hidden units. In models of the visual system, the units become edge detectors, having Gabor filter like receptive fields, as found in the primary visual cortex V1 [46]. This is similar to independent component analysis (ICA) [45].

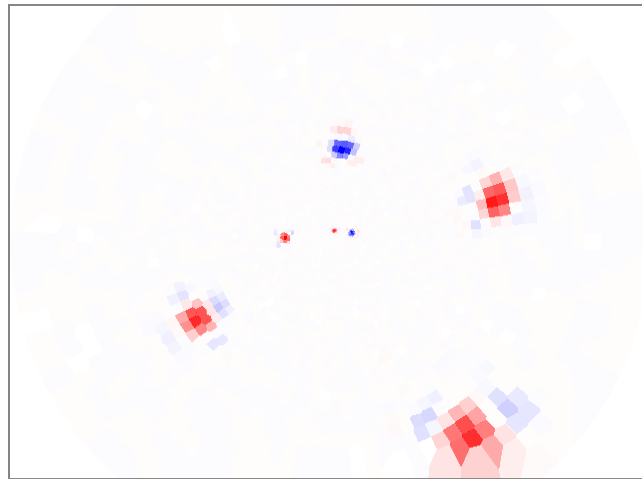


Fig. (3). Trained receptive fields of seven sample RGCs. Blue denotes positive weights, red negative weights. Five cells are OFF-center ON-surround, two are ON-center OFF-surround. The input image pixels have been concatenated according to the RGC distribution shown in Fig. 2, hence, the fields become larger toward the periphery.

Fig. (3) shows the receptive fields of some hidden layer's neurons that were trained from grey-scale images according to a generative model with a synaptic cost term. Their shape is of center-surround type (Mexican hat shape). Their sizes are increasing toward the periphery due to the pixel density sampling as in Eq. 2. Fig. (4) shows a natural image and its reconstruction from the activations of the hidden layer's neurons.

A model has also been considered in which the spatio-temporal receptive fields of ganglion cells adapt within a few seconds [31]. This fast adaptation optimizes their predictive coding to rapidly changing statistics in the environment.

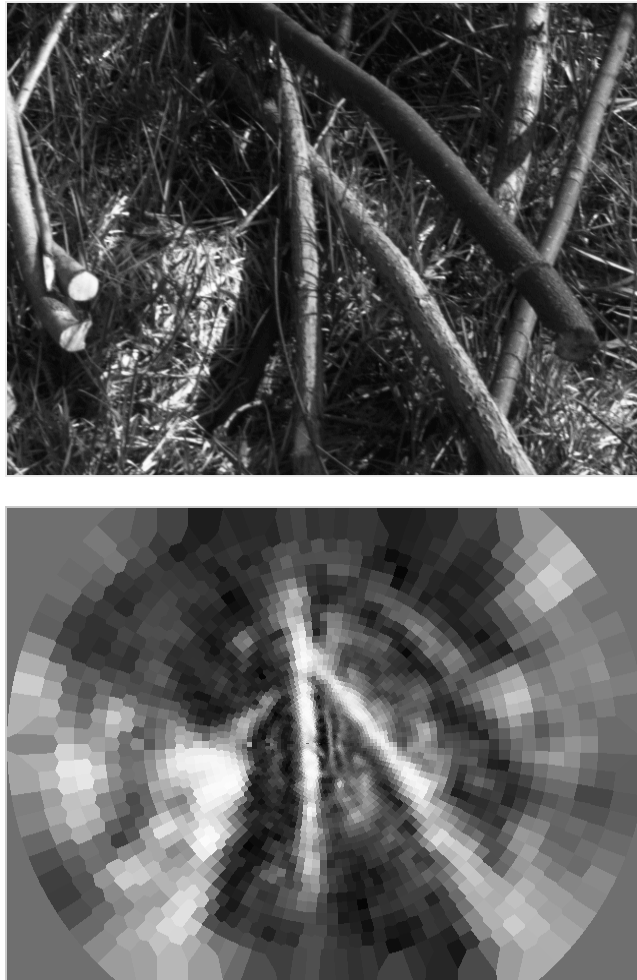


Fig. (4). Image reconstruction using foveated vision. Top: a grey-scale test image (640x480 pixel resolution). Bottom: its reconstruction by 28x28 trained RGC units, of which some are shown in Fig. (3).

The Variety of Visual Streams. Does the variety of visual streams, each filtering out specific information, fulfill some information theoretic goals? The implementation of the differing visual streams is largely due to the amacrine cells which come in a great variety of types, with differing receptive field sizes, integration times, and latencies. This means that many different kinds of stimulus correlation across space and time can be sensed and exploited for predictive/generative coding. Furthermore, generative models are mostly implemented with inhibitory feedback neurons to subtract the feedback from the original stimulus, with the difference used for optimizing the model.

Therefore, the intrinsic physiological properties of the amacrine cells largely determine the properties of the visual streams. Each amacrine cell type would then be capable of reconstructing one component of the visual stimulus.

RGCs of each visual stream specifically innervate their respective target areas in different parts of the brain. This suggests that the visual streams are genetically determined. On the other hand, there are arguments for dynamic interaction between RGCs that lead to their differentiation and to their use of all possible information channels: Developing ganglion cells start with a diffuse dendritic branching pattern and later segregate into distinguished types with restricted branching in upper or lower part of the amacrine layer. This segregation requires neural activity, supporting an active role of information processing during development [34].

3.3. Patents

Light-Sensor Chips. A patent from the research community that is biologically inspired describes a retina-like CCD light-sensor chip [35]. This chip's individual light-sensitive elements are arranged in a circular arrangement with a density of rings that decreases radially. This simple arrangement allows for easy drain of the electrical charges. It is roughly as shown in Fig. (2), but the elements are arranged radially (same number of elements in each ring), and they are arranged on a continuous surface rather than on an underlying grid.

A follow-up patent [67] describes a retina-like light-sensor chip (CCD, CMOS, MOS, etc.) with improved geometry of photo elements compared to patent [35]. The inner zone (corresponding to the fovea/foveola) has constant, highest resolution, while the outer rings have decreasing density, and the number of elements may vary between the rings.

Another patent is not closely inspired from the human retina's foveation [72]. It describes an electronic imaging device that places more resolution near the center of the image, to allow for high-quality digital zoom. The suggested arrangement of the photo elements is along a grid that either has narrower spacing near the center (horizontally and vertically) or that has edges that are curved toward the center to bring the photo elements closer together there. The geometric distortion is then corrected in the cropped digital image.

Neural Implants. Retinal implants (retinal prostheses) can bring some sight back to some blind people, but they are yet in their initial development. A patent [26] describes a retina implant which has small, closely spaced low power electrodes near the fovea, and larger widely spaced electrodes at the periphery. Other than maximizing resolution, this arrangement considers that the amount of electrical current required to stimulate the perception of light increases with distance from the fovea. Hence, larger electrodes are required to transfer the necessary current farther away from the fovea. The arrangement of electrodes is circular and radially, as suggested for the photoelements in patent [35].

Another patent [18] describes retinal implants as well as implants onto the primary visual cortex V1. It characterizes the density of the micro-contacts that decreases toward the periphery. One claim suggests a higher density in the area around the horizontal plane.

Image Compression. A method for compressing images that have a wide field of view and which are viewed center-focussed is described in patent [78]. A compression ratio of 1600:1 is achieved using a combination of a log-polar mapper, separate color- and contrast channels and further data compression. This reduces a required data transmission bandwidth. Possible applications are named as remote-controlled unmanned vehicles, micromanipulation in surgery or conducting repairs or inspections in spaces inaccessible to humans.

Another patent considers that specific regions in an image are of higher interest than others [36]. It exploits the progressive-resolution characteristics of wavelets for foveal vision. Wavelets are Gabor filter like localized wave functions that can be used at different scales (hierarchical levels) to create an image by superposition. Only few large-scale filter coefficients are required to represent an image at a low resolution, but many additional small-scale filter coefficients are required for a high resolution. These can be greatly reduced if given only for image regions that should be highly resolved, such as where the fovea is focused.

Moreover, a human's visual field is typically drawn to a focal point within a scene. This focal point can be purposely set by providing more detail there than elsewhere. Applications are for artists or for advertisers that wish to draw the viewer's attention to the item being advertised.

Interactive Displays. Foveation on a camera is no problem if it can be moved in synchrony with the eye movements of the person watching the camera images. A patent [66] describes a remote television viewing system for one viewer with the main elements of a camera, a display and an eye-tracker. While the viewer looks at the display, the eye-tracker monitors the viewer's line of sight and sends this information to the camera. Only that part of the display at which the viewer is looking is resolved at a high resolution, and hence, only the corresponding region in the field of view of the camera is sampled and transmitted at a high resolution. The remainder of the image is transmitted at a lower resolution level.

The realizations of "interactive remote vision" underlying this patent is contingent on sophisticated and affordable electronics devices that have since been developed. This patent is currently cited by 45 newer patents at freepatentsonline.com.

A more recent patent [57] describes augmentations to the data transfer and display described in patent [66]. This includes storage of the video data as digital files, data transfer by mobile phone networks and display by a virtual retinal display. It also suggests to predict the eye-position based on the upcoming video data. A follow-up patent [58] describes the necessary video stream compression dependent on the distance to the viewing focus.

Summary. Foveated vision, while abundant in biology, has not yet had a breakthrough in engineering applications. Retina-like light sensors and neural implants consider foveation faithfully. Image compression and an interesting idea for a digital zoom consider the general idea of using higher-resolved regions of interest. A challenge for foveated vision is its use in interactive displays. These may involve camera movements to record any desired point in a large angle visual field with highest resolution. Such integrated vision systems require fast camera / eye movements, an issue which we will discuss in the next section.

4. IMPLICATIONS

Foveated vision allows simultaneous recognition of a wide-angle scene and of object detail with minimal visual resources. Observing a wide field in low resolution, or a small field with high resolution, faces our brain with a similar amount of information. Problem solving algorithms use such multiresolution analysis to keep the problem size limited: only a part of the entire problem is resolved with high resolution at a time, embedded in a low-resolution representation of the rest. The focus moves until all parts are solved [51].

Saccades. To analyze an visual scene, foveated vision requires active vision that includes eye-movements, and concurrent planning and action. Our fast "ballistic" eye-movements are termed saccades. On average, we sample the visual scene with three saccades per second. In biology, their calibration is learned. Interestingly, foveation may be important for learning saccade calibration.

Saccade Learning. The geometry of saccades is adjustable: in intra-saccadic step experiments, a saccade target that is, say, at 10° to the right, is displaced to 8° during the saccade, without the participant noticing. Following some dozens of repetitions with consistent displacement, saccades aiming at 10° to the right will become shorter, moving the eyes only approximately 8° . This adaptation then happens for saccade targets around 10° right (the adaptation field), but not for other targets (e.g. saccades aiming at 10° to the left remain unchanged).

Saccade adaptation happens thus based on visual feedback, and foveation may play a role. We have proposed that the increased resolution near the fovea might give positive feedback for learning [76].

The variables on which saccadic adaptation depends have been studied intensively.

- In AI models in which the saccade metric is auto-adjusted, the geometric error of a saccade is often obtained by comparing the pre-saccadic target region of the image with the post-saccadic image center (e.g. [62]). Such a comparison of visual inputs at different retinal locations, before and after a saccade, however, might be implausible for biological learning. Foveation makes such a comparison difficult, because the same object looks very different whether it is in the periphery or in the fovea. Instead, a study indicates that correspondences between peripheral and foveal object appearances are learnt (and can be altered) [13]. This implies that saccades are functional before objects at different retinal locations can be compared.
- Since saccades are generally very unprecise, they are usually followed by small corrective saccades. It was shown that saccade adaptation occurs independently of these corrective saccades [44].
- Only the saccade target contributes to saccadic adaptation, but the background is not important. Hence, a stimulus near the fovea, but not in the periphery, is relevant [61].
- Saccade adaptation is similar for point targets and large targets [4].
- Saccades also adapt when the saccade error is always the same; i.e. they do not "explore" via trial and error, what would lead to better or worse performance. The distance by which the target is displaced matters: there is a maximal change of saccade gain (for adaptation toward smaller saccades) for a target at -4° away from the eye landing position [60]; however, these noisy data are also compatible with a constant gain change as well as with a gain change that becomes smaller for larger target deviations.
- The displaced target must be visible within 400msec after the saccade, to yield effective adaptation [4].
- We conjecture that the rate of saccade adaptation scales with the strength of the visual feedback signal; hence, we would predict that a low contrast of the target object would lead to smaller adaptation rates. Literature on this is scarce, possibly because adaptation rates differ considerably from experiment to experiment, and from individual to individual. Exponential fits showed rate constants varying two- to four-fold in identical conditions [73].
- A preliminary result shows that an acoustic reinforcement given, e.g. during those saccades that are too small, causes a decrease in saccade amplitude [37]. Electrically stimulating neurons in the basal ganglia immediately after saccades to a fixed direction, selectively facilitates saccades into that direction [30]. More generally, reinforcements support an action just performed.
- Generally, adaptation for reducing the amplitude of saccades (backward adaptation) is stronger than adaptation for increasing the amplitude (forward adaptation). Note that in experiments, saccades usually undershoot; possibly, because they are normally paired with small head movements that are blocked in experiments.
- The mechanisms for forward adaptation might differ from those responsible for backward adaptation. Forward adaptation is not only weaker, but unlike backward adaptation, does not transfer to antisaccade tasks [48].
- Circuits for vertical and horizontal saccades are distinct; adaptation of vertical changes has only been little explored [76].

A simple model of saccade adaptation exploits, for horizontal saccades, the geometry of visual system mapping [76]. The idea is that if the saccade is too short then the visual target remains in the same visual hemifield, while for too large saccades the target moves to the opposite hemifield. Assuming that the target lands near the fovea, the target alone

will determine whether the left or the right SC (looking at right and left hemifield, respectively) is more activated. A comparison of these activations yields the respective learning signal.

Given that the right SC is already connected with neurons controlling leftward saccades [42], a saccade is automatically elicited into the correct direction. Learning only needs to account for fine adjustments.

Siamese cats have an abnormally large representation of the central ipsilateral visual field in the SC, and they commonly squint [7]. The large ipsilateral representation in the SC shifts the normal contralateral representation to a more caudal SC position, i.e. corresponding to far periphery. The convergent strabismus may be a consequence of a saccade strategy that places the representation of the visual target on the anterior end of the SC, even if it then does not fall onto the fovea.

In healthy people, the point of fixation was found to be consistently displaced from the retinal area of highest cone density by around 50 micrometers (roughly 0.1° visual angle) into a certain direction [55].

Vergence Eye Movements. Learning vergence eye movements might also be rooted in saccade learning. Vergence movements direct the eyes to an object in depth: the eyes are inward directed if a focused object is close, and they are directed in parallel if an object is far. During learning, correct vergence movements might receive feedback if the left-eye and the right-eye view of an object co-align [21]. However, co-alignment could be defined at arbitrary relative eye positions, such as when squinting. Foveation might prevent this problem from occurring by providing a single point at which to co-align both eye's images. The mechanism of adjusting these vergence movements may resemble that of saccade learning, pointing both eyes independently at exactly the same point in space.

Object Localization, Grasping and Embodiment. Bringing an object to the fovea not only serves its identification. The gaze direction and the eyes' vergence deliver further information about the location of the object in space. In the motor cortex, some neurons respond to the sheer presentation of an object at a specific position [59]. These visual responses are tightly linked to grasping an object at the corresponding position. The alignment of visual representations with directed movement and touch forms part of "embodiment", the grounding of sensory experience in the real world.

Objects which are seen in the periphery, too, can be localized, even though less precisely. Taking into account both, gaze angle and peripherally perceived object position, the position in a head-centered or body-centered reference frame can be computed [77].

The result of such a frame of reference computation must be related to a position, as it is encoded in the motor areas of the brain. Such a relation can be learned by grasping at it. Alternatively, a saccade to it can relate its pre-saccadic peripheral position to the gaze direction when looking at it. The "where" pathway through areas in the dorsal visual cortex can thereby be trained, based on feedback due to foveation following successful saccades.

5. CONCLUSION

We have shown that foveated vision is extremely different from a photographic implementation. Despite the low peripheral resolution, our eyes give us an impression of a fully-resolved scene, the reason being that whenever we cannot identify peripheral detail, a saccade to the corresponding location will instantly give us that detail. Foveated vision does not deliver a photo for later inspection but delivers those details of a scene that are currently being asked for.

The extremely high compression factor indicates that foveation is not just fine tuning, but together with eye movements is essential to an efficient sensory-action, behavioral strategy of animals and humans.

Among individuals and species, eyes are not alike. Likewise, the four quadrants of the retina have their individual characteristics. The eyes are custom-tailored to the needs of the animal or human.

5. CURRENT & FUTURE DEVELOPMENTS

In contrast to animal or human vision, artificial vision systems mostly rely on rectangular, homogeneous input images. With the advance of intelligent machines that act autonomously, custom-tailored image systems are advisable. These need not be inspired from biology: omni-directional cameras are beneficial for robots that do not have a front and a rear side [2], such as football robots. Though, their 360° visual field limits their resolution, since with a high resolution they would produce overly large amounts of data. Hybrid solutions like focusing on small regions for further processing, possibly taken by a second camera, are reminiscent of a crude form of foveation.

There is a trend to make robots more human-like, so that we can better understand and interact with them. A

human-like geometry necessitates eye- and head movements, and encourages the use of foveated vision. Together with stereo vision, foveation yields a benefit of high precision depth estimation of the object in focus.

In summary, the step from homogeneous vision to foveated vision is large, and has not been seriously tackled by the artificial vision community, yet.

ACKNOWLEDGEMENTS

Mahtab Nazari, Sohrab Saeb, Hyundo Kim, Thomas Weisswange and Constantin Rothkopf supplied valuable feedback on the document. We acknowledge financial support by the European Union through projects FP6-2005-015803 and MEXT-CT-2006-042484 and by the Hertie Foundation.

REFERENCES

- [1] Abramov I, Gordon J, Hendrickson A, Hainline L, Dobson V, LaBossiere E. The retina of the newborn human infant. *Science* 1982; 217 (4556): 265-267.
- [2] Adorni G, Cagnoni S, Mordonini M, Sgorbissa A. Omnidirectional stereo systems for robot navigation. In Proc. 2003 omnivis03.
- [3] Aggelopoulos N, Rolls ET. Scene perception: inferior temporal cortex neurons encode the positions of different objects in the scene. *Europ J Neurosci* 2005; 22: 2903-2916.
- [4] Bahcall D, Kowler, E. The control of saccadic adaptation: implications for the scanning of natural visual scenes. *Vision Res* 2000; 40, 277996.
- [5] Balasuriya S, Siebert, P. A biologically inspired computational vision front-end based on a self-organised pseudo-randomly tessellated artificial retina. In Proc. 2005 International Joint Conference on Neural Networks (p. 3069-74).
- [6] Barnes N, Sandini G. Direction control for an active docking behaviour based on the rotational component of log-polar optic flow. In Proc. 2000; Europ Conf on Computer Vision, vol. 2 (p. 167-81).
- [7] Berman N, Cynader M. Comparison of receptive-field organization of the superior colliculus in siamese and normal cats. *J Physiol* 1972; 224 (2): 363-389.
- [8] Berson D, Dunn F, Takao M. Phototransduction by retinal ganglion cells that set the circadian clock. *Science* 2002; 295 (5557): 1070-1073.
- [9] Bonmassar G, Schwartz E. Fourier analysis and cortical architectures: the exponential chirp transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1997; 19 (10): 1080-1089.
- [10] Borghuis B, Ratliff C, Smith R, Sterling P, Balasubramanian, V. Design of a neuronal array. *J Neurosci* 2008; 28 (12): 3178-3189.
- [11] Buzas P, Blessing E, Szmajda B, Martin P. Specificity of M and L cone inputs to receptive fields in the parvocellular pathway: Random wiring with functional bias. *J Neurosci* 2006; 26 (43): 11148-11161.
- [12] Carceri S, Jacobs A, Nirenberg S. Classification of retinal ganglion cells: A statistical approach. *J Neurophysiol* 2003; 90: 1704-1713.
- [13] Cox D, Meier P, Oertelt N, DiCarlo J. 'Breaking' position invariant object recognition. *Nature Neurosci* 2005; 8 (9): 1145-1147.
- [14] Curcio CA, Sloan KR, Kalina RE, Hendrickson AE. Human photoreceptor topography. *J Comp Neurol* 1990; 292: 497-523.
- [15] Curcio CA, Sloan KR, Packer O, Hendrickson AE, Kalina RE. Distribution of cones in human and monkey retina: individual variability and radial asymmetry. *Science* 1987; 236: 579-582.
- [16] Davenport C, Detwiler P, Dacey D. Effects of pH buffering on horizontal and ganglion cell light responses in primate retina: Evidence for the proton hypothesis of surround formation. *J Neurosci* 2008; 28 (2): 456-464.
- [17] Deering, M. A photon accurate model of the human eye. In *Acm transactions on graphics* 2005; (p. 649-58).
- [18] Eckmiller, R, Suchert, S. Microcontact structure for implantation in a mammal, especially a human being. European Patent No. EP 1712253 (2006).
- [19] Felleman, D, Van Essen, D. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* 1991; 1: 1-47.
- [20] Field, G, Chichilnisky, E. Information processing in the primate retina: Circuitry and coding. *Annu Rev Neurosci* 2007; 30: 1-30.
- [21] Franz, A, Triesch, J. Emergence of disparity tuning during the development of vergence eye movements. In Proc. 2007 6th IEEE International Conference on Development and Learning.
- [22] Fukuda Y, Iwama K. Visual receptive-field properties of single cells in the rat superior colliculus. *Jpn J Physiol* 1978; 28 (3): 385-400.
- [23] Gonzalez-Hoyuela M, Barbas JA, Rodriguez-Tebar, A. The autoregulation of retinal ganglion cell number. *Development* 2001; 128 (1): 117-124.
- [24] Goodchild A, Ghosh K, Martin P. Comparison of photoreceptor spatial density and ganglion cell morphology in the retina of human, macaque monkey, cat, and the marmoset callithrix jacchus. *J Comp Neurol* 1996; 366: 55-75.

- [25] Goodhill G, Xu J. The development of retinotectal maps: A review of models based on molecular gradients. *Network: Computation in Neural Systems* 2005; 16 (1): 5-34.
- [26] Greenberg R, Williamson R, Humayan M. Variable pitch electrode array. US Patent No. 7149586 (2006).
- [27] Hammarrenger B, Lepore F, Lippe S, Labrosse M, Guillemot J, Roy M. Magnocellular and parvocellular developmental course in infants during the first year of life. *Documenta Ophthalmologica* 2003; 107 (3): 225-233.
- [28] Harada T, Harada C, Parada L. Molecular regulation of visual system development: more than meets the eye. *Genes & Development* 2007; 21: 367-378.
- [29] Hendrickson A, Drucker D. The development of parafoveal and mid-peripheral human retina. *Behavioural Brain Research* 1992; 49 (1): 21-31.
- [30] Hikosaka O, Nakamura K, Nakahara H. Basal ganglia orient eyes to reward. *J Neurophysiol* 2006; 95 (2): 567-584.
- [31] Hosoya T, Baccus S, Meister M. Dynamic predictive coding by the retina. *Nature* 2005; 436: 71-7.
- [32] Kenyon G, Marshak D. Gap junctions with amacrine cells provide a feedback pathway for ganglion cells within the retina. *Proc Biol Sci* 1998; 265 (1399): 919-925.
- [33] Kolb H. How the retina works. *American Scientist* 2003; 91: 28-35.
- [34] Kolb H, Fernandez E, Nelson R, Jones B. <http://webvision.med.utah.edu> 2008; Web page.
- [35] Kreider G, Claeys C, Debusschere I, Sandini G, Dario P, Tagliascio V. Radiation-sensitive sensor having a plurality of radiation sensitive elements arranged substantially circular with radially decreasing density. US Patent No. 5166511 (1992).
- [36] Kurapati K. Hierarchical foveation based on wavelets. US Patent No. 6535644 (2003).
- [37] Madelain L, Paeye C, Wallman J. Saccade adaptation: reinforcement can drive motor adaptation. In Proc. 2008 8th annual meeting of the vision sciences society (p. 234).
- [38] Martinez-Conde S, Macknik S, Hubel D. Microsaccadic eye movements and firing of single cells in the striate cortex of macaque monkeys. *Nature Neurosci* 2000; 3: 251-258.
- [39] Masland R. The fundamental plan of the retina. *Nature Neurosci* 2001; 4 (9): 877-86.
- [40] McLoon S. A monoclonal antibody that distinguishes between temporal and nasal retinal axons. *J Neurosci* 1991; 11(5): 1470-1477.
- [41] Monnier P, Sierra A, Macchi P, Deitinghoff L, Andersen J, Mann M, et al. RGM is a repulsive guidance molecule for retinal axons. *Nature* 2002; 419: 392-395.
- [42] Moschovakis A, Kitama T, Dalezios Y, Petit J, Brandi A, Grantyn A. An anatomical substrate for the spatiotemporal transformation. *J Neurosci* 1998; 18 (23): 10219-10229.
- [43] Nirenberg S, Meister M. The light response of retinal ganglion cells is truncated by a displaced amacrine circuit. *Neuron* 1997; 18: 637-650.
- [44] Noto C, Robinson F. Visual error is the stimulus for saccade gain adaptation. *Cog Brain Res* 2001; 12: 301-305.
- [45] Olshausen B. Learning linear, sparse, factorial codes. A.I. Memo No. 1580, 1996; Massachusetts Institute of Technology.
- [46] Olshausen B, Field D. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Res* 1997; 37: 3311-3325.
- [47] Osterberg G. Topography of the layer of rods and cones in the human retina. *Acta Ophthalmol Suppl* 1935; 6: 1-102.
- [48] Panouilleres M, Cotti J, Guillaum, A, Urquizar C, Salemm R, Munoz D, et al. Adaptation of saccadic eye movements: behavioral evidence for different mechanisms controlling saccade amplitude lengthening and shortening. In Proc. 2008 8th annual meeting of the vision sciences society (p. 234).
- [49] Perry V, Cowey A. Retinal ganglion cells that project to the superior colliculus and pretectum in the macaque monkey. *Neuroscience* 1984; 12 (4): 1125-1137.
- [50] Petrusca D, Grivich M, Sher A, Field G, Gauthier J, Greschner M, et al. Identification and characterization of a Y-like primate retinal ganglion cell type. *J Neurosci* 2007; 27 (41): 11019-11027.
- [51] Pizlo Z, Stefanov E, Saalweachter J, Li Z, Haxhimusa Y, Kropatsch W. Traveling salesman problem: a foveating pyramid model. *J Problem Solving* 2006; 1: 83-101.
- [52] Polyak S. *The retina*. 1941; Chicago: University of Chicago Press.
- [53] Polyak S. *The vertebrate visual system*. 1957; Chicago: University of Chicago Press.
- [54] Purves D, Augustine G, Katz L, LaMantia A, McNamara J, Williams S. (Eds.). 2001; *Neuroscience*. Sinauer Associates.
- [55] Putnam N, Hofer H, Doble N, Chen L, Carroll J, Williams D. The locus of fixation and the foveal cone mosaic. *J Vision* 2005; 5 (7): 632-639.
- [56] Ratliff C, Kao Y, Sterling P, Balasubramanian V. Retinal ganglion cell arrays are structured to process the excess of dark information in natural scenes. *Neuron* 2008; submitted.
- [57] Ritter R, Lauper E. Transmission and display of video data. European Patent No. EP 1186148 (2002).
- [58] Ritter R, Lauper E. Method and device for transmission of video data using line of sight - eye tracking - based compression. European Patent EP1720357 (2006).
- [59] Rizzolatti G, Luppino G. The cortical motor system. *Neuron* 2001; 31: 889-901.

- [60] Robinson F, Noto C, Bevans S. Effect of visual error size on saccade adaptation in monkey. *J Neurophysiol* 2003; 90: 1235-1244.
- [61] Robinson F, Noto C, Watanabe S. Effect of visual background on saccade adaptation in monkeys. *Vision Res* 2000; 40 (17): 2359-2367.
- [62] Rodemann T, Joublin F, Korner E. Saccade adaptation on a 2 dof camera head. In Proc. 2004 Selforganization and Adaptive Behaviour (p. 94-103).
- [63] Rodieck R. Comparative primate biology: Neurosciences. In 1988 (Vol. 4, p. 203-78). New York: Alan R. Liss: Steklis, H.D. and Erwin, J.
- [64] Rovamo J, Virsu V. An estimation and application of the human cortical magnification factor. *Exp. Brain Res* 1979; 37: 495-510.
- [65] Rowe M. Trichromatic color vision in primates. *News Physiol Sci* 2002; 17: 93-98.
- [66] Ruoff Jr C. Retinally stabilized differential resolution television display. US Patent No. 4513317 (1985).
- [67] Sandini G, Questa P, Scheffer D. Constant resolution and space variant sensor array. US Patent No. 7009645 (2006).
- [68] Sandini G, Questa P, Scheffer D, Mannucci A. A retina-like CMOS sensor and its applications. In Proc. 2000 IEEE sensor array and multichannel signal processing workshop.
- [69] Schein SJ. Anatomy of macaque fovea and spatial densities of neurons in foveal representation. *J Comp Neurol* 1988; 269: 479-505.
- [70] Schiller P, Malpeli J. Functional specificity of lateral geniculate nucleus laminae of the rhesus monkey. *J Neurophysiol* 1978; 41 (3): 788-797.
- [71] Schwartz E. Spatial mapping in the primate sensory projection: Analytic structure and relevance to perception. *Biol Cybern* 1977; 25 (4): 181-194.
- [72] Stavely D. Apparatus and method for improved-resolution digital zoom in an electronic imaging device. US Patent No. 7227573 (2007).
- [73] Straube A, Fuchs A, Usher S, Robinson F. Characteristics of saccadic gain adaptation in rhesus macaques. *J Neurophysiol* 1997; 77 (2): 874-895.
- [74] Szel A, Rohlich P, Caffè A, Juliusson B, Aguirre G, Van Veen T. Unique topographic separation of two spectral classes of cones in the mouse retina. *J Comp Neurol* 1992; 325 (3): 327-342.
- [75] Vincent B, Baddeley R, Troscianko T, Gilchrist I. Is the early visual system optimised to be energy efficient? *Network* 2005; 16 (2-3): 175-190.
- [76] Weber C, Triesch J. A possible representation of reward in the learning of saccades. In Proc. 2006 EpiRob (p. 153-160).
- [77] Weber C, Wermter S. A self-organizing map of Sigma-Pi units. *Neurocomputing* 2007; 70 (13-15): 2552-60.
- [78] Weiman C, Evans Jr J. Digital image compression employing a resolution gradient. US Patent No. 5103306 (1992).
- [79] Werblin F. Regenerative amacrine cell depolarization and formation of on-off ganglion cell response. *J Physiol* 1977; 264 (3): 767-785.
- [80] Young M, Lund R. The retinal ganglion cells that drive the pupilloconstrictor response in rats. *Brain Research* 1998; 787 (2): 191-202.
- [81] Zrenner E. Central and peripheral mechanisms of colour vision. In 1985 (p. 165-82). New York: Macmillan.