Maximizing Causal Information of Natural Scenes in Motion

Dawei W. Dong

Center for Complex Systems and Brain Sciences
Florida Atlantic University
Boca Raton, FL 33431

Email: dawei@ccs.fau.edu

(*In: Dynamics of Visual Motion Processing, Ilg UJ and Masson GS eds* pp 261-282)

Abstract

Natural scenes contain a huge amount of information if counted by spatial pixels and temporal frames. However, most of the information is redundant because the pixels and the frames are highly correlated. The optical flow, generated by motions of the objects and the observer, contributes significantly to the statistical regularity of such spatiotemporal correlations. The visual system of an animal such as the human is highly adapted to this statistical regularity such that the visual sensitivity follows the same contours as the spatiotemporal correlations of natural scenes in motion, in particular, along two axes: space and motion instead of space and time. Furthermore, vision is an active process, during which eye movements interact with visual scenes and select images that arrive on the retina: pursuits and fixations on objects significantly alter the image velocity distributions on the fovea and the periphery, which lead to the dependence of the visual sensitivity on the retinal eccentricity; saccades between objects change the natural scene statistics dynamically, which lead to the dependence of the visual sensitivity on the time relative to saccades. All of these can be accounted for by the proposed ecological theory that the visual system maximizes the causal information of the natural visual input.

1 Introduction: an approach based on information theory

Animals live in a dynamical world. Everything changes and fluctuates. It seems that nothing is certain but the passage of time. Yet, the physical world, however dynamic and probabilistic, does have some causal connections from its past to its future, governed by the law of physics. A fundamental function of the brain is to afford an animal some ability of knowing the future from the past, i.e., to acquire the causal information.

Through the sensors which are essential for the survival of an animal species, an individual who knows more about the future has an obvious evolutionary advantage over those who know less. Due to the natural selection, it is conceivable that the neural system which processes the information from those sensors tends to maximize the causal information — and equivalently, to minimize the non-causal information — for given physiological constraints on the processing capacity.

The hypothesis of this paper/chapter is that maximizing causal information is an organization principle of the visual system, in particular, in dynamically processing the motion information. To apply this principle, we first gain a good understanding of the statistical properties of the natural visual input to the brain. We then reveal how the visual processing depends on the statistics of the natural visual input. We apply this approach in two cases: first, to the stationary case in which we relate the average natural scene statistics to the average visual sensitivities; second, to the non-stationary case in which we reveal how the statistics of the natural visual input and the visual processing depend on scenes and eye movements.

2 Statistics of natural scenes: scaling to motion

In recent years, quantitative measurements have been conducted extensively on the statistical properties of the visual input — in particular, the images "natural" for the human visual system — in many aspects, such as color (Webster and Mollon 1997; Parraga et al. 1998; Ruderman et al. 1998; Tailor et al. 2000), stereo (Li and Atick 1994; Hoyer and Hyvarinen 2000), space (Burton and Moorhead 1987; Field 1987, 1994; Ruderman and Bialek 1994; Olshausen and Field 1996), and space-time (Dong and Atick 1995a, VanHateren and Ruderman 1998). The discovery most relevant to this paper is that the Fourier transform of the spatiotemporal correlation matrix, or the power spectrum, of natural scenes, is a non-separable function of spatial and temporal frequencies and exhibits an interesting scaling behavior:

$$R(f, w) \sim f^{-m-1} P(w/f)$$

where f and w are the spatial and the temporal frequencies, $-m$ is the exponent of the spatial power spectrum $\sim f^{-m}$, and $P(v)$ is the probability density distribution function of the optical flow velocity v (Dong 2001a).

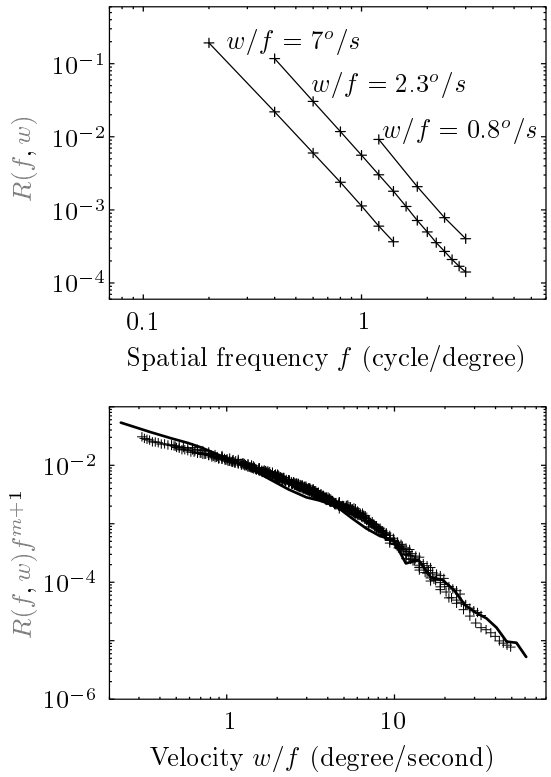

Figure 1: Scaling behavior of spatiotemporal power spectrum. **Top:** the power spectrum is plotted for three velocities — ratios of temporal and spatial frequencies — (0.8, 2.3, 7) degree/second. The curves have the same shape $\sim f^{-m-1}$. **Bottom:** the power spectrum is replotted as a function of w/f after multiplication by f^{m+1}, with $m = 2.3$. All the data points fall on a single curve. The solid curve is the probability density distribution of the optical flow velocity (adapted from Dong 2001a).

To see the scaling behavior, the measured power spectrum is plotted in Figure 1 (top) as a function of f for different w/f ratios. The curves are just a horizontal shift from each other and all of them follow a straight line in the log-log plot. In fact, when multiplying the spectrum by $m + 1$ power of f, i.e., when plotting $f^{m+1}R(f, w)$ as a function of w/f, all curves coincide very well, as shown in Figure 1 (bottom).

It has been shown earlier that such spatiotemporal correlation can be derived from the first principles, under the assumption that the motions of objects relative to the observer follow a certain velocity distributions (Dong and Atick 1995a). Supporting this physical explanation is the clear agreement between the scaled power spectrum

$f^{m+1}R(f,w)$ and the measured optical flow velocity distribution (Dong 2001a) shown together in Figure 1 (bottom).

Of course, this physical scenario has correlations higher than the pairwise second order correlation (or equivalently the spatiotemporal power spectrum). Then, what is the reason to study the second order corrlation? The second order correlation matrix, which should be thought of as a constraint on the joint probability distribution, is the simplest quantity that captures the statistical dependency in space and time of natural time-varying images. Furthermore, we believe that the second order correlation is the quantity that neurons the early stages of the visual system are able to evaluate and take advantage of in recoding the visual input (see Section 3 and Section 4).

This constraint suggests that the spatial and temporal correlations arise predominantly from motions and in general, the spatial and temporal parts of such correlations are not separable and hence visual processing cannot be fully characterized in space independently of time. The scaling behavior suggests that a natural way to separate the correlations is in space and motion — just as Newtonian (Hamiltonian) physical systems should be. More interestingly, it suggests that a better way to examine spatiotemporal tuning data from visual systems is to plot the data not as a function of f and w separately but as a function of f and w/f — if the visual processing is indeed adapted to the natural scene statistics. In this representation we expect that neural responses will exhibit more universal behavior.

3 Scaling of human visual sensitivity

We will show in the next section the theoretical connection between the spatiotemporal power spectrum of natural scenes and the response properties of visual systems. But before showing the theory, we will first show the scaling behavior of the visual sensitivity, which is strikingly similar to the scaling behavior of the natural scenes.

Figure 2 (top) shows the human visual sensitivity to spatiotemporal gratings, determined by the threshold modulation amplitude I_t. Here, instead of the usual way of plotting along spatial and temporal frequencies, the curves are plotted along the spatial frequency and the velocity, i.e., the ratio of the temporal and the spatial frequencies (Kelly 1979). Clearly, the curves for different velocities have similar shapes. In particular, the rising slopes of the curves in the log-log plot are half of $m+1$, which ensure that the natural power spectrum with a slope of $-(m+1)$ (for a constant f/w ratio, see Figure 1, top) is whitened. It was shown that the sensitivity curves approximately coincide with each other when shifted by a factor of $P(f/w)$ according to the statistics of natural scenes (Dong 2001a). This scaling behavior in human

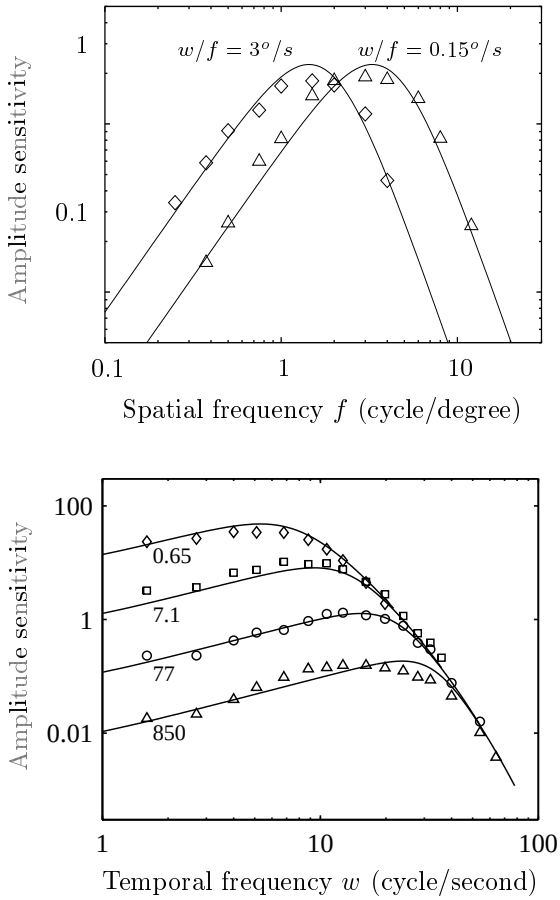

Figure 2: **Top:** the spatiotemporal visual sensitivity. The amplitude sensitivity data ($1/I_t$, in the unit of trolands^{-1}, Kelly 1979) are plotted as a function of spatial frequency for two w/f ratios — 0.15, 3 degree/second (at a fixed mean light level of 300 troland). The curves at low spatial frequencies have the positive slopes which exactly decorrelates the natural image power spectrum. **Bottom:** the temporal visual sensitivity adapting to the mean light level. The amplitude sensitivity data ($1/I_t$, in the unit of trolands^{-1}, Kelly 1961) are plotted for four different mean light levels — 0.65, 7.1, 77, 850 trolands (at a very low spatial frequency). The curves at low temporal frequencies have the positive slopes which exactly decorrelates the natural image power spectrum. The solid lines are the theoretical prediction.

visual sensitivity shows the predominant role of image motion in visual processing as suggested by the natural scene statistics shown in Section 2.

Figure 2 (bottom) shows a non-stationary aspect of visual processing: the human visual sensitivity adapting to a simple scene statistics, I_0, the mean light level. At a given temporal frequency which is relatively low, the amplitude sensitivity is inversely proportional to I_0. Furthermore, the rising slopes of the curves in the log-log plot are approximately 1, which decorrelate (whiten) the natural scenes with the velocity distribution shown in Figure 1 (bottom) which has the asymptotic slope near -2 for reasonably high velocities. Because the sensitivity is inversely proportional to the mean light, the decorrelation is maintained at all mean light levels for relatively low spatial and temporal frequencies shown in Figure 2.

Next, let's see how we can make a theoretical connection between the scaling to motion in the statistics of natural scenes and the scaling of visual response properties shown in human visual sensitivity.

4 Maximizing causal information

It is worth to review some earlier approaches based on the information theory (Shannon and Weaver 1949). The brain has to process an enormous amount of information from the sensory input. It is reasonable to expect that neurons in the sensory pathways developed to take advantage of certain regularities and statistical properties in the sensory input to build more efficient representations of the world (Attneave 1954; Barlow 1961; Srinivasan et al. 1982; Linsker 1988; Atick and Redlich 1990). There has been a vast amount of evidence showing the connection between the properties of natural stimuli and the properties of the sensory systems, and it is quite clear that better characterization of the properties of natural signals leads to deeper understanding of neural functions (see references in Atick 1992; Dong and Atick 1995b; Simoncelli and Olshausen 2001; Zhaoping 2002).

One fundamental assumption common to most of the earlier works is that one needs to know what information/noise is **before** applying a theory (see Linsker 1989, Atick and Redlich 1990, VanHateren 1992). However, in this paper, we propose a theory which will discover what information/noise is **when** applied to the visual system. We will describe the theory briefly on the conceptual level without going into mathematical details.

At any given moment of time, let's divide the visual input S into two parts: S_- and S_+ for the past (occurred) and the future (incoming) visual input, respectively. Because of the causality, the visual system can only process and represent its input

based on S_-. Our hypothesis is that the goal of the visual processing is to transform S_- into an internal representation O such that O contains the maximum amount of information about S_+ (let's denote this information as $I(O, S_+)$). This requirement can be easily fulfilled if there is no limitation to the maximum capacity of the internal representation: just simply let $O = S_-$, so $I(O, S_+) = I(S_-, S_+)$ which contains all the information about the future from the past. However, this representation is redundant and inefficient since S_- has the spatiotemporal correlations as shown in Section 2. Given a limitation to the capacity (let's denote it as $C(O)$), one can build a better representation $O = K(S_-)$ by choosing a transfer function K which maximizes $I(O, S_+)$ while minimizing $C(O)$ or keeping $C(O)$ constant. In this representation, O will be more independent (decorrelated).

Both $I(O, S_+)$ and $C(O)$ are well defined in terms of the information theory for a given statistical ensemble of visual input S and a given transfer function K. In this chapter, we will restrict ourselves to the class of linear transfer functions and to the second order statistics.

If natural scenes had been changing randomly over time, the power spectrum would be flat in time and the causal information $I(O, S_+)$ would be zero, i.e., everything would be noise, i.e., one could not predict the future from the past. However, natural scenes are dominated by objects which move according to Newtonian (Hamiltonian) dynamics, and as a result, the temporal power spectrum is not flat — in fact, dominated by a characteristic velocity distribution, which give rise to the predictability.

Using the natural scene power spectrum R as shown in Section 2, one can derive the optimal linear transfer function (filter) K in the Fourier domain for a given capacity C: $K(f, w) = K(R, C)$.* Since the power spectrum is the product of the spatial power $f^{-(m+1)}$ and the velocity distribution $P(w/f)$, for a given $w/f = v$, the transfer function K is the same in the form of $K({f_s}^{-(m+1)}, C)$ — in which f_s is the spatial frequency f scaled by a factor of $P(v)^{-\frac{1}{m+1}}$. This explains why both the power spectrum and the visual sensitivity have the same scaling behavior. Furthermore, for low spatial frequencies, $K \sim {f_s}^{\frac{m+1}{2}}$, such that $|K|^2 R$ is constant, i.e., the output is whitened or decorrelated. The amplitude of the derived K is shown in Figure 2 as solid lines to compare with human visual sensitivity. One immediately sees the good agreement between the theoretical prediction and the experimental data.

The derived K not only covers different spatial and temporal frequencies but also holds for different background mean light intensities (Figure 2, bottom) where the

*This is true when the spatial and temporal frequencies f and w are low. When f and w are high, the optical transfer function will have an impact (see Atick 1992 and VanHateren 1992).

sensitivity K is inversely proportional to the mean light level. The maximization determines the spatiotemporal receptive field as well. The resulting receptive field is similar to the minimum phase solution (Dong and Atick 1995b), and is shown in Figure 3. This filter does decorrelate the natural visual input on average.

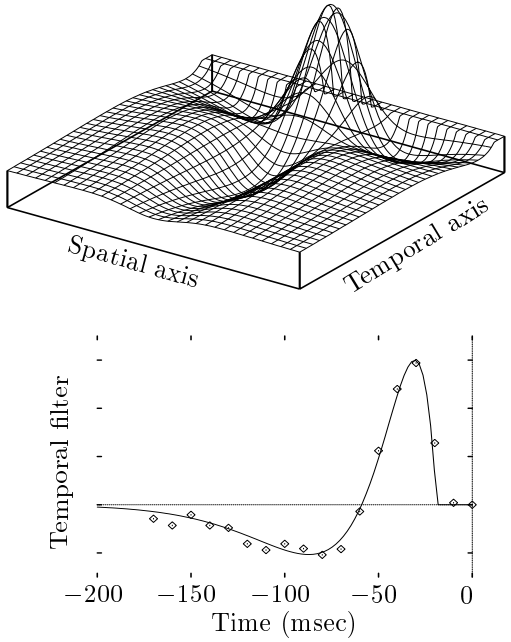

Figure 3: **Top:** the predicted spatiotemporal receptive field. **Bottom:** the temporal filter $K(t)$. The data (diamond symbol) are calculated through the reverse correlation of the spike train and the stimuli at the center of the spatial receptive field of a cat LGN cell (Cai et al. 1997). The solid line is the theoretical prediction, sliced through the spatial center of the spatiotemporal receptive field on the top (note: there is 20 msec time delay from visual stimuli to the LGN).

5 Statistics of the visual input on the retina

So far we did not take into account a very important aspect of vision. Vision is an active process. In the natural environment of animals such as humans, eye movements interact with visual scenes and select images that arrive on the retina. It is very important to study the statistical properties of such images on the retina, even for the averaged or stationary properties; since after all, those images are the ones which constitute the visual input to the brain. Furthermore, there are two important non-stationary aspects of the visual input: first, the eye movement itself is a state variable, it is important to investigate the dependency of the input statistics on eye movements; second, different scenes have different intrinsic properties and interact differently with

eye movements, it is important to investigate the dependency of the visual input statistics on scenes.

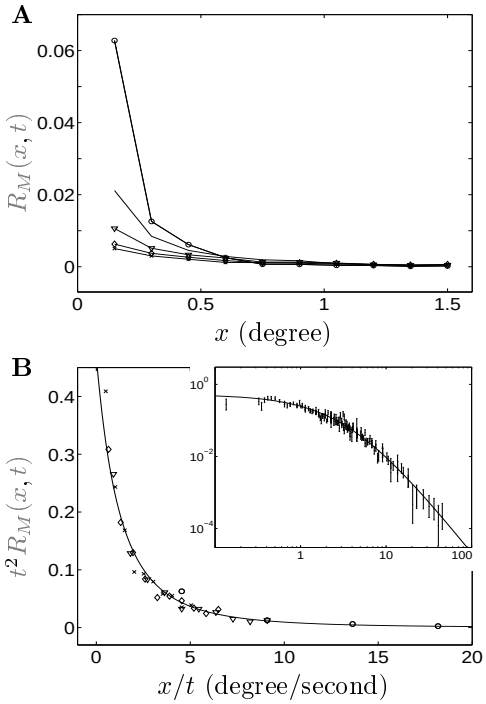

Figure 4: Measured spatiotemporal correlation of the contrast signal $R_M(x,t)$. **(A)** the curves in circle, square, triangle, diamond, and cross, for $t = 33$, 100, 167, 233, and 300 msec, respectively. (**B**) all the data points in (A) re-plotted as a function of x/t after multiplication by t^2. Inset: $t^2R_M(x,t)$ plotted in the log-log scale with standard error bars for all data points of $x < 1.5$ deg and $t < 300$ msec. Also plotted in (B) and the inset is the curve $P(x/t)$, while $P(v)$ is the best fit curve for the velocity distribution.

The interest in investigating the properties of such visual input to the brain dates back to the early days when eye movements were first recorded as human subjects explored various natural images (Yarbus 1967). But the images used then were static scenes. Natural images also change over time due to the movements of objects in a scene. Pursuits of moving objects as well as saccades between them are essential to the active visual process. It was shown in simulation that smooth tracking of objects could significantly alter the image velocity distributions on the fovea and the periphery (Eckert and Buchsbaum 1993). It was also argued that small eye movements between saccades (drifts and micro-saccades during pursuits and fixations) are essential for processing visual input when objects are otherwise stationary on the retina (Hubel 1988) and for generating input correlations necessary to the development of visual systems (Rucci et al. 2000). Yet partially due to technical difficulties, direct experimental measurements of the statistical properties of the natural visual input

to the brain, i.e., the natural time-varying images on the retina during free-viewing, have been wanting.

Recently, the statistics of natural time-varying images were studied with an account for eye movements (Dong 2001b). The recorded eye positions during free viewing were used to derive the images on the retina and to serve as state variables. For several different visual scenes, each of which was a large segment of natural time-varying images, the velocity distribution and the spatiotemporal correlation function of the images on the retina were measured. It confirmed the earlier result that the correlation function exhibits a scaling behavior in space and time, which is related to the spatial correlation function and the velocity distribution. Based on this scaling behavior, a spatially decorrelated representation (called *contrast signal* in this paper, see Appendix) was derived, which completely separates the effects of the spatial correlation function and the velocity distribution (see Figure 4) and is particularly useful for studying non-stationary statistics. It was shown that both visual scenes and eye movements give rise to non-stationary statistics. In the following sections we will summarize the non-stationary statistics and relate those to the non-stationary visual processing.

6 Non-stationary velocity distribution on the retina

During pursuits and fixations, the optical flow velocity distribution is different for different retinal eccentricities and for different visual scenes. The average optical flow velocity is lower on the fovea than on the periphery. On the fovea, more dynamic scenes have higher averages of optic flow velocity.

Figure 5 (top) shows the optical flow velocity distributions for two different scenes. Although both scenes have power law tails for high velocities, the slopes in the log-log plot are different and the average velocities are different. The more dynamic scene (picking fruits in a forest) has a longer tail for high velocities and lower probabilities at low velocities than the less dynamic scene (watching birds on a river bank). The average velocities are different by a factor more than four. The visual system has to be able to deal with a wide range of retinal image velocities due to the non-stationary velocity distribution of visual scenes.

At the bottom of Figure 5, the distributions of the average velocity for five scenes are shown for the fovea (the horizontal axis) and for 3 degrees away from the fovea (the vertical axis). This shows one more aspect of the non-stationary velocity distribution: the difference due to retinal eccentricity. For all five scenes, the average optical flow velocity is higher for more peripheral locations, indicated by the data points

clustering above the diagonal line. This is a direct confirmation that the efficient coding requires that the visual sensitivity to higher velocities increases with increasing retinal eccentricities (Eckert and Buchsbaum 1993).

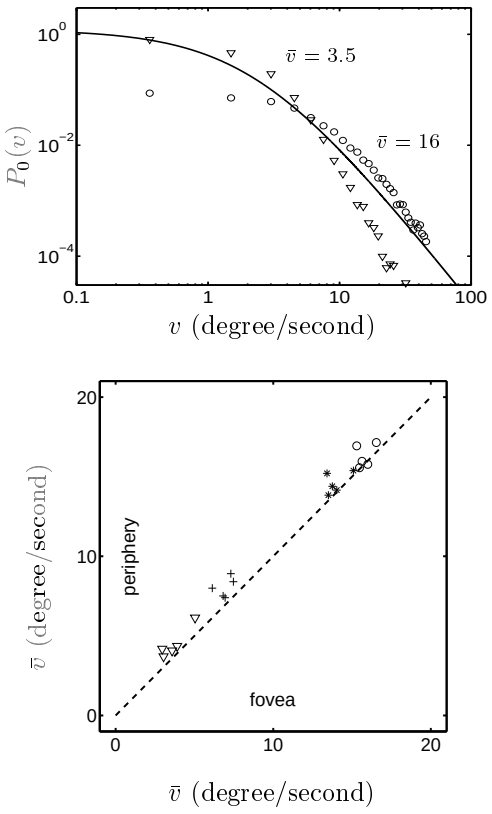

Figure 5: **Top:** the measured probability density distribution of the retinal image velocity $P_0(v)$ for two different scenes (triangle: watching birds on a river bank, circle: picking fruits in a forest). Also plotted is $P_0(v) = \frac{v_o}{\pi(v+v_o)^3}$ curve with the best fit $v_o = 2.3$ deg/sec for the velocity distribution averaged over five scenes. **Bottom:** the averages of the retinal image velocity $\bar{v}$ of the fovea (horizontal) versus the periphery (vertical). Different symbols are for different scenes and different data points of the same symbol are for different observers. The average retinal image velocity at 3 degrees away from the fovea (the vertical axis) is clearly higher than on the fovea (the horizontal axis) for all scenes.

7 Non-stationary correlations of the contrast signal

The eye movements have two effects: the spatiotemporal correlation is greatly reduced by saccadic eye movements and the velocity distribution between saccades is more concentrated in the regime of lower velocities. These findings reveal some important aspects of how information is processed by the sensory systems with the active

participation of the motor systems.

We use $R_1(x,t)$ to denote the correlations of two contrast signals with one saccade in between and $R_0(x,t)$ to denote the correlations of two contrast signals without any saccade in between. Figure 6 shows those two correlation functions (scaled by t^2 as before) of the contrast signal. From Figure 6, it is clear that

$$R_0(x,t) >> R_1(x,t) \quad \text{and} \quad R_1(x,t) \sim 0 \tag{1}$$

for the range of x/t from zero to several degrees per second. Within this range, the contrast signals at different times are uncorrelated if there is a saccade occurring in between, but correlated if there is none.

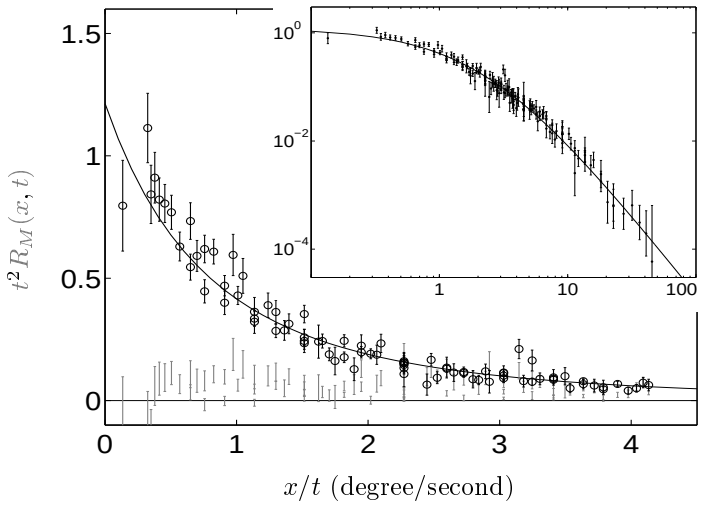

Figure 6: Measured spatiotemporal correlation functions R_1 (with saccades) and R_0 (without saccades). The scaled correlation functions $t^2R_1(x,t)$ and $t^2R_0(x,t)$ in square and circle, respectively, are shown for a small x/t range. Inset: $t^2R_0(x,t)$ is shown in the log-log plot with standard error bars for all data points of $x < 1.5$ deg and $t < 300$ msec. Also plotted in both is the curve $P_0(x/t)$, where $P_0(v)$ is the best fit curve for the velocity distribution without any saccade in between.

The other important effect of eye movements is the concentration of $R_0(x,t)$ toward the region of lower x/t. This is shown in Figure 6 (inset) in a log-log plot, in comparison to Figure 4B (inset). Also plotted in both Figure 6 and its inset is the $P_0(x/t)$ curve, which has the same shape as $P(x/t)$ in Figure 4B with the asymptotic power law of an exponent 3, but has a smaller average velocity and thus a higher peak around zero.

It has been proposed that the first stage of visual processing (retina) removes the spatial correlation and that the retinal output is the contrast signal (Atick and Redlich

1990, 1992). It is clear that saccades remove the spatiotemporal correlations of the contrast signal. However, the contrast signal still has spatiotemporal correlations between saccades. It has been proposed that the remaining correlations are removed by the second stage of visual processing (LGN) (Dong and Atick 1995b). Dan et al. (1996) directly verified the decorrelation theory of the LGN. They stimulated paralyzed and anesthetized cats with natural movies. However, since the correlations are non-stationary for awake animals, the LGN processing also has to be non-stationary to maintain decorrelation and efficient coding (Dong 2001b; Truccolo and Dong 2001). Again, this can be derived from maximizing causal information but once more again, let's take a look at the experimental facts first.

8 Eye movements and saccadic modulation in the LGN

Lee and Malpeli (1998) have recently studied the excitability of LGN neurons as the cat made saccadic eye movements under light and dark conditions. They found that a period of suppression began about 200 msec prior to the saccade, peaked about 100 msec prior to the saccade onset and quickly reversed, so that there was a peak of facilitation about 100 msec after the saccade end. As this function persisted in the dark, oculomotor signals must have manipulated LGN gain. Potential sources of these eye movement signals include sensory signals from muscle afferents in the brainstem (Donaldson and Dixon 1980; Lal and Friedlander 1990ab), corticogeniculate input (Swadlow and Weyand 1987), and pretectal influence (Schmidt 1996). Other reports also indicate that eye movement signals, via corollary discharge or muscle afferents, reach the LGN (see the review by Buisseret 1995).

More than one experiment showed that excitability around the time of eye movement changes because extra-retinal influences appear to alter the gain of retinogeniculate processing during the peri-saccadic interval (Guido and Weyand 1995; Lee and Malpeli 1998; Ramcharan et al. 2001; Reppas et al. 2002). Such phenomenon contributes not only to saccadic suppression, an attenuation of the ability to detect certain stimuli before and during the time of eye movements, but also to saccadic disinhibition, an increased excitability of the LGN after the end of a saccade. Functionally, such suppression would minimize the "blurring" effects caused by the retina rapidly sweeping by a textured environment and such disinhibition would allow the free flow of visual information at the new gaze location (Singer 1977).

Guido and Weyand (1995), Lee and Malpeli (1998) and Ramcharan et al. (2001) have noted a relatively increased probability of high frequency "bursting" following saccadic eye movements. Since these bursts are linked to the low-threshold calcium

conductance, a period of hyperpolarization must precede the burst. Our recent theoretical study showed that the hyperpolarization — even with a small amount of burst spikes — can change the temporal filtering of the LGN to help improve efficiency (Truccolo and Dong 2001). From the efficient coding point of view, a change of temporal filtering properties according to eye movement can help the visual system to get more information (see Section 9).

9 Maximizing causal information of non-stationary statistics

Although in Section 4 we apply the efficient coding theory to stationary statistics, there is no reason to believe that the theory should not be applicable to non-stationary statistics. In fact, the visual system has short-term adaptations, on the order of tens of seconds, which acts to reduce dependencies between neurons (Atick et al. 1993; Dong 1994; Webster and Miyahara 1997; Carandini and Ferster 1998; Dragoi et al. 2000). Furthermore, individual neurons adapt to changes in visual input statistics on very short time scales, on the order to seconds (Smirnakis et al. 1997; Muller et al. 1999; Brenner et al. 2000).

If the statistical properties of the visual input for animals are non-stationary, a dynamic coding is needed for the decorrelation. A dynamic coding — receptive field — adapting to non-stationary statistics, such as those generated by the scene changes and the eye movements, will be predicted from the efficient coding theory, which has been recently formulated for such fast changing, non-stationary input statistics (Dong 2001b). In this paper, the general efficient coding principles are applied to make predictions about the specific dynamic neural processing related to saccadic eye movements (but see also Section 4 for different mean light levels).

To demonstrate the method in the non-stationary case, we will derive the non-stationary optimal linear transfer function (filter), which takes into account the timing of the eye movement. Since the eye-movement information is assumed to be available to LGN cells, the relative timing τ to eye movements is taken into K, specifically, the temporal filter will be different depending on the time to the nearest saccade (note: long before saccade or long after saccade, the temporal filter will be the same). For the demonstration purpose, we also make a simplified assumption that the LGN only does temporal filtering of its input from the retina. So our LGN temporal filter $K = K(t, \tau)$ relates the LGN input $S(t)$ to the LGN output $O(\tau)$ by

$$O(\tau) = \sum_{t \leq \tau} K(t, \tau) S(t) \tag{2}$$

in which the summation is over t for $t \leq \tau$ (the causality). It is important to understand the meaning of $K(t, \tau)$: it is the temporal filter of the LGN at time τ relative to the nearest saccade (note: $\tau < 0$ is before saccade, $\tau > 0$ after saccade). Because of the causality, $K(t, \tau) = 0$ for any $t > \tau$.

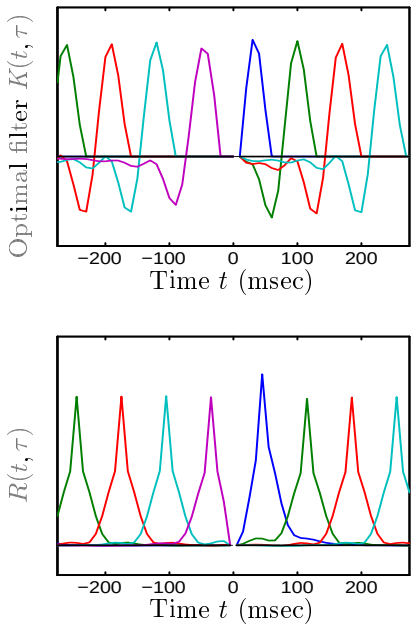

Figure 7: The predicted optimal temporal filter $K(t, \tau)$ (top) and the corresponding output correlation $R(t, \tau)$ (bottom). The filter $K(t, \tau)$ at eight different τ are plotted, in which τ is the time relative to the saccade (τ = -245, -175, -105, -35, 35, 105, 175, 245 msec from left to right). It is clear that the optimal temporal filter of the LGN depends on the time relative to saccades. The output correlations $R(t, \tau)$ are close to zero except near $t = \tau$.

A solution is shown in Figure 7 (top). The introduction of reference time τ can achieve better coding on the time scale of intervals between saccades, which is several hundred milliseconds on average. This makes the predicted temporal filter very dynamic. During and right after a saccade, the predicted receptive-field behaves as a temporal low-pass filter; whereas between two saccades, the predicted receptive-field behaves as a temporal difference (band-pass) filter. The predicted filter during saccades has a smaller response to a stimulus at certain midrange temporal frequencies, as expected from saccadic suppression. However, it has bigger responses for stimuli at relatively low and relatively high temporal frequencies. As a result, the LGN output decorrelation is maintained all the time (Figure 7, bottom).

10 Concluding remarks

The paper/chapter shows that the spatiotemporal statistics of the visual input to the brain arise from the coupling of two separate sources: the spatial statistics and the retinal image motion statistics. This indicates that the visual world is spatial scale invariant and objects at different distances move with certain velocity distributions. Furthermore, the paper shows that the statistics are non-stationary and there are two major non-stationary effects, eye movements and visual scenes, both significantly affecting the retinal image motion statistics. This paper proposes that a fundamental principle of sensory coding is to maximize the causal information. Consequently, the visual processing is shaped according to the statistics of natural scenes in motion. These findings reveal how information is processed by the visual-sensory system with the active participation of the ocular-motor system, in an "optimal" way by maximizing causal information.

The main effect of the observed saccadic eye movements is to bring different parts of a visual scene into the fovea for processing. The fovea has the highest density of cones — the photoreceptors which operate under normal daylight conditions. The size of human fovea is 5.2 deg in diameter (1.7 deg for rod-free fovea) (Wandell 1995). The average saccade amplitude during the natural viewing is 2.7 ± 1.7 deg — the right amount for bringing one part of a visual scene from right outside the fovea into the fovea. As shown in Figure 6, those saccades greatly reduce the spatiotemporal correlation of the visual input in the space-time range where the human fovea has the highest sensitivity: the optical flow velocity or x/t from 0.2 to 3 deg/sec and the spatial separation x from 0.1 to 0.5 deg (i.e. the spatial frequency from 5 to 1 cycle/deg) (Kelly 1979). Thus the saccades help to reduce the redundancy of the input, i.e., to increase the information flow through the fovea, during the natural viewing.

To code the natural visual input efficiently, the visual system needs to distinguish between the spatiotemporal changes generated by the motion of objects in a scene and the spatiotemporal changes generated by the motion of the observer — in particular saccadic eye movements. Since saccadic movements are predictable to the observer and do not carry information about the scene, visual responses containing information about saccadic movements are not efficient. Retinal responses generated by the model shown in Section 4 or the more realistic model using a spatio-temporal retinal filter derived from various experiments (Wehmeier et al. 1989) inevitably contain information about saccades and thus are less efficient than the proposed LGN model shown in Section 9. The latter changes according to saccades in order to be

more efficient and thus contains less information about saccades. Recent experiments support this prediction (Dastjerdi et al. 2003; Dastjerdi 2007).

The other eye movements between saccades (drifts and micro-saccades during pursuits and fixations) do generate the retinal image statistics, especially the velocity distributions, which are suitable for the visual system to process. It is well known that images stabilized on the retina fade away after about a second (Riggs and Ratliff 1952; Ditchburn and Ginsborg 1952). It was hypothesized that eye movements between saccades help to change the retinal image characteristics to match the properties of the neurons in the early visual pathway to facilitate information processing, Hubel (1988). The measurements in this paper show that eye movements between saccades do increase (comparing Figure 4B and Figure 6) the input signal in the velocity range where the human fovea has the highest sensitivity: the retinal image velocity or x/t from 0.2 to 3 deg/sec (Kelly 1979). However, the measurements also show that for different scenes, the retinal image statistics are different, which leave the possibility that receptive fields dynamically adapt to different scenes to process the visual input better.

It is conceivable that with the theoretical consideration outlined in this chapter/paper, carefully designed experiments can improve our understanding of sensory systems enormously by using natural stimuli and taking into account the non-stationary aspects. After all, the dynamic nature of the brain makes it more powerful in information processing in the natural environment. In particular, for the visual system, natural time-varying images should be used to probe visual responses, and the eye movements of awake/roused animals should be recorded not only to register images on the retina but also to be counted as state variables. Doing so will truly help understand the dynamics of visual motion processing.

11 Appendix

11.1 Calculation of spatiotemporal correlation

In this paper/chapter, the following conventions are used: given light intensity $S(\mathbf{x}', t')$, the correlation between two points separated by spatiotemporal distance $\mathbf{x}, t$ is

$$R(\mathbf{x}, t, t') = \langle S(\mathbf{x} + \mathbf{x}', t + t')S(\mathbf{x}', t')\rangle_{\mathbf{x}'} \tag{3}$$

in which the $\langle\rangle_{\mathbf{x}'}$ denotes the average over the spatial location $\mathbf{x}'$. For an ergodic system, the expectation value of $S(\mathbf{x}+\mathbf{x}', t+t')S(\mathbf{x}', t')$ is independent of $\mathbf{x}', t'$, but in general, it is a function of the retinal eccentricity and relative timing to the saccades

(see "Classification of eye movements"). In this paper/chapter, the focus is on the statistics near the fovea, so it is averaged over the $4^o \times 4^o$ region around the center of gaze.

When the dependence on the absolute timing is ignored

$$R(\mathbf{x}, t) = \langle \langle S(\mathbf{x} + \mathbf{x}', t + t') S(\mathbf{x}', t') \rangle_{\mathbf{x}'} \rangle_{t'} \tag{4}$$

in which the second $\langle \rangle_{t'}$ is the average over t'. This equation is used to measure the average or stationary statistics.

To understand the effect of saccades on the retinal image statistics, the correlation is calculated around the saccades. In this case, the time t' of the signal $S(x', t')$ is measured by the advance time τ to the onset of next saccade at time s. So the time $t' = s - \tau$. Ignoring the dependence on the absolute timing but accounting for the dependence on the relative timing to saccade,

$$R(\mathbf{x}, t, \tau) = \langle \langle S(\mathbf{x} + \mathbf{x}', t + t') S(\mathbf{x}', t') \rangle_{\mathbf{x}'} \rangle_{s} \tag{5}$$

in which the second $\langle \rangle_s$ is the average over all saccades. This equation is used to measure the non-stationary statistics.

The light intensity $S(\mathbf{x}, t)$ in all the equations is mean-removed and normalized, such that the $R(0, 0) = 1$.[†] For the purposes of this paper/chapter the dependence on spatial orientation is ignored and all calculations are averaged over all orientations, thus only one spatial dimension x — the radial distance — is shown in all the figures. All the calculations are also averaged over viewers and scenes, unless mentioned otherwise. Also used in all calculations are the natural units of the measurements: 1 spatial unit is 0.15 deg, 1 temporal unit is 33 msec, and 1 velocity unit is 4.5 deg/sec. But for the convenience of the readers, the units of degree and second are used in all illustrations.

11.2 Calculation of contrast signal

To further reveal the nature of the spatiotemporal correlation of the visual input, the retinal input is spatially whitened using the localized whitening filters (Atick and Redlich 1992; Atick et al. 1993; Penev and Atick 1996) calculated from the measured spatial correlation function as the following.

The spatial correlation function is calculated first by

$$R_s(\mathbf{x}, \mathbf{x}') = \langle S(\mathbf{x}, t) S(\mathbf{x}', t) \rangle_t \tag{6}$$

[†]The illumination does not affect the results in a wide range. see Conclusion and Discussion.

in which the $\langle\rangle_t$ is the average over time. Obviously it is positive definite, so all the eigenfunctions $\{E_i(\mathbf{x})\}$ are real and all the eigenvalues $\{\lambda_i\}$ are positive. The localized whitening filters are

$$K_s(\mathbf{x}, \mathbf{x}') = \sum_i E_i(\mathbf{x}) \lambda_i^{-\frac{1}{2}} E_i(\mathbf{x}') \tag{7}$$

in which the summation is over all the eigenfunctions.

Those whitening filters are approximately shift invariant within two degrees of the gaze, and simulate the spatial process of the retina — of course, the real retina does more than spatial whitening (see references in Dong and Atick 1995b; Sterling 1998; Victor 1999). The filter output is

$$S_M(\mathbf{x}, t) = \int \int K_s(\mathbf{x}, \mathbf{x}') S(\mathbf{x}', t) d\mathbf{x}' \tag{8}$$

This is the spatially whitened visual input and is called the contrast signal in this paper/chapter. The spatiotemporal correlation of the contrast signal is calculated in the same way as of the light intensity signal in the previous subsection. To make a distinction, a subscript "M" is used for the contrast signal, e.g., $R_M(\mathbf{x}, t)$.

11.3 Quantification of retinal image motion

The retinal image motion is characterized by the optical flow velocity of the visual input. It is another important statistical property of the visual input — as shown in the chapter, it is an integral part of the spatiotemporal correlation and the main contributor to the non-stationary aspects.

The velocity is determined by calculating the optical flow fields between two consecutive frames using the following method in two steps. For a small area ($1^o \times 1^o$) of image at the center of gaze in one frame, first find its translated one in the next frame, i.e., get the large movement in number of pixels (d_x, d_y), through minimizing the square-difference,

$$\langle (S(x + d_x, y + d_y, t + 1) - S(x, y, t))^2 \rangle_{xy} \tag{9}$$

in which $\langle\rangle_{xy}$ is the average over the small area. Second, calculate the sub-pixel small movements (δ_x, δ_y) by minimizing another square-difference,

$$\left\langle \left(\delta_x \frac{\partial S}{\partial x} + \delta_y \frac{\partial S}{\partial y} - \frac{\partial S}{\partial t} \right)^2 \right\rangle_{xy} \tag{10}$$

in which $\langle\rangle_{xy}$ is the average over the same area.[‡] The overall velocity $\mathbf{v}$ has $v_x = d_x + \delta_x$ and $v_y = d_y + \delta_y$ for horizontal and vertical velocities, respectively. Since the light intensity $S(x, y, t)$ on the retina is used for the calculation, the velocity $\mathbf{v}$ is the velocity of the retinal image motion, i.e., the optical flow velocity on the retina — or an operational definition of it.

Assuming that the pixel values and their partial derivatives are not degenerate, $v_x = d_x + \delta_x$ and $v_y = d_y + \delta_y$ are well determined. For rigid-body image motion, they give the optical flow velocity. For non-rigid body motion, they give a definition of an "average" optical flow velocity. The dimension of the averaging area is chosen to be 1^o, similar to the average spatial separation in calculating the correlation function. For $1^o \times 1^o$ area (400 pixels), it is very unlikely that the degenerate situation ever happens for natural time-varying images. It is straightforward to find the minimum error solutions (for similar methods, see Cafforio and Rocca 1975; Jain and Jain 1981; Horn and Schunk 1981).

But here is a note of caution. Although one never fails to find a minimum, it doesn't mean that there is no inaccuracy in calculating the optical flow velocity, e.g., in calculating the velocity component parallel to a high contrast edge. For non-rigid body motions, the "average" optical flow velocity obviously depends on the area size and location. What is used here is the average on the central part of the fovea. Furthermore, the optical flow velocity calculated through Equation (5) and (6) is by no means the same as the projected velocity of self-motion (Koenderink and VanDoorn 1987) or scene-motion (Verri and Poggio 1989).

11.4 Classification of eye movements

The information processing by the brain is dynamic, i.e., the brain actively seeks information. In the human visual system, this active and dynamic process is characterized by periods of fixations and pursuits interleaved by saccades, which change the center of gaze from one direction to another in a very short period of time, and thus change the visual input significantly. Therefore it is very interesting to quantitatively compare between saccades and other eye movements for their effects on the visual input.

In this paper/chapter, saccade is defined as an eye movement of more than 0.5 deg within 20 msec. From the measurements during a typical free viewing of one of the natural scene movies, the average saccade amplitude is $\sim 2.7 \pm 1.7$ deg, the

[‡]The three partial derivatives are estimated by $\frac{1}{2}(S(x+1,y,t) - S(x-1,y,t))$, $\frac{1}{2}(S(x,y+1,t) - S(x,y-1,t))$, and $(S(x+d_x, y+d_y, t+1) - S(x,y,t))$, respectively.

average saccade duration is $\sim 28 \pm 8$ msec, the average saccade speed is $\sim 88 \pm 29$ deg/sec, and the average inter-saccade interval is $\sim 490 \pm 470$ msec. For another subject viewing another movie, the corresponding numbers are 2.4 ± 1.5 deg, 26 ± 8 msec, 87 ± 27 deg/sec, and 520 ± 400 msec.

Acknowledgments

The author wishes to thank Dr. Theodore Weyand for many interesting discussions about the LGN function and to thank Dr. Anna Kashina for her critical reading of the manuscript. This work was supported in part by FAU under the grant No. RIA-25, by NIMH under the grant No. MH019116, and by NSF under the grant No. PHY99-07949.

References

[1] Atick JJ (1992) Could information theory provide an ecological theory of sensory processing? *Network-Compu. Neural.* **3**, 213–251.

[2] Atick JJ, Redlich AN (1990) Towards a theory of early visual processing. *Neural Comp.* **2**, 308–320.

[3] Atick JJ, Redlich AN (1992) What does the retina know about natural scenes? *Neural Comp.* **4**, 196–210.

[4] Atick JJ, Li Z, Redlich AN (1993) What does post-adaptation color appearance reveal about cortical color representation. *Vision Res.* **33**, 123–129.

[5] Attneave F (1954) Some informational aspects of visual perception. *Psychol Rev* **61**, 183–193.

[6] Barlow HB (1961) Possible principles underlying the transformation of sensory messages. *In: Sensory Communication* (Rosenblith WA, ed, MIT Press, Cambridge) 217–234.

[7] Brenner N, Bialek W, VanSteveninck RR (2000) Adaptive Rescaling Maximizes Information Transmission. *Neuron* **26**, 695–702.

[8] Buisseret P (1995) Influence of extraocular-muscle proprioception on vision. *Physiological Reviews* **75**, 323–338.

[9] Burton GJ, Moorhead IR (1987) Color and spatial structure in natural scenes. *Applied Optics* **26**, 157–170.

[10] Cafforio C, Rocca F (1975) Methods for measuring small displacements of television images. *IEEE Info. Theo.* **22**, 573–579.

[11] Cai DQ, DeAngelis GC, Freeman RD (1997) Spatiotemporal receptive-field organization in the lateral geniculate-nucleus of cats and kittens. *J of Neurophysiol* **78**, 1045–1061.

[12] Carandini M, Ferster D (1998) A tonic hyperpolarization underlying contrast adaptation in cat visual cortex. *Science* **276**, 949–952.

[13] 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(10)**, 3351–3362.

[14] Dastjerdi M (2007) Efficient representation of natural visual input in the thalamus. *Ph D thesis, University Microfilms International, Ann Arbor, MI.*

[15] Dastjerdi M, Weyand TG, Dong DW (2003) The spatiotemoral receptive field (STRF) properties of the lateral geniculate nucleus (LGN) in the awake cats during free-viewing natural time-varying images. *Society for Neuroscience Abstract* **29**, 229.3.

[16] Ditchburn RW, Ginsborg BL (1952) Vision with a stabilized retinal image. *Nature* **170**, 36–37.

[17] Donaldson IML, Dixon RA (1980) Excitation of units in the lateral geniculate and contiguous nuclei of the cat by stretch of extrinsic ocular muscles. *Exp. Brain Res.* **38**, 245–255.

[18] Dong DW (1994) Associative decorrelation dynamics: a theory of self-organization and optimization in feedback networks. *In: Advances in Neural Information Processing Systems 7* (Tesauro G, Touretzky DS, Leen TK, eds, MIT Press, Cambridge, MA), 925–932.

[19] Dong DW (2001a) Spatiotemporal inseparability of natural images and visual sensitivities. *In: Computational, neural & ecological constraints of visual motion processing* (Zanker JM, Zeil J, eds, Springer Verlag, Berlin Heidelberg New York), 371–380.

[20] Dong DW (2001b) Effects of eye movements on information processing: Visual input statistics during free-viewing natural time-varying images. *Invest ophth & Vis Sci* **42**, S3347.

[21] Dong DW, Atick JJ (1995a) Statistics of natural time-varying images. *Network-Compu. Neural.* **6**, 345–358.

[22] Dong DW, Atick JJ (1995b) Temporal decorrelation: a theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network-Compu. Neural.* **6**, 159–178.

[23] Dragoi V, Sharma J, Sur M (2000) Adaptation-induced plasticity of orientation tuning in adult visual cortex. *Neuron* **28**, 287–298.

[24] Eckert MP, Buchsbaum G (1993) Efficient coding of natural time varying images in the early visual system. *Phil. Trans. R. Soc. Lond. B* **339**, 385–395.

[25] 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.

[26] Field DJ (1994) What is the goal of sensory coding. *Neural Computation* **6**, 559–601.

[27] Guido W, Weyand TG (1995) Burst responses in thalamic relay cells of the awake, behaving cat. *J. Neurophysiol.* **74**, 1782–1786.

[28] Horn BKP, Schunk BG (1981) Determining optical flow. *Artificial Intelligence* **17**, 185–203.

[29] Hoyer PO, Hyvarinen A (2000) Independent component analysis applied to feature extraction from colour and stereo images. *Network-Compu. Neural.* **3**, 191–210.

[30] Hubel DH (1988) Eye, Brain, and Vision. (Freeman and Company, New York).

[31] Jain JR, Jain AK (1981) Displacement measurement and its application in interframe image coding. *IEEE T. Commun.* **29**, 1799–1808.

[32] Kelly DH (1961) Visual responses to time-dependent stimuli .1. amplitude sensitivity measurements. *Journal of the Optical Society of America* **51**, 422–429.

[33] Kelly DH (1979) Motion and vision. II. Stabilized spatio-temporal threshold surface. *J. Opt. Soc. Am.* **69**, 1340–1349.

[34] Koenderink JJ, VanDoorn AJ (1987) Facts on Optical Flow. *Bio. Cybern.* **56**, 247–254.

[35] Lee D, Malpeli JG (1998) Effects of saccades on the activity of neurons in the cat lateral geniculate nucleus. *J. Neurophysiol.* **79**, 922–936.

[36] Lal R, Friedlander MJ (1990a) Effect of eye position changes on retinogeniculate transmission in the cat. *J. Neurophysiol.* **63**, 502-522.

[37] Lal R, Friedlander MJ (1990b) Effect of passive eye movement on retinogeniculate transmission in the cat. *J. Neurophysiol.* **63**, 523–538.

[38] Li Z, Atick JJ (1994) Efficient stereo coding in the multiscale representation. *Network-Compu. Neural.* **5**, 1–18.

[39] Linsker R (1988) Self-organization in a perceptual network. *Computer* **21**, 105–117.

[40] Linsker R (1989) An application of the principle of maximum information preservation to linear systems. *In: Advances in Neural Information Processing Systems 1* (Touretzky DS, ed, Morgan Kaufman, San Mateo, CA), 186–194.

[41] Muller JR, Metha AB, Krauskopf J, Lennie P (1999) Rapid Adaptation in Visual Cortex to the Structure of Images. *Science* **285**, 1405–1408.

[42] Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive-field properties by learning a sparse code for natural images. *Nature* **381**, 607–609.

[43] Parraga CA, Brelstaff G, Troscianko T (1998) Color and luminance information in natural scenes. *J Opt Soc Am* **15**, 563–569.

[44] Penev PS, Atick JJ (1996) Local feature analysis: A general statistical theory for object representation. *Network-Compu. Neural.* **7**, 477–500.

[45] Ramcharan EJ, Gnadt JW, Sherman SM (2001) The effects of saccadic eye movements on the activity of geniculate relay neurons in the monkey. *Visual Neuroscience* **18**, 253–258.

[46] Reppas JB, Usrey WM, Reid RC (2002) Saccadic eye movements modulate visual responses in the lateral geniculate nucleus. *Neuron* **35**, 961–974.

[47] Riggs LA, Ratliff F (1952) The effects of counteracting the normal movements of the eye. *J. Opt. Soc. Am.* **42**, 872–873.

[48] Rucci M, Edelman GM, Wray J (2000) Modeling LGN responses during free-viewing: A possible role of microscopic eye movements in the refinement of cortical orientation selectivity. *J Neurosci* **20**, 4708–4720.

[49] Ruderman DL, Bialek W (1994) Statistics of natural images: scaling in the woods. *Phy. Rev. Let.* **73(6)**, 814–817.

[50] Ruderman DL, Cronin TW, Chiao CC (1998) Statistics of cone responses to natural images: implications for visual coding. *J Opt Soc Am* **15**, 2036–2045.

[51] Schmidt M (1996) Neurons in the cat pretectum that project to the dorsal lateral geniculate nucleus are activated during saccades. *J. Neurophysiol.* **76**, 2907–2918.

[52] Shannon, CE, Weaver W (1949) A mathematical theory of communication. *Univ. of Illinois Press, Urbana.*

[53] Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. *Ann. Rev. Neurosci.* **24**, 1193–1216.

[54] Singer W (1977) Control of thalamic transmission by corticofugal and ascending reticular pathways in the visual system. *Phys. Reviews* **57**, 386–420.

[55] Smirnakis SM, Berry MJ, Warland DK, Bialek W, Meister M (1997) Adaptation of retinal processing to image contrast and spatial scale. *Nature* **386**, 69–73.

[56] Srinivasan MV, Laughlin SB, Dubs A (1982) Predictive coding - a fresh view of inhibition in the retina. *Proc. R. Soc. B* **216**, 427–459.

[57] Sterling P (1998) Retina. *In: The Synaptic Organization of the Brain* (Shepherd GM, ed, Oxford, New York) 205–253.

[58] Swadlow HA, Weyand TG (1987) Corticogeniculate neurons, corticotectal neurons, and suspected interneurons in visual cortex of awake rabbits: Receptive field properties, axonal properties, and effects of EEG arousal. *J. Neurophysiol.* **57**, 977–1001.

[59] Tailor DR, Finkel LH, Buchsbaum G (2000) Color-opponent receptive fields derived from independent component analysis of natural images. *Vision Research* **40**, 2671–2676.

[60] Truccolo WA, Dong DW (2001) Dynamic temporal decorrelation: information theoretic and biophysical model of the functional role of lateral geniculate nucleus. *Neurocomputing* **38-40**, 993–1001.

[61] VanHateren JH (1992) Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. *J. Comp. Physiol. A* **171**, 157–170.

[62] VanHateren JH, Ruderman DL (1998) Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *P Roy Soc Lond B Bio* **265**, 2315–2320.

[63] Verri A, Poggi T (1989) Motion Field and Optical Flow: qualitative properties. *IEEE T. Pat. Ana. Mac. Intel.* **11**, 490–498.

[64] Victor JD (1999) Temporal aspects of neural coding in the retina and lateral geniculate. *Network-Comp. Neural.* **10**, R1–R66

[65] Wandell BA (1995) Foundations of Vision (Sinauer, Sunderland).

[66] Webster MA, Mollon JD (1997) Adaptation and the color statistics of natural images. *Vision Res.* **37**, 3283–3298.

[67] Webster MA, Miyahara E (1997) Contrast adaptation and the spatial structure of natural images. *J Opt Soc Am A* **14**, 2355–2366.

[68] Wehmeier U, Dong DW, Koch C, VanEssen DC (1989) Modeling the mammalian visual system. *In: Methods in Neuronal Modeling: from Synapses to Networks* (Koch C, Segev I, eds, MIT Press, Cambridge), 335–360.

[69] Yarbus AL (1967) Eye Movements and Vision (Plenum, New York).

[70] Zhaoping L (2002) Optimal Sensory Encoding. *In: The Handbook of Brain Theory and Neural Networks (Second Edition)* (Arbib MA, ed, MIT Press, Cambridge, MA), 815–819.