

Sampling and pixels

CS 178, Spring 2014

Begin 4/22, finished 4/24.



Marc Levoy
Computer Science Department
Stanford University

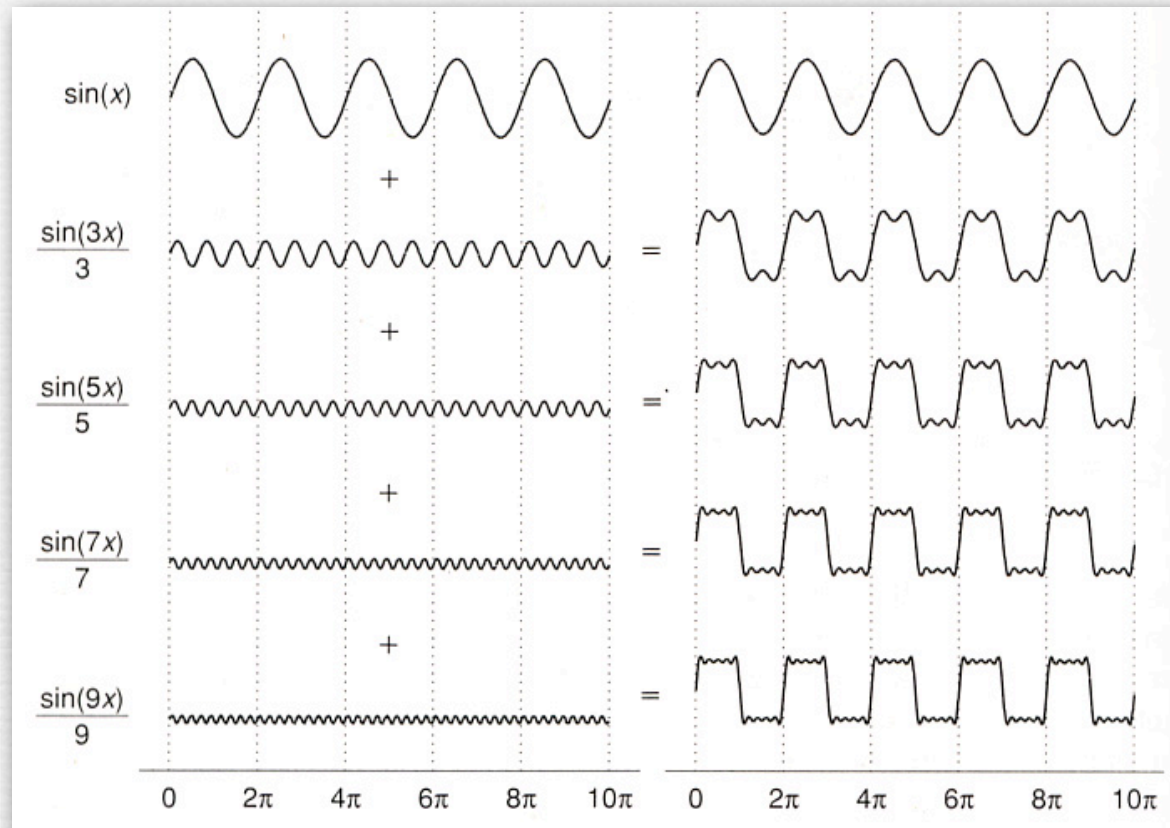
Why study sampling theory?

- ◆ Why do I sometimes get moiré artifacts in my images?
- ◆ What is an antialiasing filter?
- ◆ How many megapixels is enough?
- ◆ How do I compute circle of confusion for depth of field?
- ◆ Is Apple's "Retina Display" just hype?
- ◆ What do MTF curves in lens reviews mean?
- ◆ What does Photoshop do when you downsize/upsample?
- ◆ What's the difference between more pixels and more bits?

Outline

- ◆ frequency representations of images
 - filtering, blurring, sharpening
 - MTF as a measure of sharpness in images
 - ◆ resolution and human perception
 - the spatial resolution of typical display media
 - the acuity of the human visual system
 - the right way to compute circle of confusion (C)
 - ◆ sampling and aliasing
 - aliasing in space and time
 - prefiltering using convolution to avoid aliasing
 - prefiltering and sampling in cameras and Photoshop
-
- ◆ sampling versus quantization

Frequency representations



(Foley)

- ◆ a sum of sine waves, each of different wavelength (*frequency*) and height (*amplitude*), can approximate arbitrary functions
- ◆ to adjust horizontal position (*phase*), replace with cosine waves, or use a mixture of sine and cosine waves

Frequency representations

- ◆ Fourier series: any continuous, integrable, periodic function can be represented as an infinite series of sines and cosines

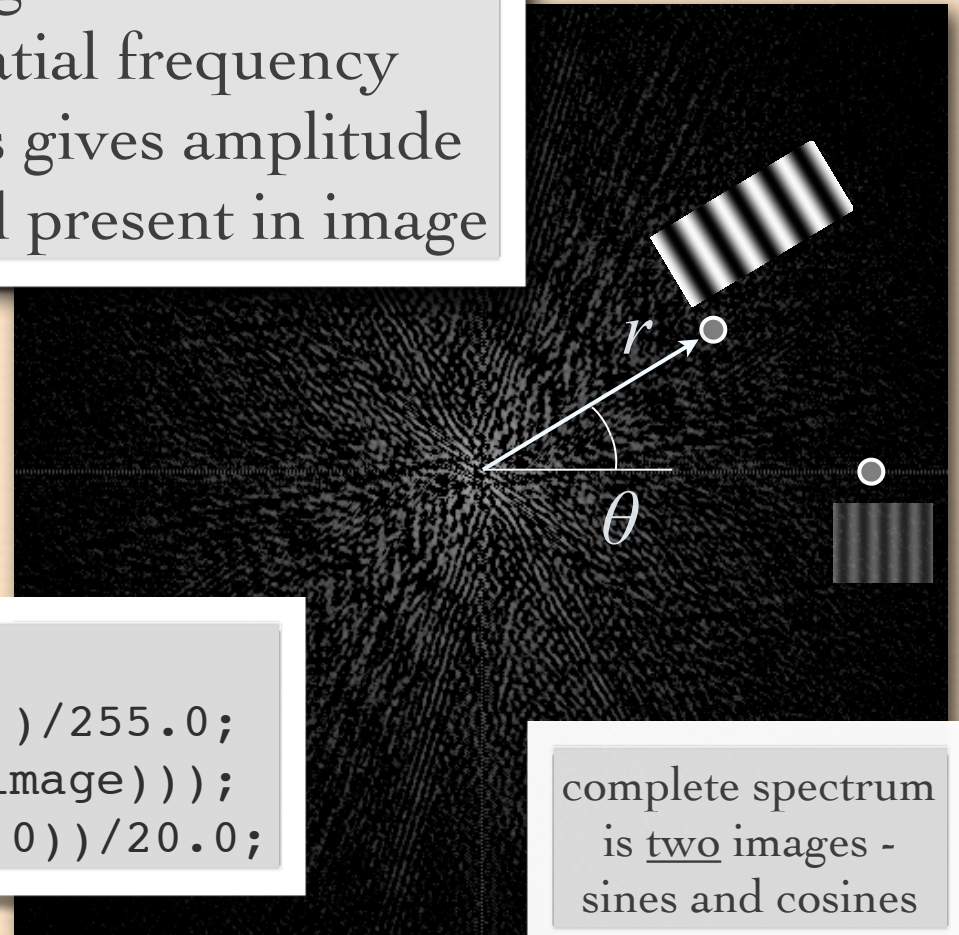
$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(nx) + b_n \sin(nx)]$$

Not responsible on exams
for orange-tinted slides

- ◆ a sum of sine waves, each of different wavelength (*frequency*) and height (*amplitude*), can approximate arbitrary functions
- ◆ to adjust horizontal position (*phase*), replace with cosine waves, or use a mixture of sine and cosine waves

Fourier transforms of images

- θ gives angle of sinusoid
- r gives spatial frequency
- brightness gives amplitude of sinusoid present in image



```
% In Matlab:  
image = double(imread('flower.tif'))/255.0;  
fourier = fftshift(fft2(iffshift(image)));  
fftimage = log(max(real(fourier),0.0))/20.0;
```

complete spectrum
is two images -
sines and cosines

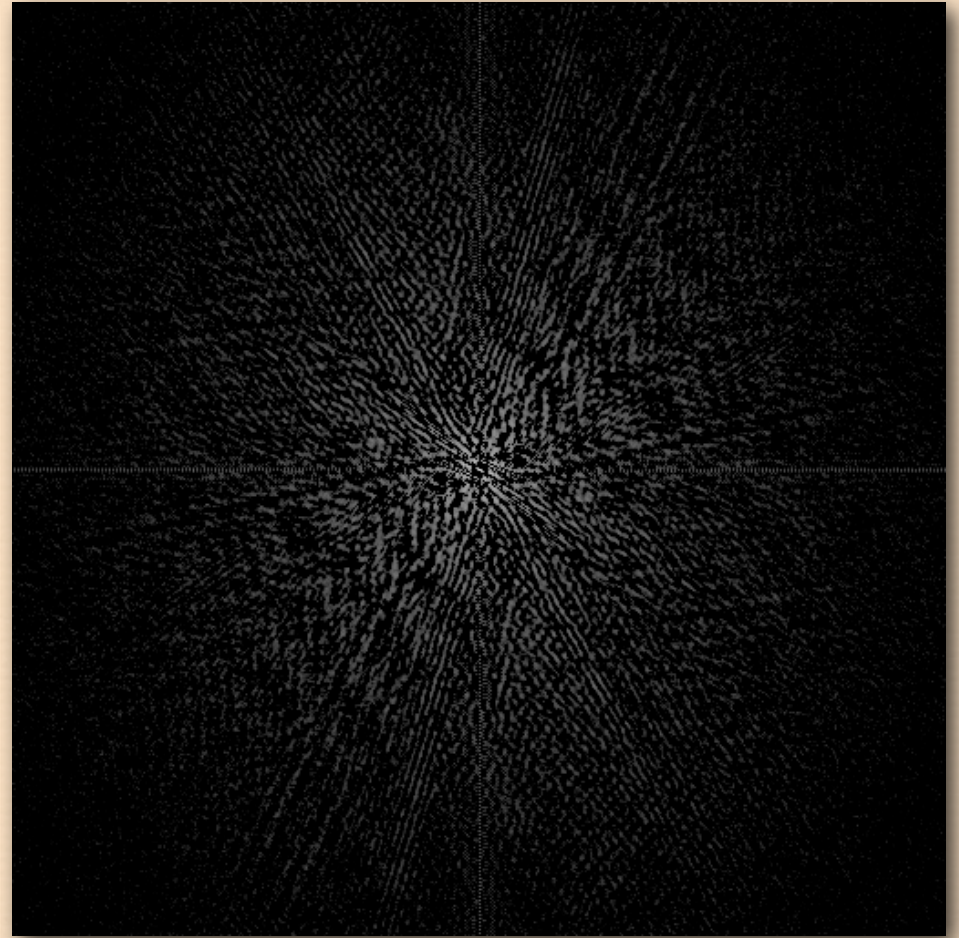
image

spectrum

A typical photograph

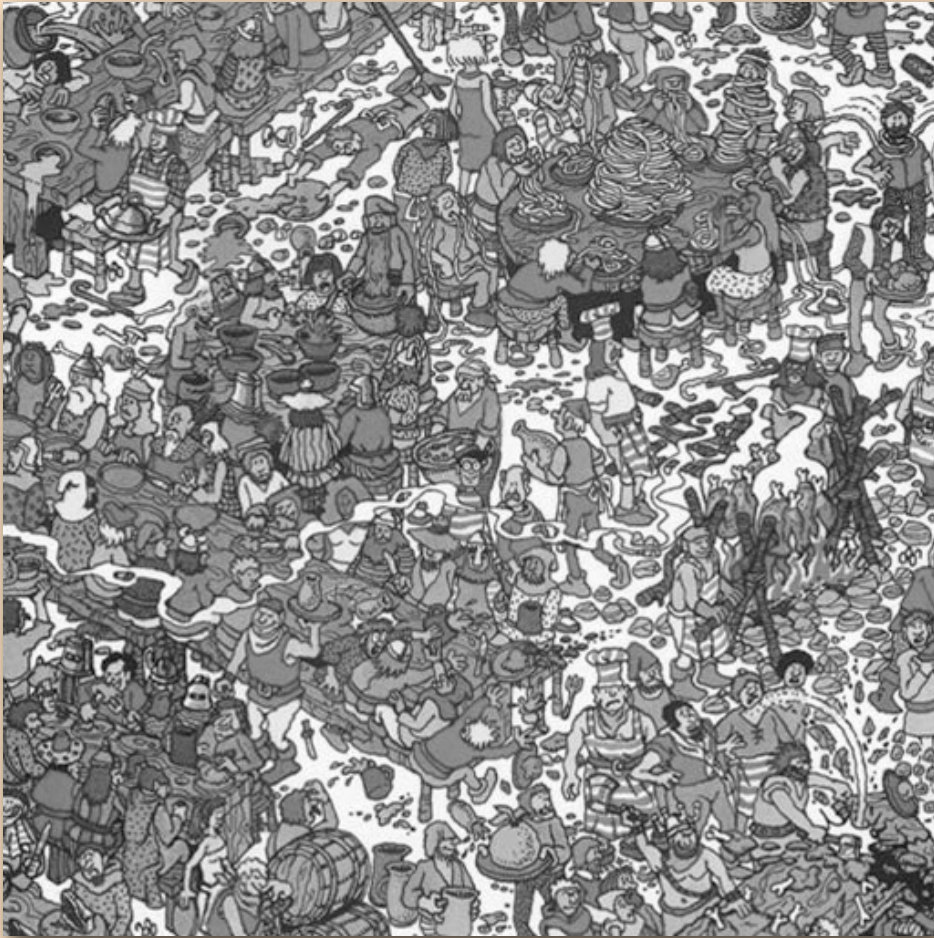


image

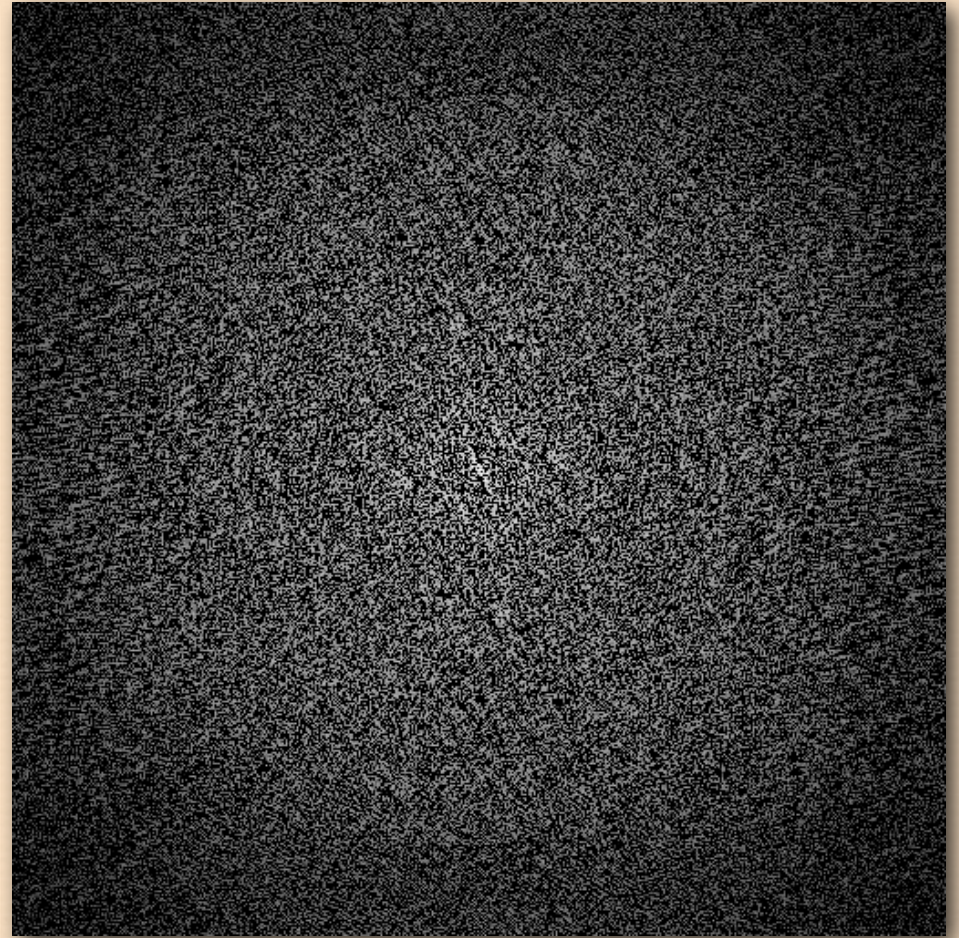


spectrum

An image with higher frequencies



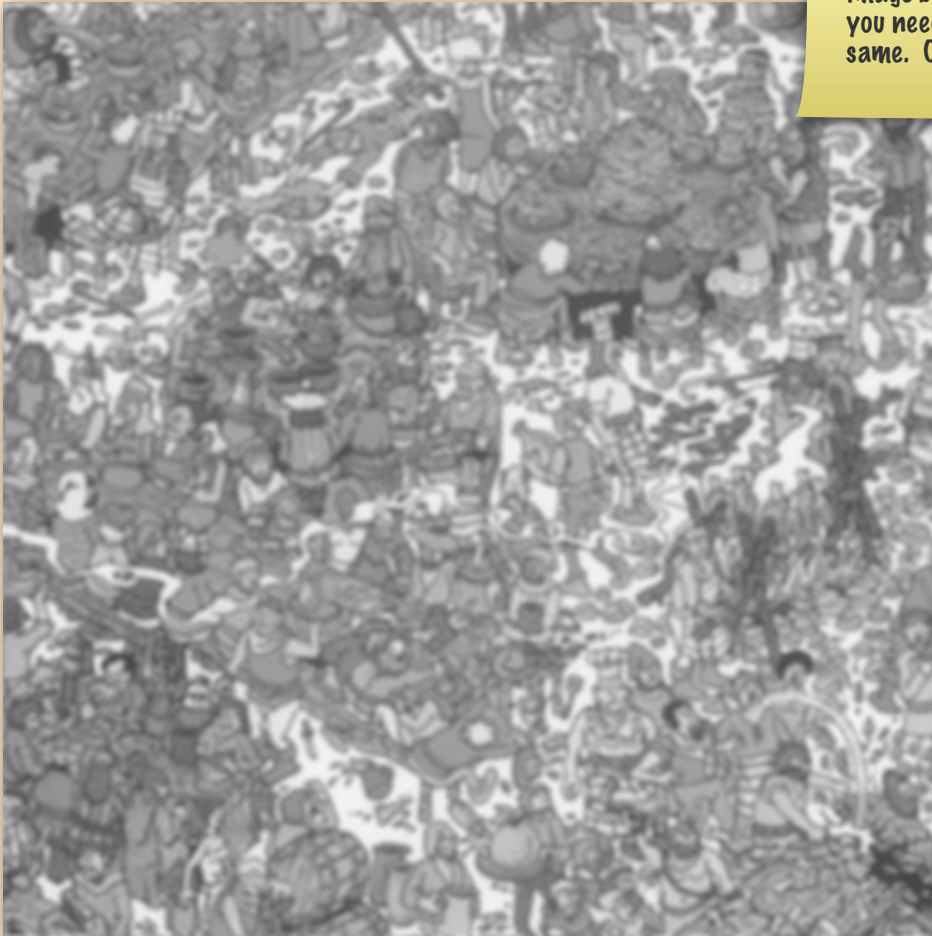
image



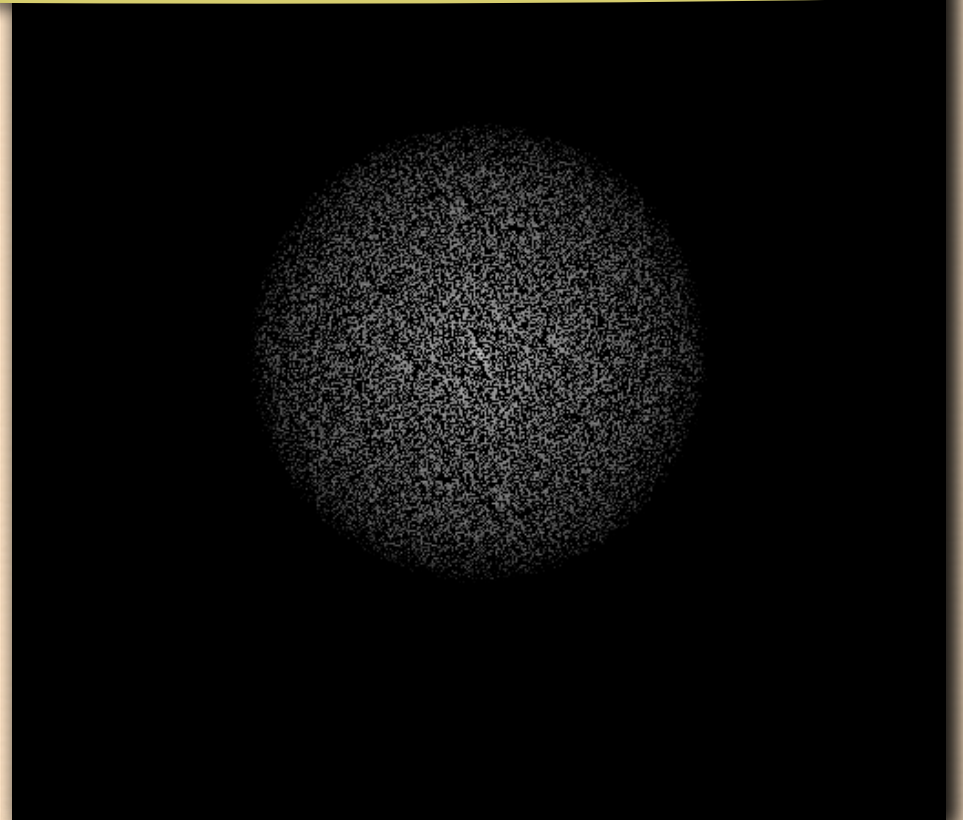
spectrum

Blurring in the Fourier domain

I didn't want to introduce too many technicalities into the lecture, but if you blur an image by erasing or attenuating selected frequencies in the spectrum as I've done here, you need to boost the remaining frequencies so that the sum of all frequencies stays the same. Otherwise, the image (on the left) will get dimmer, which you don't want.



image

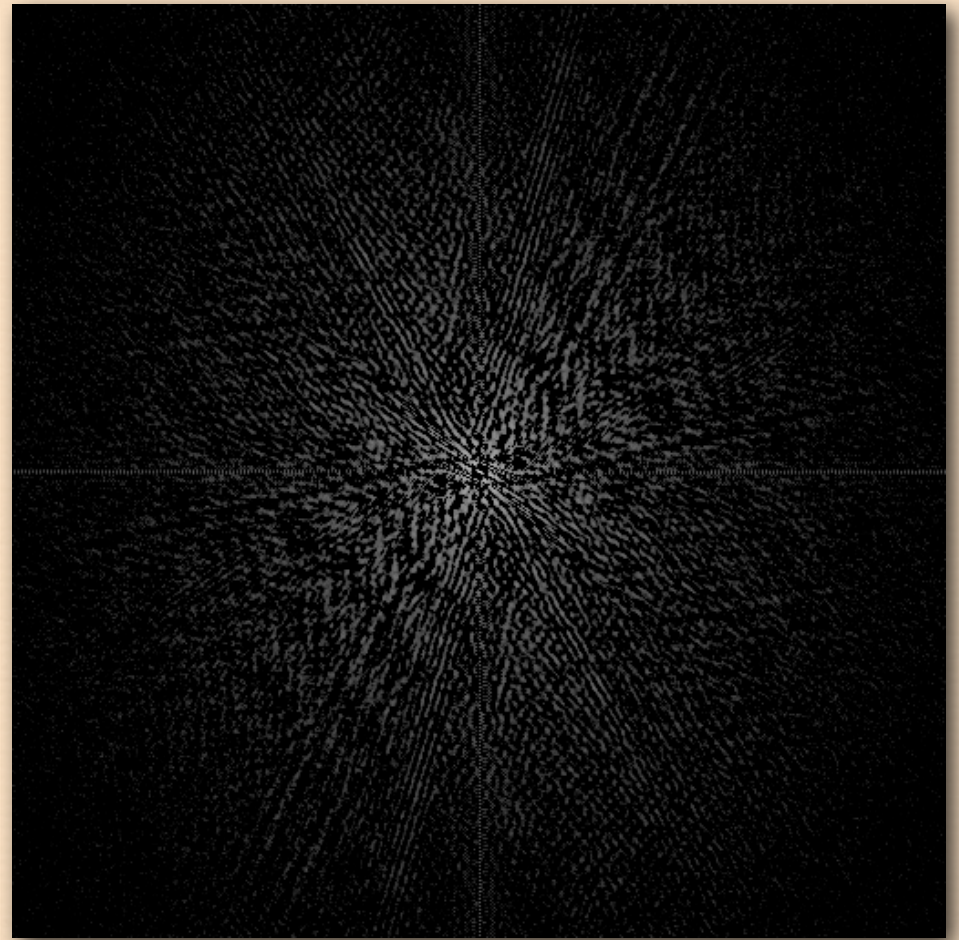


spectrum

Original flower



image

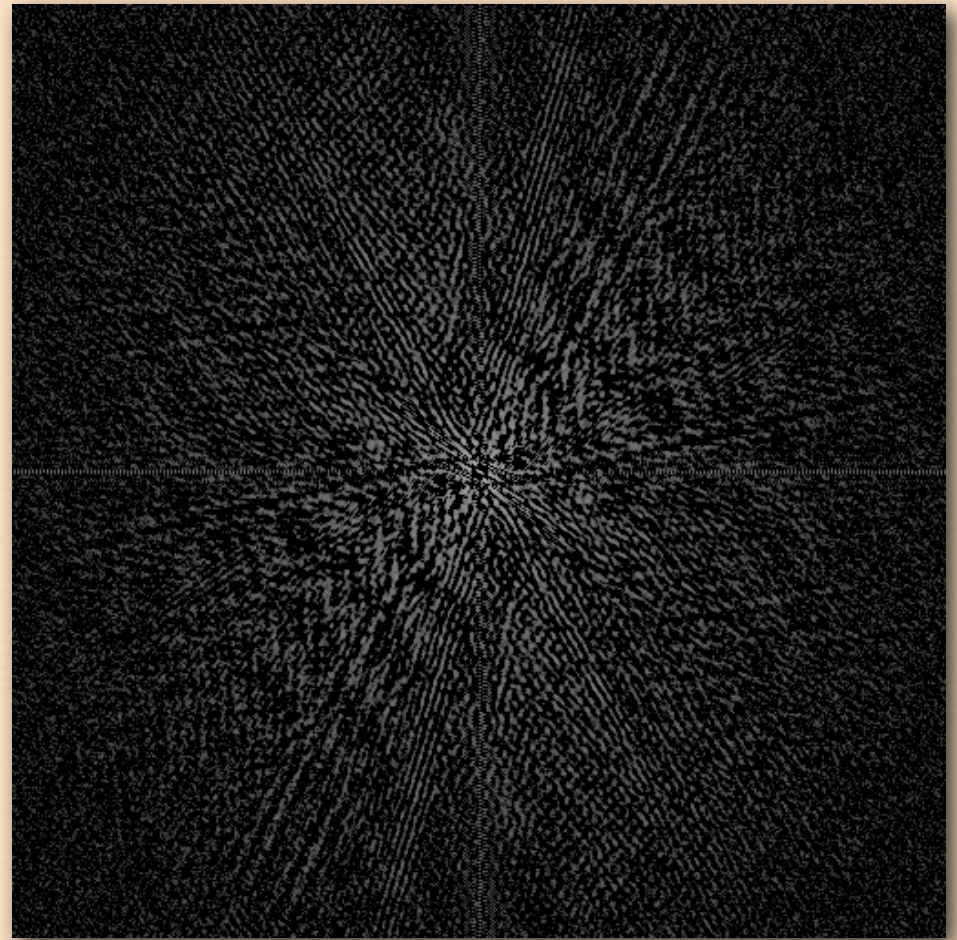


spectrum

Sharpening in the Fourier domain



image

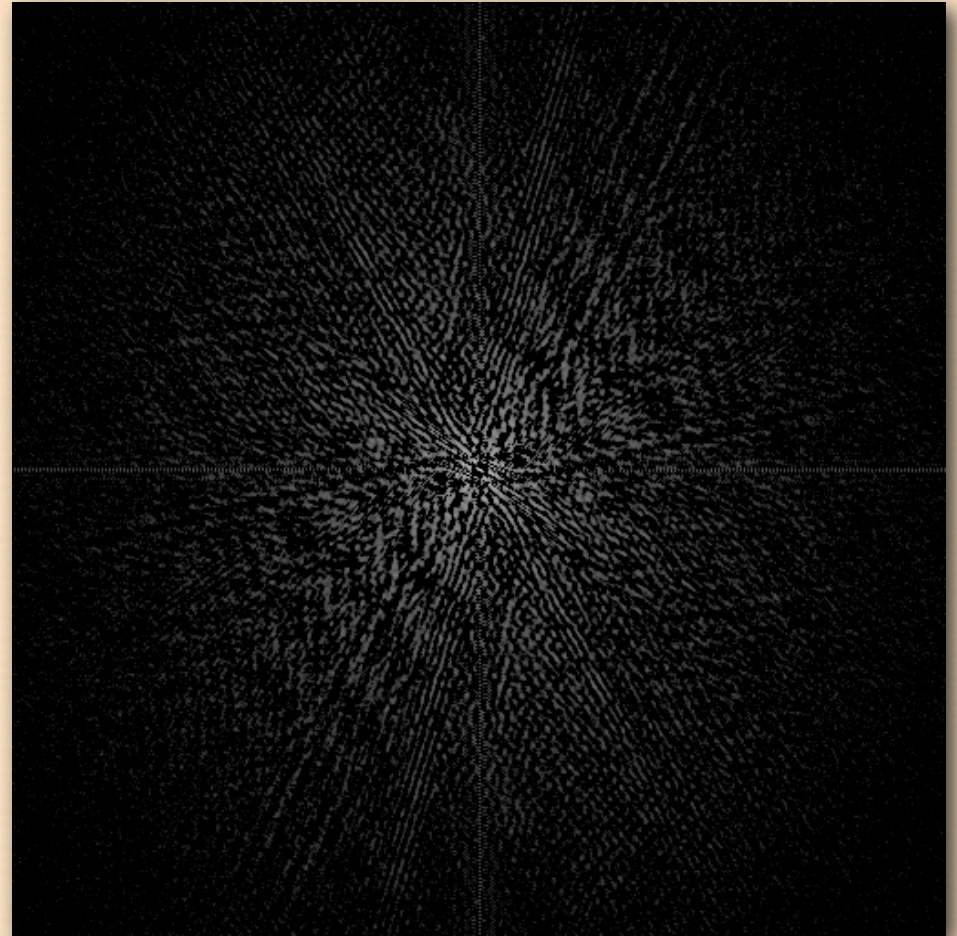


spectrum

Q. What does this filtering operation do?



image

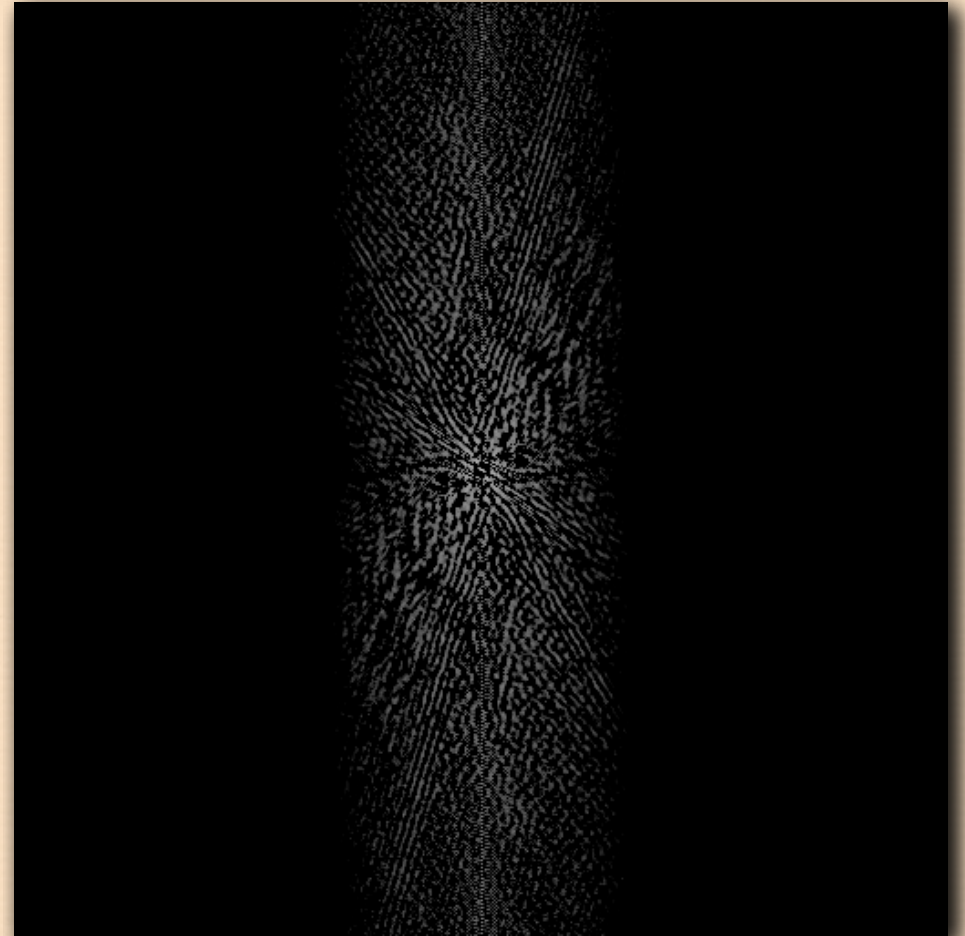


spectrum

Q. What does this filtering operation do?



image

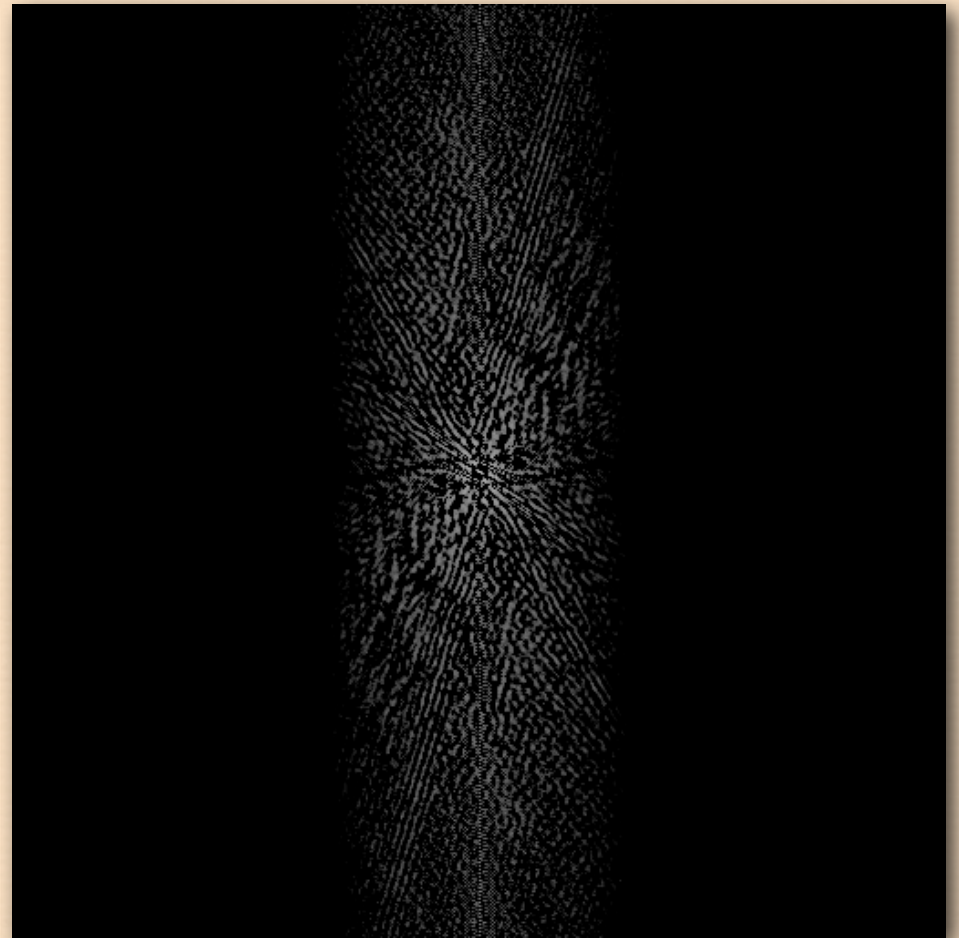


spectrum

Blurring in x , sharpening in y



image

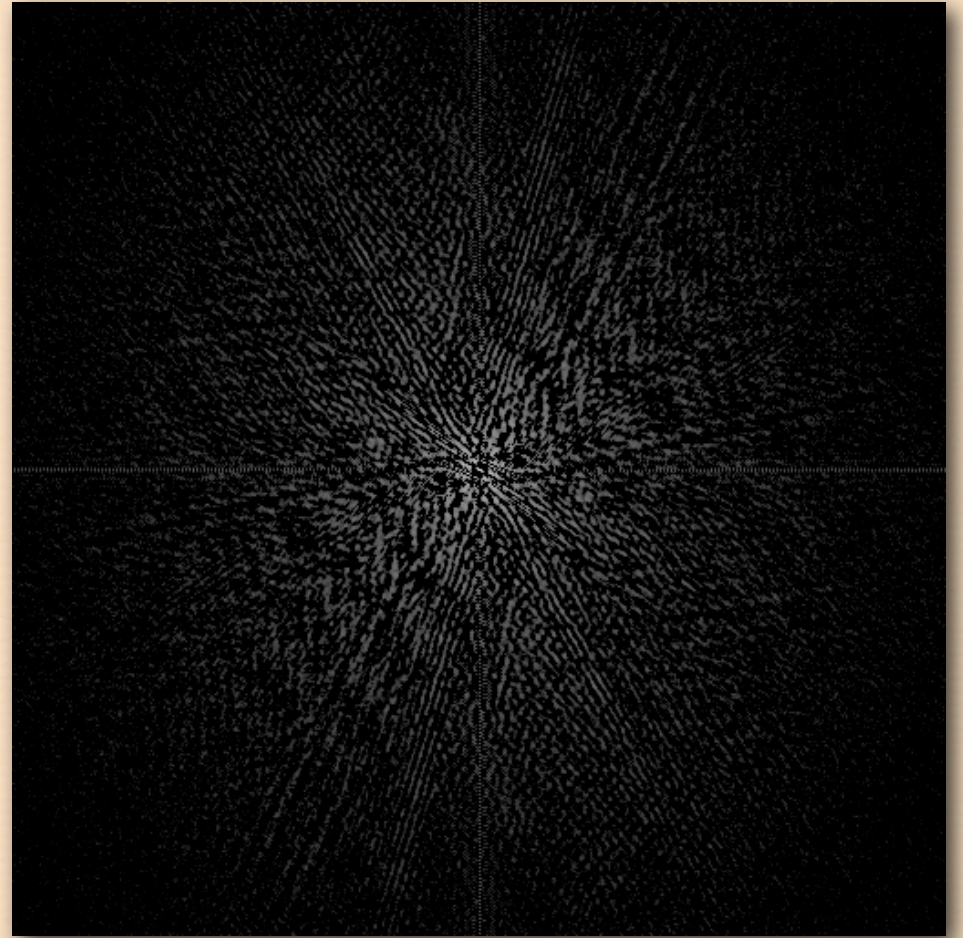


spectrum

Original



image

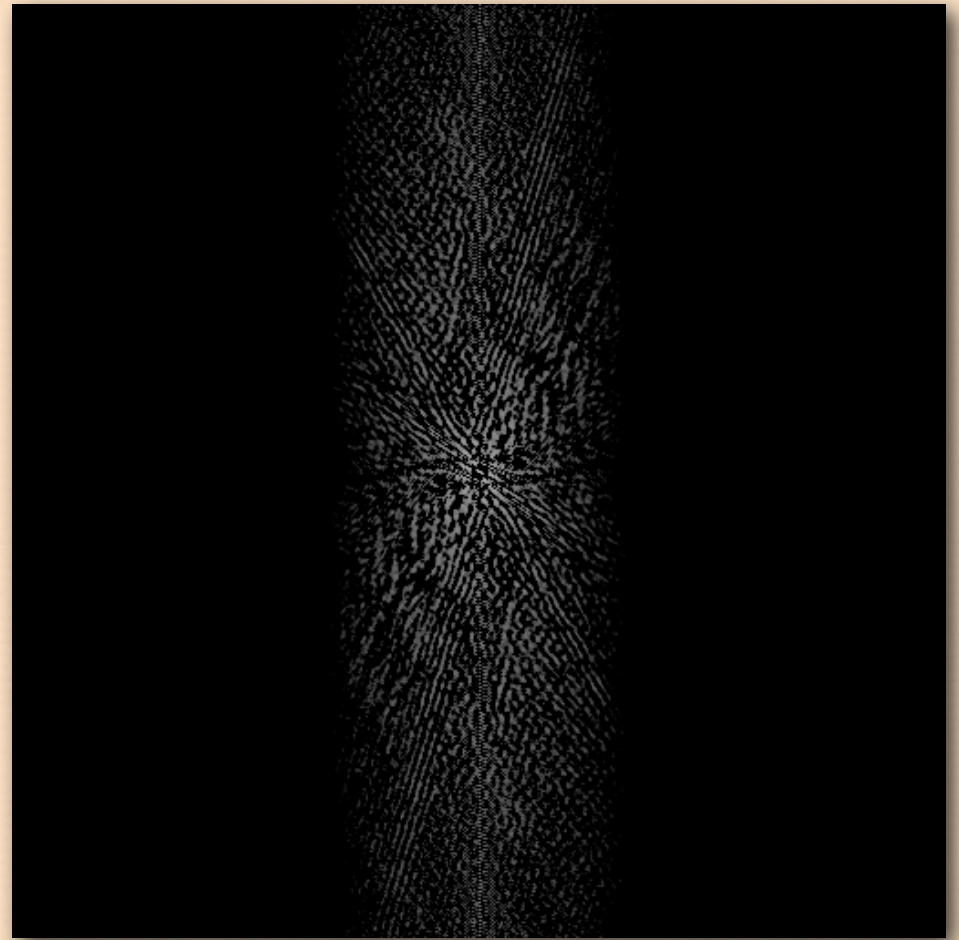


spectrum

Blurring in x , sharpening in y



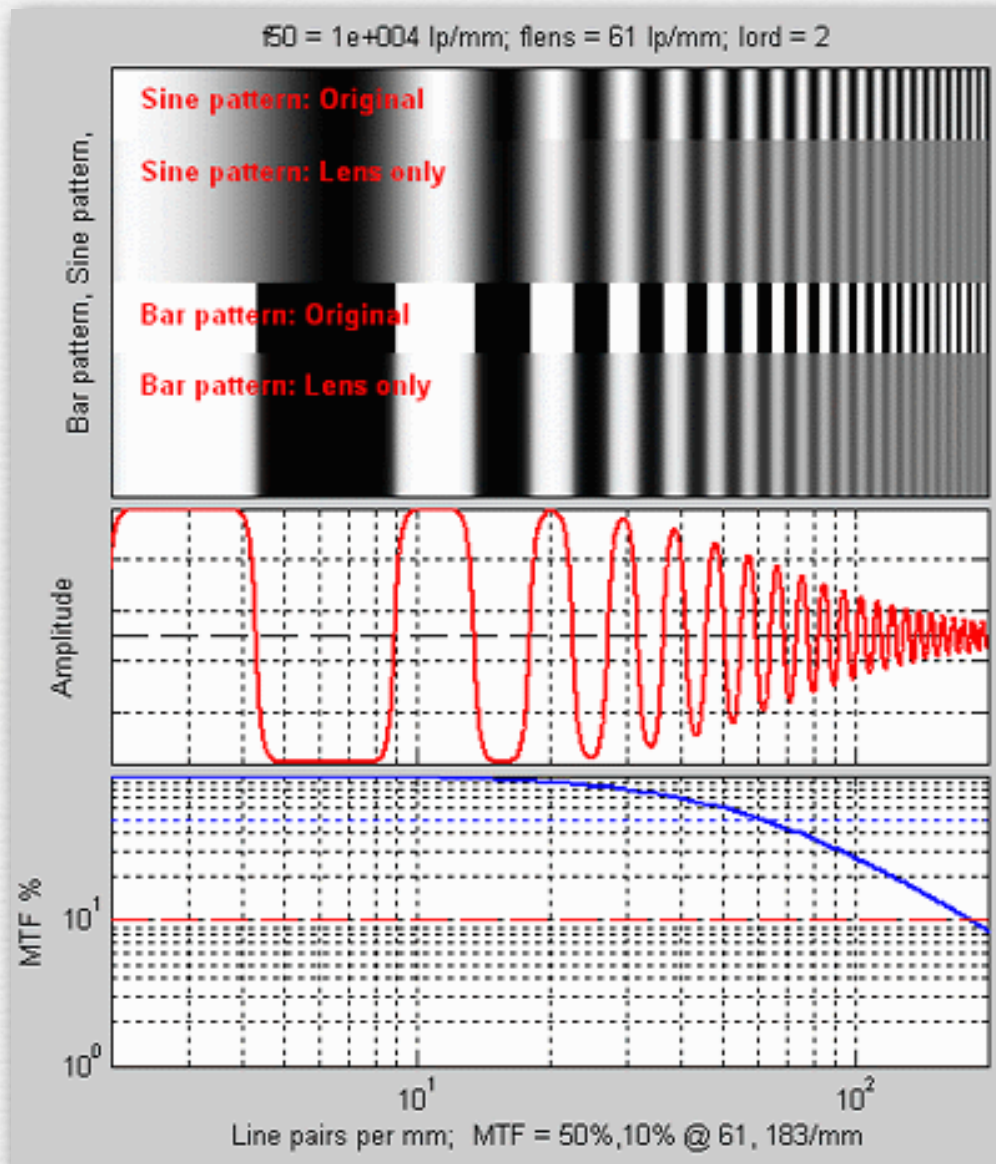
image



spectrum

argh, astigmatism!

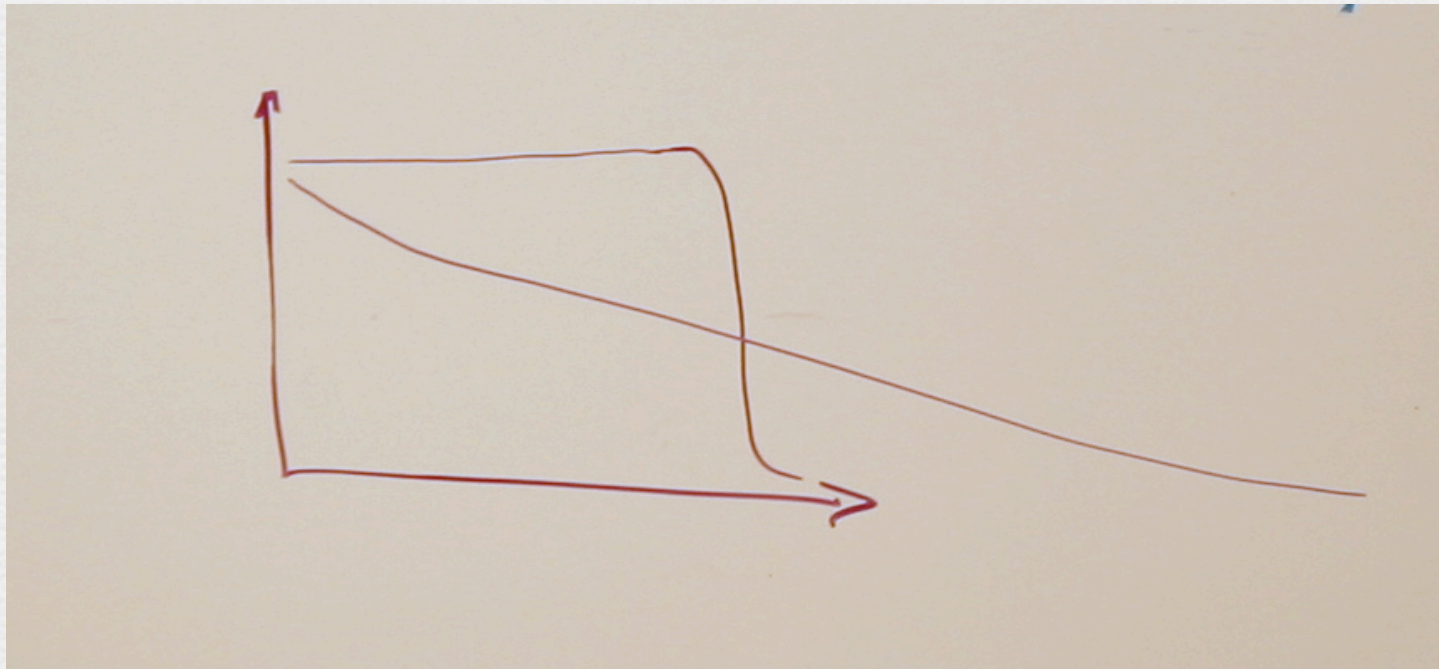
Describing sharpness in images: the modulation transfer function (MTF)



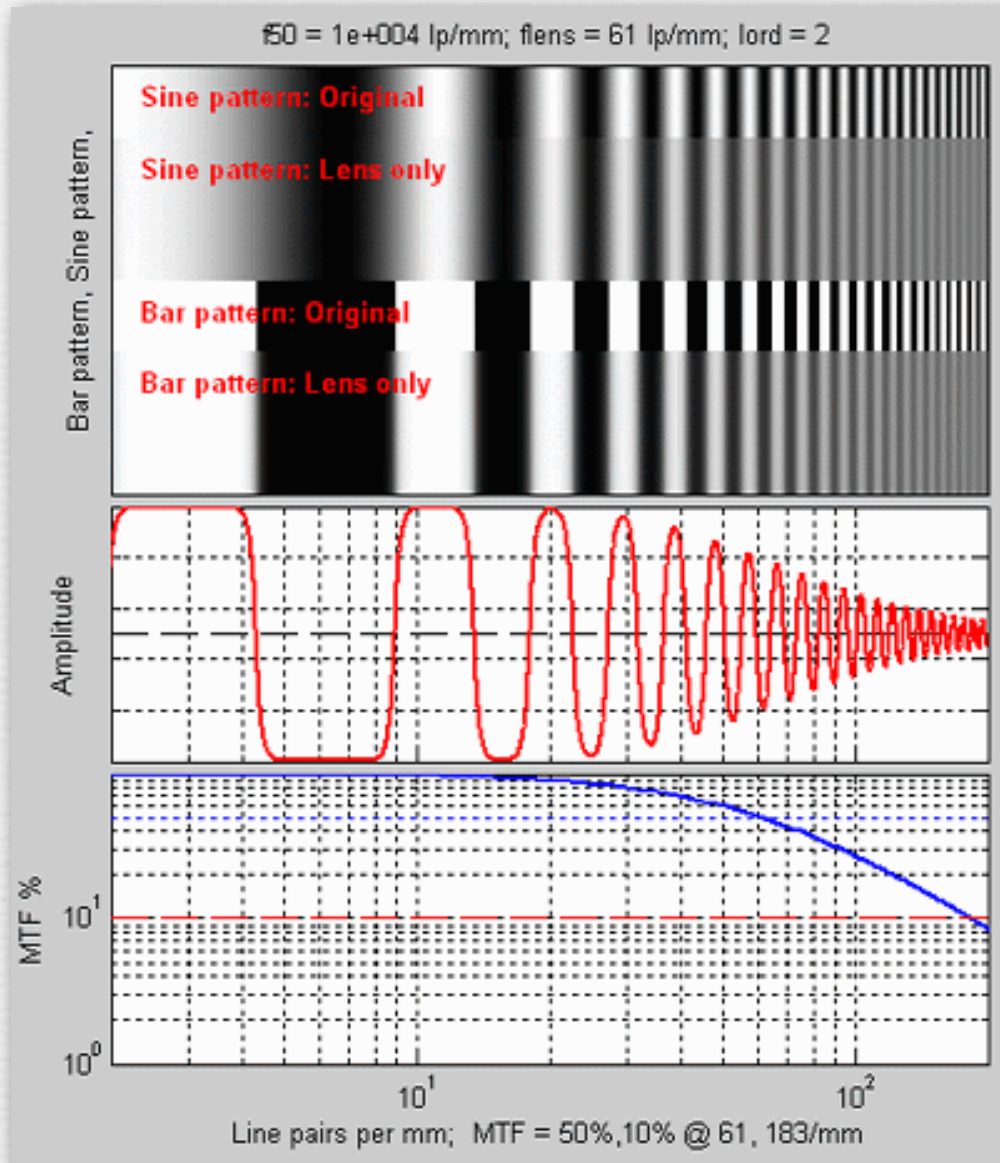
- ◆ the amount of each spatial frequency that can be reproduced by an optical system
- ◆ loss may be due to misfocus, aberrations, diffraction, manufacturing defects, nose smudges, etc.
- ◆ MTF is contrast at each frequency relative to original signal

Two different MTF curves

- ◆ in one curve, contrast stays high, but drops off at a relatively low resolution
- ◆ in the other curve, higher-resolution features are preserved, but contrast is lower throughout



Sharpness versus contrast



Restoring images

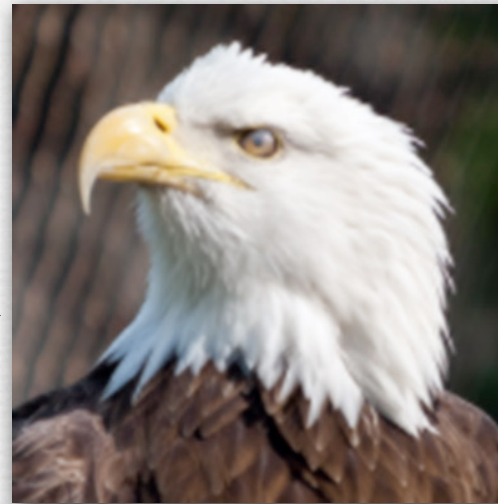
see the ringing?

(cambridgeincolour.com)

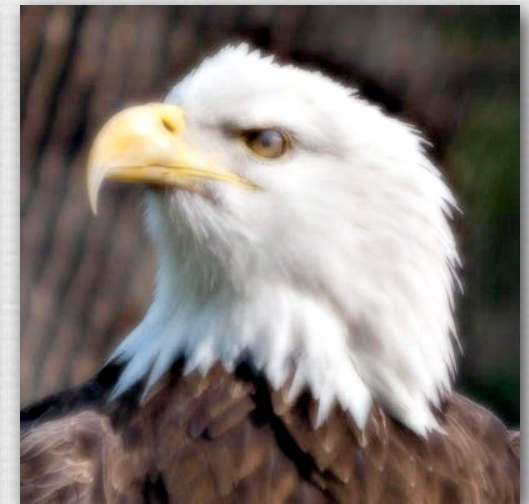


original

poor
resolution



poor
contrast



degraded

restored

In practice, since blurring and loss of contrast both involve attenuation of high frequencies (although to different extents), and restoration involves boosting those attenuated frequencies, then both sharpening and restoration of contrast should produce both some amount of ringing and some amount of noise enhancement. In these slides the ringing is evident, but not the noise enhancement. We would have to blow the images up to see the noise.

Recap

As I mentioned in class, “high-pass” might be better called “high-boosted”, if the goal is sharpening. The term “high-pass” is historical, and in the context of this course should be taken to mean that the filter strengthens high frequencies relative to low frequencies.

- ◆ any image can be equivalently represented by its Fourier transform, a.k.a. frequency or spectral representation
 - weighted sum of sine and cosine component images
 - each having a frequency, amplitude, and orientation in the plane
- ◆ filtering, for example blurring or sharpening, can be implemented by amplifying or attenuating selected frequencies
 - i.e. modifying the contrast of selected sine or cosine components relative to others, while maintaining same average over all components
 - attenuating high frequencies \approx *low-pass-filtering* \approx blurring
 - attenuating low frequencies \approx *high-pass filtering* \approx sharpening
- ◆ MTF measures preservation of frequencies by an optical system
 - subjective image quality depends on both sharpness and contrast
 - both can be restored, but at a price (in ringing or noise)

Questions?

Spatial resolution of display media

$$\text{pitch} = \Delta x \quad \begin{array}{c} \downarrow \\ \hline \hline \\ \uparrow \end{array} \quad \begin{array}{c} \blacksquare \\ \square \\ \blacksquare \\ \square \end{array} \quad \text{density} = 1/\Delta x$$

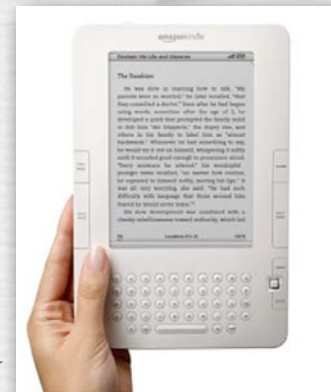
◆ Example #1: Macbook Air (laptop)

- 900 pixels on 7" high display
- $\Delta x = 7"/900 \text{ pixels} = 0.0078"/\text{pixel}$
- $1/\Delta x = 129 \text{ dpi}$ (dots per inch)

Line printers are 300 dpi.
This is why we don't like reading on laptops.

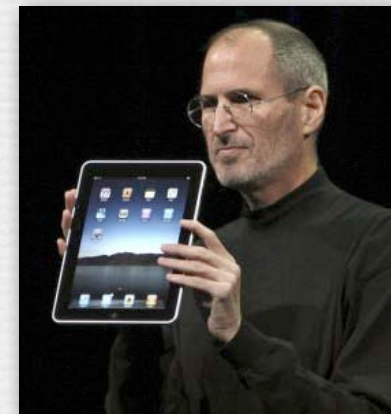
◆ Example #2: Kindle 2

- 800 pixels on 4.8" high display
- $1/\Delta x = 167 \text{ dpi}$



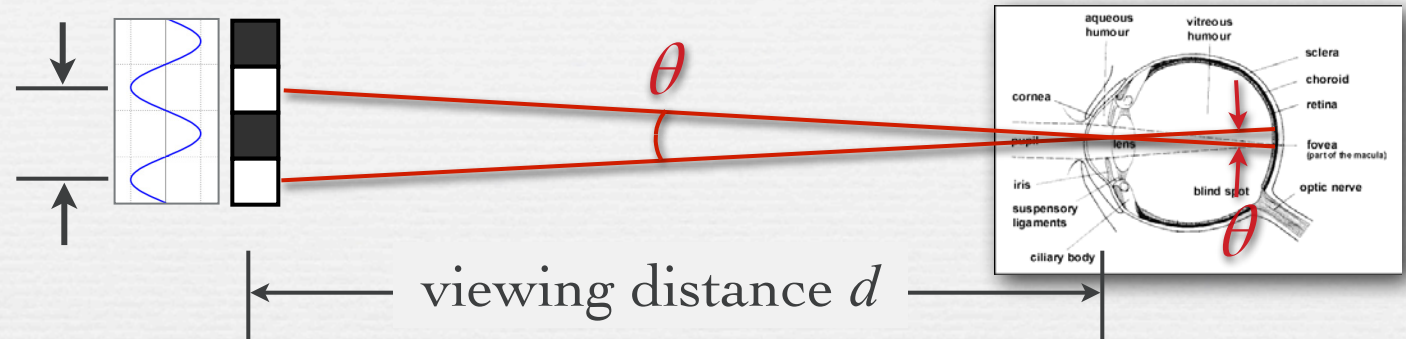
◆ Example #3: ~~iPad~~ iPad3

- ~~1024~~ 2048 pixels on 7.8" high display
- ~~132~~ 263 dpi



Spatial frequency on the retina

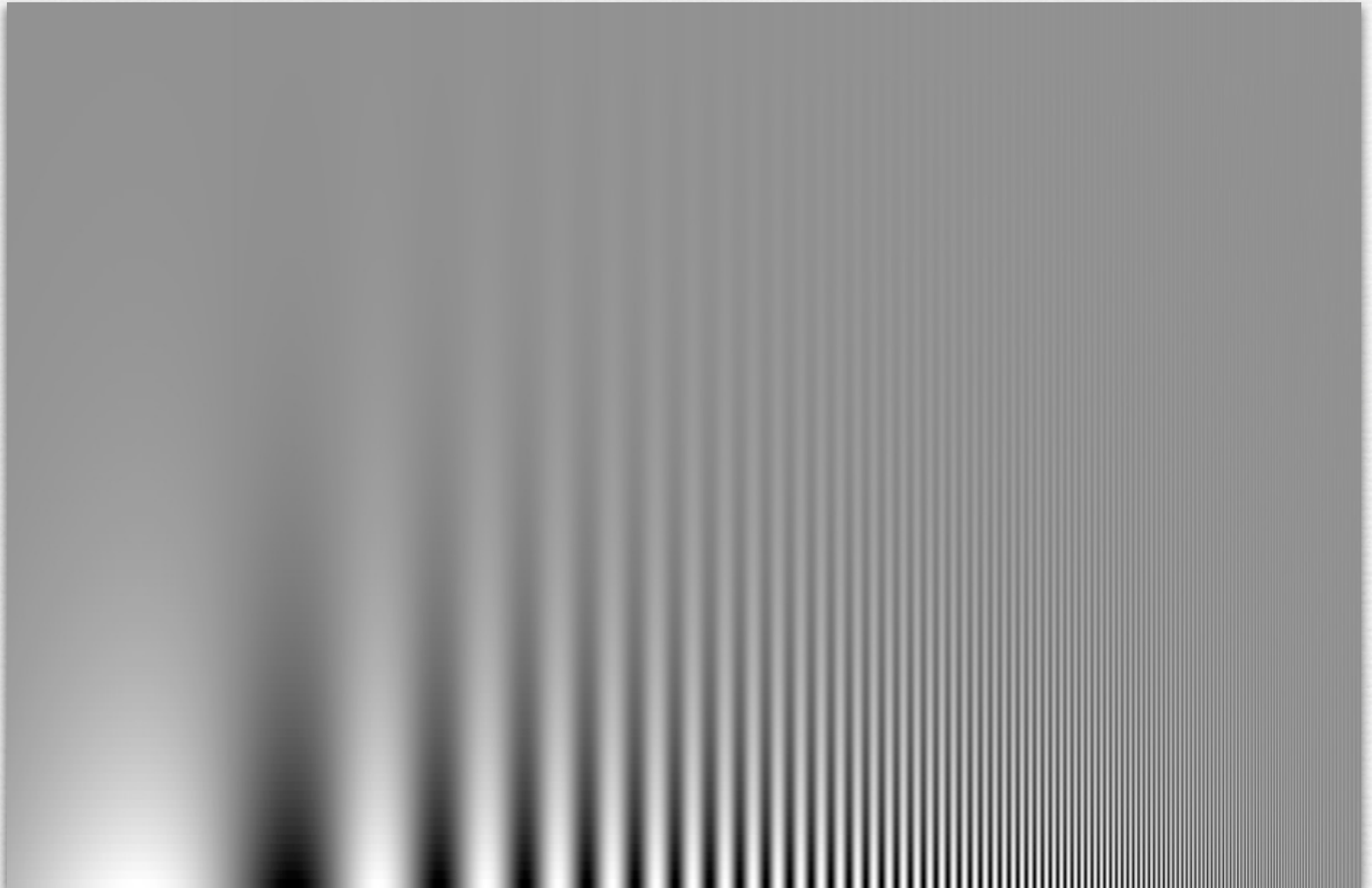
assume the minimum period p of a sine wave is a black-white pixel pair (“line pair”)



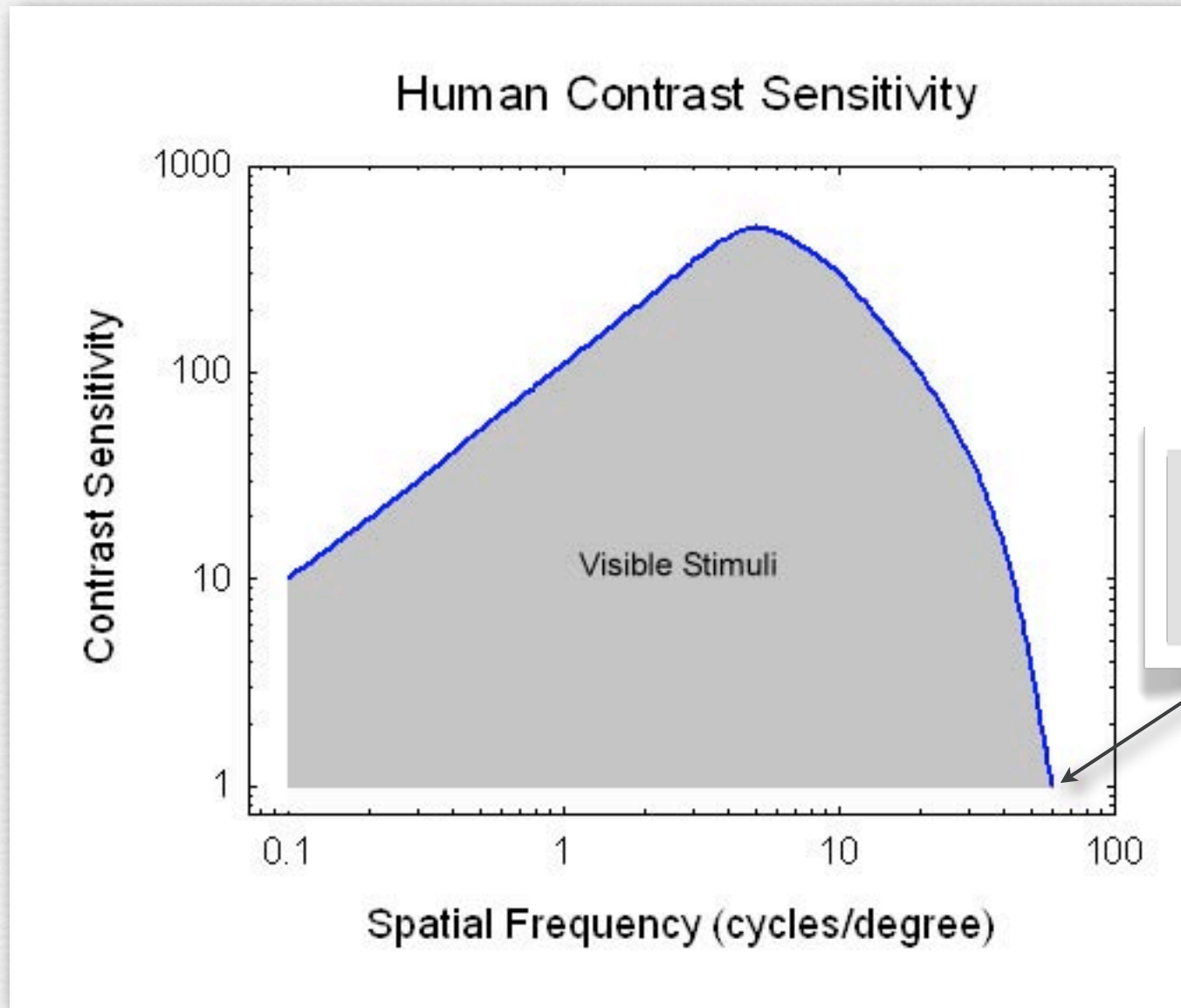
- ◆ Example #1: Macbook Air viewed at $d = 18''$
 - 900 pixels on 7'' high display, $p = 2 \times 0.0078''$
 - retinal arc $\theta = 2 \arctan (p / 2d) = 0.05^\circ$
 - spatial frequency on retina $1/\theta = 20$ cycles per degree

Q. What is the acuity of the human visual system?

Human spatial sensitivity (Campbell-Robson Chart)



Human spatial sensitivity

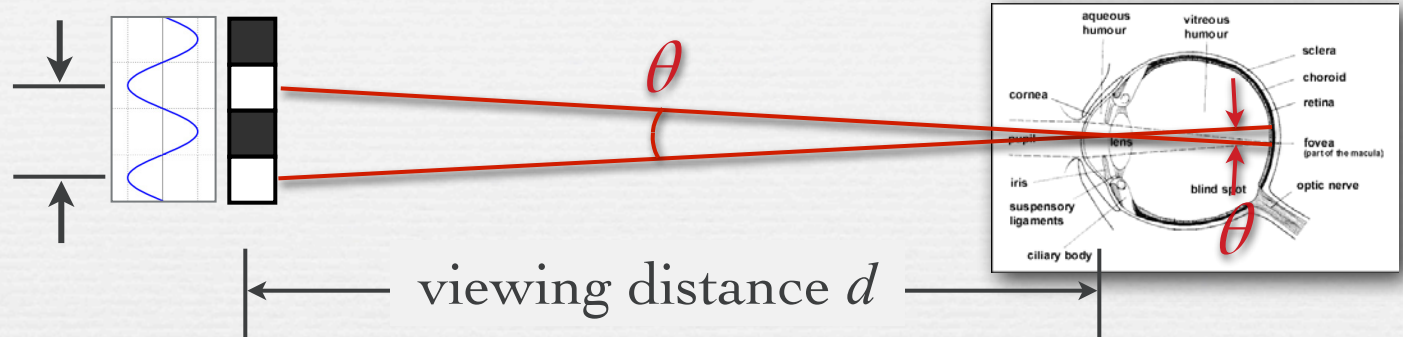


(horizontal axis
not comparable
to image on
previous slide)

cutoff is at
about 50 cycles
per degree

Spatial frequency on the retina

assume the minimum period p of a sine wave is a black-white pixel pair



- ◆ Example #1: Macbook Air viewed at $d = 18''$
 - 900 pixels on 7'' high display, so $p = 2 \times 0.0078''$
 - retinal arc $\theta = 2 \arctan (p / 2d) = 0.05^\circ$
 - spatial frequency on retina $1/\theta = 20$ cycles per degree

not nearly as high as human acuity



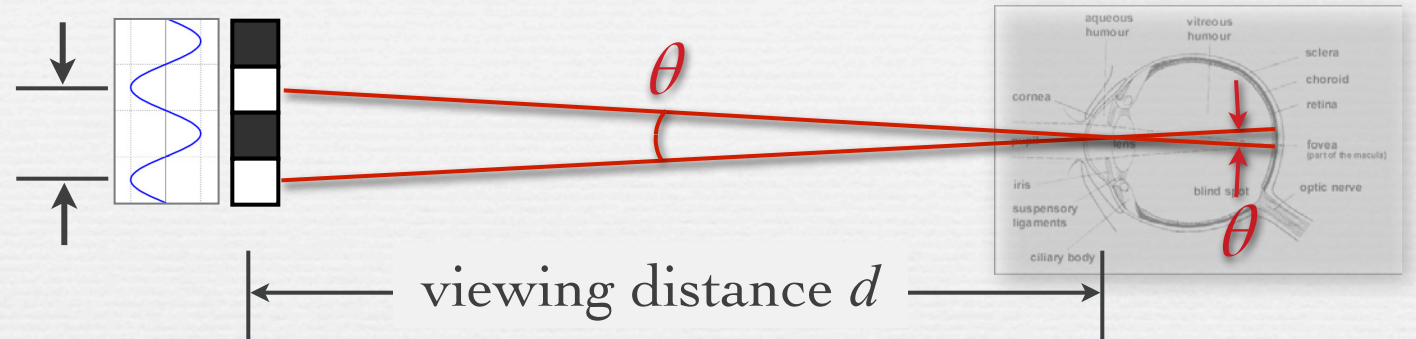
(Graham Flint)

Balboa Park, San Diego

(original is 40K × 20K pixels, Gates Hall print is 72" × 36")

Spatial frequency on the retina

assume the minimum period p of a sine wave is a black-white pixel pair



- ◆ Example #1: Macbook Air viewed at $d = 18''$
 - 900 pixels on 7'' high display, $p = 2 \times 0.0078''$
 - retinal arc $\theta = 2 \arctan (p / 2d) = 0.05^\circ$
 - spatial frequency on retina $1/\theta = 20$ cycles per degree
- ◆ Example #2: gigapixel photo viewed at $d = 48''$
 - 20,000 pixels on 36'' high print, $p = 2 \times 0.0018''$
 - spatial frequency on retina $1/\theta = 232$ cycles per degree

much finer than human acuity

Human acuity & circle of confusion

- ◆ the maximum allowable circle of confusion (C) in a photograph can be computed from human spatial acuity projected onto the intended display medium
 - depends on viewing distance
- ◆ Example: photographic print viewed at 12"
 - max human acuity on retina $1/\theta \approx 50$ cycles per degree
 - minimum detectable retinal arc $\theta \approx 0.02^\circ$
 - minimum feature size $p = 2 \times 12'' \times \tan(\theta / 2) = 0.0043''$ (0.1mm)
- ◆ assume 5" \times 7" print and Canon 5D II (5616 \times 3744 pixels)
 - 5" / 3744 pixels = 0.0017"/pixel (0.04mm)
 - therefore, circle of confusion can be 2.5 pixels wide before it's blurry
 - $C = 6.4\mu$ per pixel \times 2.5 pixels = 16 μ

To clarify, "minimum feature size p " means a pixel should be at least this big (0.0043") on a print viewed from 12 inches away in order to match human acuity. The figure of 0.0017" further down means that Canon 5D II pixels appear this big on a 5x7 print. Since 0.0017 is smaller than 0.0043, the printed pixels are smaller by 2.5x than we can actually perceive. Conversely, features in the scene can blur out to 2.5 pixels in the camera before we'll see that blur in a 5x7 print (viewed from 12 inches away). Thus, when shooting, we can treat the circle of confusion (C in the depth of field formula) as being 2.5 pixels wide rather than the usual 1 pixel, leading to a deeper depth of field. (A pixel on the Canon 5D II is 6.4 microns wide.)

- ◆ the maximum allowable circle of confusion (C) in a photograph can be computed from human spatial acuity projected onto the intended display medium
 - depends on viewing distance
- ◆ Example: photographic print viewed at 12"
 - max human acuity on retina $1/\theta \approx 50$ cycles per degree
 - minimum detectable retinal arc $\theta \approx 0.02^\circ$
 - minimum feature size $p = 2 \times 12'' \times \tan(\theta / 2) = 0.0043''$ (0.1mm)
- ◆ assume 5" \times 7" print and Canon 5D II (5616 \times 3744 pixels)
 - 5" / 3744 pixels = 0.0017"/pixel (0.04mm)
 - therefore, circle of confusion can be 2.5 pixels wide before it's blurry
 - $C = 6.4\mu$ per pixel \times 2.5 pixels = 16 μ

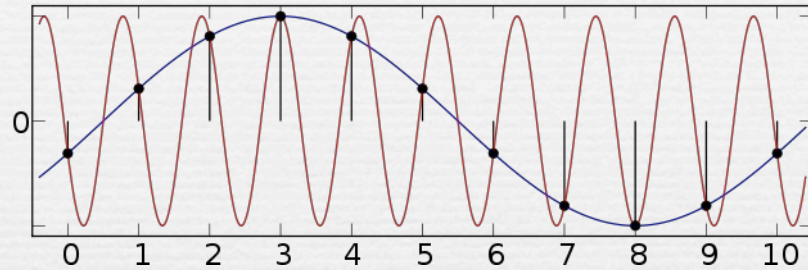
Recap

- ◆ spatial resolution of display media is measured by
 - pitch (distance between dots or pixels) or density (dots per inch)
- ◆ effect on human observers is measured by
 - retinal angle (degrees of arc) or frequency (cycles per degree)
 - depends on viewing distance
- ◆ human spatial acuity is about 50 cycles per degree
 - depends on contrast
 - convert back to pitch to obtain circle of confusion for depth of field, and this conversion depends on viewing distance

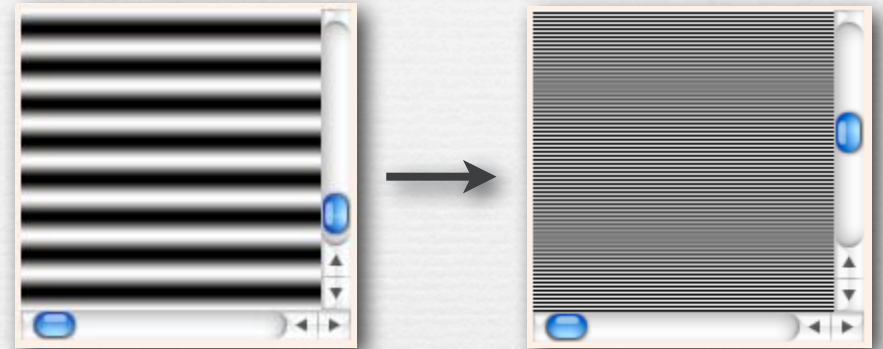
Questions?

Sampling and aliasing

abstract function



spatial aliasing in images

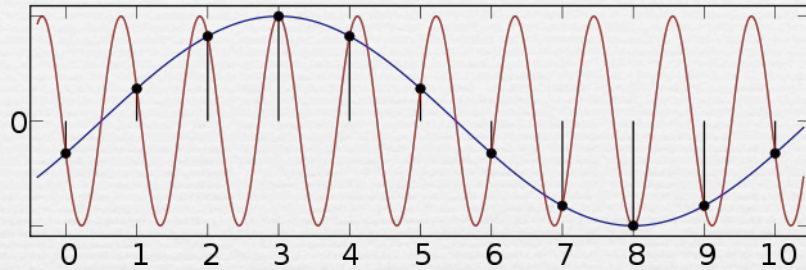


(<http://ptolemy.eecs.berkeley.edu/eecs20/week13/moire.html>)

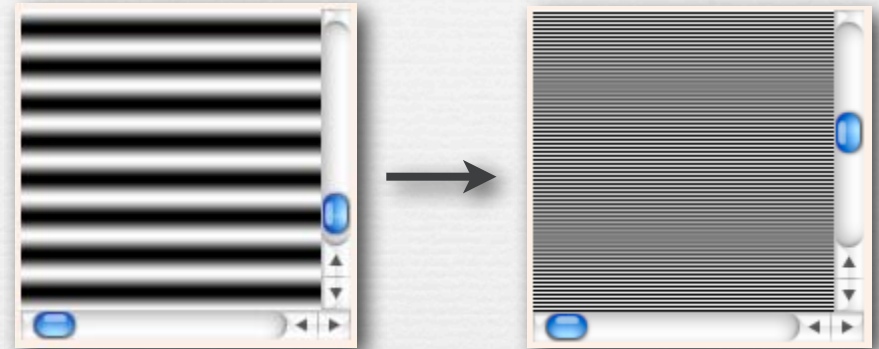
- ♦ aliasing is high frequencies masquerading as low frequencies due to insufficiently closely spaced samples

Sampling and aliasing

abstract function

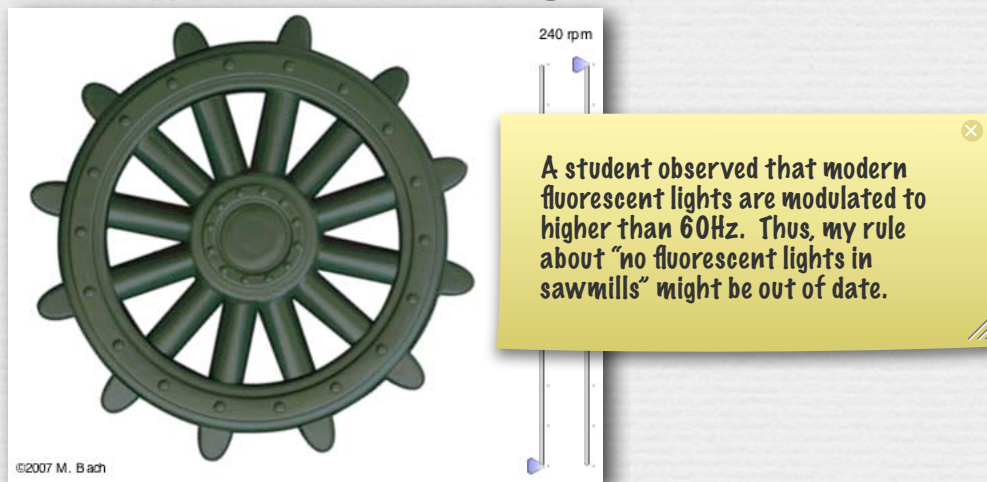


spatial aliasing in images



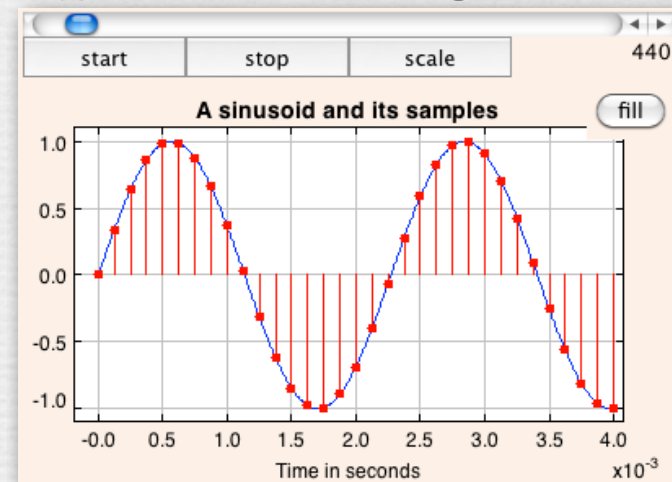
(<http://ptolemy.eecs.berkeley.edu/eecs20/week13/moire.html>)

temporal aliasing



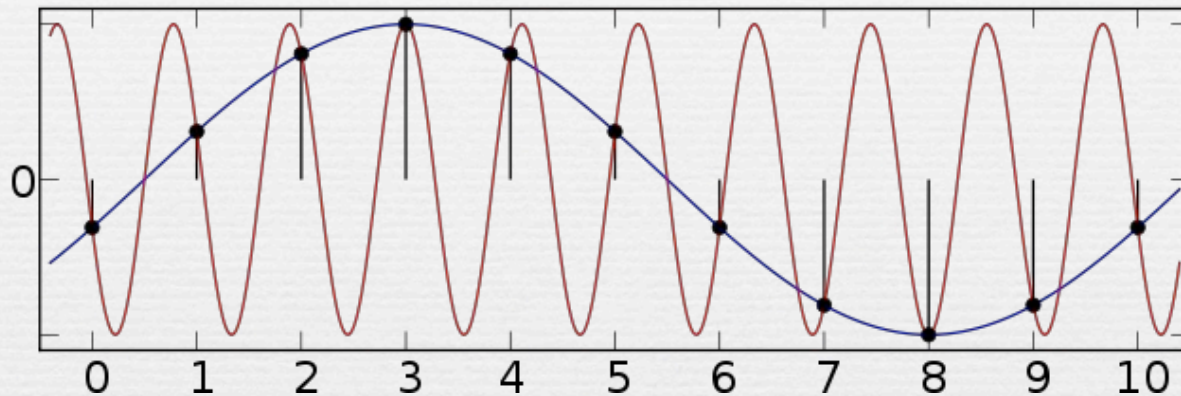
(http://www.michaelbach.de/ot/mot_wagonWheel/index.html)

temporal aliasing in audio



(<http://ptolemy.eecs.berkeley.edu/eecs20/week13/aliasing.html>)

Fourier analysis of aliasing

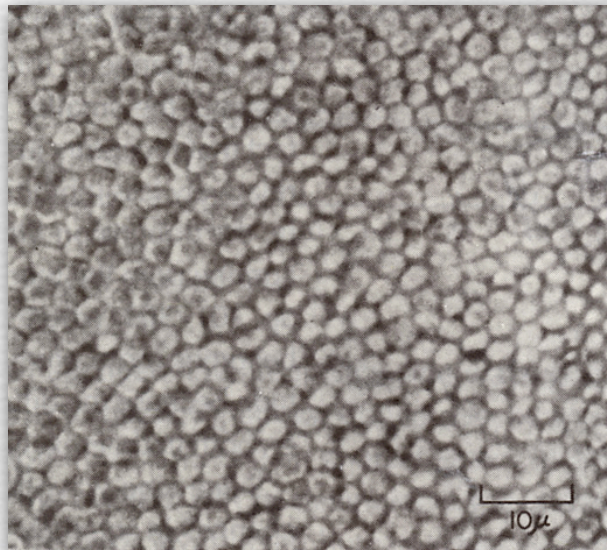


- ◆ Nyquist-Shannon sampling theorem: a function having frequencies no higher than n can be completely determined by samples spaced $1 / 2n$ apart

$$f_{\text{sampling}} > 2 \times f_{\text{cutoff}}$$

Retinal sampling rate

- ◆ the human retina consists of discrete sensing cells
- ◆ therefore, the retina performs sampling
- ◆ sampling theory says $f_{\text{sampling}} > 2 \times f_{\text{cutoff}}$
- ◆ if observed human cutoff is 50 cycles per degree, then its sampling rate must be > 100 samples per degree
- ◆ this agrees with observed retinal cell spacing!



(Cornsweet)

spacing between L,M cone cells is $1\mu \approx 30$ arc-seconds ($1/120^\circ$)

Retinal sampling rate

- ◆ the human retina consists of discrete sensing cells
- ◆ therefore, the retina performs sampling
- ◆ sampling theory says $f_{\text{sampling}} > 2 \times f_{\text{cutoff}}$
- ◆ if observed human cutoff is 50 cycles per degree, then its sampling rate must be > 100 samples per degree
- ◆ this agrees with observed retinal cell spacing!

yes, almost equal to human acuity

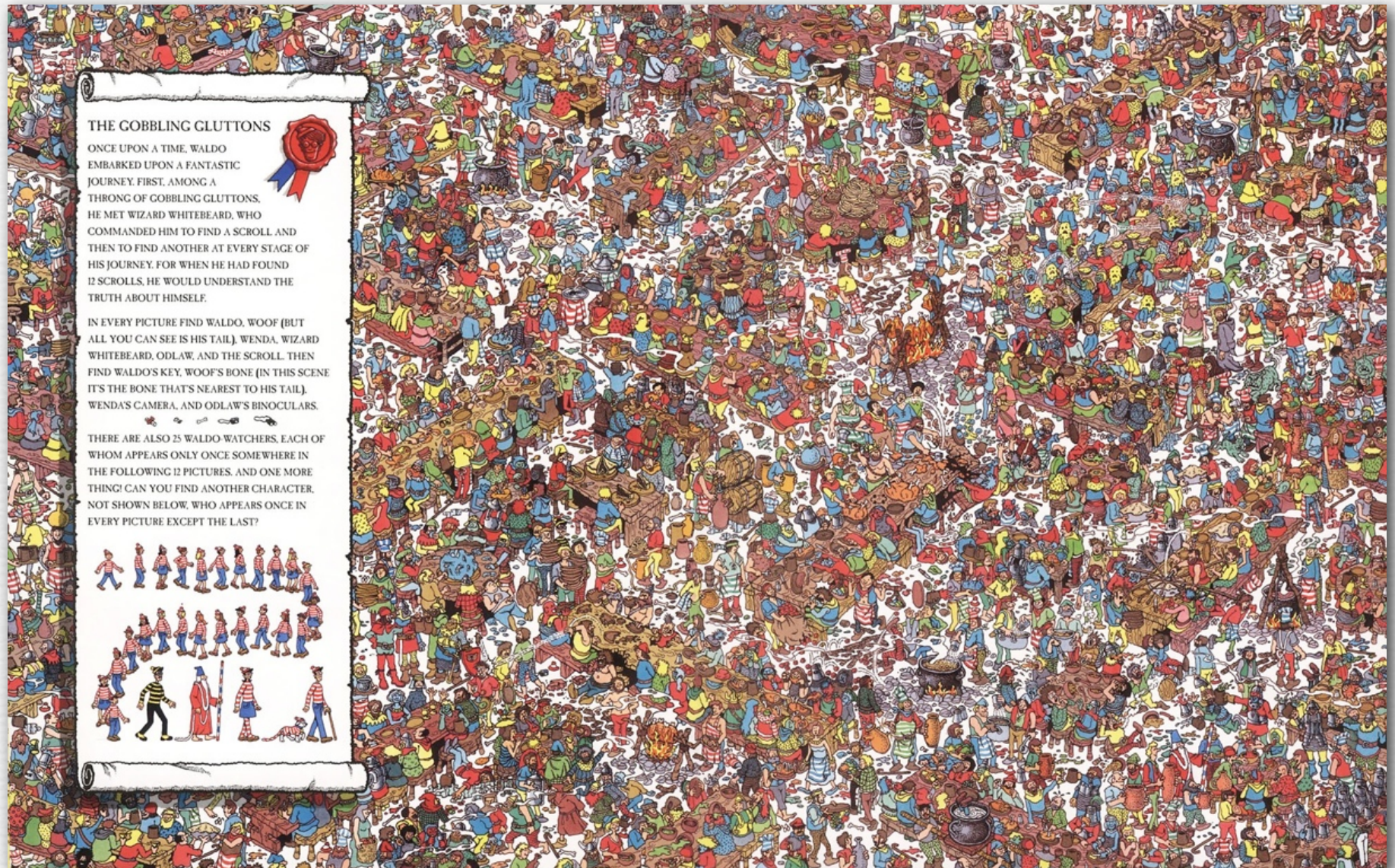
- ◆ Example #3: iPhone 4 “Retina Display” viewed at 12” inches
 - 960 pixels on 2.94” high display
 - $1/\Delta x = 326$ dpi
 - spatial frequency on retina = 34 cycles per degree



Aliasing in photography

- ◆ a lens creates a focused image on the sensor
- ◆ suppose the sensor measured this image at points on a 2D grid, but ignored the imagery between points?
 - a.k.a. *point sampling*

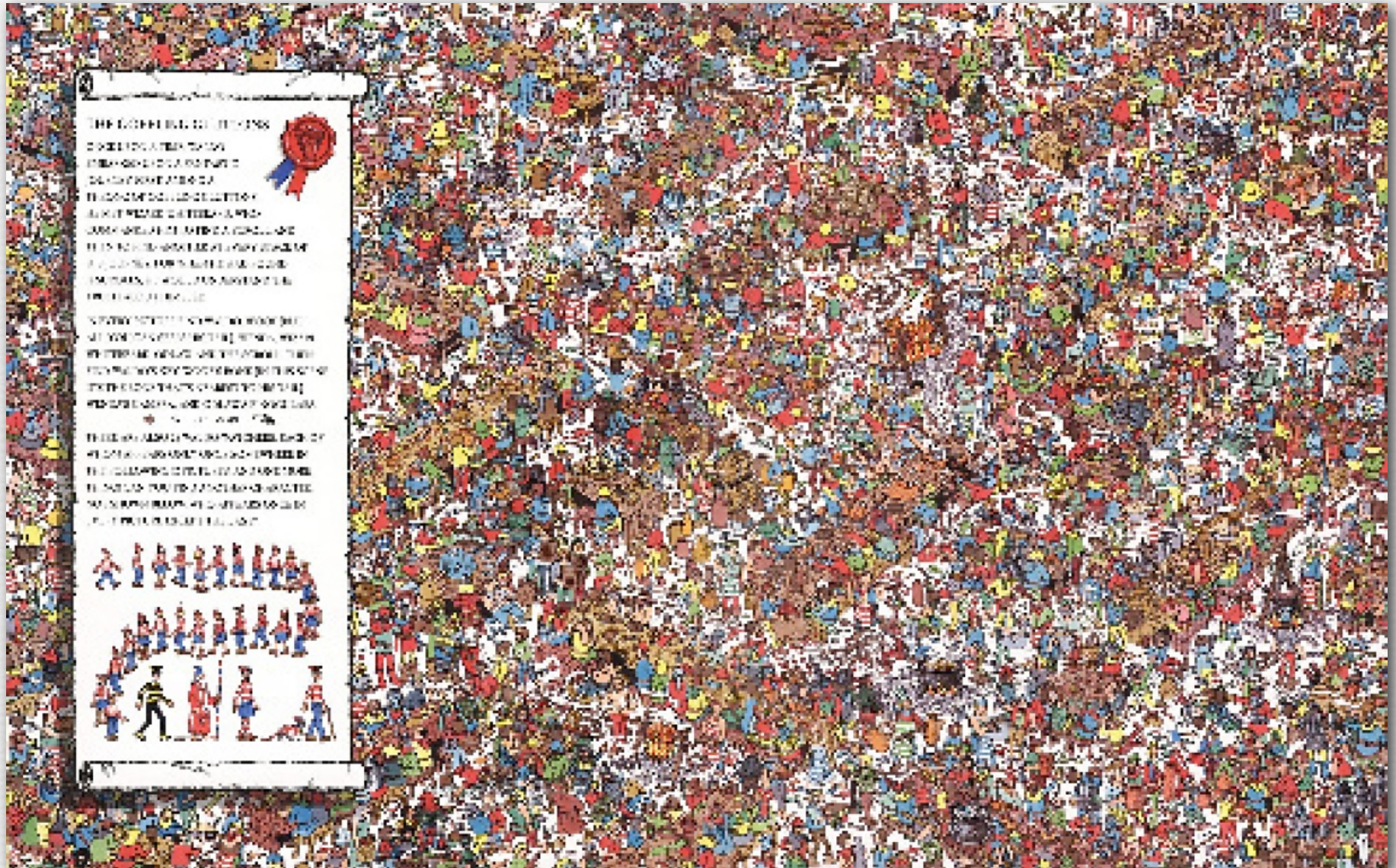
Simulation of point sampling



(Classic Media)

digital image, 1976 x 1240 pixels

Simulation of point sampling



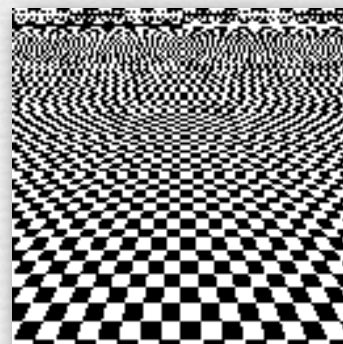
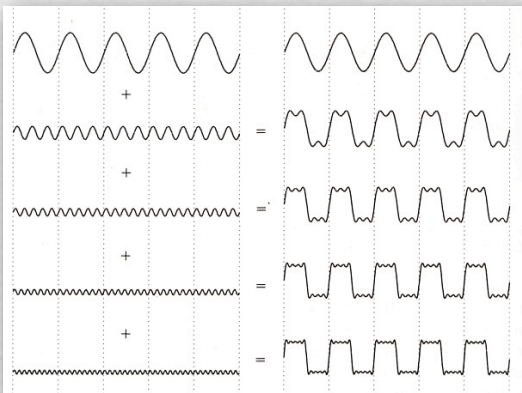
every 4th pixel in x and y , then upsized using pixel replication

Prefiltering to avoid aliasing

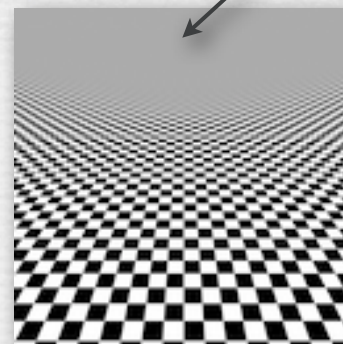
- ◆ before sampling, remove (or at least attenuate) sine waves of frequency greater than half the sampling rate

$$f_{cutoff} < \frac{1}{2} f_{sampling}$$

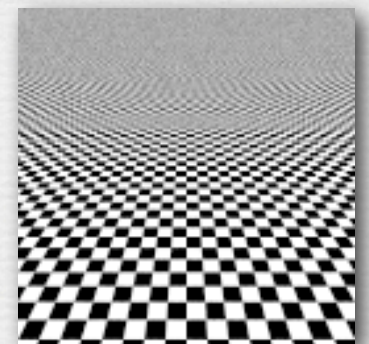
replace removed waves with their average intensity (gray in this case)



unfiltered



prefiltered



partially pre-filtered

Methods for prefiltering

- ◆ method #1: frequency domain
 1. convert image to frequency domain
 2. remove frequencies above f_{cutoff} (replace with gray)
 3. convert back to spatial domain
 4. perform point sampling as before
 - conversions are slow
 - not clear how to apply this method to images as they enter a camera

- ◆ method #2: spatial domain
 1. blur image using *convolution*
 2. perform point sampling as before
 - direct and faster
 - equivalent to method #1 (proof is beyond scope of this course)

Convolution in 1D

- ◆ replace each input value with a weighted sum of itself and its neighbors, with weights given by a filter function

$$f[x] * g[x] = \sum_{k=-\infty}^{\infty} f[k] \cdot g[x - k]$$

input signal $f[x]$

1	3	0	4	2	1
---	---	---	---	---	---

filter $g[x]$

2	1
---	---

output $f[x] * g[x]$

--	--	--	--	--	--

Convolution in 1D

A technicality I ignored during lecture is that if I really run this convolution from $-\infty$ to $+\infty$, then the non-zero part of the output should be bigger than the non-zero part of the input, and will taper towards zero over a band as wide as the filter. In practice, Photoshop and other programs clip the output to the size of the input, and sometimes also try to fix this band, so that your image doesn't end up with a dark border after convolution.

- ◆ replace each input value with a weighted sum of itself and its neighbors, with weights given by a filter function

$$f[x] * g[x] = \sum_{k=-\infty}^{\infty} f[k] \cdot g[x - k]$$

input signal $f[x]$

1	3	0	4	2	1
---	---	---	---	---	---

1	2
---	---

notice that the filter gets flipped when applied

output $f[x] * g[x]$

7					
---	--	--	--	--	--

Convolution in 1D

- ◆ replace each input value with a weighted sum of itself and its neighbors, with weights given by a filter function

$$f[x] * g[x] = \sum_{k=-\infty}^{\infty} f[k] \cdot g[x - k]$$

input signal $f[x]$

1	3	0	4	2	1
---	---	---	---	---	---

1	2
---	---

output $f[x] * g[x]$

7	3				
---	---	--	--	--	--

Convolution in 1D

- ◆ replace each input value with a weighted sum of itself and its neighbors, with weights given by a filter function

$$f[x] * g[x] = \sum_{k=-\infty}^{\infty} f[k] \cdot g[x - k]$$

input signal $f[x]$

1	3	0	4	2	1
---	---	---	---	---	---

1	2
---	---

output $f[x] * g[x]$

7	3	8			
---	---	---	--	--	--

More convolution formulae

- ◆ 1D discrete: defined only on the integers

$$f[x] * g[x] = \sum_{k=-\infty}^{\infty} f[k] \cdot g[x - k]$$

- ◆ 1D continuous: defined on the real line

$$f(x) * g(x) = \int_{-\infty}^{\infty} f(\tau) \cdot g(x - \tau) d\tau$$



(FLASH DEMO)

<http://graphics.stanford.edu/courses/cs178/applets/convolution.html>

More convolution formulae

- ◆ 1D discrete: defined only on the integers

$$f[x] * g[x] = \sum_{k=-\infty}^{\infty} f[k] \cdot g[x - k]$$

- ◆ 1D continuous: defined on the real line

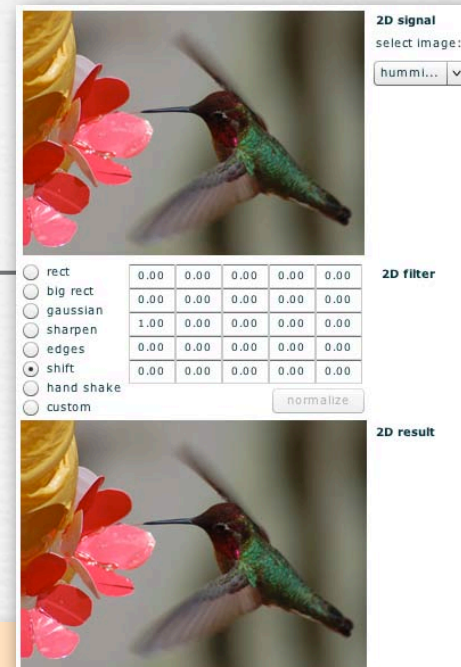
$$f(x) * g(x) = \int_{-\infty}^{\infty} f(\tau) \cdot g(x - \tau) d\tau$$

- ◆ 2D discrete: defined on the x, y integer grid

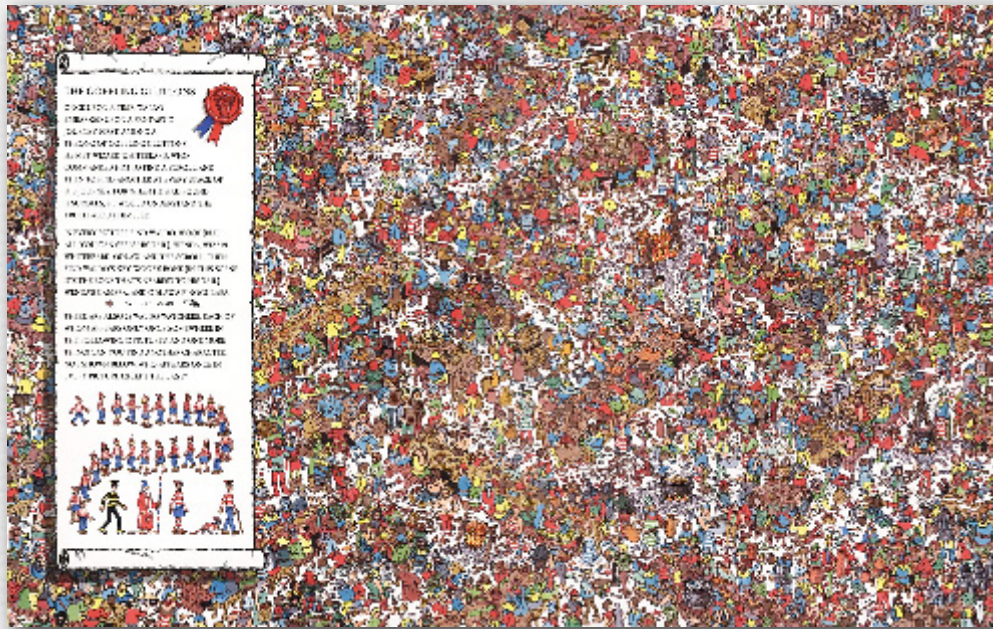
$$f[x,y] * g[x,y] = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} f[i,j] \cdot g[x - i, y - j]$$

- ◆ 2D continuous: defined on the x, y plane

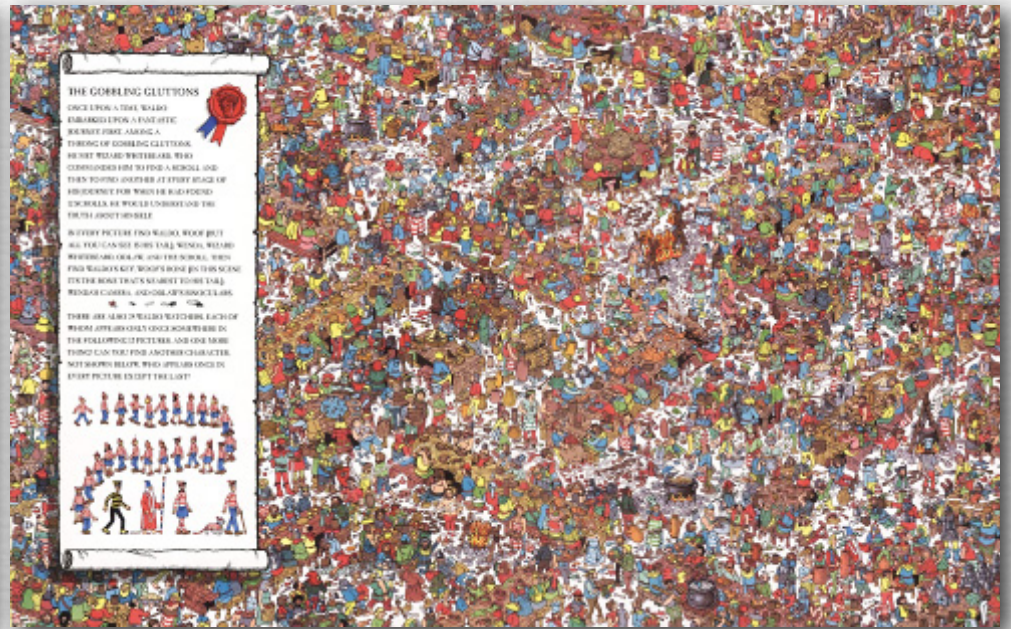
$$f(x,y) * g(x,y) = \int_{\tau_1=-\infty}^{\infty} \int_{\tau_2=-\infty}^{\infty} f(\tau_1, \tau_2) \cdot g(x - \tau_1, y - \tau_2) d\tau_1 d\tau_2$$



Prefiltering reduces aliasing



every 4th pixel in x and y



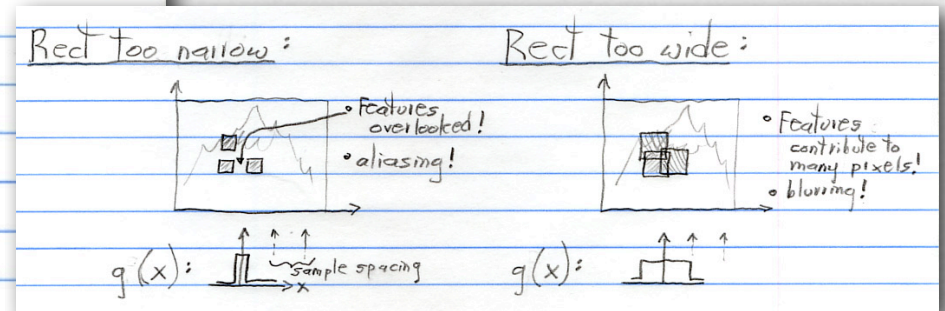
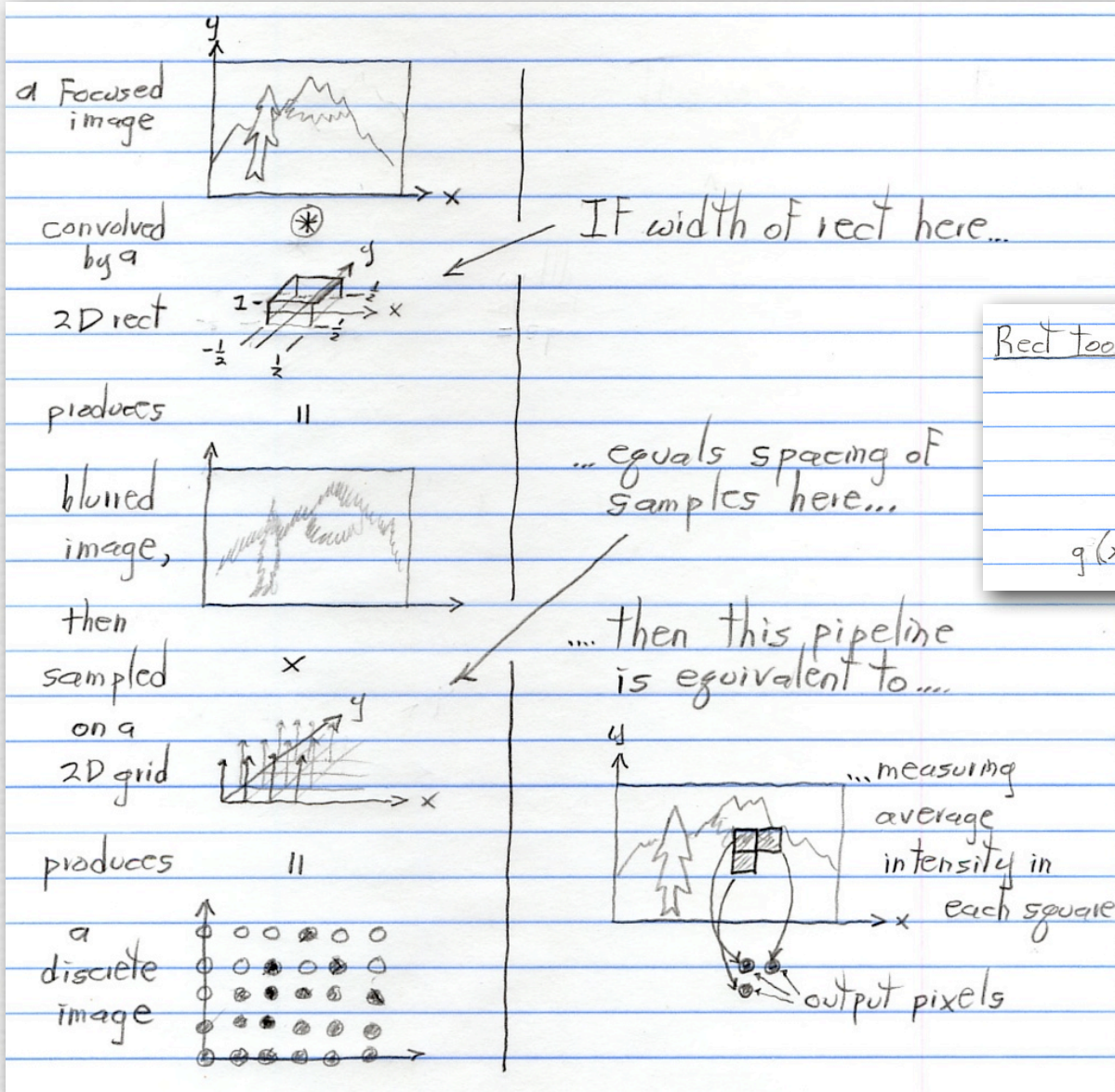
convolved by 4×4 pixel rect,
then sampled every 4th pixel

Prefiltering & sampling in photography

- ◆ photography consists of convolving the focused image by a 2D rect filter, then sampling on a 2D grid
 - each point on this grid is called a *pixel*
- ◆ if convolution is followed by sampling, you only need to compute the convolution at the sample positions
 - for a rect filter of width equal to the sample spacing, this is equivalent to measuring the average intensity of the focused image in a grid of abutting squares
 - this is exactly what a digital camera does
- ◆ the width of the rect is typically equal to the spacing between sample positions
 - narrower leaves aliasing; wider produces excessive blur

Prefiltering & sampling in photography (contents of whiteboard)

By the way, if you make the rect narrower than the canonical case (which runs from $-1/2$ to $+1/2$ in X), then you must make it taller (than $+1$ in Y), so that its area stays unity (1.0). Otherwise, as in Fourier filtering, your image will get dimmer, which is an unintended result. For the same reason, you saw me adjust the weights of the center column of pixels when I built an astigmatic filter in the convolution applet.

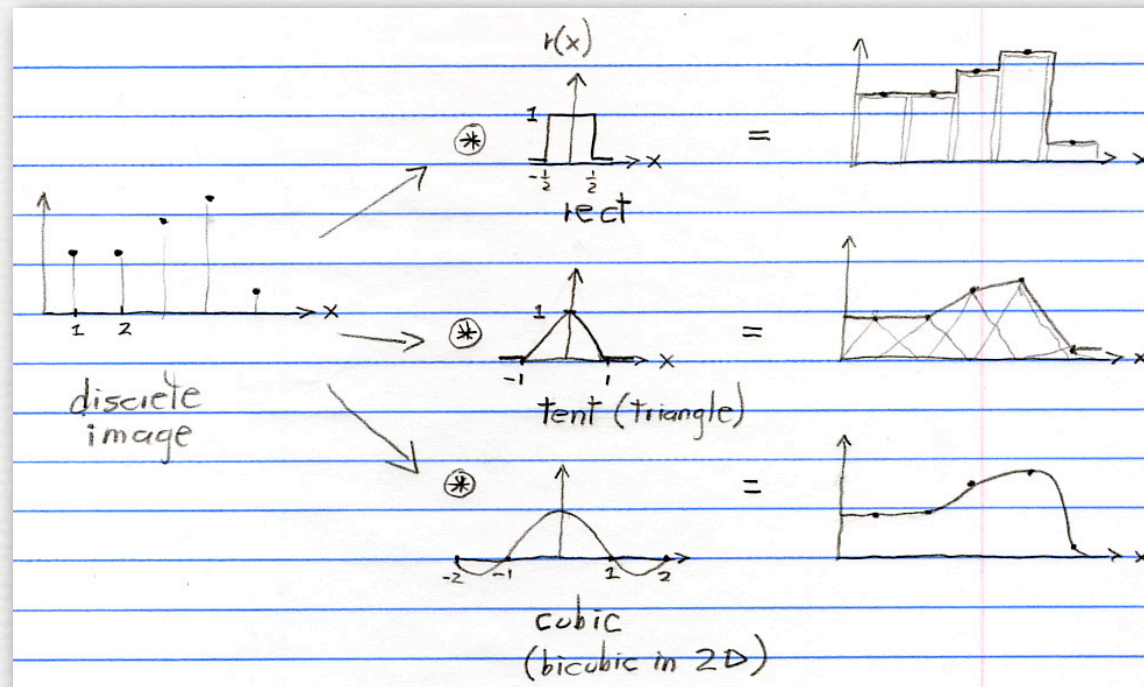


Upsizing/downsizing in Photoshop

- ◆ *resampling* is the conversion of a discrete image into a second discrete image having more or fewer samples
 1. interpolate between samples using convolution
 2. if downsizing, blur to remove high frequencies
 3. point sample at the new rate
- these steps can be simplified into a single discrete convolution

I didn't explain in class how one might combine these steps into a single discrete convolution, and I won't hold you responsible for knowing it. Briefly, steps 1 and 2 are both convolutions, and convolution is associative. Thus, $(f \circledast r) \circledast g \equiv f \circledast (r \circledast g)$, where f is the input image, r is the reconstruction filter (see next slide), and g is a blurring filter (such as are shown in the online convolution applet). This equation says that the two filters r and g can be convolved with each other, thereby producing a single filter, sometimes called the resampling filter ρ (rho), which is larger in non-zero extent than either r or g , and that can be applied to the input image, followed by point sampling at the new rate.

Interpolation via convolution (contents of whiteboard)

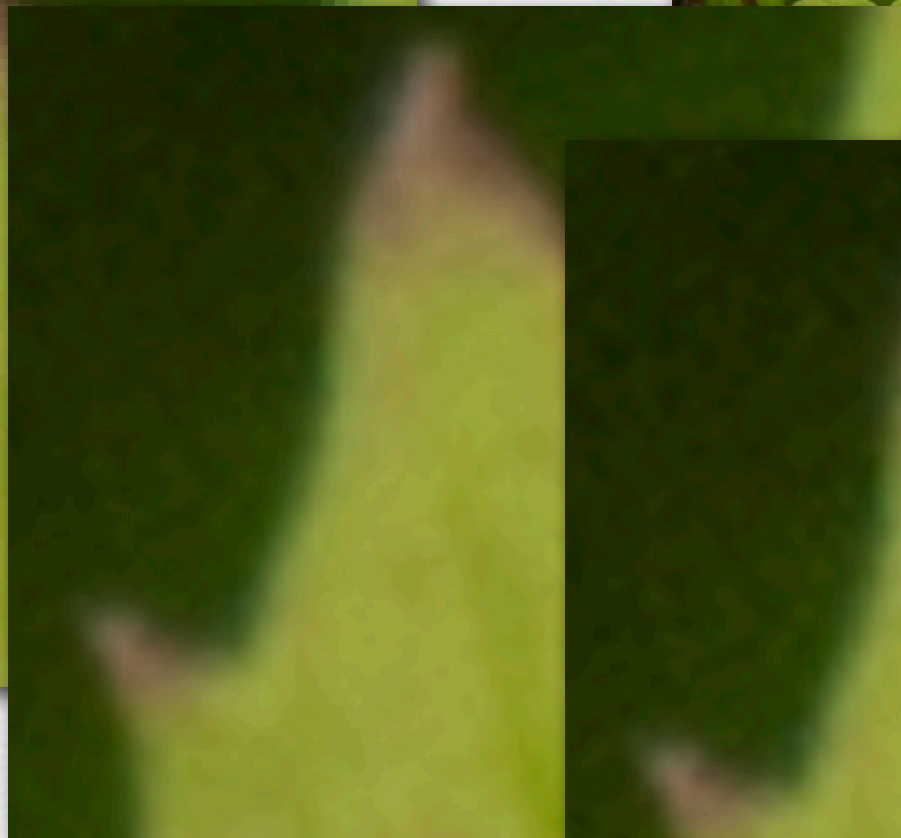


- ◆ if the input is a discrete (i.e. sampled) function, then convolution can be treated as placing a vertically-scaled copy of the filter $r(x)$ at each sample position as shown, summing the results, and dividing by the area under the filter (1.0 in the cases shown)
- ◆ the effect is to interpolate between the samples, hence reconstructing a continuous function from the discrete function

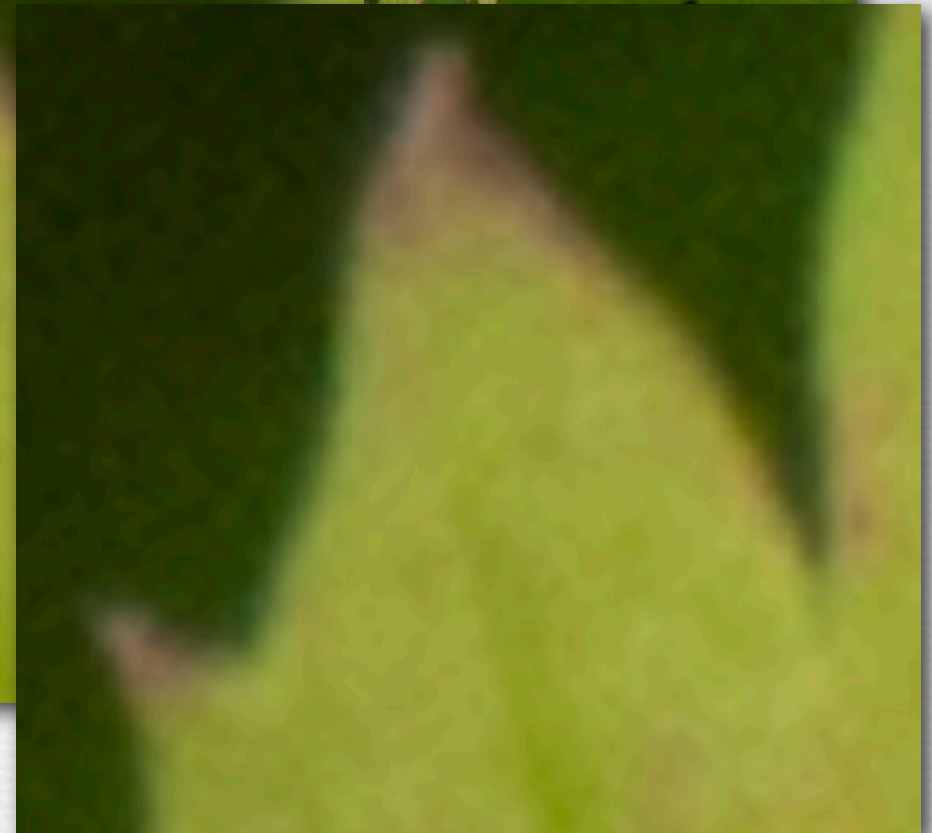
Upsizing by 16:1



nearest neighbor
(a.k.a. pixel replication)



bilinear

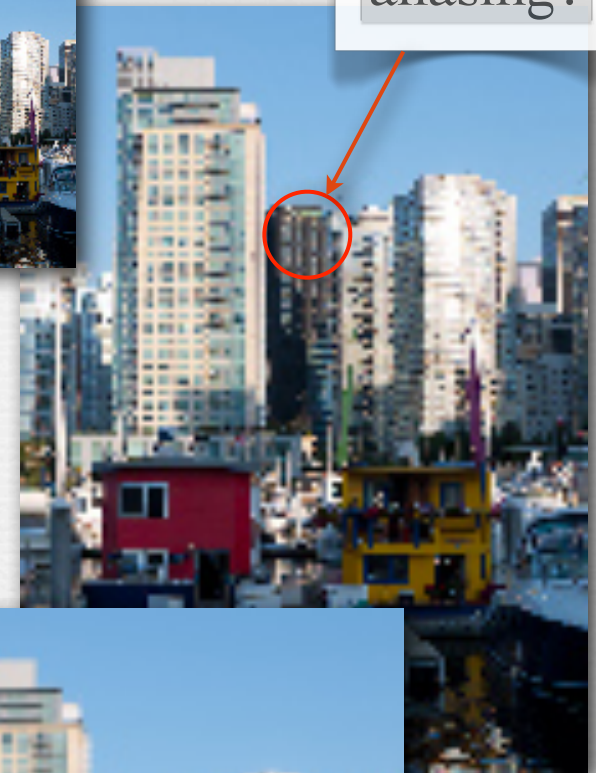


bicubic

Downsizing by 1:6



nearest neighbor
(point sampling)



aliasing!

bicubic



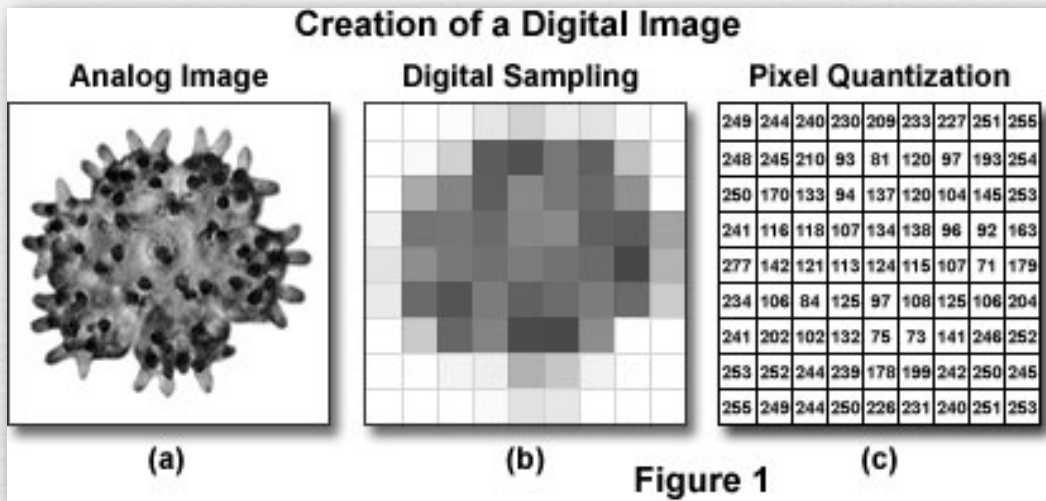
I used pixel replication in this blowup, which is included solely so that you can more clearly see the pixels in the actual downsized image (the small one at left).

Recap

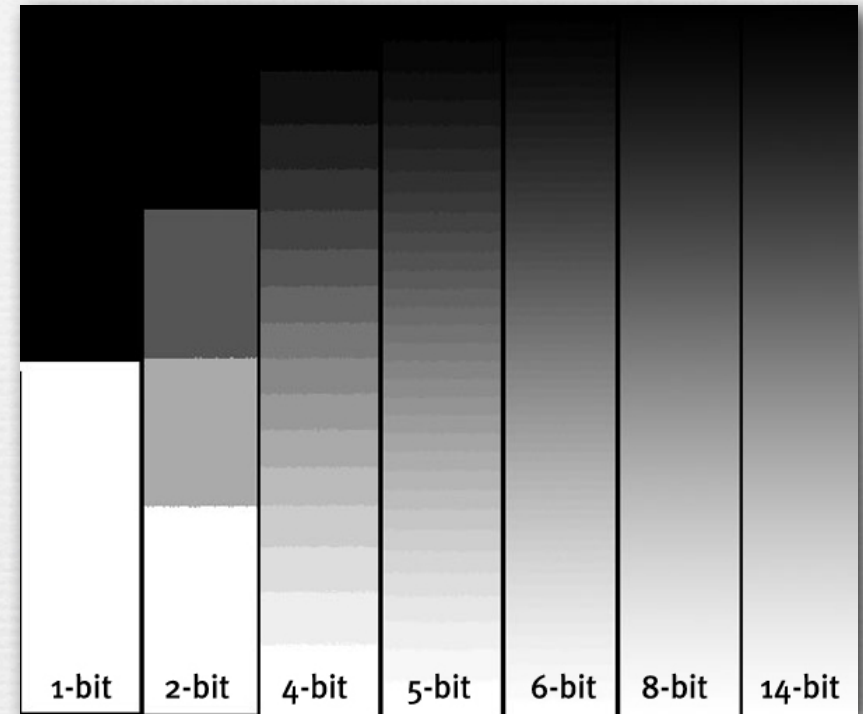
- ◆ *aliasing* is high frequencies masquerading as low frequencies due to insufficiently closely spaced samples
- ◆ reduce aliasing by prefiltering the input before sampling
 - implement by multiplication in the frequency domain
 - or convolution in the spatial domain
 - in the spatial domain, the prefilter is denoted $g(x)$
- ◆ in digital photography:
 - $g(x)$ is a pixel-sized rect, thus averaging intensity over areas
 - if the rect is too small, aliasing occurs; solve with antialiasing filter

Questions?

Sampling versus quantization



(<http://learn.hamamatsu.com/articles/digitalimagebasics.html>)



(Canon)

- ◆ an image is a function $\vec{f}(\vec{x})$
 - typically $(\vec{x}) = (x, y)$ and $\vec{f} = (R, G, B)$
- ◆ we sample the domain (\vec{x}) of this function as pixels
- ◆ we quantize the range \vec{f} of this function as intensity levels

Example

8 bits × R,G,B =
24 bits per pixel



Canon 1D III,
300mm, f/3.2

Example

8 bits \times R,G,B =
24 bits per pixel



Example

6 bits × R,G,B =
18 bits per pixel



Example

5 bits × R,G,B =
15 bits per pixel



Example

4 bits × R,G,B =
12 bits per pixel



Example

$3 \text{ bits} \times R, G, B =$
9 bits per pixel



Dithering

256 colors (8 bits) uniformly distributed across RGB cube, patterned dithering in Photoshop



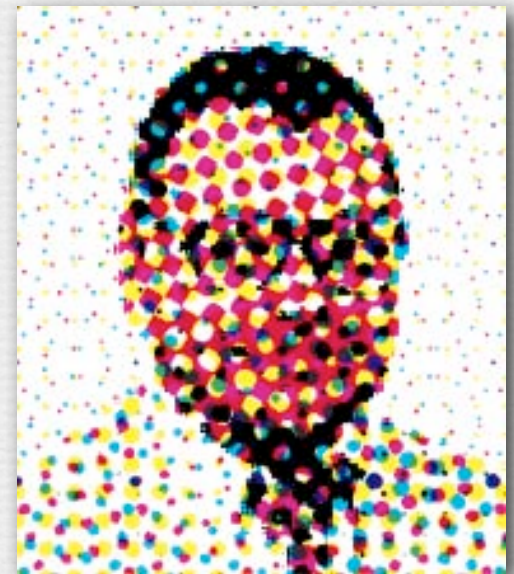
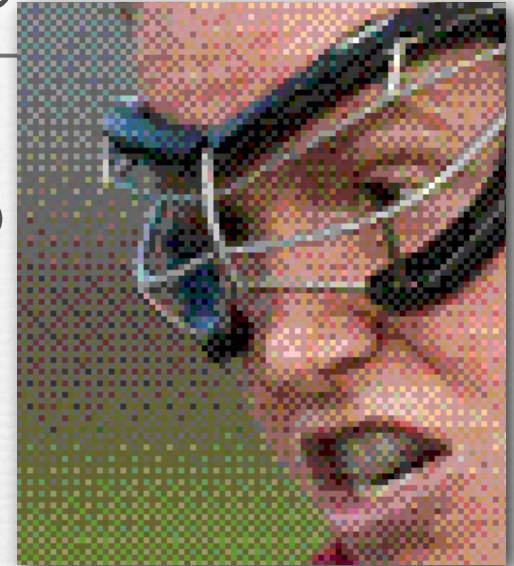
Dithering

256 colors (8 bits) adaptively distributed across RGB cube, patterned dithering in Photoshop



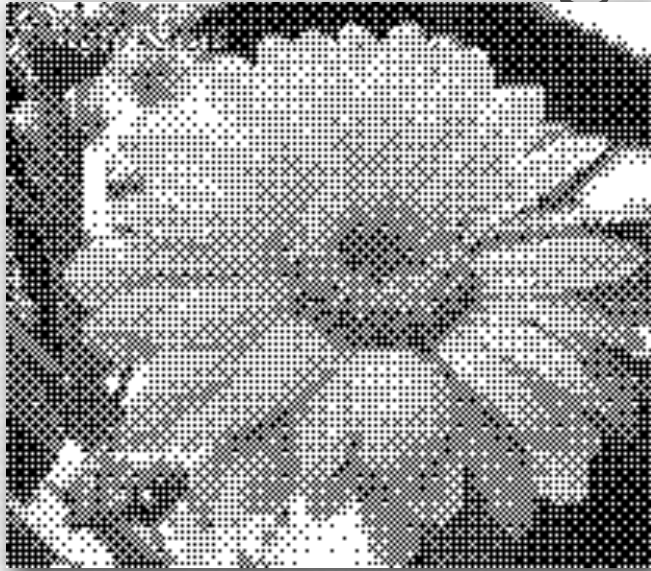
Dithering versus halftoning

- ◆ *dithering* for display (on a screen)
 - palette of a few hundred colors (uniform or adaptive)
 - flip some pixels in each neighborhood to the next available color in the palette to approximate intermediate colors when viewed from a distance
- ◆ *halftoning* for printing (on paper)
 - palette of only 3 or 4 colors (primaries)
 - print each primary as a grid of dots, superimposed but slightly offset from the other primaries, and vary dot size locally to approximate intermediate colors
- ◆ both techniques are applicable to full-color or black and white imagery
- ◆ both trade off spatial resolution to obtain more colors, hence to avoid quantization (contouring)



(wikimedia)

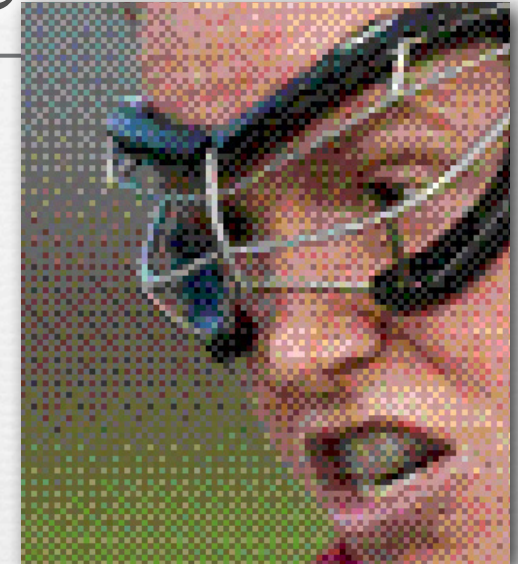
Dithering versus halftoning



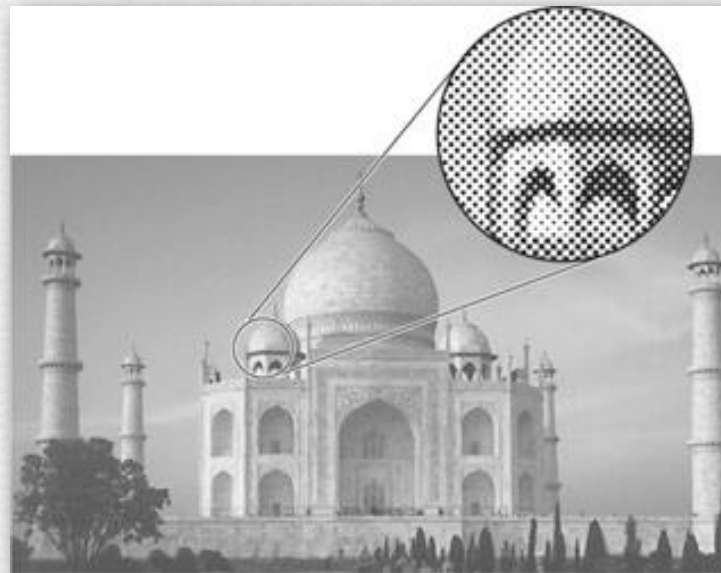
binary dithering



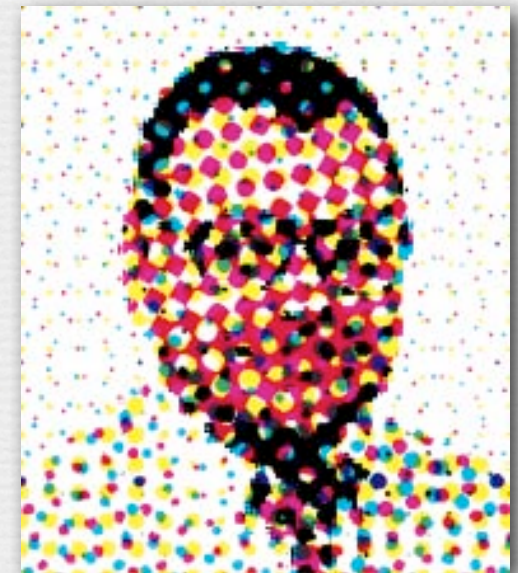
grayscale dithering



color dithering



grayscale halftoning



color halftoning

(see <http://bisqwit.iki.fi/jutut/colorquant/> for more examples)

Recap

- ◆ *sampling* describes where in its domain you measure a function
 - for uniformly spaced samples, you can specify a *sampling rate*
 - if the sampling rate is too low, you might suffer from *aliasing*
 - you can reduce aliasing by *prefiltering*

- ◆ *quantization* describes how you represent these measurements
 - for uniformly spaced levels, you can specify a *bit depth*
 - if the bit depth is too low, you might suffer from *contouring*
 - you can reduce contouring by *dithering* (if displaying the image on a screen) or *halftoning* (if printing it on paper)

Questions?

Slide credits

◆ Pat Hanrahan

◆ Cornsweet, T.N., *Visual Perception*, Kluwer Academic Press, 1970.