IN-61-CR
78073
p-16

Separability of Spatiotemporal Spectra of Image Sequences

Michael P. Eckert, Gershon Buchsbaum, and Andrew B. Watson*

University of Pennsylvania
School of Engineering and Applied Science
Department of Bioengineering
220 S. 33rd St.
Philadelphia, PA 19104-6392

*NASA Ames Research Center
MS 262-2
Moffett Field, CA 94035-1000

Phone: (215) 898-5767
FAX: (215) 573-2071

Email: gershon@eniac.seas.upenn.edu

Separability of Spatiotemporal Spectra of Image Sequences

Summary

We calculated the spatiotemporal power spectrum of 14 image sequences in order to determine the degree to which the spectra are separable in space and time, and to assess the validity of the commonly used exponential correlation model found in the literature. We expand the spectrum by a Singular Value Decomposition into a sum of separable terms and define an index of spatiotemporal separability as the fraction of the signal energy that can be represented by the first (largest) separable term. All spectra were found to be highly separable with an index of separability above 0.98. The power spectra of the sequences were well fit by a separable model of the form:

$$P(k,f) = \frac{ab/(4\pi^3)}{((a/2\pi)^2 + k^2)^{3/2}((b/2\pi)^2 + f^2)}$$

where k is radial spatial frequency, f is temporal frequency, and a,b are spatial and temporal model parameters which determine the effective spatiotemporal bandwidth of the signal. This power spectrum model corresponds to a product of exponential autocorrelation functions separable in space and time.

Introduction

The statistics of images and image sequences have been extensively studied for image coding and compression applications [1,2], as well as for the development of models of biological image processing [3, 4]. An exponential autocorrelation function has been shown to be a good model for temporal frame to frame correlations of image sequences [e.g., 5, 6, 7, 8], and for spatial correlations within each frame [e.g., 2, 3, 9].

This paper focuses on the separability of the spatiotemporal statistics of image sequences and on the validity of using a separable exponential autocorrelation model for the spatiotemporal statistics. The autocorrelation function is uniquely related to the power spectrum via a Fourier transform, and either is valid as a description of the statistics.

The spectra of 14 image sequences were calculated. The sequences represented a small ensemble of possible motion activity. The sequences were selected for a range of motion activity. For example, a fast camera pan represents the maximum image motion activity and a small moving object with a static background represents the least activity. Sequences with motion activity between these extremes had slight camera motion and some object motion.

Calculation of Image Statistics

We collected 14 image sequences (256x256x64 @ 8 bits/pixel, 30 frames/second with no scene cuts) from a video disc which contained scenes from a broadcast TV source. Each frame was originally sampled at 512x512 pixels/screen, but adjacent pixels were averaged, and the image was subsampled to 256x256 pixels/screen. The sample mean of each sequence was removed to reduce low frequency bias in the calculations.

The sample power spectrum, $P(k_1,k_2,f)$ of each sequence, $x(n_1,n_2,t)$, is the squared magnitude of the Discrete Fourier Transform, calculated as

$$P(k_1,k_2,f) = \frac{1}{256\cdot 256\cdot 64}\left|\sum_{n_1=0}^{255}\sum_{n_2=0}^{255}\sum_{t=0}^{63} x(n_1,n_2,t)e^{-j2\pi(k_1n_1+k_2n_2+ft)}\right|^2 \qquad (1)$$

where k_1,k_2 are spatial frequencies, f is temporal frequency, n_1,n_2 are spatial locations, and t is time measured in frame number.

We converted the two spatial frequency dimensions, k_1 and k_2, into one radial frequency dimension, k, by averaging in 32 annuli around the spatial frequency origin as illustrated in Fig. 1. In this manner, the spatial frequency range of 0-127 cycles/screen of k_1 and k_2 is represented by 32 annuli in bands of 4 cycles/screen. Averaging the spatial spectra in annuli is equivalent to assuming a circularly symmetric spatial autocorrelation function. This autocorrelation function is not separable in the two spatial dimensions, but is considered a better fit than the corresponding separable autocorrelation function for most images [9].

Figure 1 near here

The average magnitude of the power spectrum in each annulus can be obtained by summing over the power spectrum, $P(k_1,k_2,f)$, in the annulus indexed by k and normalizing by the number of sample points, A(k), within each annulus

$$P(k,f) = \frac{1}{A(k)} \sum_{k_1^2 + k_2^2 \geq k^2}^{k_1^2 + k_2^2 < (k+4)^2} P(k_1,k_2,f) \qquad k = 0,4,8, \ldots 124 \qquad (2)$$

where

$$A(k) = \sum_{k_1^2 + k_2^2 \geq k^2}^{k_1^2 + k_2^2 < (k+4)^2} 1 \qquad k = 0,4,8, \ldots 124 \qquad (3)$$

The resulting 14 sample spectra were described in terms of a 33 (temporal frequency) x 32 (spatial frequency) matrix, **P**, with the spatial frequency axis ranging from 0-127 cycles/screen in steps representing bands of 4 cycles/screen, and the temporal frequency axis ranging from 0-15 Hz in steps of 15/32 Hz each.

Models of space-time statistics

The most commonly used statistical model for intraframe correlations and frame to frame correlations is an exponential correlation model in both space and time,

$$R(\nu) = e^{-a|\nu|} \qquad (4)$$

$$R(\tau) = e^{-b|\tau|} \qquad (5)$$

where ν represents a two-dimensional spatial lag, τ represents temporal lags, and a,b are spatial and temporal parameters. A separable formulation for the spatiotemporal correlation of image sequences is found as a product of (4) and (5). An equivalent description of the statistics is the power spectrum, which for the exponential correlation function of Eqs. 4 and 5 would be

$$S(k) = \frac{a/2\pi}{((a/(2\pi))^2 + k^2)^{3/2}} \qquad k \geq 0 \qquad (6)$$

$$T(f) = \frac{b/2\pi^2}{(b/(2\pi))^2 + f^2} \qquad -\infty < f < \infty \qquad (7)$$

where k is radial spatial frequency, f is temporal frequency, and a is a spatial parameter with units of cycles/screen, and b is a temporal parameter with units of Hertz. The parameters a and b describe the effective spatial and temporal bandwidth of the signal. A spatial power spectrum (Eq. 6) has 85% of its power in the frequency band $k \le a$. The temporal power spectrum (7) has 90% of its power in the band $f \le |b|$. A separable spatiotemporal power spectrum is formed as the product of Eqs. (6) and (7).

$$P(k,f) = \frac{ab/(4\pi^3)}{((a/2\pi)^2 + k^2)^{3/2}((b/2\pi)^2 + f^2)} \quad . \qquad (8)$$

Singular Value Decomposition and Index of Separability

A space-time separable spectrum is modeled as the product of a spatial and temporal spectrum (as in Eq. 8). In this section we define an index of separability for an arbitrary spectrum, P(k,f), based on a Singular Value Decomposition.

Any m x n matrix, **D**, with $m \ge n$ may be expanded into a sum of terms by a Singular Value Decomposition [10, 11],

$$\mathbf{D} = \sum_{i=1}^{n} \sqrt{\lambda_i}\,\mathbf{v}_i\mathbf{u}_i \qquad (9)$$

where

$\lambda_1 \ge \lambda_2 \ge \ldots \lambda_n$ are the real non-negative eigenvalues of the n-th order symmetric matrix $\mathbf{S} = \mathbf{D}^T\mathbf{D}$. $\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_n$ are normalized, orthogonal row eigenvectors associated with the corresponding eigenvalues $\lambda_1 \ge \lambda_2 \ge \ldots \lambda_n$ of **S**. $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n$ are normalized, orthogonal column eigenvectors associated with the corresponding eigenvalues $\lambda_1 \ge \lambda_2 \ge \ldots \lambda_n$ of the m-th order symmetric matrix $\mathbf{Q} = \mathbf{D}\mathbf{D}^T$, where **Q** can have a maximum of n nonzero eigenvalues which are the same as those of S. In the case of duplicate eigenvalues, an orthonormal combination of eigenvalues can be selected.

Approximating **D** by the first term of the decomposition,

$$\mathbf{D}' = \sqrt{\lambda_1}\mathbf{v}_1\mathbf{u}_1 \quad , \tag{10}$$

gives the minimum mean squared error separable approximation to **D**, where the mean squared error is

$$e = \sum_{i=1}^{n} \sum_{j=1}^{m} (d-d')^2_{ij} \tag{11}$$

where d_{ij} and d_{ij}' are the elements of **D** and **D'** respectively. Noting that

$$\sum_{i=1}^{n} \sum_{j=1}^{m} d^2_{ij} = \sum_{i=1}^{n} \lambda_i \tag{12}$$

and

$$\sum_{i=1}^{n} \sum_{j=1}^{m} d'^2_{ij} = \lambda_1$$

the mean square error between the approximate matrix **D'** and the true matrix **D** is determined by the eigenvalues as,

$$e = \lambda_2 + \lambda_3 + \ldots \lambda_n. \tag{13}$$

We define an index of separability, α, as the relative energy share of **D'**,

$$\alpha = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \ldots \lambda_n} \tag{14}$$

Since $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n \geq 0$, α will range from 1/n for the most inseparable spectrum to 1 for a completely separable spectrum. The eigenvalues represent the energy carried by each term of the expansion in Eq. 9. The index of separability, α, is simply the fraction of the total energy carried by the first and largest term in the expansion, the term which constitutes the best separable approximation.

We applied the Singular Value Decomposition to the spatiotemporal spectra by considering each spectrum as a matrix **P** of dimension 33x32. As shown in Eq. 9, **P** can be expanded as

$$\mathbf{P} = \sum_{i=1}^{32} \sqrt{\lambda_i}\,\mathbf{t}_i\mathbf{s}_i \qquad (15)$$

where $\mathbf{s}_i$ are now orthonormal row vectors representing spatial spectra and $\mathbf{t}_i$ are orthonormal columns vectors representing temporal spectra in each term of the sum. A separable approximation of the form

$$\mathbf{P}' = \sqrt{\lambda_1}\,\mathbf{t}_1\mathbf{s}_1 \qquad (16)$$

exists where $\mathbf{s}_1$ and $\mathbf{t}_1$ represent the spatial and temporal components of the separable approximation. The normalized energy share of this term is α, the index of separability. Examination of α for the spatiotemporal spectra of the 14 image sequences (Table 1) shows that for 13 out of the 14 sequences, $\alpha > 0.993$, which constitutes a high degree of separability [10]. Though for one sequence, the separability was low ($\alpha = 0.982$). This suggests that a space-time separable model, such as Eq. 8, may adequately describe the spatiotemporal spectrum of image sequences since the assumption of separability is valid. The extraction of nearly all the energy with the separable term is also significant for perceptual reasons since small fractions of image energy can markedly affect the perception of some images [12].

Calculation of Model Parameters

Since the spatiotemporal spectra of the image sequence, **P**, are all highly separable, we need only determine whether the model of Eq. 8 adequately characterizes the frequency distribution of the spectra and find the spatial and temporal parameters *a* and *b*. This will determine whether the commonly used model defined by a separable exponential autocorrelation in space and time is satisfactory.

We find the model parameters *a* and *b* by minimizing the mean squared error between the actual signal spectra, **P**, of Eq. 2, and the analytical separable model of Eq. 8.

$$\min[\ (\mathbf{P} - P(k,f))^2\ ] \qquad (17)$$

The optimal parameters *a*,*b* for each of the sequences were calculated using the Nelder-Meade simplex algorithm [13]. The mean squared error between the analytical separable model (8) and the true spectrum, expressed as a percentage of the average squared power of the spectrum is small

(0.03% < mse < 4.7%) and is given in Table 1. The parameters *a* and *b* determine the effective bandwidth for the spatiotemporal power spectrum. Fig. 2 illustrates the relationship between the parameters *a* and *b* for all 14 sequences, and thus the simultaneous spatial and temporal bandwidths. All of the pairs of *a* and *b* are located within a well defined range for this ensemble such that no sequence contains both high spatial and high temporal frequencies.

Figure 2 near here

The separable kernel in the model of Eq. 8 is based on theoretical considerations, mainly statistical properties of Markov processes as models for image signals. It is interesting to investigate how this theoretical separable model captures the functional shape of the spectra in spatial and temporal frequency compared to the empirically derived separable kernels derived by the Singular Value Decomposition. The empirically derived kernels are not constrained by a predetermined functional shape as is the theoretical model. We compare the spatial and temporal components of the analytical separable model to the corresponding components of the separable approximation (Eq. 16). Four examples are shown in Figures 3 and 4. The model provides a good fit for the sample signal spectra in all frequency ranges. (Note that the ordinate scale is logarithmic, so the contribution to the mean squared error is small at high frequencies.) This finding is consistent with the applicability of the models of Eqs. 6 and 7 in earlier studies of spatial and temporal statistics [2, 5, 7, 8, 9].

Figures 3 and 4 near here

Discussion

We calculated the spatiotemporal power spectra of 14 image sequences to investigate whether these spectra are separable in space and time. Using a newly defined index of separability, we show that a separable approximation for the spectra derived from the Singular Value Decomposition extracts over 98% of the signal energy (Table 1). We also investigated whether the space-time separable exponential model commonly used in the literature provides a reasonable description of the statistics of image sequences. This exponential model is equivalent to the space-time separable power spectrum model of Eq. 8. We show that this model provides a good analytical description of the spectrum of image sequences.

For this ensemble of image sequences, no sequence possessed both high spatial and high temporal frequencies (Fig. 2). This property may be a result of spatial blurring caused by motion.

If so, it is not an inherent property of the image sequence, but rather caused by the low pass temporal filtering of the camera. The visual system also temporally low pass filters images (mainly due to photoreceptor integration time), so this property holds true for a signal perceived by the visual system as well. This limitation on signal spatiotemporal bandwidth may be useful for perceptually based image coding and processing applications [14].

Applications of the model to image processing accrues both the advantages and limitations of using autocorrelation and power spectrum methods. As descriptions of images, the autocorrelation and power spectra are global in the sense that they represent a calculation averaged over the entire image or image sequence. This averaging does not retain the phase spectrum of images and removes local nonstationarities and hence specific local details of images. Also, the separable model may not apply to local sections of image sequences even though the global spectrum of the sequence is separable. In those cases where the autocorrelation and power spectrum methods are applicable, the assumption of separability enables considerable mathematical simplicity. Any methods of image processing developed for spatial only or temporal only processing using Eq. 6 and 7 can be extended in a straightforward manner to spatiotemporal processing with Eq. 8.

* Michael P. Eckert was supported by the NASA Graduate Fellowship Program. Partial support for this paper was provided by grants NSF 8351637 and AFOSR 91-0082.

References

1. Jayant, N.S. and Noll, P. **Digital Coding of Waveforms: Principles and Applications to Speech and Video**. Prentice-Hall. 1985.
2. Rosenfeld, A. and Kak, C.C. **Digital Picture Processing**. Academic Press. 1982.
3. Srinivisan, M.V., Laughlin, S.B., and Dubs, A. "Predictive coding: A fresh view of inhibition in the retina". Proc. Roy. Soc. Lond. B, Vol. 216, pp. 427-459, 1982.
4. Field, D.J. "Relations between the statistics of natural images and the response properties of cortical cells". J. Opt. Soc. Am. A, Vol. 4, No. 12, pp. 2379-2394, 1987.
5. Connor, D.J. and Limb, J.O. "Properties of frame-difference signals generated by moving images". IEEE Trans. Comm., Vol. 10, pp. 1564-1575, 1974.
6. Franks, L.E. "A model for the random video process". Bell Sys. Tech. J., pp. 609-630, 1966.
7. Coll, D.C. and Choma, G.K. "Image activity characteristics in broadcast television". IEEE Trans. Comm., pp. 1201-1206, 1976.

8. Chen, W. and Hein, D. "Recursive temporal filtering and frame rate reduction for image coding". IEEE J. Special Areas Comm., Vol. SAC-5, No. 7, pp. 1155-1165, 1987.

9. Jain, A.K. "Partial differential equations and finite-difference methods in image processing, Part 1: Image representation". J. Optim. Theory Appl., Vol. 23, No. 1, pp. 65-91, 1977.

10. Treitel, S. and Shanks, J.L. "The Design of Multistage Separable Planar Filters". IEEE Trans. Geo. E., Vol. GE-9, pp. 10-27, 1971.

11. Dudgeon, D.E. and Mersereau, R.M. **Multidimensional Digital Signal Processing**. Prentice-Hall. 1984.

12. Gonzalez, R.C. and Wintz, P. **Digital Image Processing**. Addison Wesley. 1987.

13. Dennis, J.E. and Woods, D.J. "Optimization in Microcomputers: The Nelder-Meade Simplex Algorithm". In **New Computing Environments: Microcomputers in Large-Scale Computing**. SIAM, ed. (Wouk, A.), pp. 116-122. 1987.

14. Watson, A.B. "Perceptual-components architecture for digital video". J. Opt. Soc. Am. A, Vol. 7, No. 10, pp. 1943-1954, 1990.

Figure Captions:

Figure 1. Conversion from two dimensions of spatial frequency to one dimension of radial spatial frequency is done by averaging the spectrum in annuli around the spatial frequency origin. The temporal frequency axis is denoted by *f*.

Figure 2. Scatter plot of the parameters *a* and *b* for all sequences. The parameters *a* and *b* are measures of the effective spatial and temporal bandwidths of the signal spectrum. No spectrum had both a large spatial and large temporal bandwidth within the spatial and temporal frequency spans of the sequences.

Figure 3. The magnitude of the spatial component of the spectrum derived by the Singular Value Decomposition (stars) compared with the analytical model (solid line).

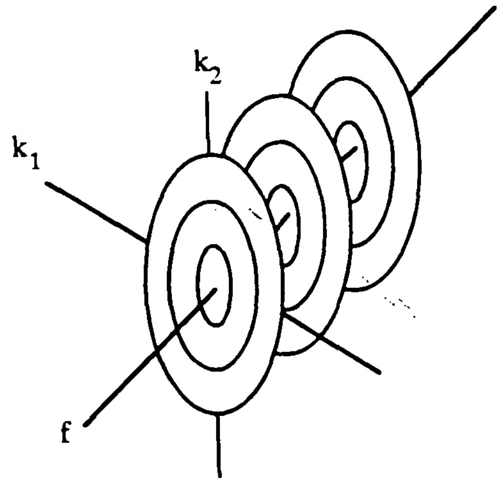
Figure 4. The magnitude of the temporal component of the spectrum derived by the Singular Value Decomposition (stars) compared with the analytical model (solid line).

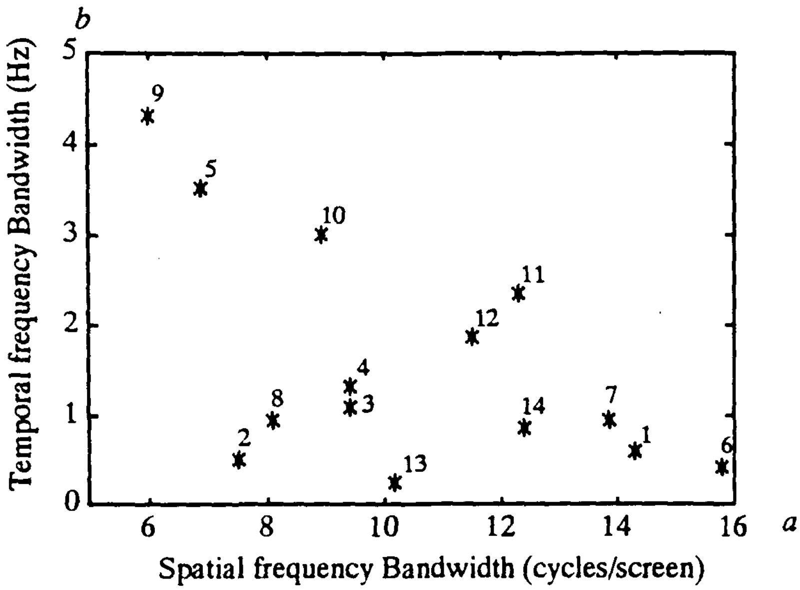
Table 1. Description of image sequences and results of calculations.

	Sequence Number	Motion Type	Index of Separability α	Spatial Parameter a	Temporal Parameter b	mse (%)
1	IJ01300	1,a	0.999	14.33	0.59	0.01
2	IJ04454	1,b	0.999	7.54	0.51	0.04
3	IJ10833	2,a	0.993	9.45	1.08	0.09
4	IJ10897	2,a	0.995	9.42	1.30	0.07
5	IJ11907	1,c	0.999	6.91	3.50	4.70
6	IJ12100	1,a	0.999	15.80	0.41	0.03
7	IJ12164	1,b	0.999	13.85	0.92	0.06
8	IJ12426	2,b	0.998	8.10	0.92	0.04
9	IJ14461	3,a	0.998	6.00	4.30	0.70
10	IJ15300	3,b	0.997	8.93	2.99	4.10
11	IJ17830	1,c	0.982	12.30	2.32	0.40
12	IJ07860	1,c	0.993	11.50	1.85	0.60
13	IJ33960	1,a	0.999	10.20	0.24	0.005
14	IJ30229	1,b	0.999	12.40	0.85	0.06

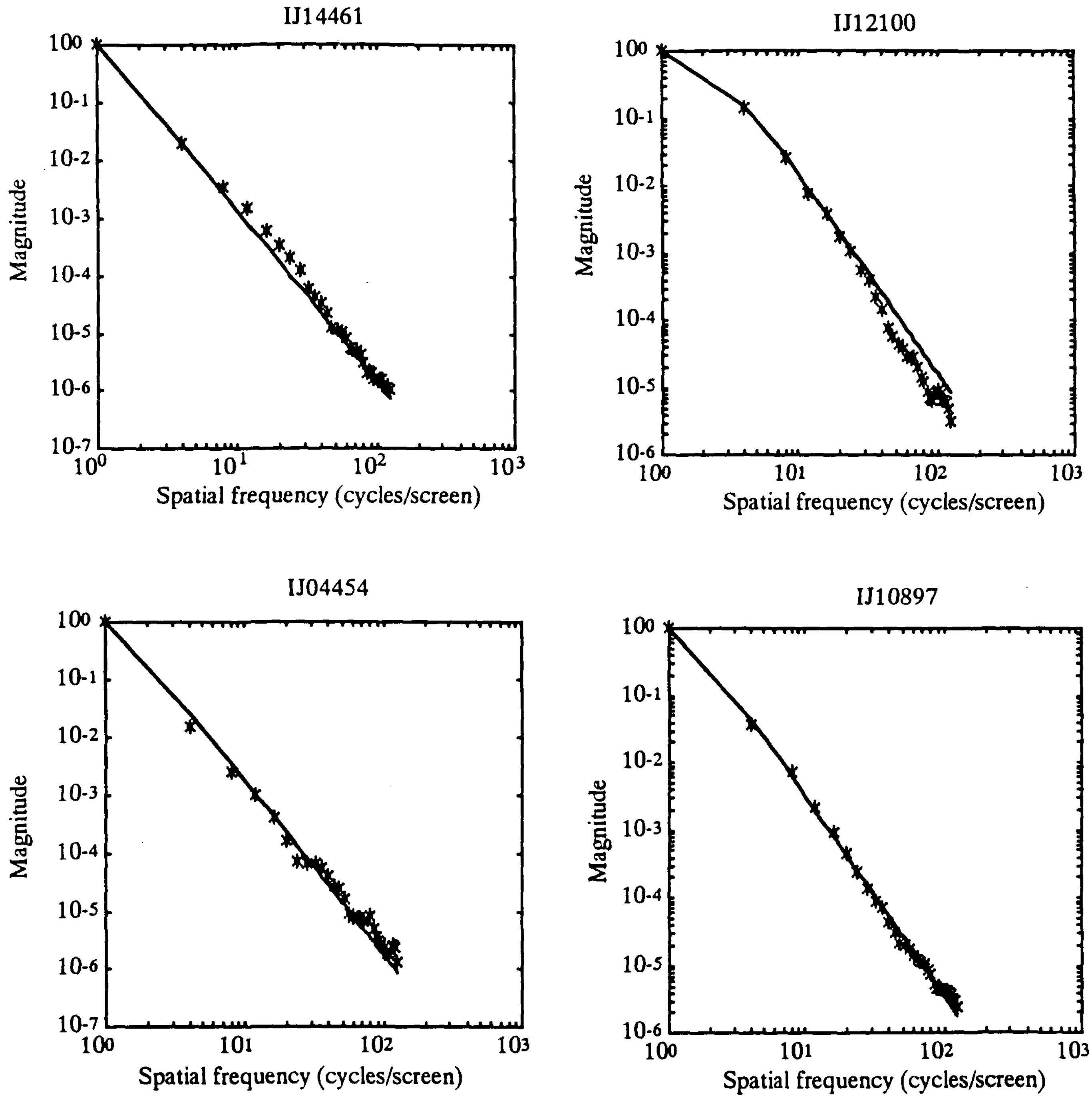
α : Index of separability, unitless (0.982,0.999)
a : Spatial parameter, cycles/screen (6.0,15.80)
b : Temporal parameter, Hertz, (0.24,4.30)
mse : The mean squared error between the actual spectrum and the model with the parameters a,b of Eq. 8. The mse is expressed as the percentage of the average power of the sequence.

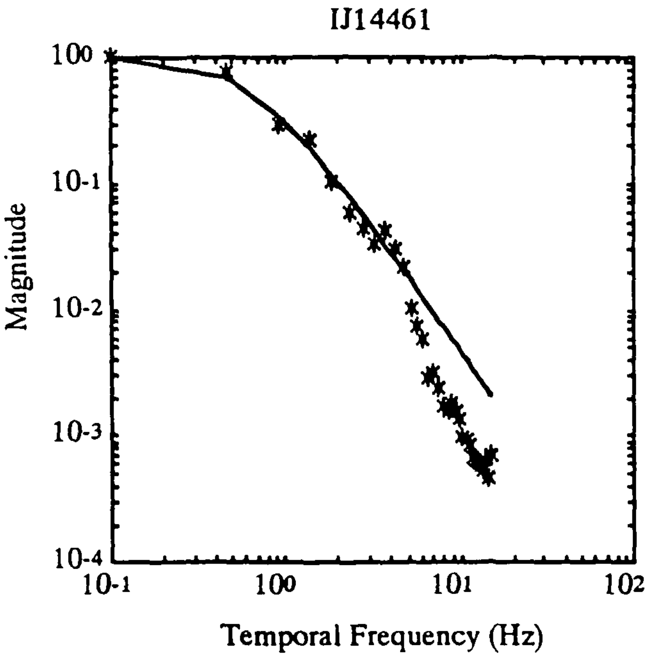
1. No camera motion
2. Some camera motion
3. Much camera motion
a. Little object motion
b. Some object motion
c. Much object motion

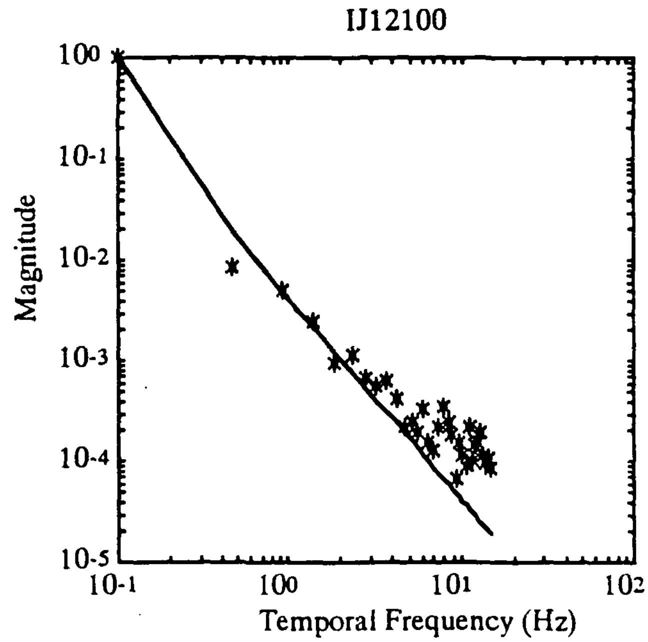

k_2
k_1
f

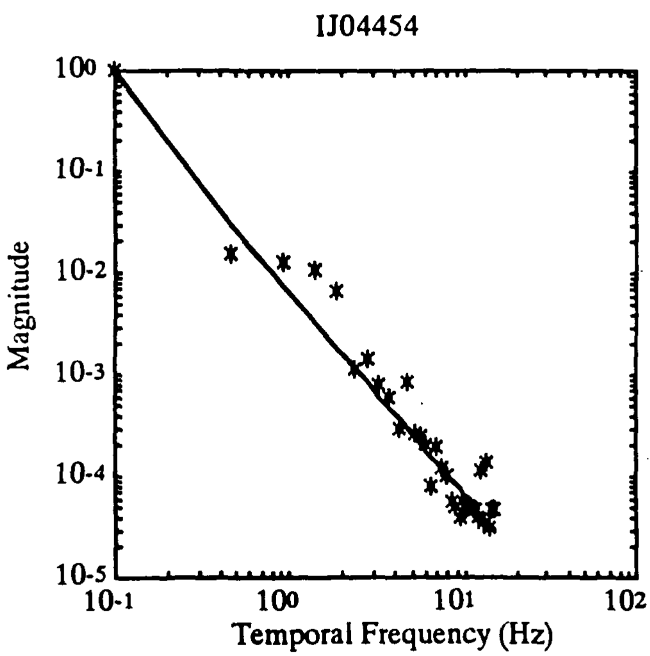


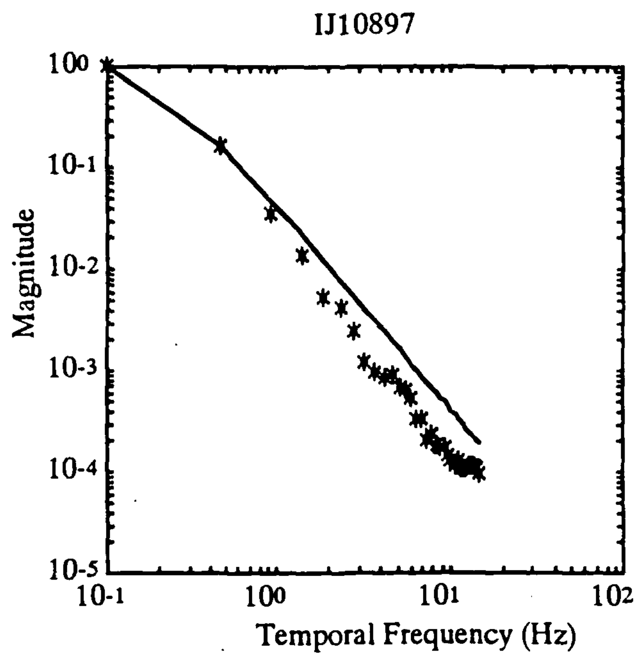
b
5
4
3
2
1
0
9
5
10
11
12
4
8
3
14
7
2
1
6
13
6
8
10
12
14
16
a
Temporal frequency Bandwidth (Hz)
Spatial frequency Bandwidth (cycles/screen)


IJ14461
IJ12100
IJ04454
IJ10897
Magnitude
Spatial frequency (cycles/screen)


IJ14461
Magnitude
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-1}
10^{0}
10^{1}
10^{2}
Temporal Frequency (Hz)


IJ12100
Magnitude
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-1}
10^{0}
10^{1}
10^{2}
Temporal Frequency (Hz)


IJ04454
Magnitude
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-1}
10^{0}
10^{1}
10^{2}
Temporal Frequency (Hz)


IJ10897
Magnitude
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-1}
10^{0}
10^{1}
10^{2}
Temporal Frequency (Hz)