



6.869/6.819 Advances in Computer Vision

Bill Freeman, Phillip Isola







Weds, March 3, 2021

Problem set 1 due today Problem set 2 out today (due in a week).

- Comment on problem 2 of ps2
- Fourier transforms and signal processing, continued
- Fourier processing by human visual system
- Spatial digital filters



Projector







If we know u_c , u_p , f, and d, we can infer depth Z_c



















F recognizes that this is the pattern that the projector projected at $u_p=32$ pixe

= 32



Visualizing the image Fourier transform

$$F[u,v] = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} f[n,m] \exp\left(-2\pi j \left(\frac{un}{N} + \frac{vm}{M}\right)\right)$$

The values of F[u,v] are complex.

Using the real and imaginary components:

Or using a polar decomposition:

 $F[u,v] = A[u,v] \exp(j\theta[u,v])$ Amplitude

 $F[u,v] = Re\{F[u,v]\} + jImag\{F[u,v]\}$

Phase



14

Visualizing the image Fourier transform







real



imaginary



magnitude

phase



15



Images



The DFT Game: find the right pairs





(Solution in the class notes)

The DFT Game: find the right pairs







The inverse Discrete Fourier transform

2D Discrete Fourier Transform (DFT) transforms an image f [n,m] into F [u,v] as:

$$F[u,v] = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} f[n,m] \exp\left(-2\pi j \left(\frac{un}{N} + \frac{vm}{M}\right)\right)$$

The inverse of the 2D DFT is:

$$f[n,m] = \frac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} F[u,v] \exp\left(+2\pi j \left(\frac{un}{N} + \frac{vm}{M}\right)\right)$$

How does summing waves ends up giving back a picture?













#2: Range [1.89e-007, 0.226] Dims [256, 256]





#1: Range [0, 1] Dims [256, 256]



#2: Range [4.79e-007, 0.503] Dims [256, 256]





#1: Range [0, 1] Dims [256, 256]



#2: Range [8.5e-006, 1.7] Dims [256, 256]







#2: Range [3.85e-007, 2.21] Dims [256, 256]





#1: Range [0, 1] Dims [256, 256]



#2: Range [8.25e-006, 3.48] Dims [256, 256]





#1: Range [0, 1] Dims [256, 256]



#2: Range [1.39e-005, 5.88] Dims [256, 256]







#2: Range [6.17e-006, 8.4] Dims [256, 256]





#1: Range [0, 1] Dims [256, 256]



#2: Range [9.99e-005, 15] Dims [256, 256]





#1: Range [0, 1] Dims [256, 256]



#2: Range [8.7e-005, 19] Dims [256, 256]





#1: Range [0, 1] Dims [256, 256]

4052.



#2: Range [0.000556, 37.7] Dims [256, 256]





8056.



#2: Range [0.00032, 64.5] Dims [256, 256]







#2: Range [0.000231, 91.1] Dims [256, 256]







#2: Range (0.00109, 146) Dims (256, 256)





49190.



#2: Range [0.00758, 294] Dims [256, 256]



65536.





#2: Range [4.43e-015, 255] Dims [256, 256]



Now, an analogous sequence of images, but selecting Fourier components in descending order of magnitude.



3

indow Help	<u>د</u>
3	
#2: Range [0 237_0 54!	51
Dims [256, 256]	·1
ionco of image	s but colocting
JEILE UL IIIAge	s, but selecting
descending or	der of magnitud





5	




































J Figure 14								
File	Edit	View	Insert	Tools	Desktop	W		
D	2 6	6		ତ୍ର୍	(†) (†)	1		



#1: Range [0, 1] Dims [256, 256]





Figure 15								
File	Edit	View	Insert	Tools	Desktop	W		
۵	2	6		ର୍ ପ	. 🖑 🖲	1		





























🍌 i	igu	re	20								
File	Edi	t	View	Ī	nsei	t	Τo	ols	Des	¢op	7
	F		9	1	3		2	Q	<u>سمج</u>	۲	
			_								
				#	:1:	Ra	ang	ge (0.5,	1.5]
					D	im	IS	[251	6, 25	56]	







DFT













DFT















Some visual areas...







Figure 1. Stimulus presentation scheme. The stimuli were originally calibrated to be seen at a distance of 150 cm in a 19" display.



Campbell & Robson chart



What do you think you should see when looking at this image?

$\mathbf{I}[n,m] = A[n]\sin(2\pi f[m]m/M)$





$\mathbf{I}[n,m] = A[n]\sin(2\pi f[m]m/M)$



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20



Things that are very close and/or large are hard to see Things far away are hard to see





Vasarely visual illusion



Input visual stimulus



Frequency filtering of human visual system





bandpass filtered output



Center-surround spatial filtering of human visual system is subtracting less positive intensity at the corners, giving a bright line there





Today: A collection of useful filters



Low-pass filters



High-pass filters





Low pass-filters



 $h_{N,M}[n,m] = \begin{cases} 1 & \text{if } -N \le n \le N \text{ and } -M \le m \le M \\ 0 & \text{otherwise} \end{cases}$



2M+1





256X256

What does it do?

- Achieve smoothing effect (remove sharp features)

mean	
$\frac{1}{21}$	
mean	

256X256

• Replaces each pixel with an average of its neighborhood







The box filter is separable as it can be written as the convolution of two 1D kernels

 $h_{N,M}[n,m] = h_{N,0} \circ h_{0,M}$









256X256



Requires N+N sums, instead of N*N





If you convolve two boxes:



The convolution of two box filters is not another box filter. It is a triangular filter.



Gaussian filter

In the continuous domain:





Gaussian filter

 $g(x, y; \sigma) =$

Discretization of the Gaussian:

$$\frac{1}{2\pi\sigma^2}\exp{-\frac{x^2+y^2}{2\sigma^2}}$$

At 3σ the amplitude of the Gaussian is around 1% of its central value

 $g[m,n;\sigma] = \exp{-\frac{m^2 + n^2}{2\sigma^2}}$



 $g[m,n;\sigma] = \exp{-\frac{m^2 + n^2}{2\sigma^2}}$





Scale







Gaussian filter










Properties of the Gaussian filter

 $g(x, y; \sigma) = \frac{1}{2\sigma}$

- The n-dimensional Gaussian is the only completely circularly symmetric operator that is separable.
- The (continuous) Fourier transform of a Gaussian is another gaussian

$$\frac{1}{\pi\sigma^2}\exp-\frac{x^2+y^2}{2\sigma^2}$$

 $G(u,v;\sigma) = \exp(-2\pi^2(u^2 + v^2)\sigma^2)$



Properties of the Gaussian filter

 $g(x, y; \sigma) = \frac{1}{2\sigma}$

The convolution of two n-dimensional gaussians is an n-dimensional gaussian.

 $g(x, y; \sigma_1) \circ g(x, y)$

where the variance of the result is the sum

$$\sigma_3^2 = \sigma_1^2 +$$

(it is easy to prove this using the FT of the gaussian)

$$\frac{1}{\pi\sigma^2}\exp-\frac{x^2+y^2}{2\sigma^2}$$

$$\sigma_2$$
; σ_2) = $g(x, y; \sigma_3)$

$$\sigma_2^2$$



Discretization of the Gaussian

when working with discretized gaussians.



There are very efficient approximations to the Gaussian filter for certain values of σ with nicer properties than

 $g_5[n] = [0.0183, 0.3679, 1.0000, 0.3679, 0.0183]$



Binomial filter

the gaussian coefficients using only integers.

The simplest blur filter (low pass) is

Binomial filters in the family of filters obtained as successive convolutions of [1 1]

- Binomial coefficients provide a compact approximation of

 - 1 1



Binomial filter

$b_2 = [1 \ 1] \circ [1 \ 1] = [1 \ 2 \ 1]$ $b_3 = [1 1] \circ [1 1] \circ [1 1] = [1 3 3 1]$

 $b_1 = [1 \ 1]$



Binomial filter

 b_1 b_2 1 2 1 3 b_3 6 b_4 4 b_5 1 15 20 b_6 1 6 35 35 21 21 b_7 b_8 1 8 28 56 70 56 28 8 1 $\sigma_8^2 = 2$



 $\sigma_1^2 = 1/4$ $\sigma_2^2 = 1/2$ $\sigma_{3}^{2} = 3/4$ $\sigma_{4}^{2} = 1$ $\sigma_{5}^{2} = 5/4$ $\sigma_{6}^{2} = 3/2$ $\sigma_{7}^{2} = 7/4$



- Sum of the values is 2ⁿ
- The variance of b_n is $\sigma^2 = n/4$
- The convolution of two binomial filters is also a binomial filter

With a variance:

 $\sigma_n^2 + \sigma_m^2 = \sigma_{n+m}^2$

gaussian)

Properties of binomial filters

- $b_n \circ b_m = b_{n+m}$
- These properties are analogous to the gaussian property in the continuous domain (but the binomial filter is different than a discretization of a





$b_2 = [1, 2, 1]$



B2[n]

The simplest approximation to the Gaussian filter is the 3-tap kernel:



B2[n] versus the 3-tap box filter

[1 2 1]

-10

[1 1 1]



Which one is better?



 $[1, 1, 1] \cap [..., 1, -1, 1, -1, 1, -1, ...] = [..., -1, 1, -1, 1, -1, 1, ...]$ $[1, 2, 1] \cup [..., 1, -1, 1, -1, 1, -1, ...] = [..., 0, 0, 0, 0, 0, 0, ...]$

B2[n]



B2[n]

$b_{2,2} = b_{2,0} \circ b_{0,2} = \begin{bmatrix} 1 & 2 & 1 \end{bmatrix} \circ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$



What about the opposite of blurring?





















Hybrid Images

Oliva & Schyns







Hybrid Images













Hybrid Images





















http://cvcl.mit.edu/hybrid_gallery/gallery.html



DR(MADRS)

High pass-filters

Finding edges in the image



Edge strength

Edge orientation:

Edge normal:

Image gradient:

$$\nabla \mathbf{I} = \left(\frac{\partial \mathbf{I}}{\partial x}, \frac{\partial \mathbf{I}}{\partial y}\right)$$

Approximation image derivative:

$$\frac{\partial \mathbf{I}}{\partial x} \simeq \mathbf{I}(x, y) - \mathbf{I}(x - 1, y)$$

 $E(x,y) = |\nabla \mathbf{I}(x,y)|$

 $\theta(x, y) = \angle \nabla \mathbf{I} = \arctan \frac{\partial \mathbf{I} / \partial y}{\partial \mathbf{I} / \partial x}$ $\mathbf{n} = \frac{\nabla \mathbf{I}}{|\nabla \mathbf{I}|}$



[-1, 1]

h[m,n]



g[m,n]

$\begin{bmatrix} -1 & 1 \end{bmatrix}$ $\frac{\partial \mathbf{I}}{\partial x} \simeq \mathbf{I}(x, y) - \mathbf{I}(x - 1, y)$



f[m,n]





g[m,n]

[-1, 1]⊤

=

[-1 1]^T



f[m,n]





Back to the image





Reconstruction from 2D derivatives

In 2D, we have multiple derivatives (along *n* and *m*)



and we compute the pseudo-inverse of the full matrix.



Reconstruction from 2D derivatives







Editing the edge image





Thresholding edges











2D derivatives

There are several ways in which 2D derivatives can be approximated.



Robert-Cross operator:

And many more...

$$\begin{bmatrix} 1 - 1 \end{bmatrix}$$





Issues with image derivatives

Derivatives are sensitive to noise lacksquare

some regions (e.g., object boundaries, ...)

• If we consider continuous image derivatives, they might not be define in



Derivatives

We want to compute the image derivative: $\frac{\partial f(x,y)}{\partial x}$ If there is noise, we might want to "smooth" it with a blurring filter $\frac{\partial f(x,y)}{\partial x} \circ g(x,y)$

But derivatives and convolutions are linear and we can move them around:

$$\frac{\partial f(x,y)}{\partial x} \circ g(x,y)$$

 $= f(x, y) \circ \frac{\partial g(x, y)}{\partial x}$



Gaussian derivatives $g(x, y; \sigma) = \frac{1}{2\pi\sigma^2} \exp{-\frac{x^2 + y^2}{2\sigma^2}}$

The continuous derivative is:

- $g_x(x,y;\sigma) = \frac{\partial g(x,y;\sigma)}{\partial x} =$

 - $= \frac{-x}{\sigma^2} g(x, y; \sigma)$









Gaussian Scale







Derivatives of Gaussians: Scale





Orientation

$$g_x(x,y) = \frac{\partial g(x,y)}{\partial x} = \frac{-x}{2\pi\sigma^4} e^{-\frac{x^2 + y^2}{2\sigma^2}}$$

$$g_{y}(x,y) = \frac{\partial g(x,y)}{\partial y} = \frac{-y}{2\pi\sigma^{4}}e^{-\frac{x^{2}}{2\sigma^{4}}}$$











Orientation

$$g_x(x,y) = \frac{\partial g(x,y)}{\partial x} = \frac{-x}{2\pi\sigma}$$





What about other orientations not axis aligned?



 $g_{y}(x,y) = \frac{\partial g(x,y)}{\partial y} = \frac{-y}{2\pi\sigma^{4}}e^{-\frac{x^{2}+y^{2}}{2\sigma^{2}}}$








Orientation

$$g_x(x,y) = \frac{\partial g(x,y)}{\partial x} = \frac{-x}{2\pi\sigma^4} e^{-\frac{x^2+y^2}{2\sigma^2}} \int_{0}^{1} \int_{0}^{1}$$

The smoothed directional gradient is a linear combination of two kernels

$$u^T \nabla g \otimes I = (\cos(\alpha)g_x(x,y) + \sin(\alpha)g_y(x,y)) \otimes I(x,y) =$$

Any orientation can be computed as a linear combination of two filtered images

$$= \cos(\alpha)g_x(x,y) \otimes I(x,y) + \sin(\alpha)g_y(x,y) \otimes I(x,y)$$

Steereability of gaussian derivatives, Freeman & Adelson 92



Orientation



 $+\sin(\alpha)$

$\cos(\alpha)$



Steereability of gaussian derivatives, Freeman & Adelson 92



Discretization Gaussian derivatives

There are many discrete approximations. For instance, we can take samples of the continuous functions. In practice it is common to use the discrete approximation given by the binomial filters.

Convolving the binomial coefficients with [1, -1]



111

Discretization 2D Gaussian derivatives

and then convolve them.

One example is the Sobel-Feldman operator:

$$Sobel_x = \begin{bmatrix} 1 & 0 \end{bmatrix} -$$

$$Sobel_{y} = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

As Gaussians are separable, we can approximate two 1D derivatives

$$\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$





$$g_{x^{n},y^{m}}(x,y;\sigma) = \frac{\partial^{n+m}g(x,y)}{\partial x^{n}\partial y^{m}} = \left(\frac{-\sigma}{\sigma}\right)^{n+m}g(x,y)$$

 $\frac{-1}{\sigma\sqrt{2}}\right)^{n+m} H_n\left(\frac{x}{\sigma\sqrt{2}}\right) H_m\left(\frac{y}{\sigma\sqrt{2}}\right) g(x,y;\sigma)$



g

$$g_{x^{n},y^{m}}(x,y;\sigma) = \frac{\partial^{n+m}g(x,y)}{\partial x^{n}\partial y^{m}} = \left(\frac{\partial^{n+m}g(x,y)}{\partial x^{n}\partial y^{m}}\right)$$



 $\frac{-1}{\sigma\sqrt{2}}\right)^{n+m}H_n\left(\frac{x}{\sigma\sqrt{2}}\right)H_m\left(\frac{y}{\sigma\sqrt{2}}\right)g(x,y;\sigma)$



114





 $g_{x^{n},y^{m}}(x,y;\sigma) = \frac{\partial^{n+m}g(x,y)}{\partial x^{n}\partial y^{m}} = \left(\frac{-1}{\sigma\sqrt{2}}\right)^{n+m} H_{n}\left(\frac{x}{\sigma\sqrt{2}}\right) H_{m}\left(\frac{y}{\sigma\sqrt{2}}\right)g(x,y;\sigma)$





 $g_{x^{n},y^{m}}(x,y;\sigma) = \frac{\partial^{n+m}g(x,y)}{\partial x^{n}\partial y^{m}} = \left(\frac{-1}{\sigma\sqrt{2}}\right)^{n+m} H_{n}\left(\frac{x}{\sigma\sqrt{2}}\right) H_{m}\left(\frac{y}{\sigma\sqrt{2}}\right)g(x,y;\sigma)$







117

Image sharpening filter





Image sharpening filter

Subtract away the blurred components of the image:



This filter has an overall DC component of 1. It de-emphasizes the blur component of the image (low spatial frequencies).

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} - \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$



Input image



Sharpened









Input image



Sharpened

Input image









