

Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics

GERHARD KRIEGER¹, INGO RENTSCHLER¹, GERT HAUSKE²,
KERSTIN SCHILL¹ and CHRISTOPH ZETZSCHE¹

¹*Institut für Medizinische Psychologie, Ludwig-Maximilians-Universität, 80336 München,
Goethestr. 31, Germany*

²*Lehrstuhl für Nachrichtentechnik, Technische Universität München, 80290 München,
Arcisstr. 21, Germany*

Received 23 August 1999; revised 22 March 2000; accepted 28 March 2000

Abstract—Based on an information theoretical approach, we investigate feature selection processes in saccadic object and scene analysis. Saccadic eye movements of human observers are recorded for a variety of natural and artificial test images. These experimental data are used for a statistical evaluation of the fixated image regions. Analysis of second-order statistics indicates that regions with higher spatial variance have a higher probability to be fixated, but no significant differences beyond these variance effects could be found at the level of power spectra. By contrast, an investigation with higher-order statistics, as reflected in the bispectral density, yielded clear structural differences between the image regions selected by saccadic eye movements as opposed to regions selected by a random process. These results indicate that nonredundant, intrinsically two-dimensional image features like curved lines and edges, occlusions, isolated spots, etc. play an important role in the saccadic selection process which must be integrated with top-down knowledge to fully predict object and scene analysis by human observers.

1. INTRODUCTION

The analysis of complex objects or scenes by a sequence of rapid saccadic eye movements is a prototypical example of the efficiency biological agents can achieve with respect to sensorimotor information processing. In general, this efficiency can be regarded to result from an evolutionary adaptation to the regularities of the complex terrestrial environment. A well known example for such an adaptation is the structure of receptive fields along the visual pathway. During recent years, a steadily increasing amount of evidence has shown that basic properties like lateral inhibition (van Hateren, 1992) and orientation selectivity (Olshausen and Field, 1996), as well as more sophisticated mechanisms like cortical gain control (Zetsche

et al., 1999), can all be understood as an adaptation of the neural connectivity to the environmental structure (for review see Atick (1992); Zetzsche and Krieger (1999)).

While adaptation takes place at various structural levels and on different time scales, there is a common underlying principle: the optimized allocation of a limited amount of neural processing resources to environmental data and behavioral tasks of nearly unlimited complexity. This principle is also the basis of the efficient selection strategy of the saccadic system. Since it would not be affordable for any reasonable system to process its complete optical surrounding with maximum spatial resolution, the system uses inhomogeneous spatial sampling. The majority of the information processing machinery is concentrated into a narrow range of a few degree of the visual field, the *fovea centralis*, whereas the remainder, orders of magnitude larger in area, is served by a much coarser representation (Dowling, 1987) with limited processing power (Rentschler and Treutwein, 1985; Strasburger *et al.*, 1994). However, such a strategy can only work without the risk of a complete miss of important information, if there are efficient mechanisms available which can rapidly redirect the foveal region to the important locations within the visual scene (cf. also Rao and Ballard (1995)). This is an example for the allocation of limited resources on a short time scale. The saccadic system can allocate the foveal processing resources within parts of a second by rapid ballistic movements with velocities of up to 700 deg/s. However, while many details of the saccadic system, like the timing parameters, the motoric control, or the reflexive mechanisms for the selection of transient targets, are quite well understood today, the powerful strategies by which the system manages to select the relevant parts out of complex scenes are still a challenging research problem.

We try to tackle this problem by a multi-disciplinary approach (Zetzsche *et al.*, 1998), which involves participants with expertise in experimental psychology, visual psychophysics and modelling, statistical signal processing, and knowledge based system design. By this, we cover the research areas which are, in our view, necessary for the analysis, understanding, and modelling of a complete sensorimotor loop like the saccadic eye movement system. In this paper we concentrate on the analysis of the bottom-up aspects of this system. The final goal of the project is the provision of a model which can take natural images as input and derives predictions about the fixations of human observers on these images.

For our investigations, we have first obtained an empirical database by recording the eye movements of human observers on a variety of artificial and natural images (Section 2). From these data, we want to derive suggestions about the processing mechanisms in early vision which are involved in the signalling of those 'salient' image locations which can be selected by saccadic fixations. Central for our research and modelling is statistical information theory. Hence, we investigate the differences between the statistical properties of those image regions which were selected by fixations, as opposed to the average statistics of randomly selected

regions (Section 3). The results of this statistical analysis are then used to derive a suggestion for a specific kind of nonlinear 'saliency-detector' in early vision (Section 4).

2. METHODS

Images have been chosen from seven different topical classes (portraits, landscapes, technics, architecture, computer graphics, polygons, and textured polygons) to enable a differentiated analysis, and to avoid any bias on the relevant image features. Every class contained 7 images, hence a total number of 49 images has been investigated. All images were grey-level images of size 512×512 .

Eleven paid subjects with a mean age of 27 years participated in the experiment. They were naive with respect to the objectives of the study. The subjects were instructed to view each image as carefully as possible, to be able to perform subsequent tests with the images. The subjects were first shown a subset of 10 randomly selected images, which served as training images, and recordings were evaluated separately for these images. After a short pause and a recalibration, the main experiment was performed with the 49 images of the set. Each image was presented to the subject for five seconds, and eye movements were recorded during the whole presentation.

The experiments took place in a moderately illuminated room. Stimuli were presented on a 21 inch video monitor (CONRAC 7550 C21) with 1024×768 pixels spatial resolution at a frame rate of 100 Hz. Screen background luminance was set to 3 cd/m^2 ; the luminance of the stimuli was 25 cd/m^2 . Subjects viewed the screen binocularly from a distance of 80 cm. The total image size was 18×18 deg. Head movements were restricted by a biteboard and a forehead rest. Eye movements were measured with an SRI Generation 5.5 Purkinje-image eyetracker (Crane and Steele, 1985) sampled at 400 Hz. Its frequency response is better than 250 Hz with a noise level equivalent to about 20 arc seconds rms. The eyetracker can follow saccades of 15 degrees or more without losing the eye. Each session started with a calibration procedure: the subject sequentially fixated 9 positions arranged on a circular array of 8 degrees radius. The eyetracker behaved linearly within this range. Static accuracy of the eyetracker was better than 0.1 degrees. Experiments were controlled by a 486 PC, which also performed automatic off-line analysis of the eye movement data in which saccadic latencies and saccade start and end positions were determined.

Small regions around the fixation points of the subjects were extracted for statistical analysis (cf. Fig. 1). Several types of statistics were computed for these regions and compared to the 'average' statistics, as obtained from randomly selected regions. Differences in the statistics should indicate which image features are preferred in the selection process.

3. RESULTS

From Fig. 1 it becomes apparent that the locations of saccadic eye movements are not randomly distributed across natural scenes. However, it is also evident that even if the saccadic selection process would be completely deterministic, there is a high degree of variability in the data. Because of this, and since we assume that statistical redundancies of the images have an important influence on the regions selected for fixation, we have investigated the statistical properties of the fixated image regions. For our analysis, we used the 7 images shown in Fig. 1.

3.1. Analysis of spatial variance

In a first step, we measured the signal variability σ_{eye}^2 of fixated image regions and compared it to the variance σ_{rand}^2 of regions, which had been chosen by a random number generator:

$$\frac{\sigma_{\text{eye}}^2}{\sigma_{\text{rand}}^2} = \frac{\sum_{\vec{x}_{\text{eye}}} \sum_{\vec{x} \in W} [u(\vec{x}_{\text{eye}} - \vec{x}) - \bar{u}]^2}{\sum_{\vec{x}_{\text{rand}}} \sum_{\vec{x} \in W} [u(\vec{x}_{\text{rand}} - \vec{x}) - \bar{u}]^2}. \quad (1)$$

Here $u(\vec{x})$ is the image intensity at \vec{x} and \bar{u} is the mean intensity value within a local window W , which in our analysis spanned 1/8 of the total image size. The vectors \vec{x}_{eye} and \vec{x}_{rand} indicate the center of fixated and randomly selected image regions, respectively.

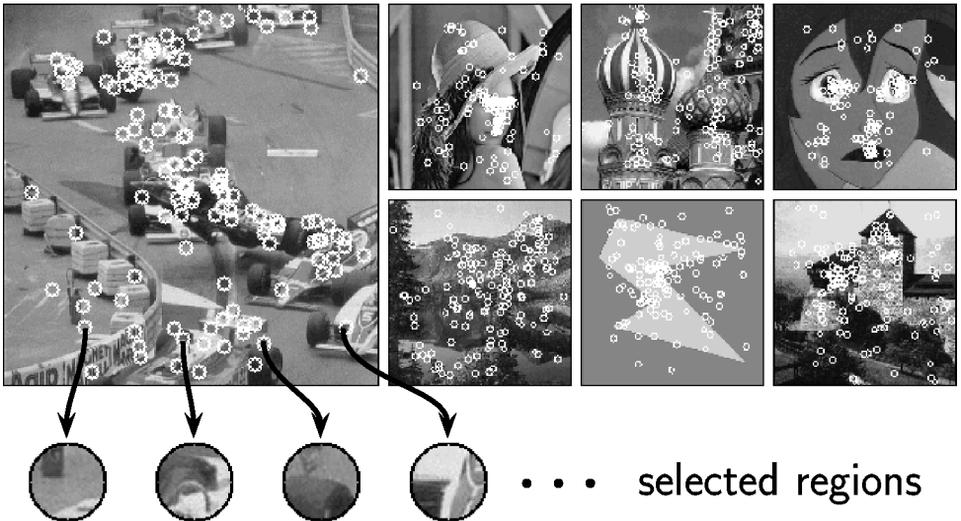


Figure 1. Test images used in our experiments. The small circles indicate positions of saccadic fixations.

As can be seen from Table 1, for most images the regions selected by eye movements had a significantly higher amount of signal variability. The only exception was the lake scene of Fig. 1, where the variance was approximately equal for both the randomly selected and the fixated regions. For the whole set of images we obtained a variance ratio of $\sigma_{\text{eye}}^2/\sigma_{\text{rand}}^2 = 1.35$.

3.2. Second-order statistics

The analysis of spatial variance is only a very crude approximation of the statistical dependencies within a local window. A more detailed characterization may be obtained by the autocorrelation function or its frequency domain counterpart, the power spectrum. In Fig. 2 we compare the power spectra of fixated and randomly chosen image regions. It turns out that the power spectra do not substantially differ in their basic shape. Similar results have been reported in (Reinagl and Zador, 1999), which also found a strong influence of the local contrast (variance) but only a small reduction of the second-order correlations in the fixated patches. Second-order statistics are hence insufficient for a clear-cut characterization of the structural properties of the fixated image regions. However, since there is evidence that the selection process in saccadic scene analysis cannot be reduced to mere variance selection, we have considered higher-order statistics of the fixated regions.

3.3. Higher-order statistics

One measure for higher-order dependencies is provided by the third-order cumulant c_3^u , which basically calculates the expectation of an image which has been multiplied by two shifted copies of itself. For a zero-mean stationary random process $\{u(\vec{x})\}$, the third-order cumulant c_3^u is defined as:

$$c_3^u(\vec{x}_1, \vec{x}_2) = E[u(\vec{x}) \cdot u(\vec{x} + \vec{x}_1) \cdot u(\vec{x} + \vec{x}_2)]. \quad (2)$$

As for autocorrelations, higher-order statistics are often more conveniently investigated in the frequency domain. The Fourier transform of c_3^u is known as the bispectrum C_3^U :

$$C_3^U(\vec{f}_1, \vec{f}_2) = \mathcal{F}\{c_3^u(\vec{x}_1, \vec{x}_2)\}. \quad (3)$$

Table 1.
Analysis of spatial variance

	Monaco	Lena	Moscow	Tears	Lake	Poly	Castle	Σ
N_{fix}	125	136	131	139	159	140	141	971
σ_{eye}^2	1879	1598	2065	2640	930	716	2586	1753
σ_{rand}^2	1400	1438	1812	1059	957	512	2002	1299
$\sigma_{\text{eye}}^2/\sigma_{\text{rand}}^2$	1.34	1.11	1.14	2.49	0.97	1.40	1.29	1.35

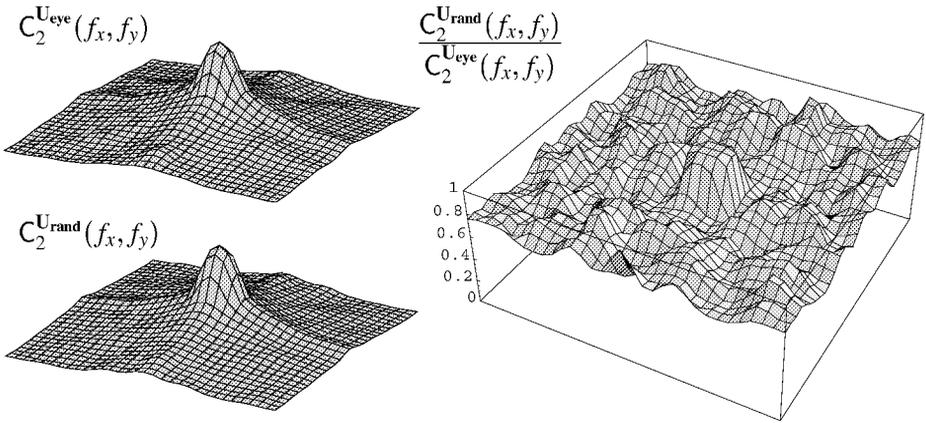


Figure 2. Power spectra of image regions selected by eye fixations (upper left) and by a random number generator (lower left). The two power spectra do not substantially differ in their basic shape. Their ratio (right) is consistently ≤ 1 , which is in agreement with our analysis of the variance. For low frequencies the ratio is close to one, thus indicating an approximately equal contribution in both cases. For higher spatial frequencies fixated image regions contribute more power than randomly chosen regions. This is consistent with a slightly steeper decay of the autocorrelation function as reported in Reinagl and Zador (1999).

Alternatively, the Fourier-Stieltjes representation of $\{u(\vec{x})\}$ offers the possibility to express the bispectrum directly in terms of the frequency components $dU(\vec{f})$ (Nikias and Petropulu, 1993):

$$E[dU(\vec{f}_1) \cdot dU(\vec{f}_2) \cdot dU^*(\vec{f}_3)] = \begin{cases} C_3^U(\vec{f}_1, \vec{f}_2) \cdot d\vec{f}_1 d\vec{f}_2 & \text{if } \vec{f}_3 = \vec{f}_1 + \vec{f}_2 \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

From this equation it becomes apparent that the bispectrum measures the statistical dependencies between three frequency components, the sum of which equals zero. A direct computation in the frequency domain can also be derived from this notation (Nikias and Petropulu, 1993).

In Fig. 3 we compare the bispectra of fixated image regions to those of randomly chosen areas. Unlike the second-order spectra of Fig. 2, the two bispectra differ in their basic shape: For the randomly chosen regions we obtain a star-like shape of the bispectrum, which indicates strong statistical dependencies between frequency components aligned to each other with respect to orientation. This is consistent with our systematic investigations of the bispectra of a large set of natural images (Krieger *et al.*, 1997). In contrast, the bispectrum of fixated regions exhibits sections with more circular structures, thus indicating a strong simultaneous contribution from frequency components of different orientations within the local windows.

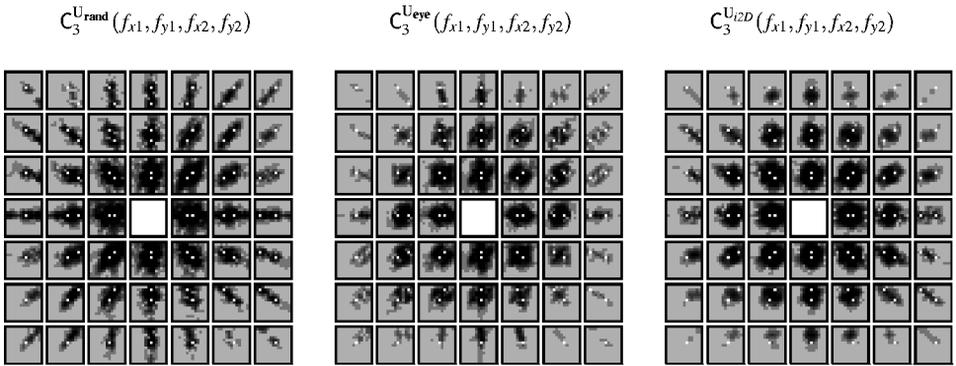


Figure 3. Bispectra of image regions selected either by a random number generator (left), by saccadic eye movements (middle), or by an *i2D*-operator (right). Shown are several sections of the bispectral magnitude $|C_3^{\text{U}}([f_{x1}, f_{y1}], [f_{x2}, f_{y2}])|$ with $(f_{x1}, f_{y1}) = \text{const.}$ (cf. the coordinate axes depicted in Fig. 4). For the random sample (C_3^{Urand} , left), the bispectral ‘energy’ is concentrated in those regions where the frequency components are aligned to each other with respect to orientation. In contrast, the bispectrum of fixated image regions (C_3^{Ueye} , middle) shows an ‘energy’ distribution which is more circular around the void points $[f_{x1}, f_{y1}] = [0, 0]$, $[f_{x2}, f_{y2}] = [0, 0]$ and $[f_{x1} + f_{x2}, f_{y1} + f_{y2}] = [0, 0]$. This indicates that the eye (as opposed to the random number generator) selects more often image regions with strong statistical dependencies between frequency components of *different* orientation. Shown on the right is the bispectrum C_3^{UizD} of curved image regions, which have been extracted from the seven images depicted in Fig. 1 by *i2D*-selective operators tuned to high curvatures (cf. Section 4.3). The image regions selected by *i2D*-operators show also the more circular bispectral ‘energy’-distribution found for fixated image regions.

4. DISCUSSION

4.1. Interpretation of the statistical results

How can we interpret these results? First, the findings of Section 3.1 indicate that the saccadic selection system avoids the fixation of image regions with little structural content (low variance), such as regions with approximately constant luminance. Similar results have been reported by (Mannan *et al.*, 1996; Reinagl and Zador, 1999). This fits with the basic operation of lateral inhibition in low level vision, since neurons with center-surround antagonism suppress the response to constant regions and enhance the response to varying signals like edges, lines etc. Such a basic distinction of structure vs. non-structure cannot differentiate between different types of structure, however. We thus applied the standard second-order analysis of the data via power spectra, but could not obtain a substantial difference between fixated and randomly selected regions. However, we were able to find significant differences in the higher-order statistics. The statistical structure of image regions is dominated by the presence of statistical dependencies between frequency components with aligned orientations, and this is consistent with the frequent occurrence of straight edges or lines. The bispectra of fixated

image regions show a different pattern. Here the bispectral ‘energy’ is not concentrated on frequency components with aligned orientations, but is more or less evenly distributed on the possible combinations. This implies that the saccadic selection system avoids image regions, which are dominated by a single oriented structure. Instead it selects regions containing different orientations, like occlusions, corners, etc.

4.2. The concept of intrinsic dimensionality

These statistical observations fit well with the idea of an information-theoretic hierarchy in terms of *intrinsic dimensionality* (Zetzsche and Barth, 1990a, b; Zetzsche *et al.*, 1993). The concept of intrinsic dimensionality relates the degrees of freedom provided by a signal domain to the degree of freedom actually used by a given signal and thus provides a hierarchy of local image signals in terms of different degrees of redundancy:

i0D-signals are constant, i.e. $u(x, y) = \text{const}$ within a local window.

i1D-signals can locally be approximated by a function of only one variable, i.e. $u(x, y) = u(ax + by)$. Examples are straight lines and edges. Sinusoidal gratings, the eigenfunctions of linear systems, are also members of this class.

i2D-signals are neither *i0D* nor *i1D*. Examples are corners, junctions, curved lines and edges, etc.

This hierarchy seems well suited for the description of saccadic eye movements. First, *i0D*-signals are least frequently fixated, as indicated by the results on the higher variance of the fixated regions. Second, *i2D*-signals are more attractive targets than *i1D*-signals. An early report (Mackworth and Morandi, 1967) suggested that the ‘gaze selects informative details’, and that the most informative details seem to be curved contours (*i2D*-signals). Similar results have been obtained by Bozkov *et al.* (1982) and Zetzsche and Deubel (1990) for fixations on simple polygons. In such figures the corners (*i2D*-signals) are more often fixated than the straight contour lines (*i1D*-signals), or the interior regions (*i0D*-signals). The special importance of *i2D*-signals has also been emphasized in the context of redundancy reduction (Attneave, 1954) and visual object recognition (Biedermann, 1987). Furthermore, the least redundant *i2D*-information is basically sufficient for a reconstruction of the original image signal (Barth *et al.*, 1993).

The hierarchy of intrinsic dimensionality is related to the types of neural operations occurring in early vision, which can be seen as the result of an efficient adaptation to the statistical properties of the natural environment. Evolution caused not only the development of linear isotropic and orientation-selective neurons, but also of nonlinear *i2D*-selective cells known as ‘hypercomplex’ or ‘end-stopped’. In primates, about half of the neurons in area V2 of the visual cortex exhibit such a preference to *i2D*-stimuli (Orban, 1984).

All this suggests that *i2D*-signals provide significant information about complex scenes and that *i2D*-selective detectors may be important members of the class of ‘saliency detectors’. However, unlike classical filters such operators are highly non-linear devices thus requiring special techniques for their modelling and functional analysis (Zetsche and Barth, 1990a).

4.3. *i2D*-selective operators

Suitable *i2D*-selective operators can be developed by a $\Sigma\Pi$ -approach (Zetsche and Barth, 1990a, b) or by the Volterra-Wiener expansion of nonlinear functionals (Krieger and Zetsche, 1996). The Volterra series relates the input $u_1(\vec{x})$ of a nonlinear, shift invariant system to its output $u_2(\vec{x})$ in the following way (Reichardt and Poggio, 1981; Schetzen, 1989):

$$u_2(\vec{x}) = h_0 + \int h_1(\vec{x}_1) \cdot u_1(\vec{x} - \vec{x}_1) \cdot d\vec{x}_1 + \iint h_2(\vec{x}_1, \vec{x}_2) \cdot u_1(\vec{x} - \vec{x}_1) \cdot u_1(\vec{x} - \vec{x}_2) \cdot d\vec{x}_1 d\vec{x}_2 + \dots \quad (5)$$

The quadratic part of equation (5) may be expressed in the frequency domain as

$$U_2(\vec{f}) = \int \tilde{U}_2(\vec{f}_1, \vec{f} - \vec{f}_1) \cdot d\vec{f}_1, \quad (6)$$

where

$$\tilde{U}_2(\vec{f}_1, \vec{f}_2) = H_2(\vec{f}_1, \vec{f}_2) \cdot U_1(\vec{f}_1) \cdot U_1(\vec{f}_2), \quad (7)$$

is the expanded output spectrum and $H_2(\vec{f}_1, \vec{f}_2)$ is the Fourier transform of the second-order Volterra kernel $h_2(\vec{x}_1, \vec{x}_2)$. Note that equation (7) may be regarded as the weighting of an AND-like conjunction between frequency components. A necessary and sufficient condition for a quadratic Volterra operator to be insensitive to *i0D*- and *i1D*-signals is given by Krieger and Zetsche (1996):

$$H_2(f_{x1}, f_{y1}, f_{x2}, f_{y2}) = 0 \quad \forall \quad f_{x1} \cdot f_{y2} = f_{y1} \cdot f_{x2}. \quad (8)$$

Systems constrained by this condition block aligned frequency components (see Fig. 4, an example is provided in Fig. 5). Thus they will yield the required selectivity as indicated by the polyspectral measurements (Fig. 3).

In order to test the predictions derived from the functional properties of *i2D*-operators, we have also computed the bispectra of image regions selected by this class of nonlinear curvature-selective saliency detectors. As can be seen from Fig. 3 (right), the image regions extracted by *i2D*-operators show strong statistical dependencies between frequency components of different orientations, resulting in a ‘circular’ bispectrum $C_3^{U_{i2D}}$ which is similar to $C_3^{U_{eye}}$. Thus, *i2D*-selective operators may indeed explain a substantial aspect of the saccadic selection process.

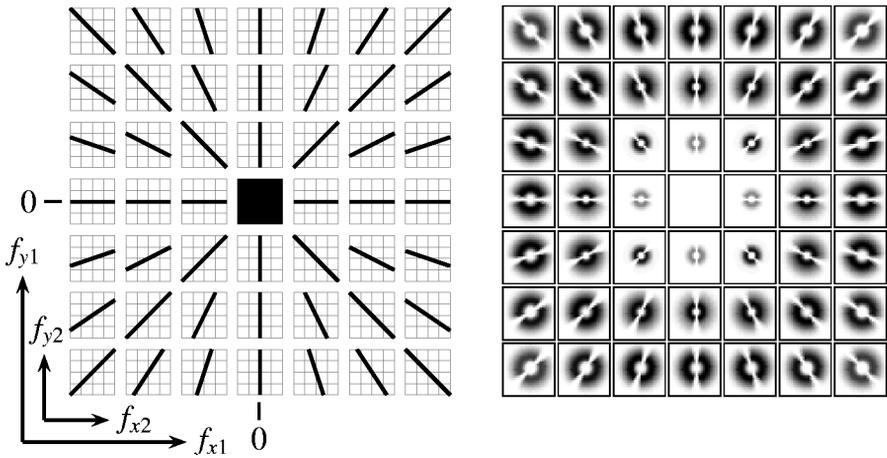


Figure 4. (Left) Illustration of the forbidden zones (indicated as black lines) for an $i2D$ -selective Volterra kernel $H_2(f_{x1}, f_{y1}, f_{x2}, f_{y2})$ in the frequency-domain. A quadratic system with its symmetric kernel vanishing in the forbidden zones blocks frequency components of equal orientation (Krieger and Zetzsche, 1996). Note also the structural relationship of the forbidden zones to the concentration of the bispectral energy $C_3^{U_{rand}}$ in Fig. 3. (Right) Prototypical example of the kernel of an $i2D$ -selective quadratic Volterra operator. Since the forbidden zones are taken into account, the operator will only respond to $i2D$ -signals. The operator will thus enhance bispectral regions with non-aligned frequency components (cf. $C_3^{U_{i2D}}$ in Fig. 3).

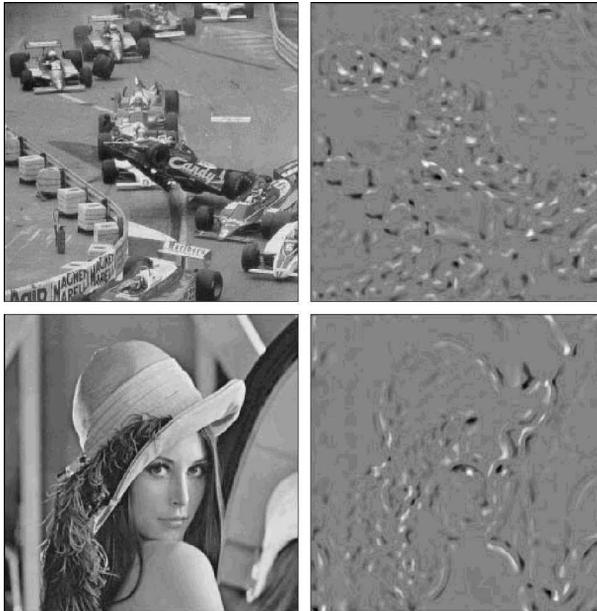


Figure 5. Application of $i2D$ -selective operators to ‘natural’ images. It is apparent that the sparsely distributed regions which are selected by such operators carry a substantial amount of information about the images. The selection is also in good qualitative agreement with the saccadic fixations on these images.

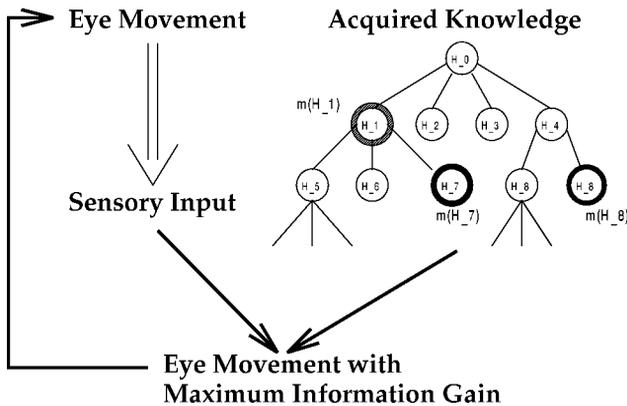


Figure 6. Integrated system for the combination of bottom-up and top-down processing. Based on the current sensory input and on the actual belief distribution in the hierarchical knowledge base, the strategy calculates the action for the greatest information gain, i.e. the eye movement with the greatest potential for a change of the actual belief distribution. A saccadic eye movement thus corresponds to the ‘best question to ask’ in order to obtain the maximum information about the scene.

However, while it may well be possible to identify nearly all potentially interesting locations within a complex scene by the use of a specific set of ‘saliency detectors’, it would usually not be an optimal strategy to systematically scan all these potential fixation regions. Rather, it must be possible to select a subset out of this ‘preselection’, and it must also be possible to fixate points outside the preselected set (note, however, that untrained subjects have great difficulties in making a saccadic eye movement to a certain position in space if they are confronted with a smooth surface of constant luminance). This implies that any model which is based solely on bottom-up processes can only provide an approximate statistical prediction of the fixations by human observers, and indeed it has been shown that top-down influences may have a substantial influence on the pattern of saccadic fixations (Yarbus, 1967). We have hence started an interdisciplinary project in which we investigate how the bottom-up/low-level feature extraction process can be combined with top-down influences from a higher-level knowledge-based scheme to obtain one integrated architecture for the whole system (cf. Fig. 6, Zetsche *et al.*, 1998; Schill *et al.*, 1999).

The knowledge-based top-down component has three main characteristics: (i) In contrast to the common assumption of a purely sensory representation of objects and scenes it is based on a sensori-motor representation, which combines sensory features with motor actions (eye movements). This representation is not deterministic, in the sense of associating each scene or object with a fixed sequence of eye movements, as in the scan-path theory (Noton and Stark, 1971), but rather each sensori-motor ‘vector’ (i.e. triple feature/eye-movement/feature) can provide different degrees of supporting evidence for the various hypotheses about the current scene. (ii) The evidences are learned by the system in a supervised training proce-

ture and their combination for specific hypotheses is obtained within the framework of Dempster–Shafer theory (Shafer, 1976), which allows the handling of conflicting hypotheses, incomplete knowledge, and the distinction between non-supporting evidence and a mere lack of knowledge. (iii) The systems strategy for the gathering of new knowledge is based on the principle of the maximum information gain (Schill, 1997). The actual belief situation is realized as the activity distribution in a hierarchical network. The potential influence of the various possible eye movements on this activity distribution is evaluated, and the eye movement with the greatest potential for a change is selected as the next action.

In this philosophy, the exploration of a scene by saccadic eye movements corresponds to a sequence of ‘best questions to ask’. These best questions are both influenced by *a priori* information about the basic statistical regularities of the environment and by cognitive higher-level representations and inference processes. Here we have presented some first approaches towards the investigation and modelling of the underlying structures and processes, but these are still far away from the efficiency and ingenuity of the biological system. Saccadic scene analysis thus remains a challenging topic for future research.

Acknowledgement

A report of a pilot study can be found in Michela (1993). The main results have first been presented on the Conf. Adaptive Behavior, Zürich, 17-21 August 1998. This study was supported by the Deutsche Forschungsgemeinschaft, SFB 462/B5 and Re337/10-2. Gerhard Krieger is now with the German Aerospace Center (DLR, Institut für Hochfrequenztechnik und Radarsysteme, Oberpfaffenhofen, Germany).

REFERENCES

- Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing? *Network* **3**, 213–251.
- Attneave, F. (1954). Some informational aspects of visual perception, *Psychol. Rev.* **61**, 183–193.
- Barth, E., Caelli, T. and Zetsche, C. (1993). Image encoding, labelling and reconstruction from differential geometry, *CVGIP: Graphical Models and Image Processing* **55** (6), 428–446.
- Biedermann, I. (1987). Recognition-by-components: a theory of human image understanding, *Psychol. Rev.* **94**, 115–147.
- Bozkov, V., Bohdanecy, Z., Radil-Weiss, T., Mitrani, L. and Yakimoff, N. (1982). Scanning open and closed polygons, *Vision Research* **22**, 721–727.
- Crane, H. and Steele, C. (1985). Generation-V dual-Purkinje-image eyetracker, *J. Opt. Soc. Amer.* **24**, 527–537.
- Dowling, J. E. (1987). *The Retina: An Approachable Part of the Brain*. Harvard University Press, Cambridge.
- Krieger, G. and Zetsche, C. (1996). Nonlinear image operators for the evaluation of local intrinsic dimensionality. [Special Issue on Nonlinear Image Processing] — *IEEE Trans. Image Processing*, **5** (6), 1026–1042.

- Krieger, G., Zetsche, C. and Barth, E. (1997). Higher-order statistics of natural images and their exploitation by operators selective to intrinsic dimensionality, in: *Proc. IEEE Signal Processing Workshop on Higher-Order Statistics*, Vol. PR08005, pp. 147–151. IEEE Computer Society, Los Alamitos, CA.
- Mackworth, N. H. and Morandi, A. J. (1967). The gaze selects informative details within pictures, *Perception and Psychophysics* **2**, 547–552.
- Mannan, S., Ruddock, K. and Wooding, D. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images, *Spatial Vision* **10**, 165–188.
- Michela, F. (1993). Merkmalsextraktion zur Blickbewegungssteuerung auf natürlichen Bildern. Lehrstuhl für Nachrichtentechnik, TU-München, Diplomarbeit.
- Nikias, C. L. and Petropulu, A. P. (1993). *Higher-Order Spectral Analysis: A Nonlinear Signal Processing Framework*. Prentice-Hall, Englewood Cliffs, NJ.
- Noton, D. and Stark, L. (1971). Scanpaths in saccadic eye movements while reviewing and recognizing patterns, *Vision Res.* **11**, 929–942.
- Olshausen, B. A. and Field, D. J. (1996). Wavelet-like receptive fields emerge from a network that learns sparse codes for natural images, *Nature* **381**, 607–609.
- Orban, G. A. (1984). *Neuronal Operations in the Visual Cortex*. Springer, Heidelberg.
- Rao, R. and Ballard, D. (1995). An active vision architecture based on iconic representations — [special issue on vision], *Artificial Intelligence* **78**, 461–505.
- Reichardt, W. and Poggio, T. A. (1981). *Theoretical Approaches in Neurobiology*. MIT Press, Cambridge.
- Reinagl, P. and Zador, A. M. (1999). Natural scene statistics at the centre of gaze, *Network: Comput. Neural. Syst.* **10**, 341–350.
- Rentschler, I. and Treutwein, B. (1985). Loss of spatial phase relationships in extrafoveal vision, *Nature* **313**, 308–310.
- Schetzen, M. (1989). *The Volterra and Wiener Theories of Nonlinear Systems*. Krieger's, Malabar, FL, updated edition.
- Schill, K. (1997). Decision support systems with adaptive reasoning strategies, in: *Foundations of Computer Science, Lecture Notes in Computer Science*, pp. 417–427. Springer, Berlin.
- Schill, K., Umkehrer, E., Beinlich, S., Krieger, G. and Zetsche, C. (1999). Knowledge-based scene analysis with saccadic eye movements, in: *Human Vision and Electronic Image Processing, Proc. SPIE*, Vol. 3644, Rogowitz, B. and Pappas, T. (Eds), pp. 520–531. SPIE, Bellingham, WA.
- Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press, Princeton.
- Strasburger, H., Rentschler, I. and L.O. Harvey, J. (1994). Cortical magnification theory fails to predict visual recognition, *Europ. J. Neurosci.* **6**, 1583–1588.
- van Hateren, J. H. (1992). Real and optimal neural images in early vision, *Nature* **360**, 68–69.
- Yarbus, A. L. (1967). *Eye Movements and Vision*. Plenum Press, New York.
- Zetsche, C. and Barth, E. (1990a). Fundamental limits of linear filters in the visual processing of two-dimensional signals, *Vision Research* **30**, 1111–1117.
- Zetsche, C. and Barth, E. (1990b). Image surface predicates and the neural encoding of two-dimensional signal variation, in: *Human Vision and Electronic Imaging: Models, Methods, and Applications, SPIE*, Vol. 1249, Rogowitz, B. (Ed.), pp. 160–177.
- Zetsche, C., Barth, E. and Wegmann, B. (1993). The importance of intrinsically two-dimensional image features in biological vision and picture coding, in: *Digital Images and Human Vision*, Watson, A. (Ed.), pp. 109–138. MIT Press, Cambridge, MA.
- Zetsche, C. and Deubel, H. (1990). Prediction of saccadic fixations by nonlinear 2d-detectors. Unpublished pilot study.
- Zetsche, C. and Krieger, G. (1999). Nonlinear neurons and higher-order statistics: new approaches to human vision and electronic image processing, in: *Human Vision and Electronic Image*

- Processing, Proc. SPIE*, Vol. 3644, Rogowitz, B. and Pappas, T. (Eds), pp. 2–33. SPIE, Bellingham, WA.
- Zetzsche, C., Krieger, G. and Wegmann, B. (1999). The atoms of vision: Cartesian or polar? *J. Opt. Soc. Amer.* **16** (7), 1554–1565.
- Zetzsche, C., Schill, K., Deubel, H., Krieger, G., Umkehrer, E. and Beinlich, S. (1998). Investigation of a sensorimotor system for saccadic scene analysis: an integrated approach, in: *Proc. 5th Intl. Conf. Soc. Adaptive Behavior*, Vol. 5, Pfeifer, R., Blumenberg, B., Meyer, J. and Wilson, S. (Eds), pp. 120–126. MIT Press, Cambridge, MA.

Copyright of Spatial Vision is the property of VSP International Science Publishers and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.