

# The world from a cat's perspective

## Statistics of natural videos

**Journal Article****Author(s):**

Betsch, Belinda Y.; Einhäuser, Wolfgang; Körding, Konrad P.; König, Peter

**Publication date:**

2004-01

**Permanent link:**

<https://doi.org/10.3929/ethz-b-000051959>

**Rights / license:**

[In Copyright - Non-Commercial Use Permitted](#)

**Originally published in:**

Biological Cybernetics 90(1), <https://doi.org/10.1007/s00422-003-0434-6>

# The world from a cat's perspective – statistics of natural videos

Belinda Y. Betsch, Wolfgang Einhäuser, Konrad P. Körding, Peter König

Institute of Neuroinformatics, ETH/University Zürich, Winterthurerstr. 190, 8057 Zürich, Switzerland

Received: 16 December 2002 / Accepted: 9 September 2003 / Published online: 19 January 2004

**Abstract.** The mammalian visual system is one of the most intensively investigated sensory systems. However, our knowledge of the typical input it is operating on is surprisingly limited. To address this issue, we seek to learn about the natural visual environment and the world as seen by a cat. With a CCD camera attached to their head, cats explore several outdoor environments and videos of natural stimuli are recorded from the animals' perspective. The statistical analysis of these videos reveals several remarkable properties. First, we find an anisotropy of oriented contours with an enhanced occurrence of horizontal orientations, earlier described in the "oblique effect" as a predominance of the two cardinal orientations. Second, contrast is not elevated in the center of the images, suggesting different mechanisms of fixation point selection as compared to humans. Third, analyzing a sequence of images we find that the precise position of contours varies faster than their orientation. Finally, collinear contours prevail over parallel shifted contours, matching recent physiological and anatomical results. These findings demonstrate the rich structure of natural visual stimuli and its direct relation to extensively studied anatomical and physiological issues.

## 1 Introduction

For a thorough understanding of the visual system it is important to know the typical input it operates on. Recent advances in information technology have spurred many studies of the characteristics of natural scenes (Field 1987; Burton and Moorhead 1987; Hancock et al. 1992; Tolhurst et al. 1992; van Hateren 1993; Eckert and Buchsbaum 1993; Ruderman and Bialek 1994; Dong and Atick 1995; van der Schaaf and van Hateren 1996; Olshausen and Field 1996). Indeed, recent

experimental results suggest that the visual system is particularly suited to process natural stimuli as compared with the processing of artificial stimuli (Rizzo et al. 1995; Dan et al. 1996). Akin to these statistical properties, anatomical and physiological studies describe regularities of cortical connections. Out of these studies the view emerges that properties of adapting systems strongly depend on the type of inputs they receive. In fact, for several analyzed species, the early visual system has been shown to be optimally adapted to its environment (Hughes 1977; Lythgoe 1979). Thus, a better characterization of properties of natural stimuli fosters a better understanding of sensory processing.

First, we want to address the question of which stimuli are "natural". An often used source of such stimuli is images from publicly available databases or photographs taken in a natural environment. Still images have the disadvantage that all temporal information is lost. However, the world is a dynamic process and confronts us with constantly changing stimuli, and a number of studies show that the inability to perceive movement (akinetopsia) comprises a range of vital abilities (Zeki 1991; Rizzo et al. 1995). It thus seems particularly relevant to analyze properties of time-varying scenes.

Second, many studies use human-selected images. Although these images depict natural scenes, it is currently unclear to what degree they might introduce an anthropological bias. For example, investigating the function of different eye movement systems, in the cat the stabilizing effect of eye movements during locomotion is found to be surprisingly small (Moeller et al. 2003). Hence the effective retinal stimulation in a cat is not a sequence of quasistatic images, as intuition might suggest. Hence, it is not obvious whether the spatial and temporal properties of human-selected natural stimuli represent the typical visual input that animals encounter in their environment. This is a particularly severe constraint as the overwhelming majority of anatomical and physiological studies of the mammalian visual system (with the recent exception of fMRI studies) concentrate on nonhuman subjects, especially monkeys and cats.

Correspondence to: W. Einhäuser  
(e-mail: weinhaeu@ini.phys.ethz.ch)

Here we take into account that the exploration strategy and the difference of vantage point of the animal may impact the statistics of natural scenes. We describe a setup for recording natural stimulus videos in outdoor environments taken by freely behaving cats. These data are analyzed by standard second-order statistics as well as by statistics in the wavelet domain. Thus this study aims at analyzing regularities of the world as seen from the cat's perspective.

## 2 Methods

### 2.1 Experimental setup

Four cats are used for collecting natural videos (Fig. 1). They wear chronic skull implants, containing microelectrodes, which are used in the context of another study. This implant is made of dental cement and carries two nuts to which we temporarily attach a CCD camera (Conrad Electronics, Hirschau, Germany). The camera is mounted on a circuit board measuring 38 mm by 38 mm, yielding a total weight of 34 g. This camera has a global gain control, which partly mimics adaptation processes of the visual system. The time constant of this adaptation process (2.5 s) does not quantitatively match the characteristics of luminance adaptation of the cat visual system precisely. However, it is an order of magnitude longer than the time constants of the processes investigated in the present study and thus does not confound any analysis presented here. The vertical distance of the camera from the eyes is about 4 cm. The camera is aligned with the visual axis of the animal, when it is looking straight ahead. We control this by examining whether objects the cat is fixating on from various distances are represented in the central image area. When the cat is very close to an object (less than about 20 cm) and eventually touching it, the images taken by the camera no longer match the visual input of the animal. Therefore, data originating from such instances are discarded. The camera contains a 10-mm lens featuring a visual angle of  $71^\circ$  in the horizontal direction and  $53^\circ$  in the vertical direction. Due to the short focal length of the lens, the whole visual field is in focus. The camera connects to a backpack carried by a scientist. It contains the recording equipment: an analog PAL-VCR (Lucky Goldstar, Seoul, Korea), a 12-V battery for power supply, and a miniature television for online control during walks outside. The whole procedure is in compliance with institutional, cantonal (Zürich), and national guidelines for experimental animal care.

### 2.2 Recording videos

We have collected color movies of an average length of 20 min, spread over the four seasons, in various weather situations and locations (forest, university campus, grassland, residential areas; see Fig. 1). Sections of the videos containing recording artifacts (e.g., glare that



**Fig. 1.** Setup. **a** The cat is exploring an outside park area on one of its walks. The red cable, twisted with the leash, connects to the VCR in the backpack. **b–e** Four typical pictures taken from the videos are shown. **b** The horizon divides the image into a bright, low-contrast upper image region (sky) and a darker, lower region of high contrast (stones). **c** The cat's view of the pond shows richly detailed plant structures and low-contrast water regions. **d** On close inspection by the cat, blades of grass are evenly spread over the entire image. **e** During a walk in the nearby forest the upper half is dominated by dark, vertically oriented trees in front of the bright sky. The lower half of the image representing the forest floor consists of many objects (branches, leaves) arranged in all possible orientations

saturates whole columns of the CCD array) are excluded from further analysis. From our database we thereby choose sequences of about 3000–5000 frames (53884 frames in total), recorded at a sampling rate of 25 frames per second. These are digitized at a resolution of  $320 \times 240$  pixels using a microVIDEO DC30 graphics card (Pinnacle Systems, Mountain View, CA) and Adobe Premiere software (San Jose, CA). The data are converted with standard algorithms from RGB to grayscale using the “*rgb2gray.m*” function of MatLab (Mathworks 2000; Natick, MA) and stored in TIFF format (8-bit). The pixel values are mapped linearly on the interval from 0 to 1 with identical scaling for all frames. We use only the central  $200 \times 200$  pixels, where the imaging properties of the camera are homogeneous, for further analysis. For the calculation, images are low-pass filtered by convolution with a Gaussian at a cutoff frequency of the transfer function of the CCD camera (0.53 cycles/deg). This is lower than the spatial acuity at the fixation spot but of similar order of magnitude at moderate eccentricities (Pasternak and Horn 1991). Contrast is then defined as the standard deviation of intensity within a  $20 \times 20$  pixel ( $4.6^\circ \times 4.6^\circ$ ) patch divided by the mean intensity in the image. Natural image data are available from the authors upon request.

### 2.3 Statistical analysis

For the spectral analysis we use standard methods (Papoulis 1991; appendix, Sect. 5.1) and a wavelet analysis (see appendix Sect. 5.2). The mean wavelet amplitudes are computed by averaging over time and/or space as described in the respective sections below. The spatiotemporal autocorrelation of the amplitude of the local Gabor components is calculated according to standard algorithms for each wavelet individually. Definitions of notation are presented in the appendix.

## 3 Results

### 3.1 Contrast at the fixation spot

A recent study reports a 20% higher contrast at fixation points selected by humans scanning still images (Reinagel and Zador 1999). Hence we compare the contrast at the center of the image, which is the direction the cat preferentially looks at (Guitton et al. 1984), with the contrast at higher eccentricities (Fig. 2b). Even using the assumption that subsequent frames are independent, only three movies indeed show significantly ( $p < 0.05$ , t-test) higher contrast values in the central part of the image. In one example (Fig. 2b, left) we find a 24% higher contrast in the central image area. A frame taken from this movie illustrates this contrast distribution (Fig. 1b): a bright sky and a dark ground form the low-contrast image regions, whereas most high-contrast objects are located in the middle image regions. Other movies, however, show a different contrast distribution. In one movie taken during a walk

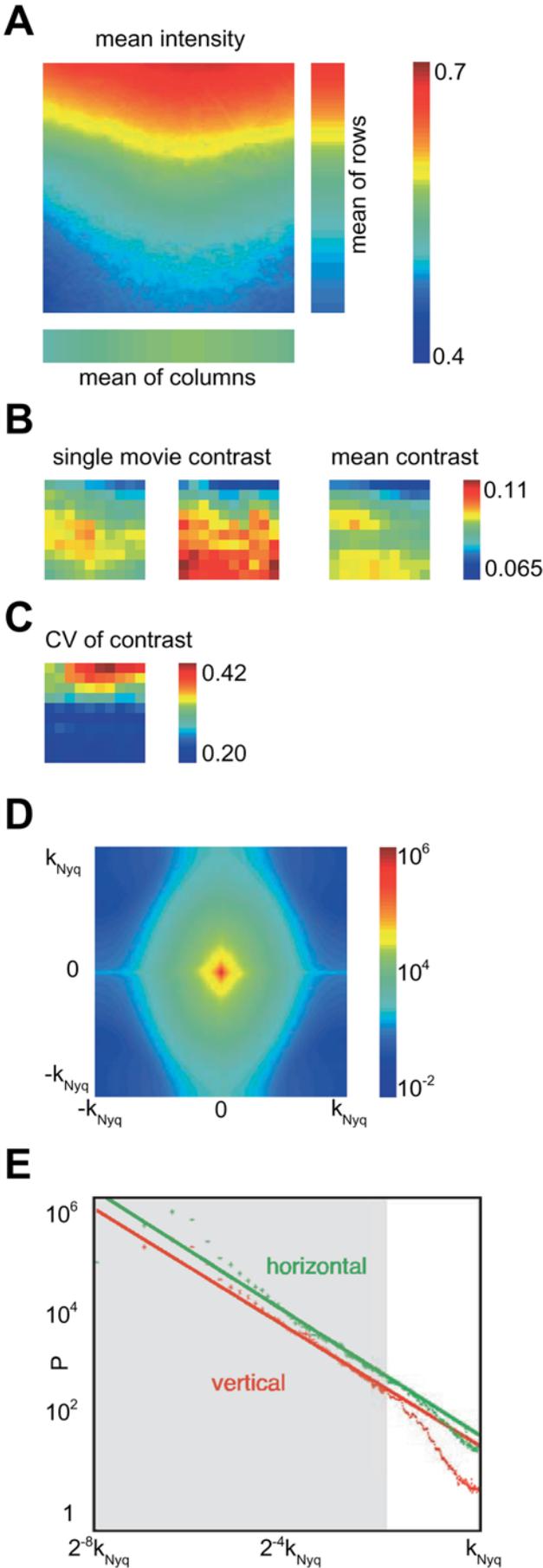
through the woods, for example, the lower image region has approximately 28% higher contrast values compared with the upper half (Fig. 2b, middle). A frame of this movie (Fig. 1e) depicts a low-contrast sky, whereas two thirds of the lower image are filled with high-contrast features like branches, leaves, and brushwood. The average contrast over the complete set of movies is shown in Fig. 2b (right), and the spatial distribution of its coefficient of variation ( $CV \equiv \text{standard deviation/mean}$ ) is illustrated in Fig. 2c. Indeed the average contrast is not homogeneously distributed over the whole region. The top margin of the image shows a band of low contrast, whereas the two thirds forming the lower half of the image shows about 16% higher contrast values. The central image area, however, does not reveal a significantly higher contrast profile. The contrast for the central  $8.8^\circ$  deviates from the mean by only 3.8%. In conclusion, in contrast to the study of human vision, here, using cats exploring a natural environment, we did not observe a bias of fixating high-contrast regions.

### 3.2 Contours at the cardinal axes

How is the distribution of oriented contours encountered by a freely behaving animal? The analysis of the stimulus sequences reveals that the local amplitude of horizontal features for the standard spatial frequency of  $\omega_2$  ( $\langle L^{(\omega_2, \varphi=90^\circ, r_0)} \rangle_{r_0 \in \Gamma}$  (see Methods and appendix Sect. 5.2) is significantly higher than the average of all orientations (6.79%,  $p < 0.001$ , paired sign test). For the other spatial frequencies we obtain values of 15.24%, 11.07%, and 4.70% ( $\omega_0, \omega_1, \omega_3$ , respectively).

This result can be illustrated by complementing the wavelet analysis with a standard frequency analysis (Fig. 2d). The shape of the power spectrum is anisotropic, and the power is highest in horizontal orientations. When plotted on a log-log scale (Fig. 2e), the falloff in power for the higher spatial frequencies along the two main axes is clearly visible. More energy belongs to the horizontal orientation (green) than to the vertical orientation (red) by a factor of 1.17. Also note that the fit is not perfect at low frequencies, indicating a limited range of scale invariance of natural images.

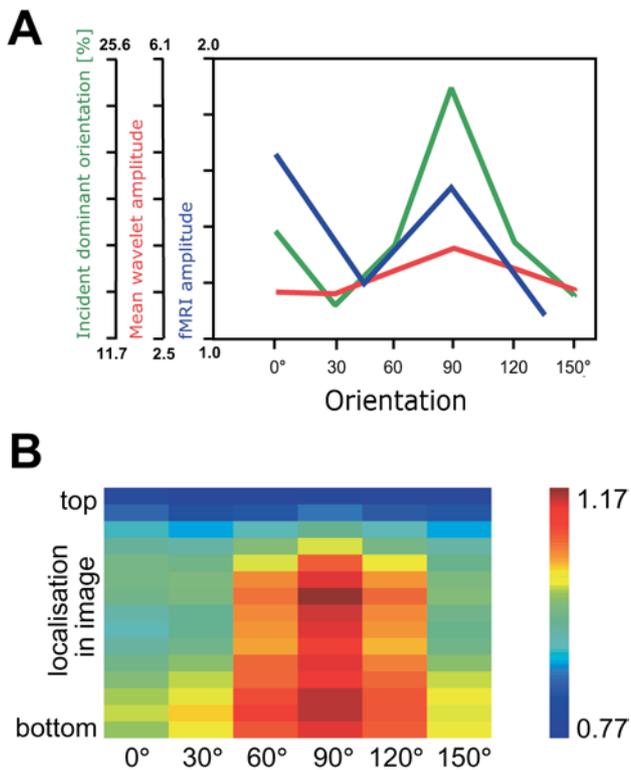
To ensure that the dominance of horizontal contours observed in the wavelet analysis is not a result of a setup artifact, we performed the following control: we mounted the camera to a pivotable tripod and took images using the camera in normal orientation and rotated by  $90^\circ$  within the image plane. The data acquisition and processing was exactly identical to that of the original setup, including recording of the camera signal to the VCR and digitizing offline from tape. For the “normal” image we found (total over all used  $\omega$ ) 22.6% of the contours to be predominantly horizontal, and for the rotated image 23.5% to be predominantly vertical (i.e., horizontal in the real world). These numbers closely match the result for the whole image set and thus exclude the possibility that the result could be a consequence of an anisotropy of the camera setup.



**Fig. 2.** Instantaneous contrast statistics. **a** Luminance in the central image region ( $200 \times 200$  pixels) averaged over the whole database is color coded. Ordinate and abscissa denote the position of the analyzed image region. The averages over columns and rows are shown beside the axis. **b** Contrast (see Methods) for  $20 \times 20$  image patches averaged over time is color coded for two single movies (*left, middle*) and averaged over the whole database (*right*). Ordinate and abscissa denote the horizontal and vertical position of the analyzed image region, respectively. The diagram shown covers the central image region of  $200 \times 200$  pixels. **c** Standard error of the mean contrast distribution shown in Fig. 2b. **d** Power spectrum averaged over all movies. Maximal power is found in the center at low spatial frequencies. **e** The two main axes of Fig. 2d are plotted on a log-log axis. The *gray shading* indicates the region unaffected by the low-pass filter, which is used for the fit. Both datasets fall approximately on a straight line and are well fitted to the  $(1/k)^{2-\eta}$ -model (see Methods) with an  $\eta$  of  $-0.10$  for the horizontal and  $0.16$  for the vertical orientation. This behavior is described by the “Power Law” (Field 1987; Burton and Moorhead 1987; Tolhurst et al. 1992; van der Schaaf and van Hateren 1996), which is compatible with scale invariance. This scaling invariance seems to be a general property of natural scenes, and it has been suggested that it is reflected in the properties of the mammalian visual system (Hughes 1977)

To ensure that the visibility of the horizon does not produce the dominance of horizontal orientations, we perform three controls. First, we calculate the incidences of the dominant orientation  $\varphi_D^{(\omega_2, r_0)}$  (Fig. 3a, green). In contrast to the above discussion, we do not calculate the mean power at each orientation but determine at each location the dominant orientation and compare the statistics of this distribution: in 24.1% of all local regions (i.e., all grid points  $r_0 \in \Gamma$  at spatial frequency  $\omega_2$ ) analyzed, the amplitude of the horizontal wavelet is largest. This is 58.6% higher than the mean of incidence of the other dominant orientations.

Thus large image areas are affected by the horizontal feature predominance, and as the horizon contributes only to a limited area within each image, it cannot fully explain this effect. Second, we perform the analysis on a subset of movies that do not contain the horizon and obtain comparable results (data not shown). Third, we investigate the spatial distribution of the mean wavelet amplitudes within the images. We are interested in the dependence of prevalent orientations along the vertical and horizontal axes. Averaging the wavelet amplitudes over image rows and orientations results in one value per column. These values have a standard deviation that is small compared to their mean ( $0.10 \ll 3.21$ ). Thus, the wavelet amplitudes show little variation in the horizontal direction. Instead, we observe a strong variation of wavelet amplitudes along the vertical axis of the image (Fig. 3b). The average amplitude of all orientations of the upper image half is 7.48% smaller than in the lower part of the image. This is a natural consequence of the spatial distribution of contrast reported above. Furthermore, the average amplitude for the horizontal orientation in the upper third of the image is even 8.64% smaller than in the lower two thirds of the image. Thus the lower image regions contribute most to the predominance of the horizontal orientation, while the upper regions of the image contribute most vertical orientations. Upright trees against



**Fig. 3.** Instantaneous wavelet statistics. **a** The mean wavelet amplitude is demonstrated (red) for the six different orientations (0° denotes the vertical orientation) averaged over the complete database. The abscissa shows the orientation of the image region ranging from 0 to 150°. The incidences of the dominant orientation (green) are illustrated for the six different orientations. The mean fMRI response amplitude in V1 (blue) plotted as a function of orientation (data replotted from Furmanski and Engel 2000; Fig. 1c). **b** The mean wavelet amplitude in all movies, averaged over columns, is plotted color coded as a function of height in the image (vertical axis) and of the orientation of the wavelet (horizontal axis)

a bright sky (Fig. 1e), for example, contribute to this effect. This third control shows that the horizon may contribute to some extent to the preponderance of horizontal orientations, but it does not explain the full effect. Taken together, these three controls show that the predominance of horizontal contours is a general property of the stimuli the animal experiences in its natural habitat.

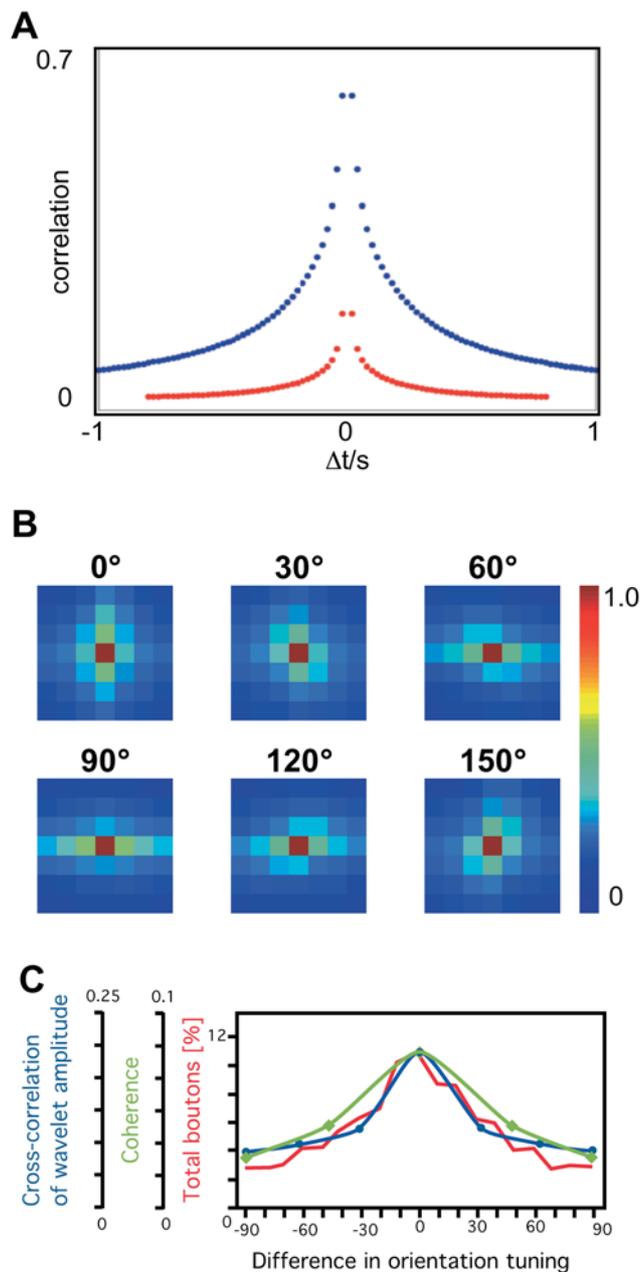
This phenomenon of a predominance of horizontal and vertical orientations has been described as the oblique effect. It is not only exemplified in statistics on natural images but has also been described in anatomical and physiological studies on visual processing. In Fig. 3a (blue line) we demonstrate this analogy in particular for data on fMRI (Furmanski and Engel 2000). The study shows a significantly greater behavioral sensitivity with larger responses in primary visual cortex to stimuli of cardinal orientation (horizontal and vertical) as opposed to oblique features. It also demonstrates that the signals induced by the two cardinal orientations are comparable. The former aspect matches our data on natural images. With respect to the latter, a difference in the results of this study is noticeable.

### 3.3 Orientation and position: reasons for “complex” type neurons

Properties of neurons found in primary visual cortex are related to the spatiotemporal statistics of natural stimuli. First, we analyze the temporal aspects of the spatiotemporal autocorrelation function of wavelet amplitude and phase (see appendix, Sect. 5.3) at zero spatial lag and average over orientations ( $\langle X_L^{(\omega_2, \varphi, r_0)}(\delta = 0, \tau) \rangle_{\varphi \in \Phi}$  and  $\langle X_\Lambda^{(\omega_2, \varphi, r_0)}(\delta = 0, \tau) \rangle_{\varphi \in \Phi}$ , respectively). Figure 4a shows the decay of the correlation functions of amplitudes and phases over time. The autocorrelation function of the local amplitude drops to 10% after 35 frames (1400 ms), whereas the correlation function of the phase reaches this level already after 4 frames (160 ms). This demonstrates that the orientation of local features changes much slower than their position. This property of natural images may serve as the basis of the emergence of complex receptive fields in primary visual cortex, as their response is invariant with respect to the fast changing feature (position) and specific to the slowly changing feature (orientation) (Einhäuser et al. 2002; Wiskott and Sejnowski 2002). Thus the improved understanding of image properties fosters our understanding of properties of neurons in the visual system.

### 3.4 Collinear contours

The Gestalt law of “good continuation” has been proposed as the underlying principle of feature binding. This inspired us to investigate the relation of orientations in neighboring image locations (Krüger 1998; Krüger and Wörgötter 2002). We analyze the spatial aspect of the spatiotemporal autocorrelation of wavelet amplitudes for different image locations. This is done at the central  $7 \times 7$  grid points  $\Gamma'$  to vary relative distances (spatial lags,  $\delta$ ). Because the peak of the spatiotemporal autocorrelation, the value we are interested in, is always located at a time lag of 0 s, only this value is used. ( $X_L^{(\omega_2, \varphi, r_0)}(\delta, \tau = 0)$ ,  $\delta \in \Gamma' \subset \Gamma$ ,  $\varphi \in \Phi$ ). The upper left graph in Fig. 4b illustrates this measure for the vertical orientation ( $\varphi = 0^\circ$ ). An anisotropic behavior is observed – the spatial decay of the correlation is different for the neighboring image locations. Neighboring image regions that lie on the same axis (here the vertical) as the examined orientation are more correlated than image regions displaced in other directions (horizontal or diagonal). The image regions on the same axis with a distance of  $6.2^\circ$  show a correlation of 33.5%, whereas the image regions displaced by the same amount in the orthogonal direction show a correlation of only 26.4%. Similar results are found for all other orientations examined. This level of correlation (33.5%) is reached along the main axis at a 1.47 times larger distance. The choice of this correlation level for comparison is somewhat arbitrary. However, very high levels of correlation are reached only for small distances when the image regions overlap, which could potentially confound results. Much lower levels of correlation lead to noisy data, making a reliable analysis difficult. Thus



**Fig. 4.** Temporal statistics. **a** The spatiotemporal correlation function for the wavelet amplitudes and phases is displayed. The abscissa shows the temporal difference in seconds, whereas the ordinate reveals the temporal correlation (spatial offset is 0 pixels) of wavelet amplitudes (*blue*) and phases (*red*), averaged over space and orientation. **b** The maximal (over relative time) spatial correlation of wavelet amplitudes is illustrated in six color-coded panels for the six different orientations ( $0^{\circ}$ – $150^{\circ}$ ). Data have been averaged over all central image regions and the complete database. Ordinate and abscissa of each panel signify the relative position of the two analyzed image subregions. **c** Correlation of wavelet amplitudes (*blue*) between neighboring image regions is illustrated as a function for the six different orientations. Coherence (*green*): The average coupling strength is illustrated between pairs of cortical recording sites as a function of the relative orientation preference of the neurons. The coupling strength is quantified as a spectral coherence (data replotted from Frien and Eckhorn 2000). Bouton distribution (*red*): The number of labeled terminals found in the sector surrounding an injection site is plotted as a function of the preferred site (data replotted from Bosking et al. 1997)

we consider a correlation level of 33.5% a reasonable choice for analysis. We compare this result with recent anatomical studies on the connectivity in primary visual cortex (Bosking et al. 1997). These authors report a 1.53 larger range (median of maximum distance at which boutons were found) of tangential connections along the axis of orientation preference as compared with tangential connections in the orthogonal direction. This number matches our results fairly well. In this respect we were interested in seeing if we could observe different correlation values along the collinearly displaced image regions when comparing the six different orientations. Indeed, Fig. 4b reveals larger correlations in collinearly displaced regions for the cardinal orientations compared to the oblique ones. Altogether those results reveal that collinear contours are more prevalent than parallel shifted contours, which matches well experimental results, as discussed below.

### 3.5 Interaction between different orientations

The relation between the statistics of natural visual stimuli, the anatomical structure of tangential connections in cortex, and the interaction of orientation-tuned neurons is illustrated in Fig. 4c. Here we juxtapose the results of recent anatomical and physiological studies with our data. The correlation of the wavelet amplitude at neighboring image regions declines with increasing differences in orientation tuning. Interestingly, a similar phenomenon can be observed in anatomical results on the bouton distribution, which is reflected in the number and in the distance of the labeled terminals (Fig. 4c, red line; data from Bosking et al. 1997). In this way, the relationship between the axis of elongation of the terminal distribution and the preferred stimulus orientation of the injection site has been quantified. Labeled boutons were consistently found to extend greater distances from the injection site along the axis of visual field that corresponds to their preferred stimulus orientation (Bosking et al. 1997; Schmidt et al. 1997). A similar dependency on the stimulus orientation could be exemplified in a physiological study measuring the coherence (Fig. 4c, green line; data from Frien and Eckhorn 2000) among recording pairs in striate cortex of awake monkeys. Similar results have been obtained in anaesthetized cats (Engel et al. 1991): the more the orientations were similar, the higher was the observed coherence.

Thus we can observe that both anatomical results and physiological data match the distribution and correlation of local image features in our natural videos.

## 4 Discussion

In this study cats explore their environment and collect natural visual stimuli. The statistical analysis of this large database allows studying properties of natural time-varying stimuli. Results from previous studies on still images, which report “scaling behavior” (Field

1987; Hancock et al. 1992; Tolhurst et al. 1992; Ruderman and Bialek 1994) and the Power-Law (Ruderman and Bialek 1994; van der Schaaf and van Hateren 1996), are also demonstrated for our dataset. Analyzing the spatial distribution of contrast, we do not find an increased contrast in the presumed direction of gaze. The precise position of contours changes more rapidly than their orientation. Finally, image features of the cardinal orientation or spatially correlated with collinear contours are more prevalent than parallel shifted contours.

#### 4.1 Contrast at the fixation spot

Previous studies on human eye movement (Reinagel and Zador 1999) demonstrated a higher contrast at the fixation points compared with randomly chosen parts of static pictures. Therefore, one might expect a higher contrast at the positions that the cats are looking at. However, the variance of contrast over movies taken in different environments outweighs the difference between the central and peripheral regions. Whereas three movies reveal a higher contrast in the central image area, movies recorded in other surroundings show a drastically different contrast distribution. Thus this aspect of the image statistics strongly depends on the explored environment. However, comparing this result with the results of Reinagel and Zador (1999), several differences in subjects and experimental setups are obvious:

In the present study we assume that the central part of the image represents the point that the cat is actually fixating on. We control this by examining whether objects the cat is fixating on from various distances are represented in the central image area. Furthermore, preliminary data with a modified setup to record eye movements in the freely behaving animal indicate that freely behaving cats make mostly eye movements of small amplitude (Moeller et al., 2003). Small eye movements, however, can potentially smooth out a contrast peak at the center of the image, and we would no longer detect this maximum in the statistics. Furthermore, allowing head movements makes correcting “head saccades” very effective (Guitton et al. 1984). Thus, it seems reasonable to assume that the animals primarily gaze into the direction of the center of the image.

Another important difference between the previous (Reinagel and Zador 1999) and the present study is the use, respectively, of humans and cats as subjects. The small fovea endows the primate visual system with a very high spatial acuity. In comparison, the region of the best visual acuity of the cat, the area centralis, is much larger than the fovea, and its spatial acuity is an order of magnitude lower (Blake 1988). Thus, the need to focus on high spatial detail might be much reduced in cats.

Finally, differences in the experimental setup, scanning static high-contrast artistic images as opposed to cluttered natural image sequences, make different results hard to judge and require further research to obtain clarity.

#### 4.2 The oblique effect

The data analysis showed a predominance of cardinal over oblique contours, with a further preference for horizontal over vertical orientations. The oblique effect is well known in the psychophysical and physiological literature. However, the number of previous publications on this issue using natural image material is surprisingly small (Switkes et al. 1978; Hancock et al. 1992; van der Schaaf and van Hateren 1996; Coppola et al. 1998b; Keil and Cristobal 2000). Those studies explore the oblique effect, though they mainly examine human-selected natural visual scenes. However, the interest in the “oblique effect” provoked many anatomical and physiological studies, extensively illuminating the orientation selectivity in different organisms and experimental conditions. The main consensus is that both cardinal-oriented contours are more abundant than oblique contours. Psychophysical studies could demonstrate a better visual performance regarding visual acuity and contrast sensitivity when using cardinal-oriented stimuli (Boltz et al. 1979; Krebs et al. 2000). Single-unit recordings in the primary visual cortex demonstrated higher response properties for stimuli at the cardinal axes, using cats (Pettigrew et al. 1968; Fregnac and Imbert 1978; Orban and Kennedy 1981; Leventhal and Schall 1983) or monkeys (Mansfield and Ronner 1978; Blakemore et al. 1981). Indeed, a recent study based on a large database of recorded cells confirms the larger fraction of neurons responding to the cardinal axes and furthermore provides evidence that horizontal contours activate more neurons than do vertical contours (Li et al. 2003). This is in line with the results reported in the present study. Stimuli at the cardinal orientations produced a larger cortical representation in optical imaging studies in ferrets (Chapman and Bonhoeffer 1998; Coppola et al. 1998a) and cats (Dragoi et al. 2001). Similarly, they induced a higher performance (faster latency, larger amplitude) when evoked potentials were measured (Arakawa et al. 2000). The cardinal effect is not purely “cortical” but could also be demonstrated for subcortical sites as the retina and the LGN, whereby retinal ganglion cells and their dendrites are preferentially arranged along the vertical and horizontal meridians (Wässle et al. 1975). A further differentiation has been described by applying stimuli with different spatial frequencies. It was shown that neurons in the macaque visual cortex that are tuned to high spatial frequency preferred the cardinal orientations (De Valois et al. 1982), whereas another study applying middle to low spatial frequencies even described a preference for oblique orientations (Nelson et al. 1984). The data presented here show that when our cats explore a natural environment, horizontal contours dominate at all spatial frequencies investigated. Thus, results on the increased representation of cardinal orientations with an emphasis on horizontal contours in primary visual cortex (Coppola et al. 1998a; Li et al. 2003) might be a natural consequence of the statistics of visual stimuli. This argument would favor the assumption that natural visual input has an instructive role in

the orientation preference of cortical neurons (Sengpiel et al. 1999). Hand in hand with this idea go single-unit studies examining the effect of offering a single visual orientation during early visual experience and demonstrating a predominance of neurons responsive to that orientation (Hirsch et al. 1970). Apparently the selection of the visual input results in an overrepresentation of the experienced image features. Thus we would suggest that the statistics of our natural visual images reflect the structure and processing of the visual system.

#### 4.3 Temporal continuity extracted by complex neurons

Analyzing the temporal properties of natural stimuli, we find that the positions (on the scale of receptive field sizes) of contours change much faster than their orientation. In a simulated neural network this property, in combination with a recently proposed learning rule, allows learning translation-invariant receptive fields (Hyvärinen and Hoyer 2000; Körding and König 2000; Einhäuser et al. 2002; Wiskott and Sejnowski 2002). This matches properties of complex neurons found in primary visual cortex (Hubel and Wiesel 1962), suggesting an unsupervised learning algorithm to provide a factorial code of independent visual features. Thus it forms a new step connecting properties of natural scenes and neural processing (Law and Cooper 1994; Olshausen and Field 1996; Dan et al. 1996; Bell and Sejnowski 1997; van Hateren and Ruderman 1998) and helps us to gain a better understanding of complex cells in the visual cortex.

#### 4.4 Anisotropy of lateral interactions

Being interested in the extent to which neighboring image regions refer to each other, we found that image regions lying on the same axis as the examined orientation are more strongly correlated. Thus collinear features are more prevalent than other combinations. This is in agreement with previous results on the statistics of human-selected still images (Krüger 1998). Furthermore, it relates closely to the connectivity of primary visual cortex. Recent anatomical studies on the topography of lateral connections and orientation maps (Gilbert and Wiesel 1989; Kisvarday et al. 1997; Schmidt et al. 1997) show that intra-areal projections in cat visual cortex preferentially connect regions of similar orientation preference and regions with a collinear alignment of preferred orientations. Similar were the results in psychophysical studies reporting lower detection thresholds of aligned Gabor patches (Gilbert and Wiesel 1990; Hess and Dakin 1997; Polat and Norcia 1998; Keeble and Hess 1999) with larger response amplitudes and higher contrast sensitivity. Likewise, physiological experiments demonstrate a dependency of neuronal activity and of synchronization (Engel et al. 1991; Gray et al. 1989; Löwel and Singer 1992; König et al. 1993; Frien and Eckhorn 2000) on the alignment of visual stimuli. Furthermore, we observe that the anatomical results on the bouton distribution (Bosking et al. 1997) and the

physiological data on coherence (Frien and Eckhorn 2000) match the distribution and correlation of local image features in our natural time series. As has been suggested for the oblique effect, we would suggest that the cortical anatomy and processing reflects the natural visual input experienced during the formation of “cortical wiring”. When observing that co-axial image regions on the cardinal axes are more strongly correlated over time than those on the oblique axes, and keeping in mind its anatomical and physiological correspondent/analogy, we would suggest that this anisotropy could be related to the stability of the processing of cardinal orientations during visual perception, as proposed in earlier studies (Dragoi 2001).

#### 4.5 Conclusion

In this study, we analyze the statistics of natural stimuli and link them to anatomical and physiological properties of the mammalian visual system. We are able to discern several parallels between the natural visual input and the structure and processing of the visual pathway.

*Acknowledgements.* We would like to thank the Swiss National Science Foundation (Grant No. 31-65415.01, BYB), Honda RI Europe (WE), Boehringer Ingelheim Fonds (KPK), and the EU/BBW (IST-2000-28127; 01.0208-1, PK) for financial support. We would like to thank John Anderson, Tom Binzegger, Andrea Benucci, Tobias Delbrück, and Christoph Kayser for inspiring discussions and comments on a previous version of the manuscript.

## Appendix

### 5.1 Spectral analysis

For the spectral analysis of each frame we use standard methods (Papoulis 1991). To avoid camera artifacts at extreme eccentricities of each digitized frame the central  $200 \times 200$  pixels are further processed. The term “image” refers to this central region throughout the paper. The mean intensity of each image is subtracted. To reduce boundary effects and to avoid any orientation bias, images are pointwise multiplied with a 2D isotropic Gaussian-window of standard deviation 70 pixels. To facilitate the subsequent Fourier transform, the resulting image is zeropadded to size  $256 \times 256$ . For the remainder of this appendix, we will refer to this zeropadded image as  $I$ ; furthermore,  $J$  denotes the set of all pixels of  $I$ ; brackets  $\langle \cdot \rangle$  denote averaging. The Fourier transform is conducted by the 2D Fast Fourier Transform function of MatLab (*fft2.m*) and the 2D power spectrum computed as

$$P_k = \left| \sum_{r \in J} I_r e^{-\frac{2\pi i k r}{256}} \right|^2, \quad (4)$$

where  $\mathbf{k}$  denotes the 2D spatial frequency. To explore scaling behavior, we fit the power spectrum along the horizontal ( $\mathbf{k} = \begin{pmatrix} k \\ 0 \end{pmatrix}$ ) and vertical ( $\mathbf{k} = \begin{pmatrix} 0 \\ k \end{pmatrix}$ ) axes with

the function  $f(k) = \left(\frac{1}{k}\right)^{2-\eta}$  in the range of spatial frequencies up to a quarter of the spatial Nyquist frequency.

## 5.2 Wavelet analysis

For wavelet analysis we use complex-valued 2D Gabor wavelets corrected for zero mean with a smooth decline toward 0. Formulated for the complete pixelset  $J$ , the wavelet corresponding to frequency  $\omega$  and center at  $r_0$  consequently has the following functional form:

$$G_r^{(\omega, \varphi, r_0)} = e^{-\frac{(r-r_0)^2}{\sigma^2}} e^{-i(\omega(r-r_0) + \frac{\pi}{2})} \quad (5)$$

$$= \frac{e^{-\frac{(r-r_0)^2}{\sigma^2}} \left\langle e^{-\frac{(r-r_0)^2}{\sigma^2}} e^{-i(\omega(r-r_0) + \frac{\pi}{2})} \right\rangle_{r \in J}}{\sum_{r \in J} e^{-\frac{(r-r_0)^2}{\sigma^2}}},$$

where  $\omega = \omega \begin{pmatrix} \sin \varphi \\ \cos \varphi \end{pmatrix}$  is the spatial frequency vector.

The representation of the dependence of the vector  $\omega$  on  $\varphi$  ( $\omega = \omega(\varphi)$ ) is omitted for readability throughout the paper. The angle  $\varphi$  runs over six different angles,  $\omega$  over four different spatial scales:  $\varphi \in \Phi \equiv \{0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ\}$  and  $\omega \in \{\omega_n = (\sqrt{2})^{(n-5)} \text{ pixel}^{-1} | 0 \leq n \leq 3\}$ . The width of the Gaussian envelope is chosen depending on  $\omega$  to yield constant bandwidth:  $\sigma = 2.5\omega^{-1}$ . For computational efficiency a complete convolution of the Gabor with the image is avoided by multiplying  $G^{(\omega, r_0)}$  by  $I$  only at specific points, which form an evenly spaced grid  $\Gamma$  and are separated by 16 pixels along each dimension (for  $\omega_2$  we chose a slightly denser grid with 14-pixel distance):

$$\Gamma = \left\{ \mathbf{r} | \mathbf{r} = \begin{pmatrix} 16n \\ 16m \end{pmatrix} | n, m \in \{1, \dots, 11\} \right\} \quad (6.a)$$

or

$$\Gamma = \left\{ \mathbf{r} | \mathbf{r} = \begin{pmatrix} 14n \\ 14m \end{pmatrix} | n, m \in \{1, \dots, 13\} \right\}, \quad (6.b)$$

respectively.

We denote the amplitude of the product as  $L^{(\omega, \varphi, r_0)}$ :

$$L^{(\omega, \varphi, r_0)} = \left| \sum_{r \in J} G_r^{(\omega, \varphi, r_0)} I_r \right| \quad (7.a)$$

referred to as ‘‘local amplitude’’ and the phase

$$\Lambda^{(\omega, \varphi, r_0)} = \text{phase} \left( \sum_{r \in J} G_r^{(\omega, \varphi, r_0)} I_r \right) \quad (7.b)$$

referred to as ‘‘local phase’’.

The mean wavelet amplitudes are computed by averaging over time and/or space as described in the respective sections in the main text (Sects. 3.2–3.5). To examine the significance level of differences of mean wavelet amplitudes, a paired sign test is used. To obtain the incidences of the dominant orientation, we

calculate for each grid point  $r_0 \in \Gamma$  and time the  $\varphi \in \Phi$  for which the local amplitude  $L^{(\omega, \varphi, r_0)}$  reaches its maximum:

$$\varphi_D^{(\omega, r_0)} = \arg \max_{\varphi \in \Phi} \left( L^{(\omega, \varphi, r_0)} \right). \quad (8)$$

## 5.3 Correlation analysis

The spatiotemporal autocorrelations  $X_L^{(\omega, \varphi, r_0)}(\delta, \tau)$  and  $X_\Lambda^{(\omega, \varphi, r_0)}(\delta, \tau)$  of local amplitude and local phase, respectively, are calculated according to standard algorithms using MatLab’s *xcorr.m* function.  $\delta, \tau$  denote the spatial and temporal lag; phases are unwrapped using MatLab’s *unwrap.m* and amplitudes subtractively normalized to zero mean before computing the correlation.

## References

- Arakawa K, Tobimatsu S, Kurita-Tashima S, Nakayama M, Kira JI, Kato M (2000) Effects of stimulus orientation on spatial frequency function of the visual evoked potential. *Exp Brain Res* 131: 121–125
- Bell AJ, Sejnowski TJ (1997) The ‘‘independent components’’ of natural scenes are edge filters. *Vision Res* 37: 3327–3338
- Blake R (1988) Cat spatial vision. *Trends Neurosci* 11: 78–83
- Blakemore CB, Garey LJ, Vital-Durand F (1981) *J Physiol Lond* 319: 378
- Boltz RL, Harwerth RS, Smith EL (1979) Orientation anisotropy of visual stimuli in rhesus monkey: a behavior study. *Science* 205: 511–513
- Bosking WH, Zhang Y, Schofield B, Fitzpatrick D (1997) Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J Neurosci* 17: 2112–2127
- Burton GJ, Moorhead IR (1987) Color and spatial structure in natural scenes. *Appl Opt* 26: 157–170
- Chapman B, Bonhoeffer T (1998) Overrepresentation of horizontal and vertical orientation preferences in developing ferret area 17. *Proc Natl Acad Sci USA* 95: 2609–2614
- Coppola DM, White LE, Fitzpatrick D, Purves D (1998a) Unequal representation of cardinal and oblique contours in ferret visual cortex. *Proc Natl Acad Sci USA* 95: 2621–2623
- Coppola DM, Purves HR, McCoy AN, Purves D (1998b) The distribution of oriented contours in the real world. *Proc Natl Acad Sci USA* 95: 4002–4006
- Dan Y, Atick JJ, Reid RC (1996) Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *J Neurosci* 16: 3351–3362
- De Valois RL, Yund EW, Hepler N (1982) The orientation and direction selectivity of cells in macaque visual cortex. *Vision Res* 22: 531–544
- Dong DW, Atick JJ (1995) Statistics of natural time-varying images. *Network* 6: 345–358
- Dragoi V, Turcu CM, Sur M (2001) Stability of cortical responses and the statistics of natural scenes. *Neuron* 32: 1181–1192
- Eckert MP, Buchsbaum G (1993) Efficient coding of natural time varying images in the early visual system. *Phil Trans R Soc B* 339: 385–395
- Einhäuser W, Kayser C, König P, Körding KP (2002) Learning the invariance properties of complex cells from their responses to natural stimuli. *Eur J Neurosci* 15: 475–486
- Engel AK, König P, Kreiter AK, Singer W (1991) Stimulus-dependent neuronal oscillations in cat visual cortex: intercolumnar interactions as determined by cross-correlation analysis. *Eur J Neurosci* 2: 588–606

- Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. *J Opt Soc Am A* 4: 2379–2394
- Field DJ, Hayes A, Hess RF (1993) Contour integration by the human visual system: evidence for a local “association field”. *Vision Res* 33: 173–193
- Fregnac Y, Imbert M (1978) Early development of visual cortical cells in normal and dark-reared kittens: relationship between orientation selectivity and ocular dominance. *J Physiol* 278: 27–44
- Frien A, Eckhorn R (2000) Functional coupling shows stronger stimulus dependency for fast oscillations than for low-frequency components in striate cortex of awake monkey. *Eur J Neurosci* 12: 1466–1478
- Furmanski CS, Engel SA (2000) An oblique effect in human primary visual cortex. *Nat Neurosci* 3: 535–536
- Gilbert CD, Wiesel TN (1989) Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *J Neurosci* 9: 2432–2442
- Gilbert CD, Wiesel TN (1990) The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. *Vision Res* 30: 1689–1701
- Gray CM, König P, Engel AK, Singer W (1989) Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* 338: 334–337
- Guitton D, Douglas RM, Volle M (1984) Eye-head coordination in cats. *J Neurophysiol* 52: 1030–1050
- Hancock PJB, Baddeley RJ, Smith LS (1992) The principal components of natural images. *Netw Comput Neural Sys* 3: 61–70
- Hess RF, Dakin SC (1997) Absence of contour linking in peripheral vision. *Nature* 390: 602–604
- Hirsch HV, Spinelli DN (1970) Visual experience modifies distribution of horizontally and vertically oriented receptive fields in cats. *Science* 168: 869–871
- Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *J Physiol Lond* 160: 106–154
- Hughes A (1977) The topography of vision in mammals of contrasting life style: comparative optics and retinal organization. In: Crescitelli F (ed) *Handbook of sensory physiology*, vol vii/5. Springer, Berlin Heidelberg New York, pp 613–756
- Hyvarinen A, Hoyer P-O (2000) Emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspaces. *Neural Comput* 12: 1705–1720
- Kapadia MK, Ito M, Gilbert CD, Westheimer G (1995) Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron* 15: 843–856
- Keeble DR, Hess RF (1999) Discriminating local continuity in curved figures. *Vision Res* 39: 3287–3299
- Keil MS, Cristobal G (2000) Separating the chaff from the wheat: possible origins of the oblique effect. *J Opt Soc Am A Opt Image Sci Vis* 17: 697–710
- Kisvarday ZF, Toth E, Rausch M, Eysel UT (1997) Orientation-specific relationship between populations of excitatory and inhibitory lateral connections in the visual cortex of the cat. *Cereb Cortex* 7: 605–618
- König P, Engel AK, Löwel S, Singer W (1993) Squint affects synchronization of oscillatory responses in cat visual cortex. *Eur J Neurosci* 5: 501–508
- Körding KP, König P (2000) Learning with two sites of synaptic integration. *Network* 11: 25–39
- Krebs WK, Essock EA, Buttrey SE, Sinai MJ, McCarley JS (2000) An oblique effect of chromatic gratings measured by color-mixture thresholds. *Perception* 29: 927–935
- Krüger N (1998) Collinearity and parallelism are significant second order relations of Gabor wavelet responses. *Neural Proc Lett* 8
- Krüger N, Wörgötter F (2002) Multi-modal estimation of collinearity and parallelism in natural image sequences. *Netw Comput Neural Sys* 13: 553–76
- Law CC, Cooper LN (1994) Formation of receptive fields in realistic visual environments according to the Bienenstock, Cooper and Munro (BCM) theory. *Proc Natl Acad Sci USA* 91: 7797–7801
- Leventhal AG, Schall JD (1983) Structural basis of orientation sensitivity of cat retinal ganglion cells. *J Comp Neurol* 220: 465–475
- Li B, Peterson MR, Freeman RD (2003) Oblique effect: a neural basis in the visual cortex. *J Neurophysiol* 90: 204–217. (Epub 2003 Feb 26)
- Löwel S, Singer W (1992) Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. *Science* 255: 209–212
- Lythgoe JN (1979) *The ecology of vision*. Oxford University Press, New York
- Mansfield RJW, Ronner SF (1978) Orientation anisotropy in monkey visual cortex. *Brain Res* 149: 229–234
- Moeller GU, Einhäuser W, Conrath J, König P (2003) Exploring the stability of the retinal image in the freely behaving cat. *Soc. Neurosci. Abstr.* 386.5
- Nelson JJ, Kupersmith MJ, Seiple WH, Weiss PA, Carr RE (1984) Spatiotemporal conditions which elicit or abolish the oblique effect in man: direct measurement with swept evoked potential. *Vision Res* 24: 579–586
- Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381: 607–609
- Orban GA, Kennedy H (1981) The influence of eccentricity on receptive field types and orientation selectivity in areas 17 and 18 of the cat. *Brain Res* 208: 203–208
- Papoulis A (1991) *Probability, random variables, and stochastic processes*, 3rd edn. McGraw-Hill, New York
- Pasternak T, Horn K (1991) Spatial vision of the cat: variation with eccentricity. *Vis Neurosci* 6: 151–158
- Pettigrew JD, Nikara T, Bishop PO (1968) Responses to moving slits by single units in cat striate cortex. *Exp Brain Res* 6: 373–390
- Polat U, Norcia AM (1998) Elongated physiological summation pools in the human visual cortex. *Vision Res* 38: 3735–3741
- Reinagel P, Zador AZ (1999) Natural scene statistics at the centre of gaze. *Netw Comput Neural Sys* 10: 341–350
- Rizzo M, Nawrot M, Zihl J (1995) Motion and shape perception in cerebral akinetopsia. *Brain* 118: 1105–1127
- Ruderman DL, Bialek W (1994) Statistics of natural images: scaling in the woods. *Phys Rev Lett* 73: 814–817
- Schmidt KE, Goebel R, Löwel S, Singer W (1997) The perceptual grouping criterion of colinearity is reflected by anisotropies of connections in the primary visual cortex. *Eur J Neurosci* 9: 1083–1089
- Sengpiel F, Stawinski P, Bonhoeffer T (1999) Influence of experience on orientation maps in cat visual cortex. *Nat Neurosci* 2: 727–732
- Switkes E, Mayer MJ, Sloan JA (1978) Spatial frequency analysis of the visual environment: anisotropy and the carpentered environment hypothesis. *Vision Res* 18: 1393–1399
- Tolhurst DJ, Tadmor Y, Chao T (1992) Amplitude spectra of natural images. *Ophthalmic Physiol Opt* 12: 229–232
- Van der Schaaf A, van Hateren JH (1996) Modelling the power spectra of natural images. *Vision Res* 36: 2759–2770
- Van Hateren JH (1993) Spatiotemporal contrast sensitivity of early vision. *Vision Res* 33: 257–267
- Van Hateren JH, Ruderman DL (1998) Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proc R Soc Lond B Biol Sci* 265: 2315–2320
- Wässle H, Levick WR, Cleland BG (1975) The distribution of the alpha type of ganglion cells in the cat’s retina. *J Comp Neurol* 159: 419–438
- Wiskott L, Sejnowski TJ (2002) Slow feature analysis: unsupervised learning of invariances. *Neural Comput* 14: 715–770
- Zeki S (1991) Cerebral akinetopsia (visual motion blindness): a review. *Brain* 114: 811–824