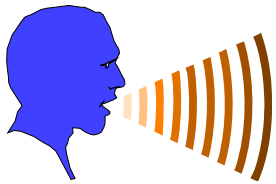





Video Processing & Communications

Fourier Analysis of Video Signals & Frequency Response of the HVS

Yao Wang
Polytechnic University, Brooklyn, NY11201

Outline

- Fourier transform over multidimensional space
- Frequency domain characterization of video signals
- Frequency response of the HVS
- Video sampling – a brief discussion

Fourier transforms over multidimensional space

- Continuous space signals
  - Defined for all real values
  - Continuous space FT (CSFT)
- Discrete space signals
  - Defined only on integers
  - Discrete space FT (DSFT)
- Sampled space signals
  - A continuous signal that is a sum of shifted impulses
  - Sampled space FT (SSFT)

- Delta function
- Linear Time/shift Invariance
- Convolution

Continuous Space Signals

- K-D Space Signals

$$\psi(\mathbf{x}), \mathbf{x} = [x_1, x_2, \ldots, x_K] \in R^k$$

- Convolution

$$\psi(\mathbf{x}) * h(\mathbf{x}) = \int_{R^k} \psi(\mathbf{x} - \mathbf{y}) h(\mathbf{y}) d\mathbf{y}$$

- Example function
  - Delta function

$$\delta(\mathbf{x}) = \begin{cases} \infty, & \mathbf{x} = \mathbf{0}, \\ 0, & \text{otherwise}, \end{cases} \quad \text{and} \quad \int_{\mathcal{R}^K} \delta(\mathbf{x})\, d\mathbf{x} = 1.$$

$$\psi(\mathbf{x}) * \delta(\mathbf{x} - \mathbf{x}_0) = \psi(\mathbf{x} - \mathbf{x}_0).$$

Continuous Space Fourier Transform (CSFT)

- Forward transform

$$\Psi_c(\mathbf{f}) = \int_{R^k} \psi(\mathbf{x}) \exp(-j2\pi \mathbf{f}^T \mathbf{x}) d\mathbf{x}$$

- Inverse transform

$$\psi(\mathbf{x}) = \int_{R^k} \Psi_c(\mathbf{f}) \exp(j2\pi \mathbf{f}^T \mathbf{x}) d\mathbf{f}$$

- Convolution theorem

$$\psi(\mathbf{x}) * h(\mathbf{x}) \Leftrightarrow \Psi_c(\mathbf{f}) H_c(\mathbf{f})$$
$$\psi(\mathbf{x}) h(\mathbf{x}) \Leftrightarrow \Psi_c(\mathbf{f}) * H_c(\mathbf{f})$$

Continuous Space Systems

- General system over K-D continuous space

$$\phi(\mathbf{x}) = T(\psi(\mathbf{x})), \mathbf{x} \in R^k$$

- Linear and Space-Shift Invariant (LSI) System

$$\alpha_1\phi_1(\mathbf{x}) + \alpha_2\phi_2(\mathbf{x}) = T(\alpha_1\psi_1(\mathbf{x}) + \alpha_2\psi_2(\mathbf{x}))$$
$$T(\psi(\mathbf{x}+\mathbf{x}_0)) = \phi(\mathbf{x}+\mathbf{x}_0)$$

- LSI systems can be completely described by its impulse response

$$h(\mathbf{x}) = T\big(\delta(\mathbf{x})\big)$$
$$\phi(\mathbf{x}) = \psi(\mathbf{x}) * h(\mathbf{x}) \Leftrightarrow \Phi_c(\mathbf{f}) = \Psi_c(\mathbf{f})H_c(\mathbf{f})$$

Discrete Space Signals

- K-D Space Signals

$$\psi(\mathbf{n}), \mathbf{n} = [n_1, n_2, \ldots, n_K] \in Z^K$$

- Convolution

$$\psi(\mathbf{n}) * h(\mathbf{n}) = \sum_{\mathbf{m} \in Z^K} \psi(\mathbf{n} - \mathbf{m}) h(\mathbf{m})$$

- Example function
  - Delta function

$$\delta(\mathbf{n}) = \begin{cases} 1, & \mathbf{n} = \mathbf{0}, \\ 0, & \text{otherwise}. \end{cases}$$

Discrete Space Fourier Transform (DSFT)

- Forward transform

$$\Psi_d(\mathbf{f}) = \sum_{\mathbf{n}\in R^K} \psi(\mathbf{n}) \exp(-j2\pi \mathbf{f}^T \mathbf{n})$$

$\Psi_d(\mathbf{f})$ is periodic in each dimension with period of 1

Fundamental period: $I^K = \{\mathbf{f}, f_k \in (-1/2, 1/2)\}$

- Inverse transform

$$\psi(\mathbf{n}) = \int_{\mathbf{f}\in I^K} \Psi_d(\mathbf{f}) \exp(j2\pi \mathbf{f}^T \mathbf{n}) d\mathbf{f}$$

- Convolution theorem

$$\psi(\mathbf{n}) * h(\mathbf{n}) \Leftrightarrow \Psi_d(\mathbf{f}) H_d(\mathbf{f})$$

$$\psi(\mathbf{n}) h(\mathbf{n}) \Leftrightarrow \Psi_d(\mathbf{f}) * H_d(\mathbf{f})$$

Frequency domain characterization of video signals

- Spatial frequency
- Temporal frequency
- Temporal frequency caused by motion

Spatial Frequency

- Spatial frequency measures how fast the image intensity changes in the image plane
- Spatial frequency can be completely characterized by the variation frequencies in two orthogonal directions (e.g horizontal and vertical)
  - f_x: cycles/horizontal unit distance
  - f_y : cycles/vertical unit distance
- It can also be specified by magnitude and angle of change

$$f_s = \sqrt{f_x^2 + f_y^2}, \quad \varphi = \arctan(f_y / f_x)$$

Illustration of Spatial Frequency

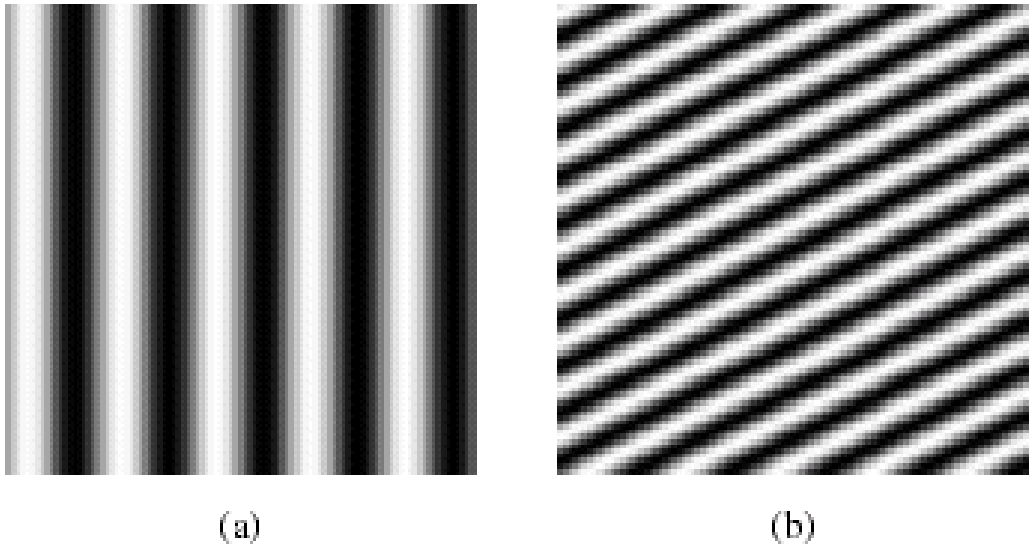
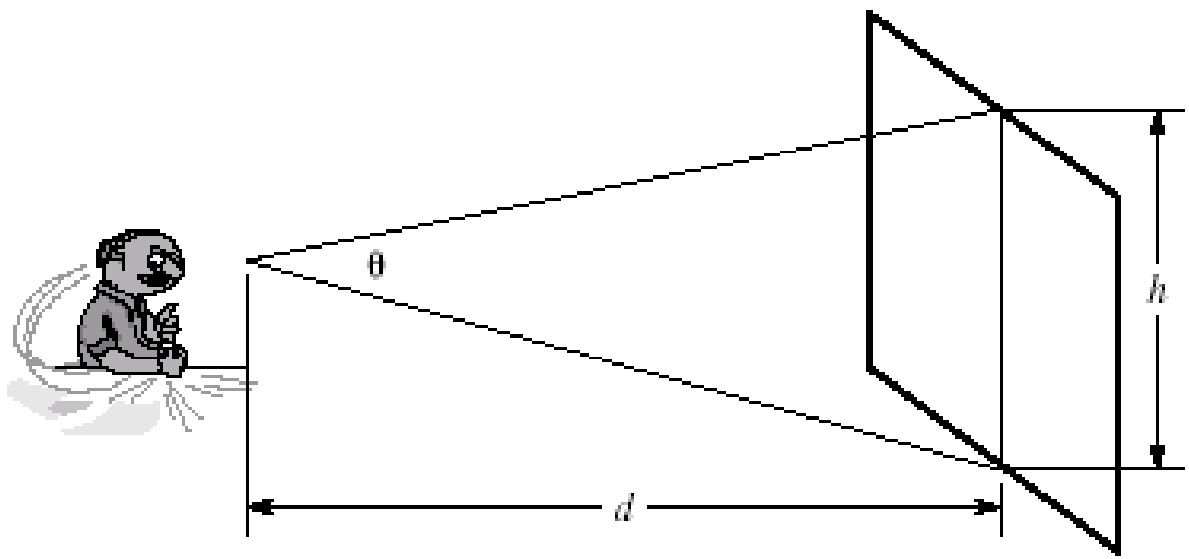


Figure 2.1 Two-dimensional sinusoidal signals: (a) $(f_x, f_y) = (5, 0)$; (b) $(f_x, f_y) = (5, 10)$. The horizontal and vertical units are the width and height of the image, respectively. Therefore, $f_x = 5$ means that there are five cycles along each row.

$$f_s = \sqrt{125}, \quad \varphi = \arctan(2)$$

Angular Frequency

Previously defined spatial frequency (cycles per pixel) depends on viewing distance



$$\theta = 2\arctan(h/2d)(\text{radian}) \approx 2\text{h}/2\text{d}(\text{radian}) = \frac{180}{\pi}\frac{h}{d}(\text{degree})$$

$$\text{f}_\theta = \frac{\text{f}_s}{\theta} = \frac{\pi}{180}\frac{d}{h}\text{f}_s(\text{cycle/degree})$$

Temporal Frequency

- Temporal frequency measures temporal variation (cycles/s)
- In a video, the *temporal frequency* is actually 2-dimensional; each point in space has its own temporal frequency
- Non-zero temporal frequency can be caused by camera or object motion

- Start simple: single object with constant velocity

Temporal Frequency caused by Linear Motion

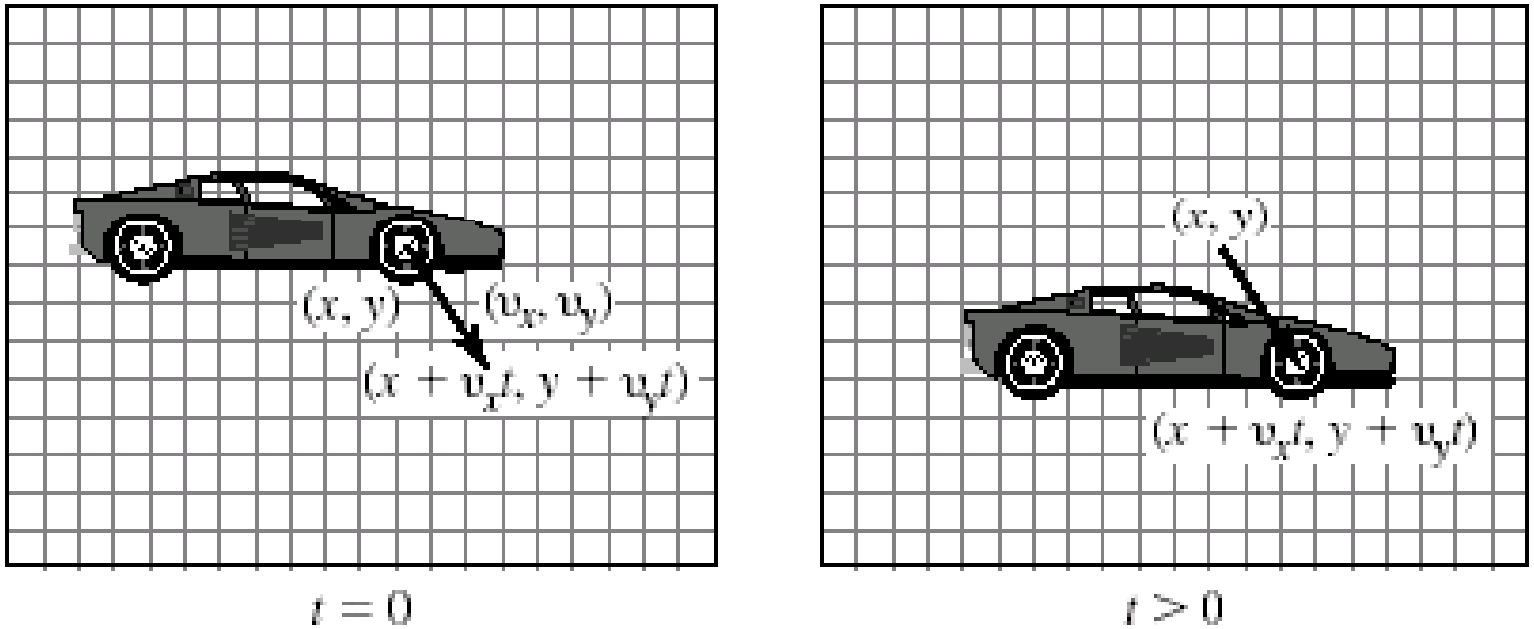


Figure 2.3 Illustration of the constant intensity assumption under motion. Every point (x, y) at $t = 0$ is shifted by $(v_x t, v_y t)$ to $(x + v_x t, y + v_y t)$ at time t, without change in color or intensity. Alternatively, a point (x, y) at time t corresponds to a point $(x - v_x t, y - v_y t)$ at time zero.

Relation between Motion, Spatial and Temporal Frequency

Consider an object moving with speed (v_x, v_y). Assume the image pattern at $t = 0$ is $\psi_0(x, y)$, the image pattern at time t is

$$\psi(x, y, t) = \psi_0(x - v_x t, y - v_y t)$$

$$\Leftrightarrow$$

$$\Psi(f_x, f_y, f_t) = \Psi_0(f_x, f_y)\delta(f_t + v_x f_x + v_y f_y)$$

Relation between motion, spatial, and temporal frequency:

$$f_t = -(v_x f_x + v_y f_y)$$

The temporal frequency of the image of a moving object depends on motion as well as the spatial frequency of the object.

Example: A plane with vertical bar pattern, moving vertically, causes no temporal change; But moving horizontally, it causes fastest temporal change

Illustration of the Relation

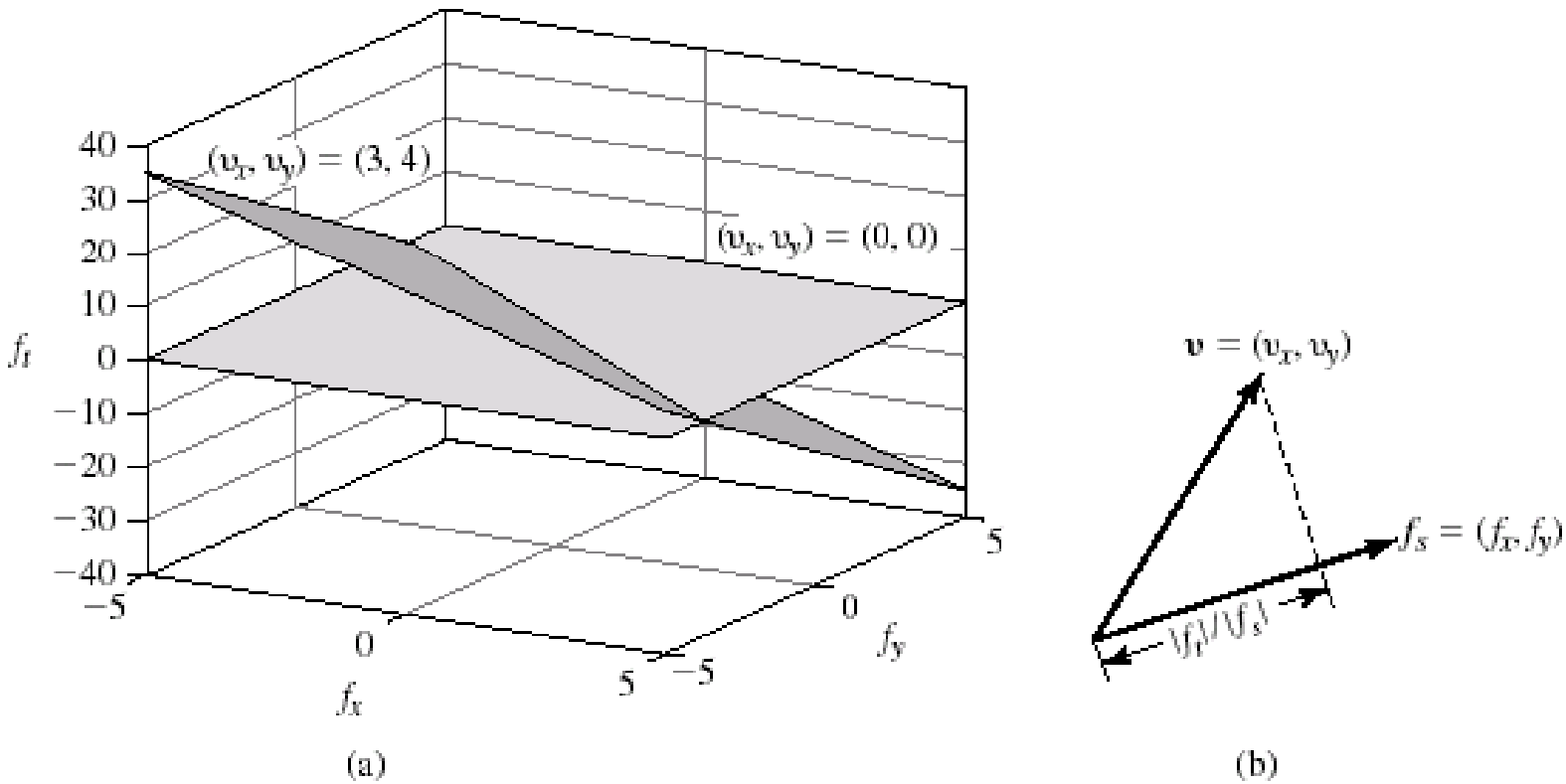


Figure 2.4 Relation between spatial and temporal frequencies under linear motions. (a) The spatiotemporal frequency plane in the (f_x, f_y, f_t) space, corresponding to two different velocity vectors; (b) the temporal frequencies is equal to the projection of the velocity onto the spatial gradient.

Frequency response of the HVS

- Temporal frequency response and flicker

$$\psi(t) = B(1 + m\cos 2\pi f t)$$

- Spatial frequency response

$$\psi(x, y, t) = B(1 + m\cos 2\pi f x)$$

- Spatio-temporal response

$$\psi(x, y, t) = B(1 + m\cos 2\pi f_x x)\cos(2\pi f_t t)$$

- Smooth pursuit eye movement

Contrast Sensitivity Function

$$\psi(t) = B(1 + m\cos 2\pi f x)$$

- B brightness, f frequency, m modulation level
- What is minimum modulation level at which sinusoidal grating is visible?
- $1/m_{min}$ at a given frequency is the *sensitivity*
- Contrast sensitivity function also known as the Modulation Transfer Function of the human eye

- Humans less sensitive to variations in chrominance

Temporal Response

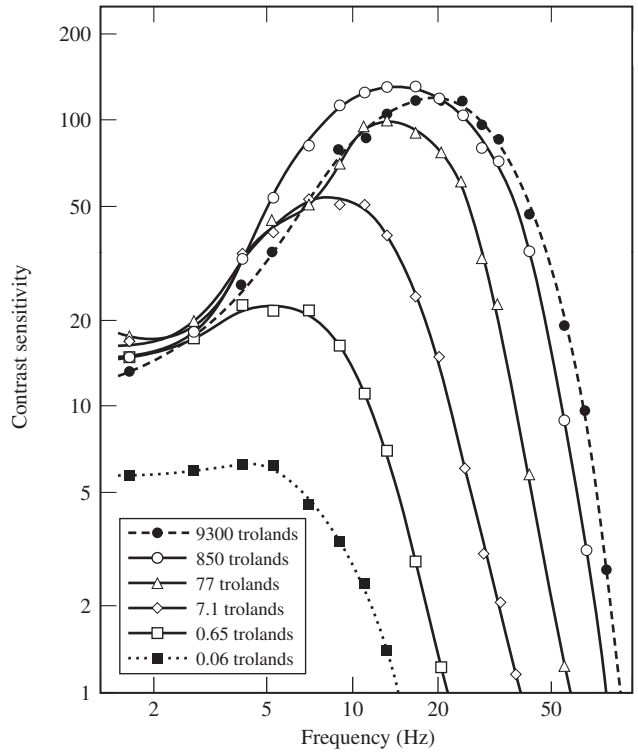


Figure 2.5 The temporal frequency response of the HVS obtained by a visual experiment. Different curves represent the responses obtained with different mean brightness levels, B, measured in trolands. The horizontal axis represents the flicker frequency f, measured in Hz. Reprinted from D. H. Kelly, Visual responses to time-dependent stimuli. I. Amplitude sensitivity measurements, *J. Opt. Soc. Am.* (1961) 51:422–29, by permission of the Optical Society of America.

Critical flicker frequency: The lowest frame rate at which the eye does not perceive flicker.

Provides guideline for determining the frame rate when designing a video system.

Critical flicker frequency depends on the mean brightness of the display:

60 Hz is typically sufficient for watching TV.

Watching a movie needs lower frame rate than TV

Spatial Response

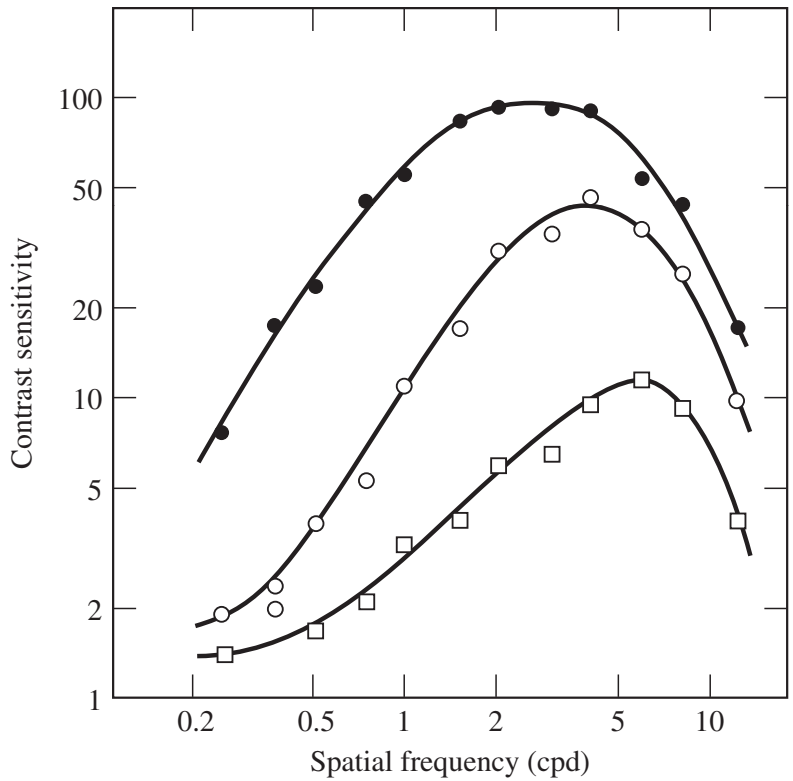
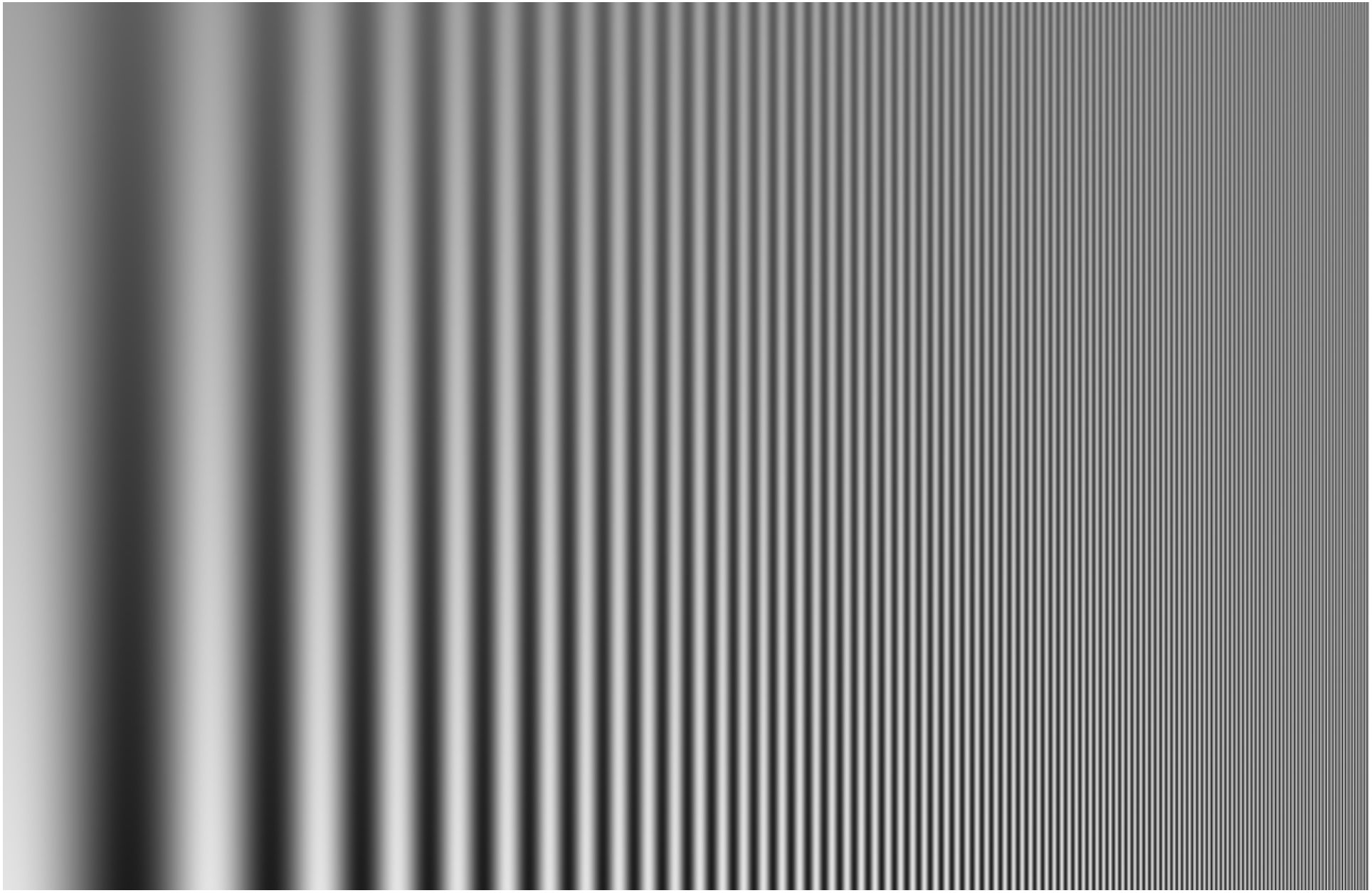


Figure 2.6 The spatial frequency response of the HVS, obtained by a visual experiment. The three curves result from different stabilization settings used to remove the effect of saccadic eye movements. Filled circles were obtained under normal, unstablized conditions; open squares, with optimal gain setting for stabilization; open circles, with the gain changed about 5 percent. Reprinted from D. H. Kelly, Motion and vision. I. Stabilized images of stationary gratings, *J. Opt. Soc. Am.* (1979), 69:1266–74, by permission of the Optical Society of America.



Provided by Amy Reibman

Spatiotemporal Response

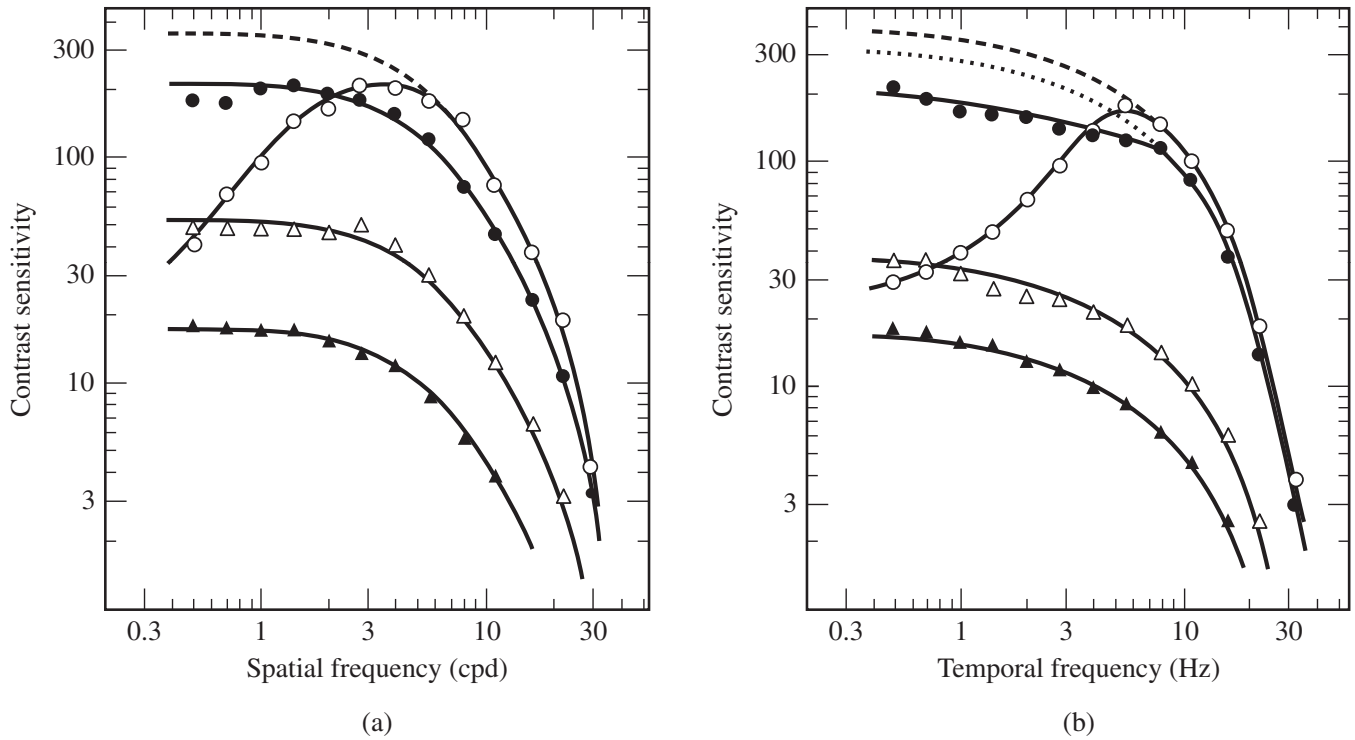


Figure 2.7 Spatiotemporal frequency response of the HVS. (a) Spatial frequency responses for different temporal frequencies of 1 Hz (open circles), 6 Hz (filled circles), 16 Hz (open triangles), and 22 Hz (filled triangles). (b) Temporal frequency responses for different spatial frequencies of 0.5 cpd (open circles), 4 cpd (filled circles), 16 cpd (open triangles), and 22 cpd (filled triangles). Reprinted from J. G. Robson, Spatial and temporal contrast sensitivity functions of the visual systems, *J. Opt. Soc. Am.* (1966), 56:1141–42, by permission of the Optical Society of America.

The reciprocal relation between spatial and temporal sensitivity was used in TV system design:

Interlaced scan provides tradeoff between spatial and temporal resolution.

Smooth Pursuit Eye Movement

- Smooth Pursuit: the eye tracks moving objects
- Net effect: reduce the velocity of moving objects on the retinal plane, so that the eye can perceive much higher raw temporal frequencies than indicated by the temporal frequency response.

Temporal frequency caused by object motion when the object is moving at (v_x, v_y):

$$f_t = -(v_x f_x + v_y f_y)$$

Observed temporal frequency at the retina when the eye is moving at $(\tilde{v}_x, \tilde{v}_y)$:

$$\tilde{f}_t = f_t + (\tilde{v}_x f_x + \tilde{v}_y f_y)$$

$$\tilde{f}_t = 0 \text{ if } \tilde{v}_x = v_x, \tilde{v}_y = v_y$$

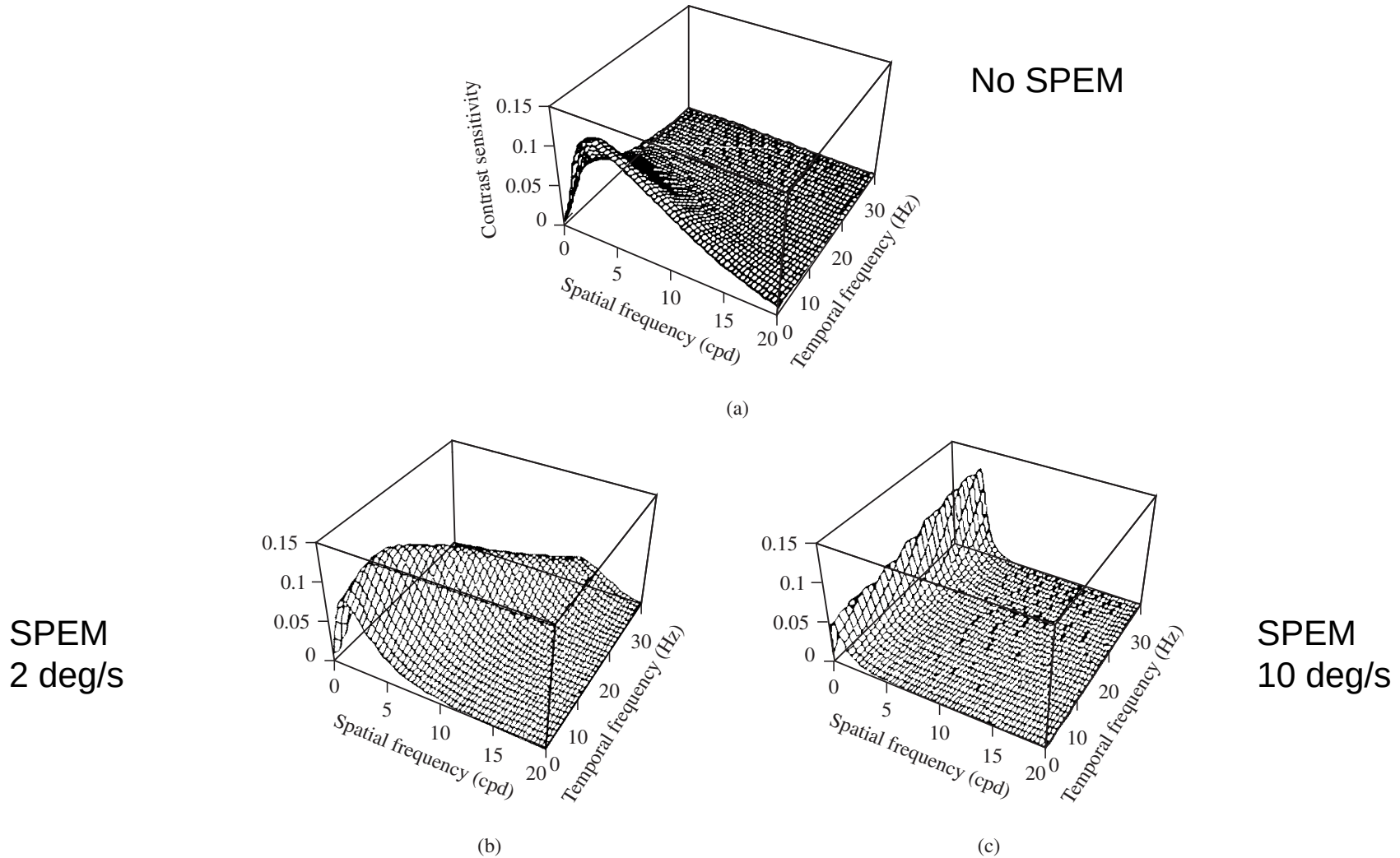


Figure 2.8 Spatiotemporal response of the HVS under smooth pursuit eye movements: (a) without smooth pursuit eye movement; (b) with eye velocity of 2 deg/s; (c) with eye velocity of 10 deg/s. Reprinted from Girod, B. "Motion compensation: visual aspects, accuracy, and fundamental limits." In Sezan, M. I., and R. L. Lagendijk, eds., *Motion Analysis and Image Sequence Processing,* Boston: Kluwer Academic Publishers, 1993, 126–52, by permission of Kluwer Academic Publishers.

Video Sampling – A Brief Discussion

- Review of Nyquist sampling theorem in 1-D
- Extension to multi-dimensions
- Prefilter in video cameras
- Interpolation filter in video displays

Nyquist Sampling Theorem in 1-D

- Given a band-limited signal with maximum frequency f_{max}, it can be sampled with a sampling rate f_s>=2 f_{max}. The original continuous signal can be reconstructed (interpolated) from the samples exactly, by using an ideal low pass filter with cut-off frequency at f_s /2.
- Practical interpolation filters: replication (sample-and-hold, 0th order), linear interpolation (1st order), cubic-spline (2nd order)
- Given the maximally feasible sampling rate f_s, the original signal should be bandlimited to f_s /2, to avoid aliasing. The desired prefilter is an ideal low-pass filter with cut-off frequency at f_s /2.
- Prefilter design: Trade-off between aliasing and loss of high frequency

Extension to Multi-dimensions

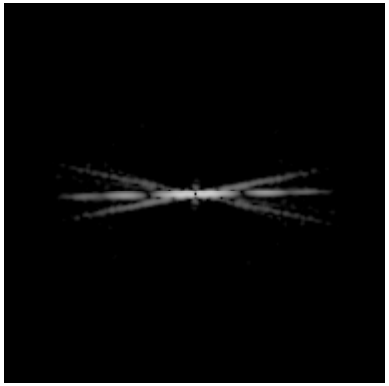
- If the sampling grid is aligned in each dimension (rectangular in 2-D) and one performs sampling in each dimension separately, the extension is straightforward:
  - Requirement: $f_{s,i} >= 2 f_{max,i}$
  - Interpolation/pre-filter: ideal low-pass in each dimension
- If the sampling grid is an arbitrary lattice, the support region of the signal spectrum must be limited within the Voronoi region of the reciprocal of the sampling lattice
  - See Chapters 3 and 4 for sampling and sampling-rate conversion for K-D signals and for video in particular.
  - Interlaced scan uses a non-rectangular lattice in the vertical-temporal plane.

Aliasing in 2D

Example provided by Amy Reibman

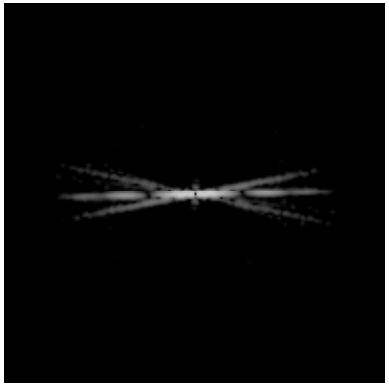
Other Examples

Fourier examples of aliasing: jaggies



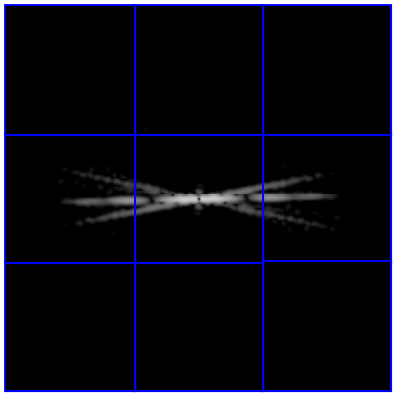
Example provided by Amy Reibman

Fourier examples of aliasing: jaggies



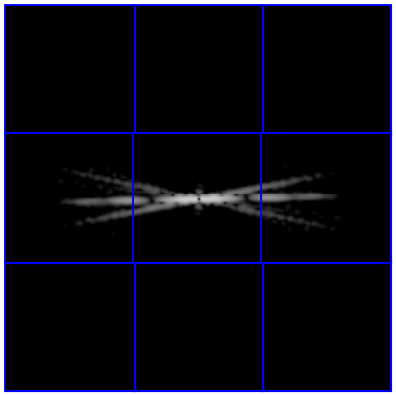
Example provided by Amy Reibman

Fourier examples of aliasing: jaggies



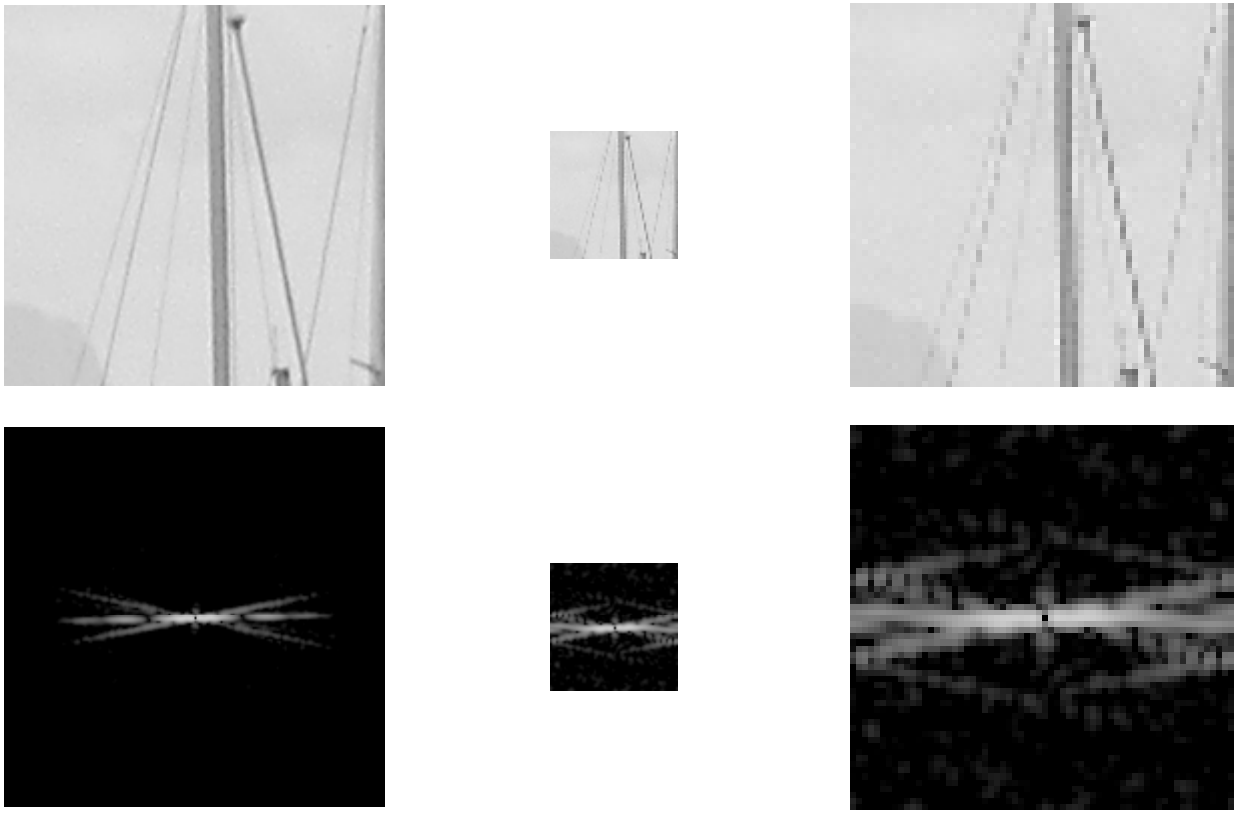
Example provided by Amy Reibman

Fourier examples of aliasing: jaggies



Example provided by Amy Reibman

Fourier examples of aliasing: jaggies



Example provided by Amy Reibman

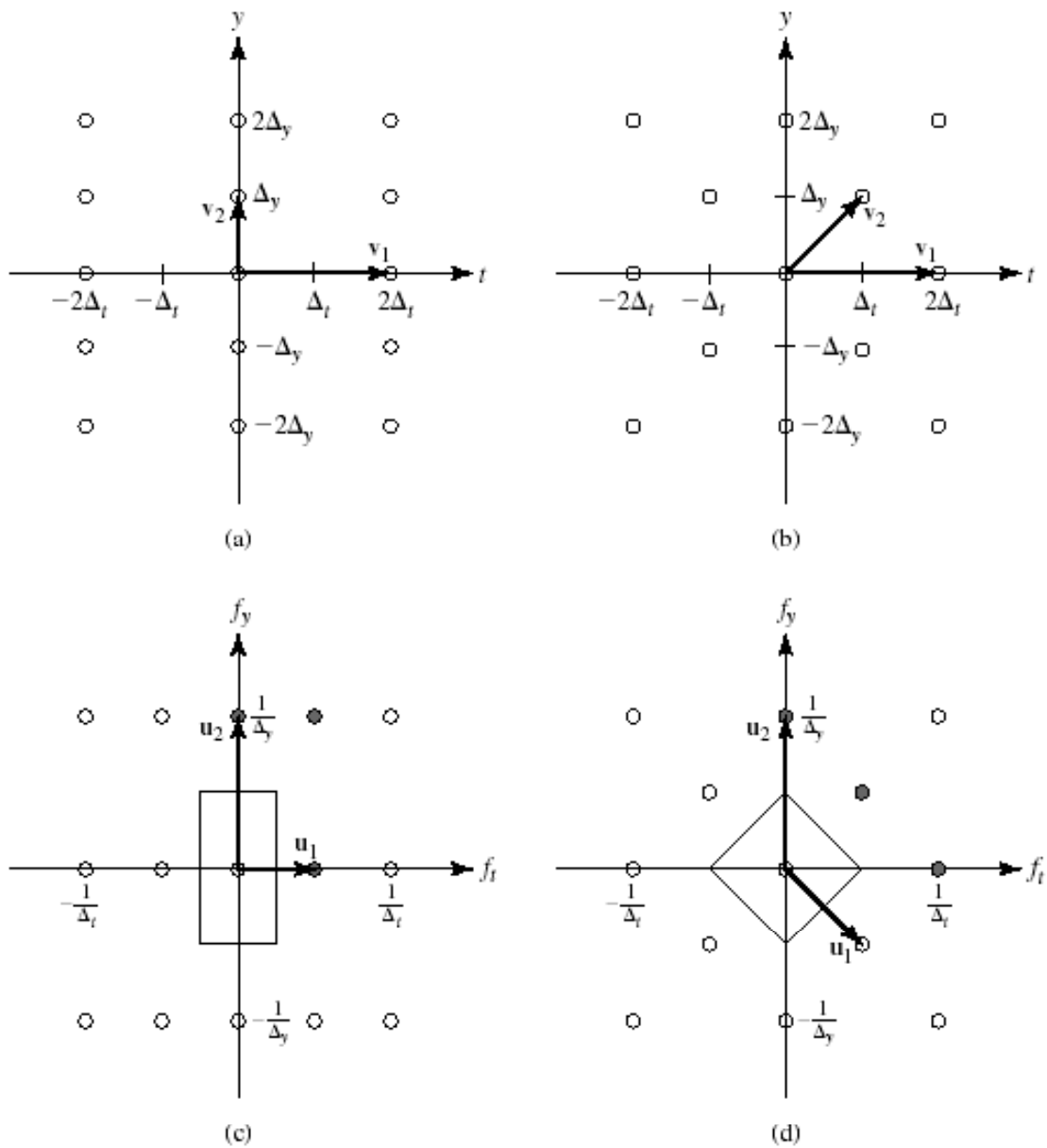


Figure 3.6 Comparison of progressive and interlaced scans: (a) sampling lattice for progressive scan; (b) sampling lattice for interlaced scan; (c) reciprocal lattice for progressive scan; (d) reciprocal lattice for interlaced scan. Filled circles in (c) and (d) indicate the nearest aliasing components.

Video Cameras

- Sampling mechanism
  - All perform sampling in time
  - Film cameras capture continuous frames on film
  - Analog video cameras sample in vertical but not horizonal direction, arrange the resulting horizontal scan lines in a 1-D continuous signal
  - Digital cameras sample in both horizontal and vertical direction, yielding pixels with discrete 3-D coordinates
- Sampling frequency (frame rate and line rate)
  - Depending on the maximum frequency in the underlying signal, the human visual thresholds, as well as technical feasibility and cost
- Prefilter
  - Controlled by temporal exposure, scanning beam, etc.
  - Digital cameras may capture at higher sampling rates and then implement explicit filtering before converting to lower resolution

Typical Camera Response

- Temporal prefilter: the value read out at any frame is the average of the sensed signal over the exposure time

- Spatial prefilter: the value read out at any pixel is a weighted integration of the signal in a small window surrounding it, called the aperture, can be approximated by a box average or a 2-D Gaussian function

Video Display

- The display device presents the analog or digital video on the screen to create the sensation of continuously varying signal in both time and space.

- With CRT, three electronic beams strike red, green, and blue phosphors with the desired intensity at each pixel location. No explicit interpolation filters are used. Spatial filtering determined by the size of the scanning beam, temporal filtering determined by the decaying time of the phosphors.

- The eye performs the interpolation task: fuses discrete frames and pixels as continuously varying, if the temporal and spatial sampling rates are sufficiently high.

Homework 2

- Reading assignment:
  - Chapter 2
  - Section 3.3, Section 3.4
- Written assignment
  - Prob. 2.1,2.2,2.5,2.6,2.7

- Spatio-temporal frequency domain is pretty empty
- Individual objects are either still (ie, defined only for zero temporal frequency) or moving at (nearly) constant velocity (ie, defined only on a plane in the 3d frequency-domain)