Statistics of Natural Time-Varying Images

Dawei W. Dong and Joseph J. Atick

Computational Neuroscience Laboratory
The Rockefeller University
1230 York Avenue
New York, NY 10021-6399

(*Network: Computation in Neural Systems* Vol 6(3) pp 345-358)

Abstract

Natural time-varying images possess substantial spatiotemporal correlations. We measure these correlations — or equivalently the power spectrum — for an ensemble of more than a thousand segments of motion pictures, and we find significant regularities. More precisely, our measurements show that the dependence of the power spectrum on the spatial frequency, f, and temporal frequency, w, is in general non-separable and is given by $f^{-m-1}F(w/f)$, where $F(w/f)$ is a nontrivial function of the ratio w/f. We give a theoretical derivation of this scaling behaviour and show that it emerges from objects with a static power spectrum $\sim f^{-m}$, appearing at a wide range of depths and moving with a distribution of velocities relative to the observer. We show that in the regime of relatively high temporal and low spatial frequencies, the power spectrum becomes independent of the details of the velocity distribution and it is separable into the product of spatial and temporal power spectra with the temporal part given by the universal power-law $\sim w^{-2}$. Making some reasonable assumptions about the form of the velocity distribution we derive an analytical expression for the spatiotemporal power spectrum which is in excellent agreement with the data for the entire range of spatial and temporal frequencies of our measurements. The results in this paper have direct implications to neural processing of time-varying images in the visual pathway.

1 Introduction

Vision for most animals is concerned with the perception of objects in a dynamic world, one that appears to be constantly changing when viewed over extended periods of time. A significant fact about natural time-varying images is that they do not change randomly over space or time; instead images at different times and spatial positions are highly correlated. For example, light intensities at nearby locations in space and time tend to be very similar, giving a luminosity profile which changes gradually in space-time and only abruptly at edges or motion ridges (illustrated in Figure 1). Knowledge of such regularities in the signal ensemble is very important since as Communication Theory (Shannon and Weaver 1949) shows it allows the design of more efficient ways to represent and transmit information.

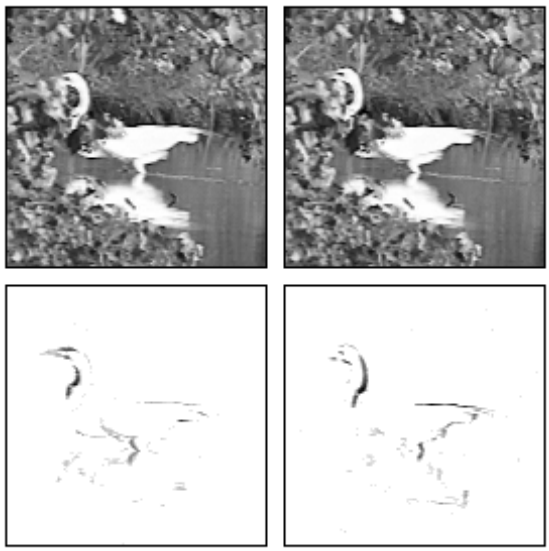

Figure 1: Natural time varying images are highly correlated in space and time. Shown on the top, are two frames of a motion scene separated by thirty three milliseconds. These two frames are highly repetitive, in fact the light intensities of most corresponding pixels are similar. Shown on the bottom, are light increase (on the left) and light decrease (on the right) of the above two snapshots indicated by greyscale of pixels (white means no change). One can immediately see that only a small portion of the image changes significantly over this time scale.

The animal brain is faced with the problem of representation and transmission of information from its senses, and hence it is reasonable to expect that neurons in the sensory pathways developed to take advantage of certain statistical regularities in incoming signals to build more suitable representations of the world. The literature

showing the connection between properties of natural stimuli and neural processing is by now vast (see references in Dong and Atick 1995), and the successes so far reinforce our belief that better characterization of properties of natural signals can result in better understanding of neural function.

The statistical properties of static images have been studied for many years (Burton and Moorhead 1987; Field 1987; Tolhurst *et al.* 1992; Hancock *et al.* 1992; Ruderman and Bialek 1994), and as a result we now know some interesting regularities of such images. Our knowledge of the regularities of time-varying images, on the other hand, has so far been very limited. This is despite the fact that interest in properties of time-varying images dates back to the early days of development of the television (Kretzmer 1952). Systematic studies have not been possible previously because the technology to capture and analyze motion pictures on computers has become available only recently in inexpensive video-boards and cameras. This is opening up a new avenue for exploration, one that we expect will attract considerable attention over the next several years.

A full characterization of the statistical regularities of natural scenes is virtually impossible. Time-varying images possess a multitude of structures and regularities at many levels of complexity. The goal of the current work is to measure certain simple aspects of these statistics, ones that we believe the visual system takes advantage of in building better representations of the dynamic world. First, we measure the joint probability distribution of the signal at two different spatiotemporal points. This is the lowest order statistical regularity which reveals the interdependency and repetitiveness of signals across space and time. Second, we systematically measure the two point correlation matrix or covariance matrix of large segments of natural time-varying images. This matrix, of course, does not give a complete characterization of the joint probability but it is simpler to measure and, as we will see below, it reveals some significant regularities. The Fourier transform of the correlation matrix, or the power spectrum, turns out to be a non-separable function of spatial and temporal frequencies and exhibits an interesting scaling behaviour. From our measurements we find

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

where $F(w/f)$ is some function of the ratio of spatial and temporal frequencies.

In section 4, we give a theoretical derivation of this scaling behaviour and show that it can be accounted for if the dominant component in the temporal signal is coming from motion of objects at many depths with static power spectra of $\sim f^{-m}$. The details of the function $F(w/f)$ depend on the distribution of relative velocities.

However, in some regimes of interest we show the details of the velocity distribution become irrelevant and the power spectrum reduces to a product of spatial and temporal spectra, with the temporal spectrum given by the universal power-law

$$\sim 1/w^2 \ .$$

Included in this regime is the region of low spatial and intermediate temporal frequencies, which is where the experiments on the LGN temporal tuning properties are done (see references in Dong and Atick 1995). Finally, taking a simple power law velocity distribution, we arrive at an analytical formula for the power spectrum that fits the measured data very well for the entire range of spatial and temporal frequencies.

2 Method and Notation

Segments of videos on 8mm video tape (NTSC format RGB) are digitized to 8 bits gray-scale using a Silicon Graphics Video board with default factory settings.* Two types of segments are analyzed. The first are segments from movies on video tapes (*e.g.* Indiana Jones, Uncommon Valor). For these movies, each segment is a $64 \times 64 \times 64$ (horizontal $\times$ vertical $\times$ temporal) image. The temporal resolution is 24 frames/second which is the rate at which movies are filmed. Since we had no control over how these movies were taken, the spatial resolution can only be estimated, which is roughly 10 degrees for a field of view of 64 pixels.

The second type of segments that we analyzed are videos made by the authors. The videos were taken by a Sony Handycam CCD-FX710 camera. The camera's zoom is fixed to give the above spatial resolution, and the shutter speed is chosen to be 1/250 second. The gain and white balance are automatically adjusted to the given environment and then fixed. Each segment is a $64 \times 64 \times 256$ (horizontal $\times$ vertical $\times$ temporal) image. The temporal resolution is 60 frames/second which is achieved using the following technique.

Since each video frame (640×480 pixels and 30 frames/second) consists of an even field and an odd field ($640 \times 240 \times 2$) which are taken at a rate of 60 fields/second, the 60 frames/second rate is achieved through separating the even and the odd fields of each frame and then using spatial linear interpolation to align them (compensate for the vertical offset). Actually, only the central 64×64 pixels of each frame are cropped carefully to compensate for offset and aspect. This technique which allows

*coring: off, aperture: 0.25, bandpass: one, color-mode: auto, chroma-agc: slow, luma-delay: 1 pixel, vertical-noise: normal, hue: 0, chroma-gain: 0.173, color-kill-threshold: -29.625db.

us to double the temporal resolution of a video is only useful for videos made by the authors.

By random sampling from the movies or videos, we have collected more than one thousand segments. The movies and videos are subjectively natural to the authors; they capture a wide range of temporal variations and motions. In taking the videos, we have tried to make sure that objects at different distances (from 2 to 40 meters) are sampled equally, and at each distance many samples are taken to cover a variety of motions.

In this paper, we adopt the following conventions: given light intensity $S(\boldsymbol{x}', t')$, the correlation between two points separated by spatiotemporal distance $\boldsymbol{x}, t$ is

$$R(\boldsymbol{x}, t) = \frac{1}{L^2 T} \int_0^{L^2} \int_0^T S(\boldsymbol{x} + \boldsymbol{x}', t + t') S(\boldsymbol{x}', t') d\boldsymbol{x}' dt', \tag{1}$$

where L^2 is the spatial size of the averaged region and T is its temporal length. The Fourier transform of $R(\boldsymbol{x}, t)$ is defined as

$$R(\boldsymbol{f}, w) = \frac{1}{L^2 T} \int_0^{L^2} \int_0^T R(\boldsymbol{x}, t) e^{i2\pi(\boldsymbol{f}\cdot\boldsymbol{x} + wt)} d\boldsymbol{x} dt. \tag{2}$$

It is obvious that both $R(\boldsymbol{x}, t)$ and $R(\boldsymbol{f}, w)$ in our definition are in the units of S squared. Since the signal S has been digitized to 8 bits, the absolute units are meaningless; thus only numerical values are presented below.

For an ergodic system, the Fourier transform of the correlation function $R(\boldsymbol{f}, w)$ can be calculated directly from the power spectrum

$$R(\boldsymbol{f}, w) = < S(\boldsymbol{f}, w) S^*(\boldsymbol{f}, w) >, \tag{3}$$

in which $S(\boldsymbol{f}, w)$ is the Fourier transform of a large segment of a spatiotemporal scene $S(\boldsymbol{x}, t)$ and the bracket $<>$ denotes averaging over many segments; the relative error for such an estimate is one over the square root of the number of segments averaged, and to reduce the effects of the limited segment window, Welch window is used to mask the data $S(\boldsymbol{x}, t)$ before Fourier transforming. (This choice does not affect the results significantly; Bartlett, Gaussian, and Hann windows give basically the same results.) Interested readers should consult Press *et al.* (1992).

3 Results

We begin by examining the pairwise probability distribution $P(S_1, S_2)$, where $S_1 = S(\boldsymbol{x}_1, t_1)$ and $S_2 = S(\boldsymbol{x}_2, t_2)$ are the light intensities at position $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ and frame

t_1 and t_2, respectively. Figure 2 shows an example of this distribution for $\boldsymbol{x}_1 = \boldsymbol{x}_2$ and $t_2 - t_1 = 33\ ms$. It is clear from this that the light intensities at the two different times are far from being statistically independent: the most probable states lie on the diagonal line $S_1 = S_2$. Many other examples of this correlation can be exhibited and they all confirm the intuitive notion that to first approximation natural scenes are quasi-static. In other words, images change gradually in time just like they change gradually in space. An information processing system, such as the brain, can use this regularity to anticipate what the signal will do in time and hence build a better representation of the incoming information.

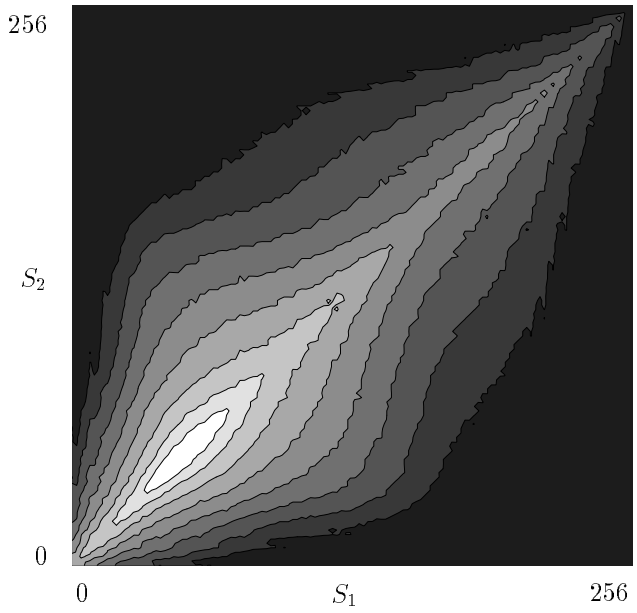

Figure 2: Measured pairwise probability distribution $P(S_1, S_2)$. S_1 and S_2 are light intensities at the same spatial location but separated in time by 33 ms. The gray scale is used for illustration purpose: white (black) represents high (low) probability. The contours are plotted at equal intervals in the logarithm of probability: $P(S_1, S_2)$ from e^{-8} to e^{-15}. The distribution is peaked around the diagonal line $S_1 = S_2$, indicating that the two light intensities are highly correlated, and most probably, are equal.

Next, we examine the simplest quantity which captures this regularity: the covariance matrix (this is the same as the correlation defined in section 2, since the mean light intensity is subtracted). For the probability distribution in Figure 2, the measured matrix is $R = 39 \begin{pmatrix} 1.0 & 0.9 \\ 0.9 & 1.0 \end{pmatrix}$. The ratio of the off-diagonal terms to the diagonal terms is close to one showing how strong S_1 and S_2 statistically depend on each other. We systematically measured the covariance matrix for different spatial and temporal separations as well as the spatiotemporal power spectrum. The spa-

tiotemporal power spectrum is easier to measure and more convenient to analyze. So from now on, we present all of our results in the spatial and temporal frequency domains.

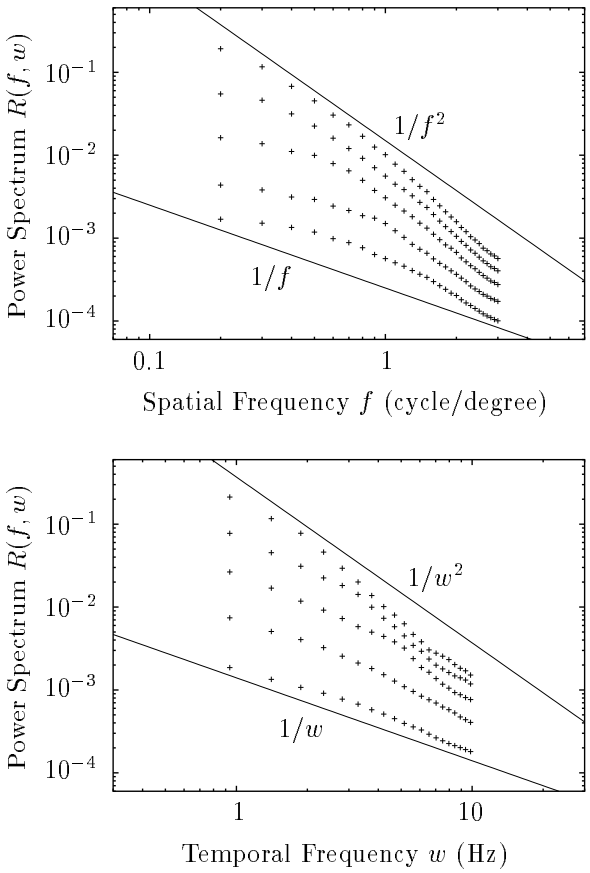

Figure 3: Measured spatial (top) and temporal (bottom) power spectra of natural time-varying images. In the top figure the temporal frequency increases from 1.4, 2.3, 3.8, 6, to 10 Hz as we go from the highest to the lowest curve, while in the bottom figure the spatial frequency increases from 0.3, 0.5, 0.8, 1.3, to 2.1 cycle/degree as we go from the highest to the lowest curve. Also shown are the lines representing the power-laws $1/f^2$, $1/f$ (top) and $1/w^2$, $1/w$ (bottom), for reference.

We have measured the spatiotemporal power spectrum $R(\boldsymbol{f}, w)$ of each time-varying image segment and then averaged over many segments. For the movie "Uncommon Valor" we have averaged over 1049 different $64 \times 64 \times 64$ (x-y-t) segments. For our video footage we have averaged over 320 different $64 \times 64 \times 256$ (x-y-t) segments. For the purposes of this paper we have ignored the dependence on spatial orientation and averaged over all orientations, thus only one spatial dimension is plotted below. In Figure 3, we show the measured $R(f, w)$ as two families of curves, one as a function

of spatial frequency (top) with temporal frequency increasing from 1.4 Hz to 10 Hz as we go from the highest to the lowest curve, and the other as a function of spatial frequency (bottom) with spatial frequency increasing from 0.3 cycle/degree to 2.1 cycle/degree as we go from the highest to the lowest curve.

Had natural scenes been random in space and time, *i.e.*, white noise, we would have gotten a flat power spectrum in both domains, *i.e.* the power lines would lie horizontally. The measurement indicates otherwise; natural scenes have more power at low frequencies and this power decreases as spatial and/or temporal frequency increases. For a given temporal frequency, the data shows that the power spectrum decreases roughly as a reciprocal power of spatial frequency:

$$R \sim \frac{1}{f^a}. \tag{4}$$

Similarly, for a given spatial frequency, the power spectrum decreases roughly as a reciprocal power of temporal frequency:

$$R \sim \frac{1}{w^b}. \tag{5}$$

Both the a and b are positive numbers. In Figure 3, on top, $1/f^2$ and $1/f$, and on bottom, $1/w^2$ and $1/w$ are plotted for reference; in the double log plot, they are straight lines.

There are some important conclusions that can be drawn from this measurement. First, it is obvious that the power spectrum cannot be separated into pure spatial and pure temporal parts, space and time are coupled in a non-trivial way. Second, underlying this data is an interesting scaling behaviour. To see this we have replotted the same data in Figure 4 (top) but as a function of w/f . The curves now become very similar. In fact, if we multiply the spectrum by a power of f, *i.e.* if we plot $f^{m+1}R(f,w)$ as a function of w/f then all curves coincide very well, as shown in Figure 4 (bottom).

The value of m that achieves this reveals a lot about the origin of this scaling behaviour. We find $m = 2.3$ makes all curves coincide very well. But this is precisely the same m that we get by measuring the static power spectrum for this collection of images (frames treated as snapshots). More precisely the static power spectrum we find is

$$R_s(f) \sim \frac{1}{f^m}. \tag{6}$$

with $m = 2.3$, in general agreement with other earlier measurements (Burton and Moorhead 1987, Field 1987, Ruderman and Bialek 1994). This observation is used

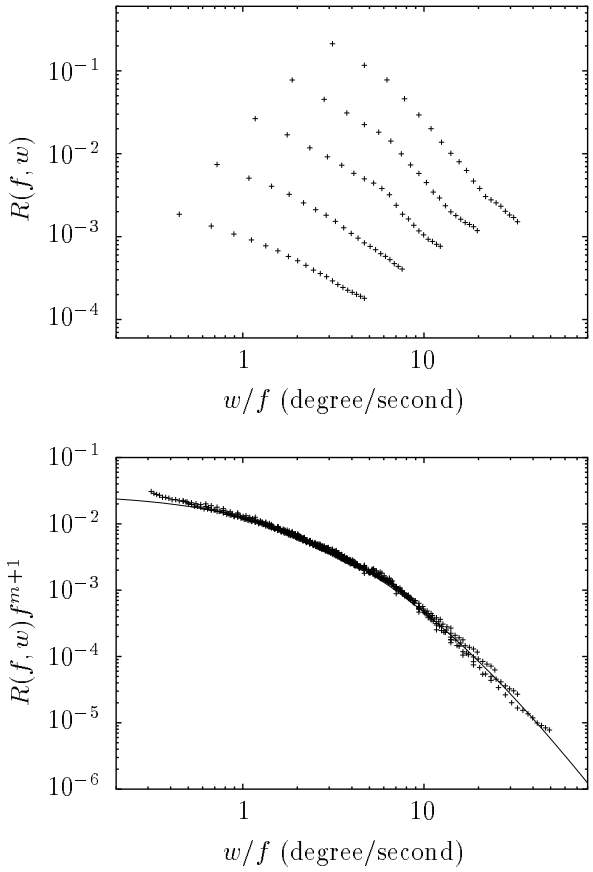

Figure 4: The data of Figure 3 for the power spectrum replotted as a function of w/f (top) and replotted after multiplication by f^{m+1} (bottom), with $m = 2.3$. All the data points fall on a single curve. The solid curve is the analytical form from Equation 19.

in the next section to explain the origin of the scaling behaviour and to derive the power spectrum of moving images starting from the static one.

4 Analysis

4.1 Deriving the power spectrum of moving images

The spatiotemporal scaling exhibited in Figure 4 (bottom) suggests that one may be able to derive the power spectrum of moving images from first principles. To see this, consider an object at distance r from the observer, moving with a relative velocity $\boldsymbol{v}$. The power spectrum of the image sequence showing this object in motion is given by:

$$R(\boldsymbol{f}, w, \boldsymbol{v}, r) = R_s(f)\delta(w - \boldsymbol{f} \cdot \boldsymbol{v}/r) \tag{7}$$

where $\delta(w - \boldsymbol{f} \cdot \boldsymbol{v}/r)$ is the Dirac (Kronecker) delta function which is normalized to one and is zero everywhere except for $w = \boldsymbol{f} \cdot \boldsymbol{v}/r$; and $R_s(f)$ is the spatial power spectrum of the image of the static object, which we assume is given by the rotationally symmetric static spectrum $R_s(f) = K/f^m$, for some normalization K and with $m = 2.3$ as shown in Figure 5. In writing this, we assume that the contributions from different distances are the same thus K does not depend on r (but see next).

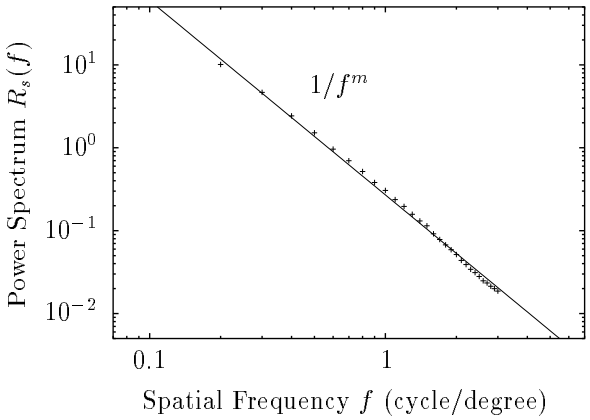

Figure 5: Measured spatial power spectrum of snap shot images. It shows that $R_s(f) \sim 1/f^m$ is a good approximation to the spectrum. In our measurement $m = 2.3$.

In long segments of natural time-varying images, we expect an object to appear moving at different relative velocities. This is due to the fact that objects do not tend to move with absolutely constant velocities and also viewers (or cameras) typically do not precisely fixate on a given moving object. Thus to recover $R(f, w)$ we must average over a distribution of velocities $P(\boldsymbol{v})$

$$R(\boldsymbol{f}, w, r) = R_s(f) \int \int \delta(w - \boldsymbol{f} \cdot \boldsymbol{v}/r) P(\boldsymbol{v}) d\boldsymbol{v} \tag{8}$$

For simplicity, we assume a rotationally invariant distribution of velocities, it is then straightforward to show that the power spectrum is also rotationally invariant, *i.e.*, $R(\boldsymbol{f}, w, r) = R(f, w, r)$ with

$$R(f, w, r) = R_s(f) \int_0^\infty \delta(w - f v_x/r) P_x(v_x) dv_x \tag{9}$$

in which v_x is the velocity component projected onto a specific direction and $P_x(v_x)$ is the one dimensional velocity distribution $P_x(v_x) = \int P(\boldsymbol{v}) dv_y$. To simplify the notation, we will henceforth drop the subscript x on v_x and P_x.

Second we integrate over distances between a range of r_1 and r_2 which represent the closest distance at which motion is observed and the farthest distance at which motion is still resolved, respectively. Thus the power spectrum we would predict is

$$R(f,w) = \frac{K}{f^m} \int_{r_1}^{r_2} \int_0^\infty \delta(w - fv/r) P(v) dr dv, \tag{10}$$

which through change of variables and integration over v is the same as:

$$R(f,w) = \frac{K}{f^{m+1}} F(\frac{w}{f}), \tag{11}$$

where

$$F(\frac{w}{f}) = \int_{r_1}^{r_2} P(\frac{w}{f} r) r dr. \tag{12}$$

and is only a function of the ratio w/f.

This immediately explains the scaling behaviour noted in the previous section (Figure 4). It shows that $f^{m+1} R(f,w)$ is a function of w/f only. It confirms the hypothesis that the dominant contribution to the spatiotemporal variability in the signal is relative motion.

4.2 Asymptotic Behaviour

While the detailed form of the function $F(\frac{w}{f})$ in Equation 12 depends on the velocity distribution, there is an interesting asymptotic regime where the detailed form of $P(v)$ is irrelevant. This asymptotic behaviour exists as long as the average velocity (absolute value) exists:

$$2 \int_0^\infty P(v) v dv = \bar{v}. \tag{13}$$

To see this, we rewrite Equation 12 for $F(w/f)$, through a change of variables as:

$$F(\frac{w}{f}) = \frac{f^2}{w^2} \int_{\frac{w}{f} r_1}^{\frac{w}{f} r_2} P(v) v dv. \tag{14}$$

in the limit of $\frac{w}{f} r_2 >> \bar{v}$ (but $\frac{w}{f} r_1 << \bar{v}$) – the regime of relatively high temporal and low spatial frequencies – the integration limits extend deep into the tails of the distribution $P(v)$ and the integral approximates the average velocity $\bar{v}$ divided by 2 and hence

$$F(\frac{w}{f}) \to \frac{f^2}{w^2} \frac{\bar{v}}{2} \tag{15}$$

and

$$R(f,w) \to \frac{K \bar{v}}{2 f^{m-1} w^2} \tag{16}$$

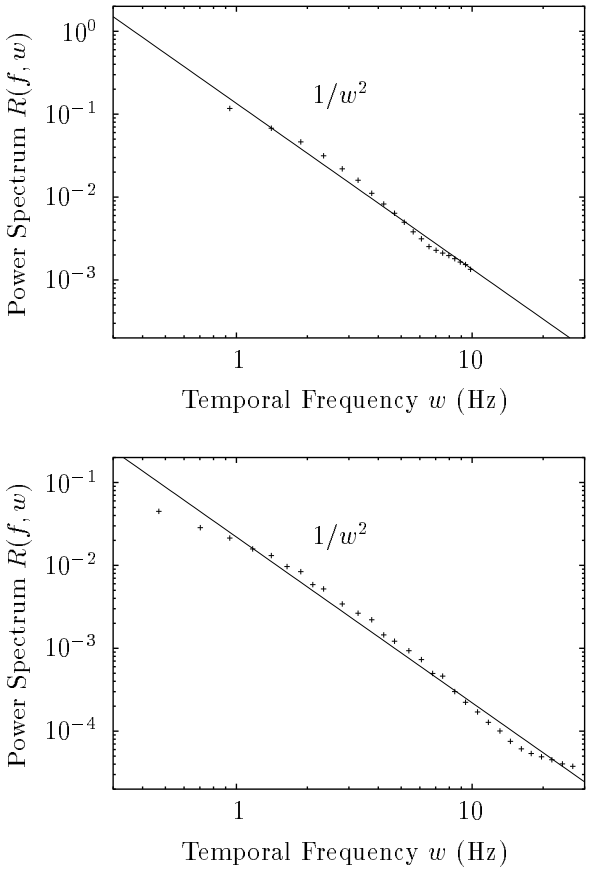

Figure 6: Measured temporal power spectra of natural time-varying images at a given spatial frequency, 0.4 cycles/degree (top) and 0.6 cycles/degree (bottom), for the ensemble of video segments from the movie "Uncommon Valor" (top) and from our video footage (bottom). In both cases, $R(f,w) \sim 1/w^2$ (solid curves) is a good approximation to the data.

Note that this asymptotic behaviour holds for any velocity distribution $P(v)$ as long as $\bar{v}$ exists.

Our measurements confirm the existence of this behaviour. Figure 6 shows the temporal power spectrum at some relatively low spatial frequencies and it indeed follows a $1/w^2$ power-law.[†]

Equation 16 also shows that in this asymptotic regime the power spectrum becomes separable in space and time and it is merely a product of $1/w^2$ in time with $1/f^{m-1}$ in space (note the spatial power is one less than the power measured for snap shot images).

[†]In the companion paper (Dong and Atick 1995), we used this behaviour in the limit of "$f \to 0$" which should be understood as the limit of small f, such as 0.4 cycle/degree.

4.3 Analytical Function

We can derive an analytical expression for $F(\frac{w}{f})$ for all w/f in Equation 12, if we choose a velocity distribution. We will make the reasonable assumption of a power law distribution:

$$P(v) \sim \frac{1}{(v+v_0)^n} \tag{17}$$

for some constant v_0 and power n. For this distribution the average velocity is $\bar{v} = v_0/(n-2)$; for a well defined distribution, $n > 2$ as required by Equation 13.

Through a change of variable and taking into account the normalization of the probability distribution, the function $F(\frac{w}{f})$ of Equation 14 reduces to:

$$F(\frac{w}{f}) = K\frac{f^2}{w^2}\frac{\bar{v}}{2}(n-2)(n-1)\int_{\frac{wr_1}{fv_0}}^{\frac{wr_2}{fv_0}} \frac{xdx}{(x+1)^n} \tag{18}$$

Finally the power spectrum is

$$R(f,w) = \frac{K\bar{v}}{2f^{m-1}w^2}\left[\frac{n-2}{(x+1)^{n-1}} - \frac{n-1}{(x+1)^{n-2}}\right]_{\frac{wr_1}{fv_0}}^{\frac{wr_2}{fv_0}} \tag{19}$$

Since K and m are determined by the power spectrum of snap shot images, there are only four parameters in this equation, r_1, r_2, v_0, and n; and they are highly constrained and cannot take arbitrary values. First of all, n has to be bigger than 2 to guarantee that the average velocity exists. Furthermore, r_1, r_2 and v_0 have to take reasonable values, since v_0 relates to the average velocity of relative motion and r_1, r_2 correspond to the closest distance at which motion is observed to the largest distance at which motion is still resolved, respectively.

The predicted power spectrum $R(f,w)$, for a reasonable set of parameters, is shown in Figure 7 and the corresponding $F(\frac{w}{f})$ is shown in Figure 4. The analytical curves fit the data very well, with most data points falling within 10% of the predicted curves. The only systematic deviations are at very high spatial and/or temporal frequencies, where the signal to noise ratio is low.

Figure 7 corresponds to the following situation: most of the objects in the movie are uniformly distributed between $r_1 = 4$ to $r_2 = 23$ meters away from the camera. While this range is narrower than the range of normal vision it may reflect some bias on the part of the movie director. For the segments that we have taken on video $r_1 = 2$ and $r_2 = 40$ meters fit the data very well. The average velocity $\bar{v}$ is 0.6 m/s which is a sensible real world value.

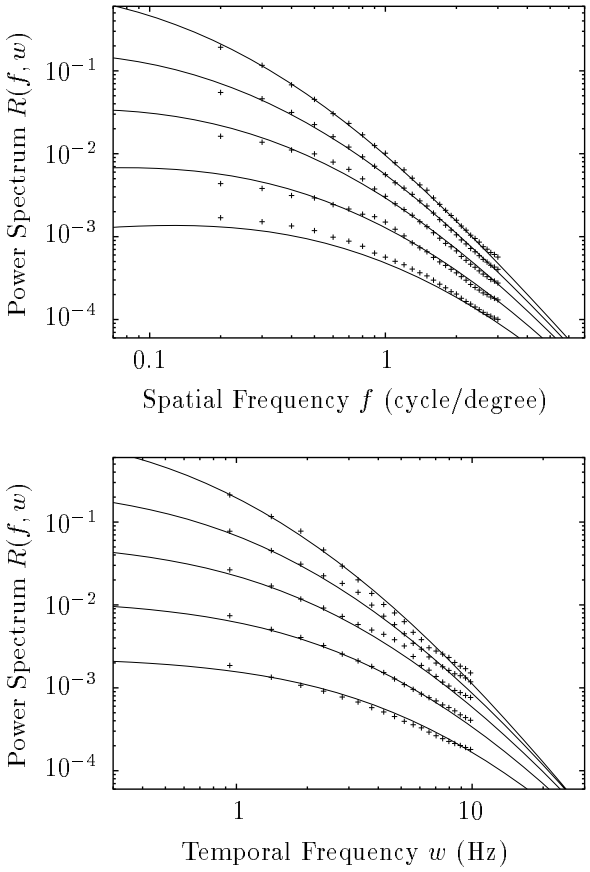

Figure 7: Measured spatial and temporal power spectra of natural time-varying images compared with the predictions of the analytic funtion in Equation 19. The parameters used are $r_1 = 4.0$ meters, $r_2 = 23$ meters, $\bar{v} = 0.6$ m/s, and $n = 3.7$.

The dependence of the predicted power spectrum on the parameters r_1 and r_2 can be explored easily, as shown is Figure 8. It is interesting to notice that for the extreme range of normal human vision of $r_1 = 0.2$ to $r_2 = 400$ meters, the predicted function is very close to

$$R(f,w) \sim \frac{1}{f^{m-1}w^2} \tag{20}$$

This is the regime where the power spectrum is independent of the parameters r_1, r_2, and n as discussed earlier. For this specific case, it is obvious from Equation 19 that when r_1 and r_2 approach zero and infinity respectively, the predicted function converges to Equation 16 exactly.

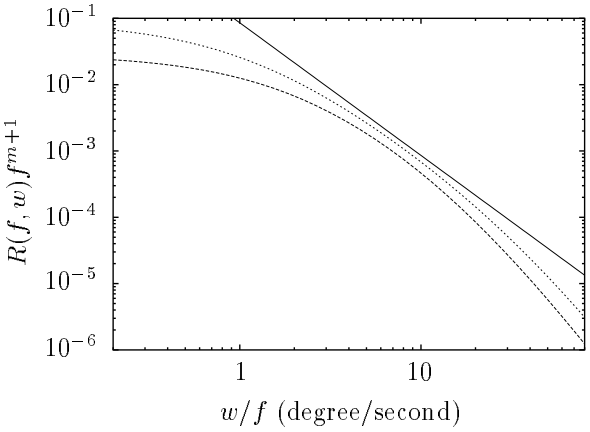

Figure 8: Asymptotic behaviour of the predicted power spectrum $R(f, w)$ plotted as $R(f, w)f^{m+1}$ versus w/f. It shows in the limit $r_1 \to 0$ and $r_2 \to \infty$, $R(f, w) \to 1/f^{m-1}w^2$, which is the solid line representing f^2/w^2. The dashed line is the curve for $r_1 = 4$ and $r_2 = 23$ meters, while the dotted line is for the intermediate range of $r_1 = 2$ and $r_2 = 40$ meters. Thus, as the range of viewing is extended the power spectrum reaches the asymptotic behaviour rather rapidly (in fact for the extreme human viewing range of $r_1 = 0.2$ and $r_2 = 400$ meters the curve cannot be distinguished from the solid curve in the w/f range shown in this figure).

5 Discussion

In this paper, we have quantified the pairwise pixel dependencies by examining the pairwise correlation matrix or equivalently its spatiotemporal power spectrum. Note, we are not suggesting that natural scenes are Gaussian signals, in fact Figure 2 shows they are not. We are merely saying that the correlation matrix, which should be thought of as a constraint on the joint distribution, is the simplest quantity that captures the statistical dependency in space and time of natural time-varying images. Furthermore, it is the quantity that we believe neurons in the early stages of the visual system seem to be able to evaluate and take advantage of in recoding the visual input in the retina and the LGN. The connection between the spatial power spectrum of natural scenes and properties of retina cells was explored in (Atick and Redlich 1990, 1992) and between the temporal domain and LGN cells in (Dong and Atick 1995).

In the companion paper (Dong and Atick 1995) we show that the temporal response properties of LGN cells in the early visual pathway of cats can be accounted for very well by the theory of temporal decorrelation of the power spectrum measured in this paper. Very briefly, the predicted temporal tuning of LGN neurons at low spatial and relatively high temporal frequencies is given by $\sim 1/\sqrt{R(f, w)}$ — and since from our measurements $R(f, w) \sim 1/w^2$ — one expects a linear function of w, which

is consistent with physiological data.

Also, our analysis in that paper suggests that in regimes where the power spectrum decouples into a spatial and temporal parts, one can characterize neural processing by giving the spatial and temporal tuning curves independently. In general, the spatial and temporal parts of $R(f, w)$ are not separable and hence receptive fields of neurons cannot be fully characterized in space independently of time. The scaling behaviour that we found in section 3 suggests a natural way for dealing with this coupling. More precisely, it suggests that a better way to examine spatiotemporal tuning data from real neurons is to plot the data not as a function of f and w separately but as a function of w/f. In this representation we expect that neurons will exhibit more universal behaviour.

The current measurement should be considered as a first order characterization of the temporal properties of time-varying images. More systematic studies can be done and the implication of those to higher neural centers should be examined. For example one could ask to what extent different types of motion (smooth pursuit vs. abrupt shifts of field of view or scene cuts) have different statistical regularities. In a sense in this paper we did not distinguish between the different types of motion. In fact with the exception of one type, the scene cut, we have tried to include a variety of motions and as such our results are ensemble averaged over all motions. We have avoided scene cuts since they can generate power law behaviour similar to what we found above for an entirely different reason. It is not difficult to show that the power spectrum of one dimensional signals dominated by edges of random phase is $1/w^2$. This is well known and is different from the effect that we have measured.

The power-law behaviour that we discovered results from two simple facts. First, static natural images at different distances, within a wide range, look equally natural and are equally important for the survival of an animal. Second, there is always a velocity distribution of relative motions either because in any given scene objects tend to move at different velocities or because animals like ourselves cannot perfectly fixate nor stabilize images on their retinas — there is always eye and head movement under natural viewing conditions. In fact the image velocity distributions measured in human experiments (Steinman and Collewijn 1980) are similar to the distributions that we find lead to excellent agreement with the measurements (Equation 17 for $P(v)$).

One aspect that was not explicitly treated in the analysis is the issue of occlusion. We believe that this effect could be important in the regime of high spatial and low temporal frequencies (the ratio of w/f is less than, say, 1 degree/second). This regime,

for the most part, is outside what we have measured and that is why we were successful in accounting for the data without taking occlusion into effect. However, we expect to need to worry about occlusion if we extend our measurements to smaller w/f ratios. Preliminary investigation into that regime shows that the effect of occlusion is also a function of (w/f) multiplied by the spatial power term, and hence could be absorbed into an effective velocity distribution. Thus the analysis presented in Section 4.1 still holds even when occlusion effects are important as long as one allows for an effective velocity distribution.

Finally, there are some other interesting works (Van Hateren 1993; Eckert and Buchsbaum 1993) related to the temporal aspects of natural images and neural coding, but none have measured nor explained the intertwined nature of the spatiotemporal power spectrum as we did here. Also, it is worth noting that similar spatiotemporal power-law exists in other types of images, such as signals of photon detectors in particle reactions, which also have different spatial scales and velocity distribution; and by measuring this spatiotemporal regularity, artificial neural networks, just like biological networks, can represent and transmit information in more efficient ways (Dong and Chan 1993).

Acknowledgments

We gratefully acknowledge the discussions with Yuen-Dat Chan, Norman Redlich and Penio Penev.

References

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

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

[3] Burton GJ, Moorhead IR, 1987. Color and spatial structure in natural scenes. Applied Optics. 26(1): 157-170.

[4] Dong DW, Atick JJ, 1995 Temporal Decorrelation: a Theory of Lagged and Nonlagged Responses in the Lateral Geniculate Nucleus. Network: Computation in Neural Systems, **6**, 159-178.

[5] Dong DW, Chan YD, 1993. Neural network for recognizing Cerenkov radiation patterns. LBL-33634. Three layer network for identifying Cerenkov radiation patterns. Proc World Congress on Neural Networks, Portland. (Hillsdale, NJ: Erlbaum Associates) 1: 312-315

[6] 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.

[7] 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.

[8] Hancock PJB, Baddeley RJ, Smith LS, 1992. The principal components of natural images.. *Network: computation in neural systems* **3** 1: 61-70.

[9] Van Hateren JH, 1993. Spatiotemporal Contrast sensitivity of early vision. Vision Res. V 33 N 2, 257-267.

[10] Kretzmer ER, 1952. Statistics of television signals. The bell system technical journal. 751-763.

[11] Press WH, Teukolsky SA, Vetterling WT, Flannery BP, 1992. Numerical Recipes: The Art of Scientific Computing, Second Edition, (Cambridge Press, Cambridge)

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

[13] Shannon, CE, Weaver W, 1949. A mathematical theory of communication. Univ. of Illinois Press, Urbana.

[14] Steinman RM, Collewijn H, 1980. Binocular retinal image motion during active head rotation. Vision research. 20: 415-429.

[15] Tolhurst DJ, Tadmor Y, Chao T, 1992. Amplitude spectra of natural images. Opthal. Physiol. Opt. 12:229-232.