MIT CSAIL

6.869: Advances in Computer Vision

MIT COMPUTER VISION

Lecture 4

Statistical Image Models

What are we tuned to?

The visual system is tuned to process structures typically found in the world.

The visual system seems to be tuned to a set of images:

Demo inspired from D. Field

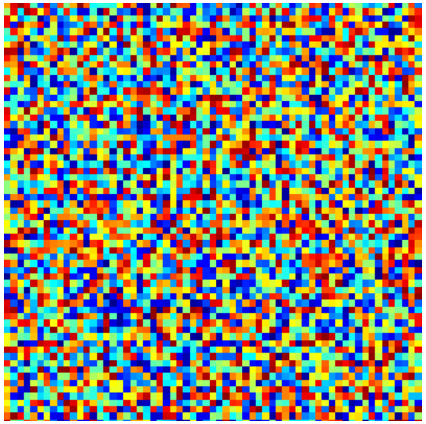
Remember these images

Did you saw this image?

Remember these images

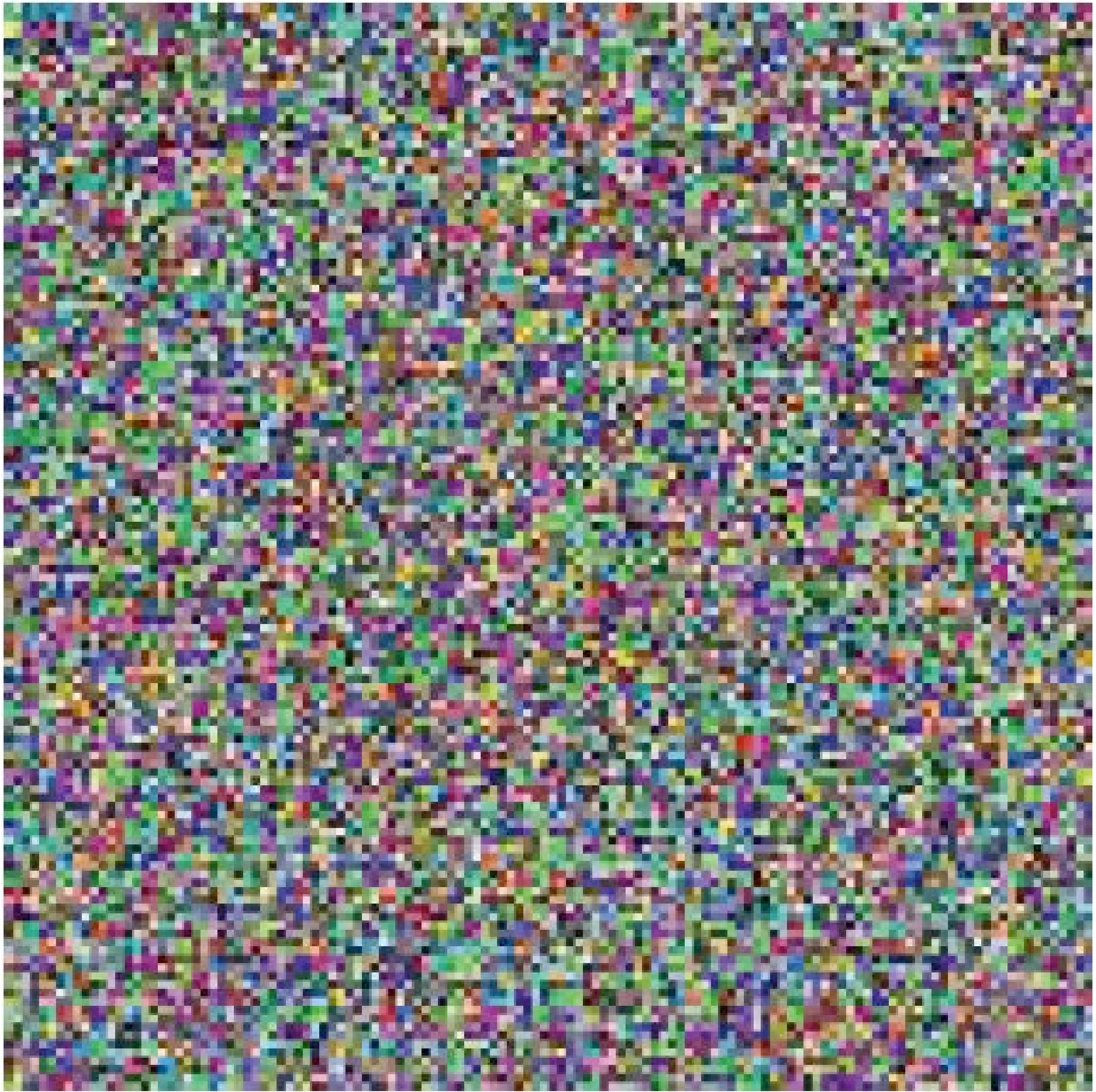
Test 2

Did you saw this image?

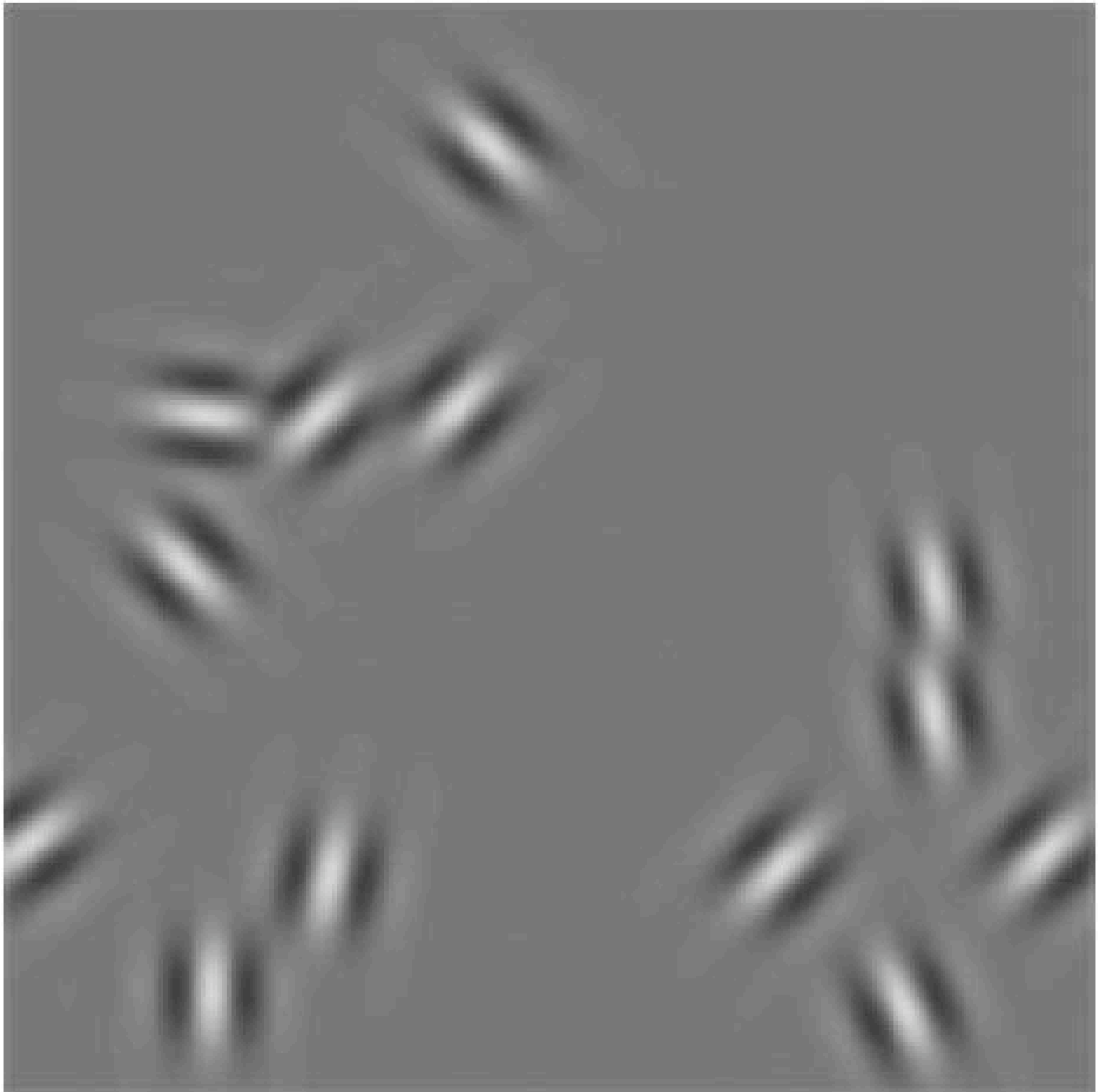


Visual Worlds

Visual Worlds



Visual Worlds



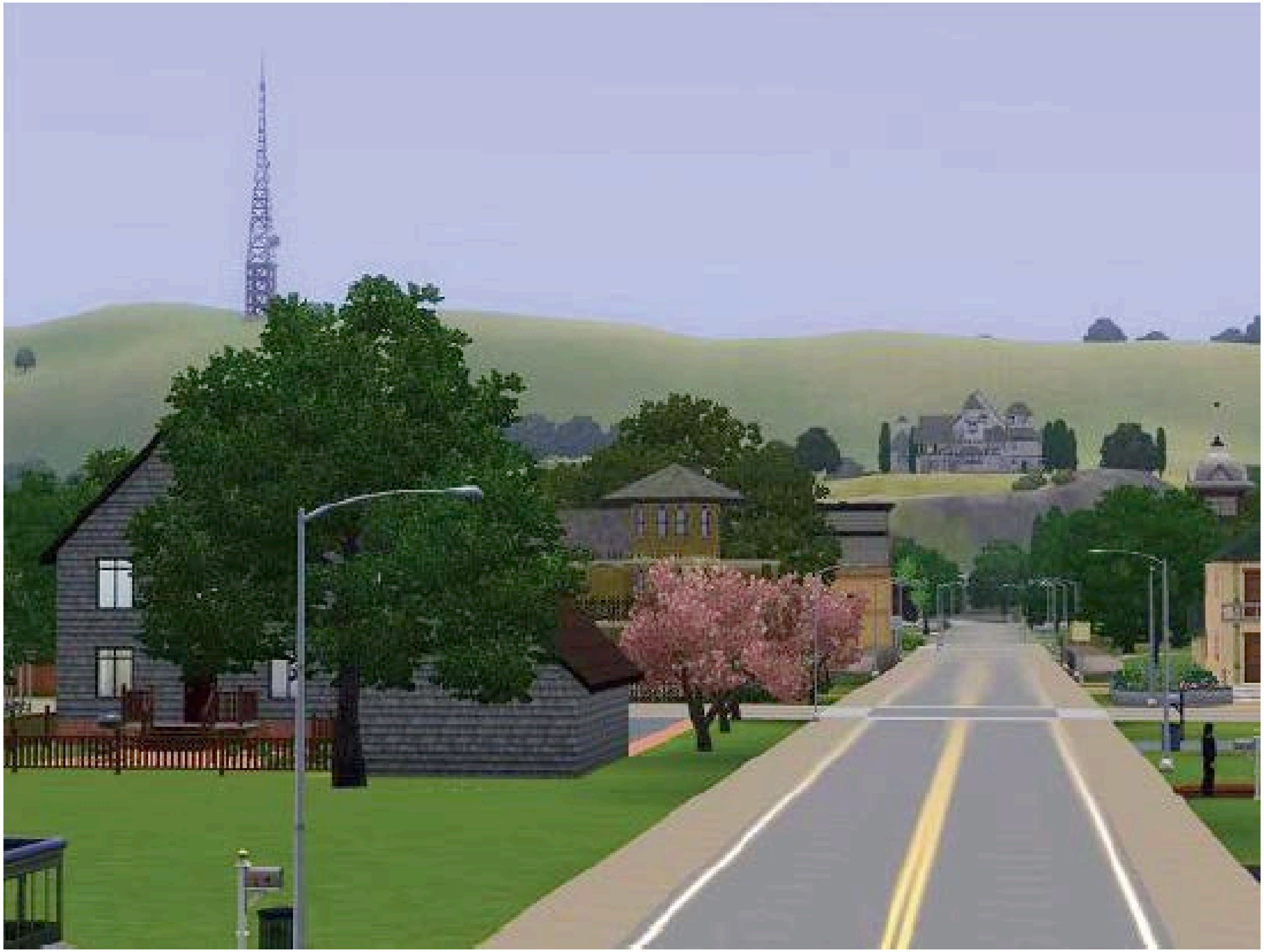
Visual Worlds

Visual Worlds

Visual Worlds

Visual Worlds

Visual Worlds



Visual Worlds

Separating images into components

=

X

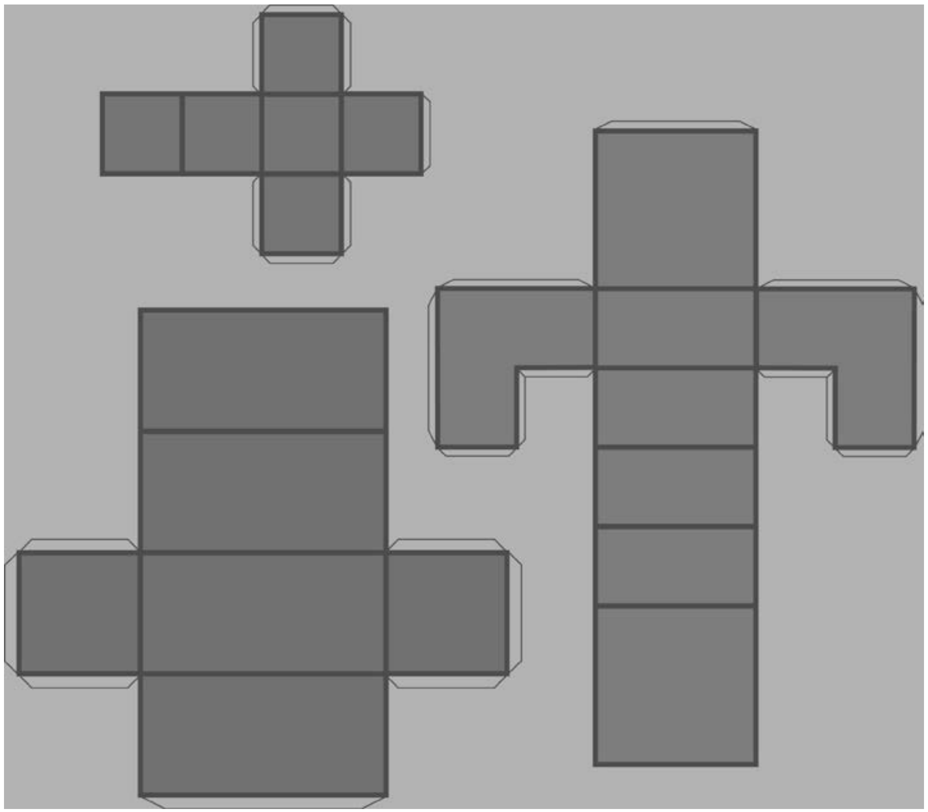
Separating images into components

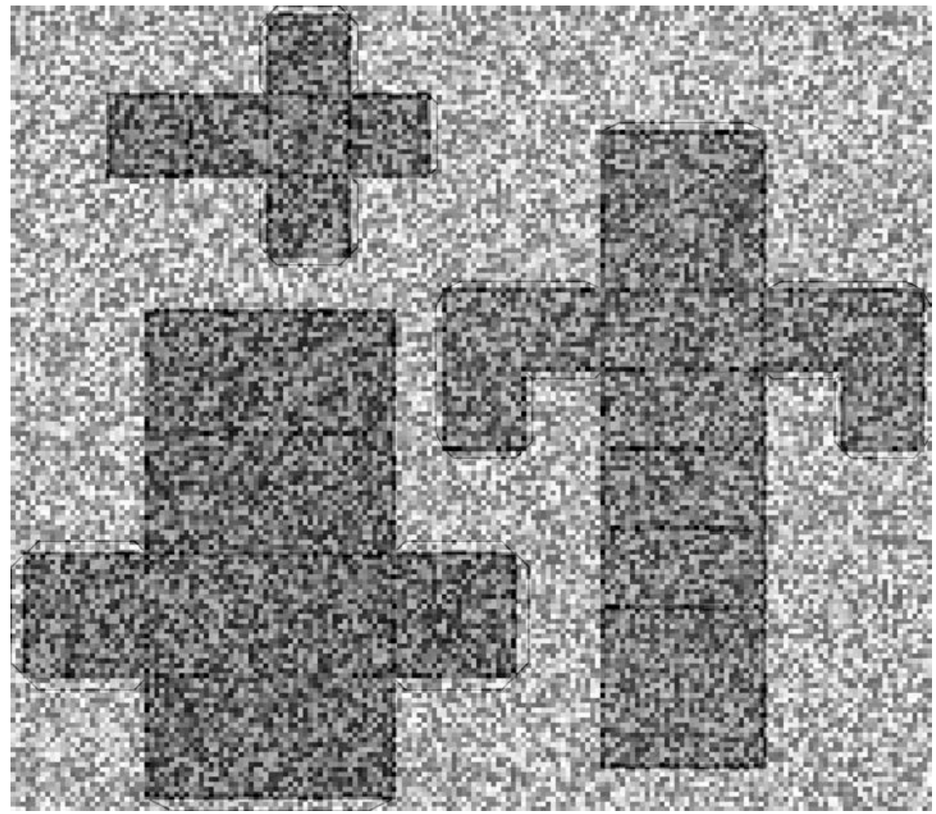
Separating images into components

=

+

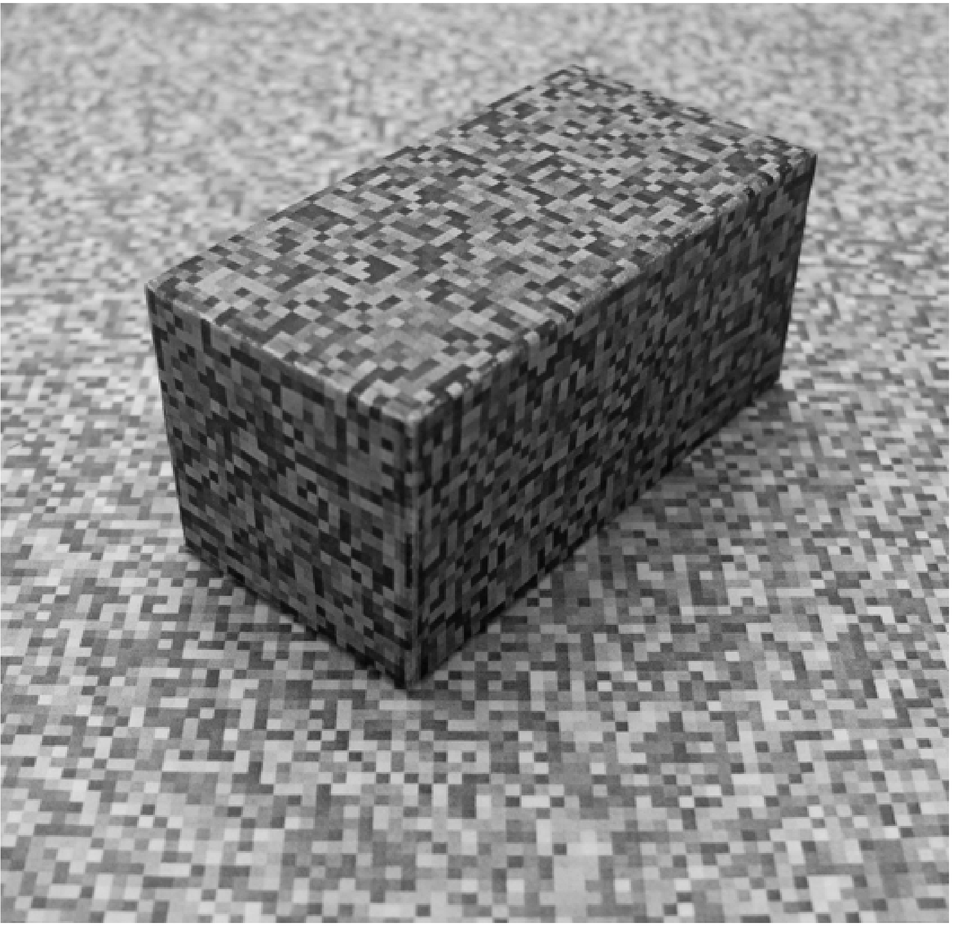
Noise on the image
vs.
noise in the world

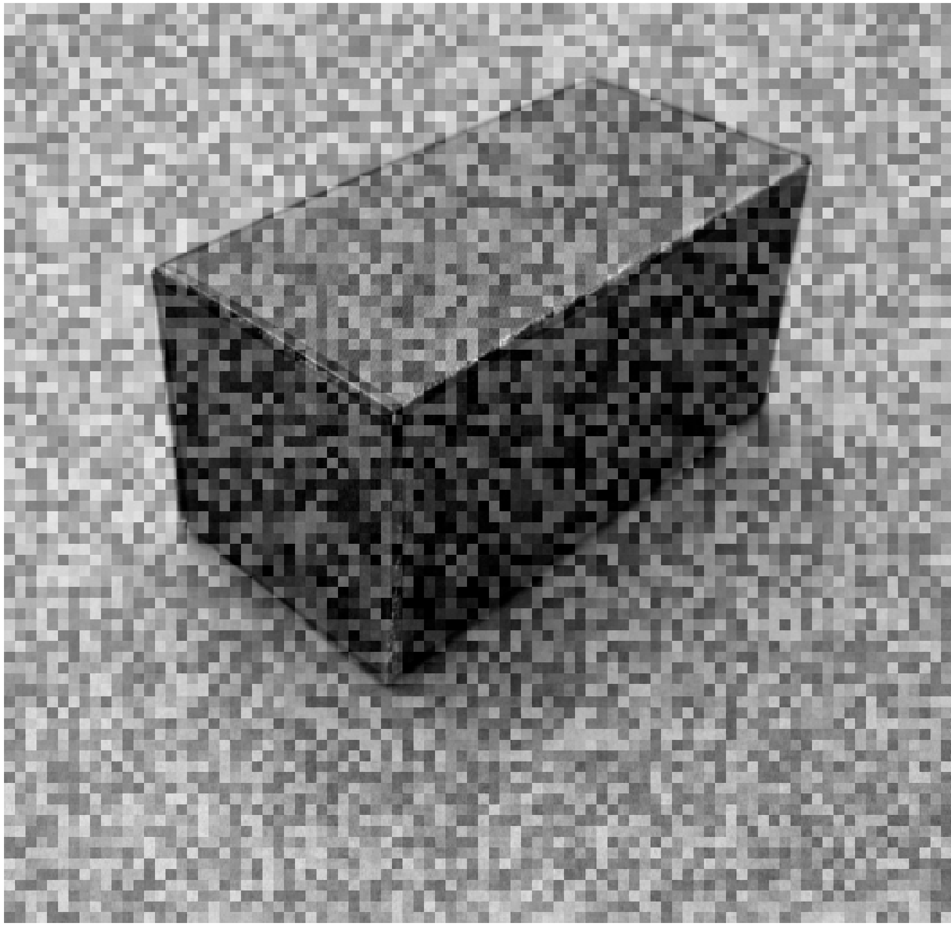




The *noise in the world*, it is called *texture* by its friends

Noise or texture?





Separating images into components

=

-

Separating images into components

=

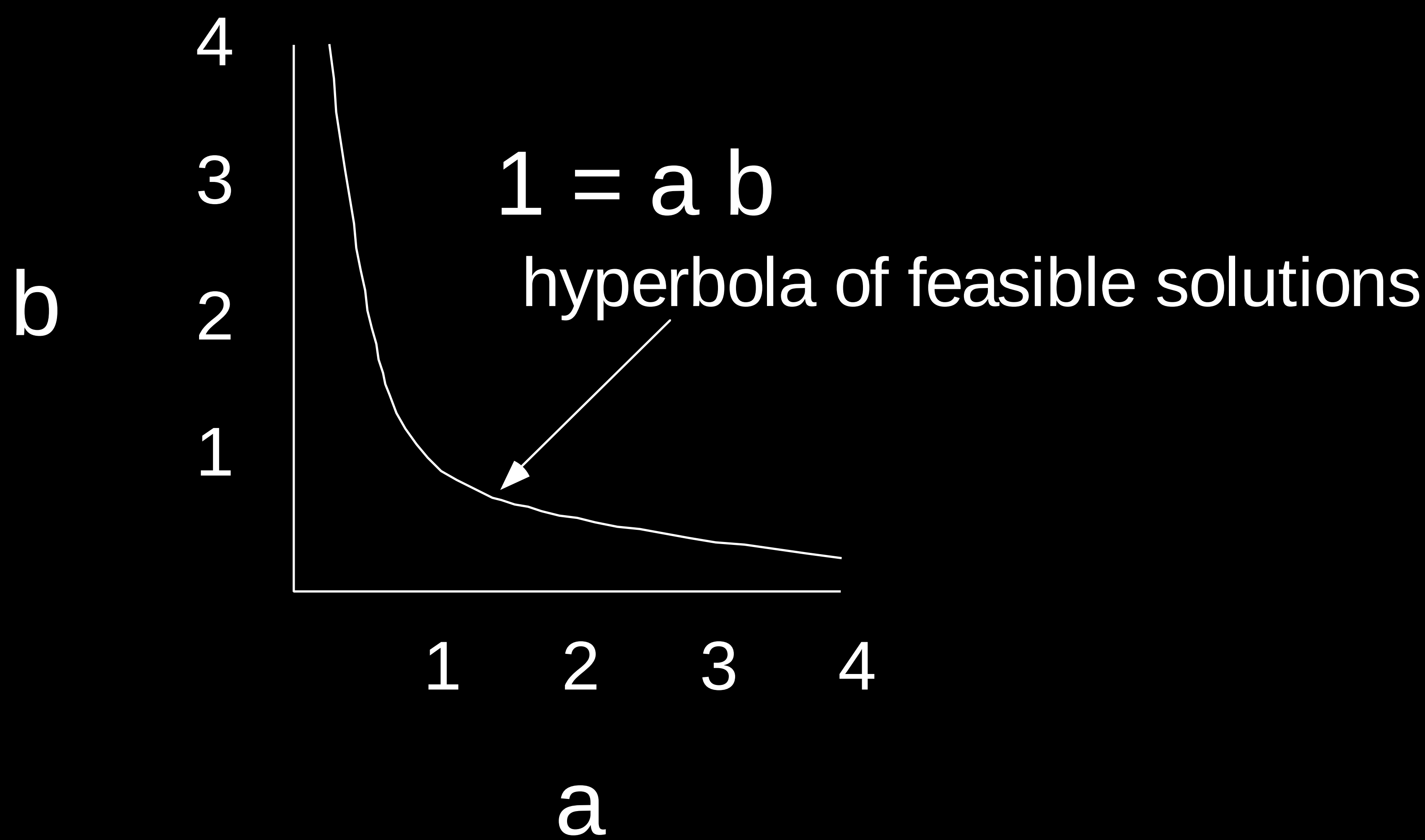

7

+

Prototypical vision problem

- Observe some product of two numbers, say 1.0
- What were those two numbers?
- Ie, 1 = ab. Find a and b.

- Compare this with the prototypical graphics problem: here are two numbers; what is their product?


1 = a b
hyperbola of feasible solutions
b
4
3
2
1
1
2
3
4
a

Bayesian approach

Want to calculate: $\max_{a,b} P(a, b \mid y = 1)$

Bayes rule

Use $P(a, b \mid y = 1) = k\, P(y=1 \mid a, b)\, P(a, b)$

Posterior probability

Likelihood function

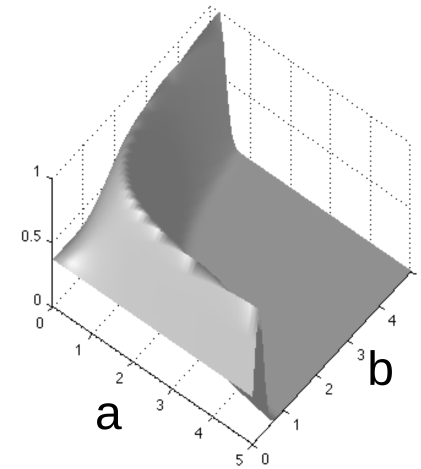
Prior probability

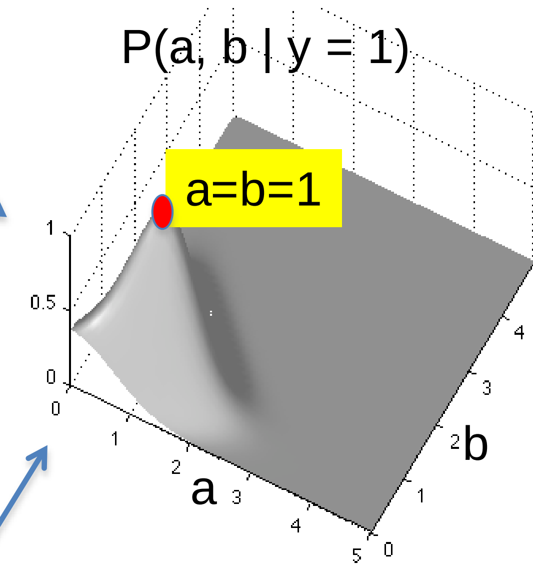
Bayesian approach

Use P(a, b | y = 1) = k P(y=1|a, b) P(a, b)

Likelihood function

$$P(y=1\,|\,a,b) = k e^{-\frac{(1-ab)^2}{2\sigma^2}}$$

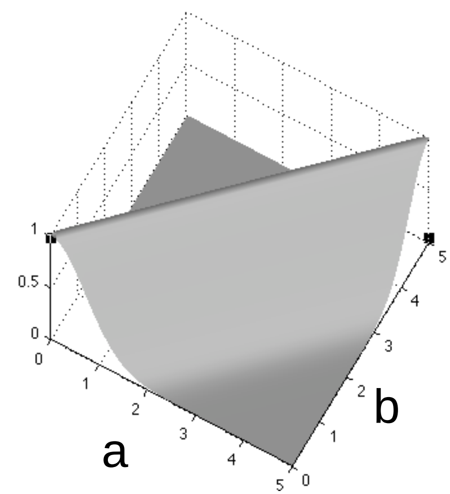




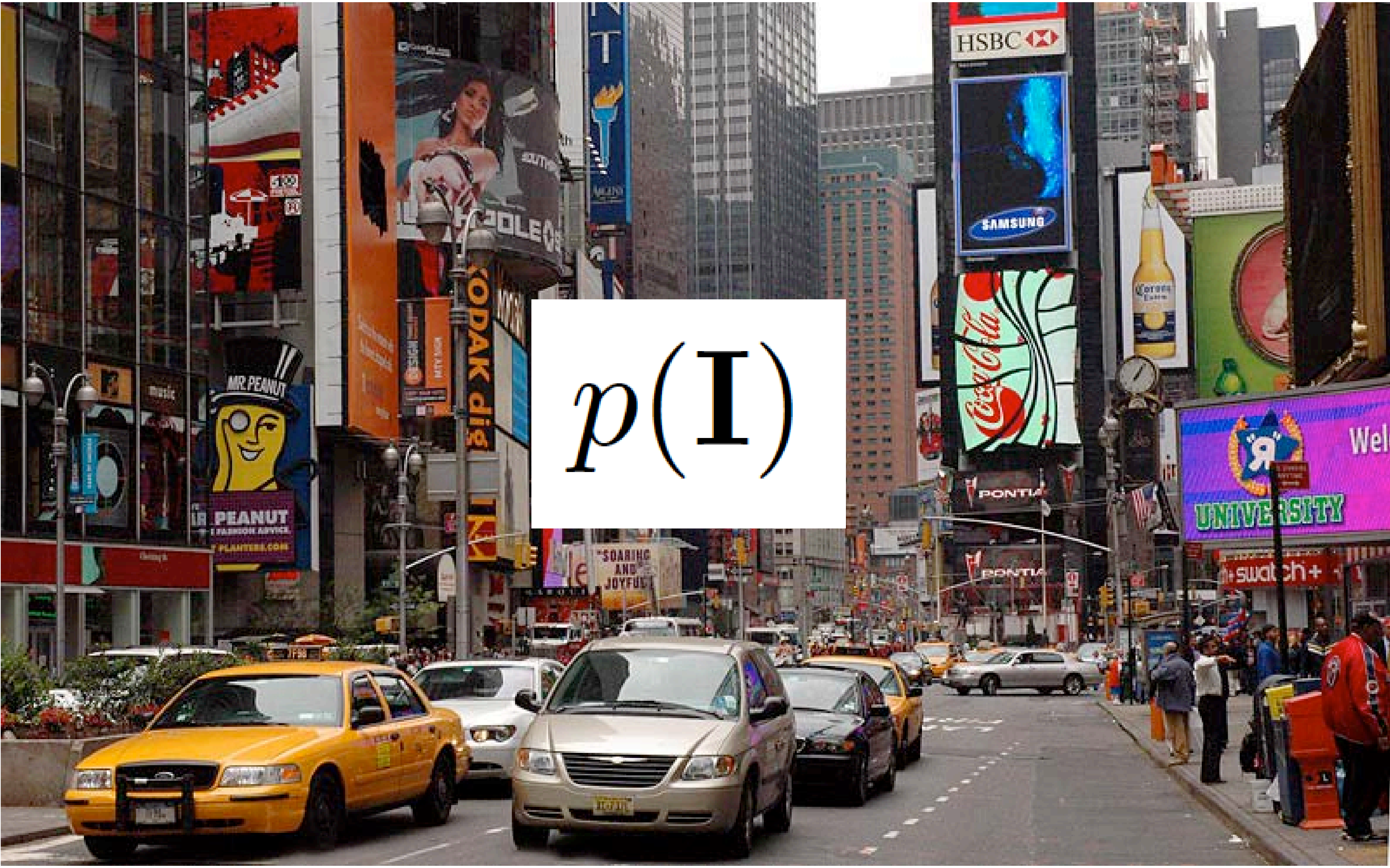
Prior probability

$$P(a,b) = k e^{-\frac{(a-b)^2}{2\sigma^2}}$$ If a>0, b>0

= 0 otherwise



Statistical modeling of images



$$p(\mathbf{I})$$

To appear in: Handbook of Video and Image Processing, 2nd edition
ed. Alan Bovik, ©Academic Press, 2005.

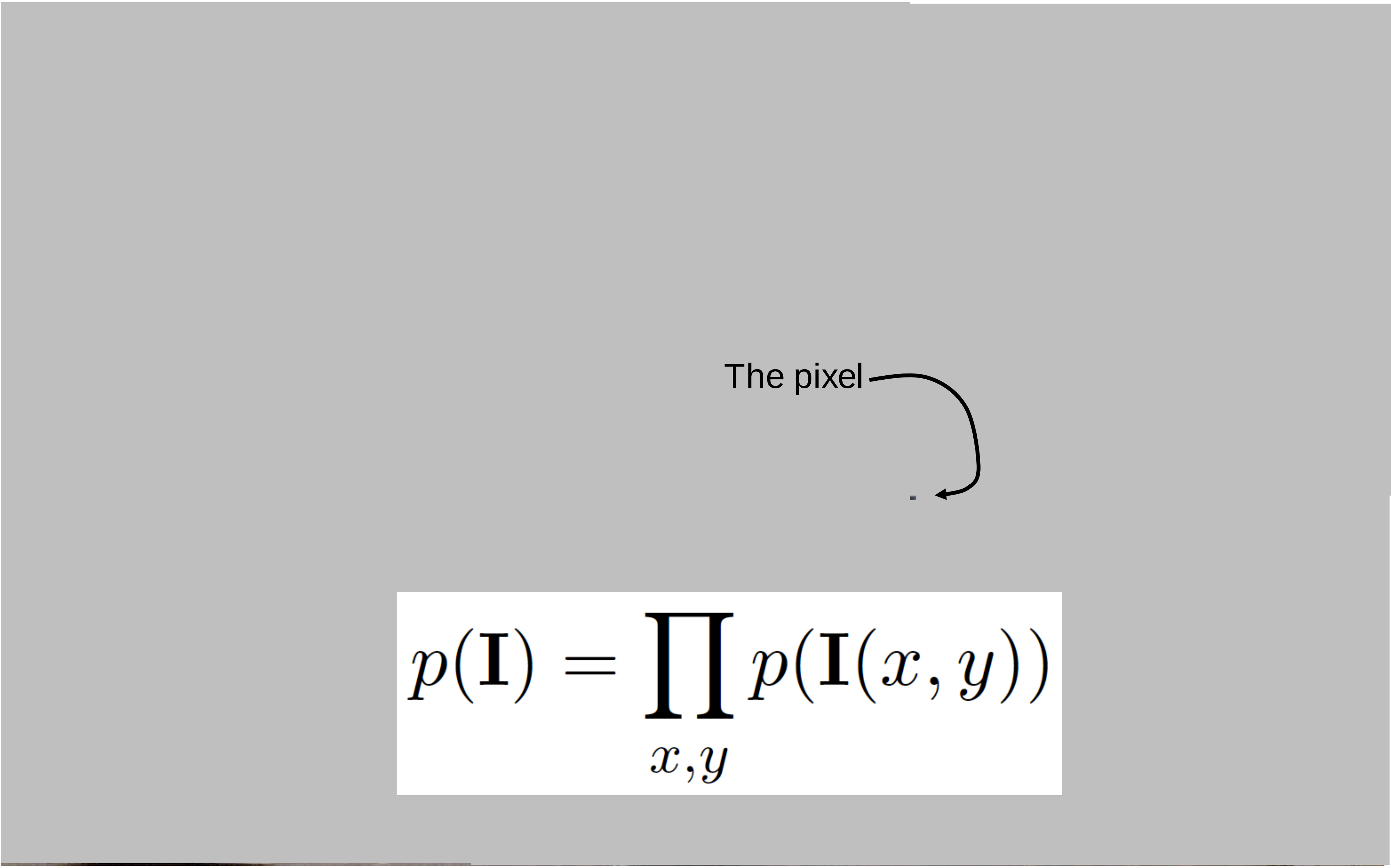
4.7 Statistical Modeling of Photographic Images

Eero P. Simoncelli

New York University

January 18, 2005

Statistical modeling of images



$$p(\mathbf{I}) = \prod_{x,y} p(\mathbf{I}(x,y))$$

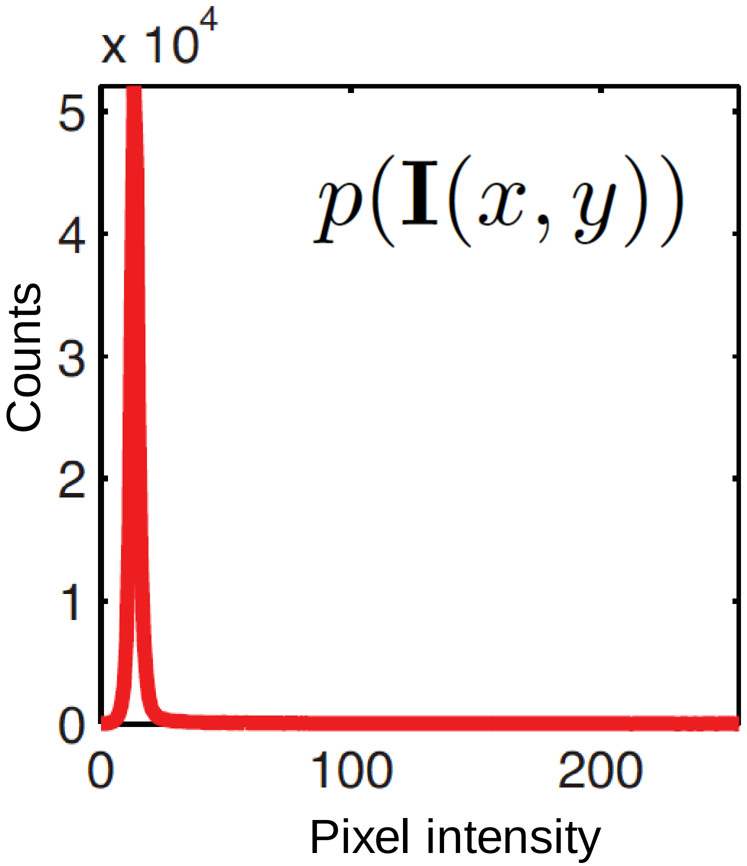
Statistical modeling of images

$$p(\mathbf{I}) = \prod_{x,y} p(\mathbf{I}(x,y))$$

Assumptions:
- Independence: All pixels are independent.
- Stationarity: The distribution of pixel intensities does not depend on image location.

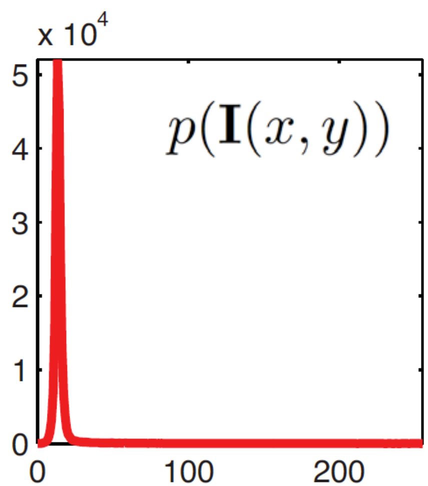
$$p(\mathbf{I}) = \prod_{x,y} p(\mathbf{I}(x,y))$$

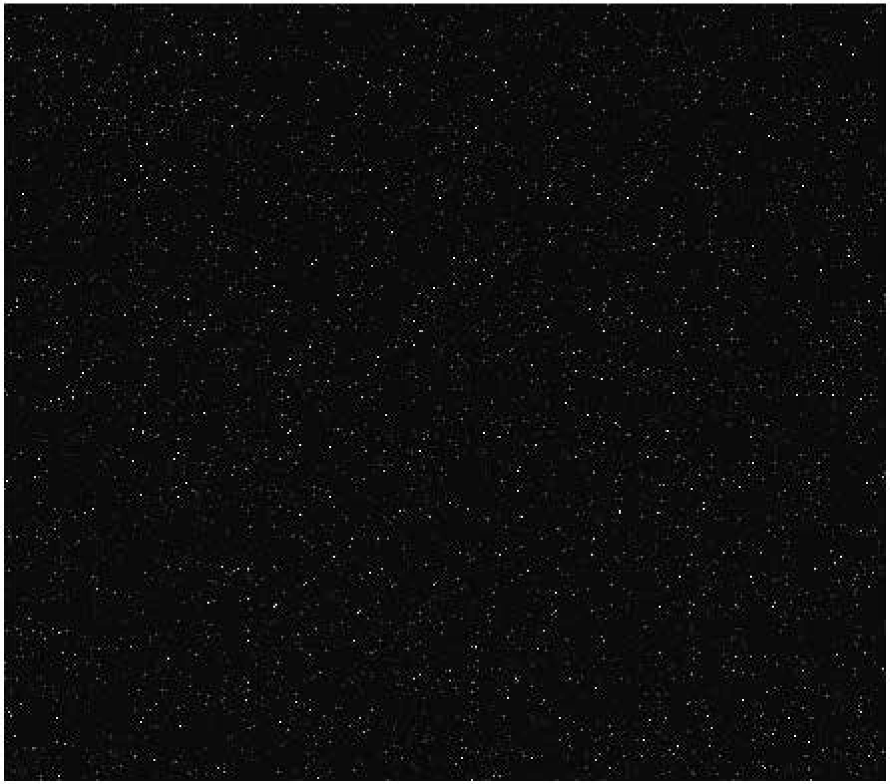
Fitting the model



Sampling new images

$$p(\mathbf{I}) = \prod_{x,y} p(\mathbf{I}(x,y))$$

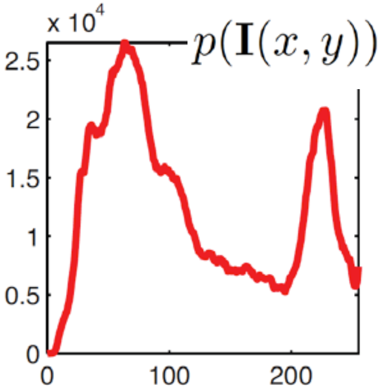


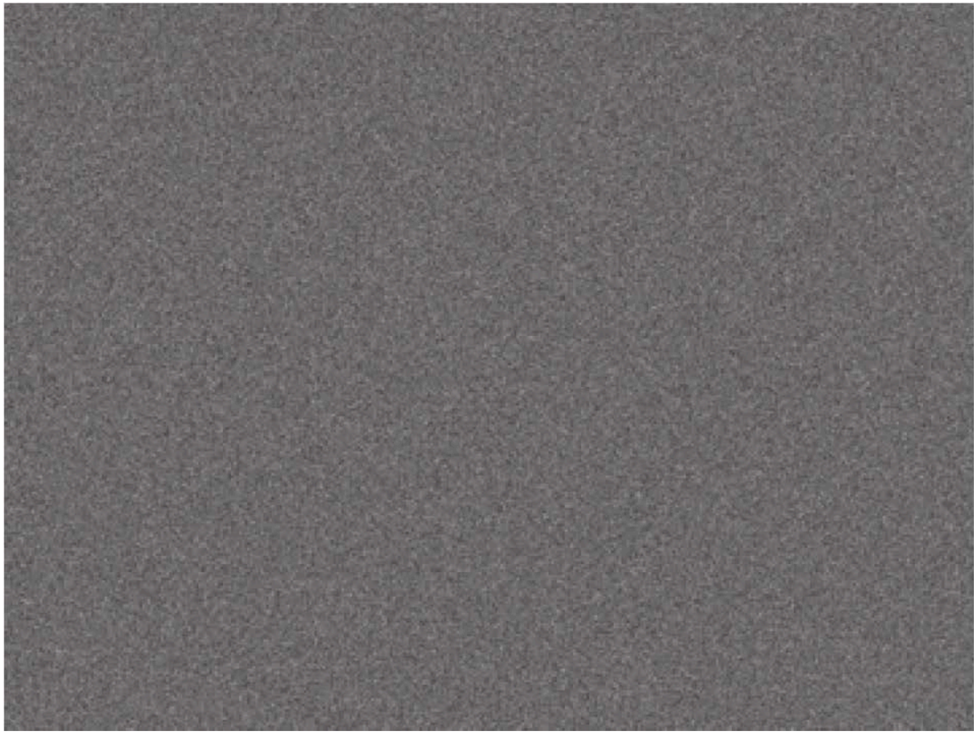


Sample

Sampling new images

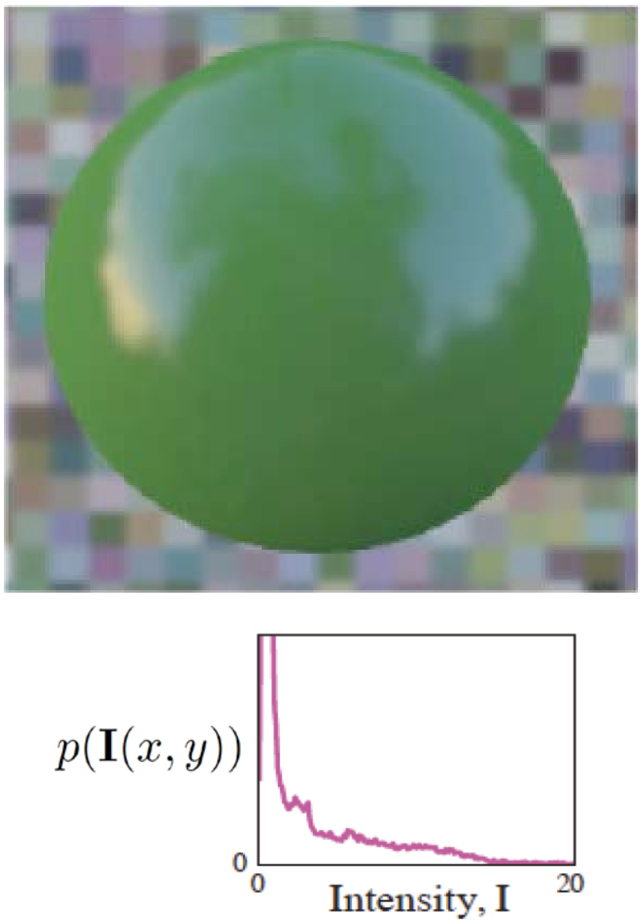
$$p(\mathbf{I}) = \prod_{x,y} p(\mathbf{I}(x,y))$$

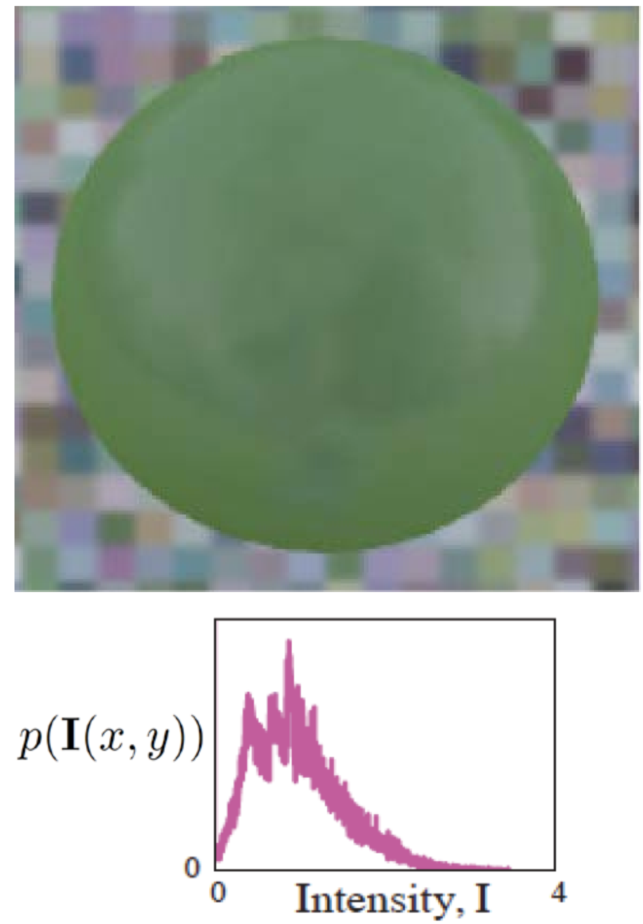




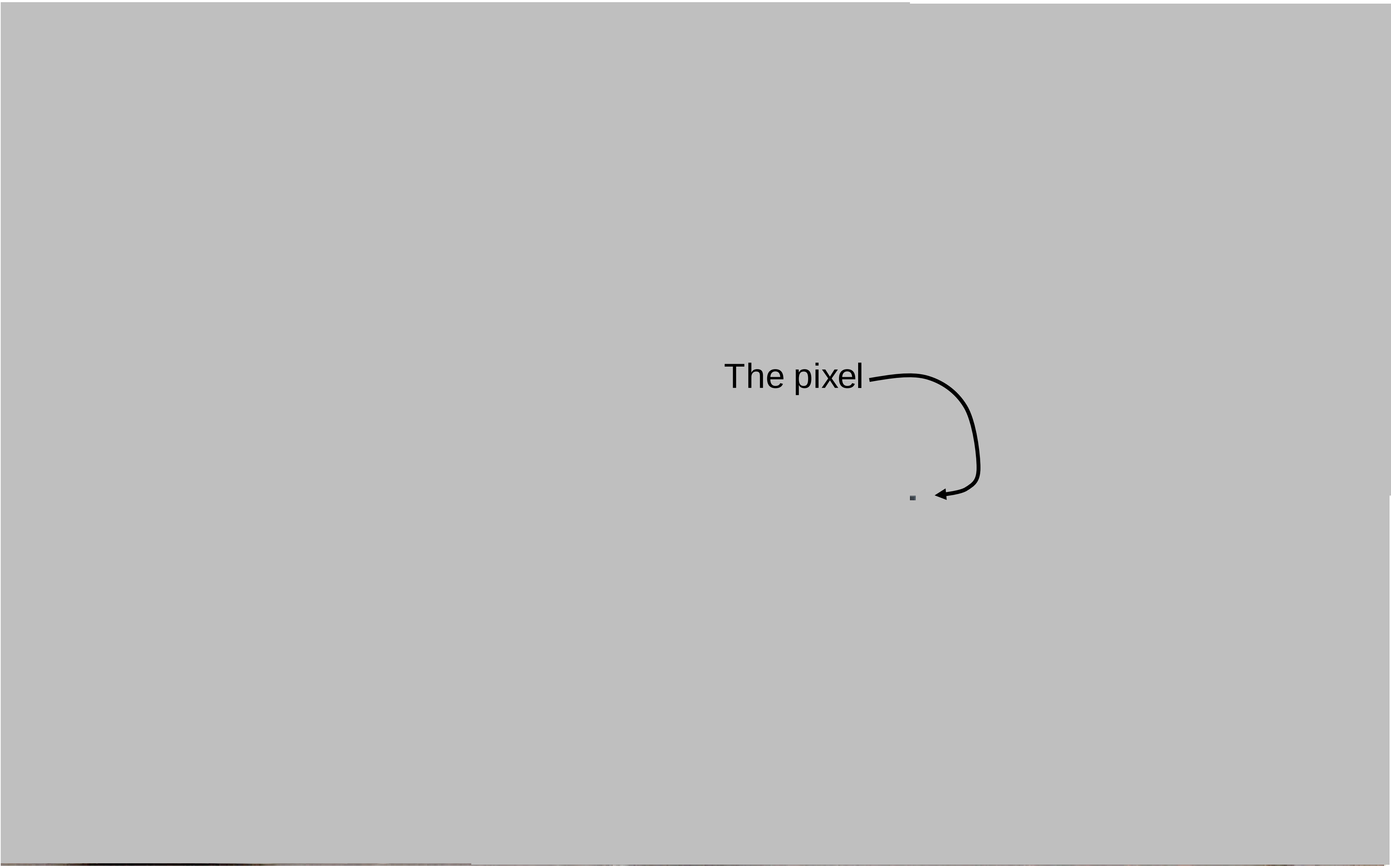
Sample

The importance of distribution of intensities

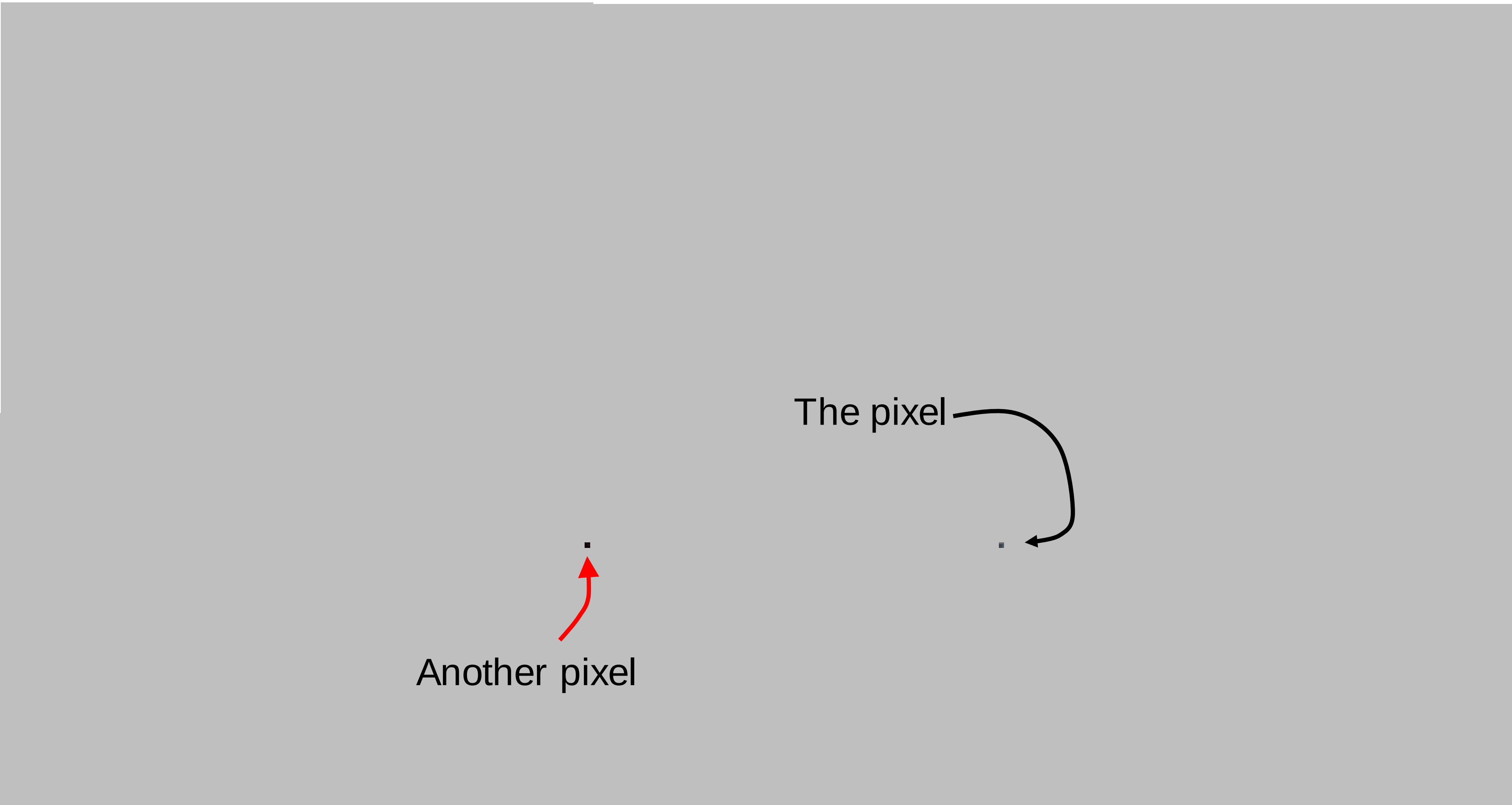




Statistical modeling of images

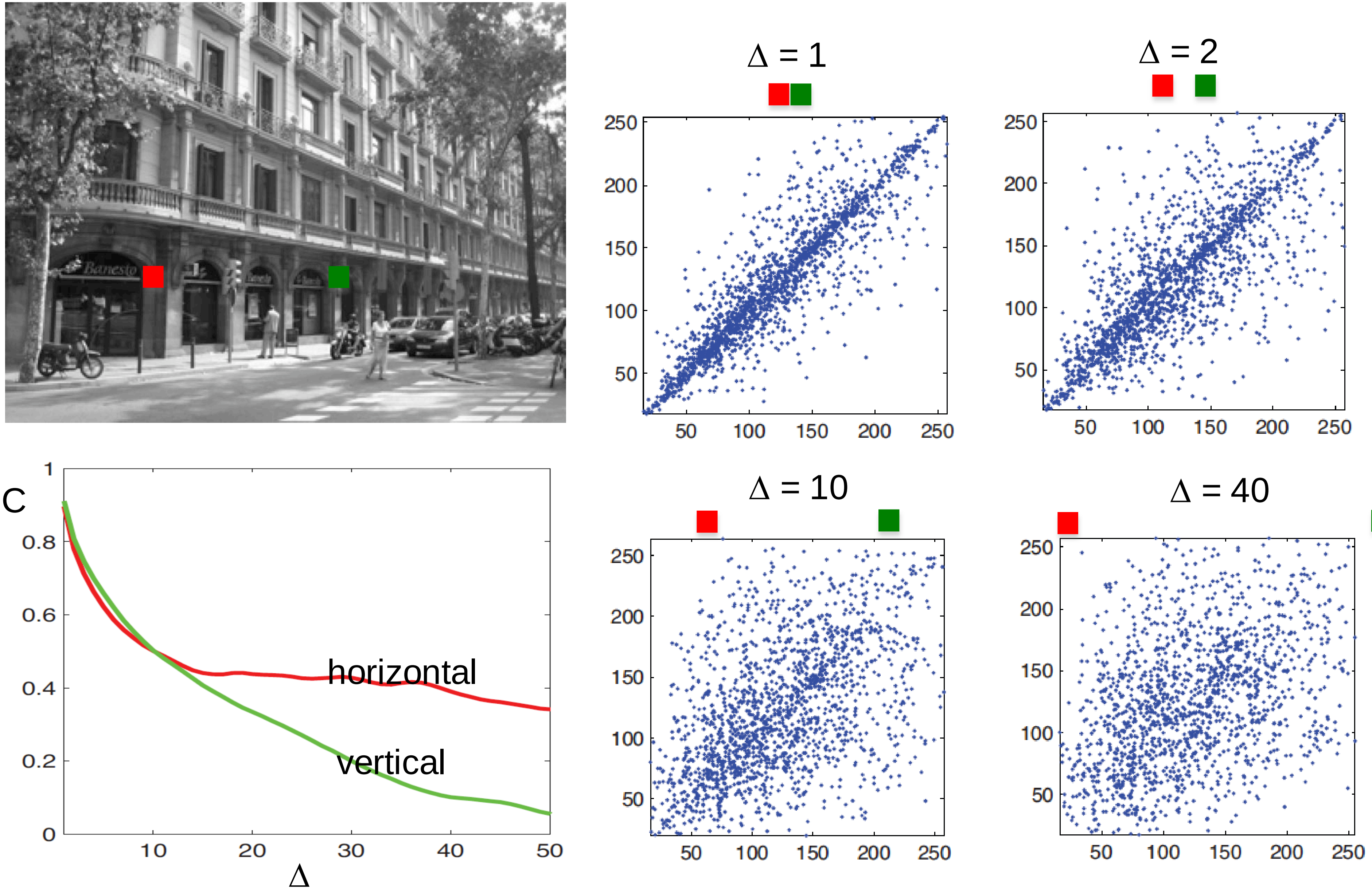


Statistical modeling of images



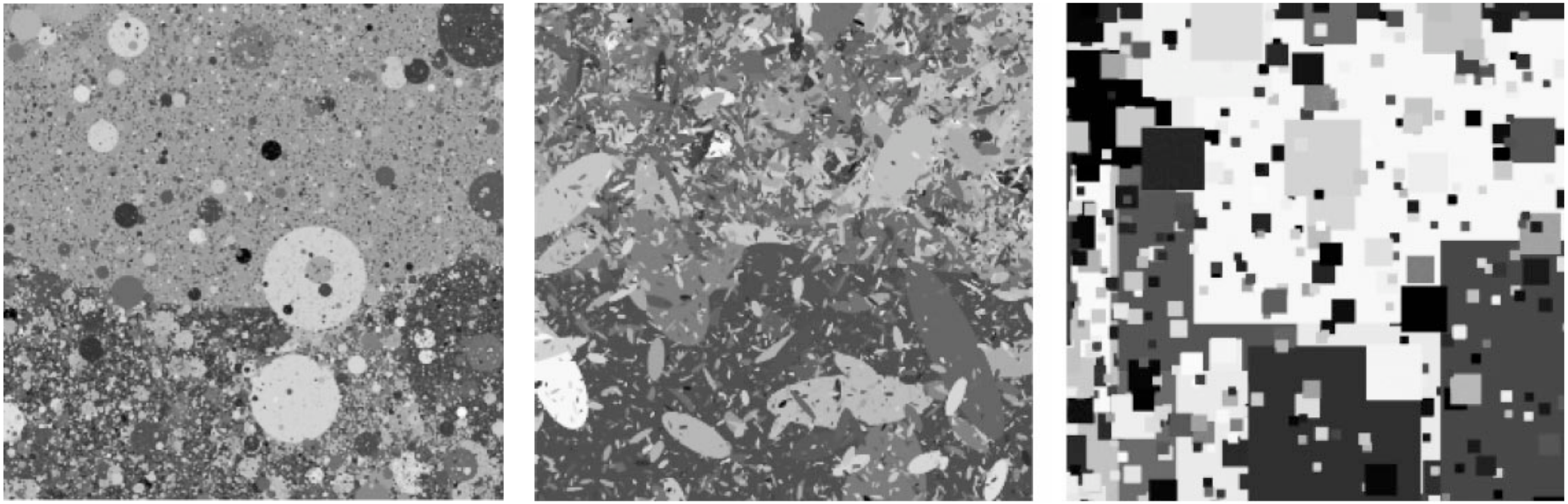
$$C(\Delta x, \Delta y) = \rho\left[\mathbf{I}(x+\Delta x, y+\Delta y), \mathbf{I}(x, y)\right]$$

$$C(\Delta x, \Delta y) = \rho\left[\mathbf{I}(x+\Delta x, y+\Delta y), \mathbf{I}(x,y)\right]$$



Dead leaves models

Introduced in the 60's by Matheron (67) and popularized by Ruderman (97)



From *Lee, Mumford and Huang 2001*

Fourier Characteristics of Images

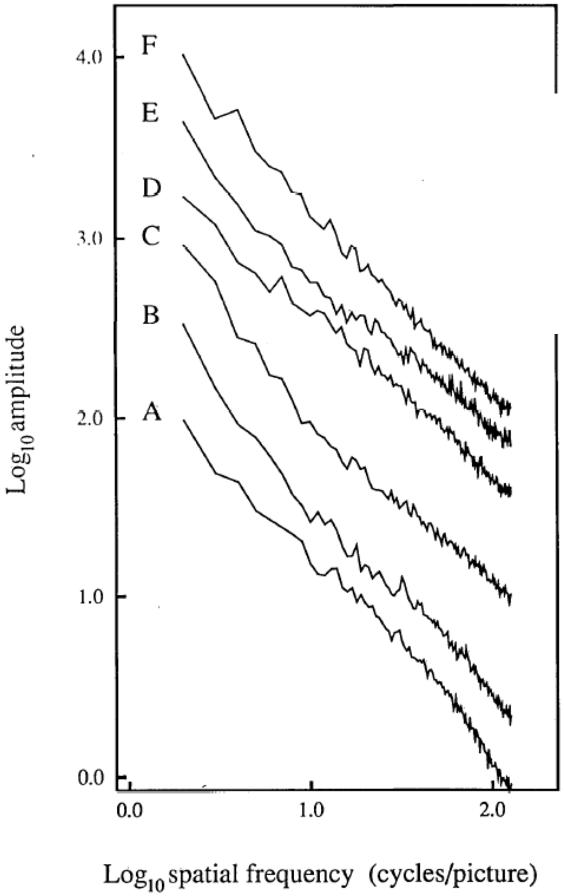
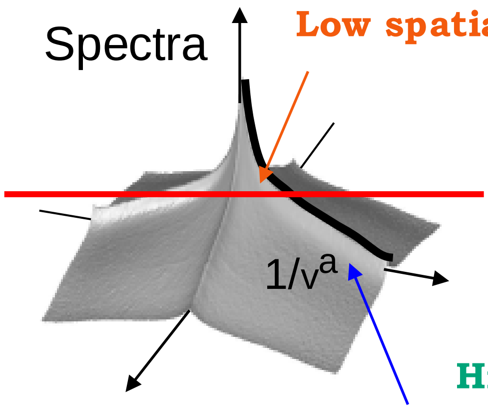
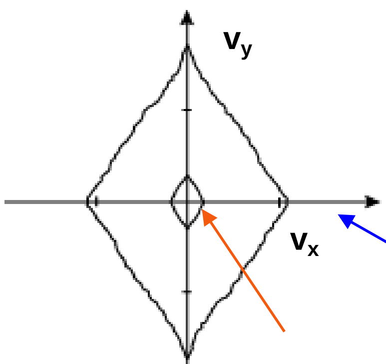


Fig. 8. Amplitude spectra for the six images A–F, averaged across all orientations. The spectra have been shifted up for clarity. On these log–log coordinates the spectra fall off by a factor of roughly $1/f$ (a slope of -1). Therefore the power spectra fall off as $1/f^2$.

Power spectra fall off as

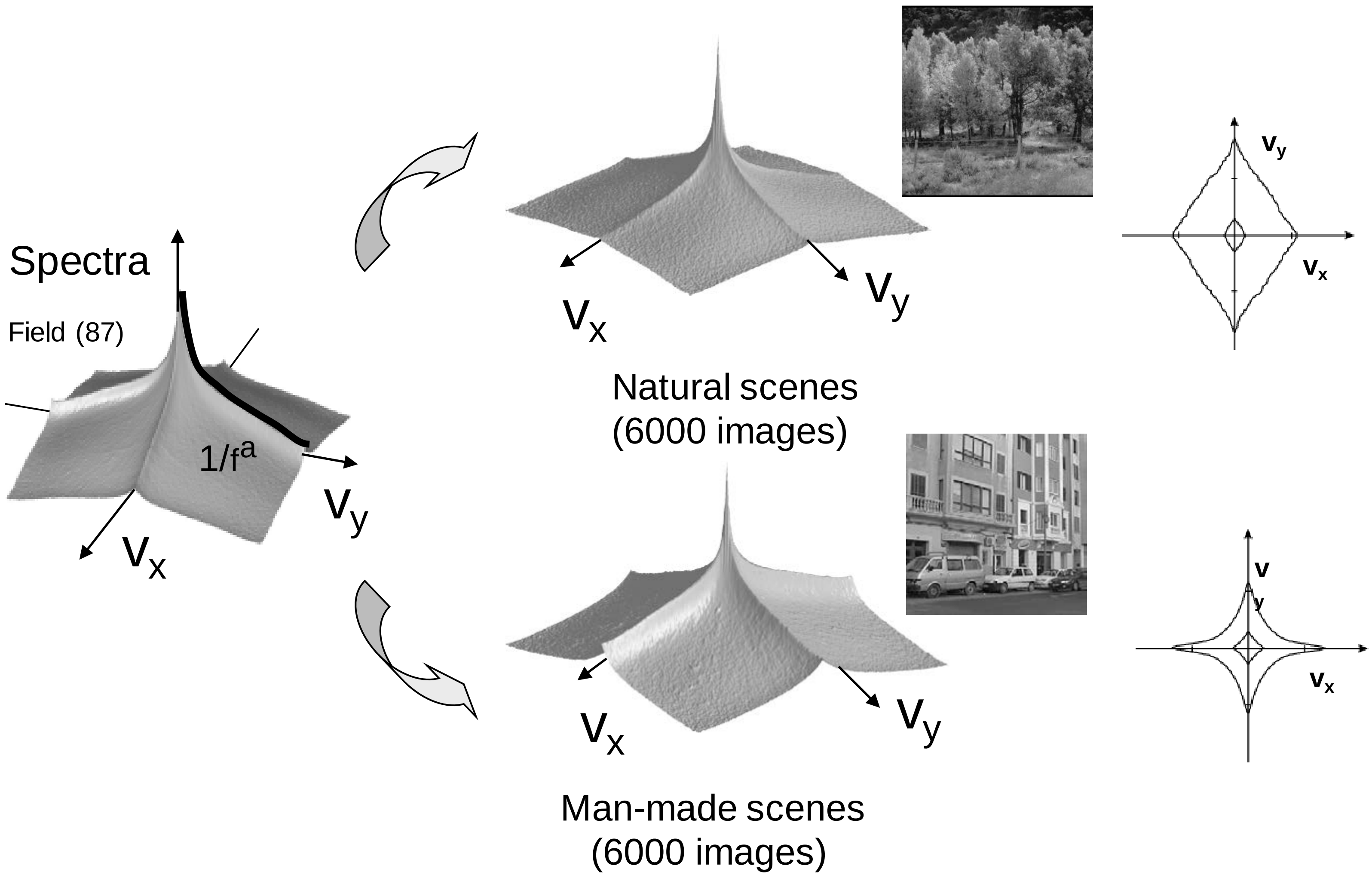
$$|\hat{\mathbf{I}}(v)| \simeq \frac{1}{|v|^{\alpha}}$$





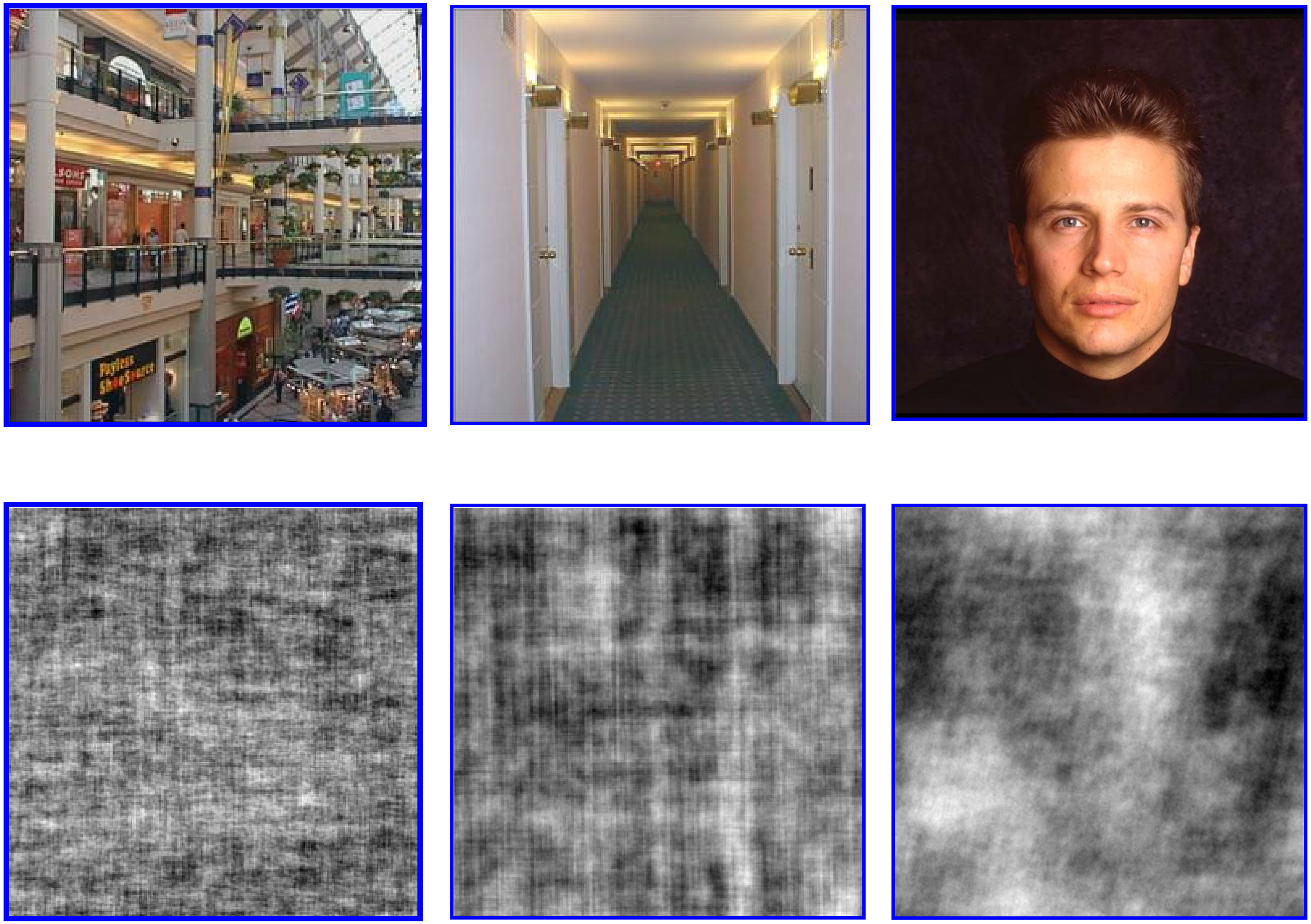
D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," J. Opt. Soc. Am. A **4**, 2379- (1987)

Fourier Characteristics of Images



Torralba and Oliva, *Statistics of Natural Image Categories*. Network: Computation in Neural Systems 14 (2003) 391-412.

Randomizing the phase

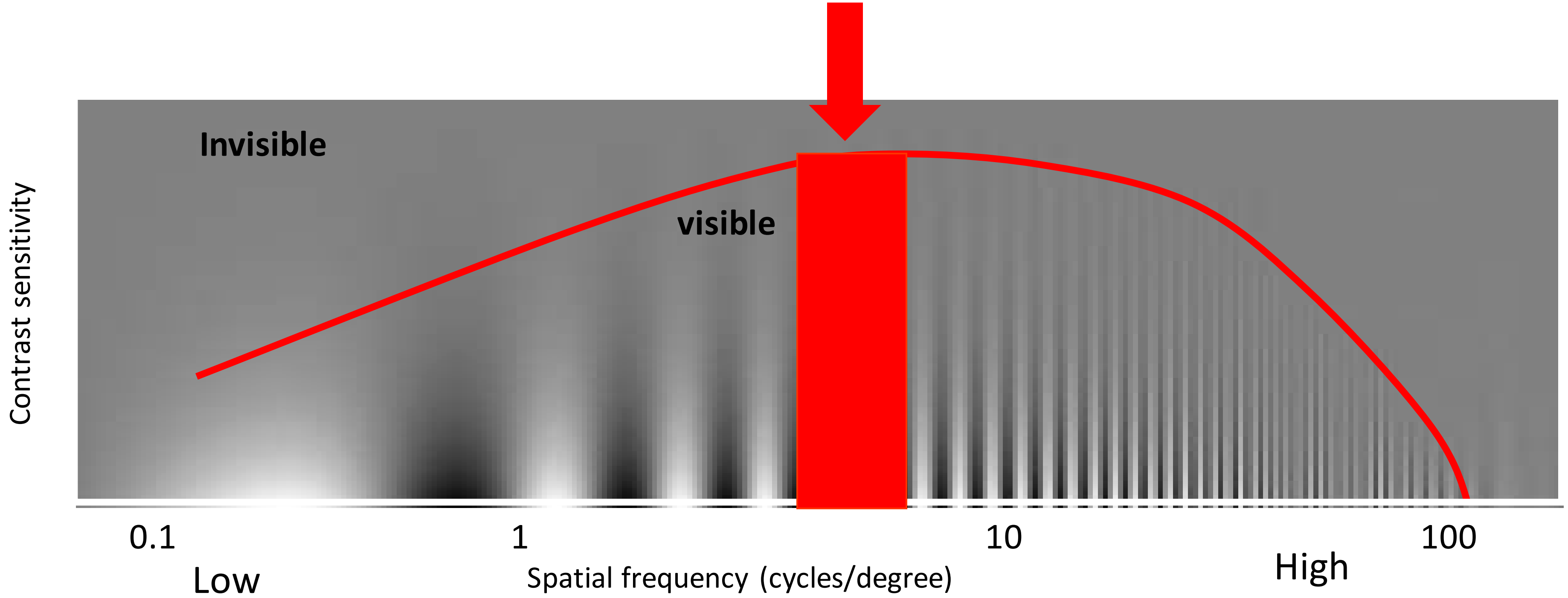


Contrast Sensitivity Function

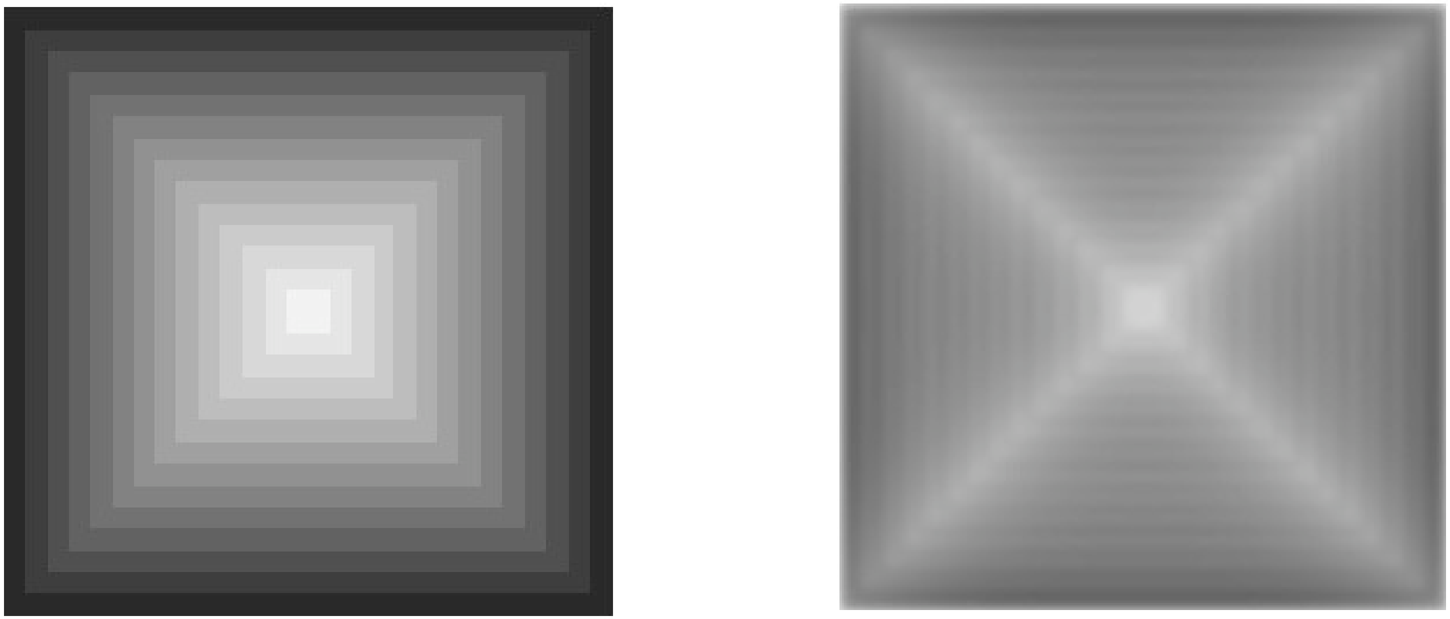
Blackmore & Campbell (1969)

Maximum sensitivity

~ **6** cycles / degree of visual angle



Laplacian



a b

An illusion by Vasarely, left, and a bandpass filtered version, right.

http://web.mit.edu/persci/people/adelson/publications/gazzan.dir/vasarely.html

Gaussian model

We want a distribution that captures the correlation structure typical of natural images.

Let **C** be the covariance matrix of the imag

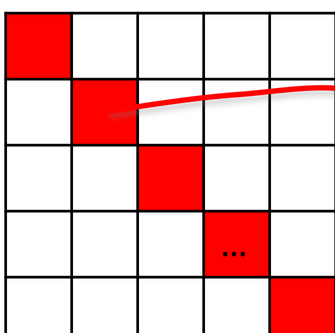
$$p(\mathbf{I}) = \exp\left(-\frac{1}{2}\mathbf{I}^T\mathbf{C}^{-1}\mathbf{I}\right) \qquad C = \begin{bmatrix} c_0 & c_1 & c_2 & & \cdots & c_{n-1} \\ c_{n-1} & c_0 & c_1 & c_2 & & \vdots \\ & c_{n-1} & c_0 & c_1 & \ddots & \\ \vdots & \ddots & \ddots & \ddots & & c_2 \\ & & & & & c_1 \\ c_1 & \cdots & & c_{n-1} & & c_0 \end{bmatrix}$$

Stationarity assumption: Symmetrical circulant matrix

Diagonalization of circulant matrices: $C = EDE^T$

The eigenvectors are the Fourier basis

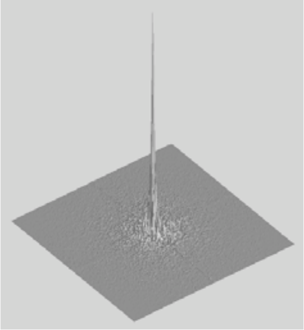
The eigenvalues are the squared magnitude of the Fourier coefficients

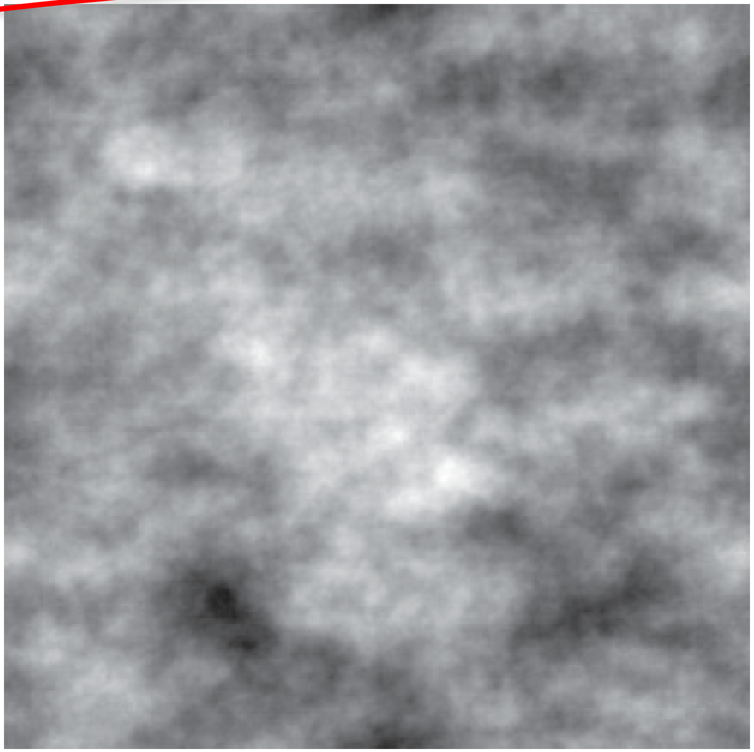
D= 

$$|\hat{\mathbf{I}}(v)|^2 \simeq \frac{1}{|v|^{2\alpha}}$$

Sampling new images

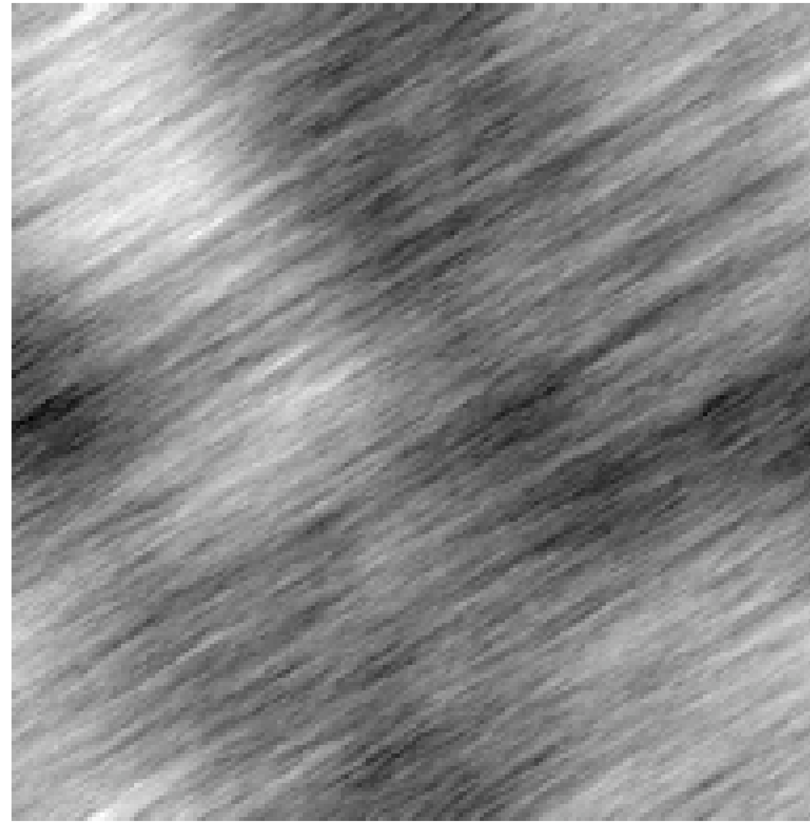
$$p(\mathbf{I}) = \exp\left(-\frac{1}{2}\mathbf{I}^T\mathbf{C}^{-1}\mathbf{I}\right)$$





Sample

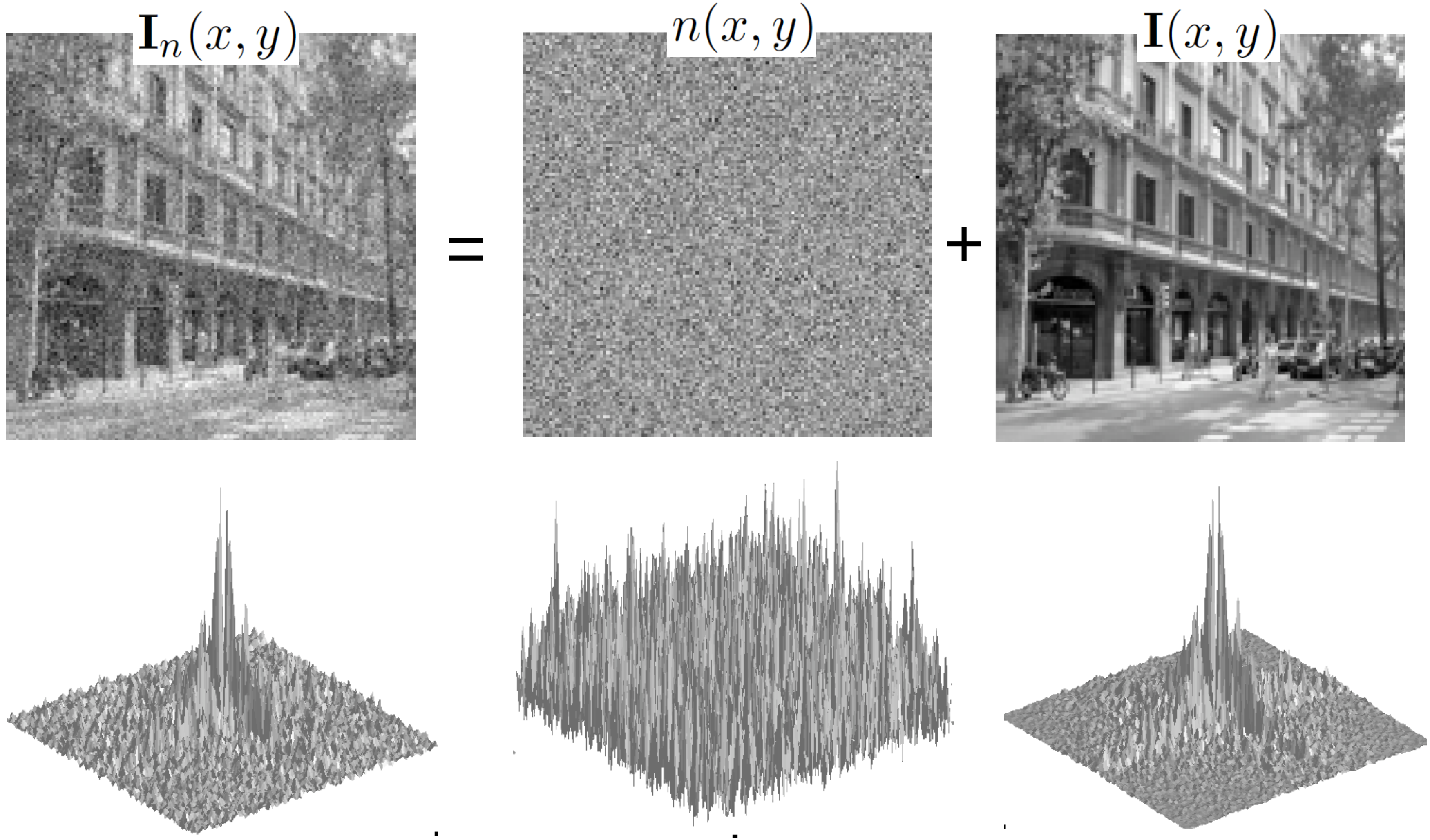
Sampling new images

Note: The average of many hair images will not give a distribution for hair images.
I believe we will get clouds again…
This representation does not encode other correlations like:
"all hairs should follow a similar orientation"

Denoising

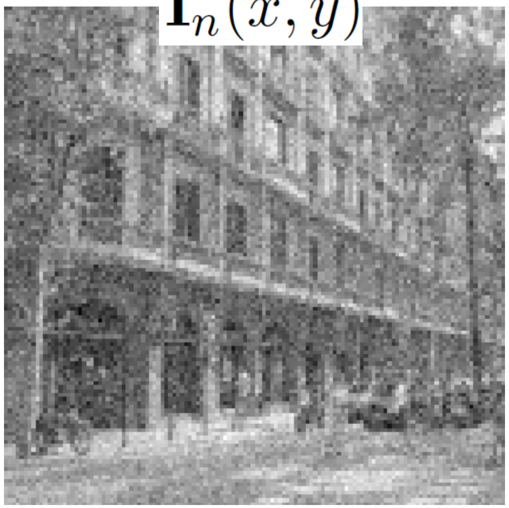
Decomposition of a noisy image



Denoising

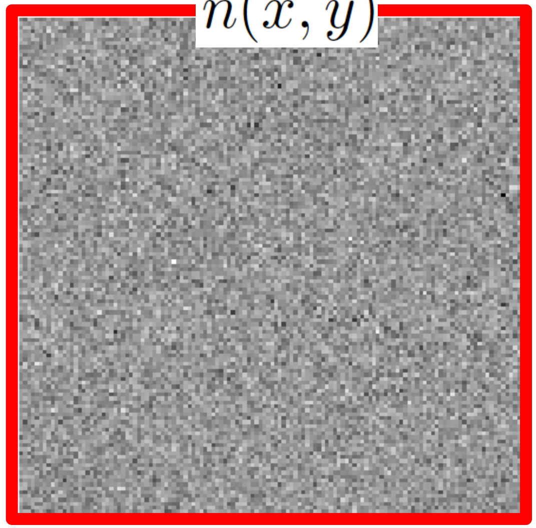
Decomposition of a noisy image

$\mathbf{I}_n(x,y)$



=

$n(x,y)$



White Gaussian noise: $N(0, \sigma_n^2)$

+

$\mathbf{I}(x,y)$

Natural image

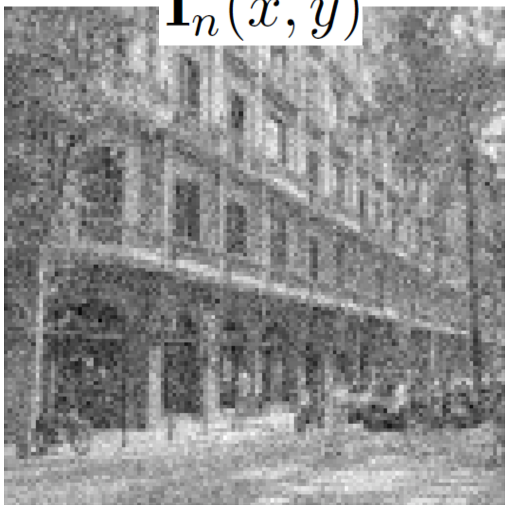
Find I(x,y) that maximizes the posterior (maximum a posteriory, MAP):

$$\max_{\mathbf{I}} p(\mathbf{I}|\mathbf{I}_n) = \max_{\mathbf{I}} \underbrace{\boxed{p(\mathbf{I}_n|\mathbf{I})}}_{\text{likelihood}} \times \underbrace{\boxed{p(\mathbf{I})}}_{\text{prior}}$$

Denoising

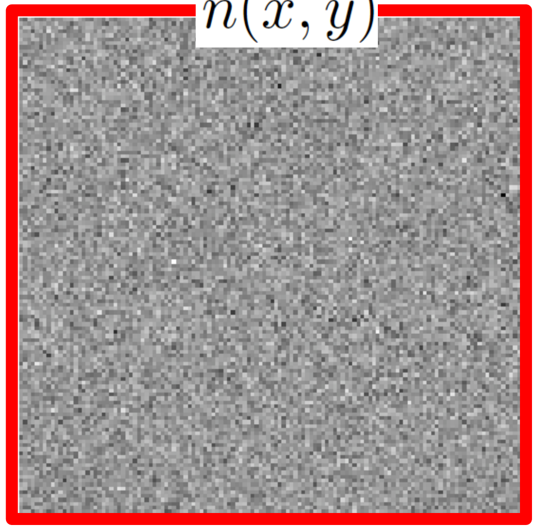
Decomposition of a noisy image

$\mathbf{I}_n(x,y)$



$=$

$n(x,y)$



$+$

$\mathbf{I}(x,y)$

White Gaussian noise: $N(0, \sigma_n^2)$

Natural image

Find I(x,y) that maximizes the posterior (maximum a posteriory, MAP):

$$\max_{\mathbf{I}} p(\mathbf{I}|\mathbf{I}_n) = \max_{\mathbf{I}} \underbrace{\boxed{p(\mathbf{I}_n|\mathbf{I})}}_{\text{likelihood}} \times \underbrace{\boxed{p(\mathbf{I})}}_{\text{prior}}$$

$$= \max_{\mathbf{I}} \boxed{\exp(-\left|\mathbf{I}_n - \mathbf{I}\right|^2 / \sigma_n^2)} \times \boxed{\exp\left(-\frac{1}{2}\mathbf{I}^T \mathbf{C}^{-1} \mathbf{I}\right)}$$

Denoising

$$\max_{\mathbf{I}} p(\mathbf{I}|\mathbf{I}_n) = \max_{\mathbf{I}} \underbrace{\boxed{p(\mathbf{I}_n|\mathbf{I})}}_{\text{likelihood}} \times \underbrace{\boxed{p(\mathbf{I})}}_{\text{prior}}$$

$$= \max_{\mathbf{I}} \boxed{\exp(-\left|\mathbf{I}_n - \mathbf{I}\right|^2 / \sigma_n^2)} \times \boxed{\exp\left(-\frac{1}{2}\mathbf{I}^T \mathbf{C}^{-1} \mathbf{I}\right)}$$

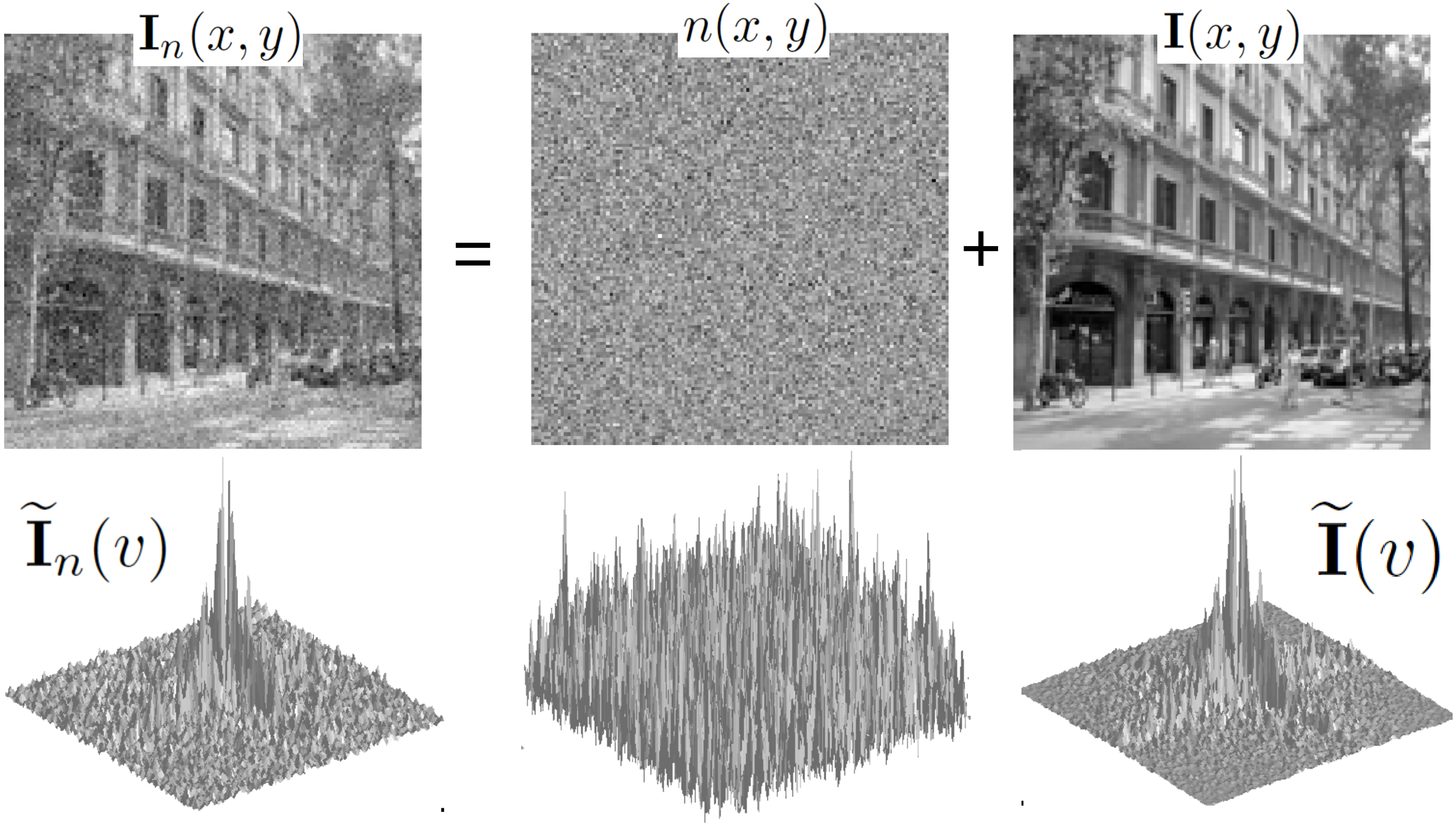
The solution is:

$$\mathbf{I} = \mathbf{C}\left(\mathbf{C} + \sigma_n^2 \mathbb{I}\right)^{-1} \mathbf{I}_n$$ (note this is a linear operation)

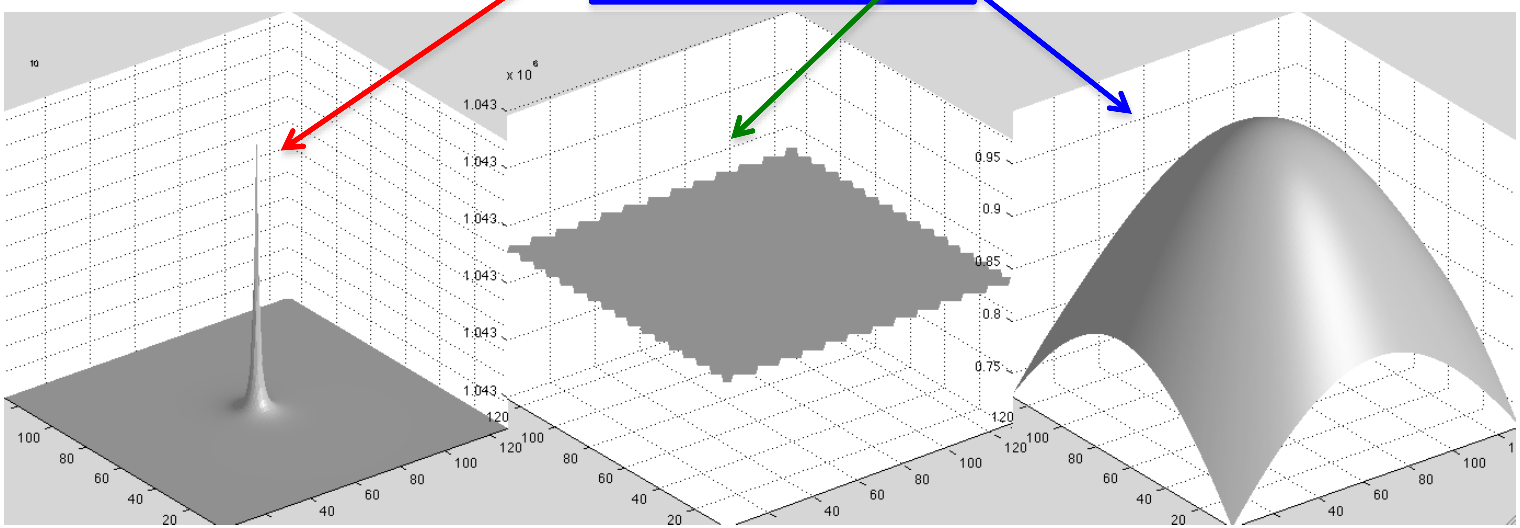
This can also be written in the Fourier domain, with C = EDE^T:

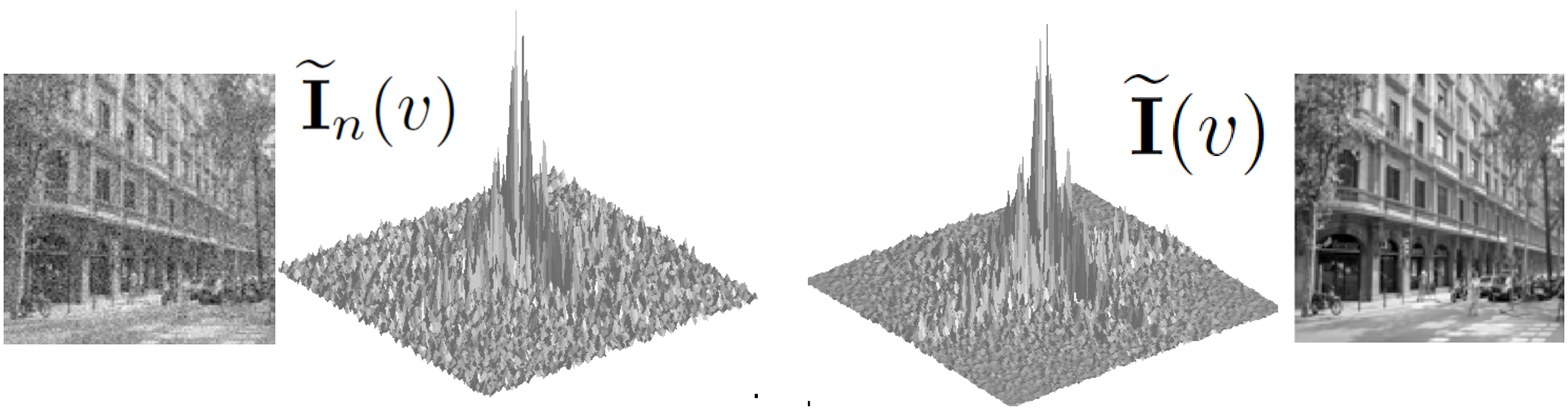
$$\widetilde{\mathbf{I}}(v) = \frac{A/|v|^{2\alpha}}{A/|v|^{2\alpha} + \sigma_n^2} \widetilde{\mathbf{I}}_n(v)$$

Decomposition of a noisy image

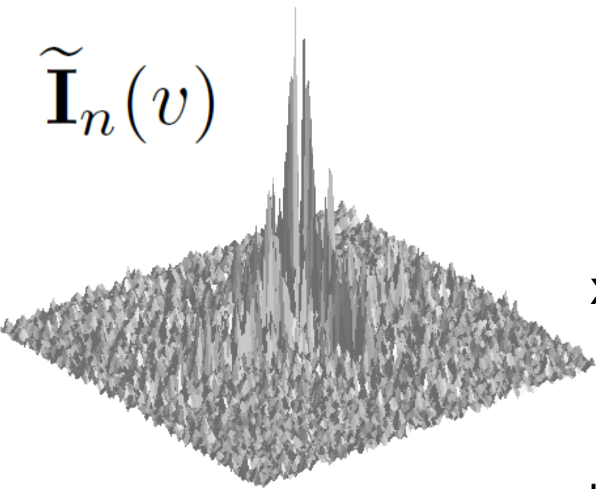


$$\widetilde{\mathbf{I}}(v) = \boxed{\frac{A/|v|^{2\alpha}}{A/|v|^{2\alpha} + \sigma_n^2}}\,\widetilde{\mathbf{I}}_n(v)$$



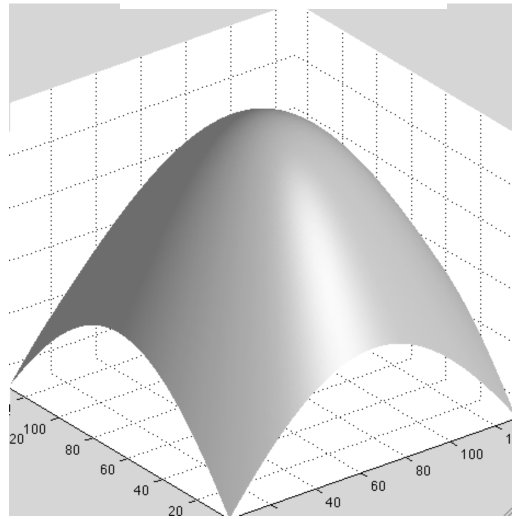


$\widetilde{\mathbf{I}}_n(v)$

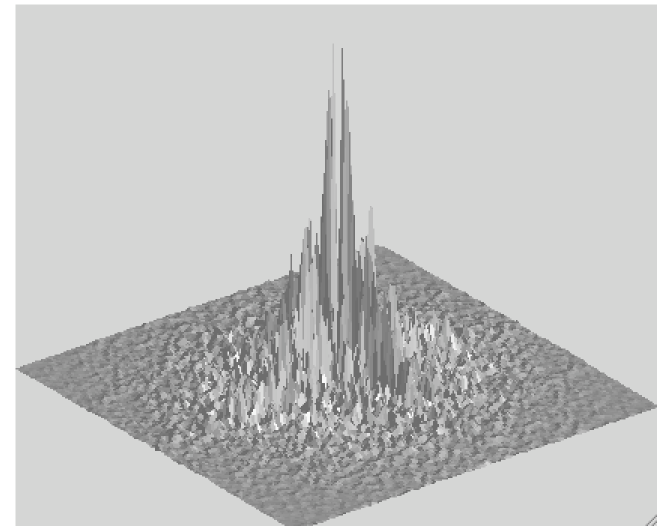


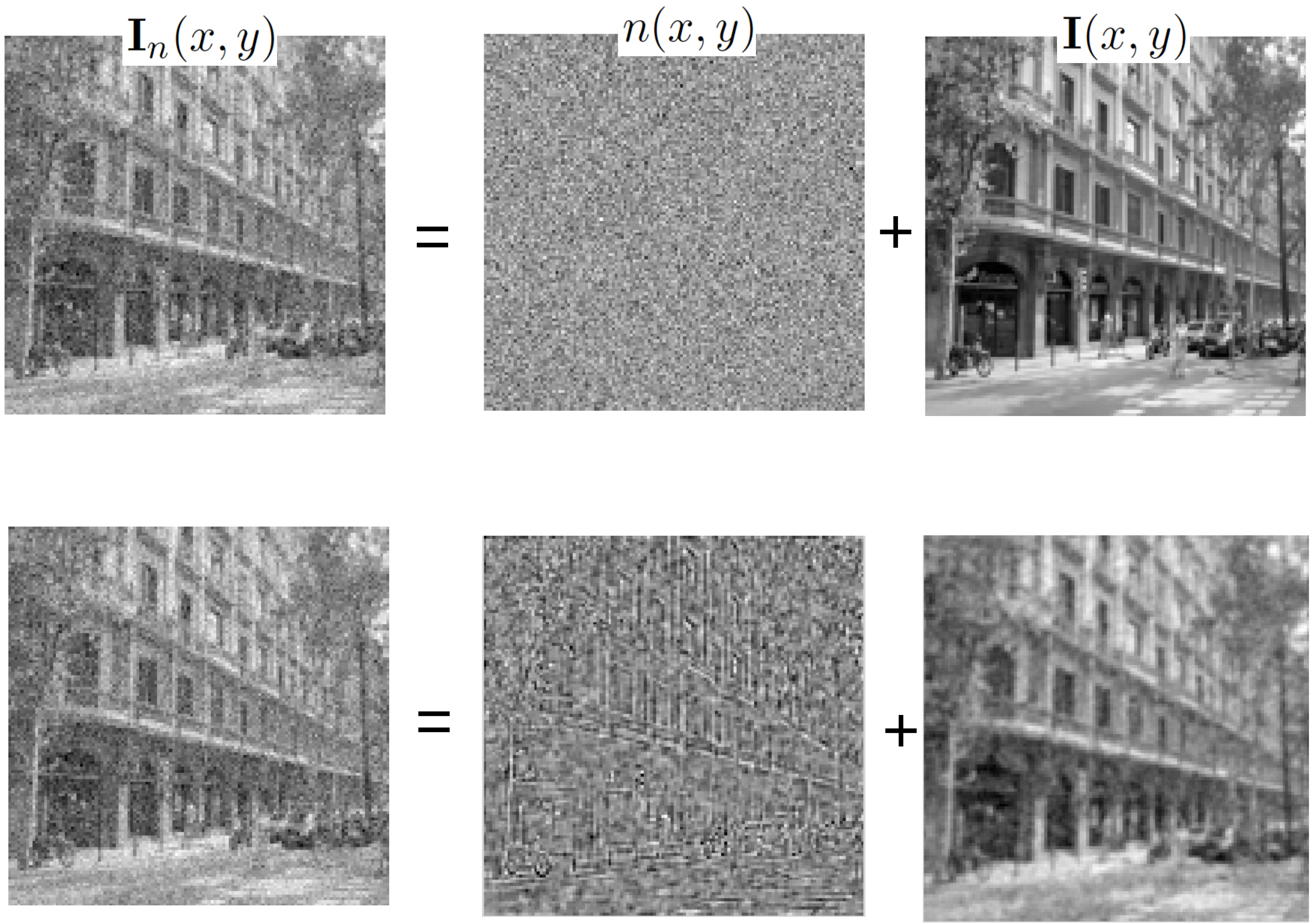
x

$$\frac{A/|v|^{2\alpha}}{A/|v|^{2\alpha}+\sigma_n^2}$$



=





$\mathbf{I}_n(x,y)$
$=$
$n(x,y)$
$+$
$\mathbf{I}(x,y)$
$=$
$+$

Statistical modeling of images

A small neighborhood

Edges

[-1 1]

[-1 1]

g[m,n]

$\otimes$

[-1, 1]

h[m,n]

=

f[m,n]

$[-1\ 1]^T$

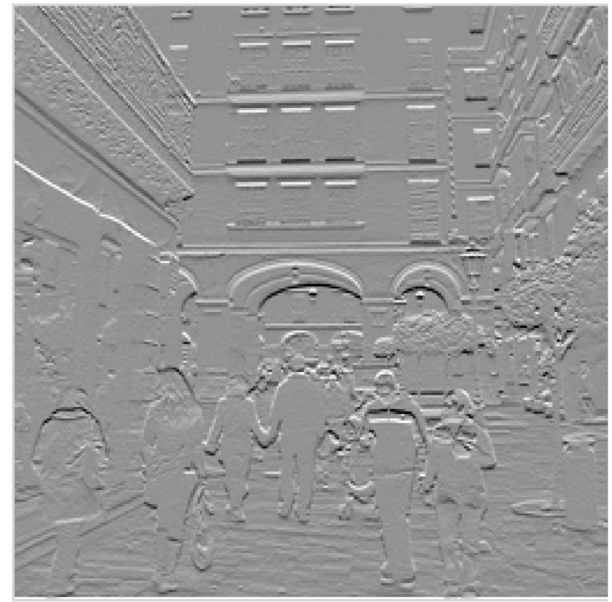
$g[m,n]$

$\otimes$

$[-1, 1]^T$

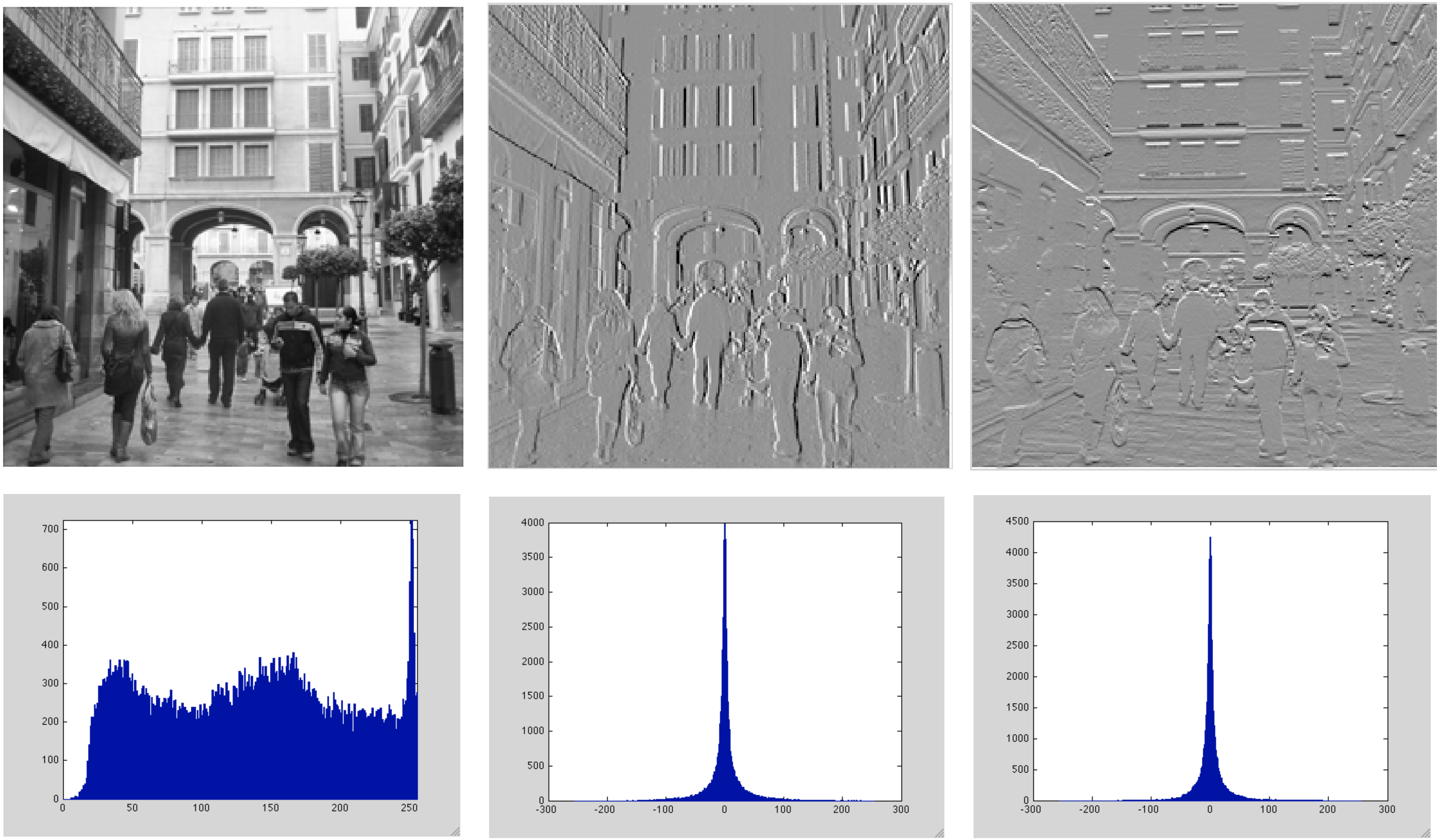
$h[m,n]$

$=$

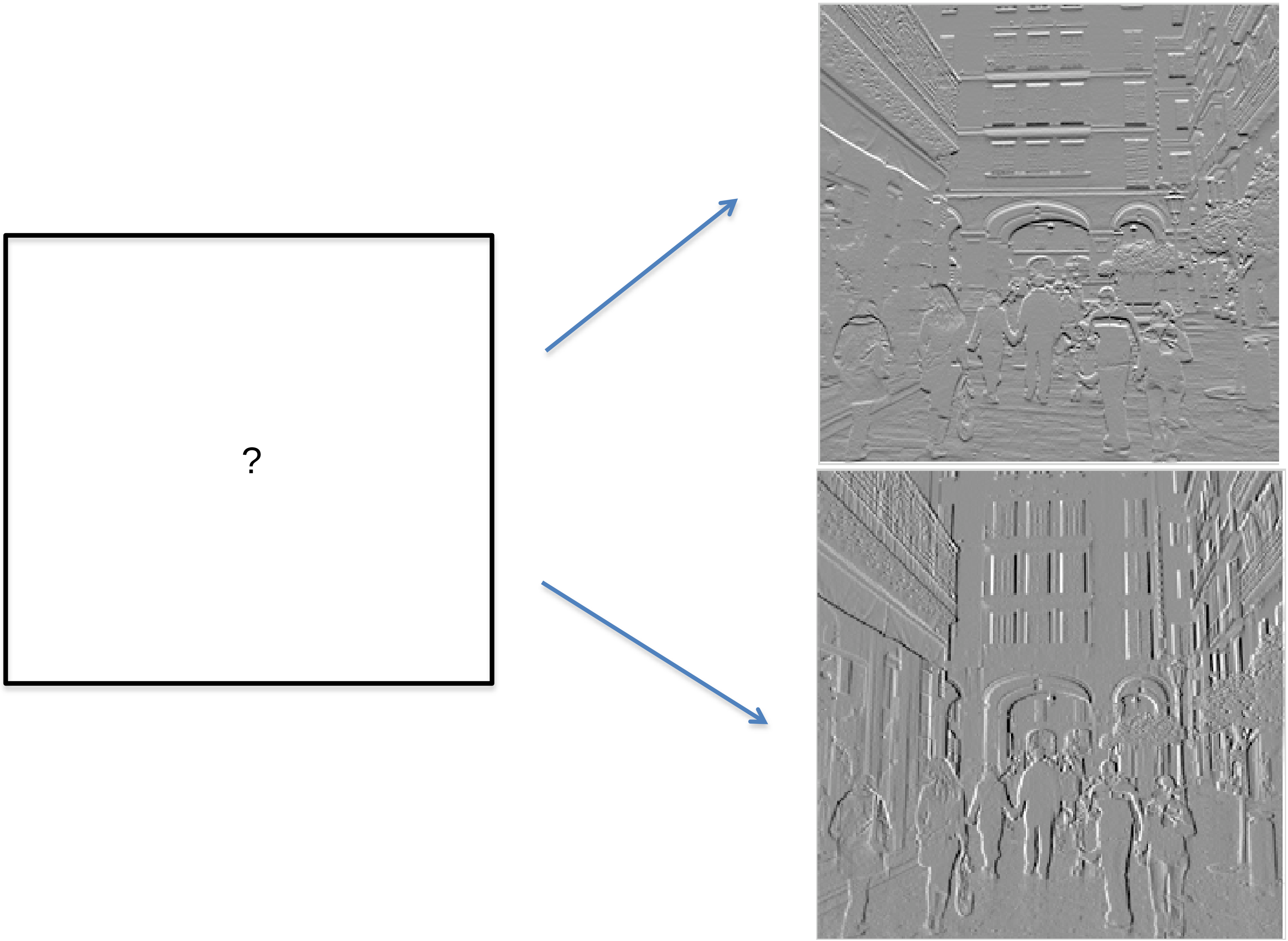


$f[m,n]$

Observation: Sparse filter response



Back to the image

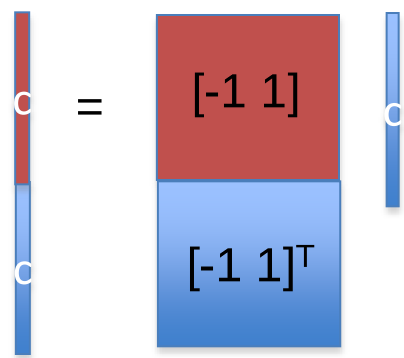


Reconstruction from derivatives

$$F = H\,G$$

$$c \quad = \quad \begin{bmatrix} 1 & -1 & & & & & & \\ & 1 & -1 & & & & & \\ & & 1 & -1 & & & & \\ & & & 1 & -1 & & & \\ & & & & 1 & -1 & & \\ & & & & & 1 & -1 & \\ & & & & & & 1 & -1 \\ & & & & & & & 1 \end{bmatrix} \quad c$$

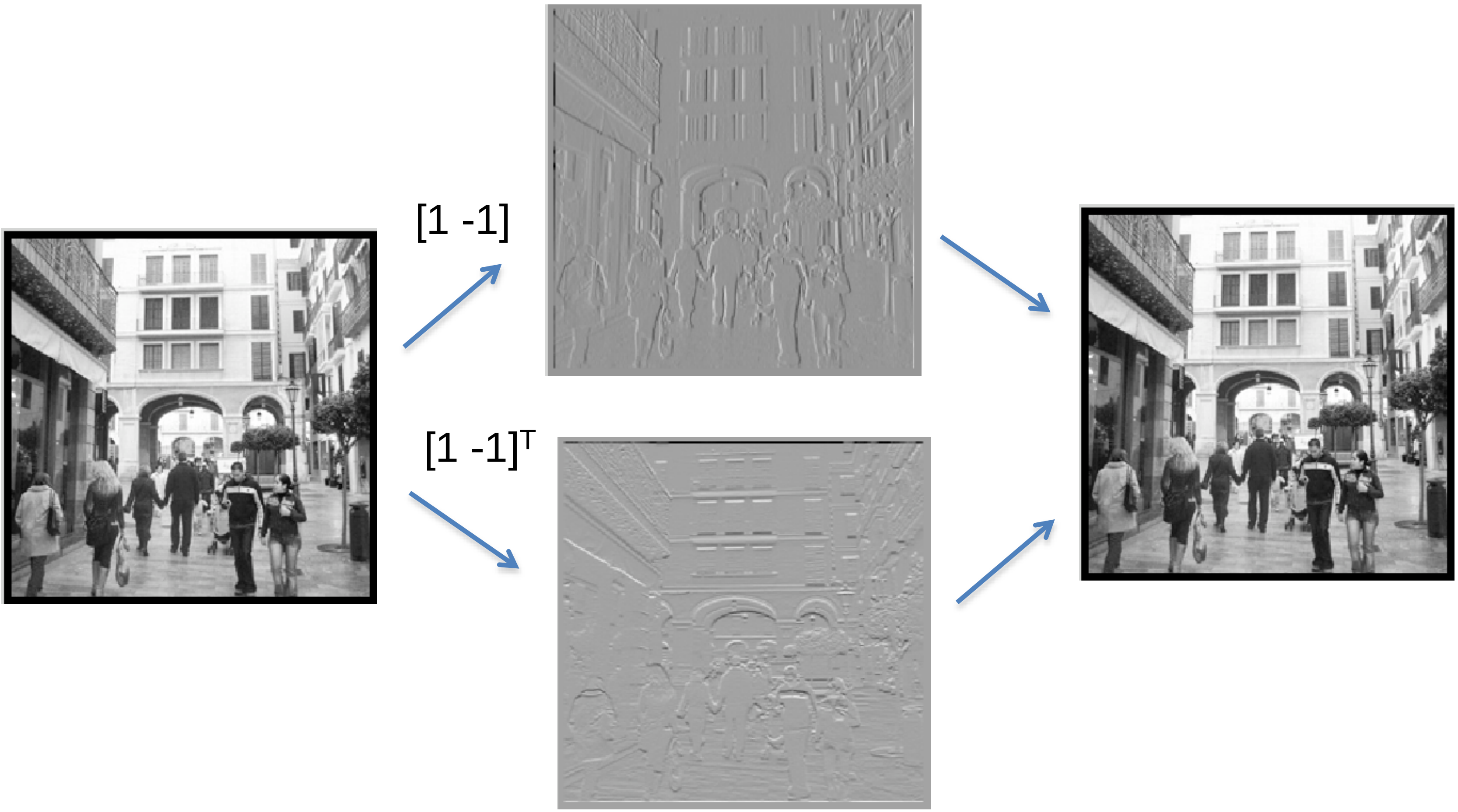
If we have multiple filter outputs:



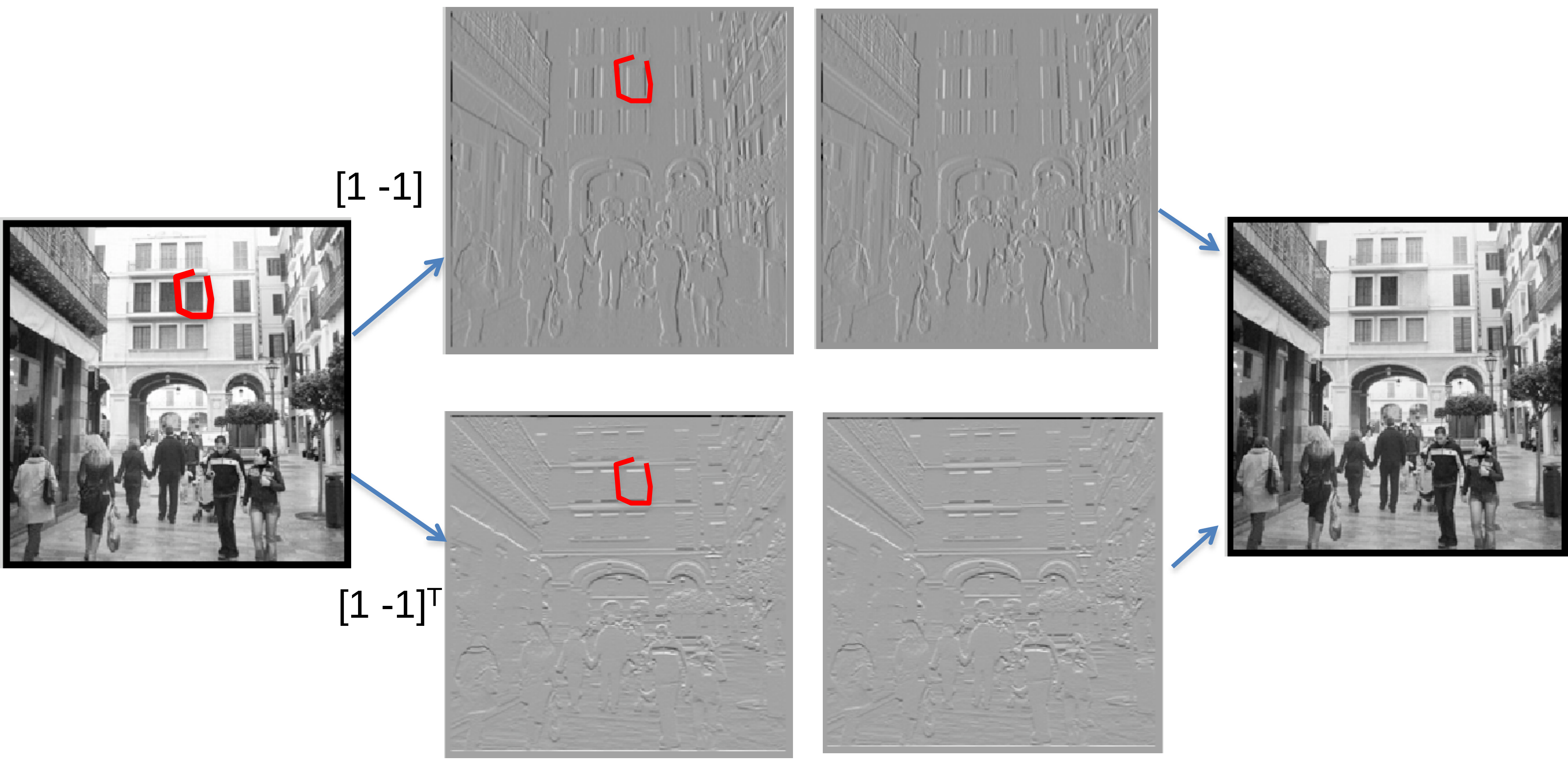
If the transformation H is not invertible, we can compute the pseudo-inverse:

$$\hat{G} = (H^T H)^{-1} H^T F$$

Reconstruction

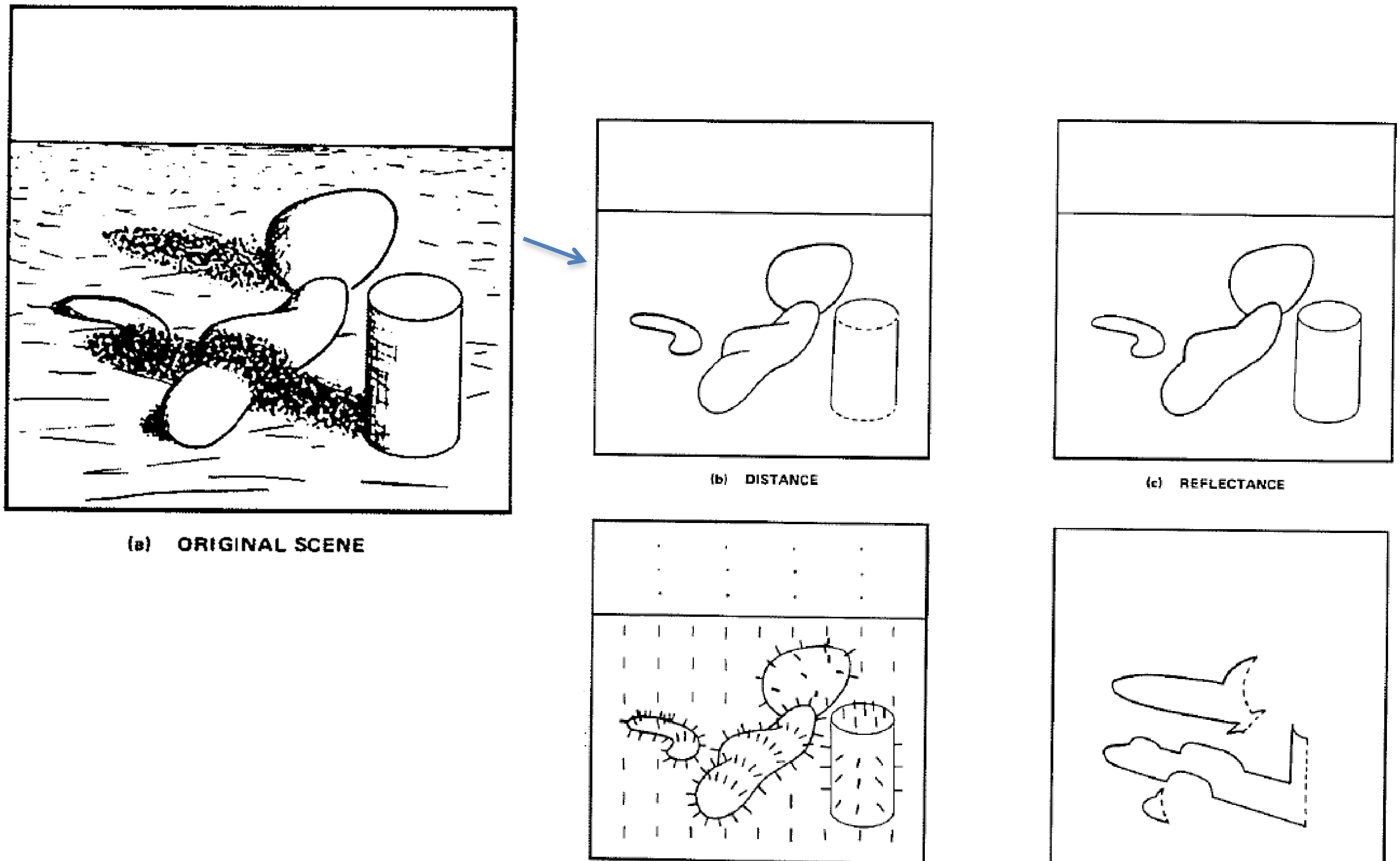


Editing the edge image



Thresholding edges

Intrinsic images



(a) ORIGINAL SCENE

(b) DISTANCE

(c) REFLECTANCE

(d) ORIENTATION (VECTOR)

(e) ILLUMINATION

Separating images into components

=

X

Table 1 The Nature of Edges

Region Intensities LA	LB	Edge Type	Region Types	Intrinsic Edges Intrinsic Values D	N	R	I
Constant	Constant	Occluding sense unknown	A B shadowed	EDGE	EDGE	EDGE RA RB	IA IB
Constant	Varying	1 Shadow	A shadowed B illuminated		NB.S	RA RB	EDGE IA IB
		2 A occludes B	A shadowed B illuminated	EDGE DA DB	EDGE NA	EDGE RA	EDGE IA
Varving	Varying	Inconsistent with domain					
Constant	Tangency	B occludes A	A shadowed B illuminated	EDGE DA DB	EDGE NB	EDGE RA RB	EDGE IA IB
Varying	Tangency	B occludes A	A B illuminated	EDGE DA DB	EDGE NB	EDGE RB	EDGE IB IA
Tangency	Tangency	Not seen from general position					

Table 1 catalogs the possible appearances and interpretations of an edge between two regions, A and B.

In this table, "Constant" means constant intensity along the edge, "Tangency" means that the tangency condition is met, and

RECOVERING INTRINSIC SCENE CHARACTERISTICS FROM IMAGES

Technical Note 157

April 1978

By: Harry G. Barrow
J. Martin Tenenbaum
Artificial Intelligence Center

The research reported herein was supported by the National Science Foundation, under NSF Grant No. ENG76-01272.

To appear in *Computer Vision Systems*, A. Hanson and E. Riseman, eds.. (Academic Press, New York, in press).

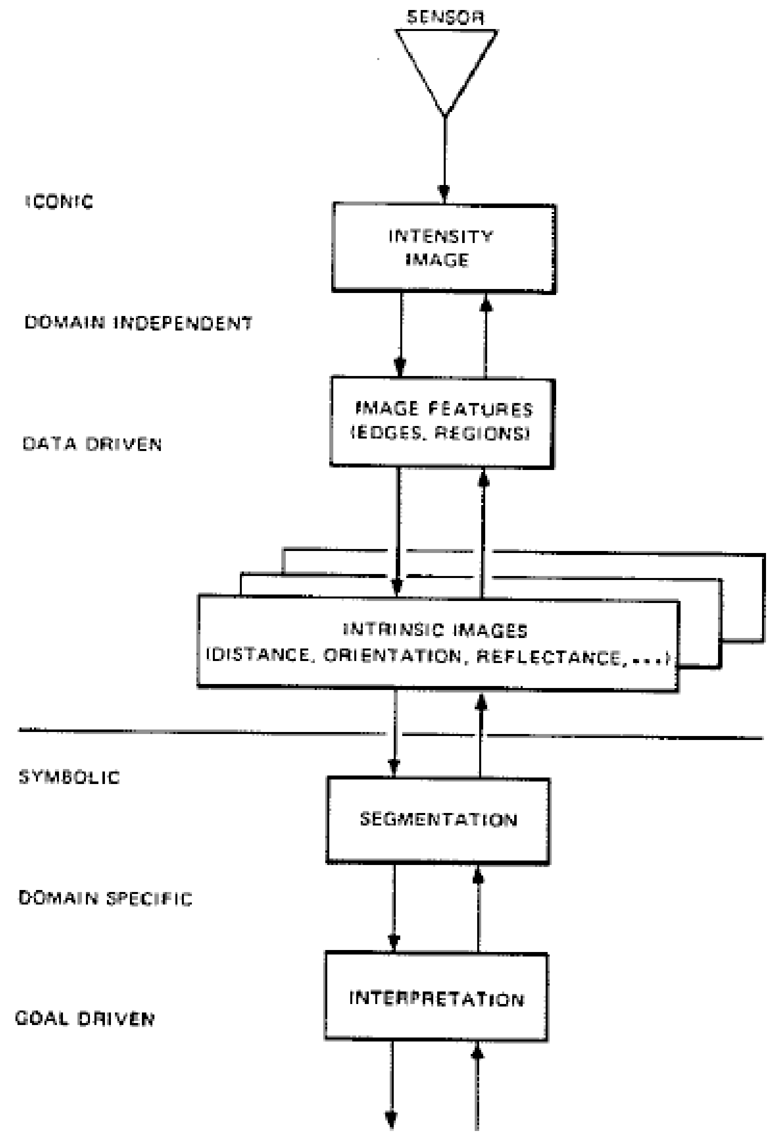
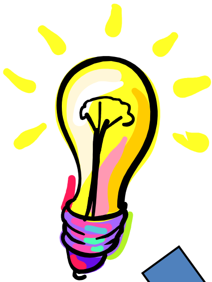
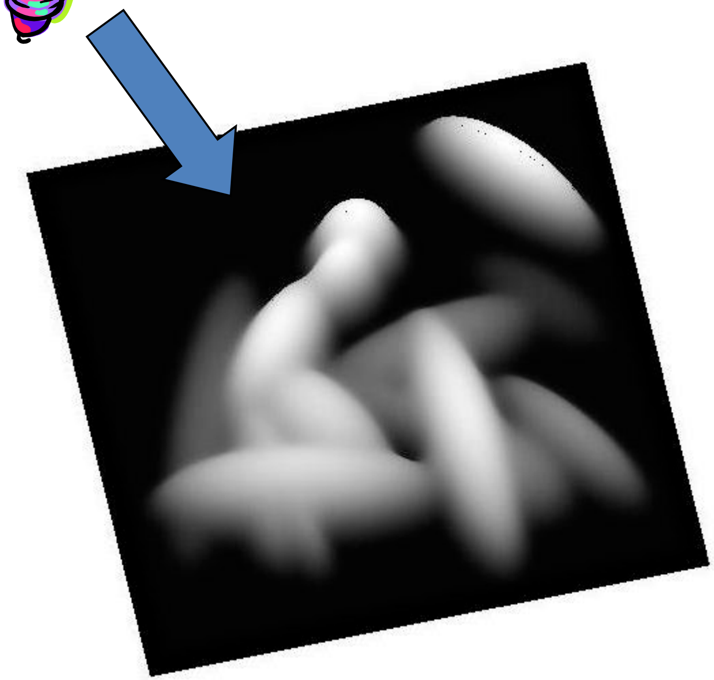


Figure 10 Organization of a visual system

Forming an Image



Illuminate the surface to get:



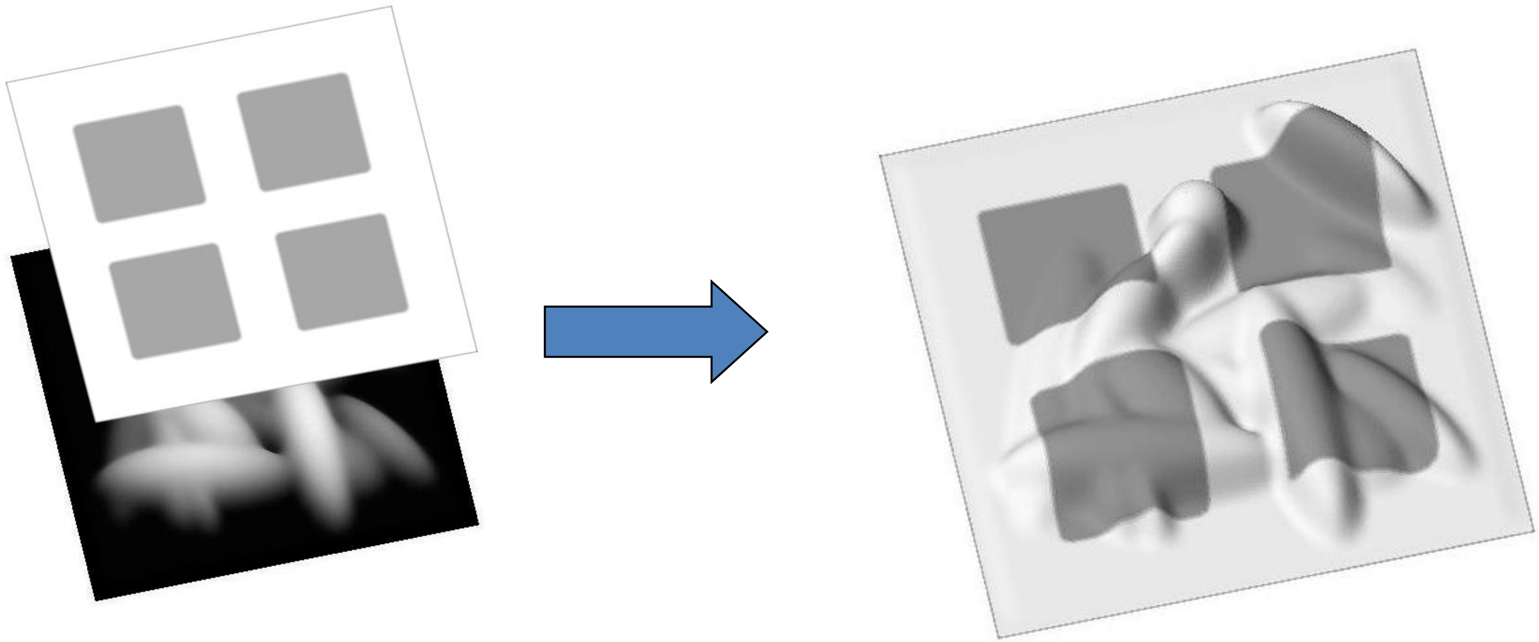
Surface (Height Map)

Shading Image

The shading image is the interaction of the shape of the surface and the illumination

Slide: Marshal Tappen

Painting the Surface

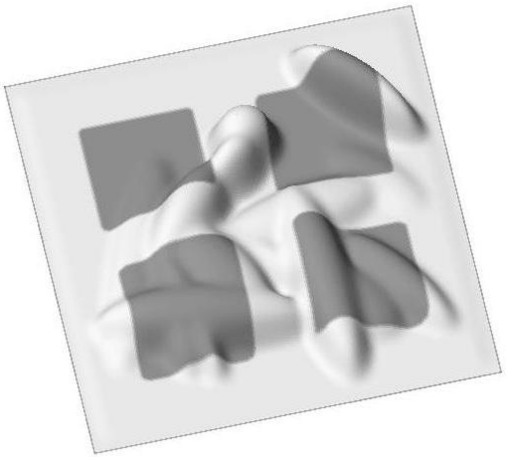


Scene Image

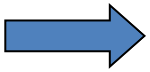
Add a reflectance pattern to the surface. Points inside the squares should reflect less light

Slide: Marshal Tappen

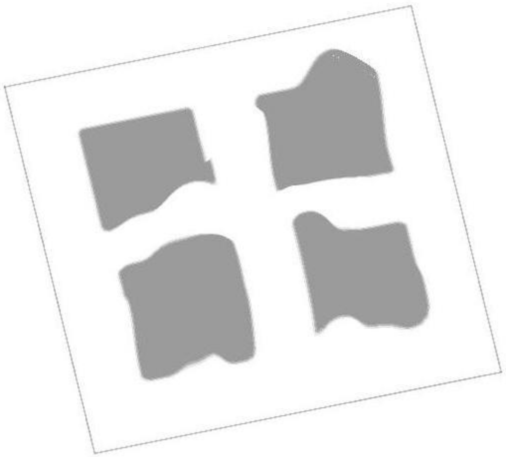
Goal



Image



Shading Image



Reflectance Image

Slide: Marshal Tappen

Retinex

E.H. Land, J.J. McCANN - Journal of the Optical society of America, 1971

Journal of the
OPTICAL SOCIETY
of AMERICA

VOLUME 61, NUMBER 1 JANUARY 1971

Lightness and Retinex Theory

EDWIN H. LAND* AND JOHN J. MCCANN
Polaroid Corporation, Cambridge, Massachusetts 02139
(Received 8 September 1970)

The reflectance tends to be constant across space except for abrupt changes at the transitions between objects or pigments. Thus a reflectance change shows itself as step edge in an image, while illuminance changes gradually over space. By this argument one can separate reflectance change from illuminance change by taking spatial derivatives: High derivatives are due to reflectance and low ones are due to illuminance.

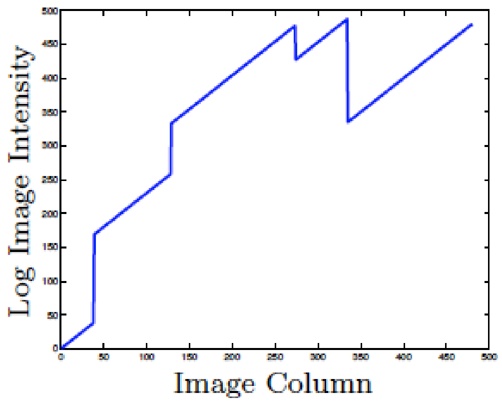
Retinex

Again, we are trying to solve an ill-posed problem:

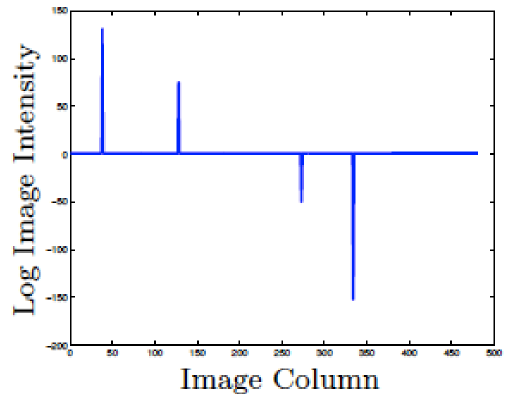
24 = ? x ?

From M. Tappen, PhD

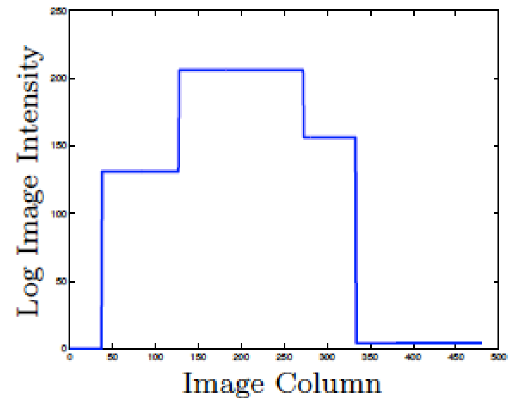
Retinex



(a) One column from the observed image.

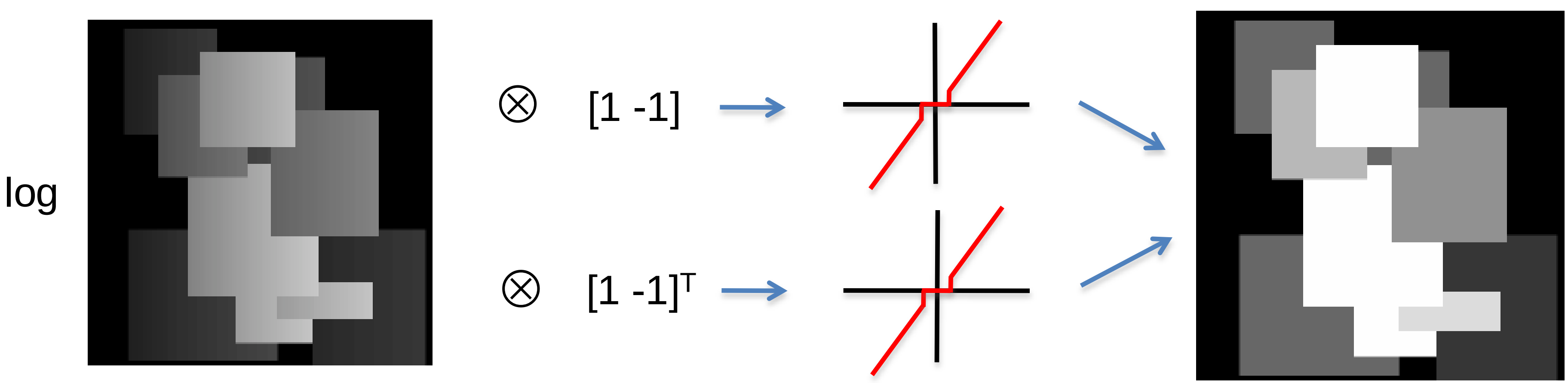


(b) The derivative of the plot from (a).



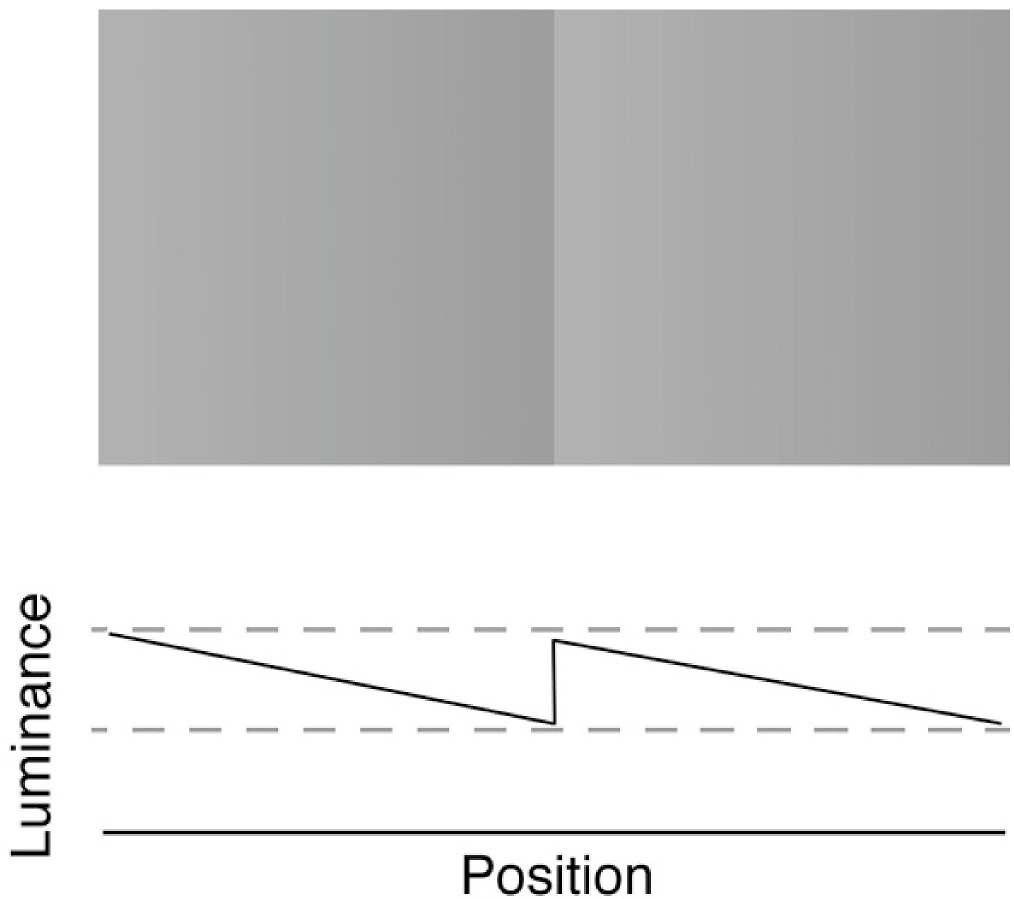
(c) The estimate of the log shading

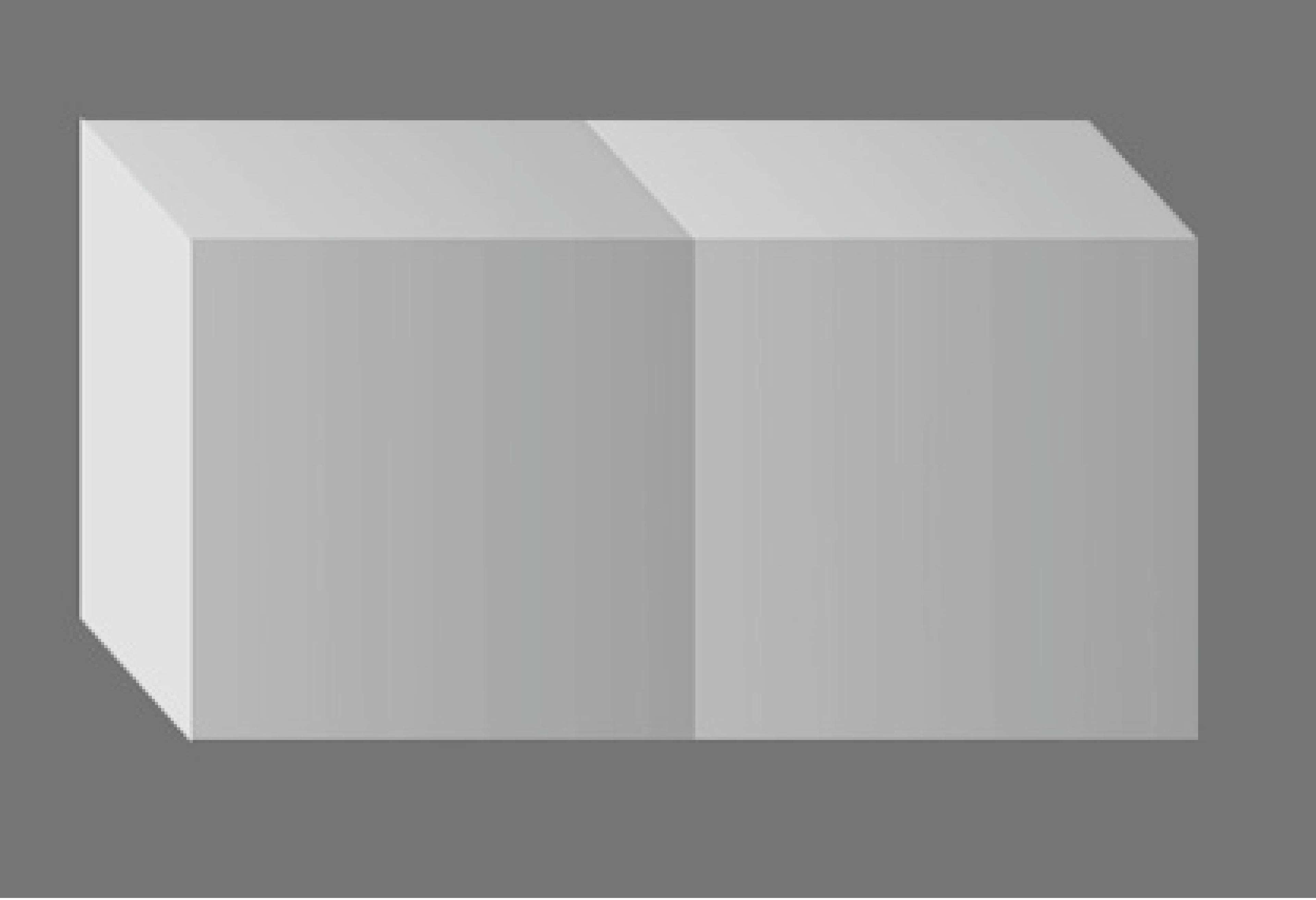
From M. Tappen, PhD

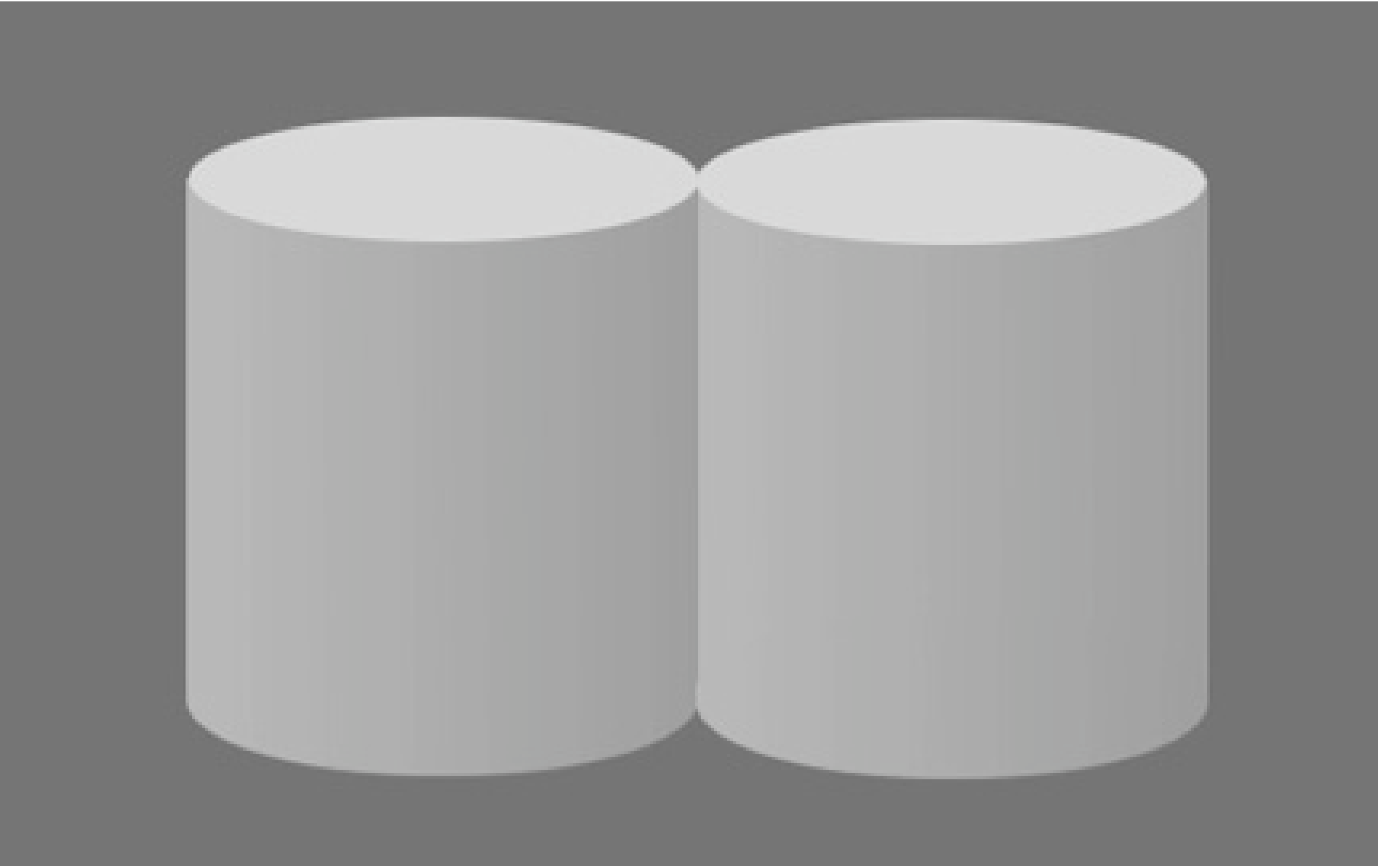


From M. Tappen, PhD

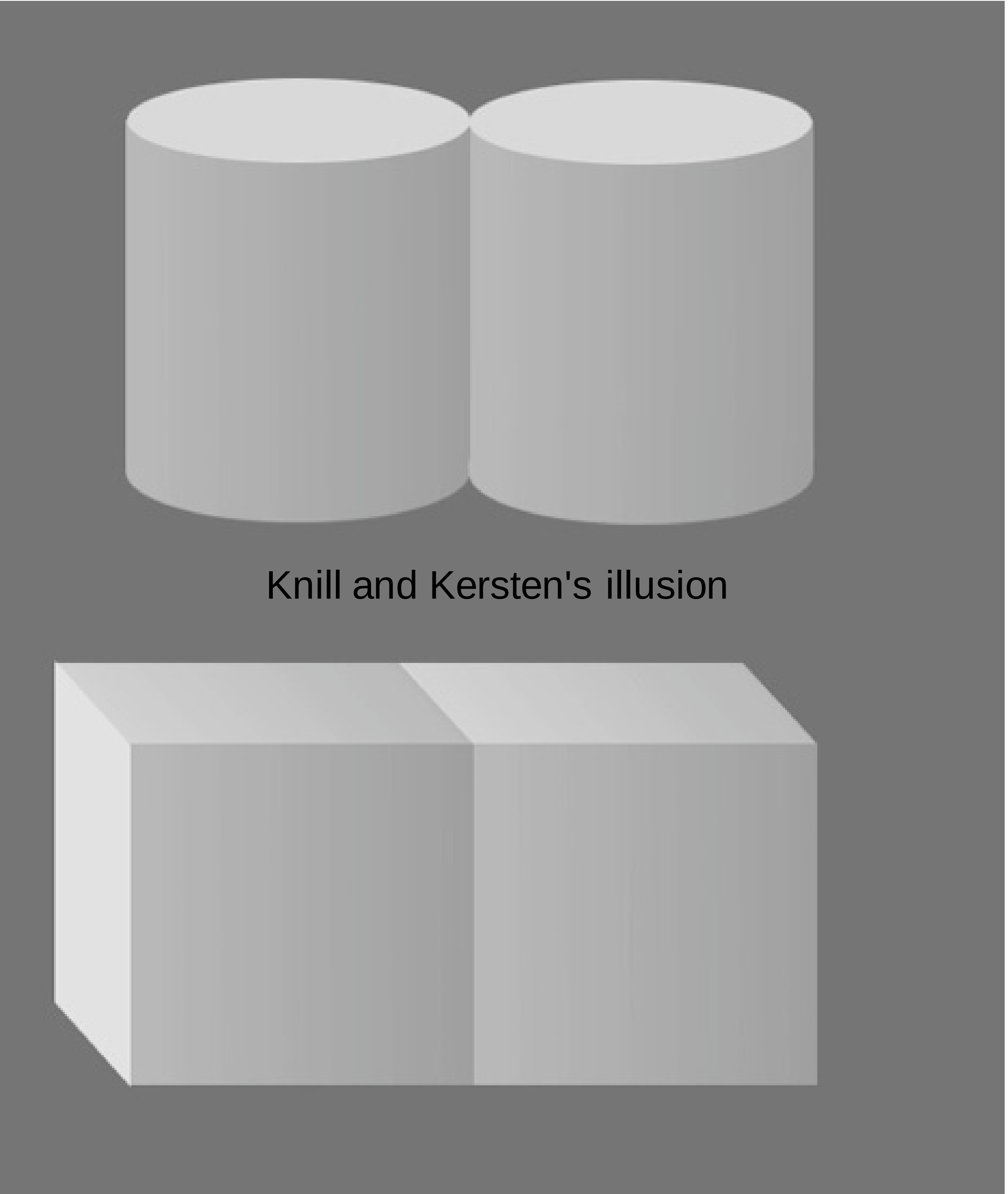
Craik-O'Brien-Cornsweet effect







Knill and Kersten's illusion



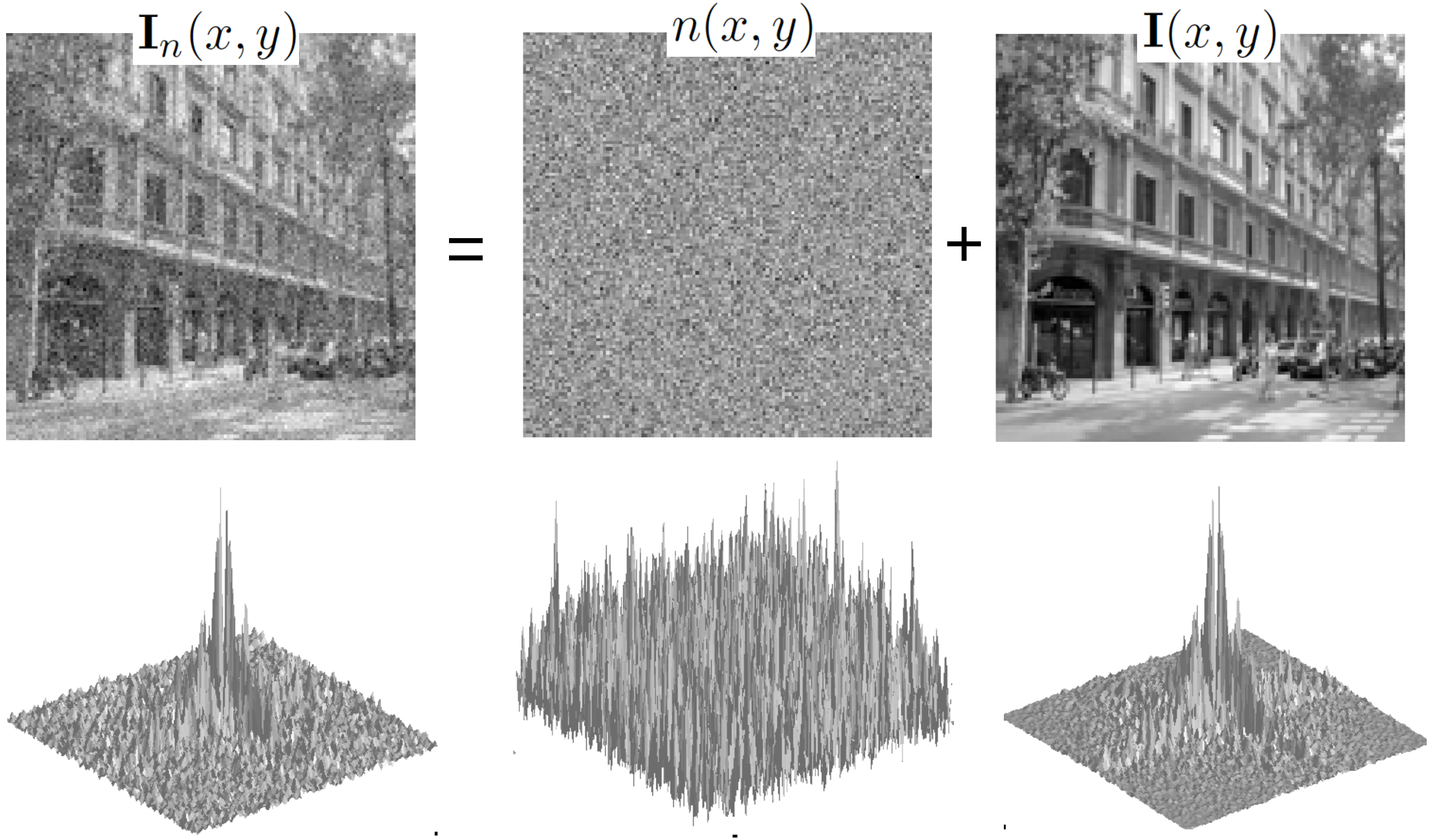
This illusion highlights the importance of scene interpretation.

← The effect is gone

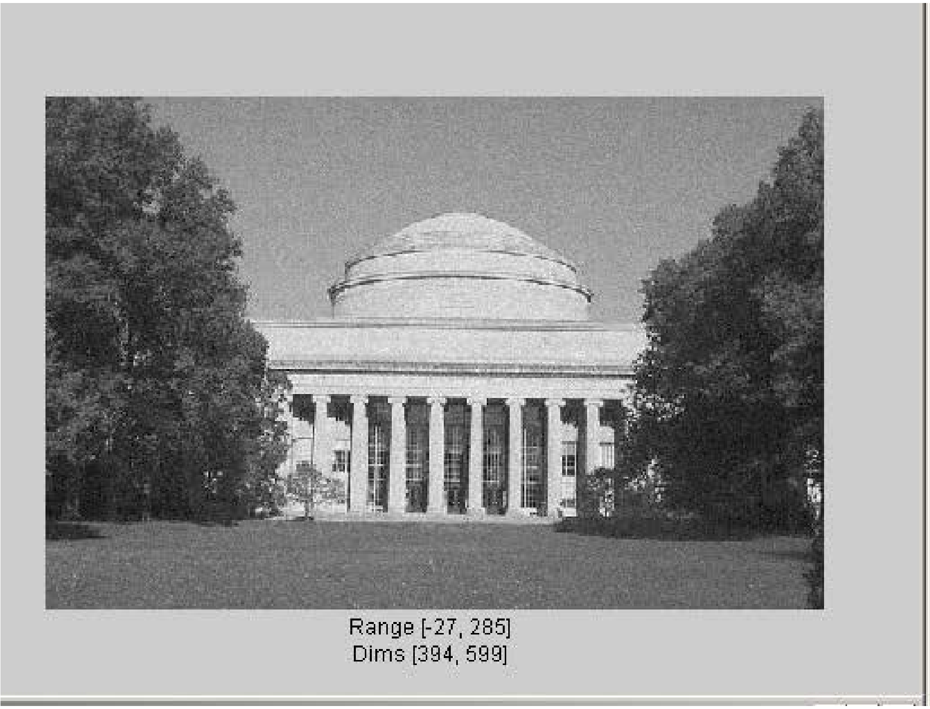
← and it comes back when the gradient is not explained by the shape.

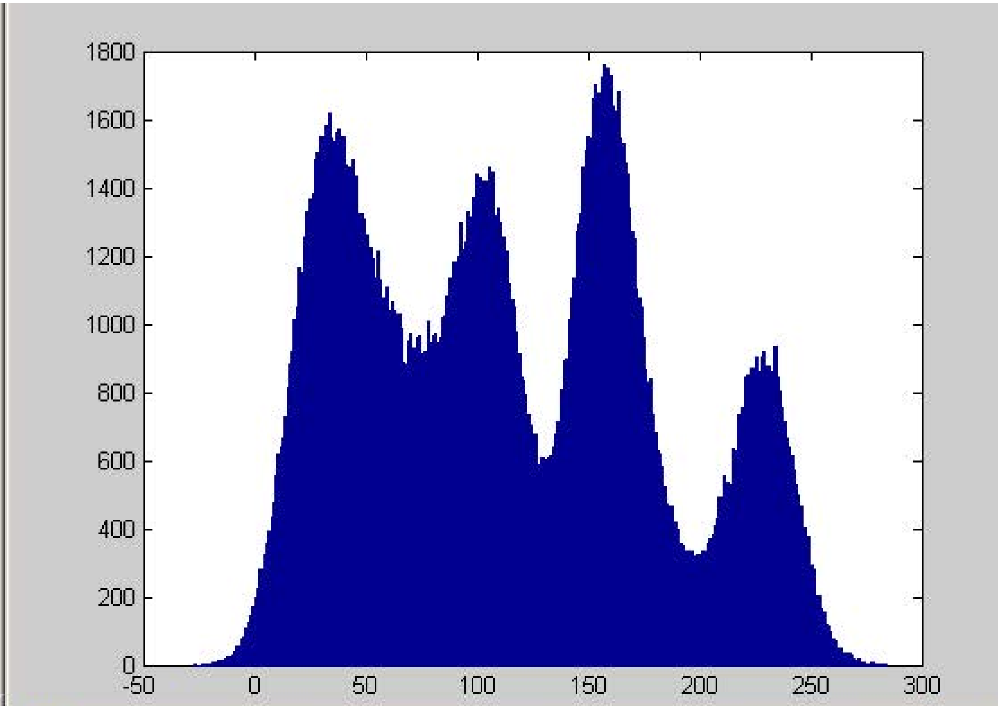
Denoising

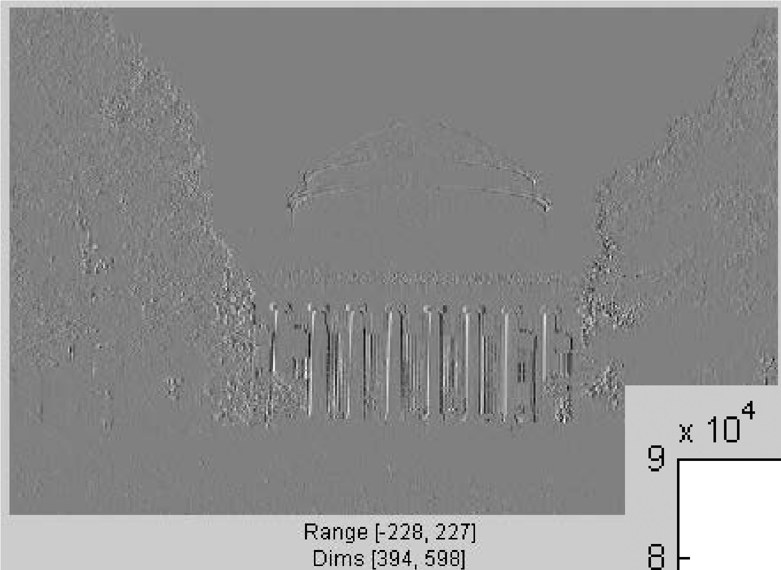
Decomposition of a noisy image



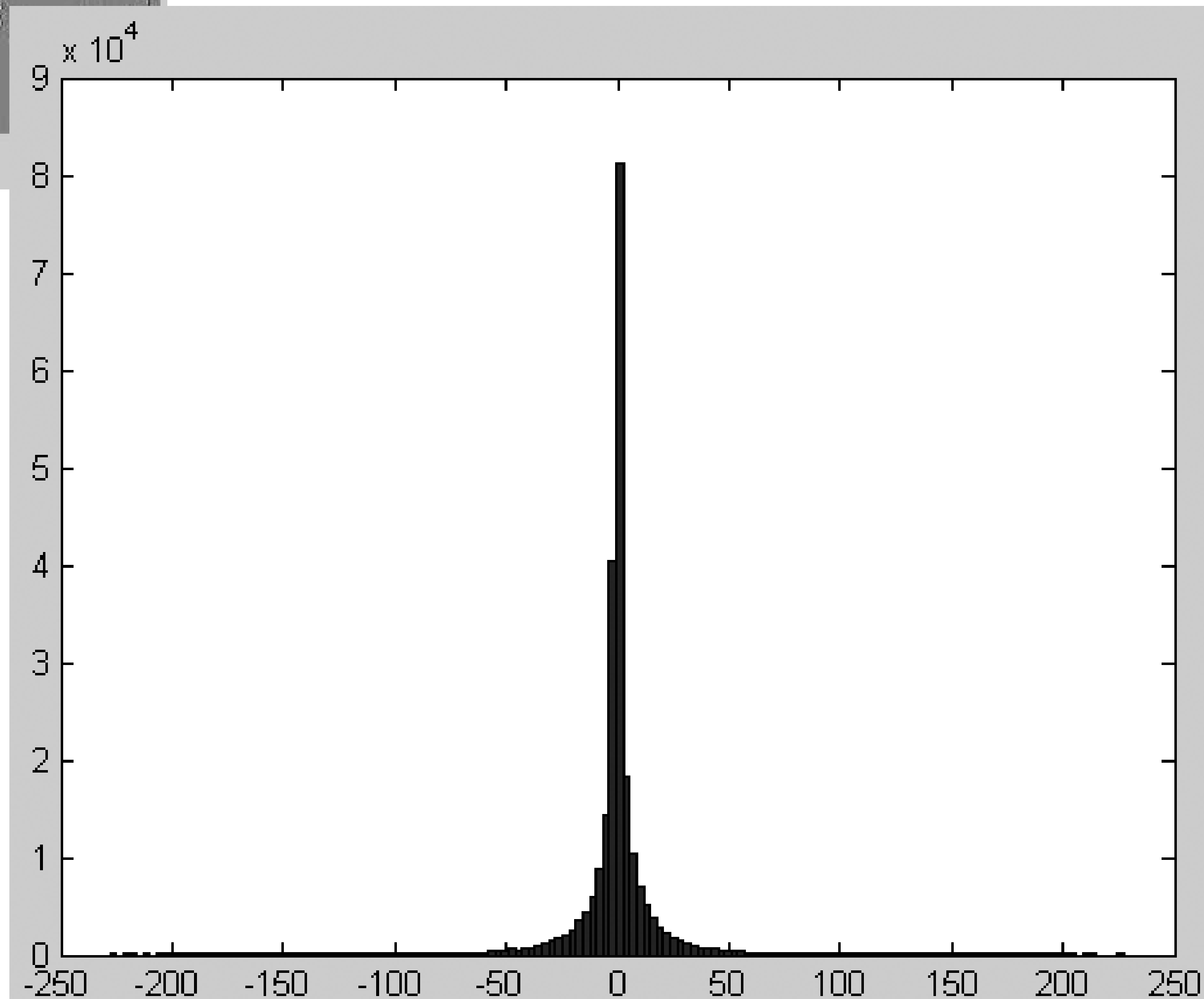
Pixel representation, noisy image histogram



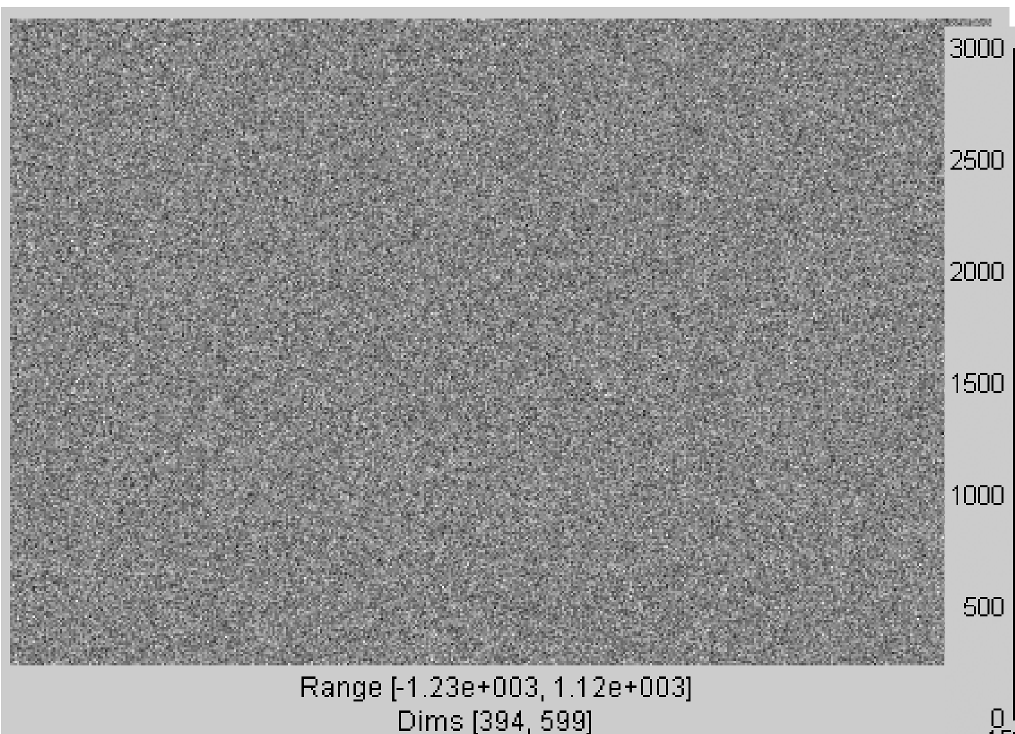


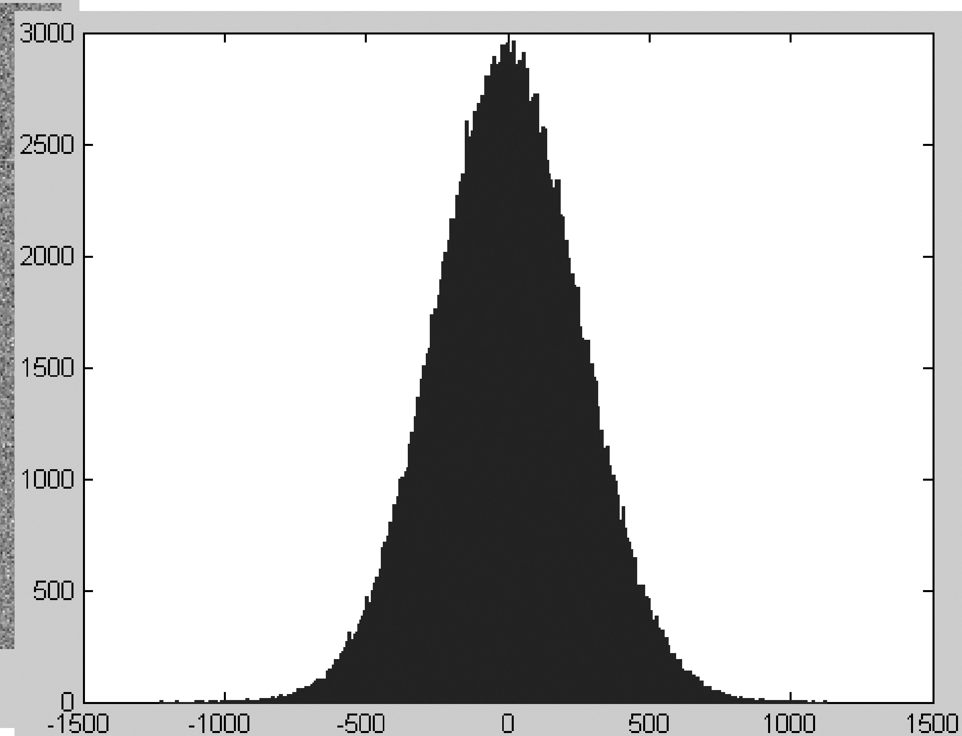


bandpassed representation image histogram

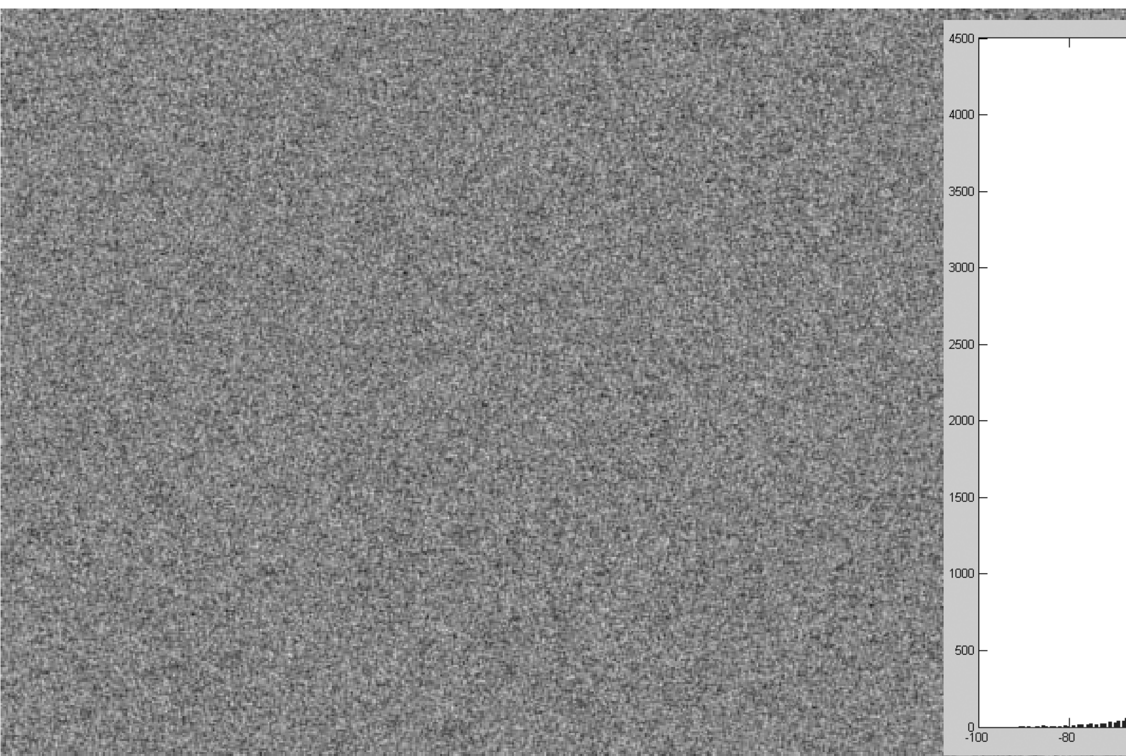


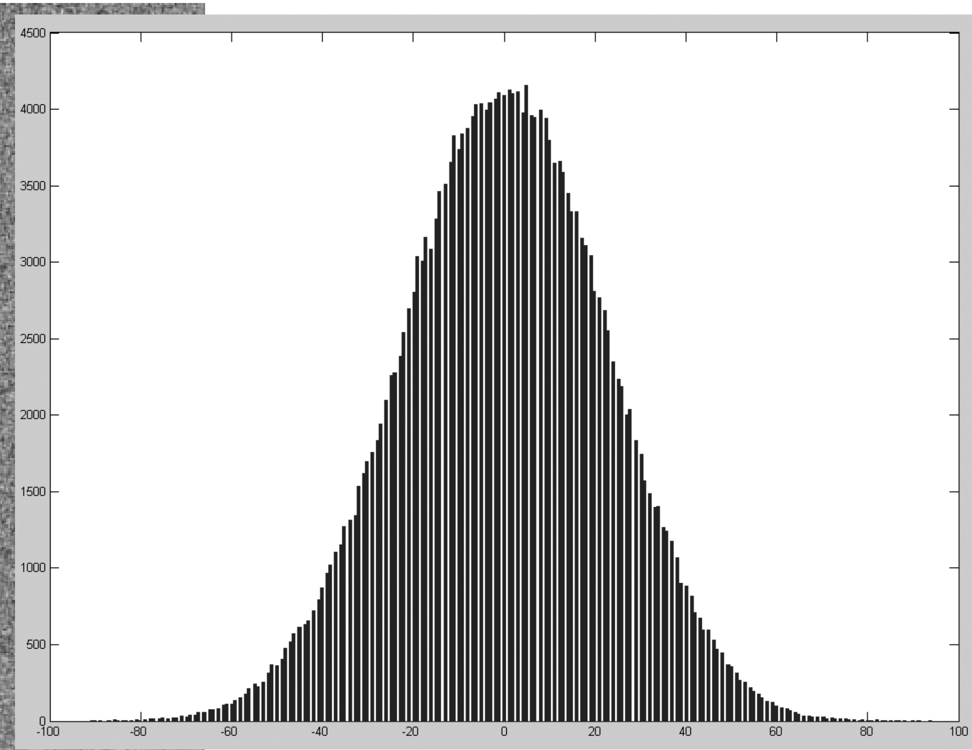
Pixel domain noise image and histogram



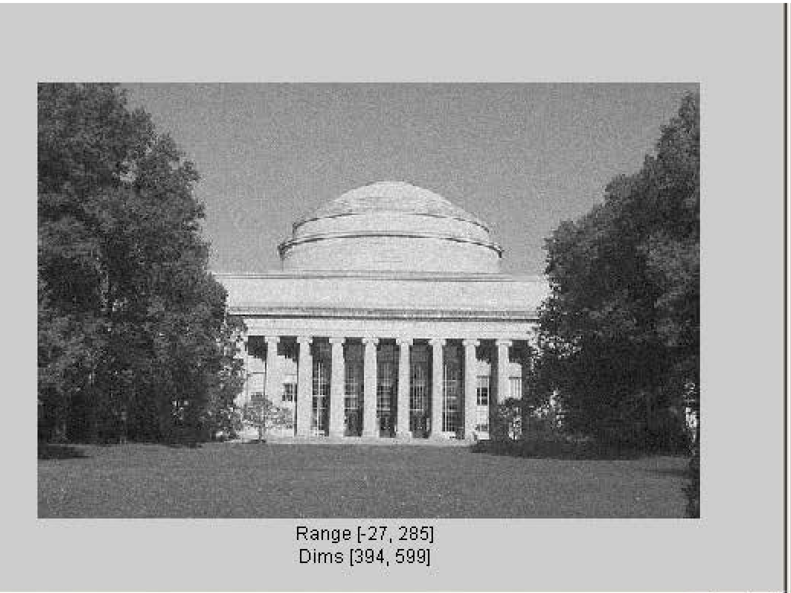


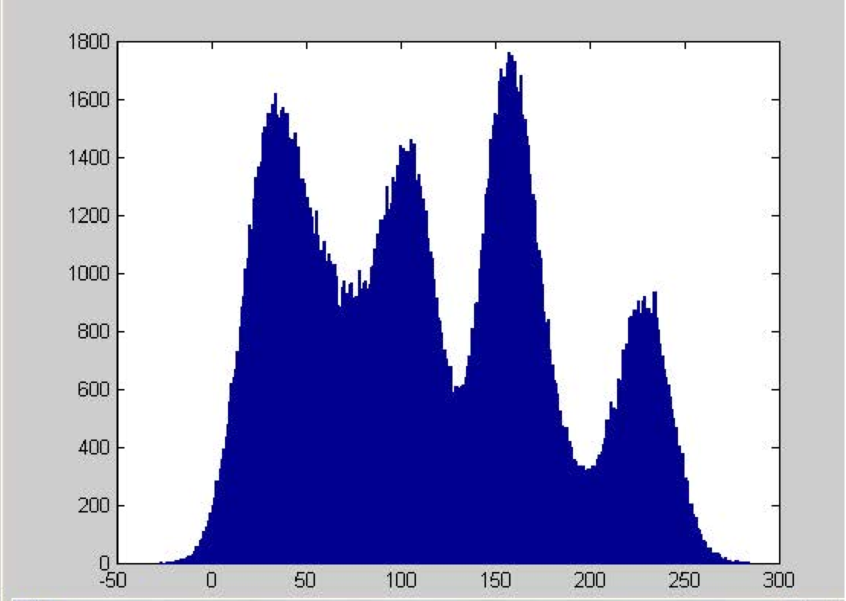
Bandpass domain noise image and histogram

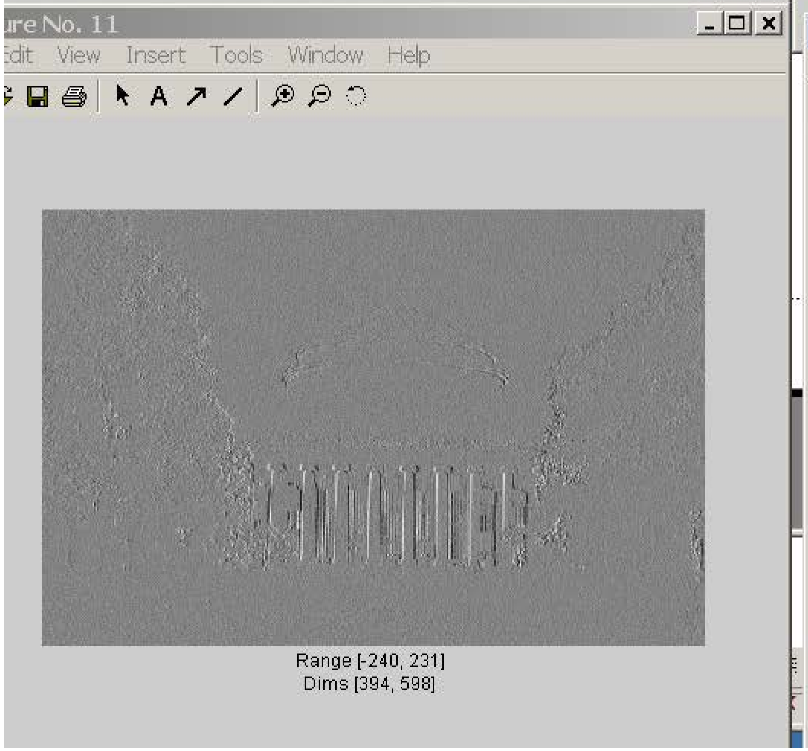


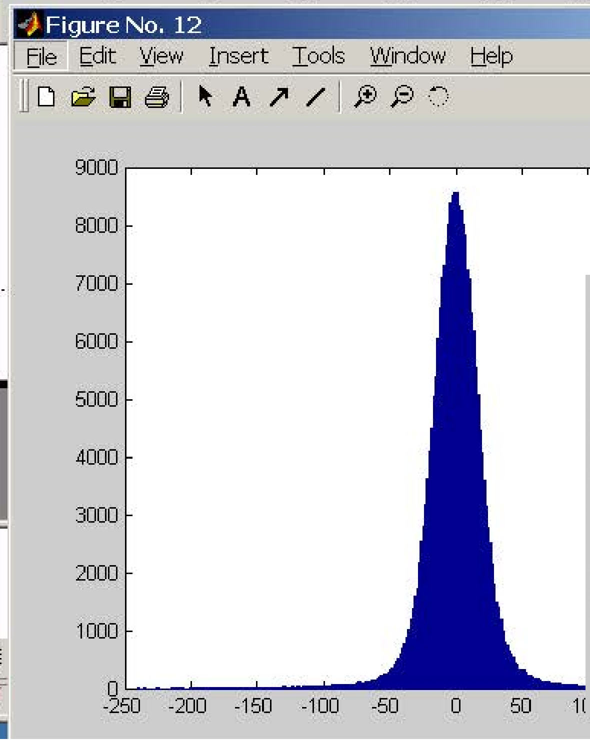


Noise-corrupted full-freq and bandpass images

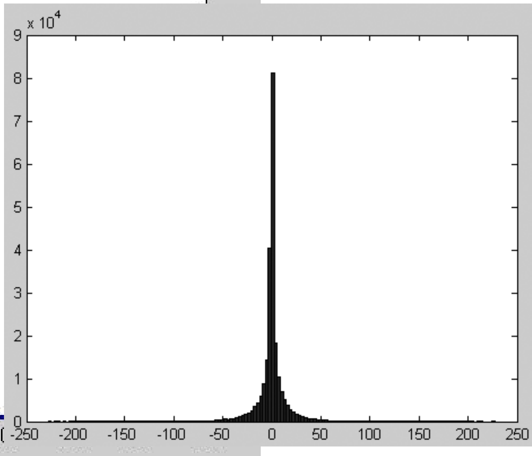








But want the bandpass image histogram to look like this



Bayesian MAP estimator for clean bandpass coefficient values

Let x = bandpassed image value before adding noise.
Let y = noise-corrupted observation.

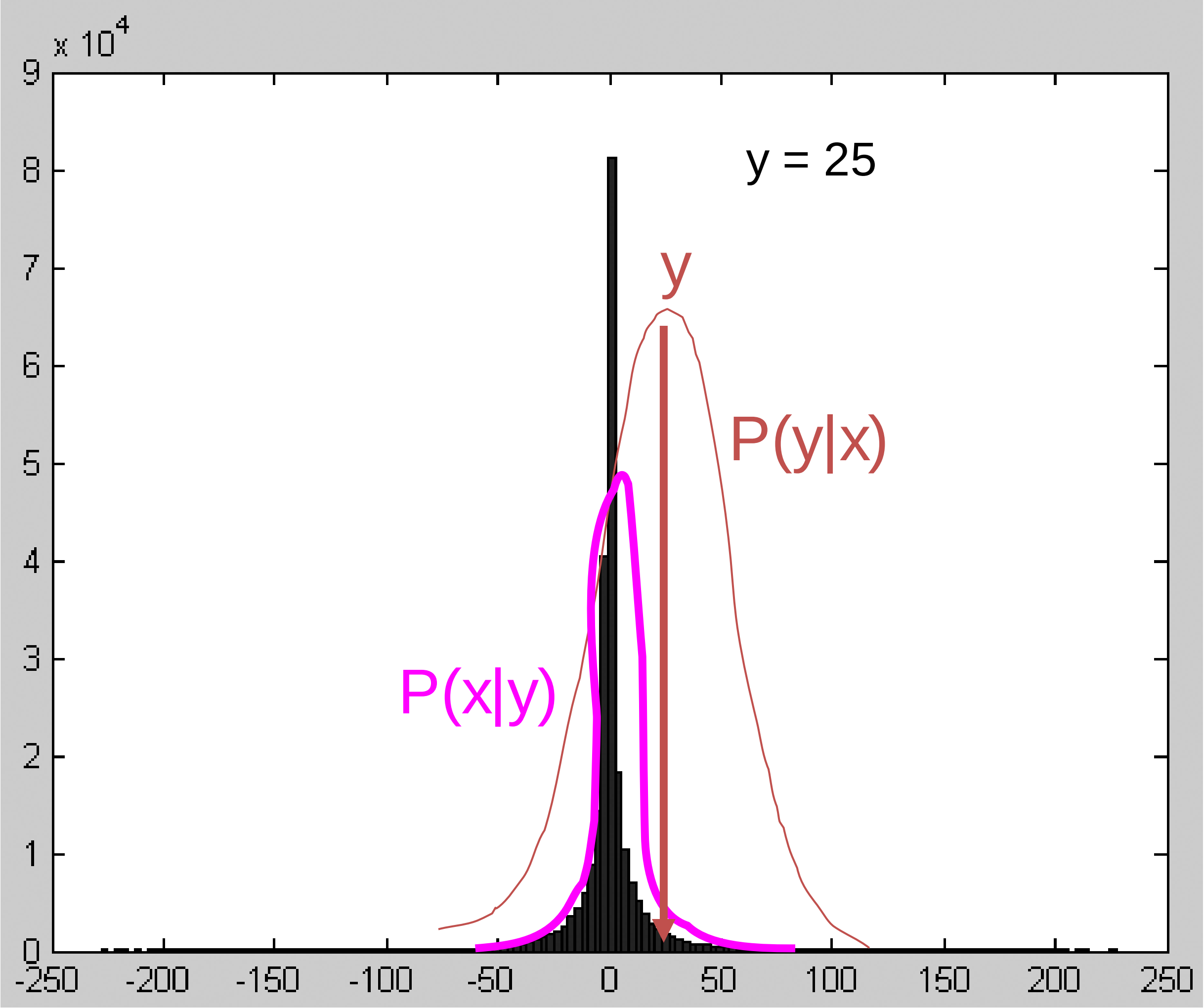
By Bayes theorem

P(x|y) = k P(y|x) P(x)

P(x)

P(y|x)

P(x|y)

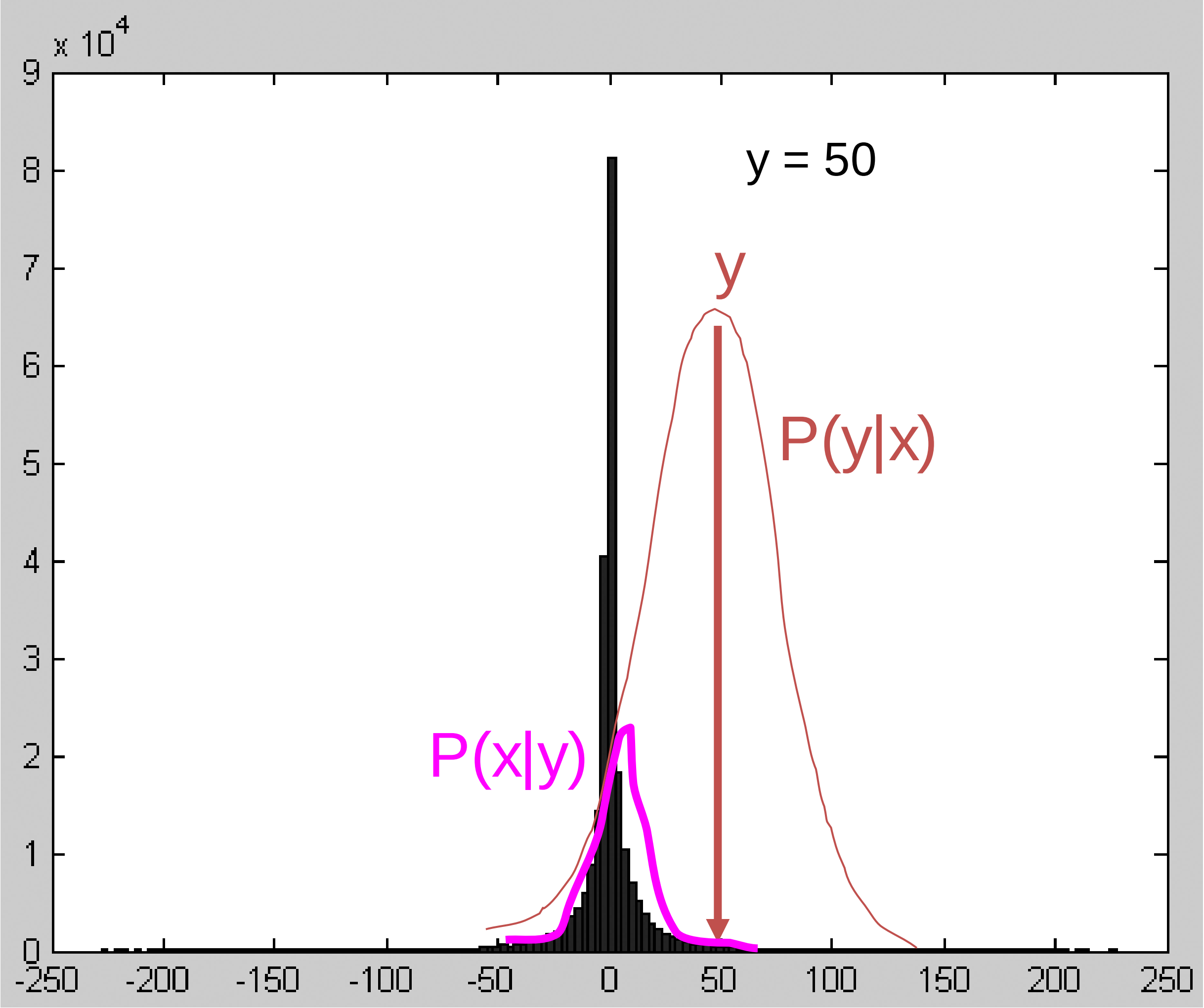


Bayesian MAP estimator

Let x = bandpassed image value before adding noise.
Let y = noise-corrupted observation.

By Bayes theorem

P(x|y) = k P(y|x) P(x)

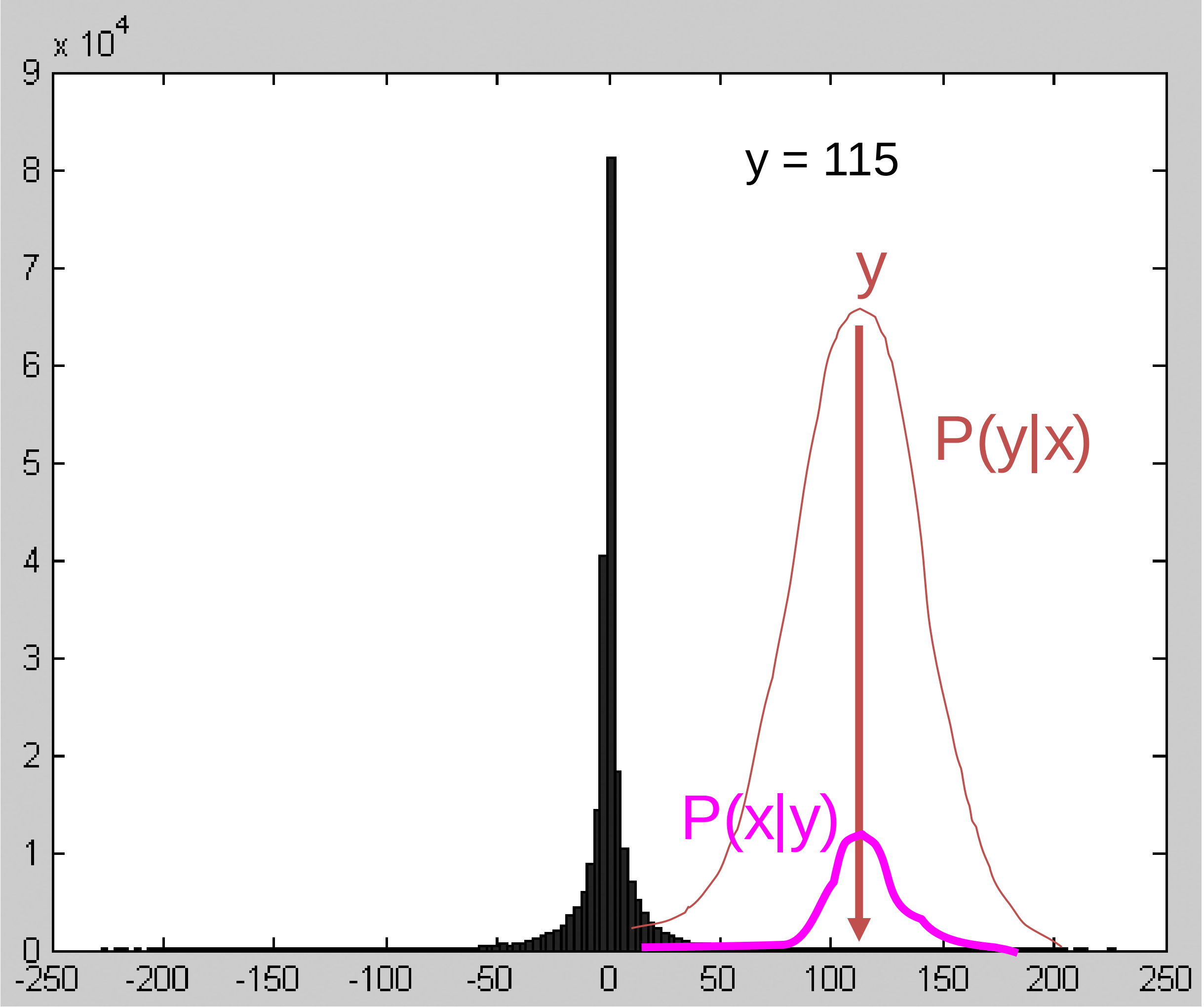


Bayesian MAP estimator

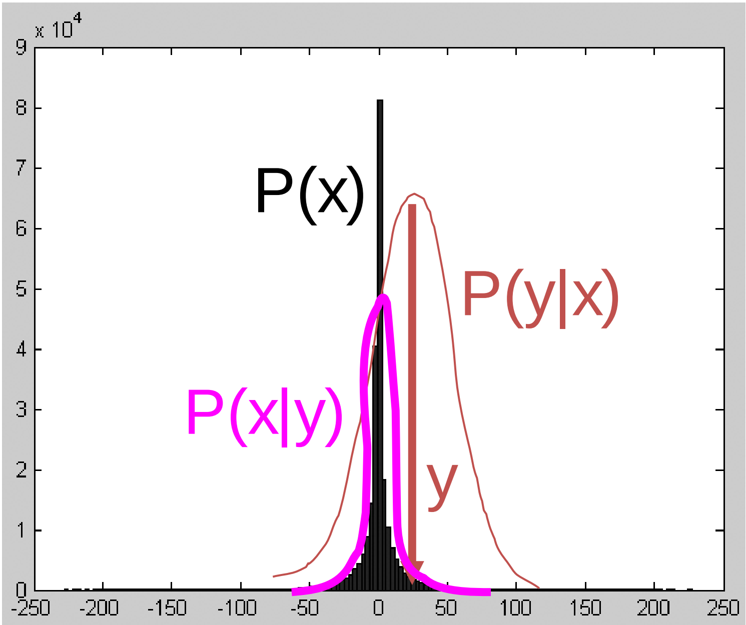
Let x = bandpassed image value before adding noise.
Let y = noise-corrupted observation.

By Bayes theorem

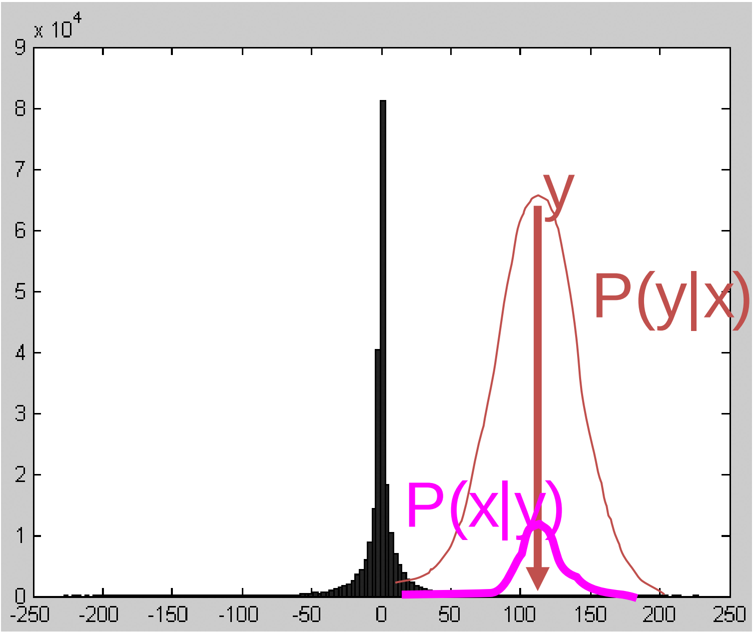
P(x|y) = k P(y|x) P(x)



y = 25



y = 115



For small y: probably it is due to noise and y should be set to 0
For large y: probably it is due to an image edge and it should be kept untouched

MAP estimate, $\hat{x}$, as function of observed coefficient value, y

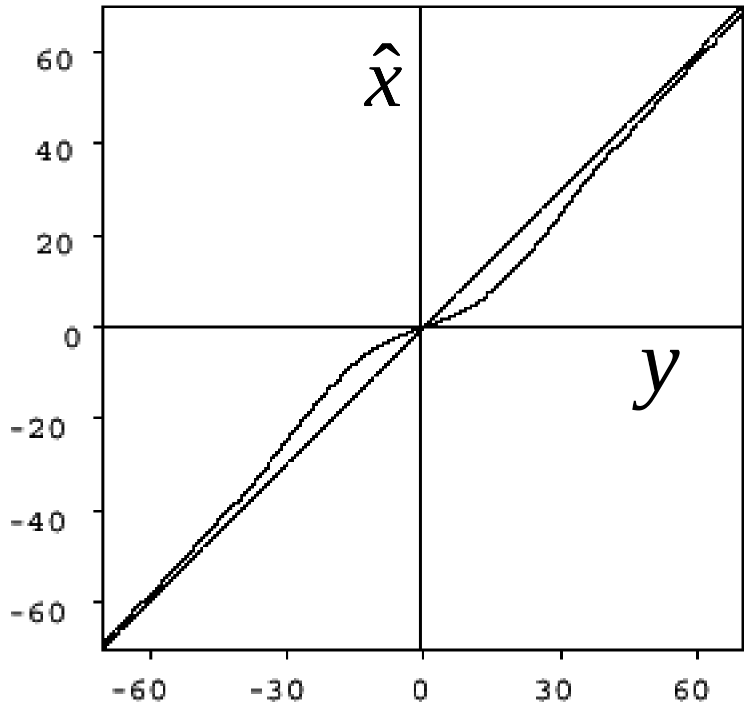
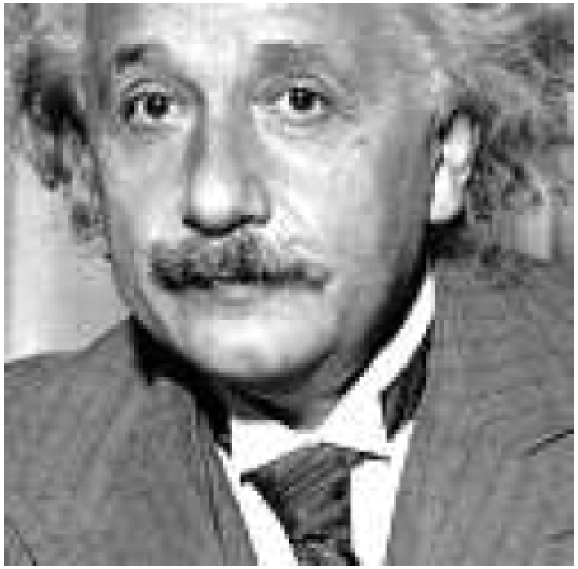


Figure 2: Bayesian estimator (symmetrized) for the signal and noise histograms shown in figure 1. Superimposed on the plot is a straight line indicating the identity function.

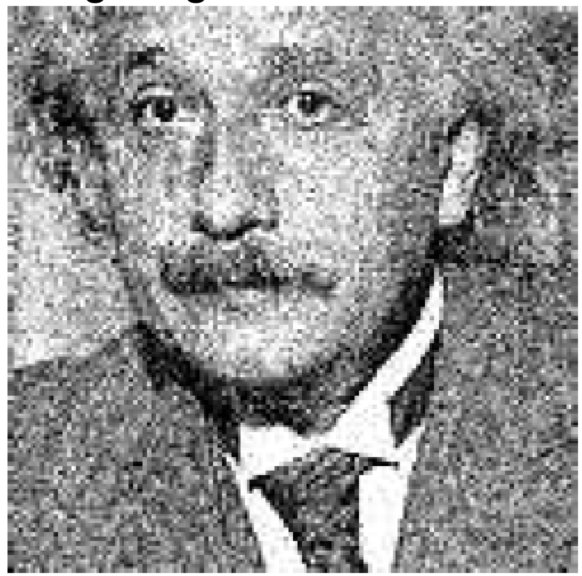
http://www-bcs.mit.edu/people/adelson/pub_pdfs/simoncelli_noise.pdf

Simoncelli and Adelson, Noise Removal via Bayesian Wavelet Coring

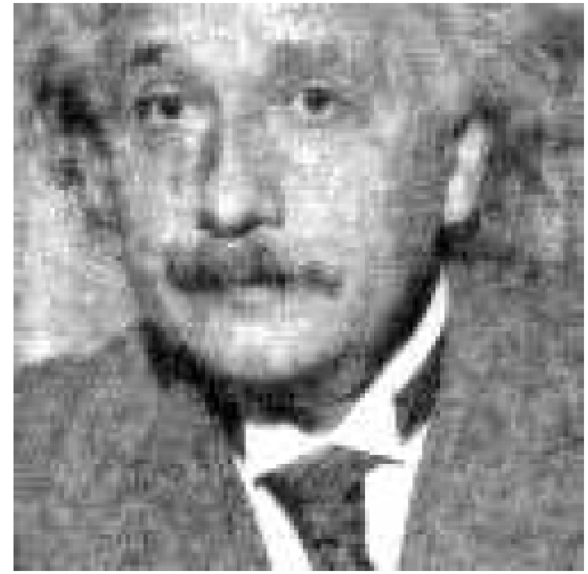
original

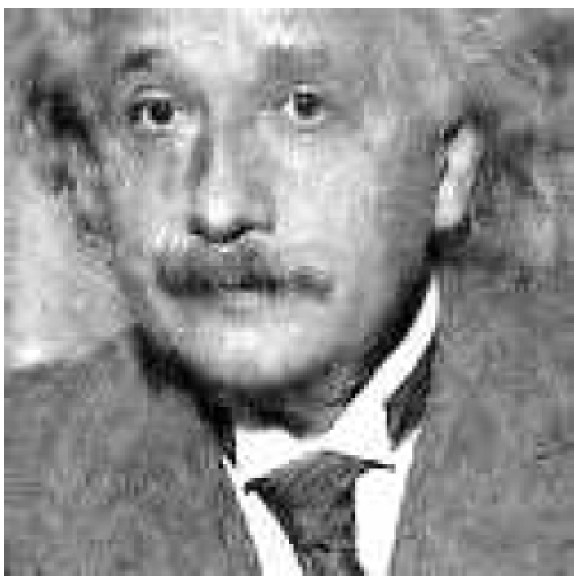


With Gaussian noise of std. dev. 21.4 added, giving PSNR=22.06



(1) Denoised with Gaussian model, PSNR=27.87





(2) Denoised with wavelet marginal model, PSNR=29.24

http://www.cns.nyu.edu/pub/eero/simoncelli05a-preprint.pdf