Spatiotemporal Inseparability of Natural Images and Visual Sensitivities

Dawei W. Dong

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

Email: dawei@dove.ccs.fau.edu

(*Computational, neural & ecological constraints of visual motion processing*, Zanker JM, Zeil J (Eds) page 371-380, 2001)

Abstract

The visual system is concerned with the perception of objects in a dynamic world. A significant fact about natural time-varying images is that they do not change randomly over space-time; instead image intensities at different times and/or spatial positions are highly correlated. We measured the spatiotemporal correlation function – equivalently the power spectrum – of natural images and we find that it is non-separable, i.e., coupled in space and time, and exhibits a very interesting scaling behaviour. This behaviour is shown to be related to the motion in the images and the power spectrum is naturally separable into a spatial term and a velocity term. The same kind of spatiotemporal coupling and scaling exists in visual sensitivity measured in physiological and psychophysical experiments. By assuming that the visual system is optimized to process information of natural images, a quantitative relationship can be derived between the power spectrum of natural images and the visual sensitivity, This reveals some interesting aspects of motion vision.

Statistics of natural time-varying images

Interest in properties of time-varying images dates back to the early days of development of the television (Kretzmer 1952). 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). The statistical properties of time-varying images, on the other hand, has been studied more carefully in recent years (Dong and Atick 1995a). We will briefly summarize the results first and later on will verify our assertion that the power spectrum is dominated by image motion and then will further relate the power spectrum to motion vision.

We collected many samples of natural time-varying images from recordings of a moving camera and analyzed their spatial-temporal statistics. We measured the spatiotemporal correlation function — or the power spectrum — for an ensemble of more than a thousand segments of motion pictures, and we find significant regularities. Figure 1 illustrates the spatial and temporal scaling behaviour found in natural images (adapted from Dong and Atick 1995a). It shows that the natural time-varying images do not change randomly over space-time; instead, image intensities at different times and/or spatial positions are highly correlated. 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:

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

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

$$R \sim \frac{1}{w^b}. \tag{2}$$

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

A straight forward inspection of Figure 1 shows that the power spectrum cannot be separated into pure spatial and pure temporal parts, space and time are coupled in a non-trivial way. A more careful examination of the power spectrum showed that spatial and temporal power spectra of natural images are intertwined in a special way

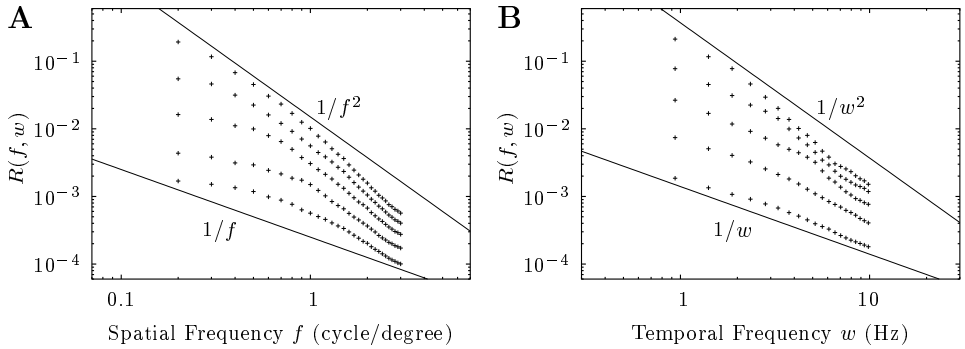

Figure 1: Measured spatial and temporal power spectra of natural time-varying images. In **(A)** the temporal frequency increases from 1.4, 2.3, 3.8, 6, to 10 Hz as we go from the highest to the lowest curve, while in **(B)** the spatial frequency increases from 0.3, 0.5, 0.8, 1.3, to 2.1 cycle/degree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

related to relative motions of objects and observer. To see this we have replotted the power spectrum in Figure 2A as a function of spatial frequency f but for fixed w/f ratio. One can see clearly that the curves for different w/f ratios are just a horizontal shift from each other and all of them follow a very precise power law, i.e., a straight line in log-log plot:

$$\frac{1}{f^{m+1}}. \tag{3}$$

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 2B, which means that

$$R(f,w) = \frac{1}{f^{m+1}}F(w/f). \tag{4}$$

This exhibits a very interesting scaling behaviour in which the power spectrum is non-separable, i.e., coupled in space and time; but is separable into two functions of the spatial frequency f and the ratio of the temporal and spatial frequencies w/f, respectively.

Spatiotemporal inseparability and motion of visual scene

The special property shown in Figure 2 and equation (4) tells a lot about the origin of the spatiotemporal power spectrum. Assuming that the dominant contribution to the spatiotemporal variability in the signal is relative motion, one can derive that

$$R(f,w) = \frac{1}{f}R_s(f)P(w/f) \tag{5}$$

in which $R_s(f)$ is the spatial power spectrum of the static objects and $P(v)$ is the probability distribution of one-dimensional image velocity v due to the relative motion of the objects to the observer (Dong and Atick 1995a).

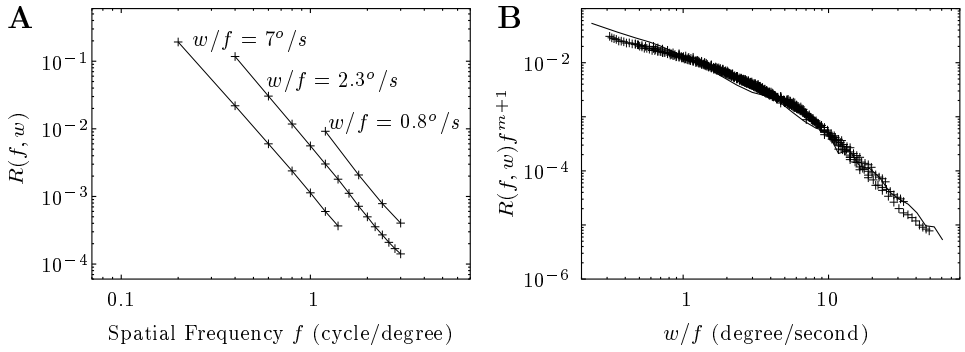

Figure 2: Scaling behaviour of spatiotemporal power spectrum. (**A**) the power spectrum is plotted for three velocities — ratios of temporal and spatial frequencies — (0.8, 2.3, 7) degree/second. (**B**) the power spectrum is replotted as a function of w/f after multiplication by f^{m+1} — all the data points fall on a single curve. The solid curve is the velocity distribution $P(v)$ multiplied by the constant C_s (see equation (7)).

By measuring the static power spectrum for this collection of images (frames treated as snapshots), we find

$$R_s(f) = \frac{C_s}{f^m}. \tag{6}$$

with $m = 2.3$ for some constant C_s — this is precisely the same m that we get from Figure 2. The measured static power spectrum is shown in Figure 3A, which is in general agreement with other earlier measurements on static images (Burton and Moorhead 1987, Field 1987, Ruderman and Bialek 1994). In most of the measurements, $m \sim 2$. Substituting equation (6) into (5), we arrive at

$$R(f,w) = \frac{C_s}{f^{m+1}} P(w/f) \tag{7}$$

This indicates that all the data points $R(f,w)f^{m+1}/C_s$ should fall on a curve which has a direct physical interpretation: the velocity density distribution of the images.

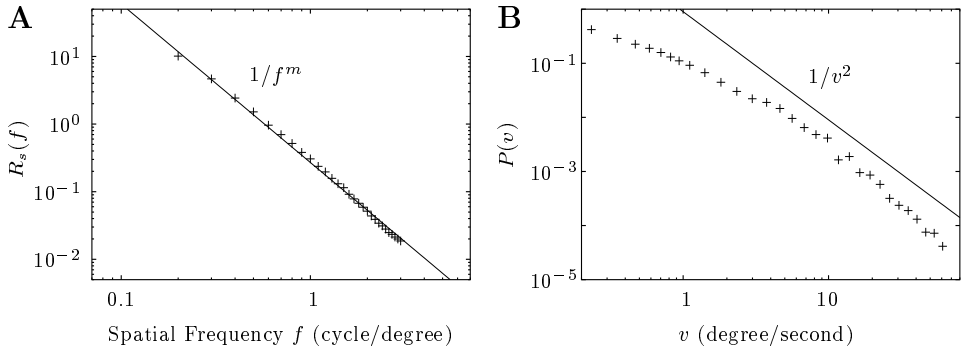

Figure 3: Spatial power spectrum and velocity distribution. (**A**) Measured spatial power spectrum of snap shot images. It shows that $R_s(f) \sim 1/f^m$ is a good approximation to the spectrum (the solid line). In our measurement $m = 2.3$. (**B**) Measured velocity density distribution $P(v)$. For reference, the $1/v^2$ is plotted (the solide line).

To verify the above claim, we measured the velocity distribution of the same collection of images by calculating the optical flow fields between frames using the

following method: for a small area of image in one frame, find its translated one in the next frame (i.e., the one with the least-square-difference) thus get the large movement in number of pixels, then calculate the sub-pixel small movements by

$$v_x \frac{\partial I(x,y)}{\partial x} + v_y \frac{\partial I(x,y)}{\partial y} = -\frac{\partial I(x,y)}{\partial t} \tag{8}$$

The one-dimensional velocity density distribution is calculated by accumulating over all possible spatial and temporal locations (for references to similar methods, e.g., Jain and Jain 1981 and Horn et al 1981). The measured $P(v)$ is shown in Figure 3B. It is interesting to see that in certain regime $P(v) \sim 1/v^2$, confirming some similar velocity distributions proposed earlier (Van Hateren 1993, Dong and Atick 1995a).

The $P(v)$ curve multiplied by the constant C_s is plotted with $R(f,w)f^{m+1}$ in Figure 2B (m and C_s are determined from the static power spectrum). It confirms the equation (7).

Visual sensitivities to moving stimuli

Given the measured spatiotemporal power spectrum, it is not difficult to predict the filter which is optimized for transmitting information from nature scenes. Do the measurements of visual systems' responses to moving stimuli agree with the theoretical prediction?

We have derived earlier (Dong and Atick 1995b) that the optimal coding requires decorrelation when the signal-to-noise ratio is high and requires smoothing where noise is significant. We derived the following relationship for the visual sensitivity K and the power spectrum R in the presence of noise power N:

$$K = \frac{1}{\sqrt{R}} \frac{1}{(1+N/R)^{\frac{3}{2}}}. \tag{9}$$

This relationship is compared with experimental data from both cat and human. In those experiments, the stimuli are gratings of spatial frequency f moving at velocity v (which is also characterized as changing at temporal frequency $w = fv$).

As shown in the previous section, the spatiotemporal power spectrum of natural images has a special form which is not separable in space and time but separable in space and velocity, i.e., equation (7) can be rewritten as the product of two functions which only depend on the spatial frequency f and the velocity $v = w/f$, respectively:

$$R(f,v) = \frac{C_s}{f^{m+1}} P(v) \tag{10}$$

Assuming white noise power N, the equation (9) and (10) predict: for any fixed spatial frequency f, the contrast sensitivity K only depends on the velocity v; and for any fixed velocity v, the contrast sensitivity K only depends on the spatial frequency f.

In general the $P(v)$ is not a simple power-law function. But there is a regime of interest the $P(v)$ approximates a universal power-law $\sim 1/v^2$ (see Figure 3B). 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 1995b). It is easy to show that our theoretical prediction in this case is

$$K(v) \sim \frac{\sqrt{v^2 C_f}}{(1 + v^2 C_f)^{\frac{3}{2}}}, \tag{11}$$

in which C_f only depends on f and is a constant for the contrast sensitivity curve of different velocities. This is verified with experimental data from cat LGN (Troy 1983a) where single cell recordings were made for contrast sensitivity to a grating of fixed spatial frequency moving at various velocities (Figure 4A). It is clear that the agreement is very good.

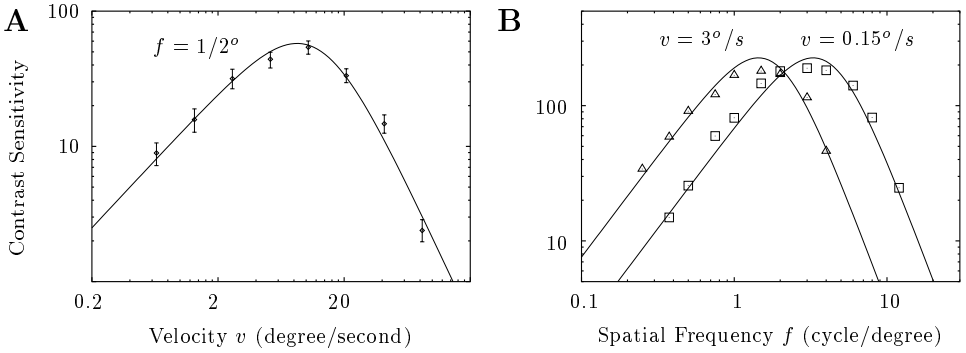

Figure 4: Comparison between predicted contrast sensitivity and experimental data. (**A**) Predicted contrast sensitivity curve for different velocity (solid curve) and cat physiological data (diamond symbols and error bars from twenty-seven X LGN cells of nine cats measured by Troy 1983a), (**B**) Predicted spatiotemporal contrast sensitivity curves (solid curves) and human psychophysical data, triangle symbols and square symbols are measured by Kelly (1979) for two velocities — w/f ratios (0.15, 3) degree/second.

Similarly, for fixed velocity v, the theory predicts (assuming $m \sim 2$):

$$K(f) \sim \frac{\sqrt{f^3 C_v}}{(1 + f^3 C_v)^{\frac{3}{2}}}, \tag{12}$$

in which C_v only depends on v and is a constant for a contrast sensitivity curve of spatial frequency. Not only so, for different velocity v, the curves in the log-log plot are only parallel shifts from each other along the spatial frequency axis. Again, as shown in Figure 4B, the theory agrees with experimental data (Kelly 1979) very well.

In fact, except at very high and very low spatial or temporal frequencies, the entire spatiotemporal contrast sensitivity surface can be represented by one curve over a single variable which is the spatial frequency scaled by velocity density distribution function (Dong 1997).

Discussion

We should point out that while our predictions show that, in general, the human visual sensitivity is strongly space-time coupled, we do predict a regime where decoupling is a good approximation. This is based on the fact that in certain regime the velocity density distribution approximates a universal power-law $\sim 1/v^2$ thus the power spectrum of natural images is separable into spatial and temporal parts. In a previous work we have used this decoupling to model response properties of cat LGN cells and where we have shown that these can be accounted for by the theoretical prediction based on the power spectrum in that regime (Dong and Atick 1995b). The cat LGN data used in current paper were generated from moving gratings and different from the data we modeled earlier which were generated from temporal modulation of receptive field centers. The striking agreements in both cases give very strong support of the theory that the visual system is optimized to process spatiotemporal information of natural images.

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 illustrated suggests a natural way for dealing with this coupling. More precisely, it suggests that a better way to examine spatiotemporal tuning data from real visual system is to plot the data not as a function of f and w separately but as a function of w/f. This is a natural representation for motion vision. In this representation we expect that vision will exhibit more universal behaviour as we have shown for the human psychophisical data. We should point out that this scaling behaviour is expected to break down for very high temporal and spatial frequency where the effect of the temporal and spatial modulation function of the eye (Campbell and Gubisch 1966, Schnapf and Baylor 1987) cannot be ignored. Also this scaling behaviour might not show up in single cell since in real visual system cells could have different spatiotemporal receptive fields, for example there are different cell types and even same type of cells could tune to different spatial frequencies in LGN (Troy 1983b). So an ecological theory of relating natural image properties with visual processes need to make distinctions about single cells and ensembles (Atick

1992) and different type of cell ensembles can participate in overall efficient coding (Van Essen and Anderson 1990, Li 1992, Eckert and Buchsbaum 1993).

Finally, we want to comment on the asymptotic behaviours of our measured power spectrum. While the detailed form of the function $F(w/f)$ depends on the image velocity distribution $P(v)$ which in turn depends on the distributions of motion velocities and spatial distances of visual objects relative to observer, there is an interesting asymptotic regime where the form of $P(v)$ is $\sim 1/v^2$. This gives rise to the asymptotic behaviour at relative low spatial frequency and relative high temporal frequency:

$$R(f,w) \sim \frac{1}{f^{m-1}w^2} \tag{13}$$

On the other end of this, i.e., relative high spatial frequency and relative low temporal frequency:

$$R(f,w) \sim \frac{1}{f^{m+1}} \tag{14}$$

In general, from equation (10), whenever $P(v)$ can be approximated as $\sim 1/v^a$, the power spectrum can be approximated as

$$R(f,w) \sim \frac{1}{f^{m+1-a}w^a} \tag{15}$$

Our measured power spectrum does not include a regime of

$$R(f,w) \sim \frac{1}{f^m w^2} \tag{16}$$

with $m \geq 2$, such as what was observed by Eckert etc 1992. On the other hand, another simple relationship is implied in the coupled spatiotemporal behavior: when temporal power spectrum is measured for a single spatial point (Van Hateren 1997) which moves randomly across static image, it is not difficult to show that

$$R(w) \sim \frac{1}{w^{m-1}} \tag{17}$$

in which the temporal power index is simply the spatial power index minus one. This behavior can also be derived by integrating the spatiotemporal power spectrum over all spatial frequencies. For more discussion about the origins of scaling in natural images see Ruderman 1997.

References

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

[2] Burton GJ, Moorhead IR, 1987. Color and spatial structure in natural scenes. Applied Optics. 26 : 157–170.

[3] Campbell FW, Gubisch RW, 1966. Optical quality of the human eye. J. Physiol. 186: 558–578.

[4] Dong DW, 1997. Spatiotemporal coupling and scaling of natural images and human visual sensitivities. In: Advances in Neural Information Processing Systems 9 (Mozer MC, Jordan MI, Petsche T, eds) MIT Press, Cambridge, MA pp 859–865.

[5] Dong DW, Atick JJ, 1995a. Statistics of natural time-varying images. Network-Comp. Neural 6: 345–358.

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

[7] Eckert MP, Buchsbaum G, Watson AB, 1992. Separability of spatiotemporal spectra of image sequences. IEEE T. Pattern Anal. 14 : 1210–1213.

[8] 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.

[9] Field DJ, 1987. Relations between the statistics of natural images and the response properties of cortical cells.. J. Opt. Soc. Am. A : 2379–2394.

[10] Hancock PJB, Baddeley RJ, Smith LS, 1992. The principal components of natural images. Network-Comp. Neural 3 : 61–70.

[11] Horn BKP, Schunk BG (1981) Determining optical lofw Artificial Intelligence 17: 185.

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

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

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

[15] Li Z, 1992. Different retinal ganglion cells have different functional goals. Int J of Neural Systems 3: 237–248.

[16] Ruderman DL, 1997. Origins of scaling in natural images. Vision Res. 37 : 3385–3398.

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

[18] Schnapf JL, Baylor DA, 1987. How photoreceptor cells respond to light. Scientific American 256 : 40–47.

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

[20] Troy JB, 1983a. Spatio-temporal interaction in neurons of the cats dorsal lateral geniculate nucleus. J. Physiol. 344: 419–423.

[21] Troy JB, 1983b. Spatial contrast sensitivities of X and Y type neurons in the cats dorsal lateral geniculate nucleus. J. Physiol. 344: 399–417.

[22] Van Essen DC, Anderson CH, 1990. Information processing strategies and pathways in the primate retina and visual cortex. In: Introduction to Neural and Electronic Networks (Zotnetzer SF, Davis JL, Lau C, eds) Academic Press, Orlando, FL pp 43–72.

[23] Van Hateren JH, 1993. Spatiotemporal Contrast sensitivity of early vision. Vision Res. 33 : 257–267.

[24] Van Hateren JH, 1997. Processing of natural time series of intensities by the visual system of the blowfly. Vision Res. 37 : 3407–3416.