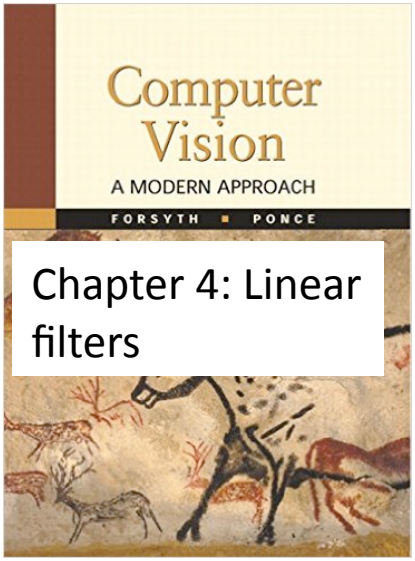
6.819 / 6.869: Advances in Computer Vision

MIT COMPUTER VISION

Basics of Image Processing II:

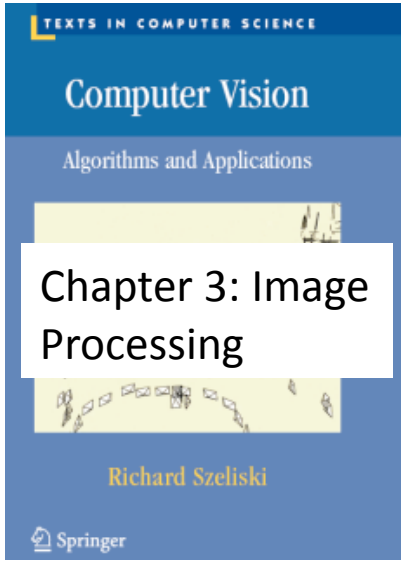
Multi-scale representations (pyramids, gabors, wavelets)



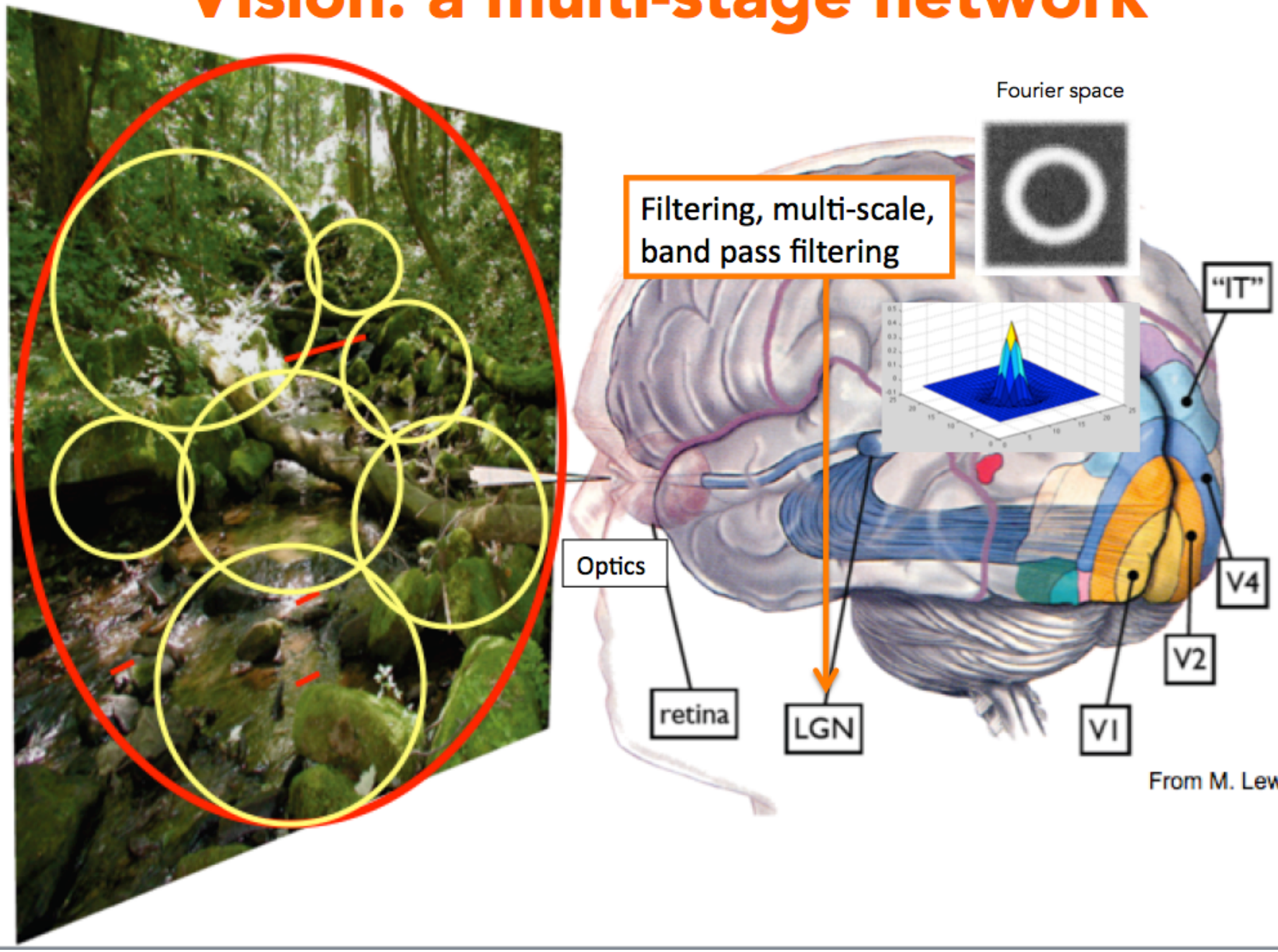
Instructor: Aude Oliva

Lecture TR 9:30AM – 11:00AM
(Room 34-101)

Website:
http://6.869.csail.mit.edu/fa15/

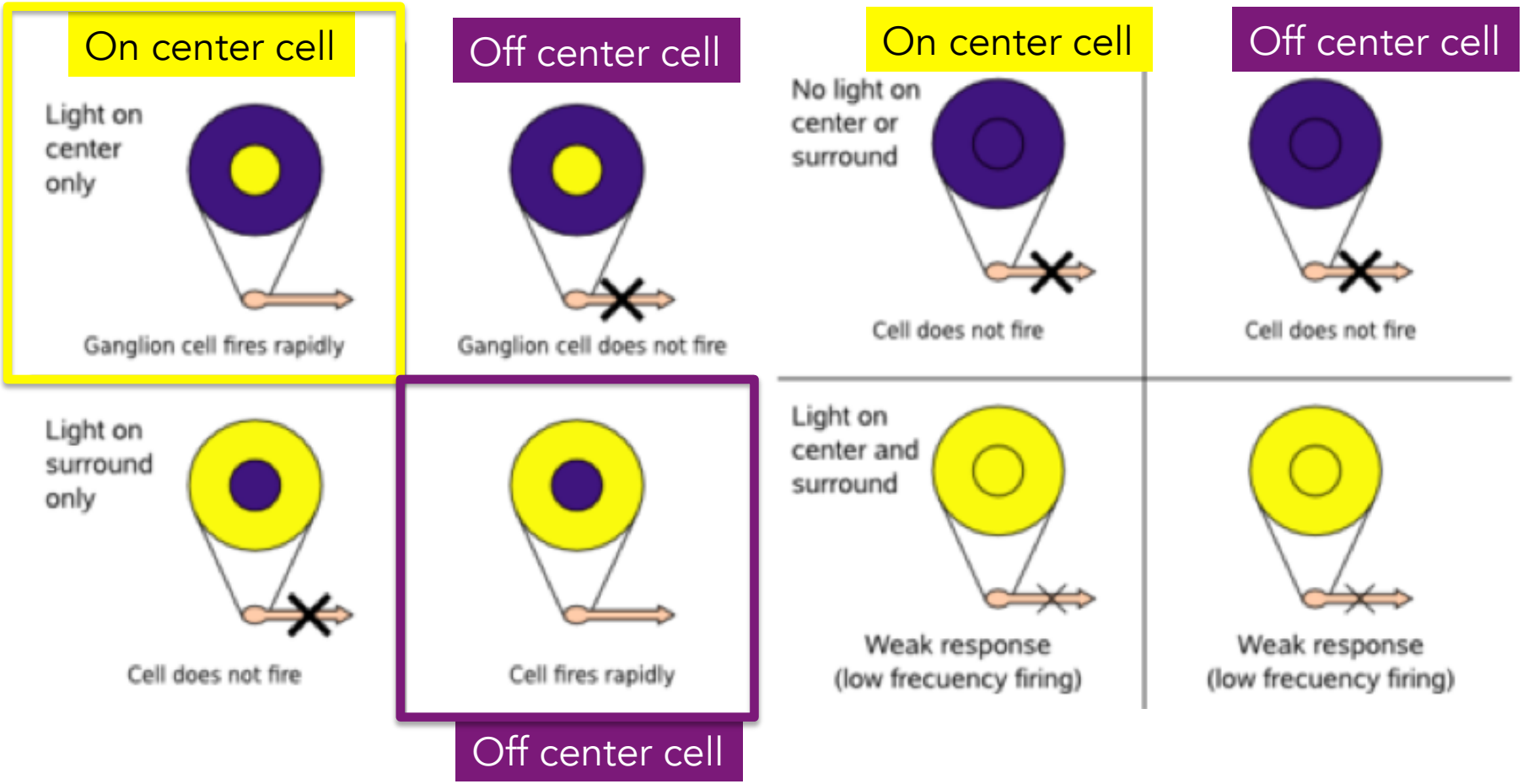


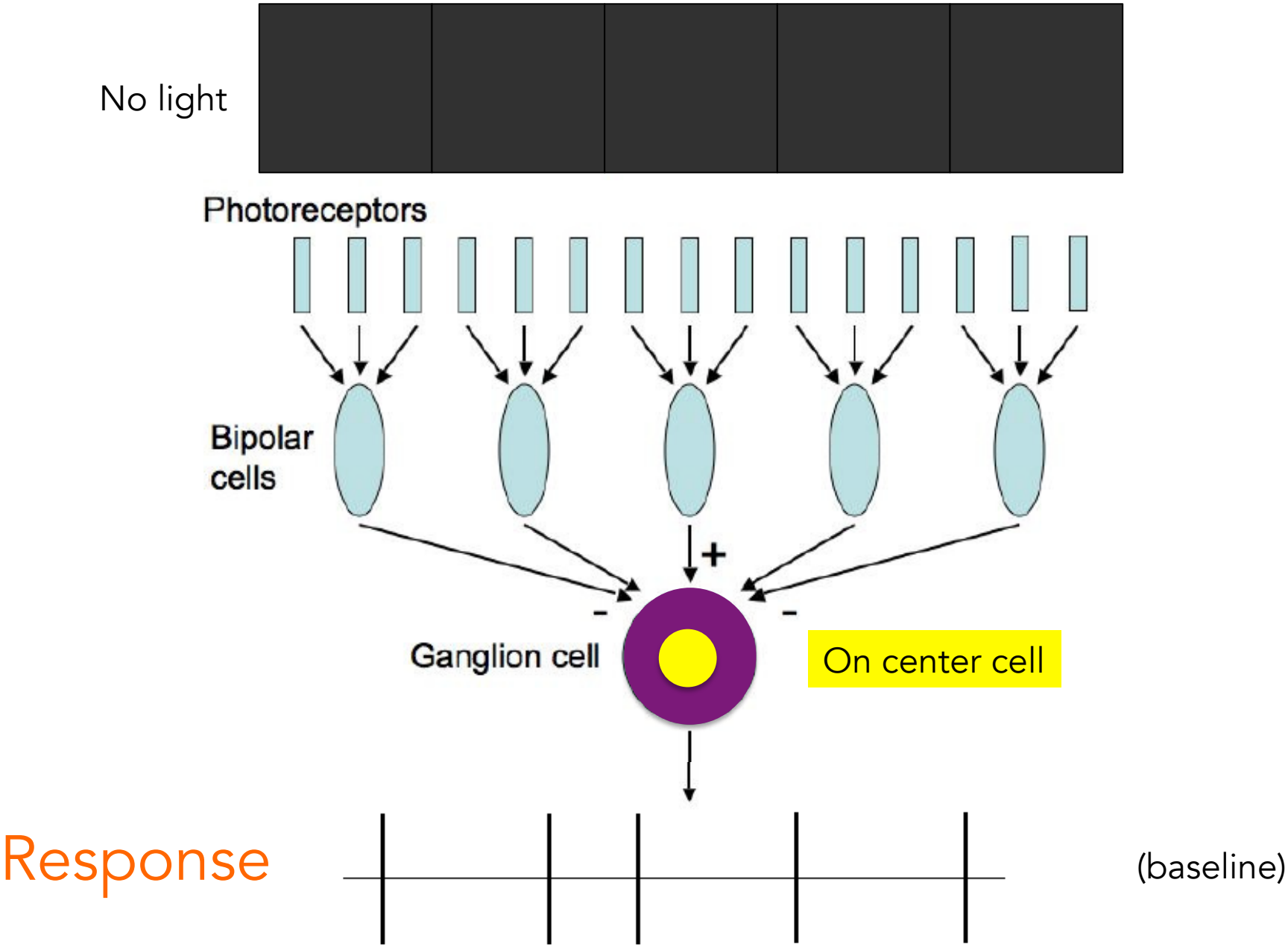
Vision: a multi-stage network

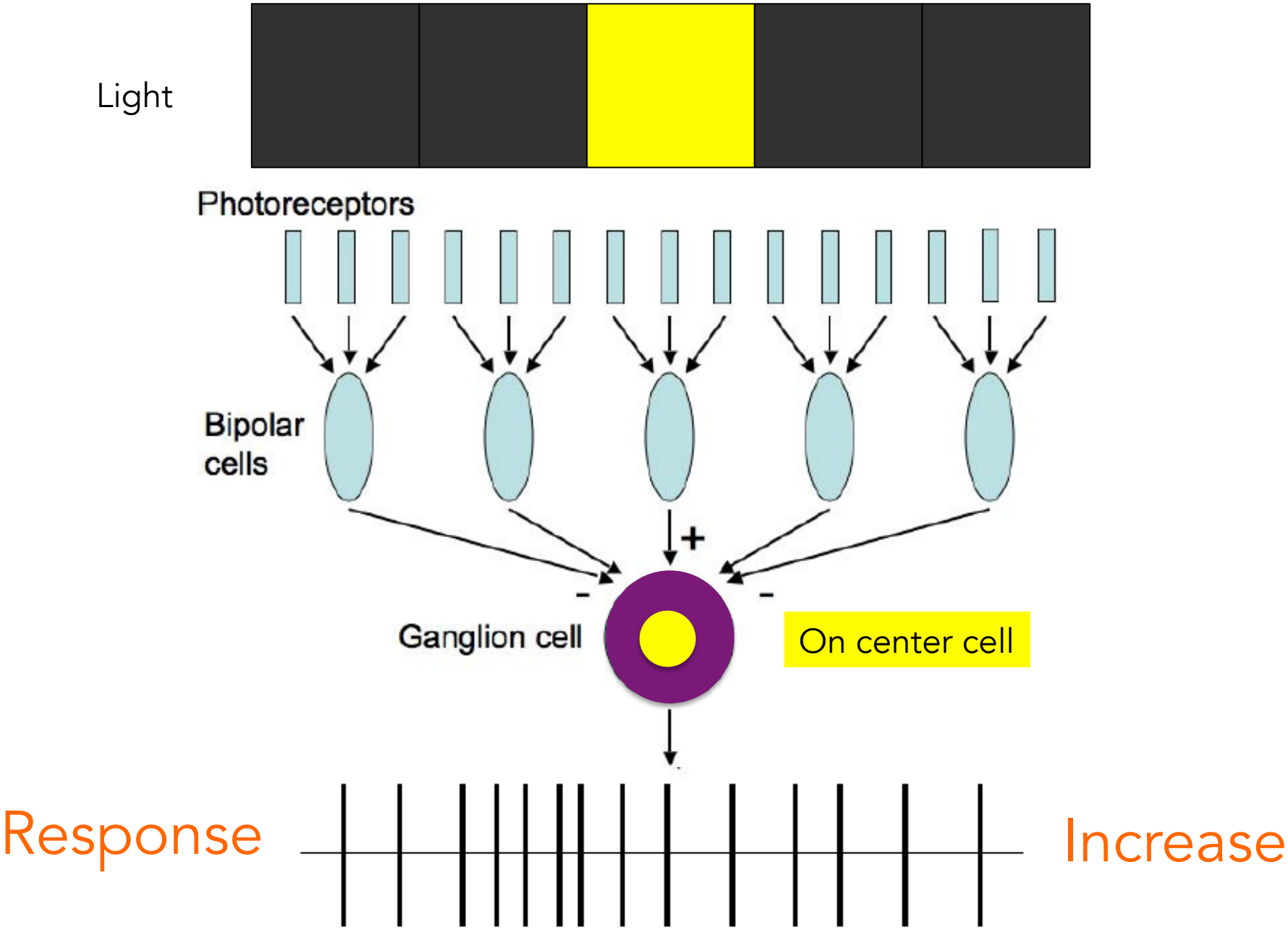


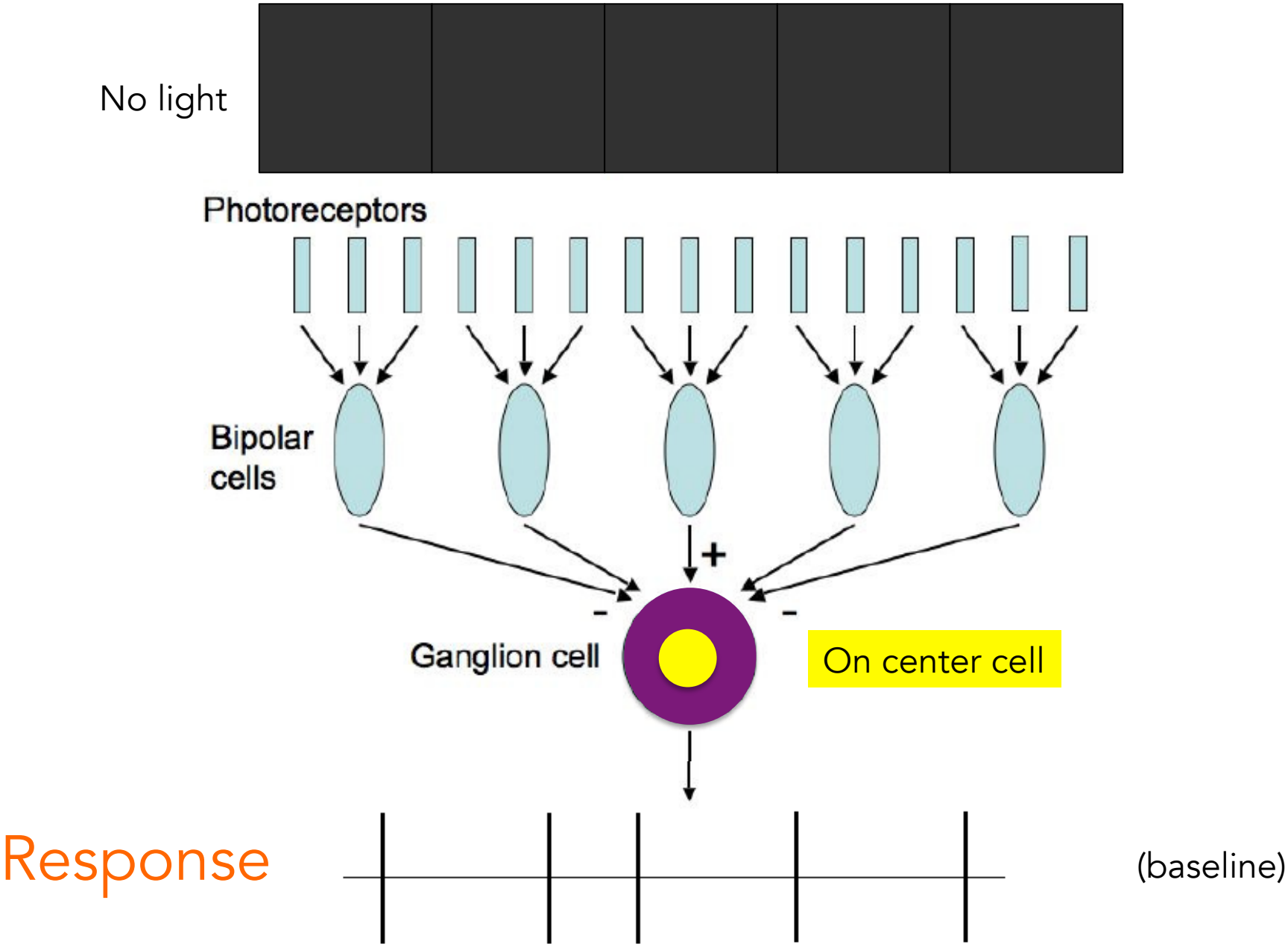
From M. Lewicky

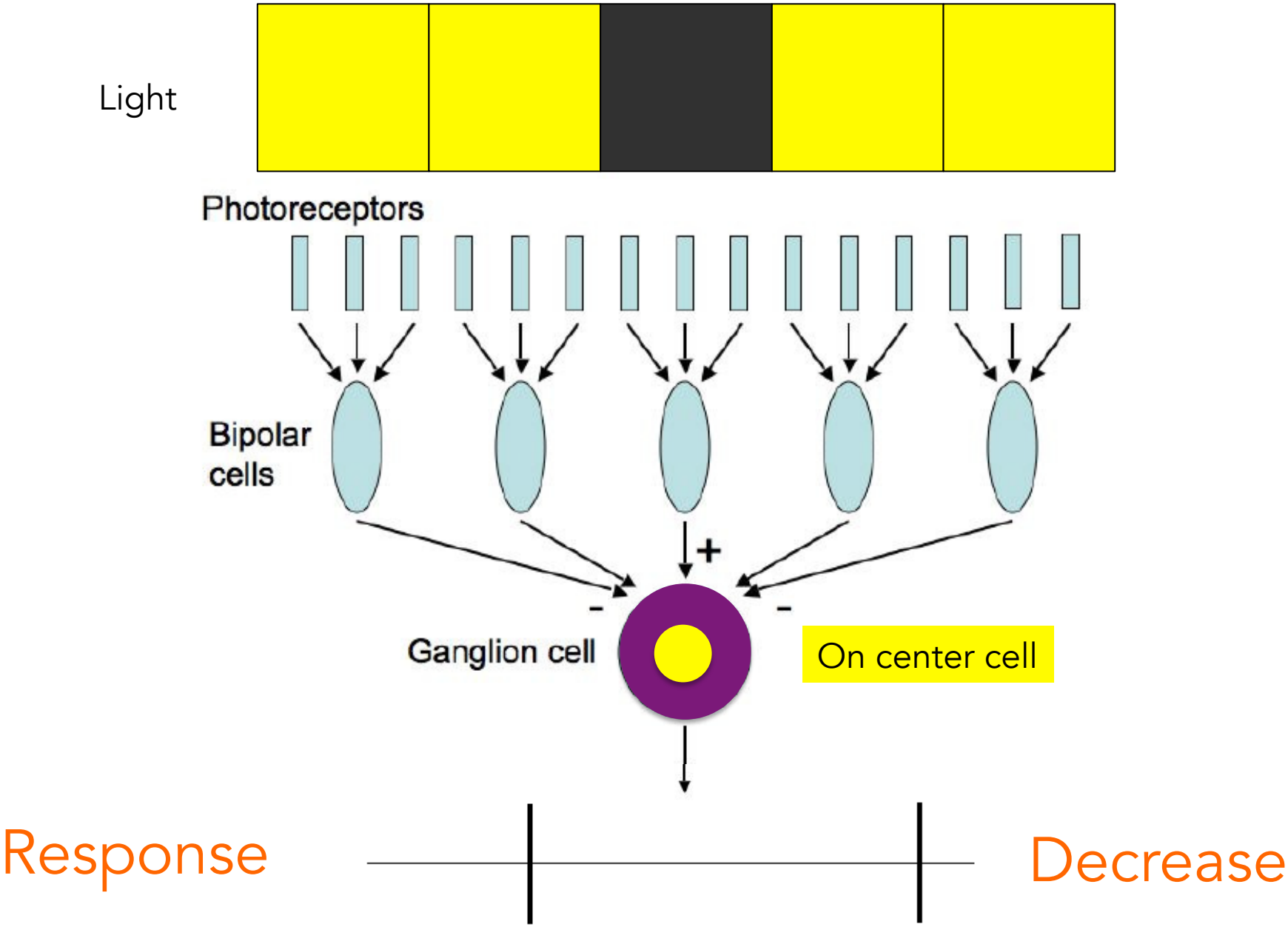
Model of Retinal and LGN cells

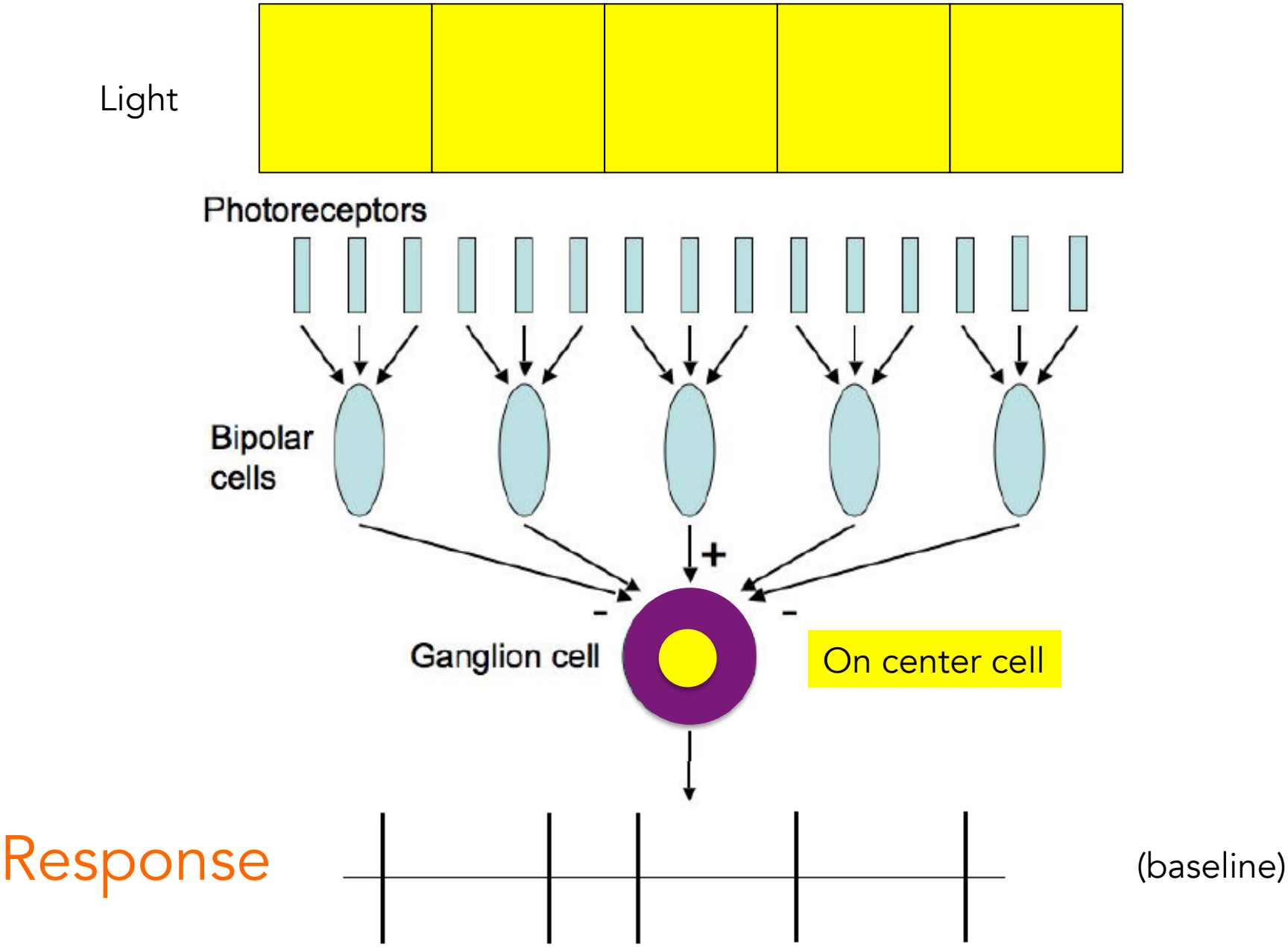



No light
Photoreceptors
Bipolar cells
+
-
-
Ganglion cell
On center cell
Response
(baseline)

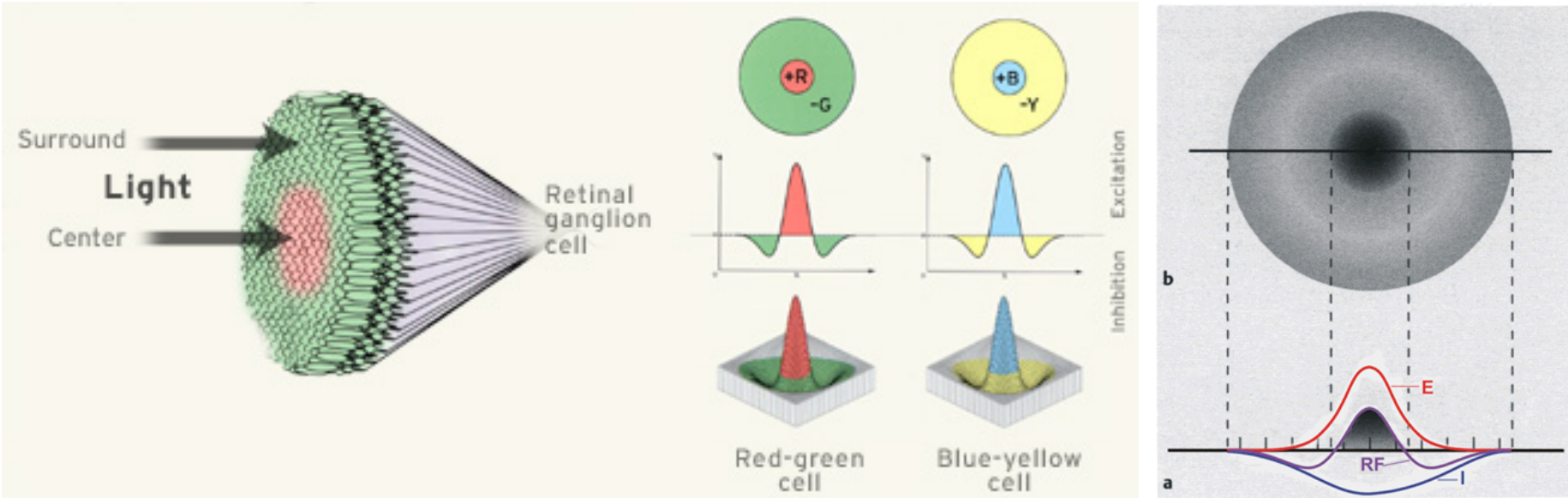

Light
Photoreceptors
Bipolar cells
+
-
-
Ganglion cell
On center cell
Response
Increase


No light
Photoreceptors
Bipolar cells
+
-
-
Ganglion cell
On center cell
Response
(baseline)


Light
Photoreceptors
Bipolar cells
+
-
-
Ganglion cell
On center cell
Response
Decrease

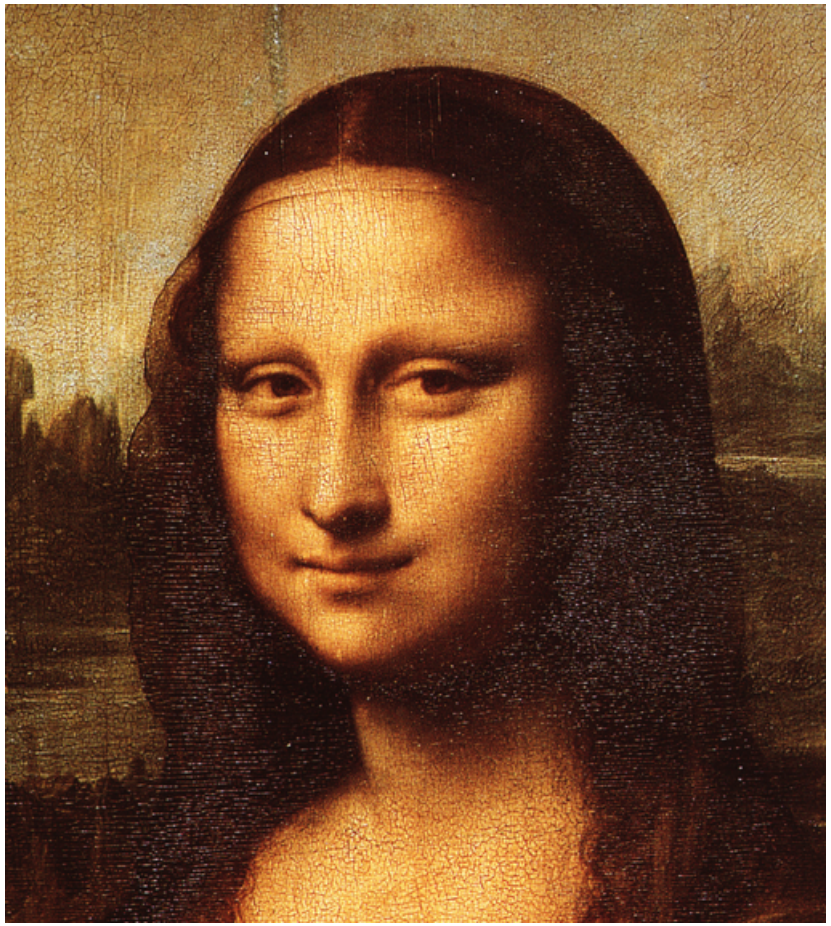

Light
Photoreceptors
Bipolar cells
+
-
-
Ganglion cell
On center cell
Response
(baseline)

Difference of Gaussian Model



- The output of a retinal ganglion cell is a weighted sum of its inputs. You can treat the cell as a shift-invariant linear system
- **Difference of Gaussian (DoG):** A model of the operation performed by retinal and ganglion cells (LGN). The DoG model supposes that the neural response results from the combined signal of two separates mechanisms
- LGN and retinal neurons have circular receptive fields: **they respond equally well to all stimulus orientations, and at different spatial frequency (scale)**
- Multiple DoG is called a **Laplacian Pyramid**

Subtle expression
Leonardo Da Vinci's Mona Lisa



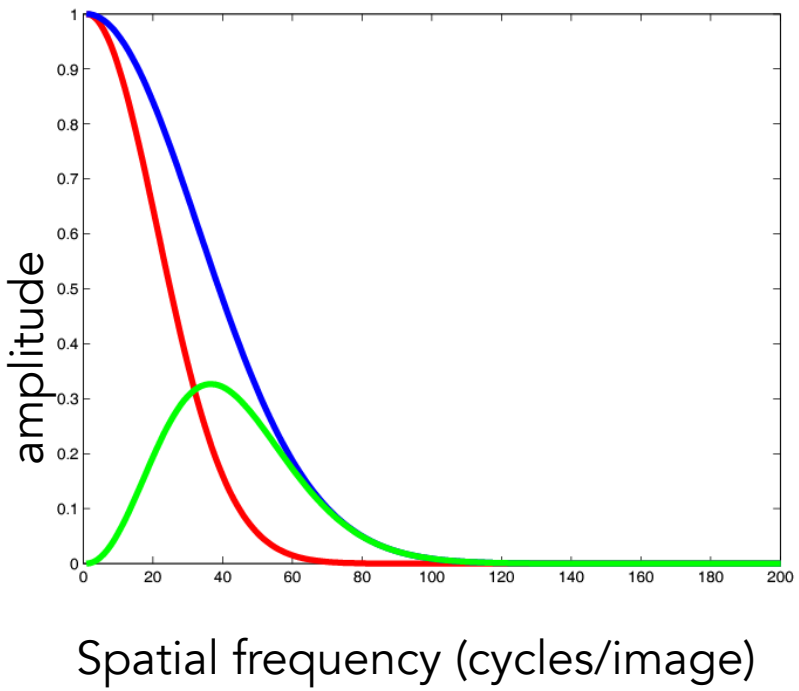
Mona Lisa Painting – Da Vinci

Difference of Gaussian – DoG

band-pass filter

SubbandDoGLaplacian/subbanddecomposition.m
Note: second part of the code

Gaussians and DoG

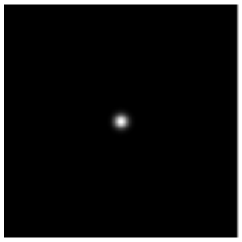


Red
(low pass filter)

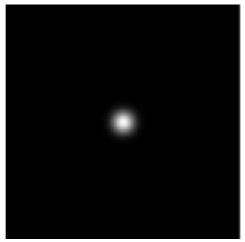
Blue
(low pass filter)

Green
= R - B
(band pass filter)

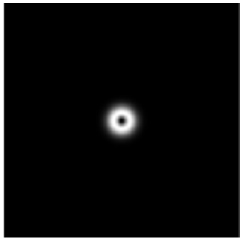
Red filter
Spectrum

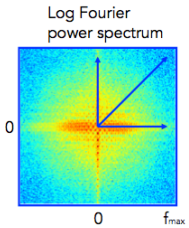


Blue filter
Spectrum



Green filter
Spectrum

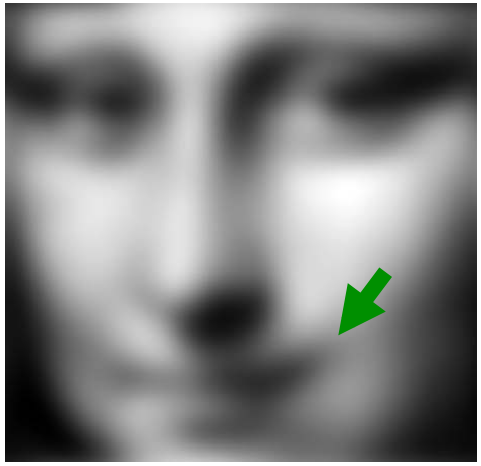




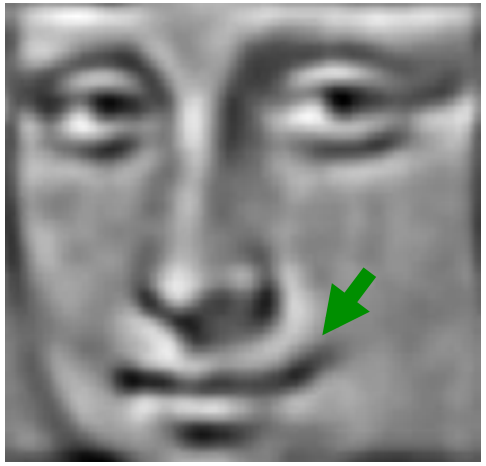
Subtle expression

Leonardo da Vinci's Mona Lisa

Smile



Smile



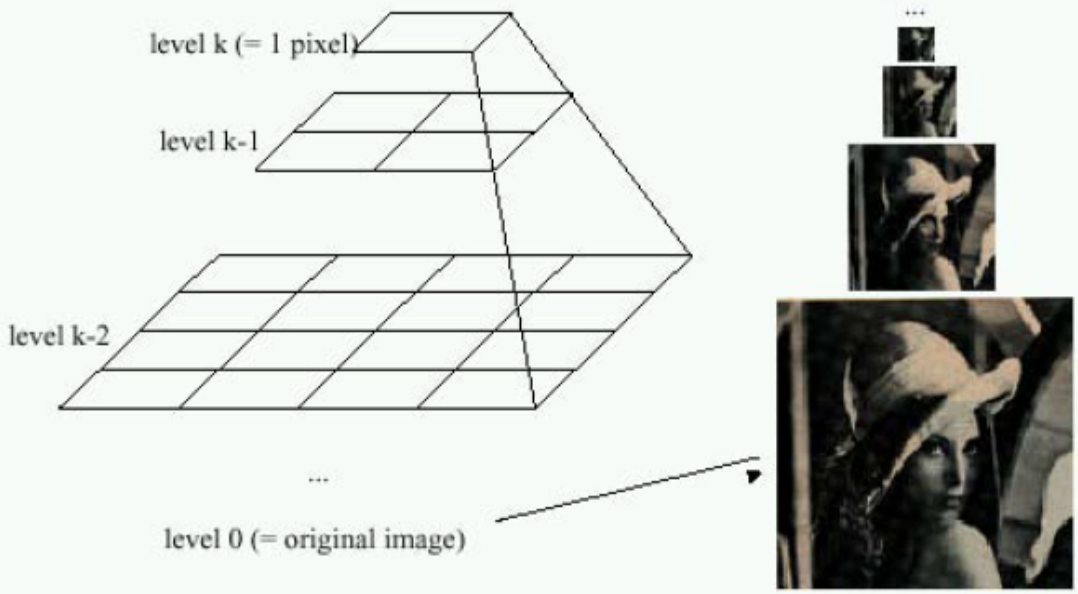
No smile

Low — Spatial frequency — High

Margaret Livingstone

Image Pyramids

Idea: Represent NxN image as a "pyramid" of 1x1, 2x2, 4x4,…, 2^kx2^k images (assuming N=2^k)



Known as a **Gaussian Pyramid** [Burt and Adelson, 1983]

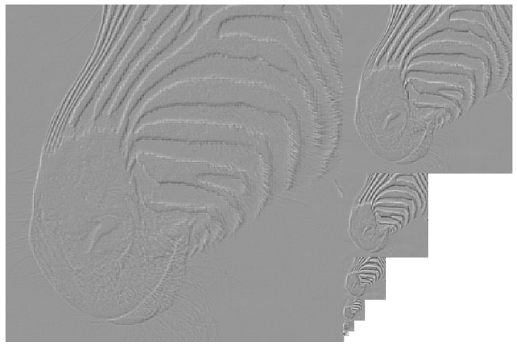
- In computer graphics, a *mip map* [Williams, 1983]
- A precursor to *wavelet transform*

Image pyramids

- Gaussian

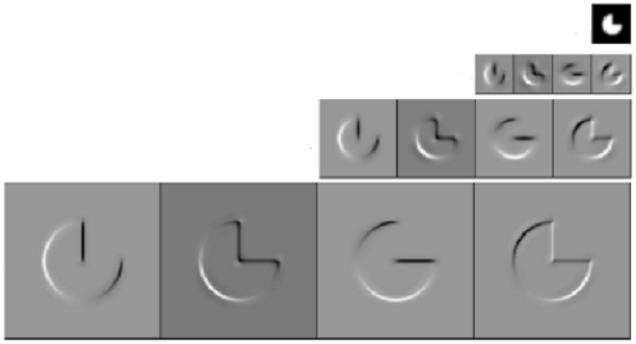
Progressively blurred and subsampled versions of the image. Adds scale invariance to fixed-size algorithms.

- Laplacian



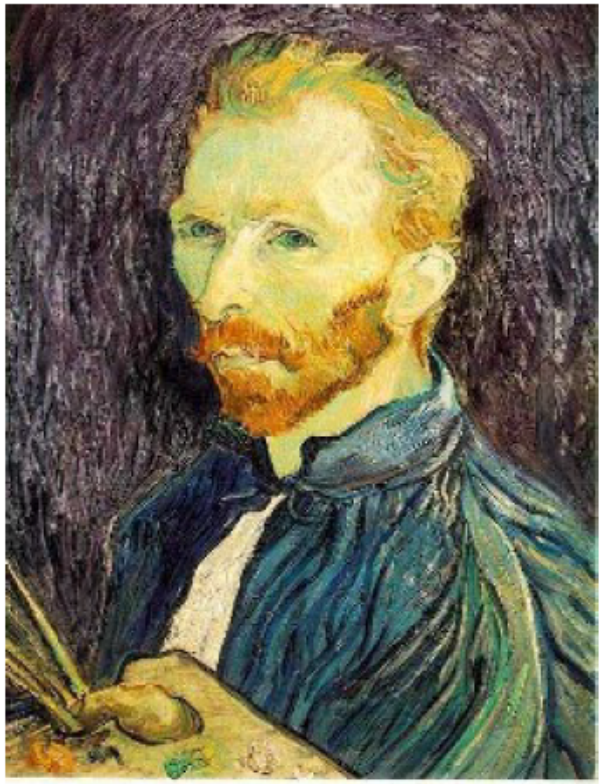
Shows the information added in Gaussian pyramid at each spatial scale. Useful for noise reduction & coding.

- Steerable pyramid



Shows components at each scale and orientation separately. Non-aliased subbands. Good for texture and feature analysis.

Image sub-sampling

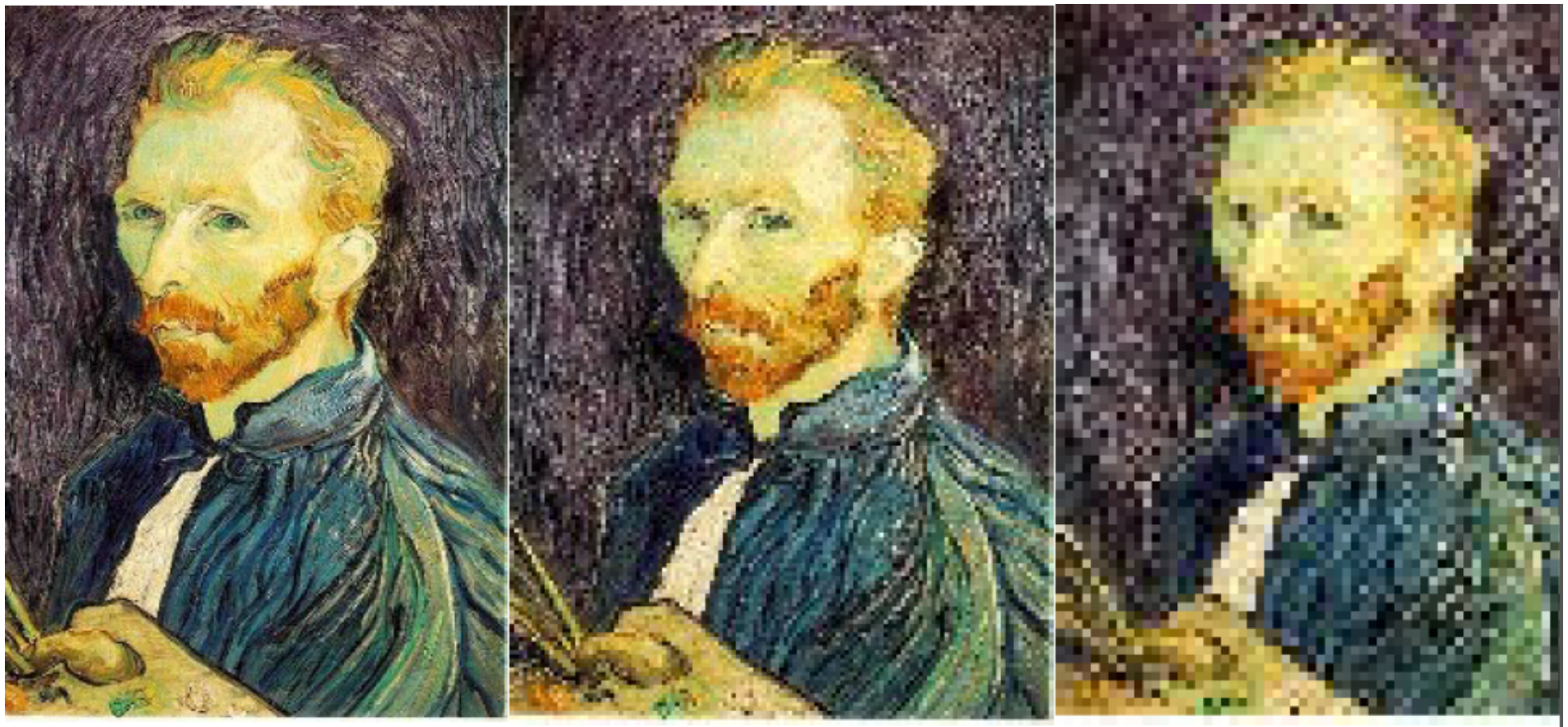


1/4

1/8

Throw away every other row and column to create a 1/2 size image
- called *image sub-sampling*

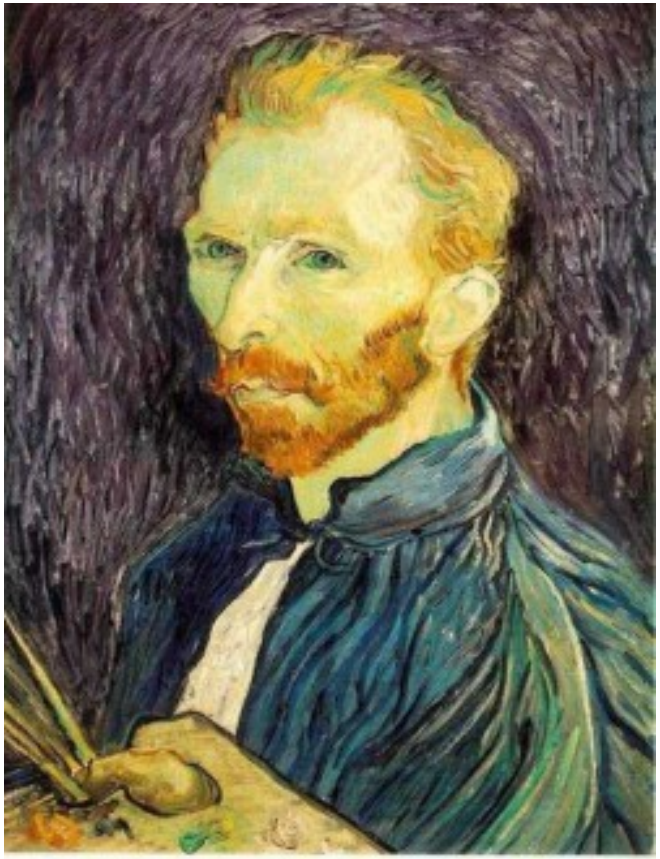
Image sub-sampling



1/2 1/4 (2x zoom) 1/8 (4x zoom)

Why does this look so bad?

Gaussian (lowpass) pre-filtering



Gaussian 1/2

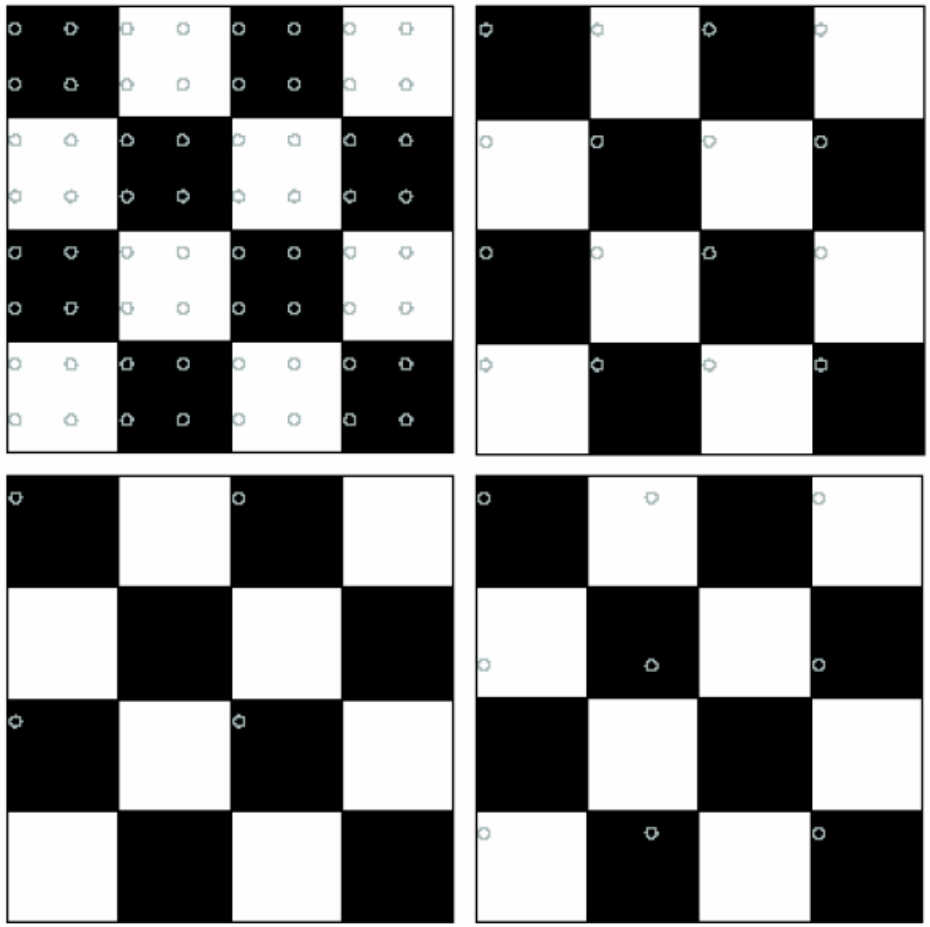
G 1/4

G 1/8

Solution: filter the image, *then* subsample

Slide by Steve Seitz

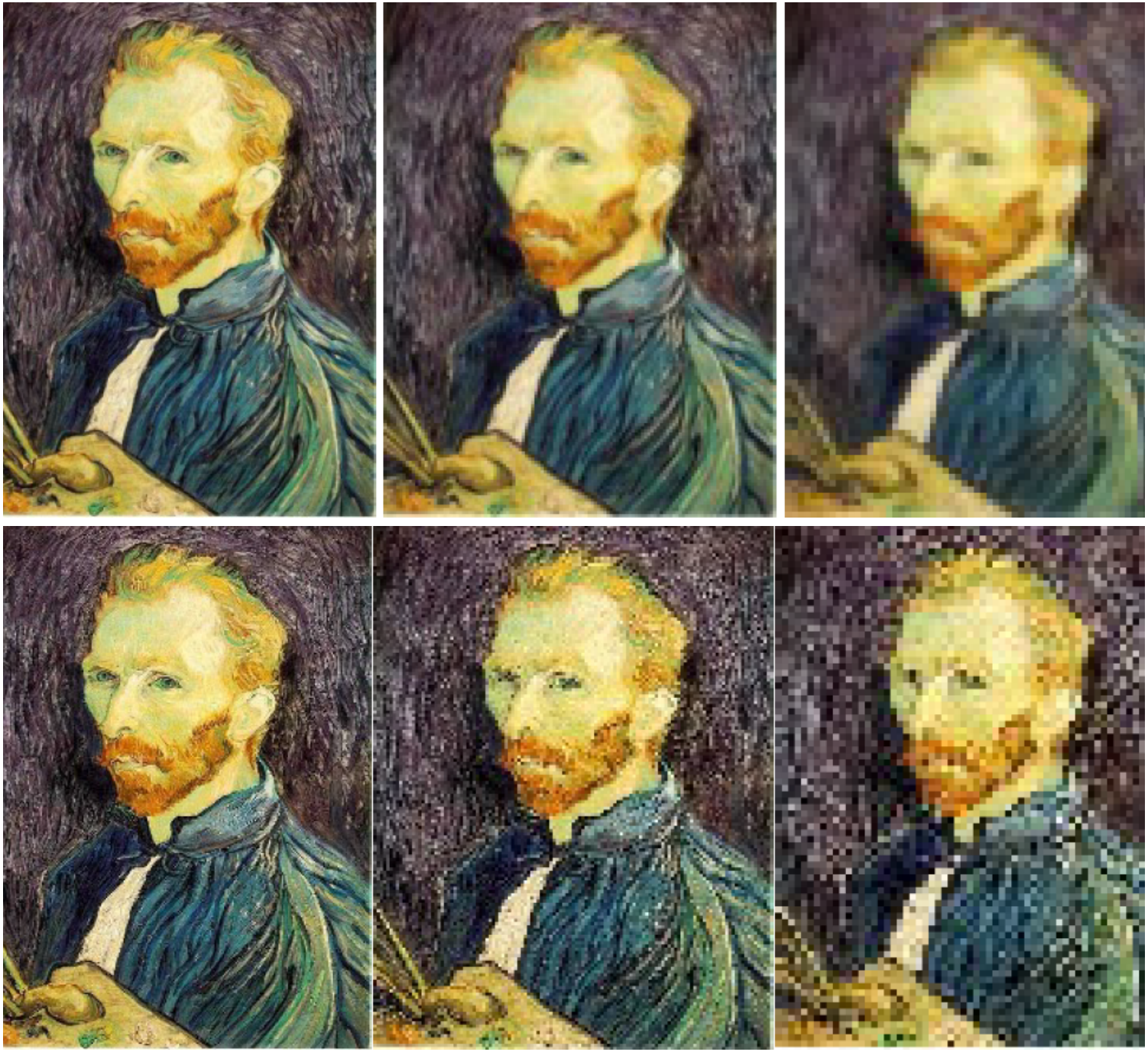
Sampling



Good sampling:
- Sample often or,
- Sample wisely

Bad sampling:
- see aliasing in action!

The two checkerboards on the top illustrate a sampling procedure that appears to be successful: the grey circle represent the samples: if they are sufficient, they represent the detail of the images. The bottom is unsuccessful: the sampling suggest they are fewer checks that there are. This illustrates 1) **successful sampling sample data often enough**; 2) unsuccessful sampling cause high spatial frequency information to appear as lower frequency information



512 256 128 64 32 16 8

Each level has half the resolution and the quarter of the pixels

Gaussian pyramid

The Gaussian pyramid:
Burt & Adelson (1983)

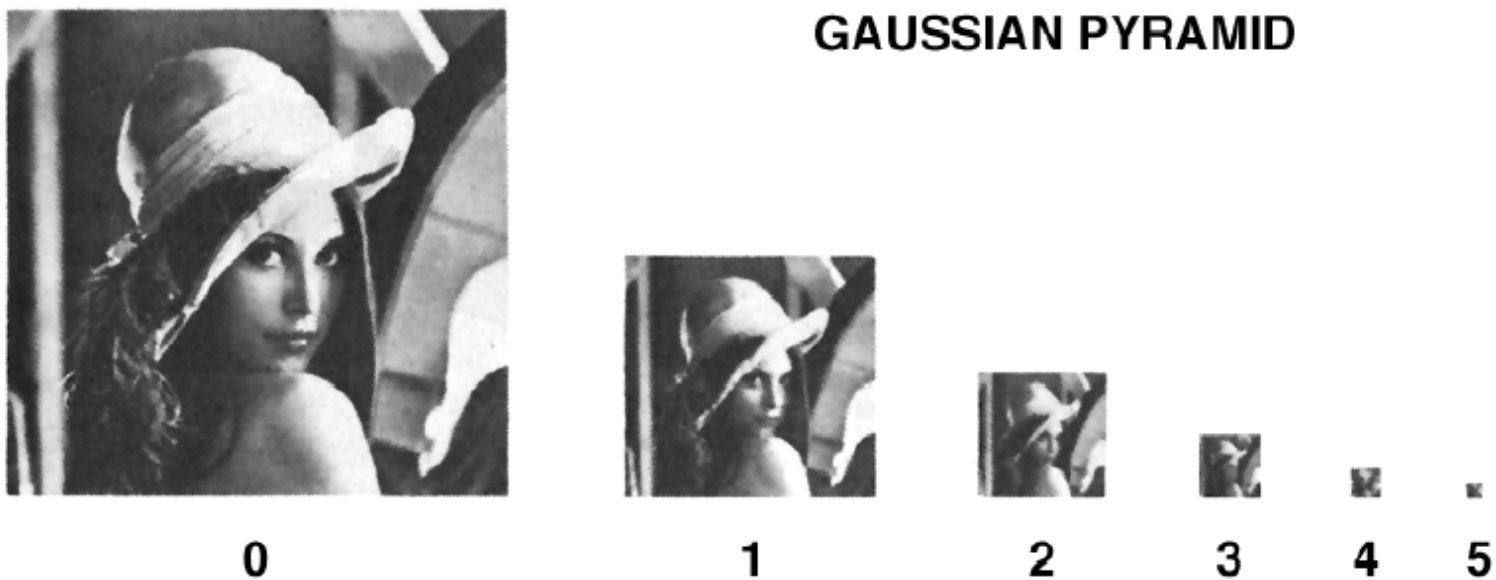


Fig. 4. First six levels of the Gaussian pyramid for the "Lady" image The original image, level 0, meusures 257 by 257 pixels and each higher level array is roughly half the dimensdons of its predecessor. Thus, level 5 measures just 9 by 9 pixels.

http://persci.mit.edu/pub_pdfs/pyramid83.pdf

IEEE TRANSACTIONS ON COMMUNICATIONS, VOL. COM-31, NO. 4, APRIL 1983

Gaussian pyramids used for

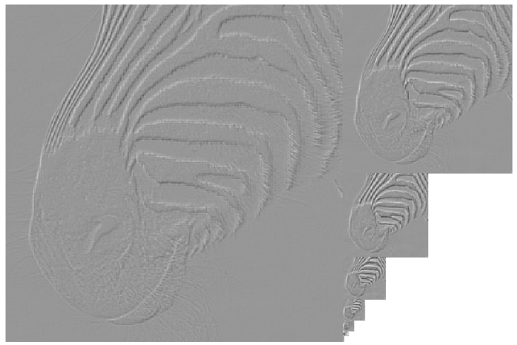
- up- or down- sampling images.
- Multi-resolution image analysis
  - Look for an object over various spatial scales
  - Coarse-to-fine image processing: form blur estimate or the motion analysis on very low-resolution image, upsample and repeat. Often a successful strategy for avoiding local minima in complicated estimation tasks.

Image pyramids

- Gaussian

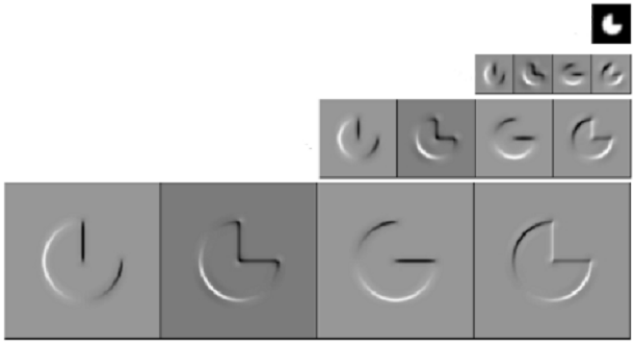
Progressively blurred and subsampled versions of the image. Adds scale invariance to fixed-size algorithms.

- Laplacian



Shows the information added in Gaussian pyramid at each spatial scale. Useful for noise reduction & coding.

- Steerable pyramid



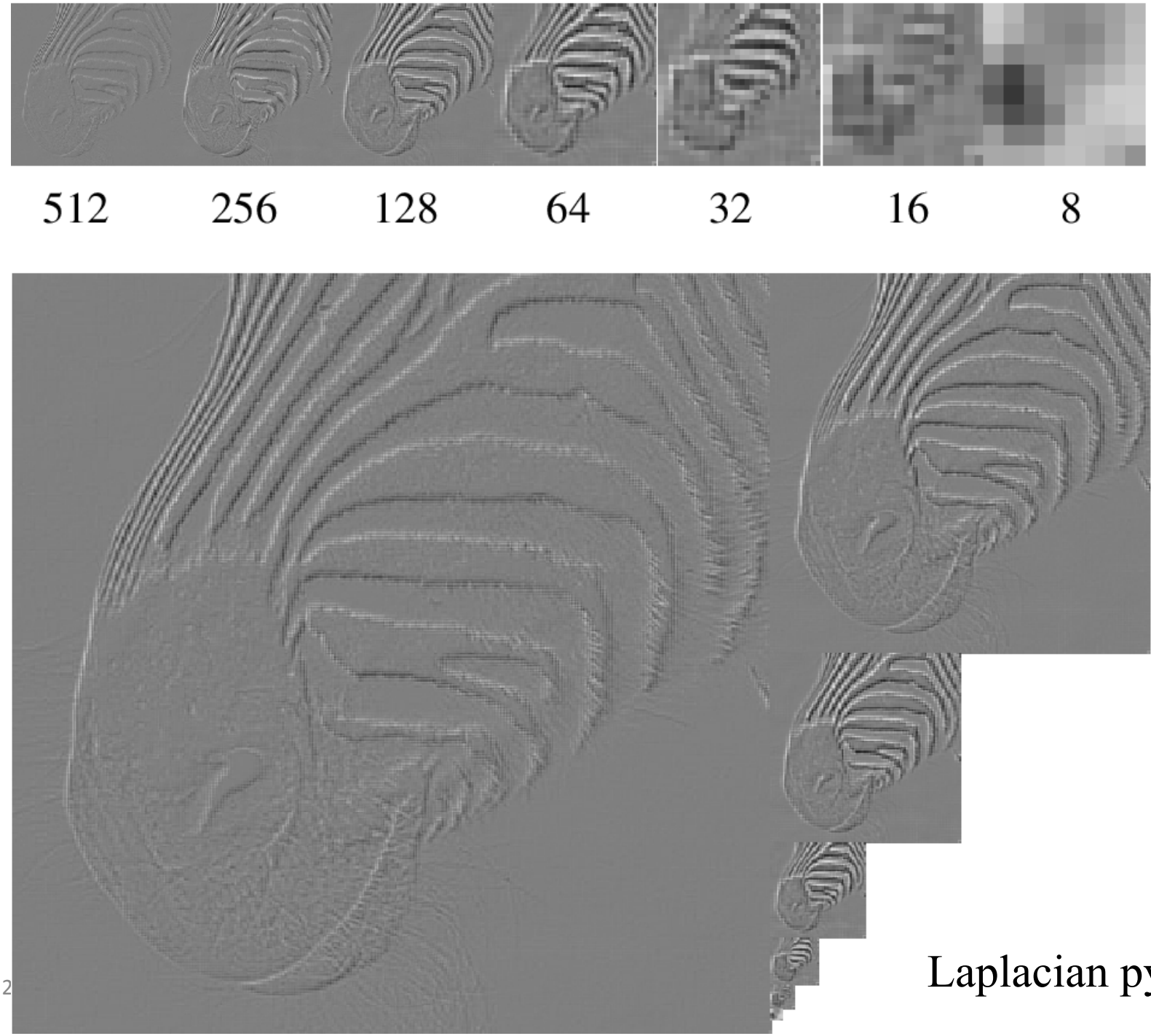
Shows components at each scale and orientation separately. Non-aliased subbands. Good for texture and feature analysis.

The Laplacian Pyramid

- Synthesis
  - Compute the difference between upsampled Gaussian pyramid level and Gaussian pyramid level.
  - band pass filter - each level represents spatial frequencies (largely) unrepresented at other level.

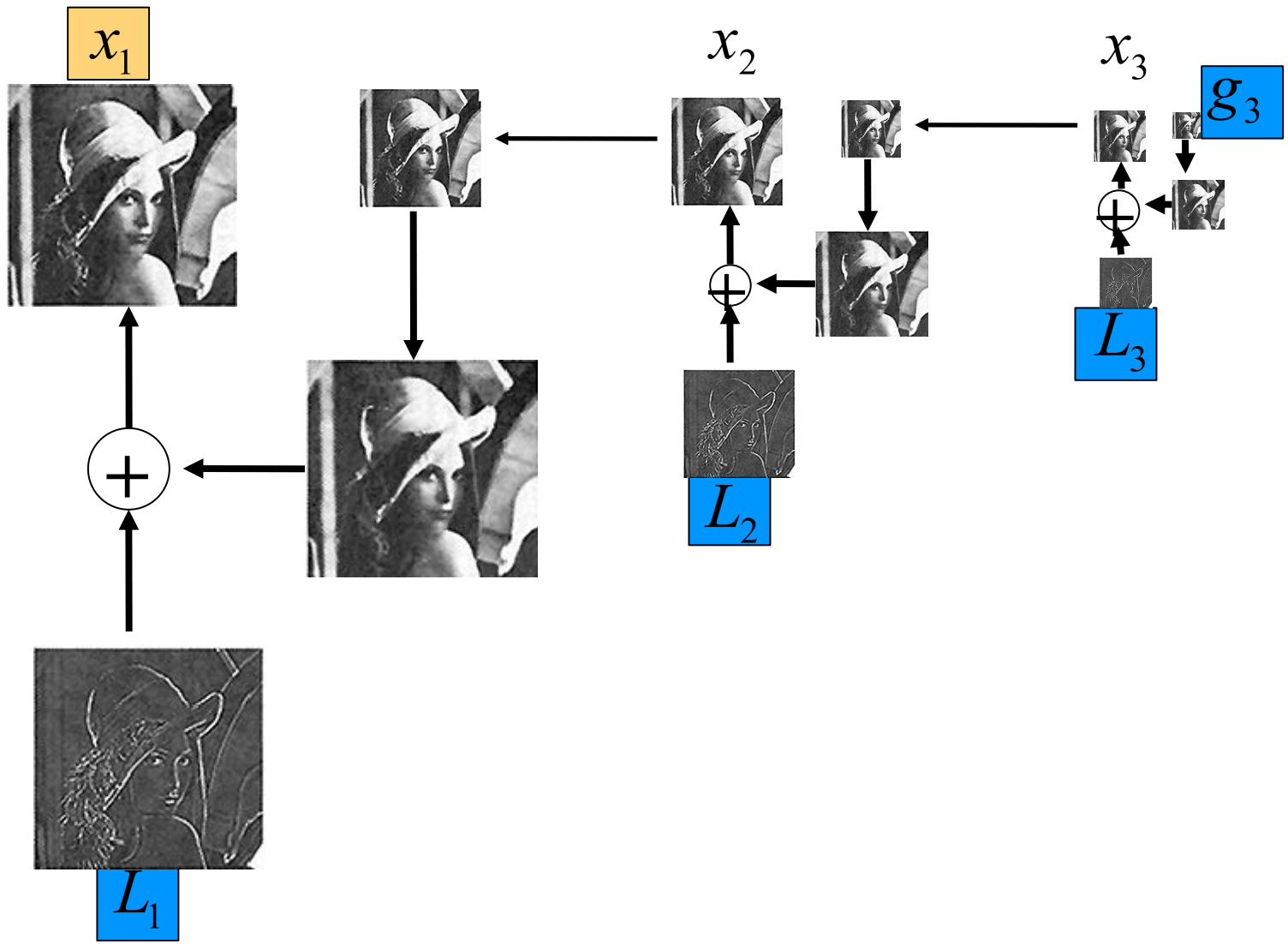
- Applications:

Texture synthesis; Image compression; Noise removal; Image blending

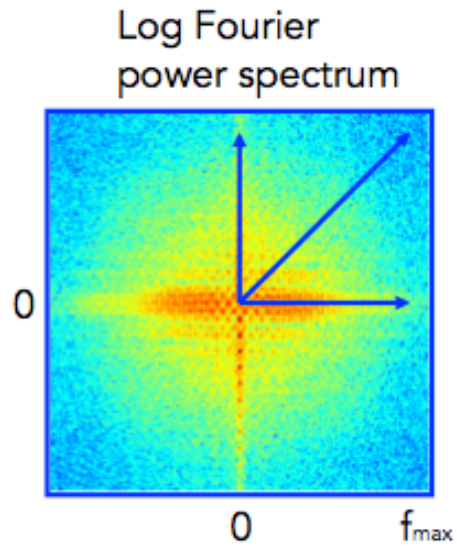


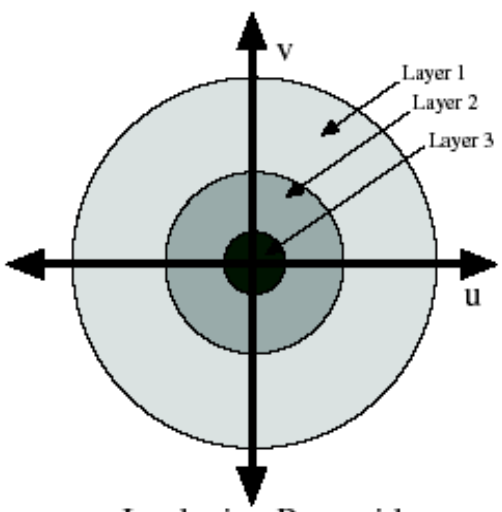
Laplacian pyramid

Laplacian pyramid reconstruction algorithm: recover x_1 from L_1, L_2, L_3 and g_3



Laplacian Pyramid in spectral (fourier) domain



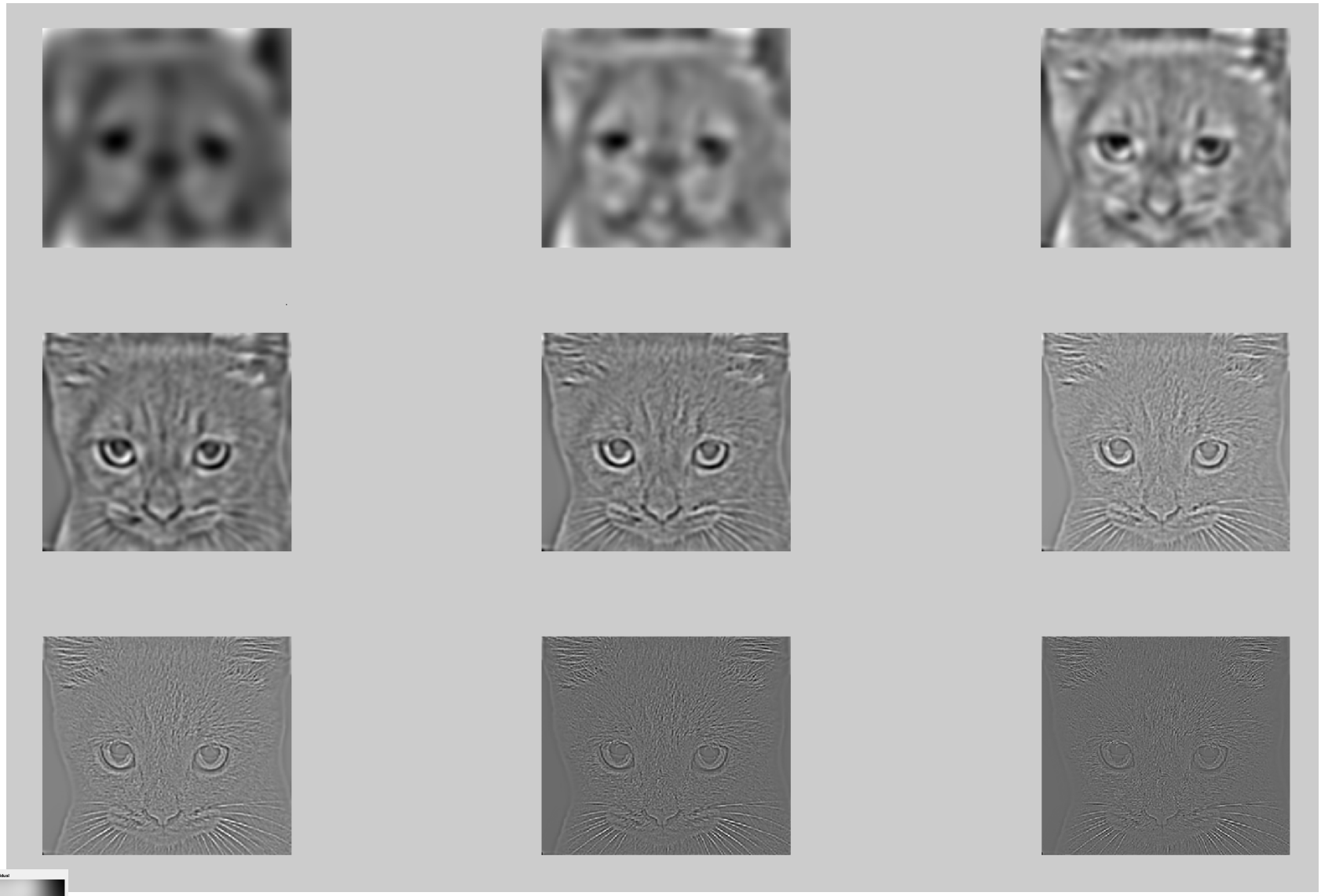


Laplacian Pyramid

- Each layer of a laplacian pyramid consists in the elements of a smoothed and resampled image.
- **Band pass filters** - each level represents spatial frequencies (largely) unrepresented at other levels
- The fourier transform of each layer is an annulus
- Laplacian pyramid is orientation independent

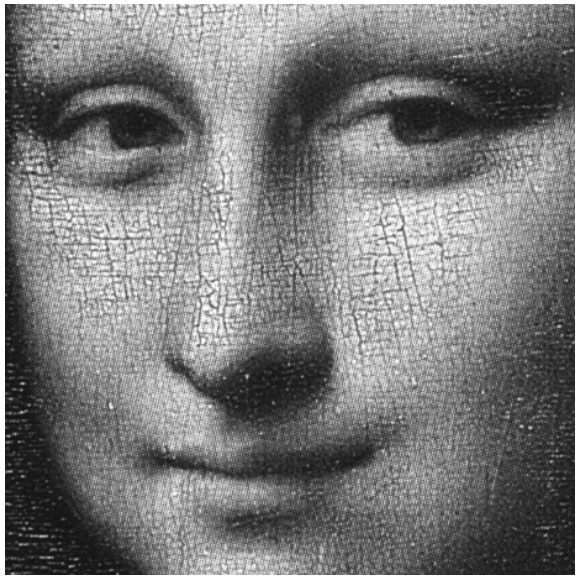
SubbandDoGLaplacian/subbanddecomposition.m

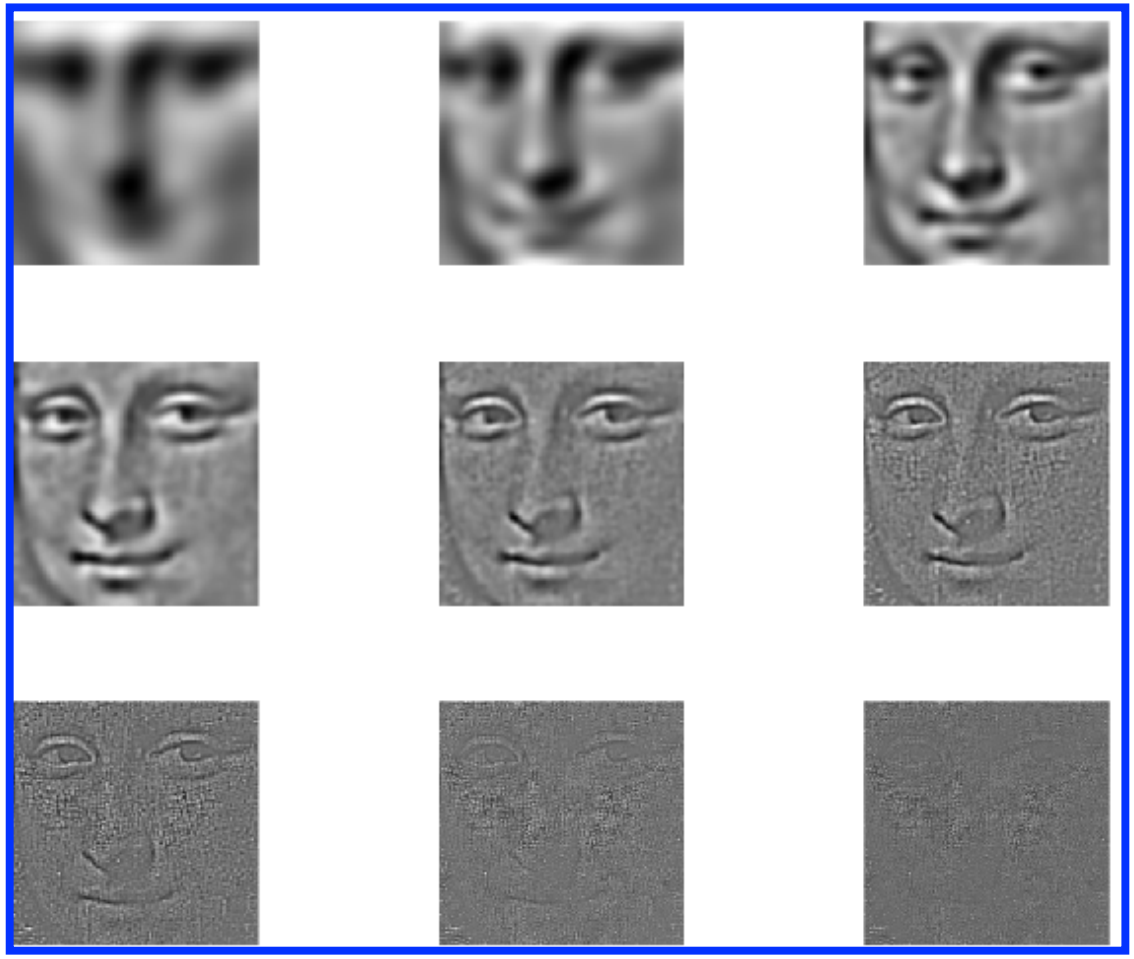
What will the Laplacien pyramid show?



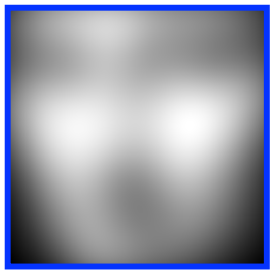

Low pass residual

Laplacien pyramid

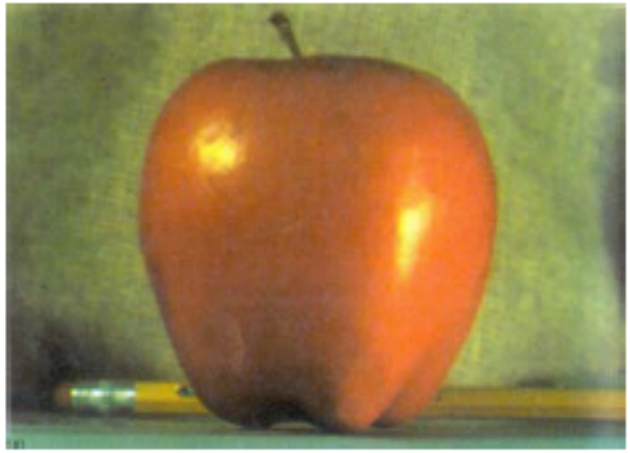




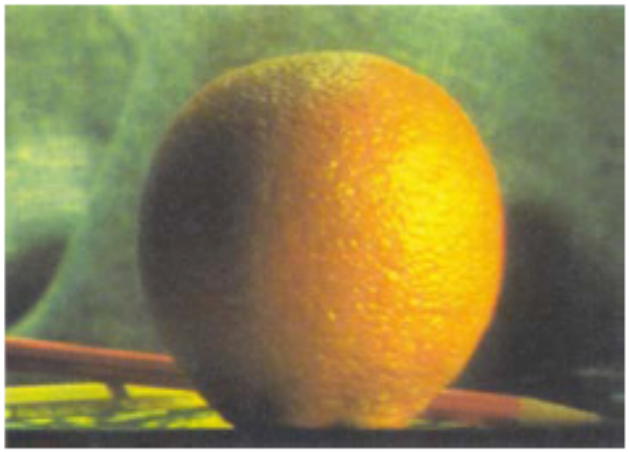
Residual



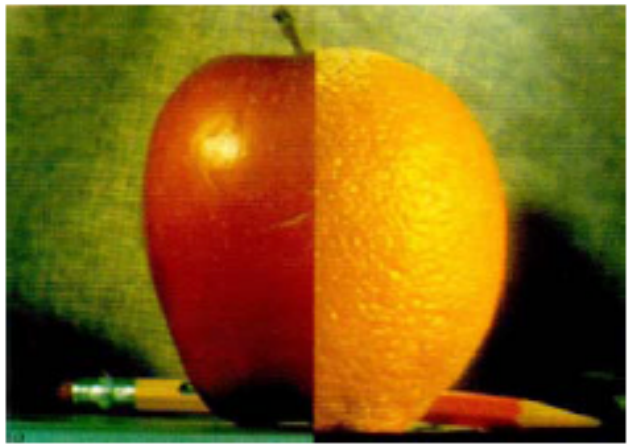
Laplacien pyramid:
Use for Image blending

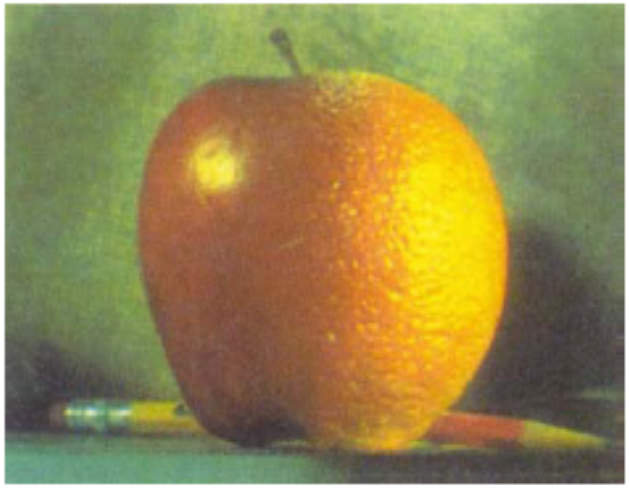


(a)



(b)





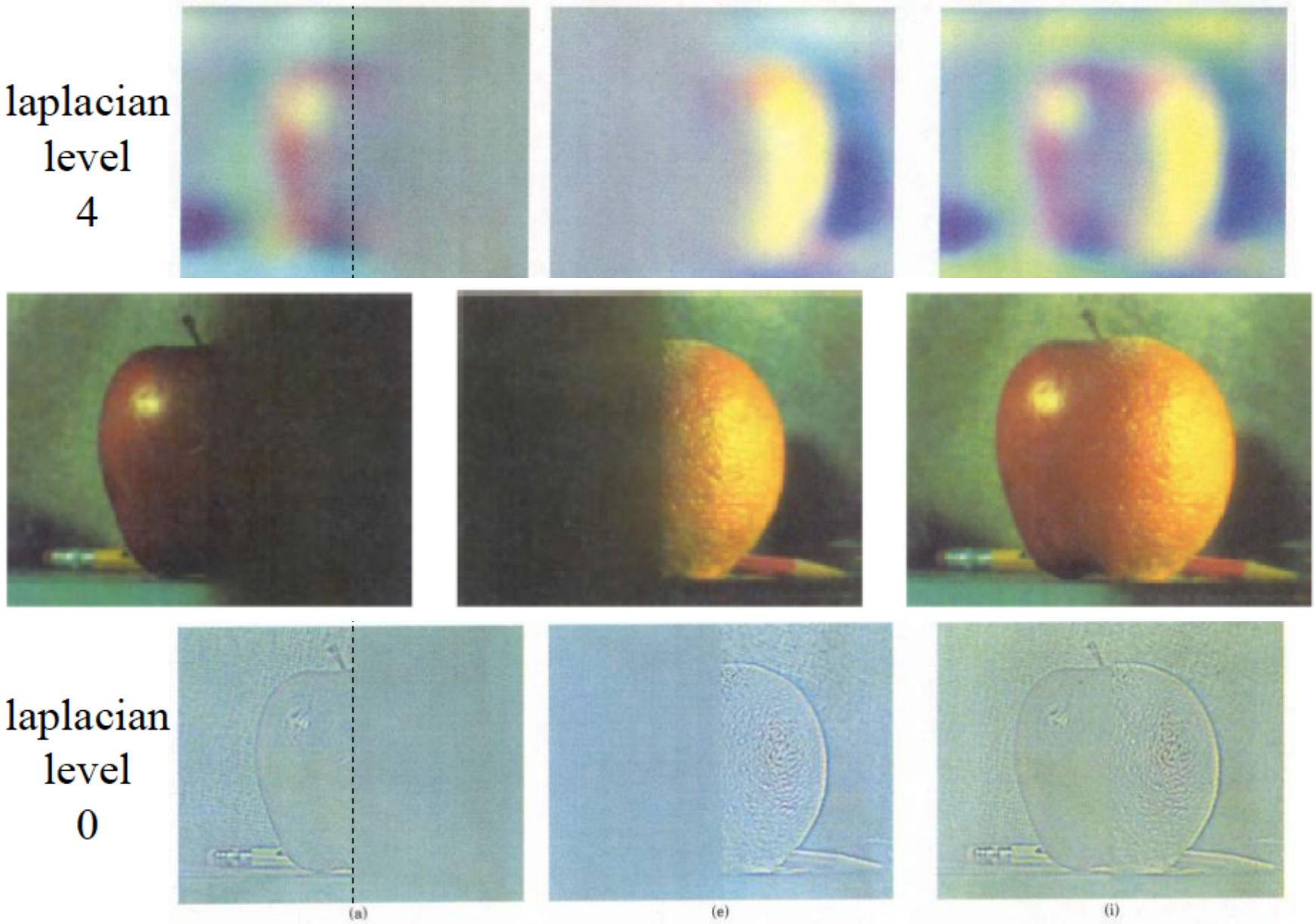
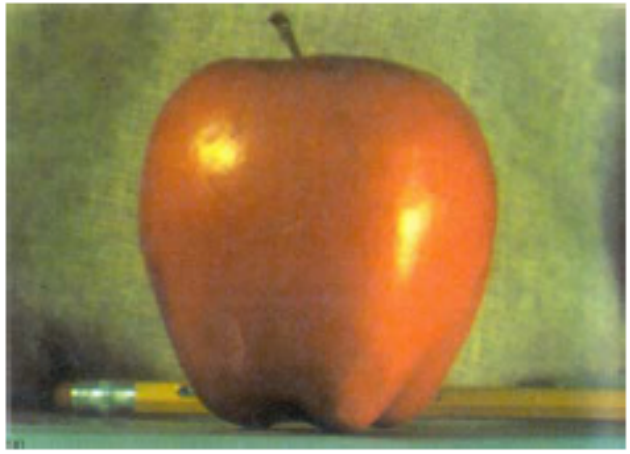
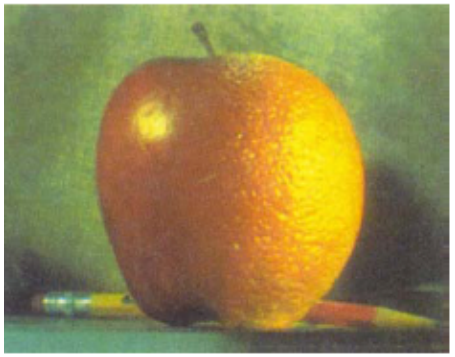

laplacian level 4
laplacian level 0
(a)
(e)
(i)
left pyramid
right pyramid
blended pyramid

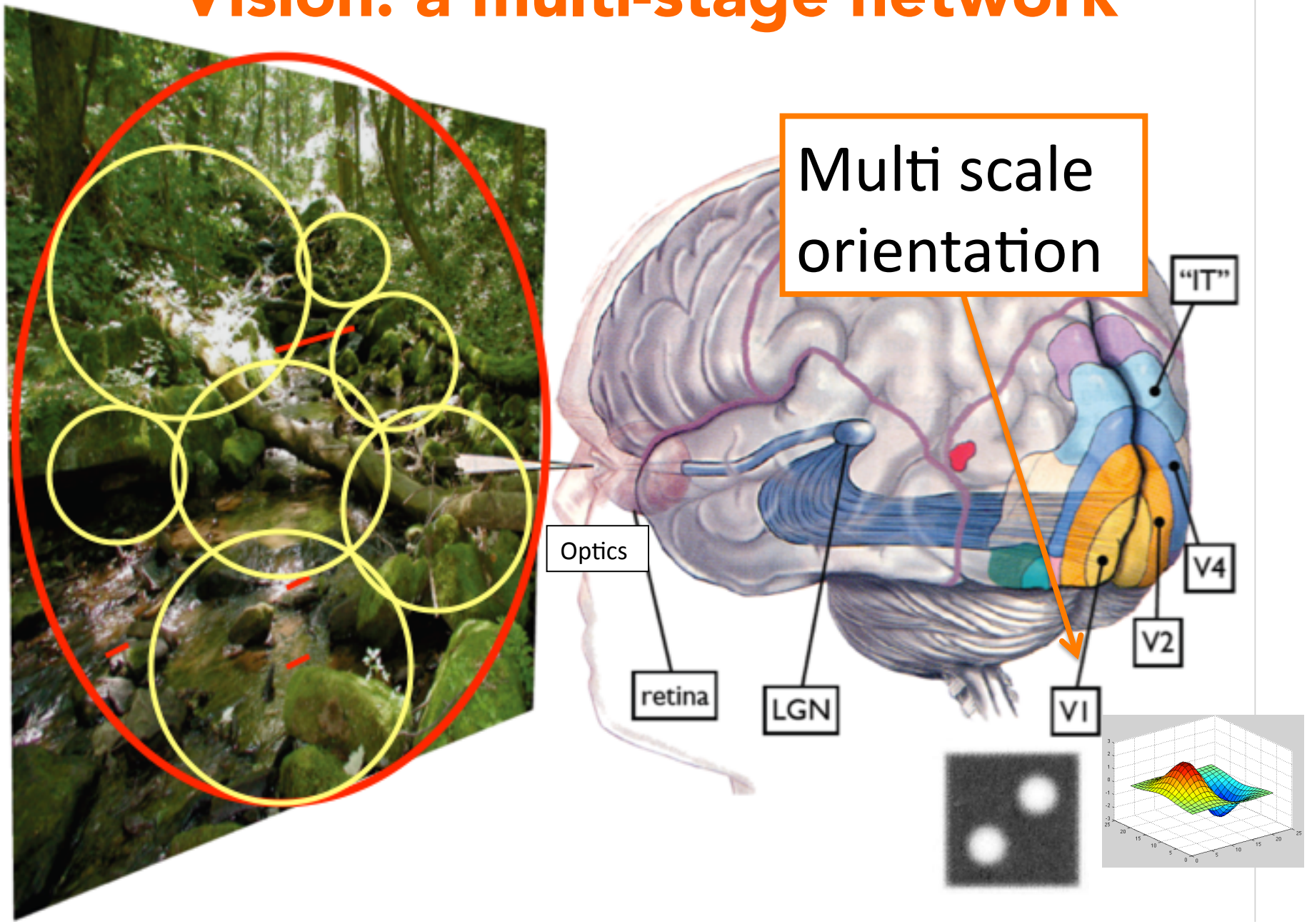
Image blending



- Build Laplacian pyramid for both images: LA, LB
- Build Gaussian pyramid for mask: G
- Build a combined Laplacian pyramid: L(j) = G(j) LA(j) + (1-G(j)) LB(j)
- Collapse L to obtain the blended image

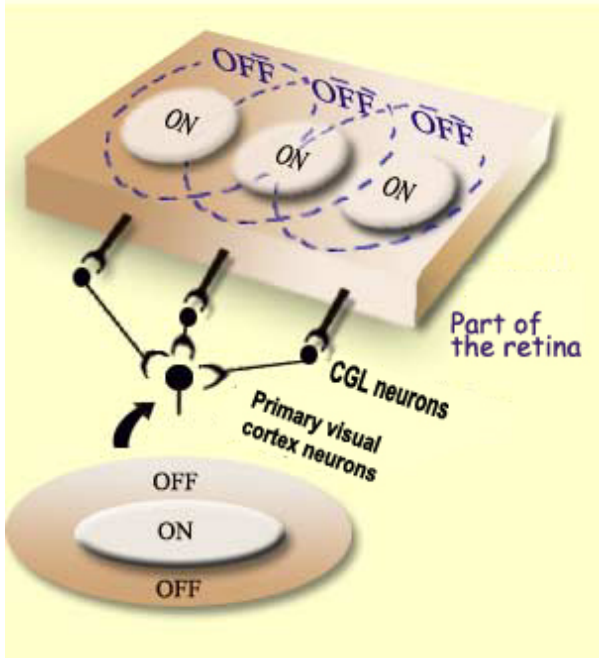


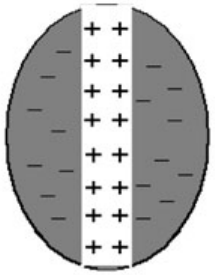
Vision: a multi-stage network



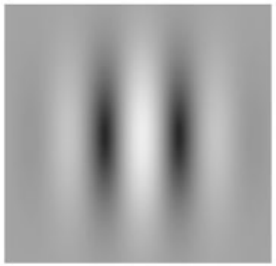
Oriented representation in V1

- Simple-cell receptive fields (RFs) are constructed from the output of LGN cells
- They are selective to **oriented contours** and edges





RECEPTIVE FIELD STRUCTURE

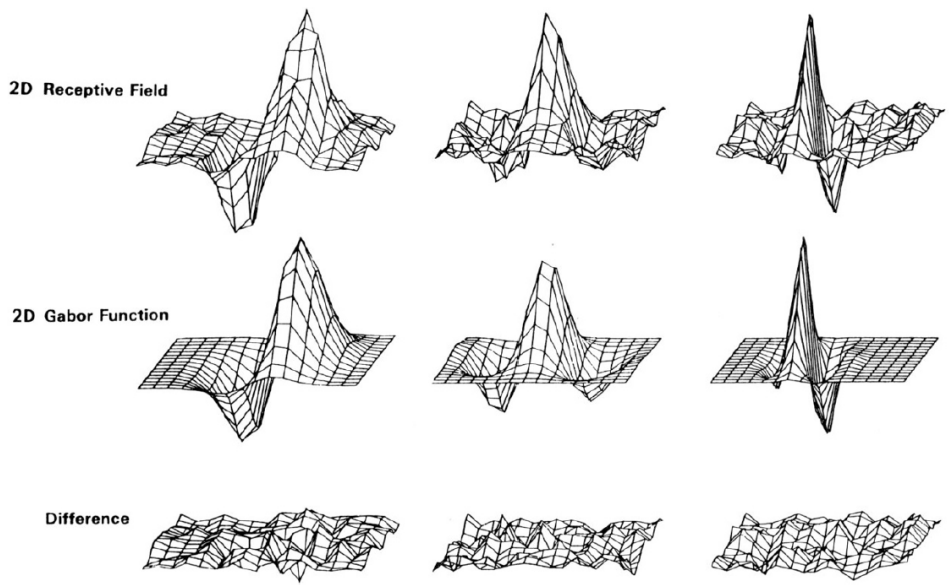


GABOR PATCH

V1 cells as Gabor Filters

The receptive fields of simple cells in V1 reflect the orientation and spatial frequency preferences of neurons. One way to model this, is to use a Gabor function, which is basically a two-dimensional gaussian modulated by a sinusoid.

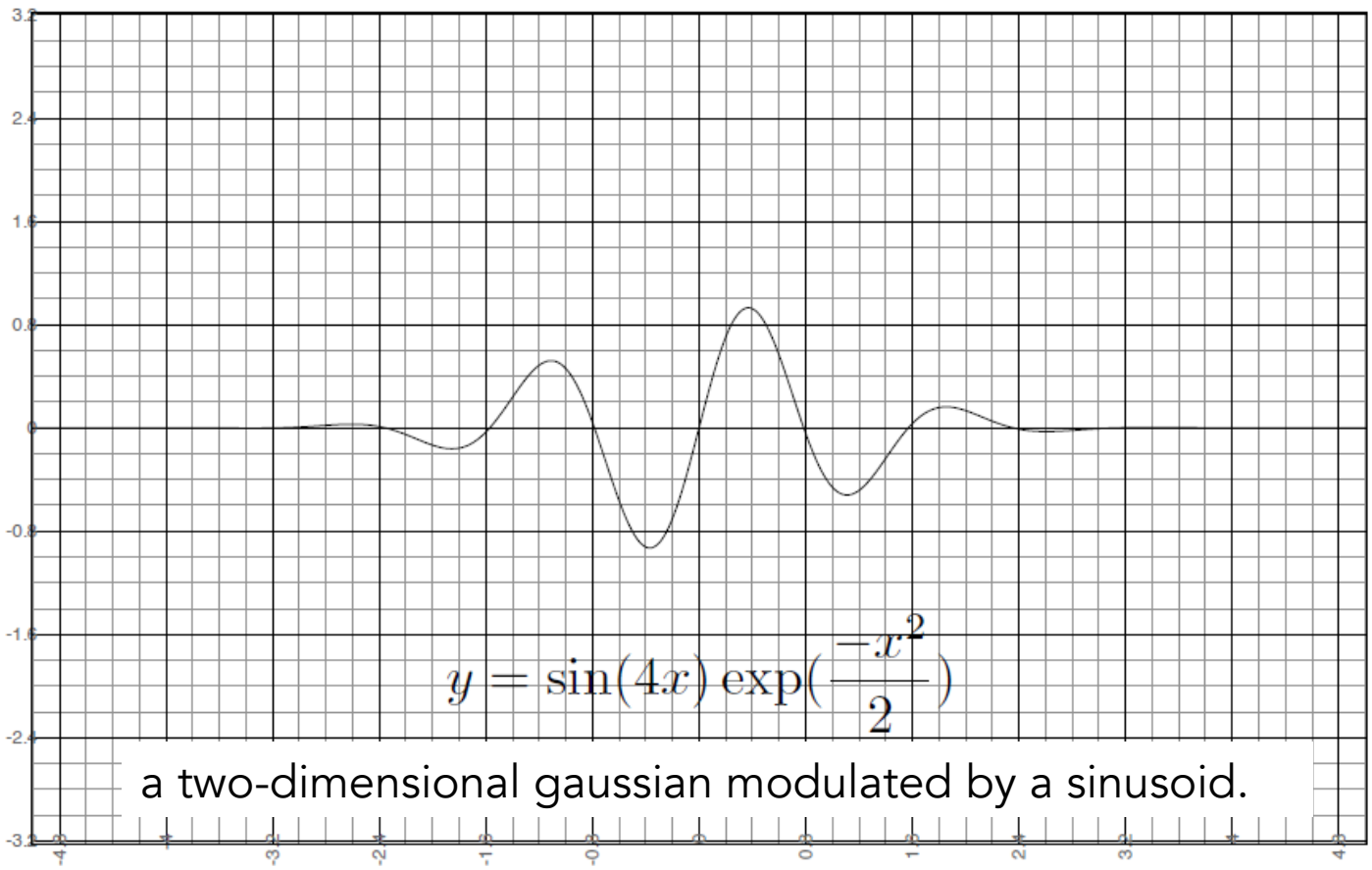
2D Receptive fields in primary visual cortex



Fit of 2D Gabor wavelet is indistinguishable from noise.

figure from Daugman, 1990
data from Jones and Palmer, 1987

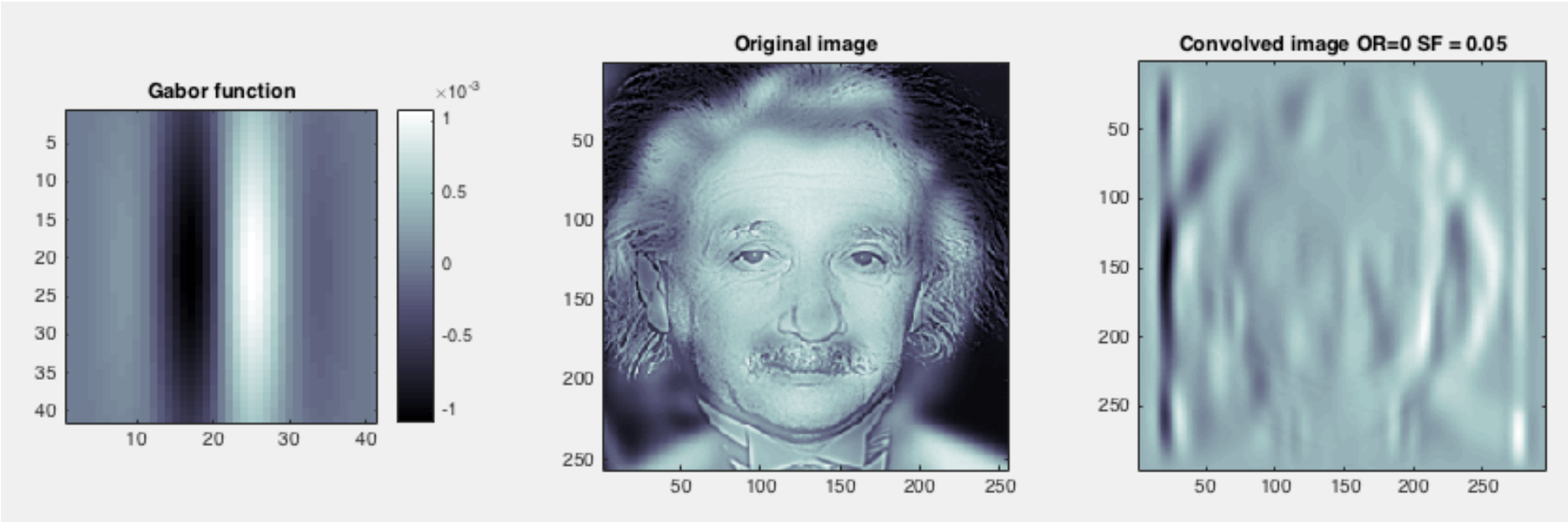
What is a Gabor filter?



Human visual system contains Gabor-filter cells

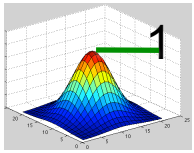
Gabor filters: **Orientation** x spatial **frequency** filtering

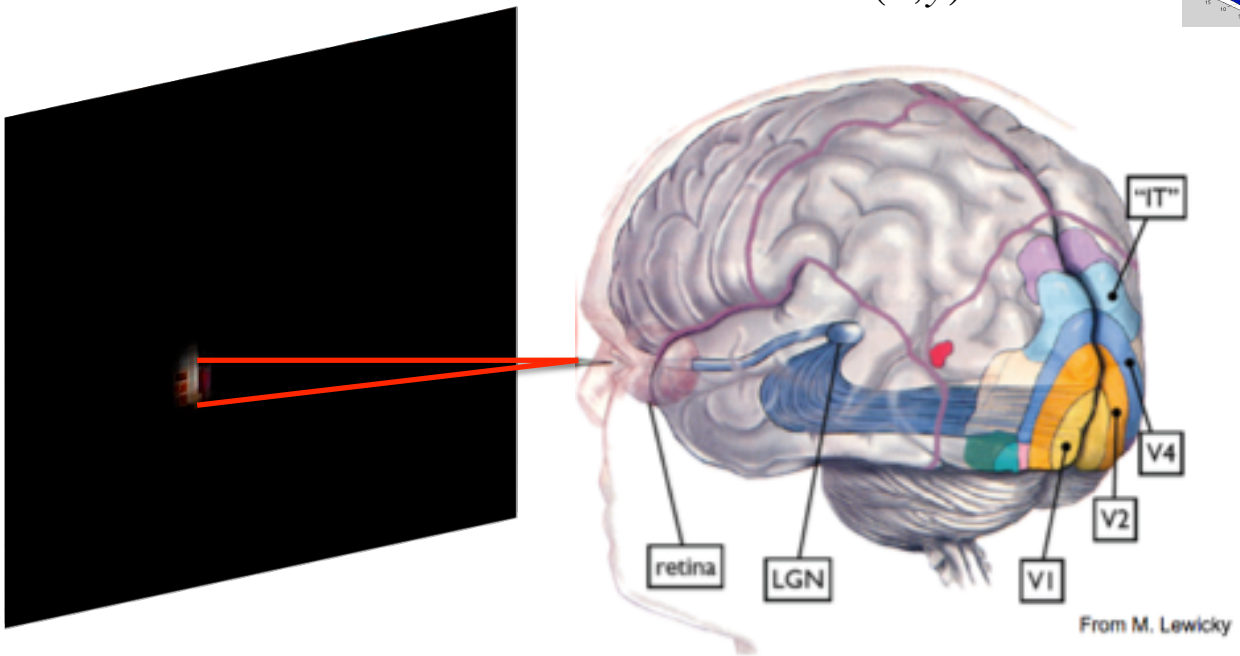
Gabor/OneGabor.m



The image through a Gaussian window

$$h(x,y) = e^{-\frac{x^2+y^2}{2\sigma^2}}$$



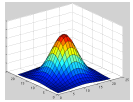


From M. Lewicky

Analysis of local frequency:

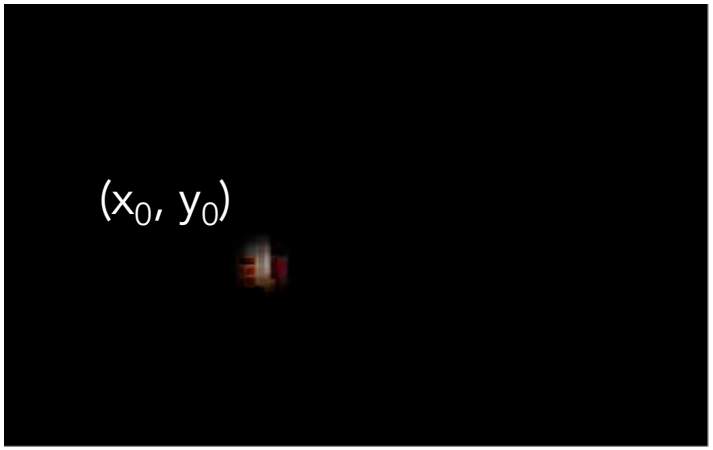
to obtain a Gabor function …

Multiply a Fourier basis function (shown here in the x direction) $e^{j2\pi u_0 x}$

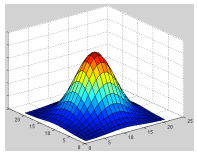
by a spatially localizing Gaussian window $e^{-\frac{(x-x_o)^2+(y-y_o)^2}{2\sigma^2}}$ 

to obtain a Gabor function or Gabor wavelet:

$$\psi(x,y) = e^{-\frac{x^2+y^2}{2\sigma^2}} e^{j2\pi u_0 x}$$



$$h(x,y;x_0,y_0) = e^{-\frac{(x-x_o)^2+(y-y_o)^2}{2\sigma^2}}$$



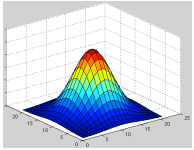
A Gabor function is complex-valued, but we can look at the real and imaginary parts to examine cosine or sine local filters:

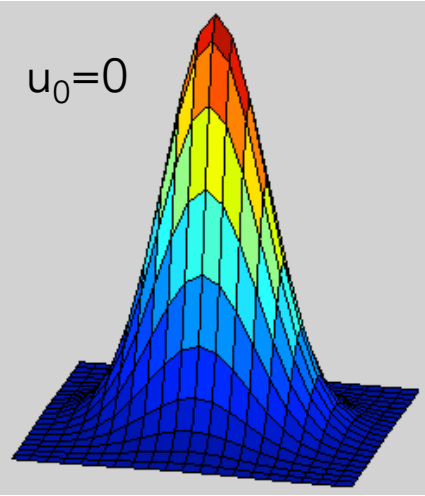
$$\psi_c(x,y) = e^{-\frac{x^2+y^2}{2\sigma^2}} \cos(2\pi u_0 x)$$

$$\psi_s(x,y) = e^{-\frac{x^2+y^2}{2\sigma^2}} \sin(2\pi u_0 x)$$

Gabor wavelets

$$h(x,y;x_0,y_0) = e^{-\frac{(x-x_o)^2+(y-y_o)^2}{2\sigma^2}}$$

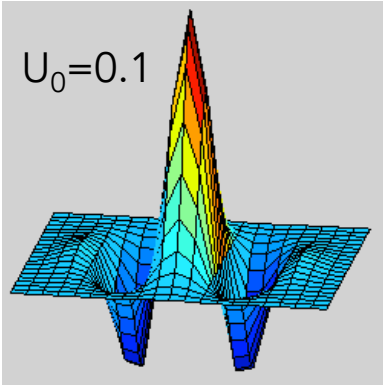


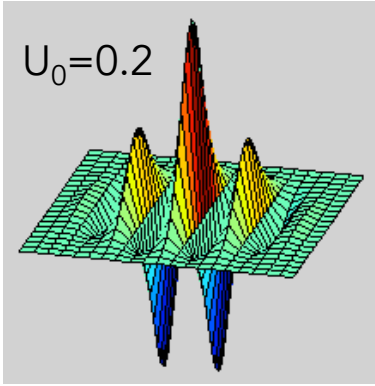


The Gaussian window is a Gabor function for a zero frequency sinusoid

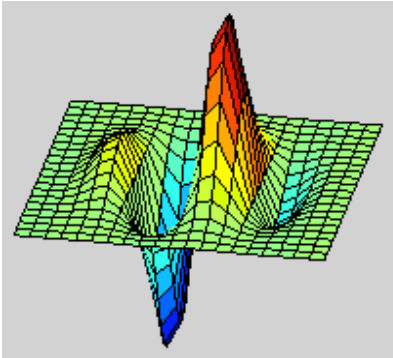
Spatial frequency lower

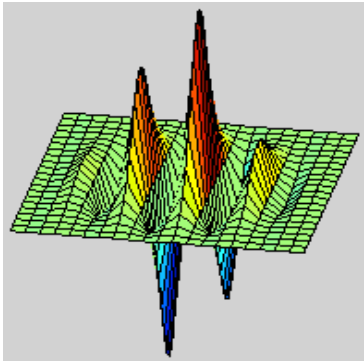
Spatial frequency higher



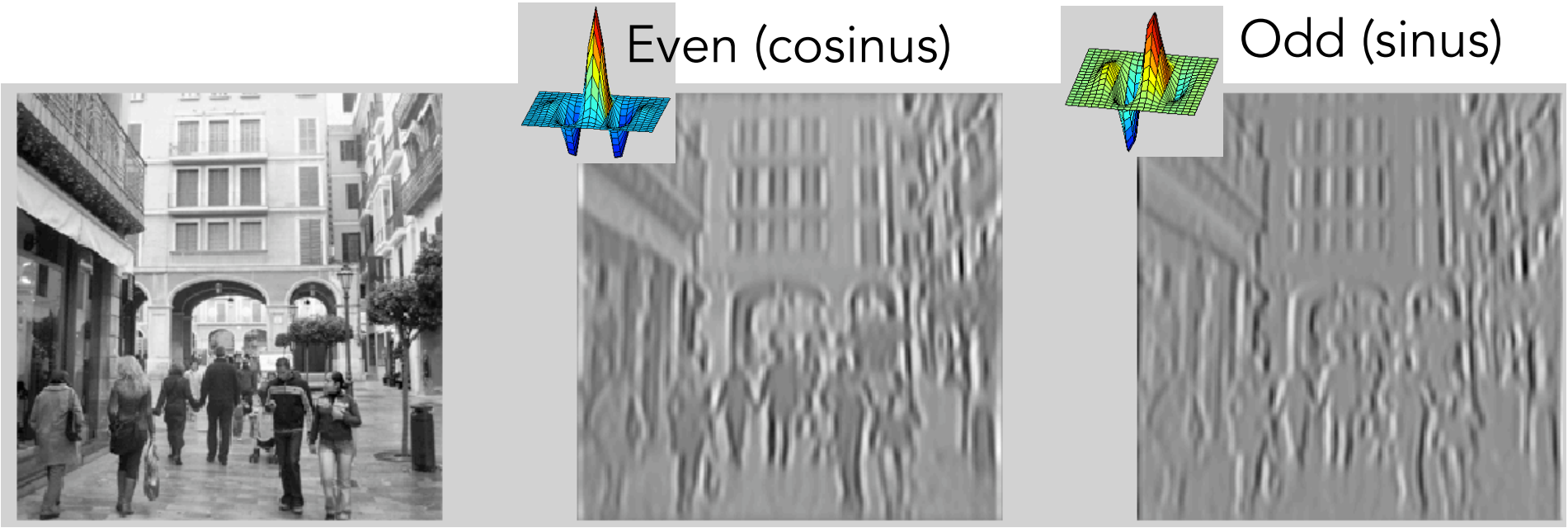


Cosine local filter

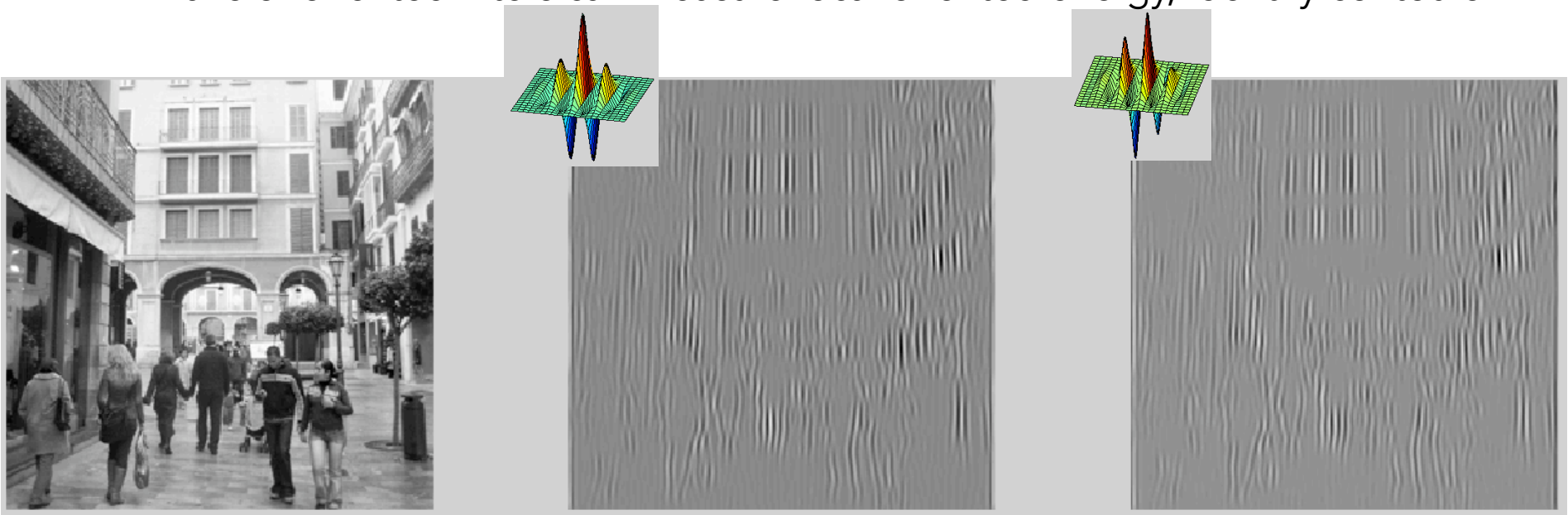


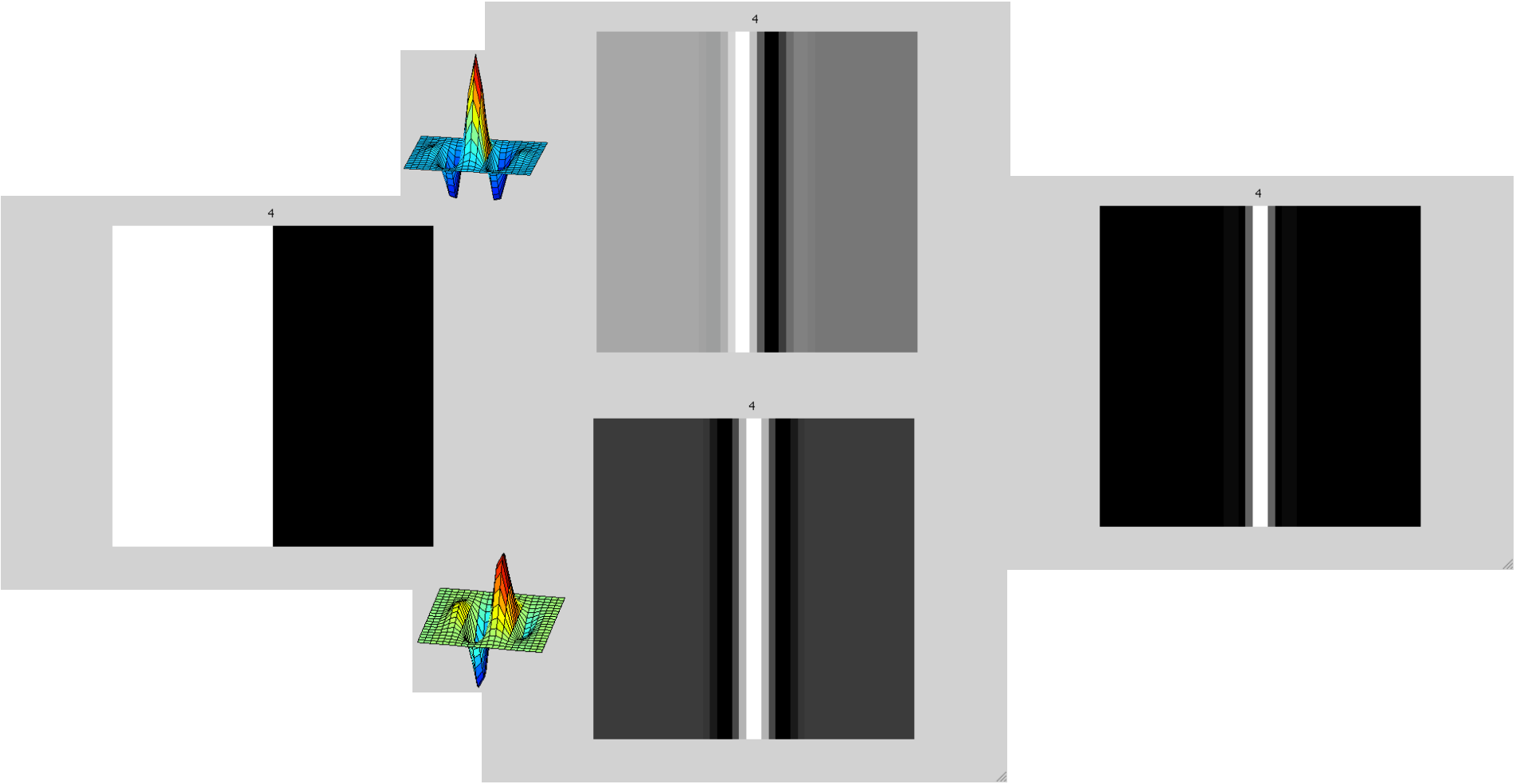


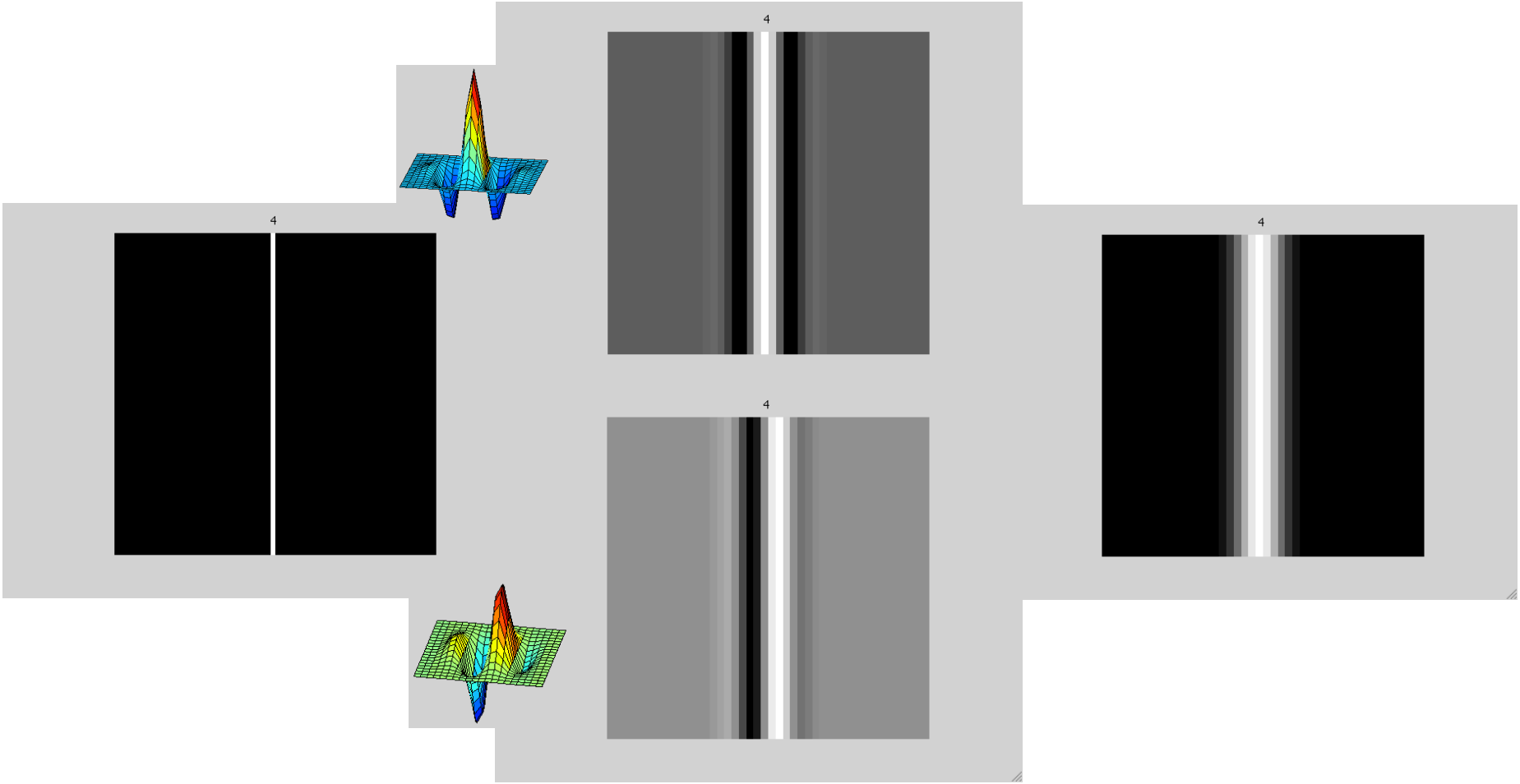
Sine local filter



Pairs of oriented filters can measure local oriented energy, identify contours




4
4
4
4


4
4
4
4

Gabor wavelet:

$$\psi(x,y) = e^{-\frac{x^2+y^2}{2\sigma^2}} e^{j2\pi u_0 x}$$

Tuning filter orientation:

$$x' = \cos(\alpha)x + \sin(\alpha)y$$

$$y' = -\sin(\alpha)x + \cos(\alpha)y$$

Space

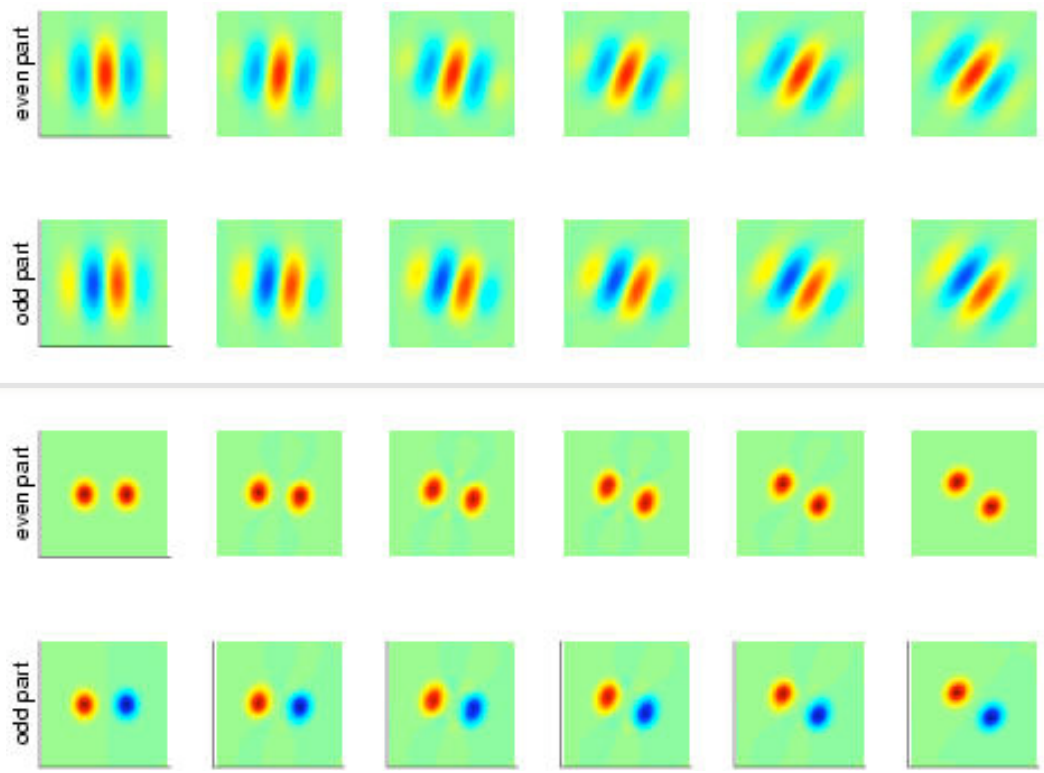
Real

Imag

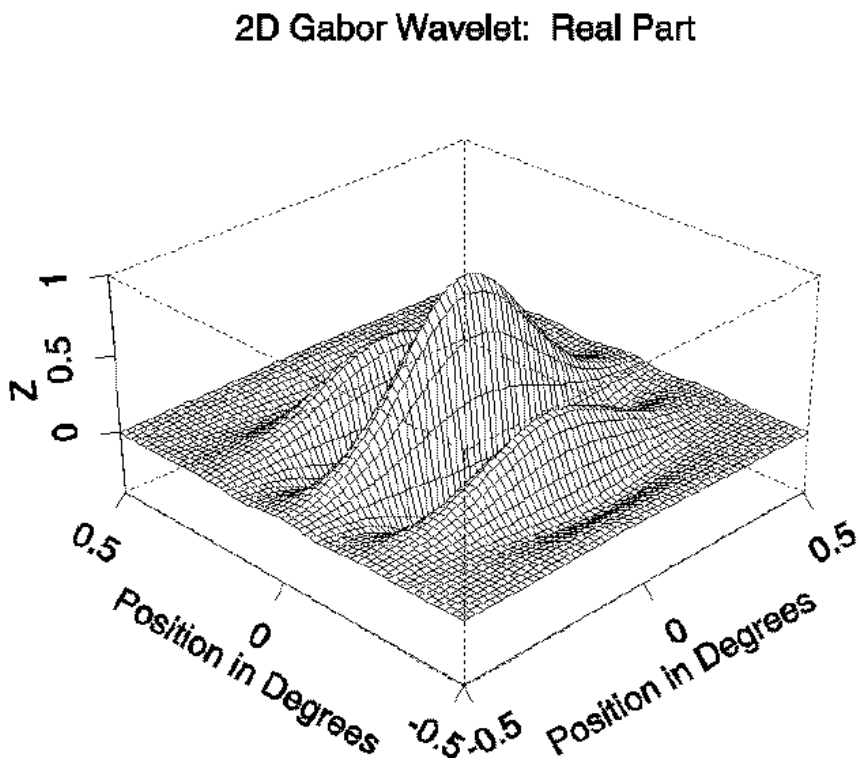
Fourier domain

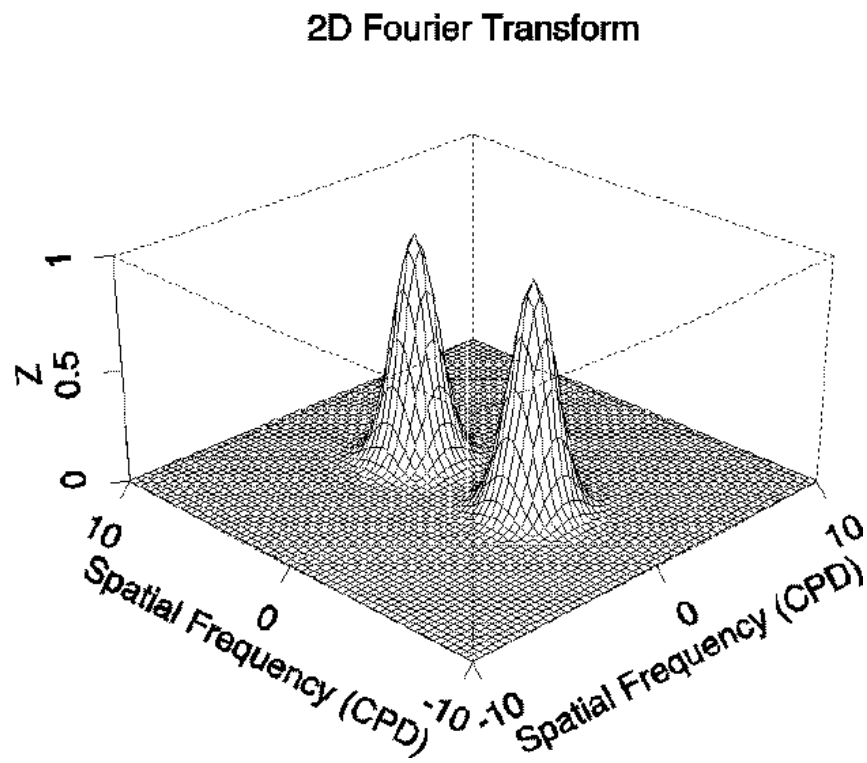
Real

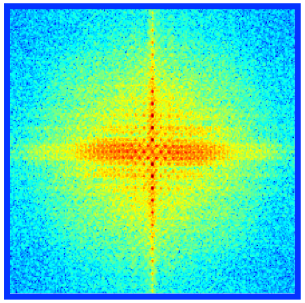
Imag



Gabor Filters: a bank of them



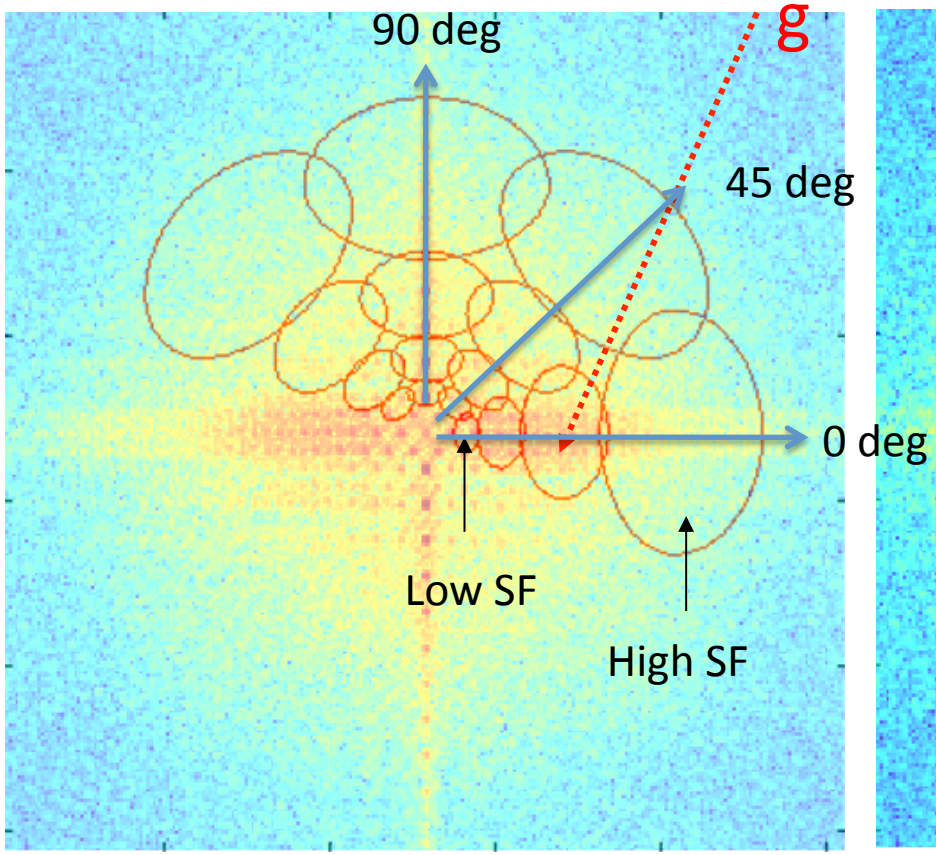


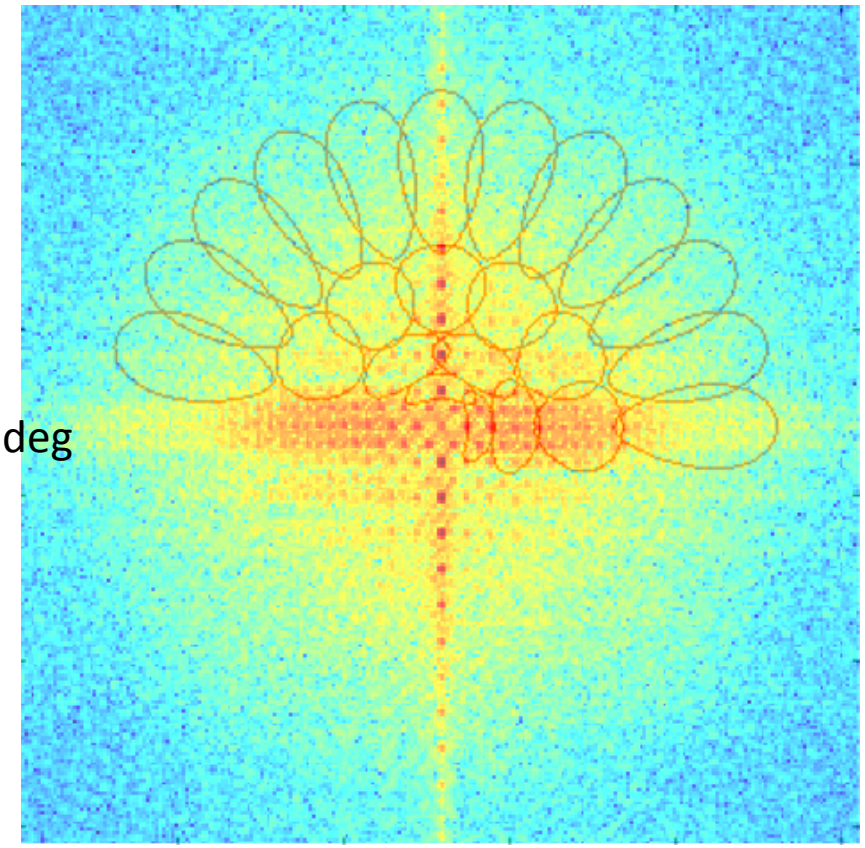


Gabor Filter Bank

Not for image reconstruction.. It does not cover the entire space

Gabor/
[createBankGabor](Gabor/createBankGabor.m).m





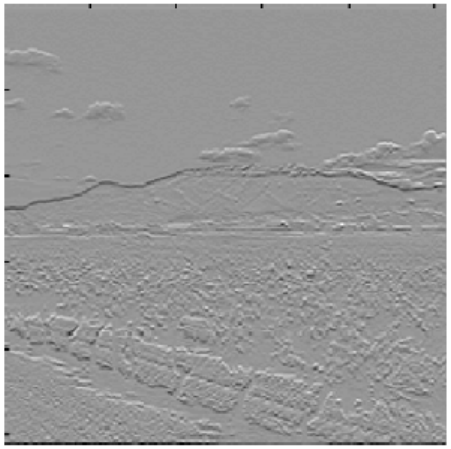
or = [4 4 4 4];

4 spatial scales with 4 orientations each

or = [12 6 3 2];

Convolution in 2-d

$$* \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} =$$

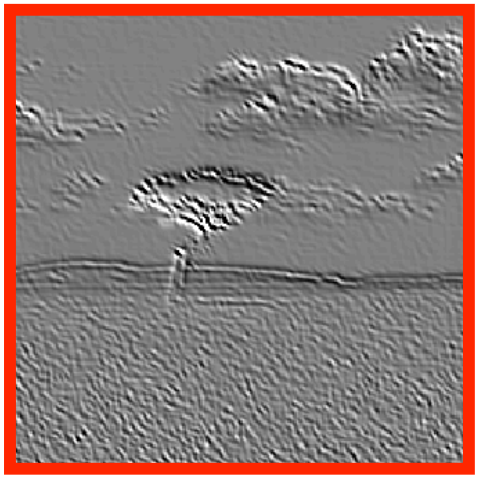


Which filters will be most useful?

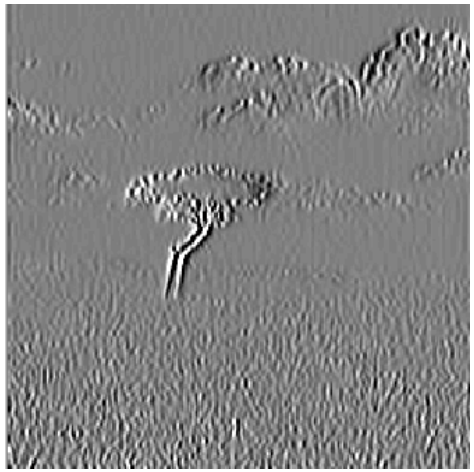
$* \; g \; =$

$* \; g_i \; =$

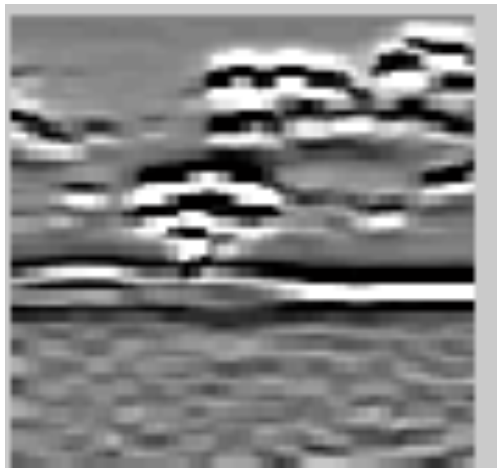
Energy in each sub-band representation



…



…



Sum of all values

$* \; g_0$

$=$

$*\ g$

$=$

$*\ g$

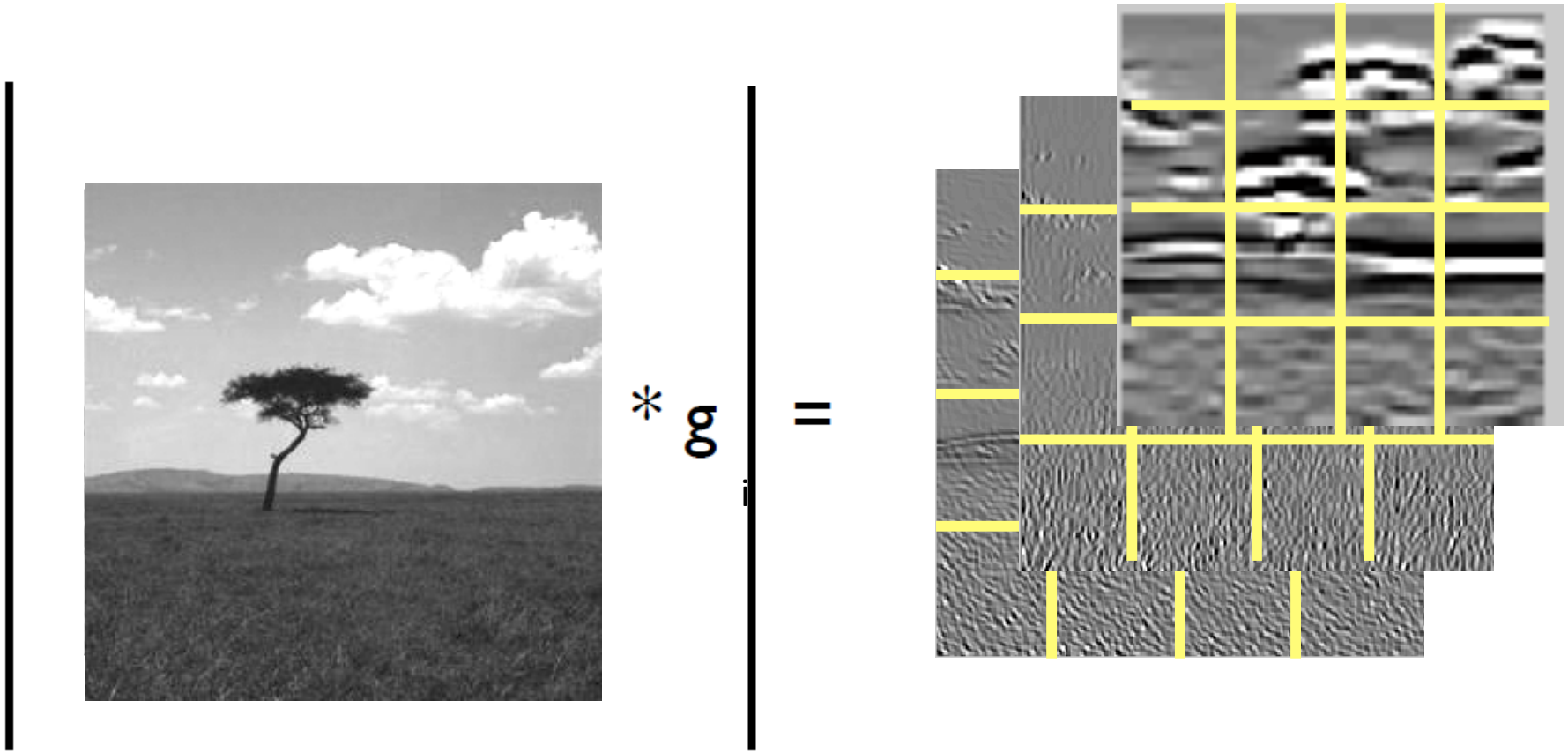
$=$

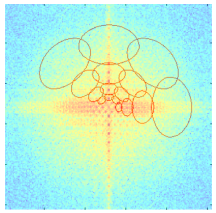
$*\ g$

$=$

$*\ g$

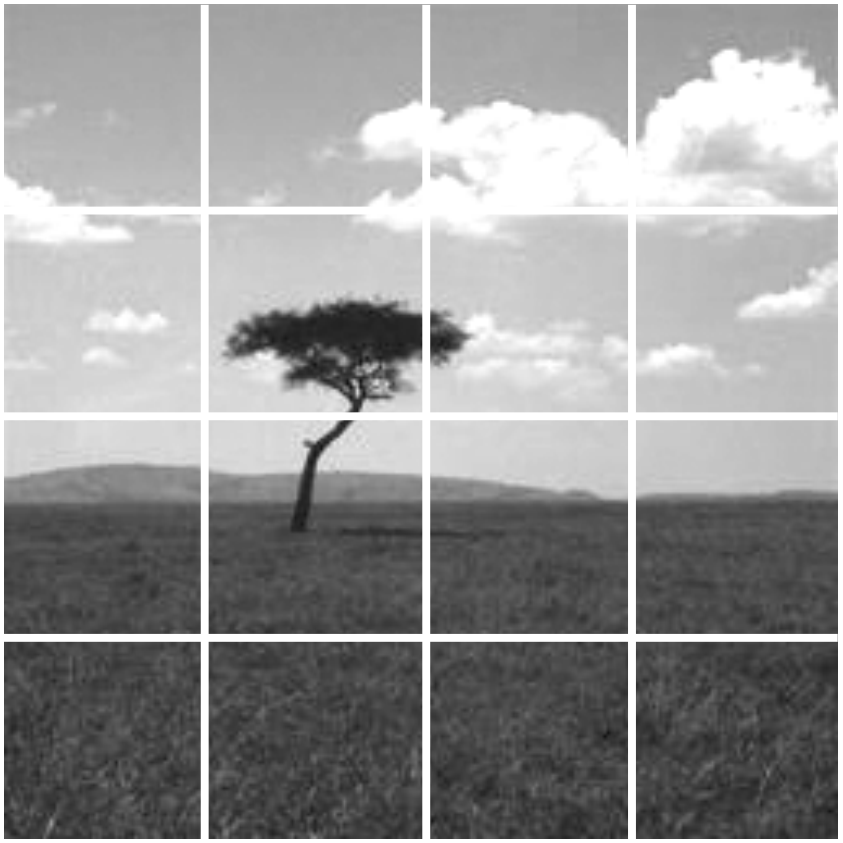
$=$

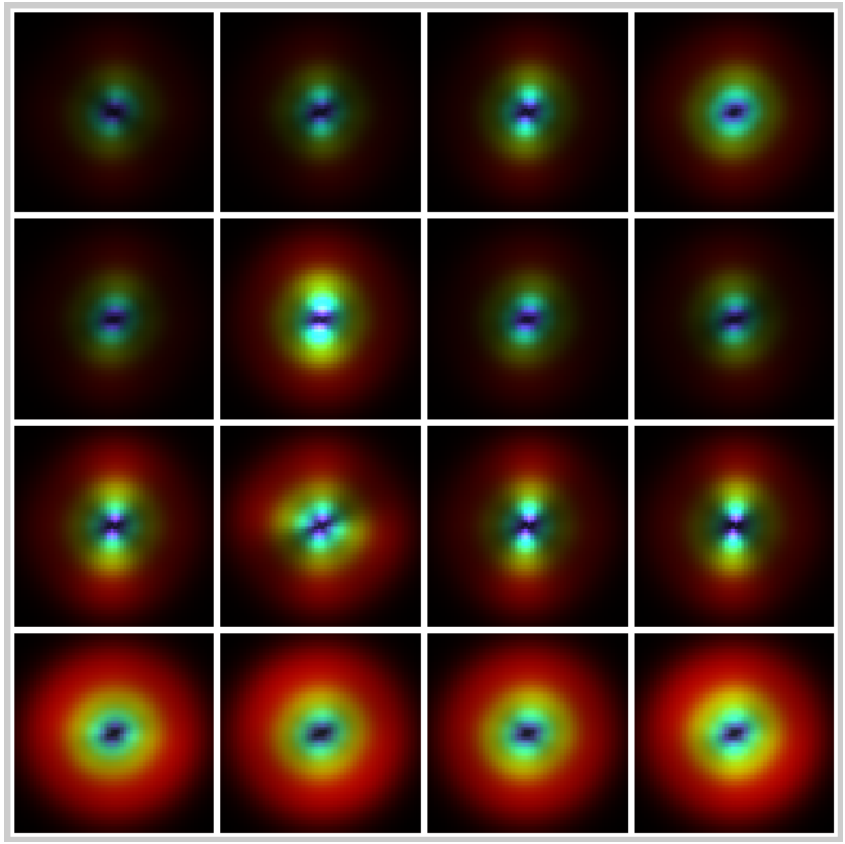




16 Gabors, 4 x 4 spatial grid -> 256 features

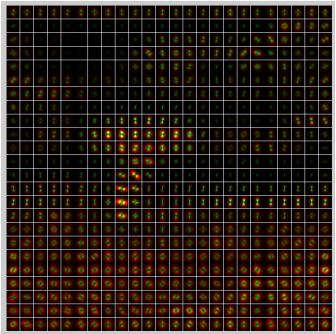
Representation of GIST in spectral domain (visualization)

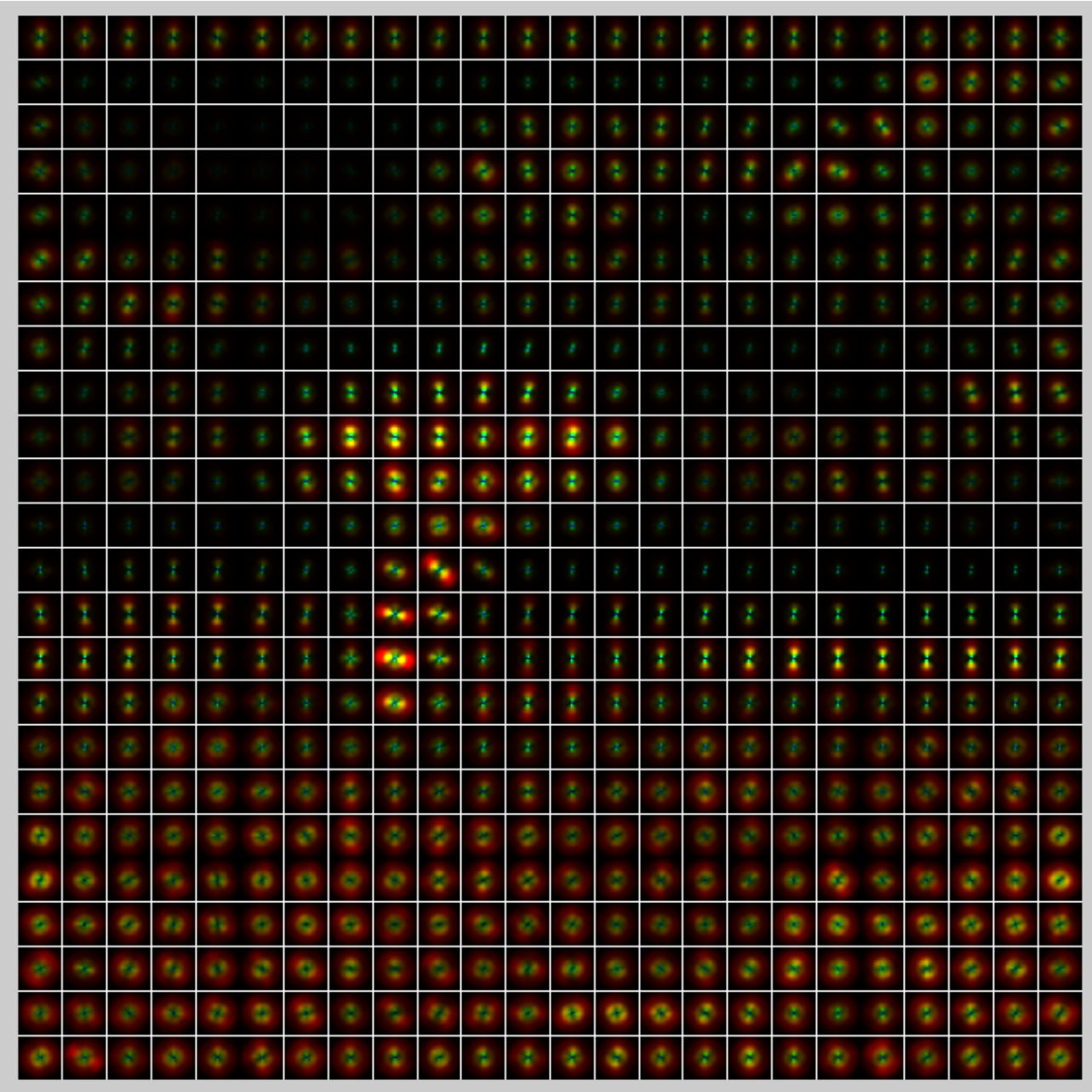




[8 8 8 8] grid 4 x 4

GIST/demogist.m





Download MATLAB Toolbox for the LabelMe Image Database

Scene recognition

Gist descriptor

Here we provide a function to compute the gist descriptor as described in: Aude Oliva, Antonio Torralba. Modeling the shape of the scene: a holistic representation of the spatial envelope. International Journal of Computer Vision, Vol. 42(3): 145-175, 2001. To compute the gist descriptor on an image you can use the function LMgist. Here is one example that reads one image and computes the descriptor.

```
% Load image
img = imread('demo1.jpg');

% Parameters:
param.imageSize = 128;
param.orientationsPerScale = [8 8 8 8];
param.numberBlocks = 4;
param.fc_prefilt = 4;

% Computing gist:
[gist, param] = LMgist(img, '', param);

% Visualization
figure
subplot(121)
imshow(img)
title('Input image')
subplot(122)
title('Descriptor')
showGist(gist, param)
```

You can also compute the gist for a collection of images:

```
gist = LMgist(D, HOMEIMAGES, param);
```

The output is an array of size [Nscenes Nfeatures], where Nscenes = length(D).

Estimation of the horizon line using the Gist descriptor

The goal is to estimate the location of the horizon line on an image. This function uses the approach described in:

- A. Torralba, P. Sinha. *Statistical context priming for object detection.* ICCV 2001.

nages and annotations in the LabelMe database.
stics, dealing with synonyms, etc. This page

ıre that "git" is installed on your machine and
ıd line. You can refresh your copy to the latest

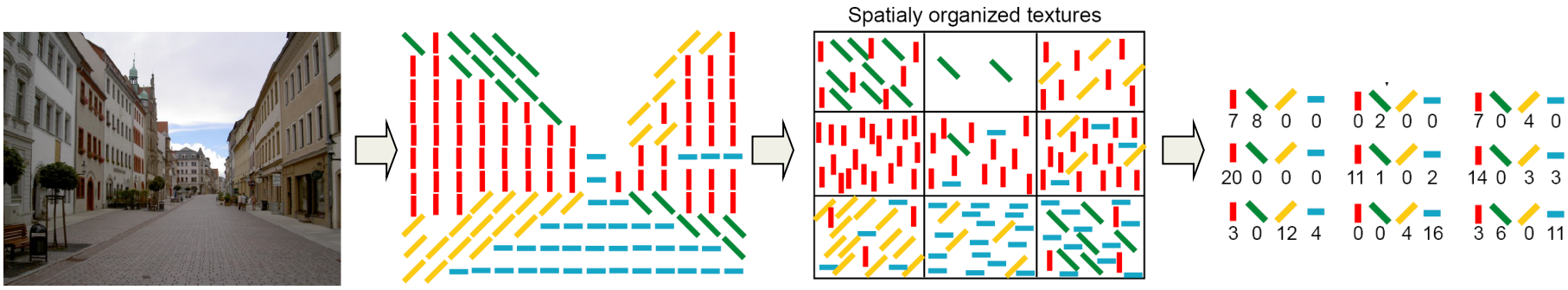
ıreciate if you cite:

)08. (paper.pdf)

A. Barriuso and A. Torralba.
Notes on image annotation.
arXiv:1210.3448 [cs.CV] October, 2012. (paper.pdf)

http://labelme.csail.mit.edu/Release3.0/browserTools/php/matlab_toolbox.php

GIST features represent local image texture



GIST Feature: Retrieve similar images

GIST/demogist.m (Demo 2)

Download labelme toolbox

GIST Feature: Principal Components

Demo 3: Projection onto first two principal components

How small is 32x32 ?

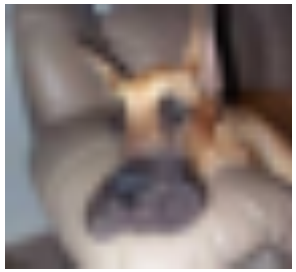
Your typical megapixel picture

Flickr (240x150 pixels)

Google (132x90 pixels)

Windows (90x90 pixels)

Tiny images (32x32 pixels)



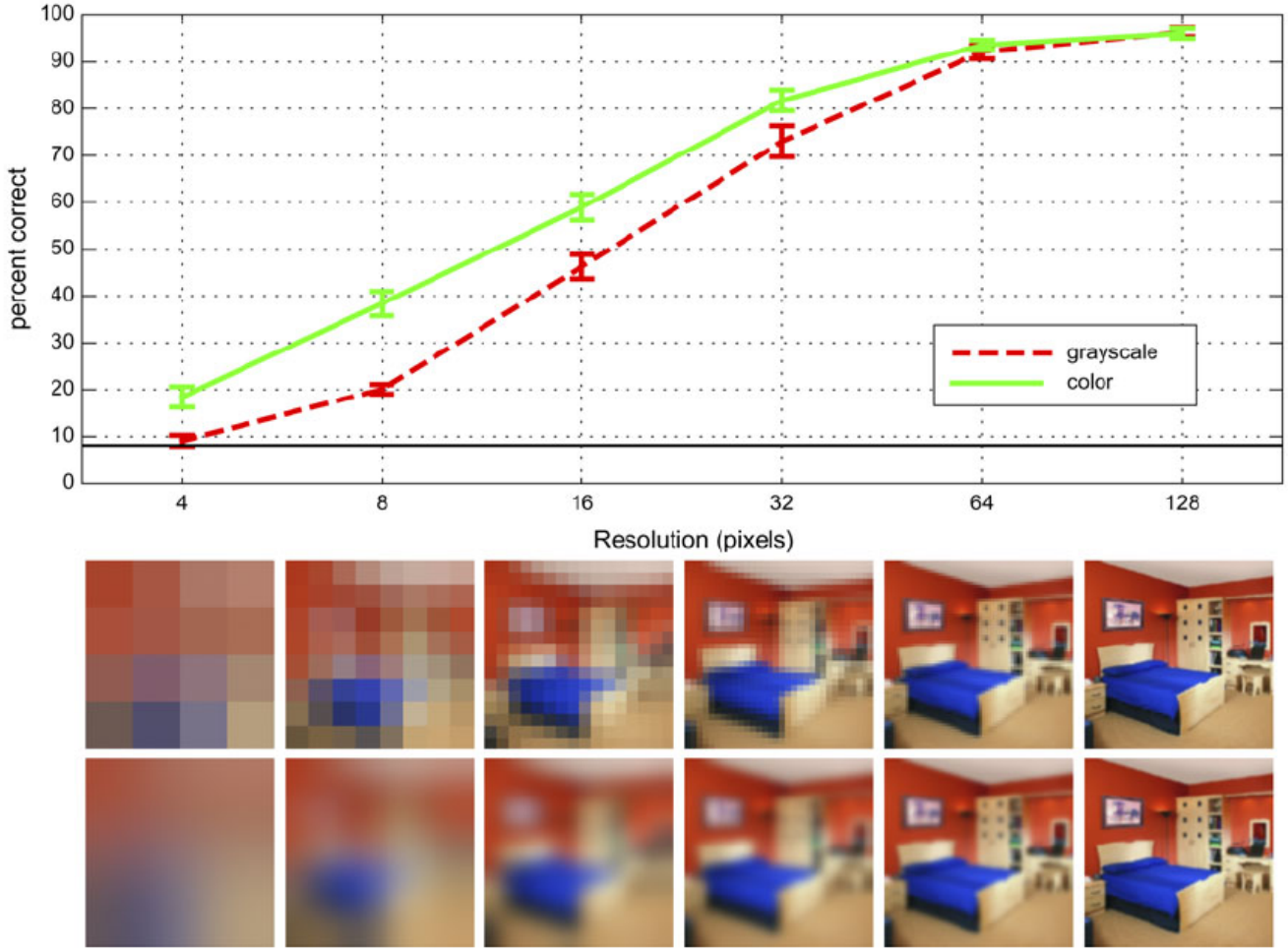
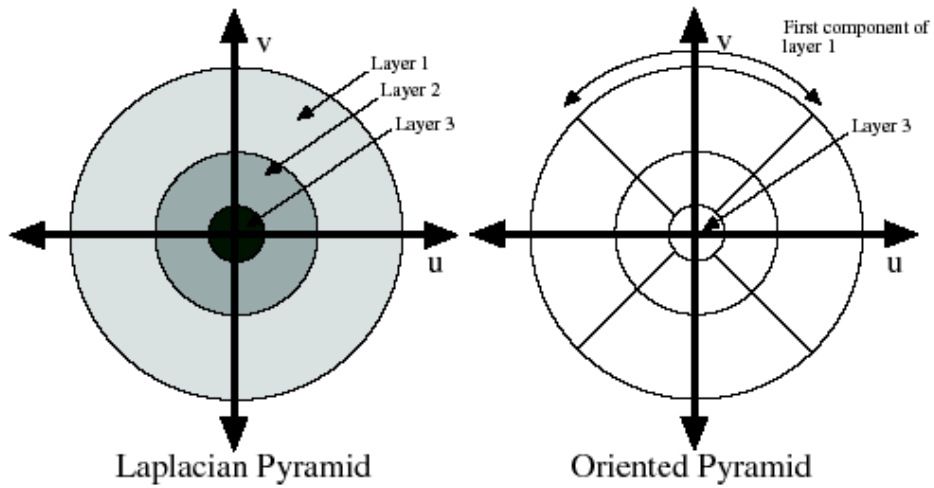


Fig. 2. (Color online) Scene recognition as a function of image resolution. Error bars represent 1 S.E. of the mean, obtained from 12 participants for each condition. The vertical axis represents the correct recognition rate, and the horizontal axis corresponds to the image resolution in a logarithmic scale. The black horizontal line represents chance level. The two rows of images illustrate the amount of information available at each resolution. The top row shows the downsampled images at each resolution (from 4×4 to 128×128 pixels), and the second row shows the images upsampled to 256×256 pixels that were shown to the participants.

Torralba (2009)

Oriented Pyramids in spectral space

- Apply an oriented filter to determine orientations at each layer
  - Important for texture synthesis
  - this represents image information at a particular scale and orientation, similar to V1 cells in the human brain



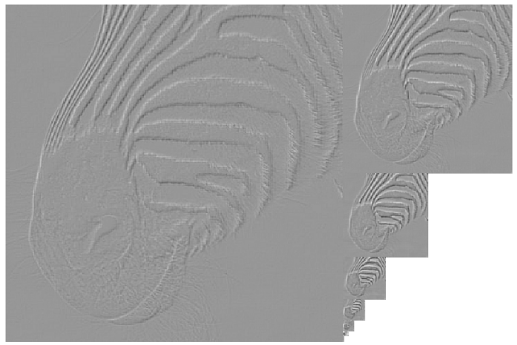
An oriented pyramid cut each annulus into a set of wedges. If (u,v) fourier space is represented in polar coordinates, each wedge corresponds to an interval of radius values and an interval of angle values.

Image pyramids

- Gaussian

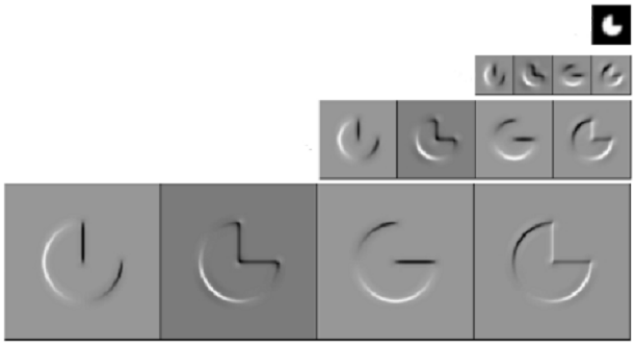
Progressively blurred and subsampled versions of the image. Adds scale invariance to fixed-size algorithms.

- Laplacian



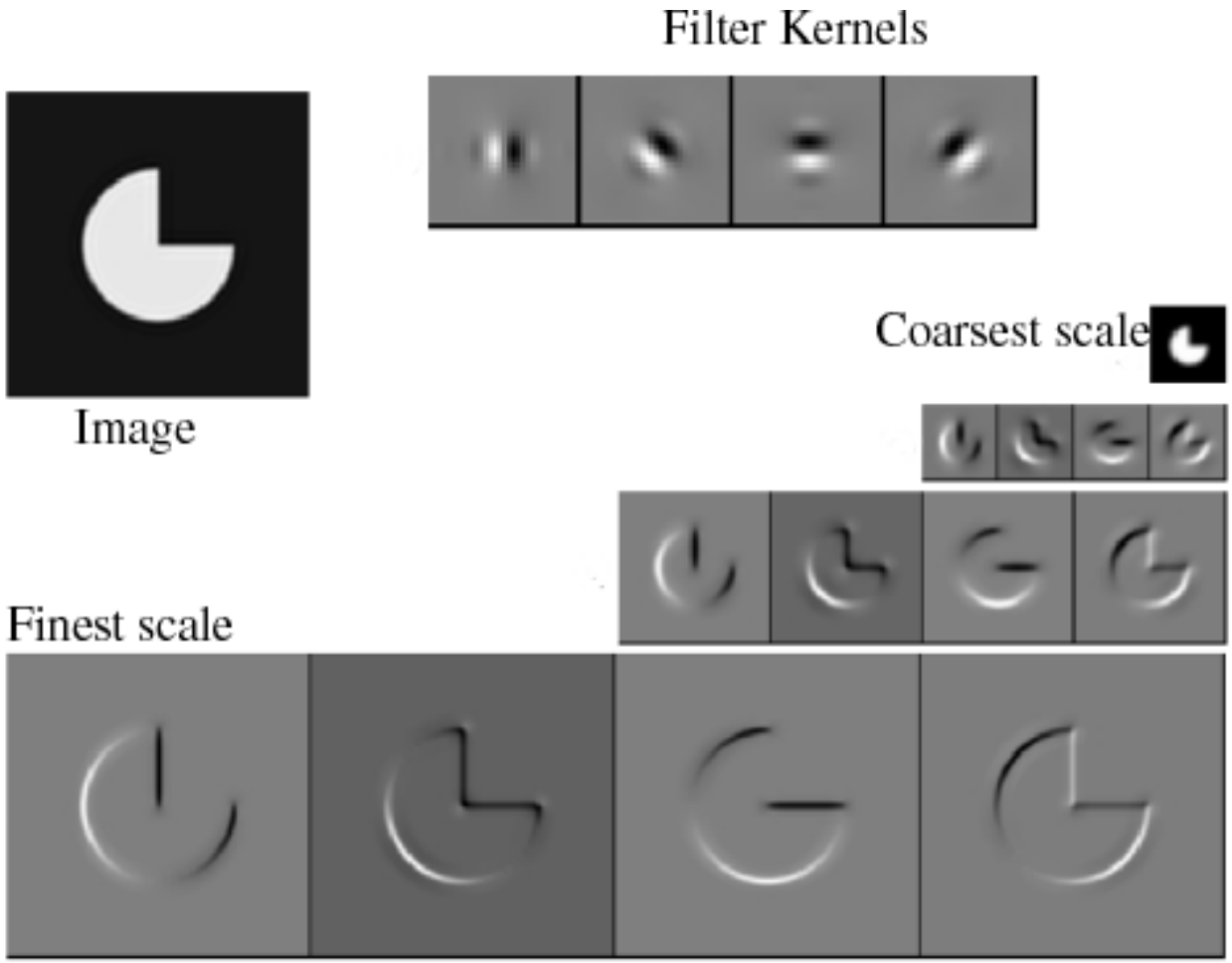
Shows the information added in Gaussian pyramid at each spatial scale. Useful for noise reduction & coding.

- Steerable pyramid



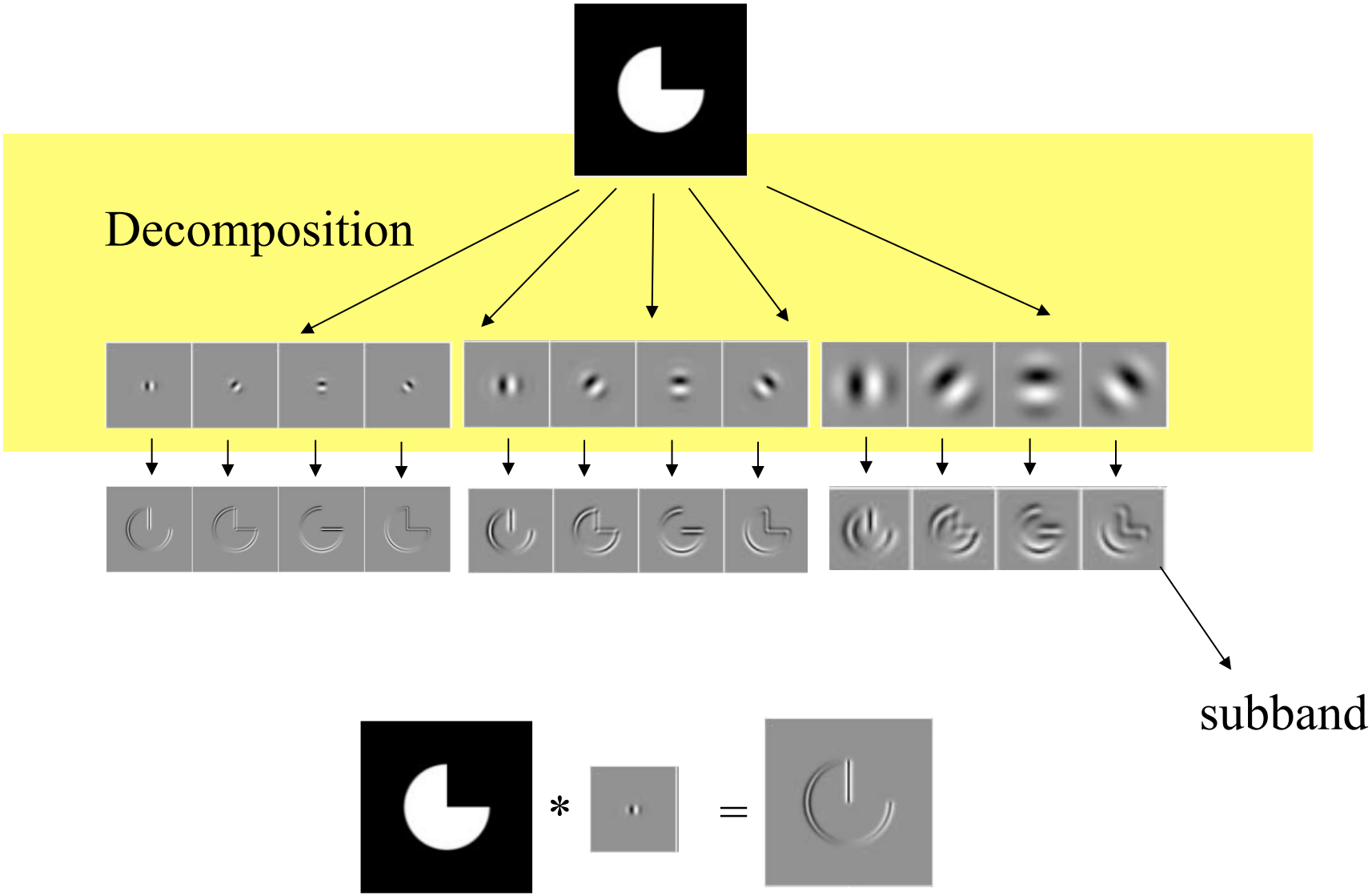
Shows components at each scale and orientation separately. Non-aliased subbands. Good for texture and feature analysis.

Steerable Pyramids



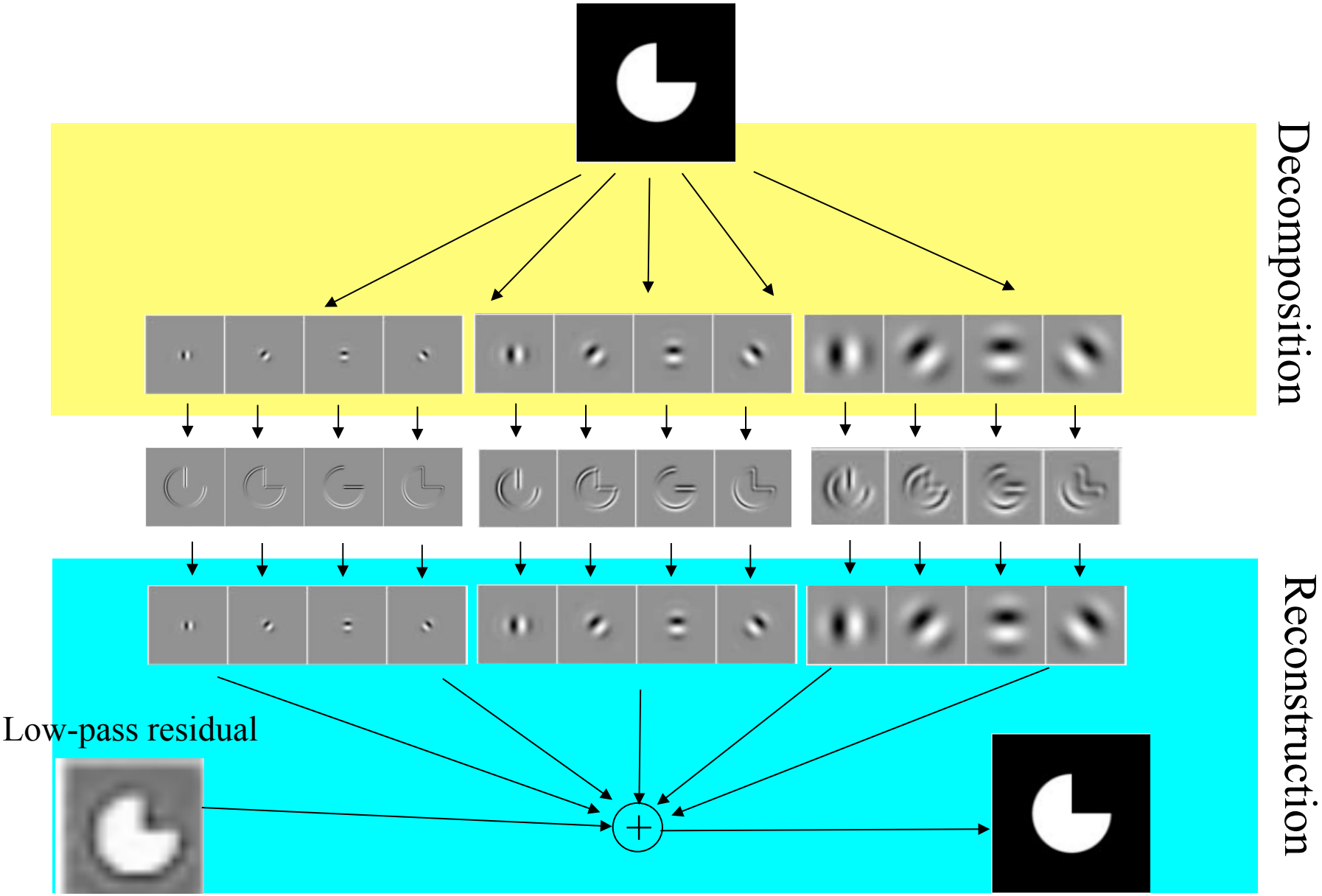
Simoncelli, Freeman, Adelson

The steerable pyramid

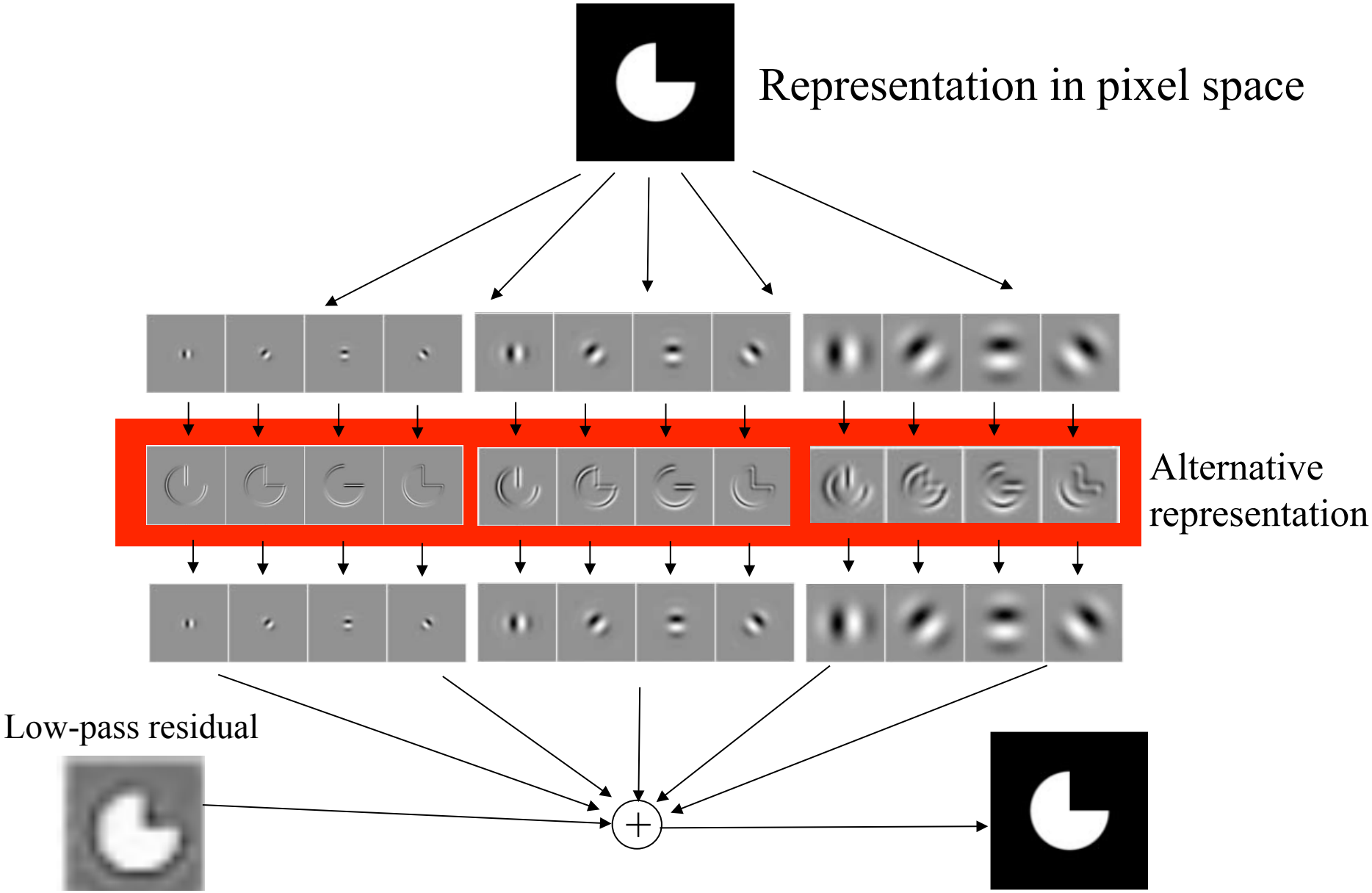


Simoncelli, Freeman, Adelson

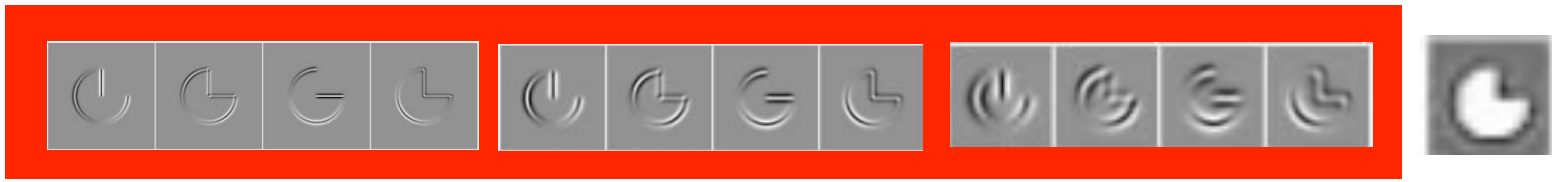
The steerable pyramid



The steerable pyramid



The steerable pyramid



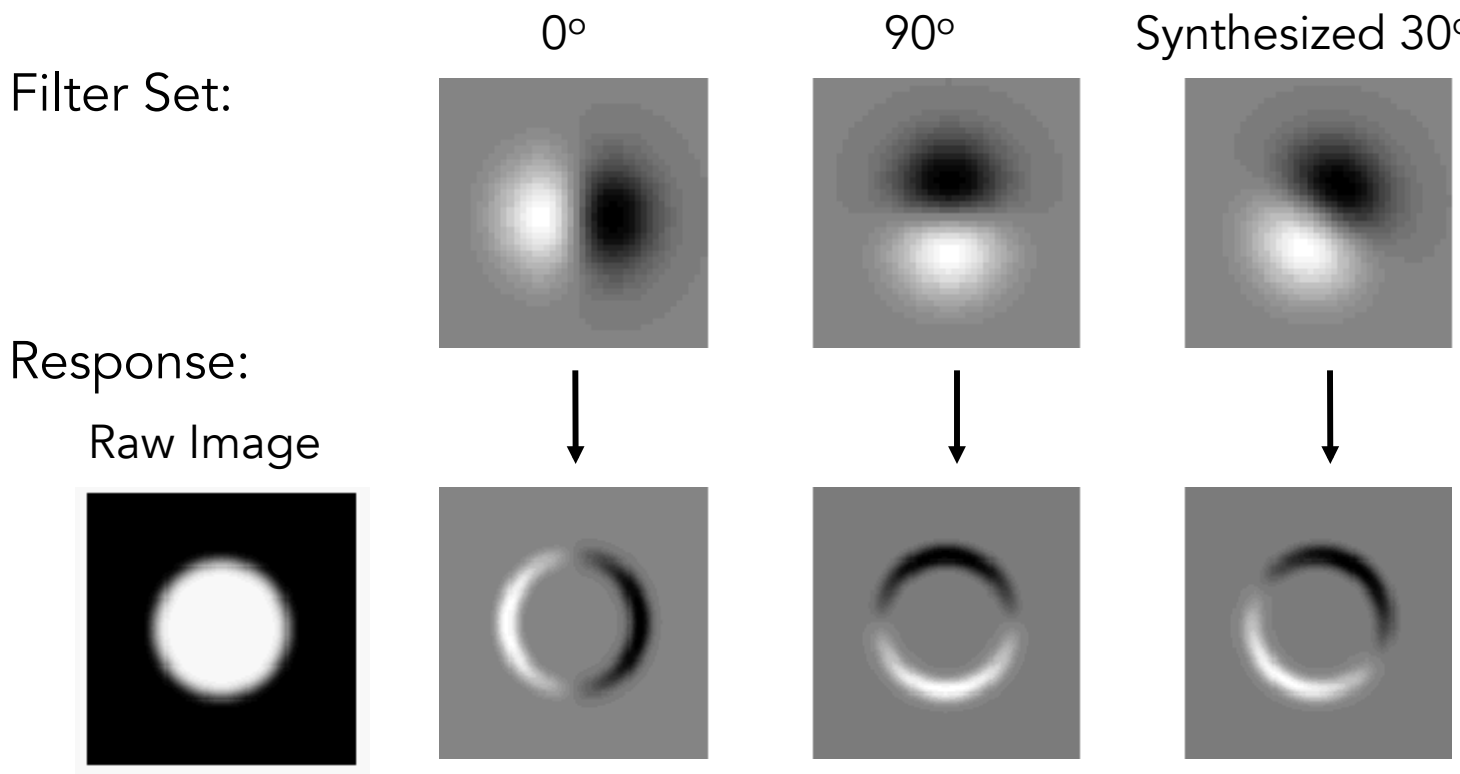
-Preserves all image information (we can go back to the image)

-Provides more independent channels of information than pixel values (we can mess with each band independently)

Why are they steerable?

"Steerability"-- the ability to synthesize a filter of any orientation from a linear combination of filters at fixed orientations.

$$G_\theta^1 = \cos(\theta)G_0^1 + \sin(\theta)G_{90}^1$$

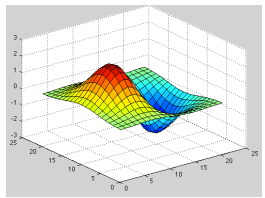


Taken from:
W. Freeman, T. Adelson, "The Design and Use of Sterrable Filters", IEEE Trans. Patt, Anal. and Machine Intell., vol 13, #9, pp 891-900, Sept 1991

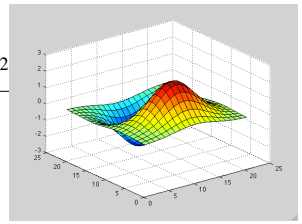
Steerable filters

Derivatives of a Gaussian:

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



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



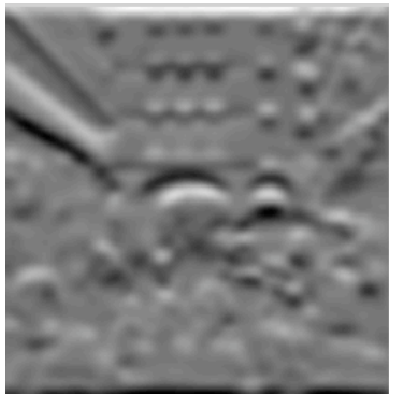
An arbitrary orientation can be computed as a linear combination of those two basis functions:

$$h_\alpha(x,y) = \cos(\alpha) h_x(x,y) + \sin(\alpha) h_y(x,y)$$

The representation is "shiftable" in orientation: We can interpolate any other orientation from a finite set of basis functions.

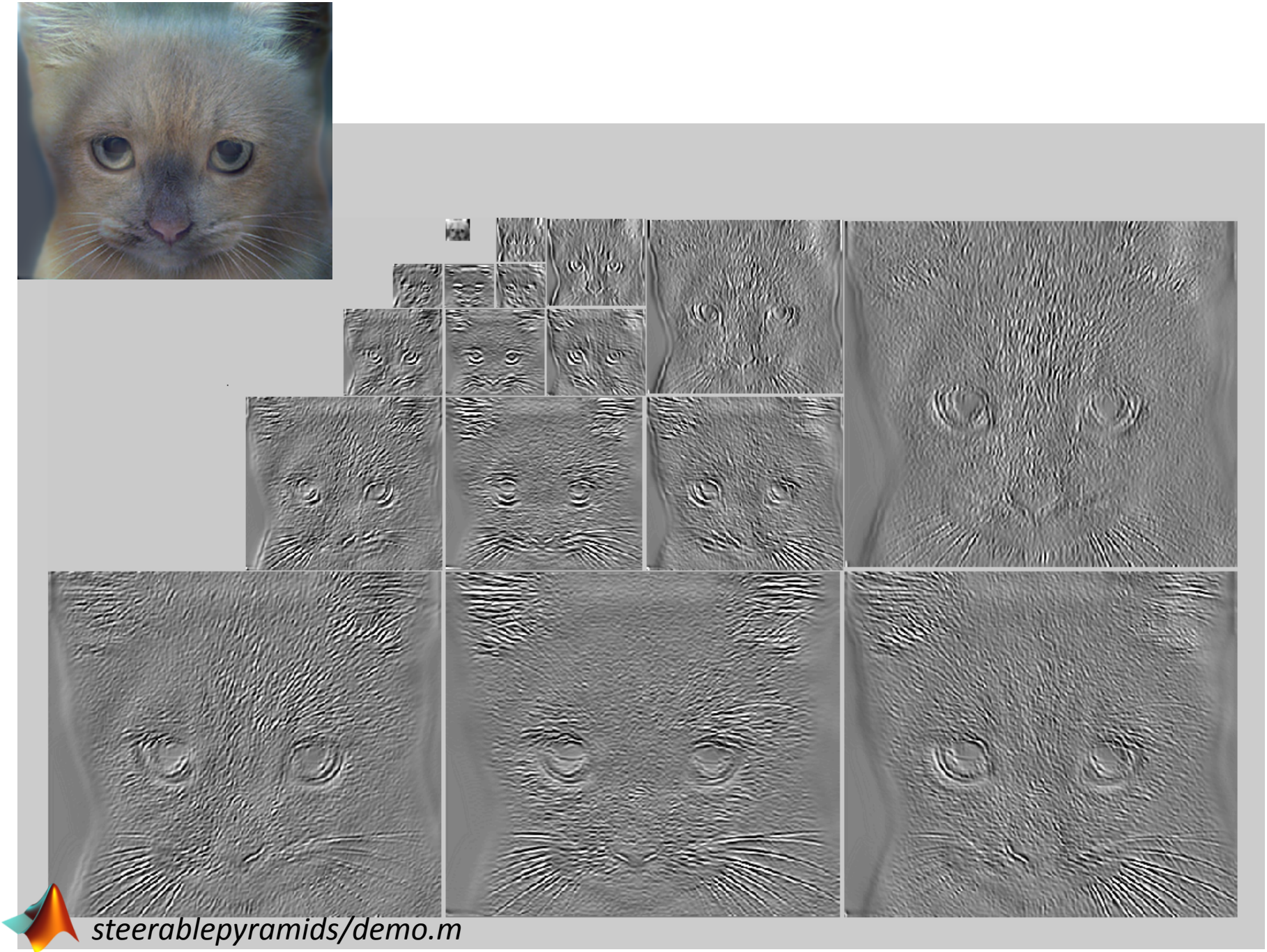
cos(α)

+sin(α)



=

Freeman & Adelson 92



steerablepyramids/demo.m

Monroe

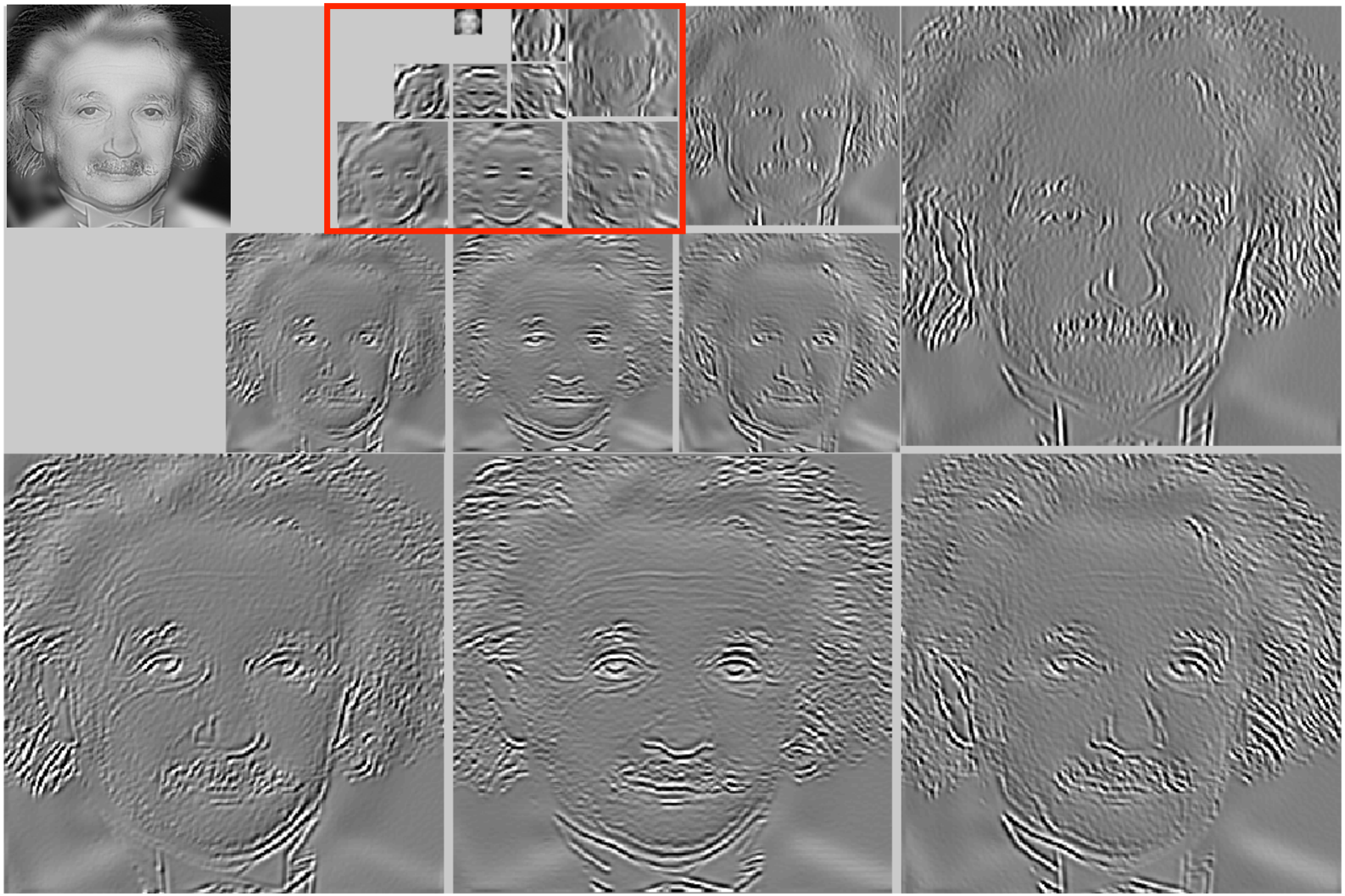
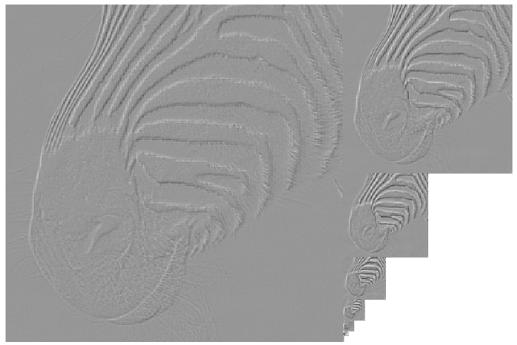


Image pyramids

- Gaussian

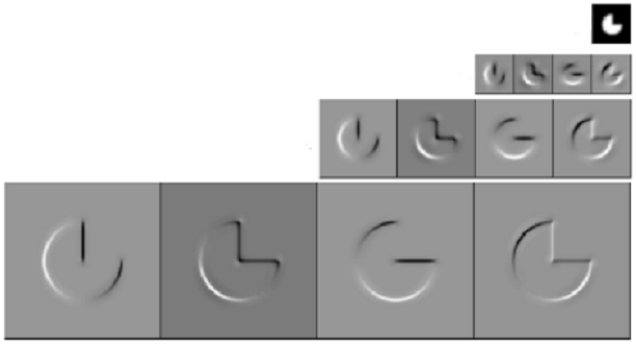
Progressively blurred and subsampled versions of the image. Adds scale invariance to fixed-size algorithms.

- Laplacian



Shows the information added in Gaussian pyramid at each spatial scale. Useful for noise reduction & coding.

- Steerable pyramid



Shows components at each scale and orientation separately. Non-aliased subbands. Good for texture and feature analysis.

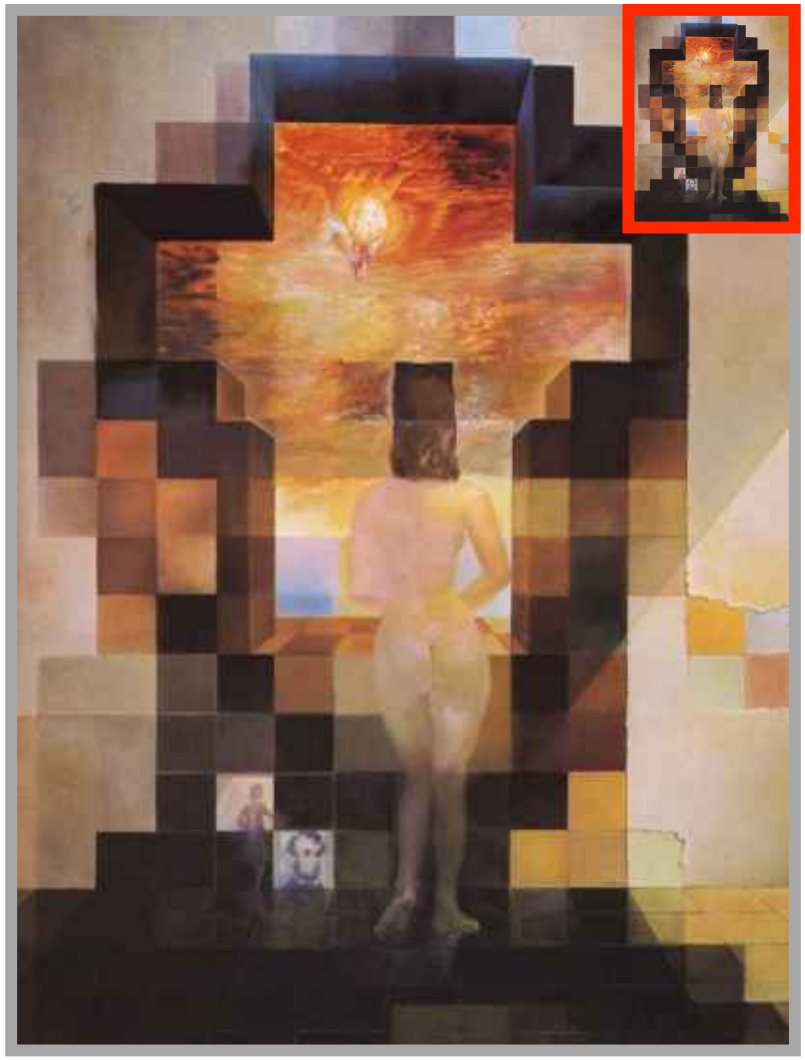
Why use these representations?

- Handle real-world size variations with a constant-size vision algorithm.
- Remove noise
- Analyze texture
- Recognize objects, scenes
- Label image features

Visual Perception

For computer vision systems to "understand art" other properties of human visual perception must be taken into account

Salvador Dali



Changes of Meaning with Scale

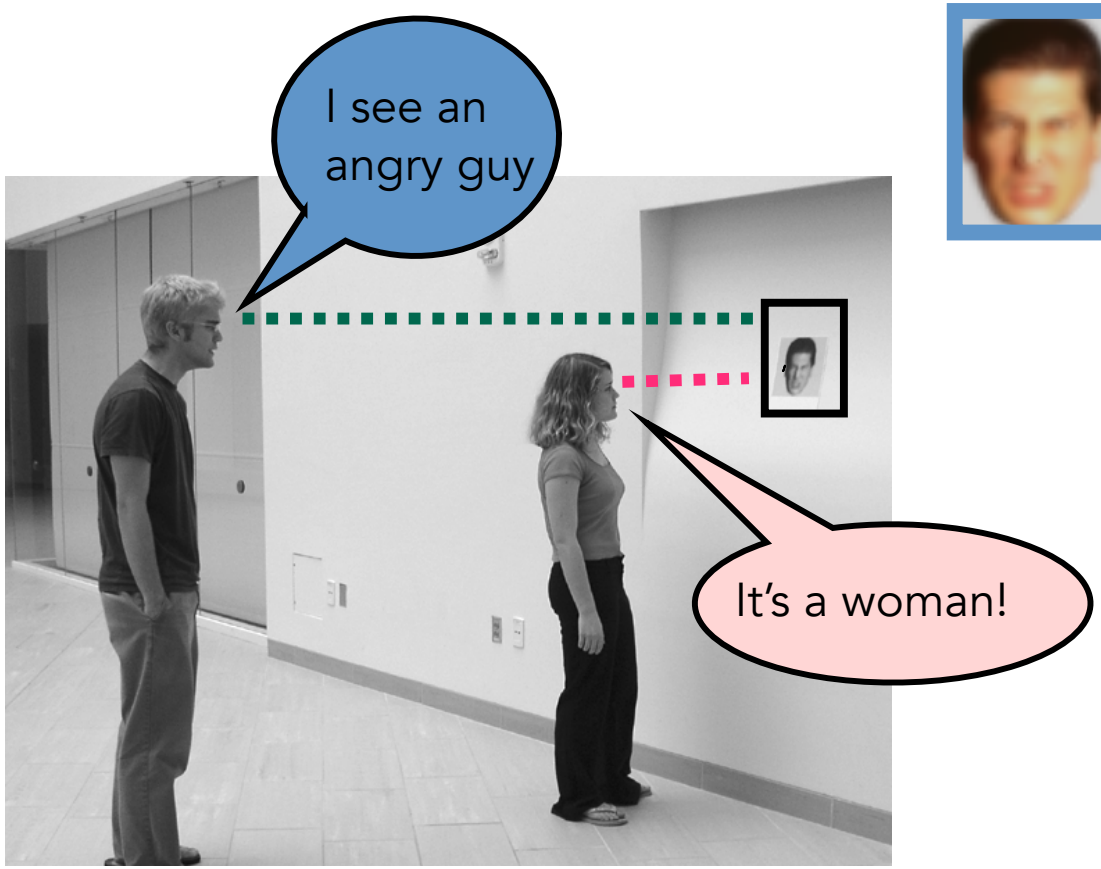
Strong percept of a woman

Strong percept of a skull

Looking at the image one way, a skull appears, comprised of the woman's form and her reflection in a mirror (the outline of the skull itself).

"All Is Vanity", by C. Allan Gilbert 1873-1929.

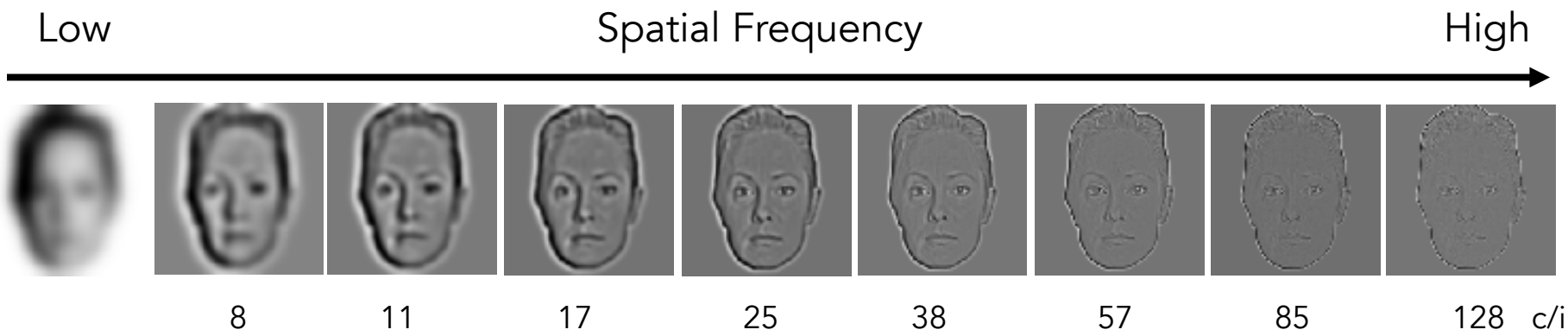
Hybrid Images

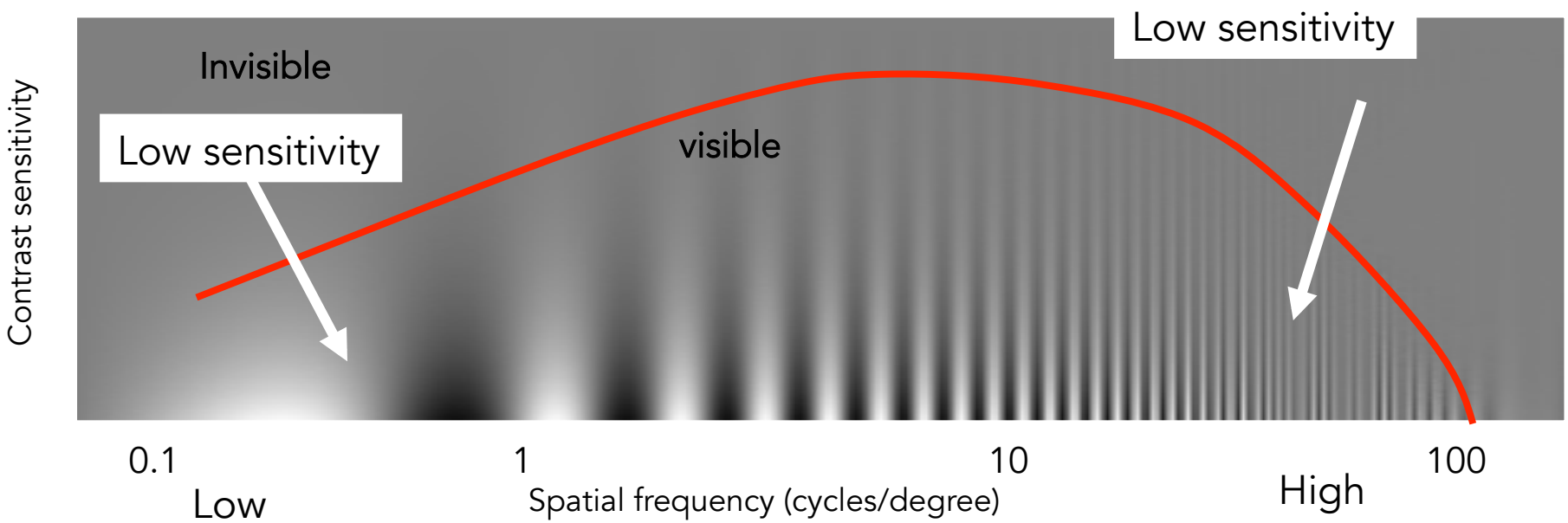


From Far Away

Up Close

Multiscale subband decomposition + Human Contrast Sensitivity Function





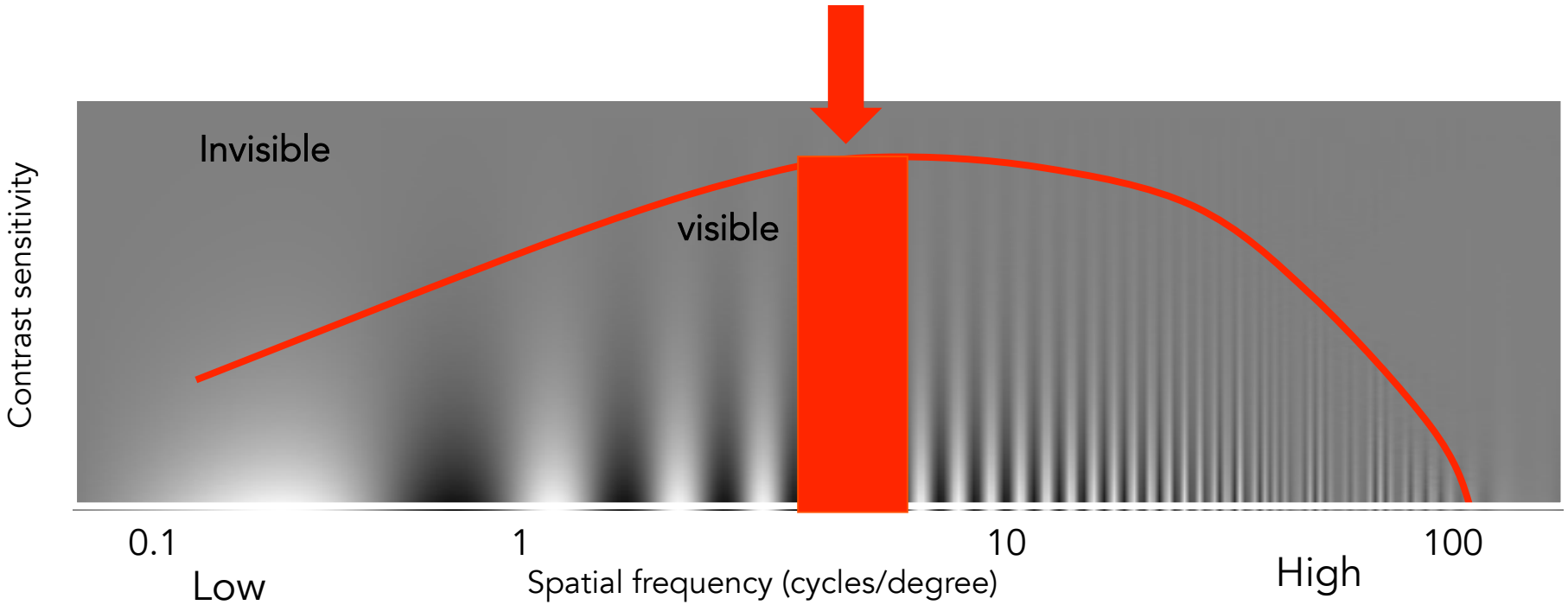
Contrast Sensitivity Function

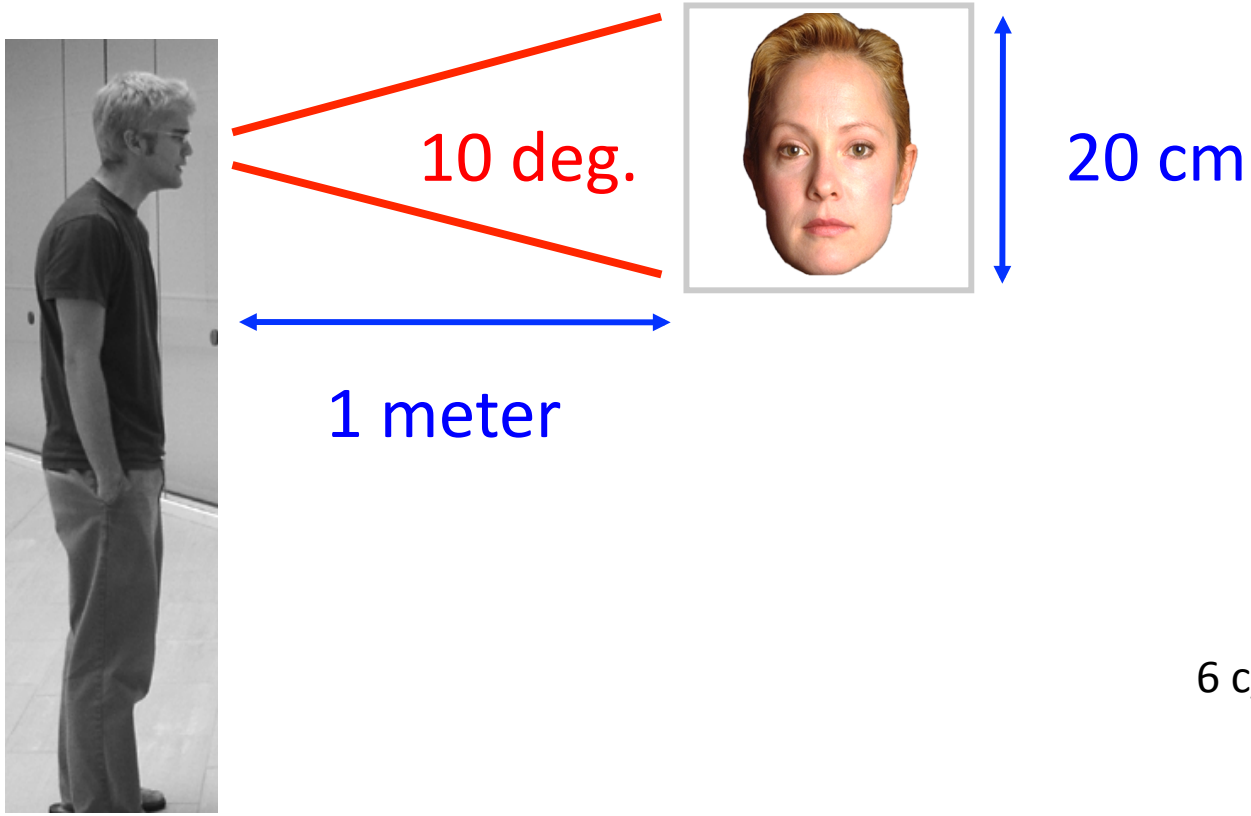
Low

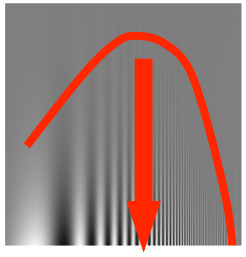
High

Maximum sensitivity

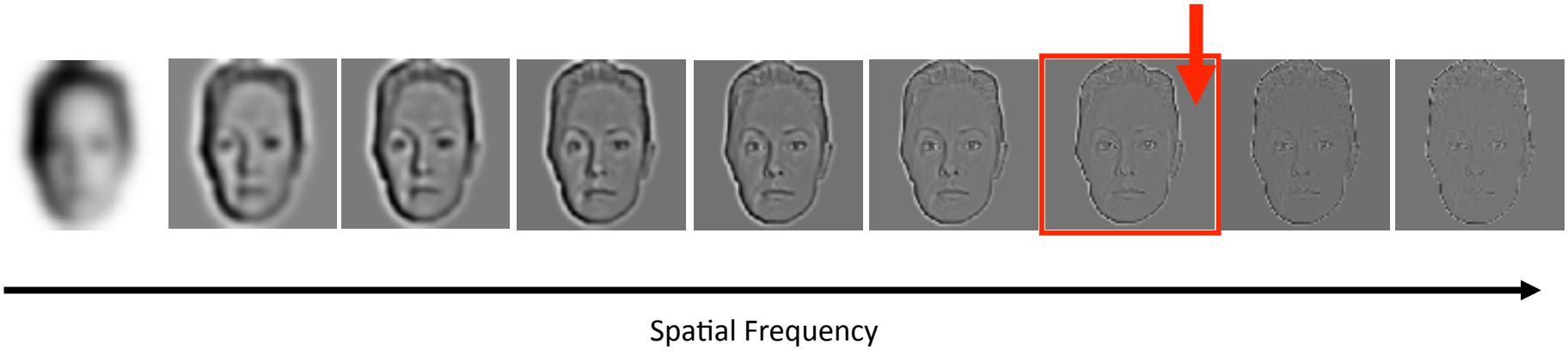
~ 6 cycles / degree of visual angle

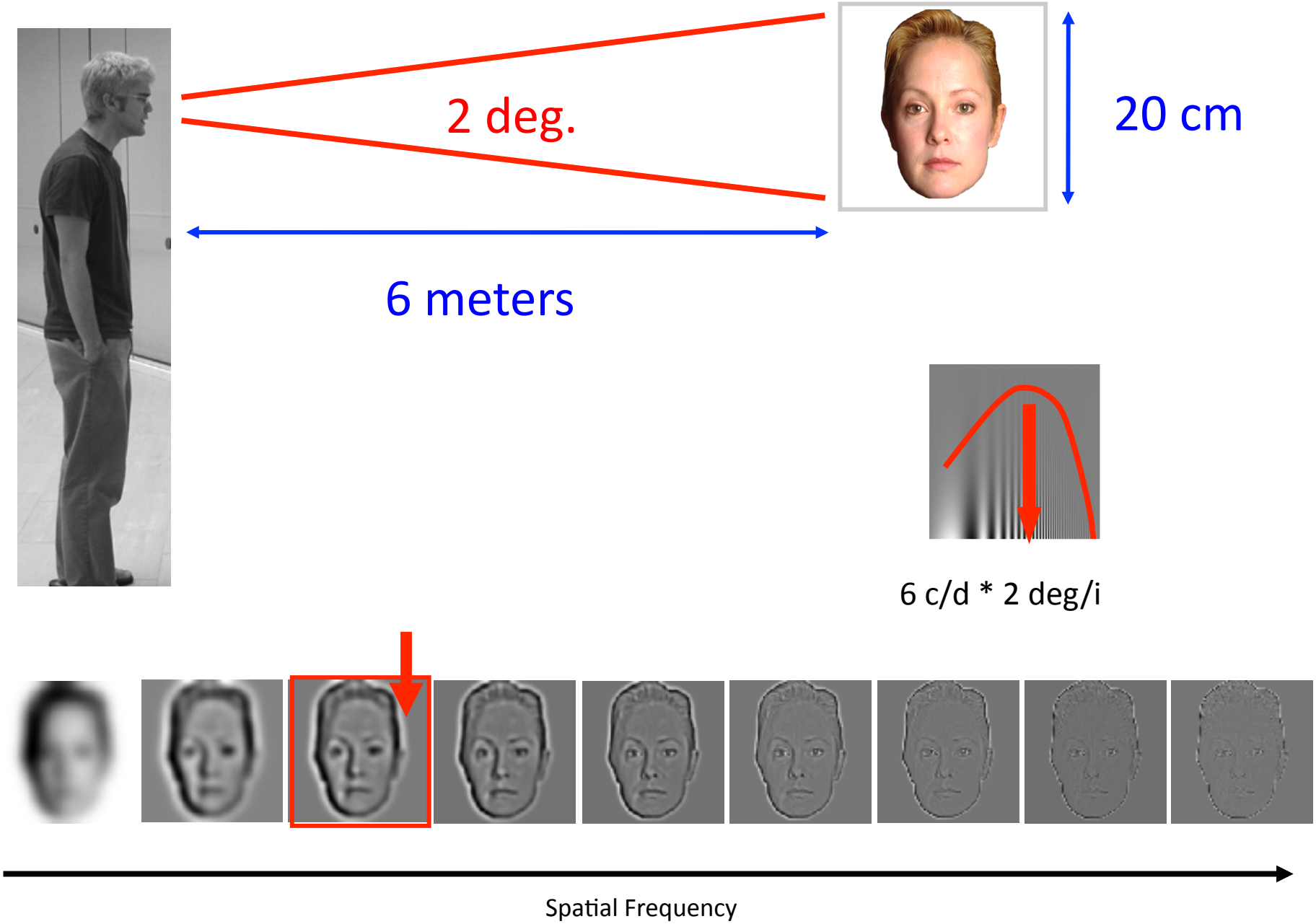




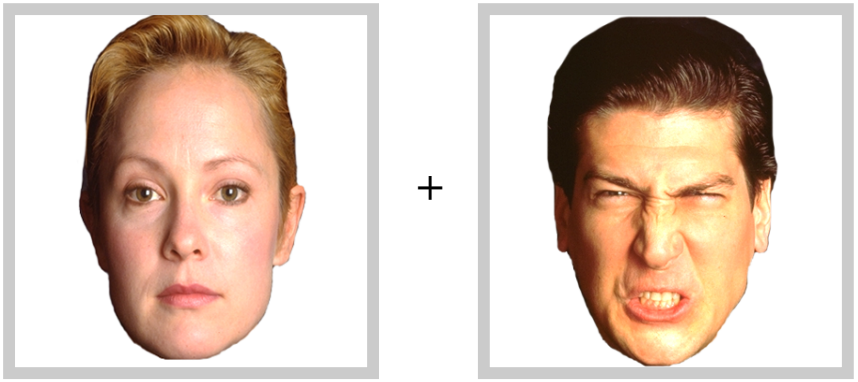


6 c/d * 10 deg/i




2 deg.
20 cm
6 meters
6 c/d * 2 deg/i
Spatial Frequency

Perception of hybrid images

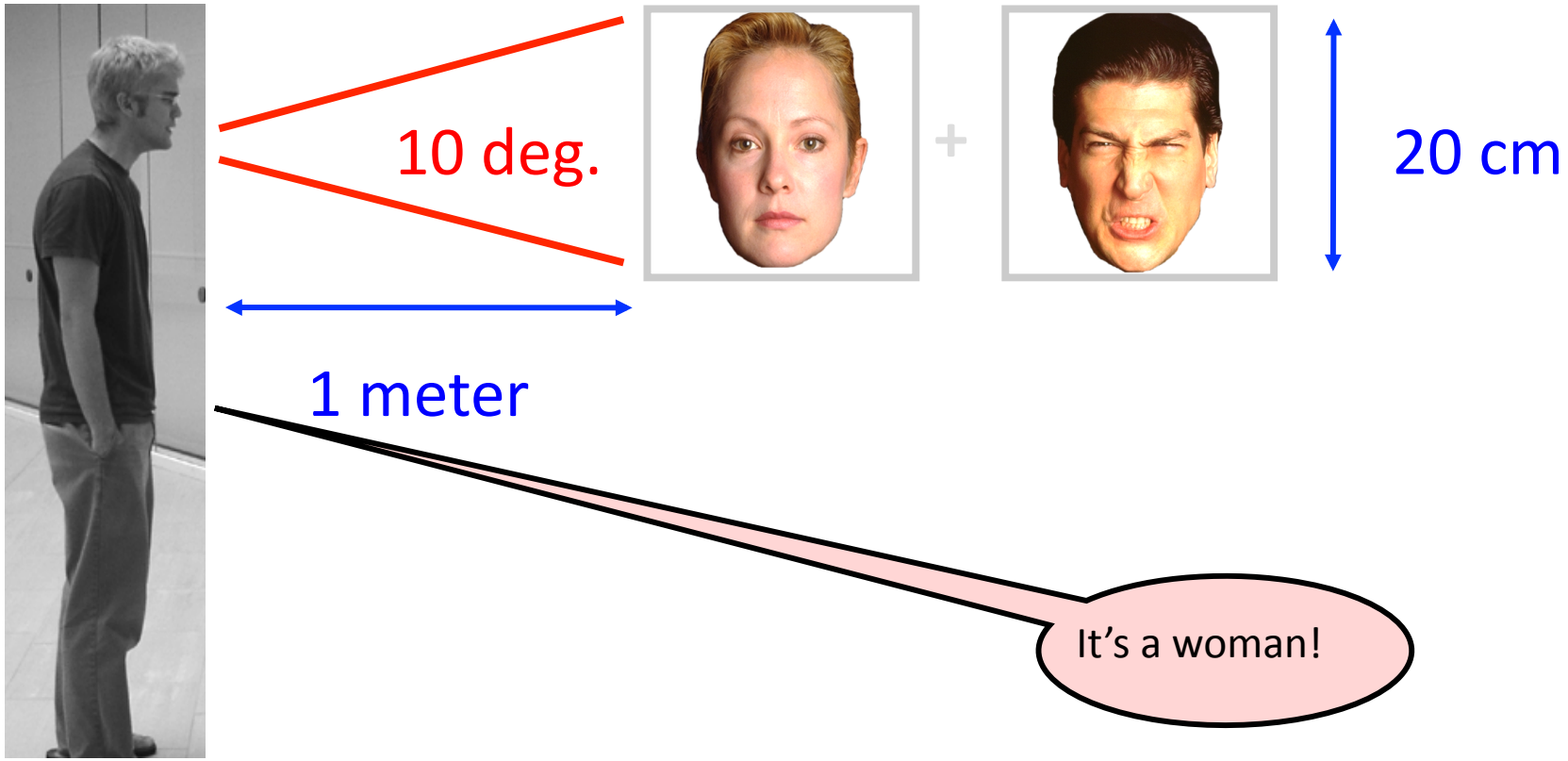


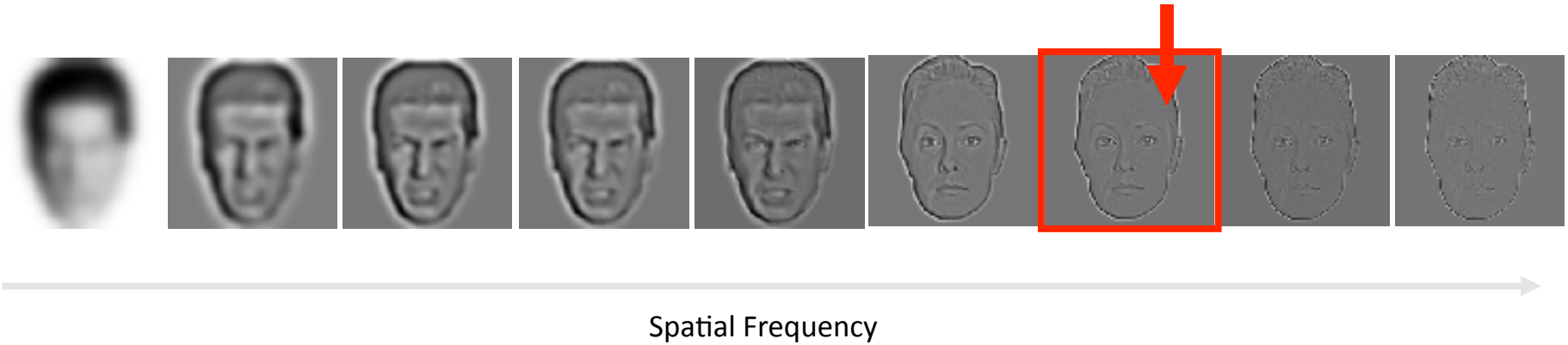
+

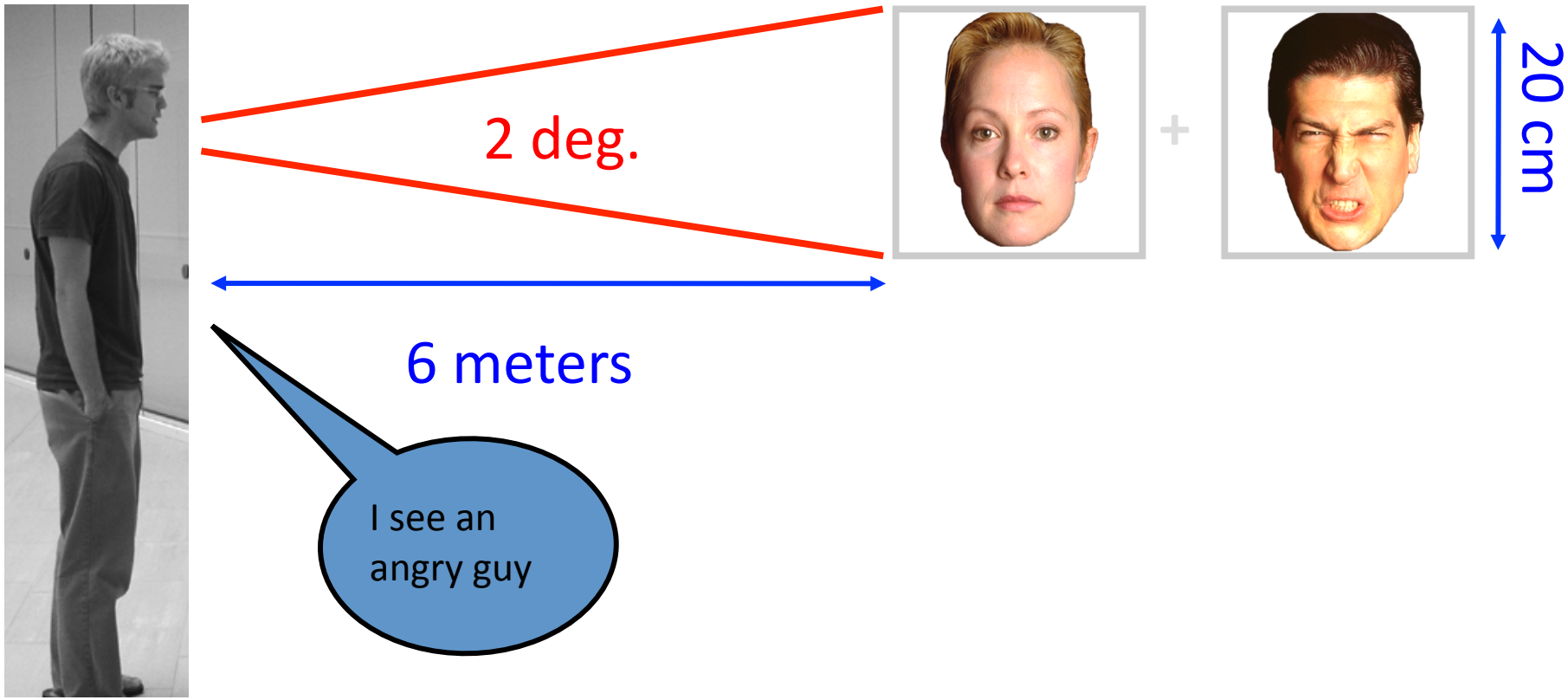
Male dominance

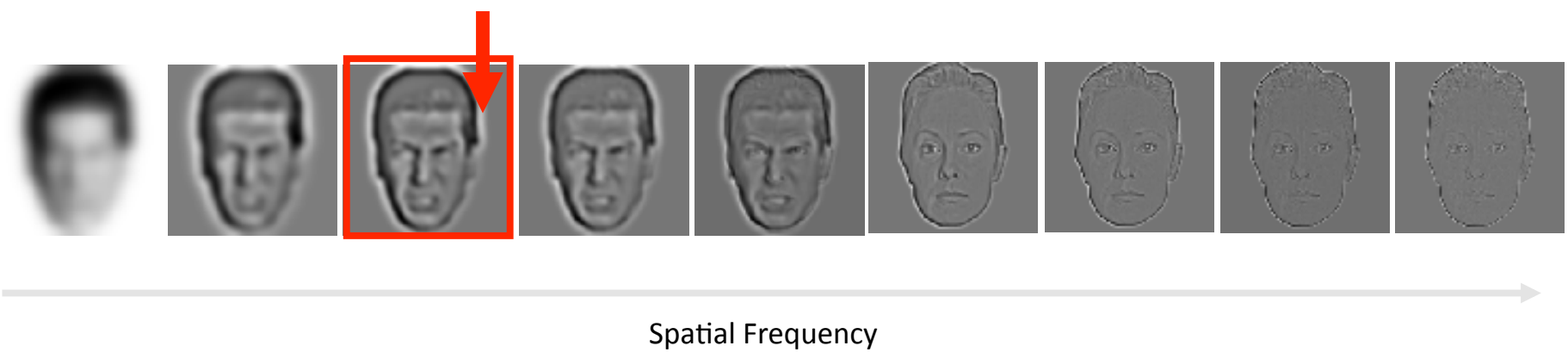
Female dominance

A man or a woman ?

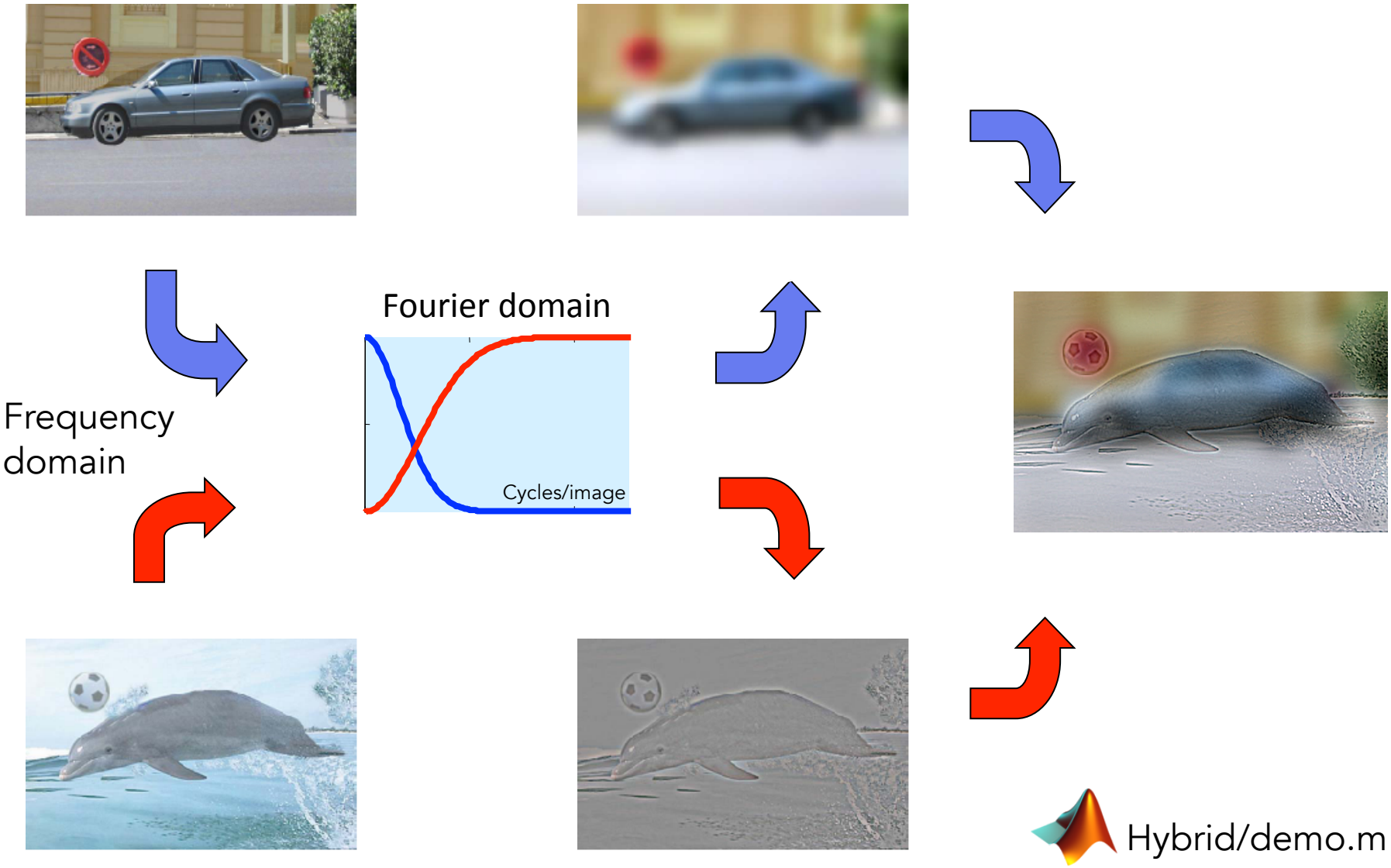

10 deg.
+
20 cm
1 meter
It's a woman!


Spatial Frequency


2 deg.
6 meters
20 cm
I see an
angry guy


Spatial Frequency

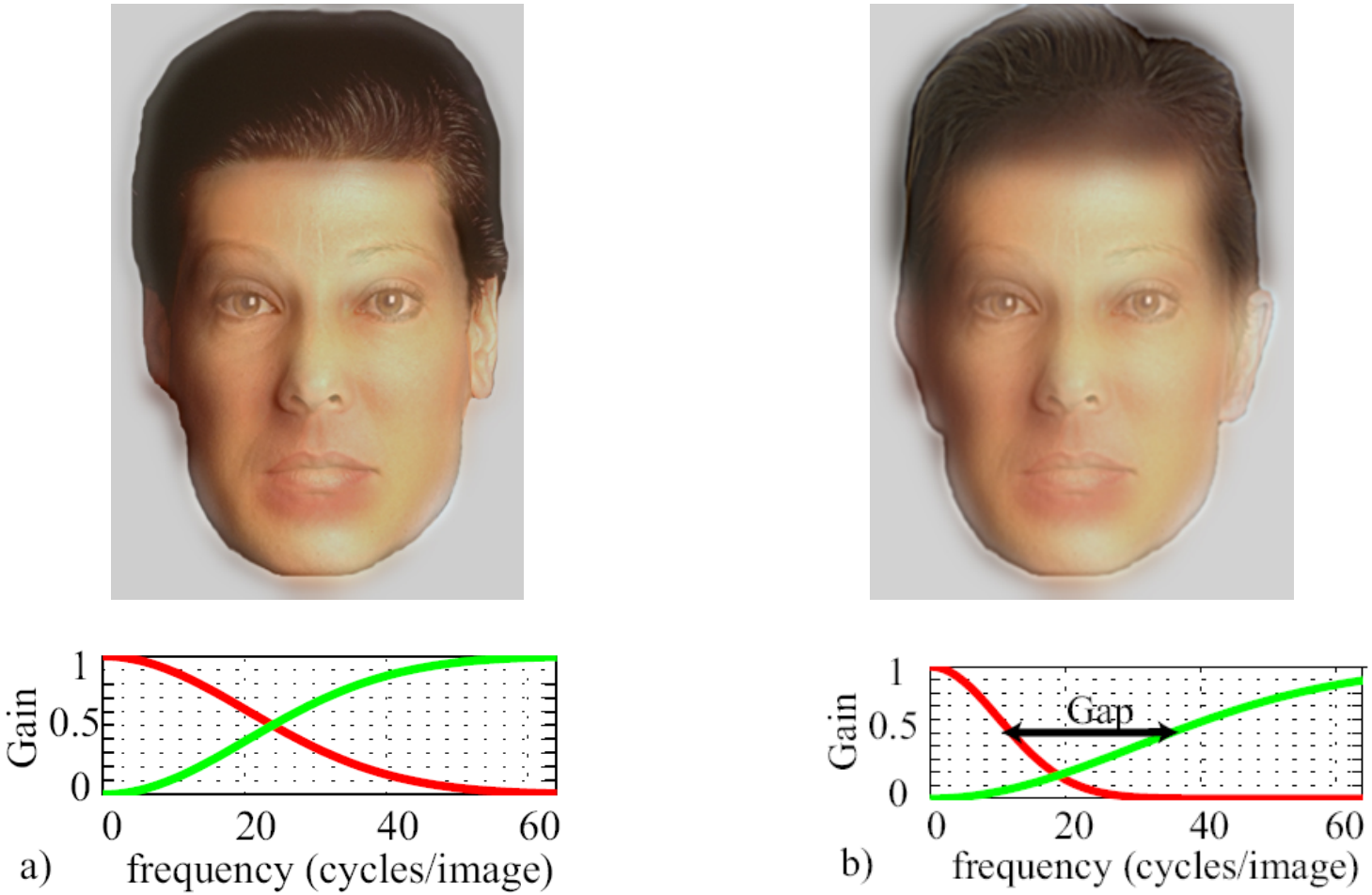
Perception of hybrid images



Hybrid/demo.m

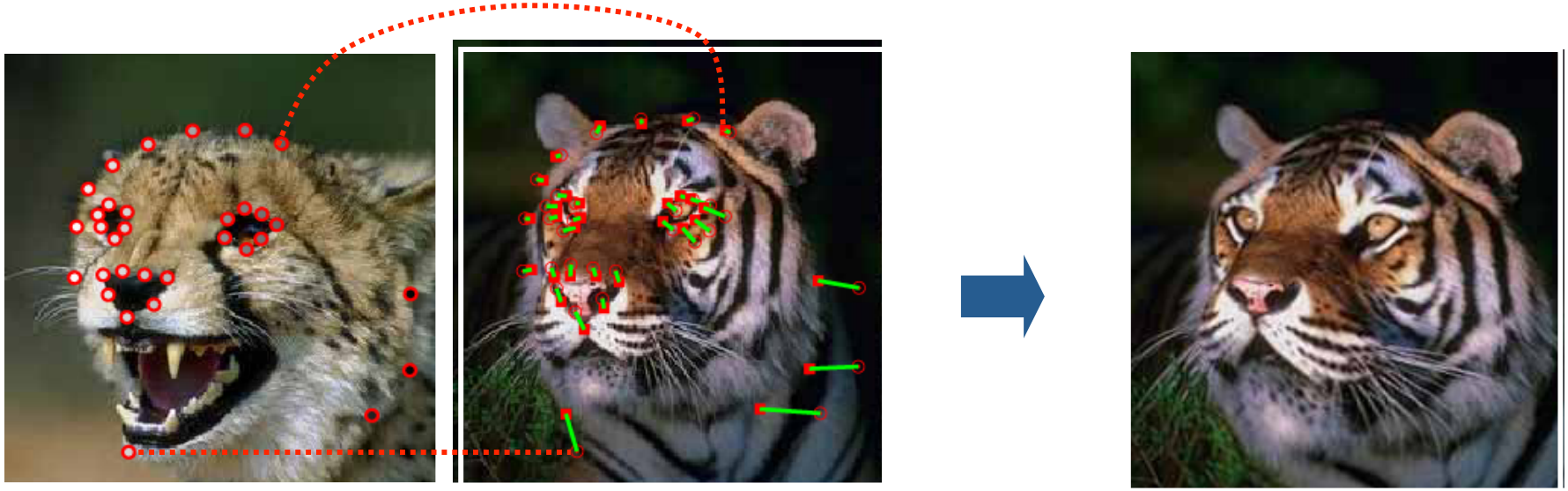
Frequency Gap

Hybrid/demo.m

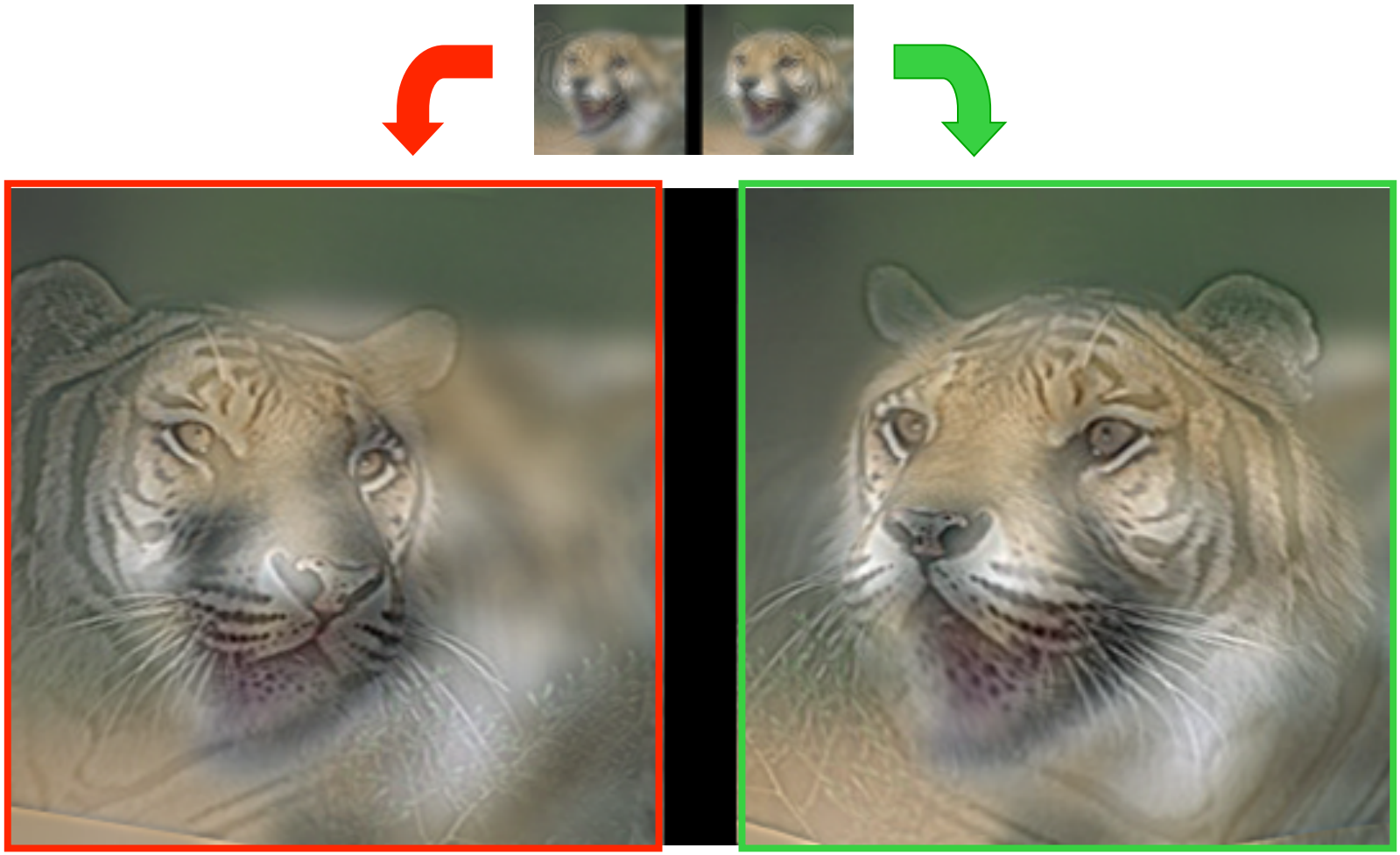


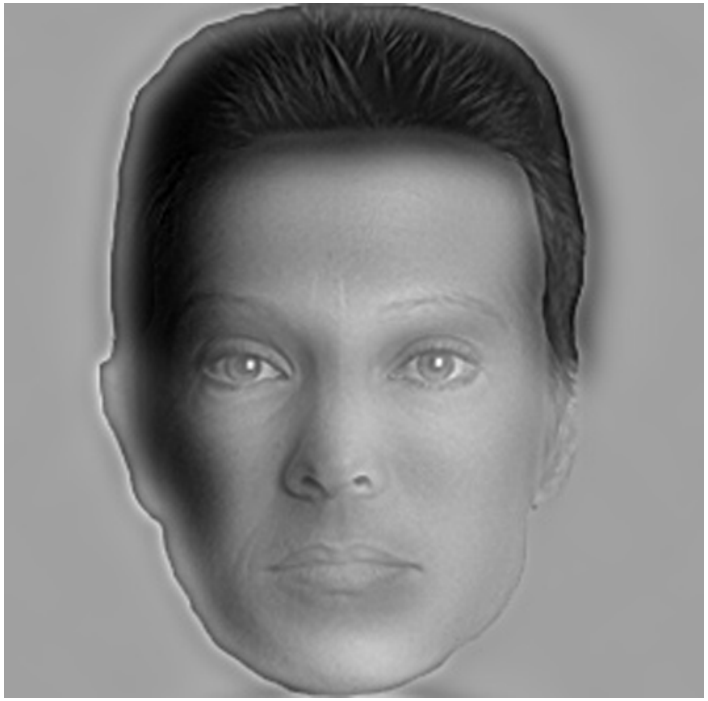
Alignment

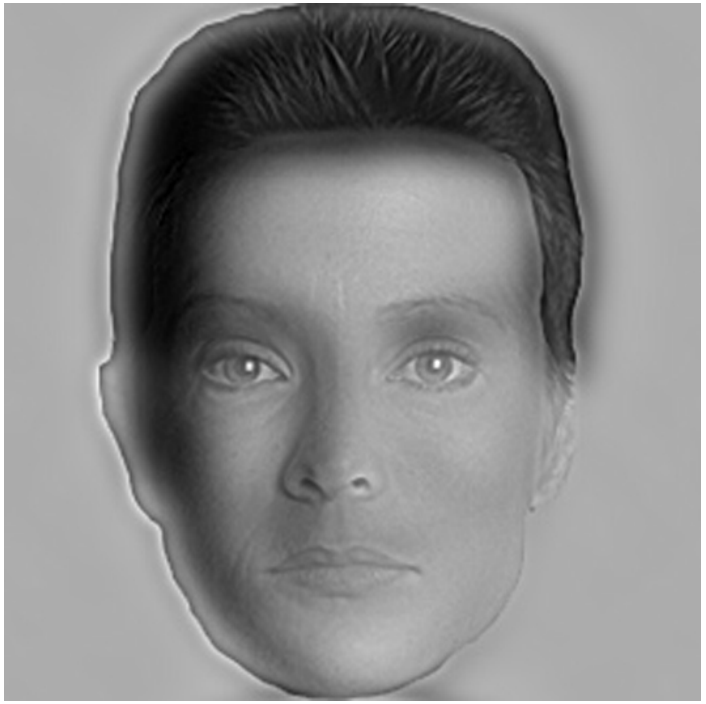
Alignment

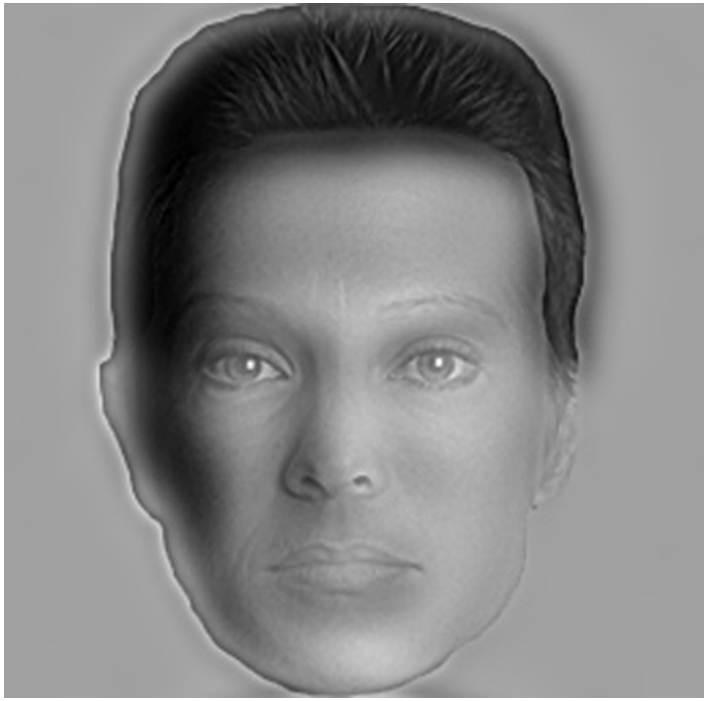


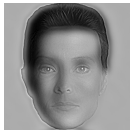
Alignment

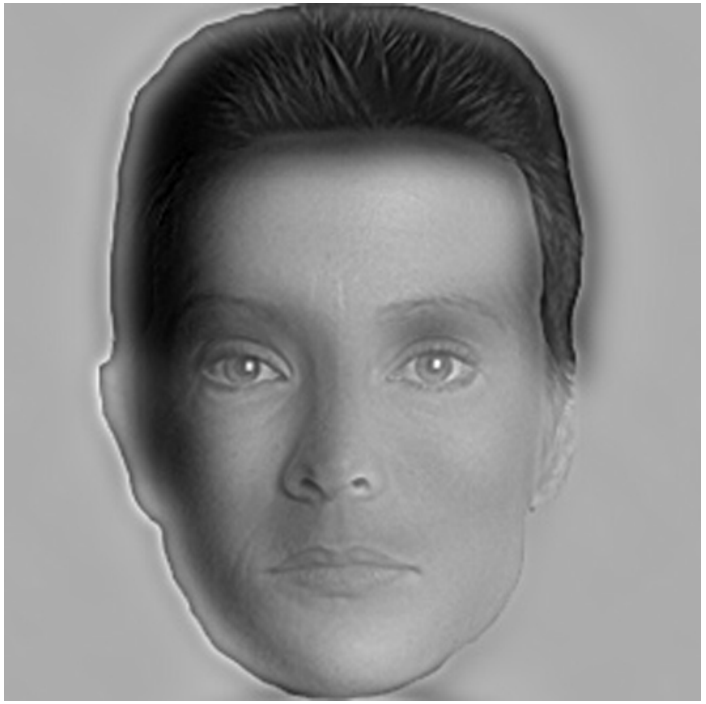










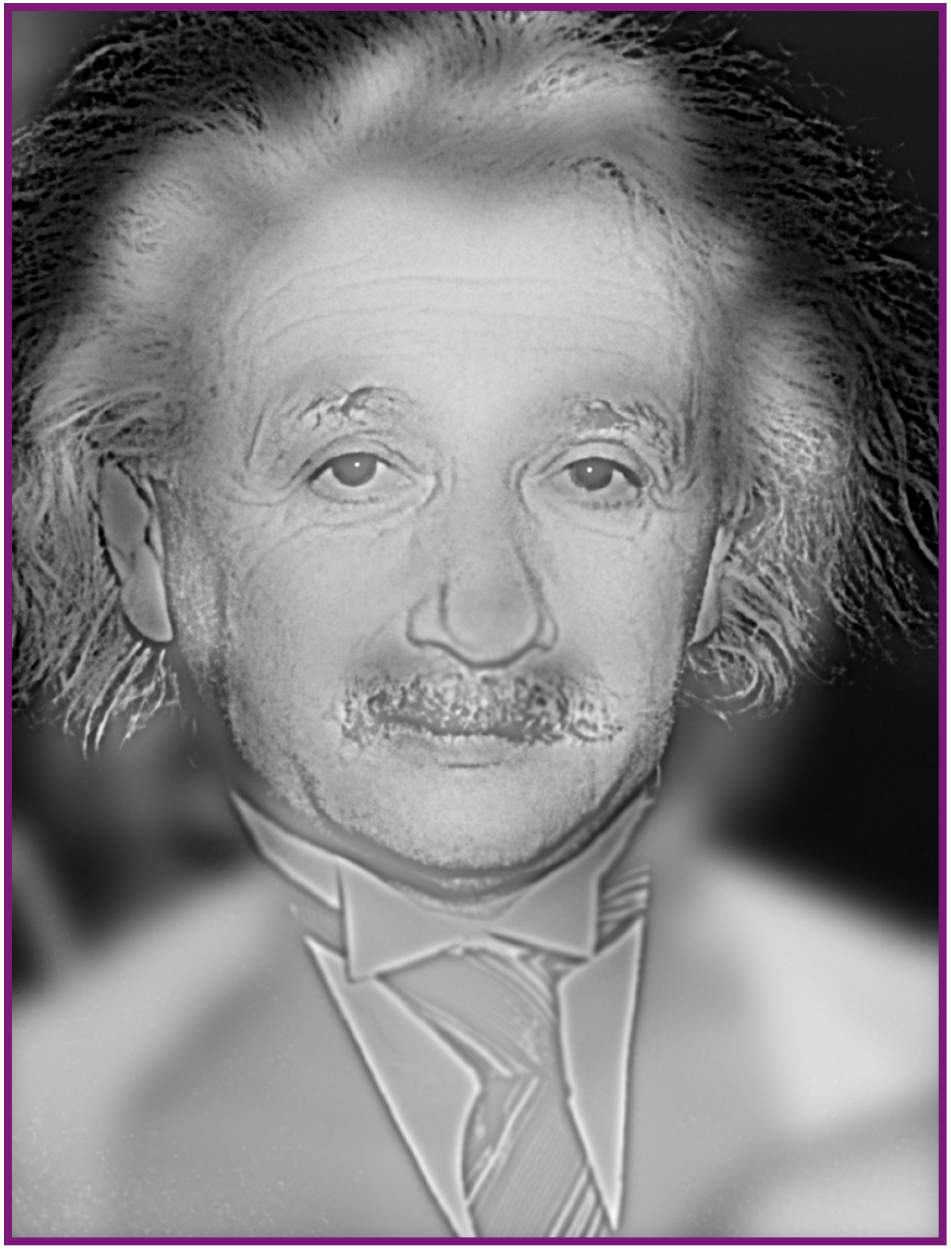


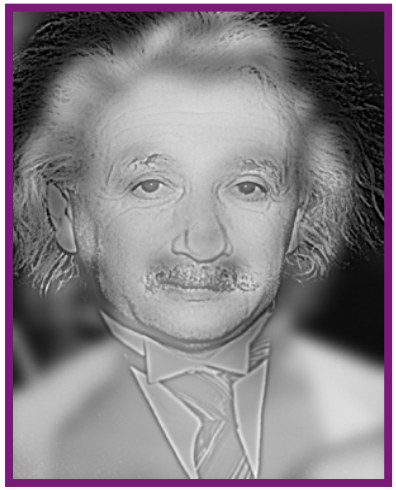
Changing expression

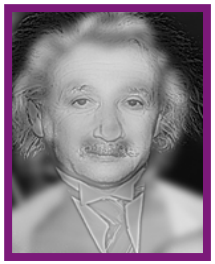
Changing expression

Changing expression

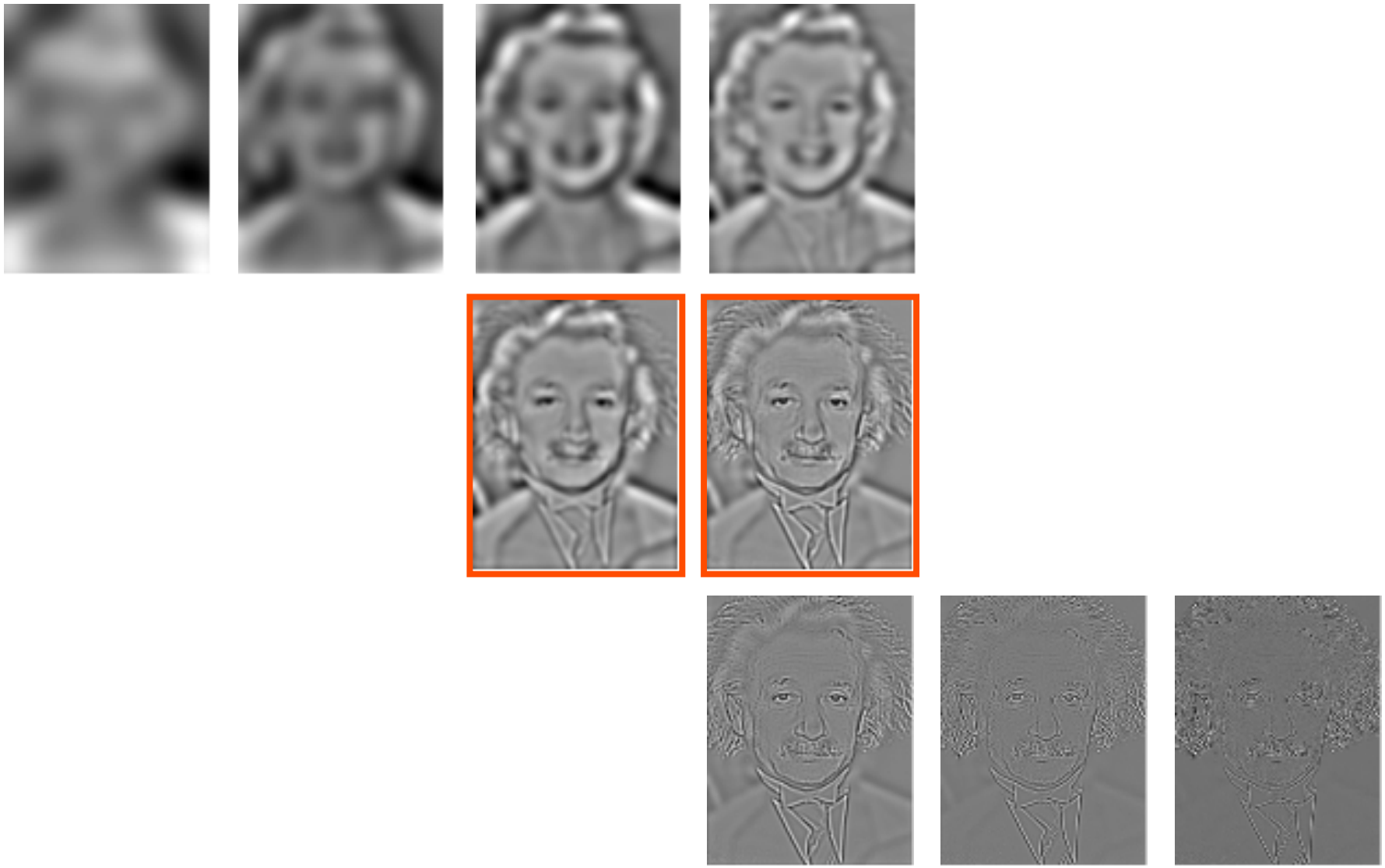
Sad ⟷ Surprised



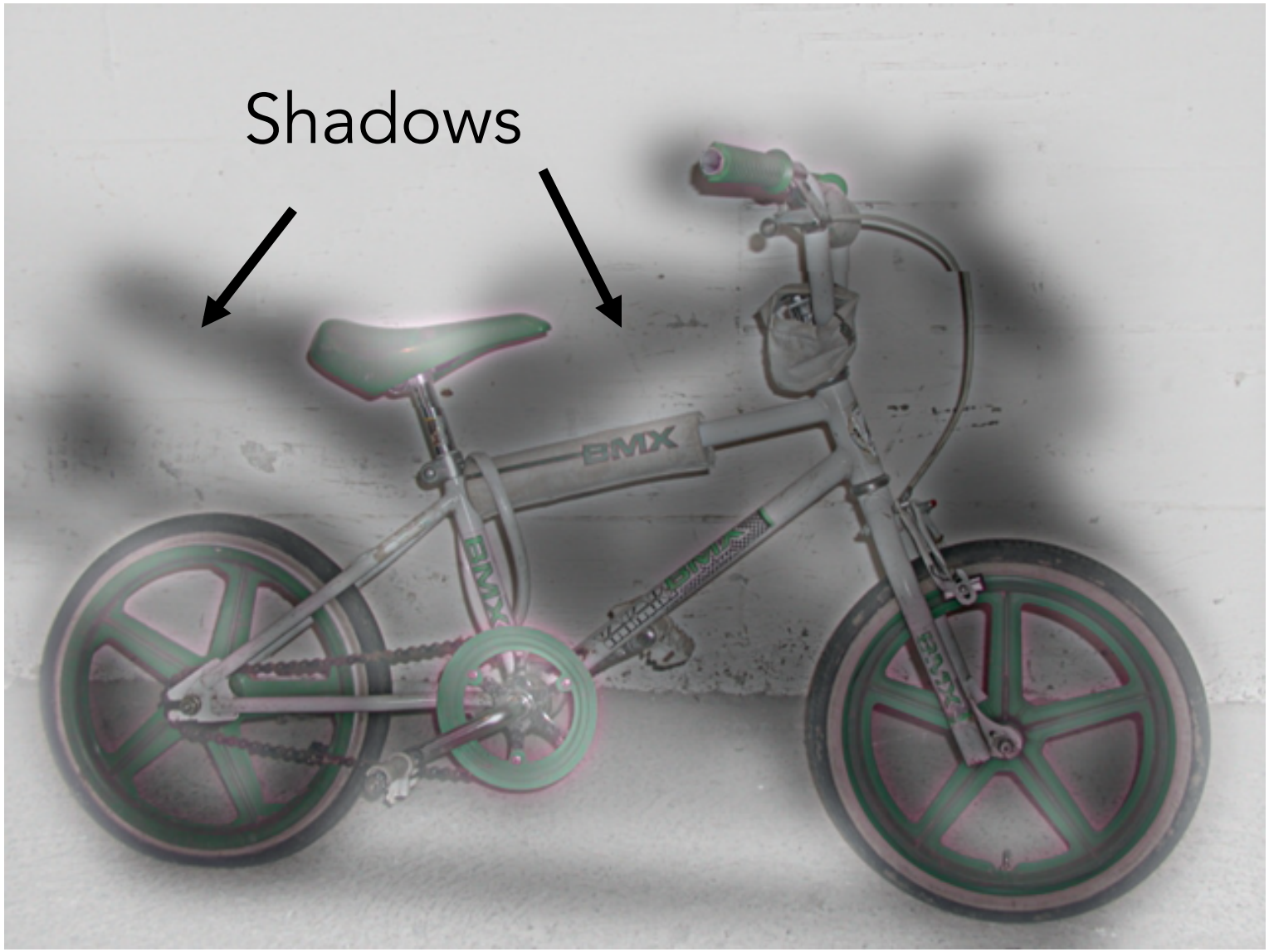




Subband decomposition of hybrid

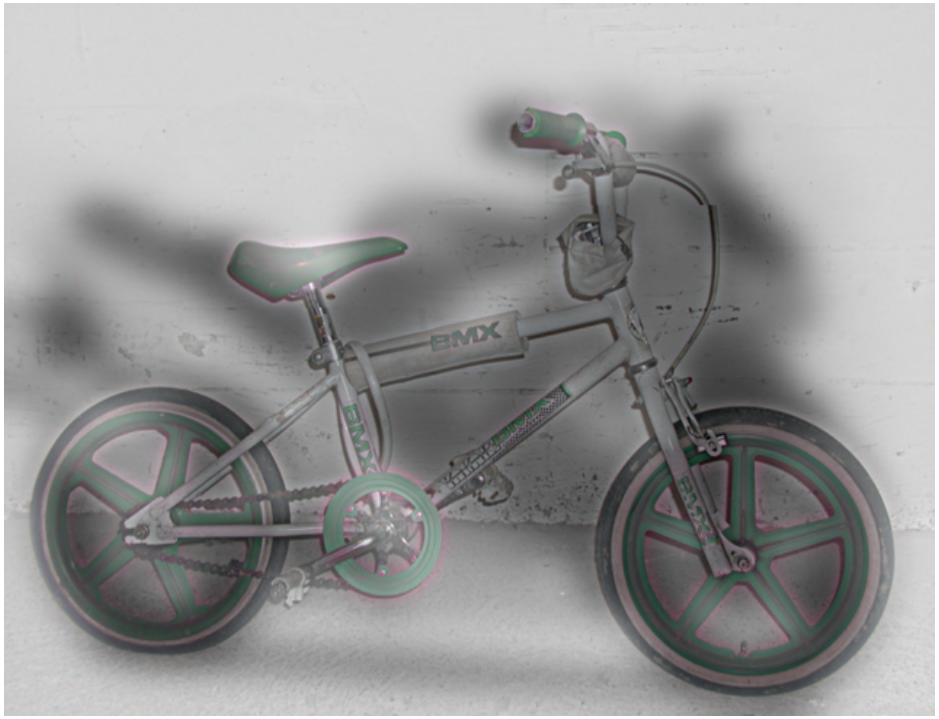


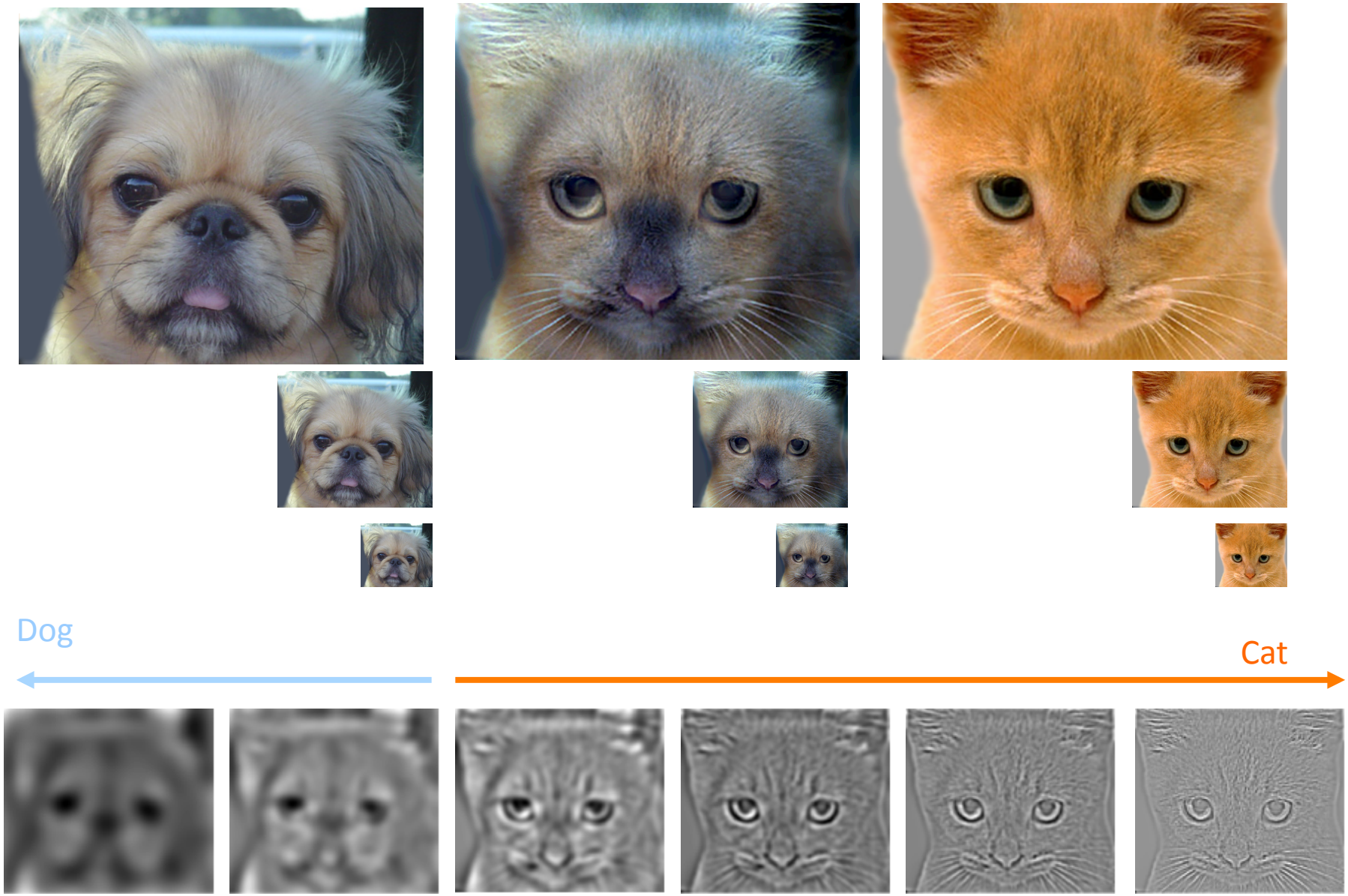
Perceptual Grouping



Perceptual Grouping

Reorganization of the shadows as a motorcycle




Dog
Cat

Texture superposition

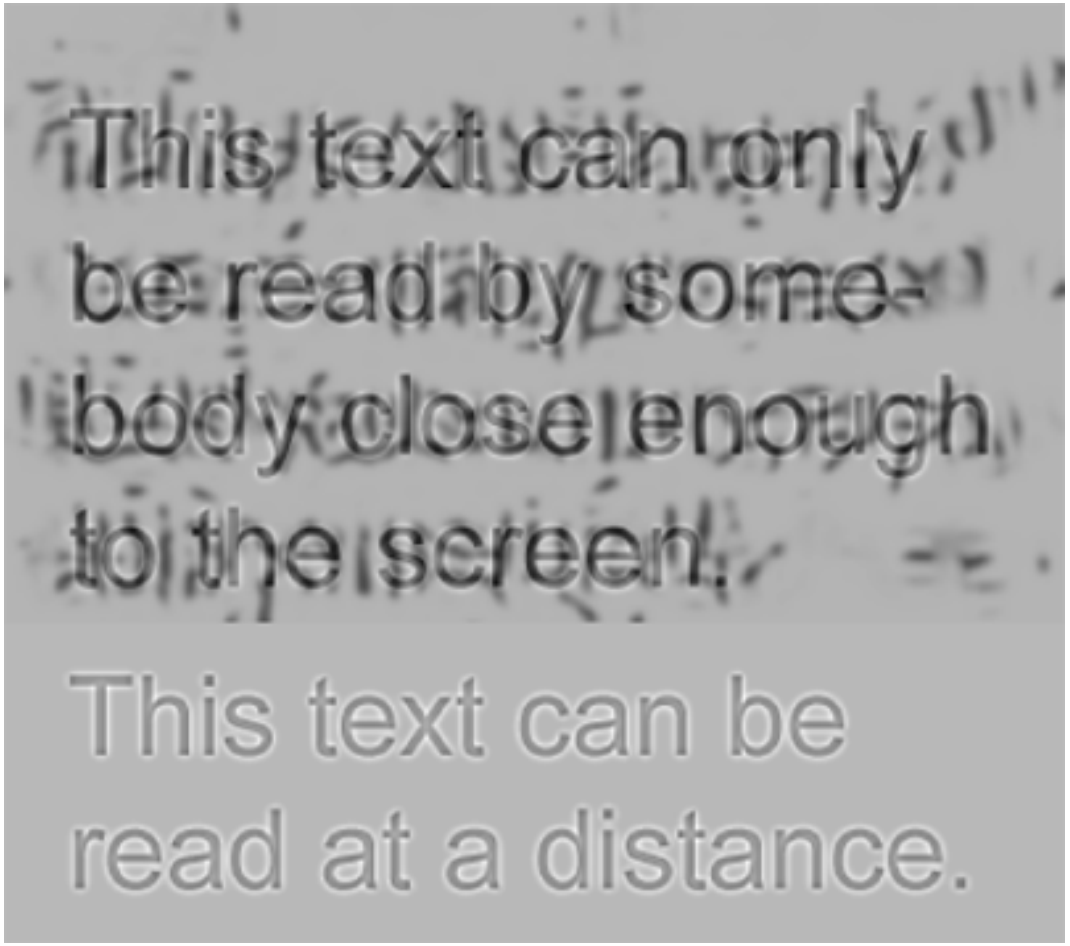
Texture superposition

Original | Cat mask | Transparency | Hybrid

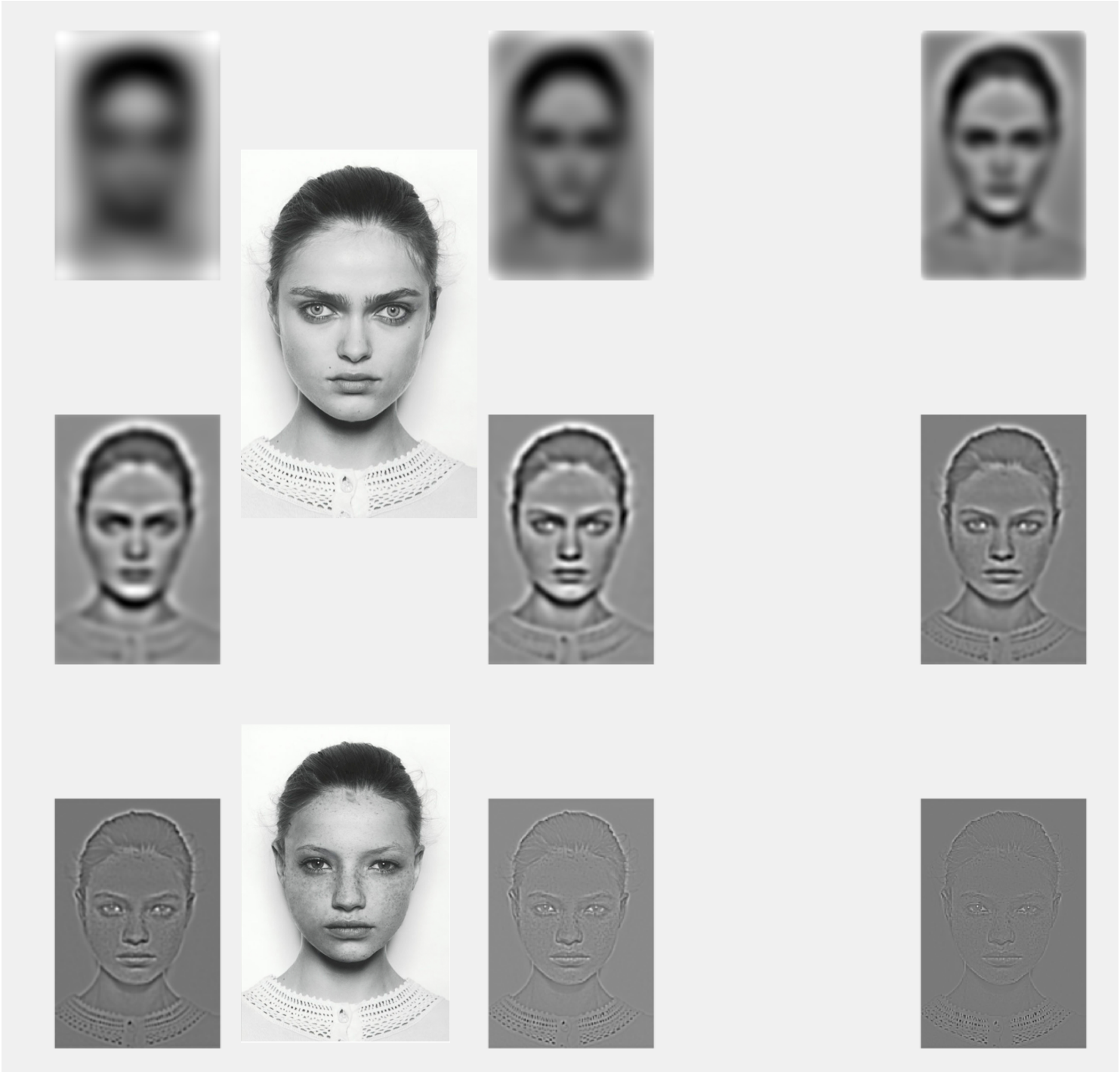
Private font

Low spatial frequencies in an hybrid can be used to mask text that will be readable otherwise

Private font



Low spatial frequencies in an hybrid can be used to mask text that will be readable otherwise



Project: MIT Museum interested in a 3D Hybrid Illusion

Exhibition at the MIT Museum

Contact Aude: oliva@mit.edu

Additional Slides
Human Perception

- principles of multi resolution perception not implemented in artificial vision systems (yet!)-
(not covered in class)

Real world perception

In the world

Far — *Close*

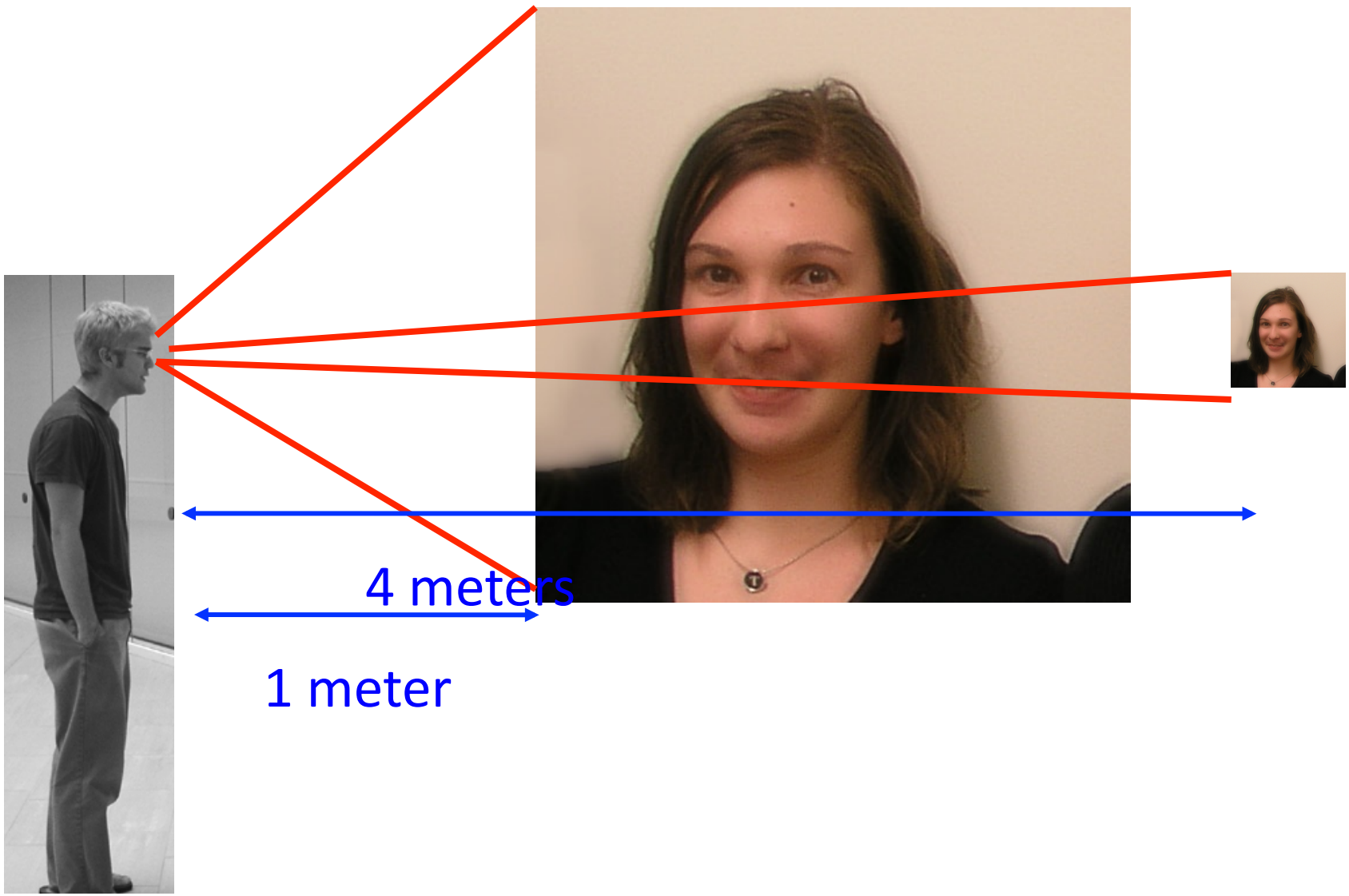
Approaching: Gaining information

Receding: Loosing information

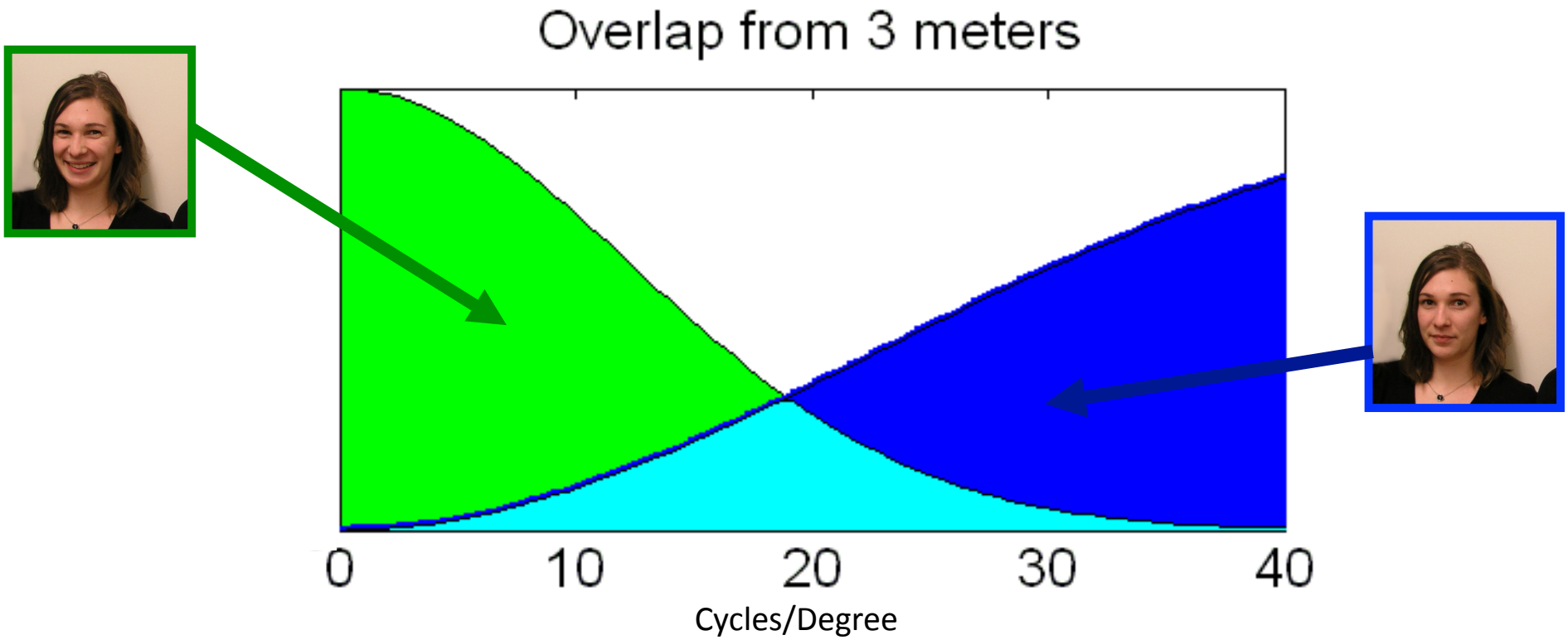
In scale space

Low spatial frequency (blurry) — *Full spatial frequency range (sharp)*

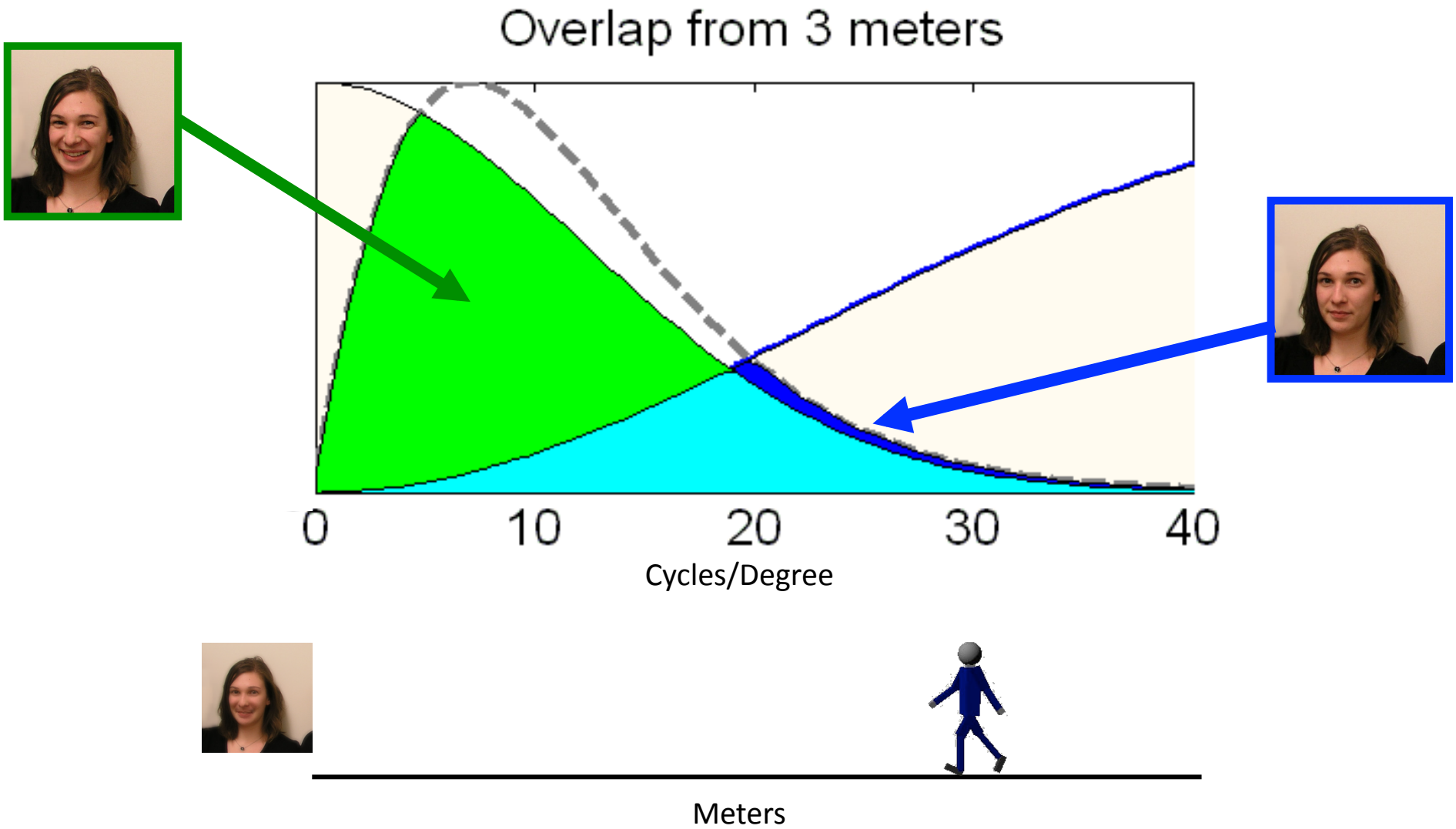
Hybrid Images & Distance



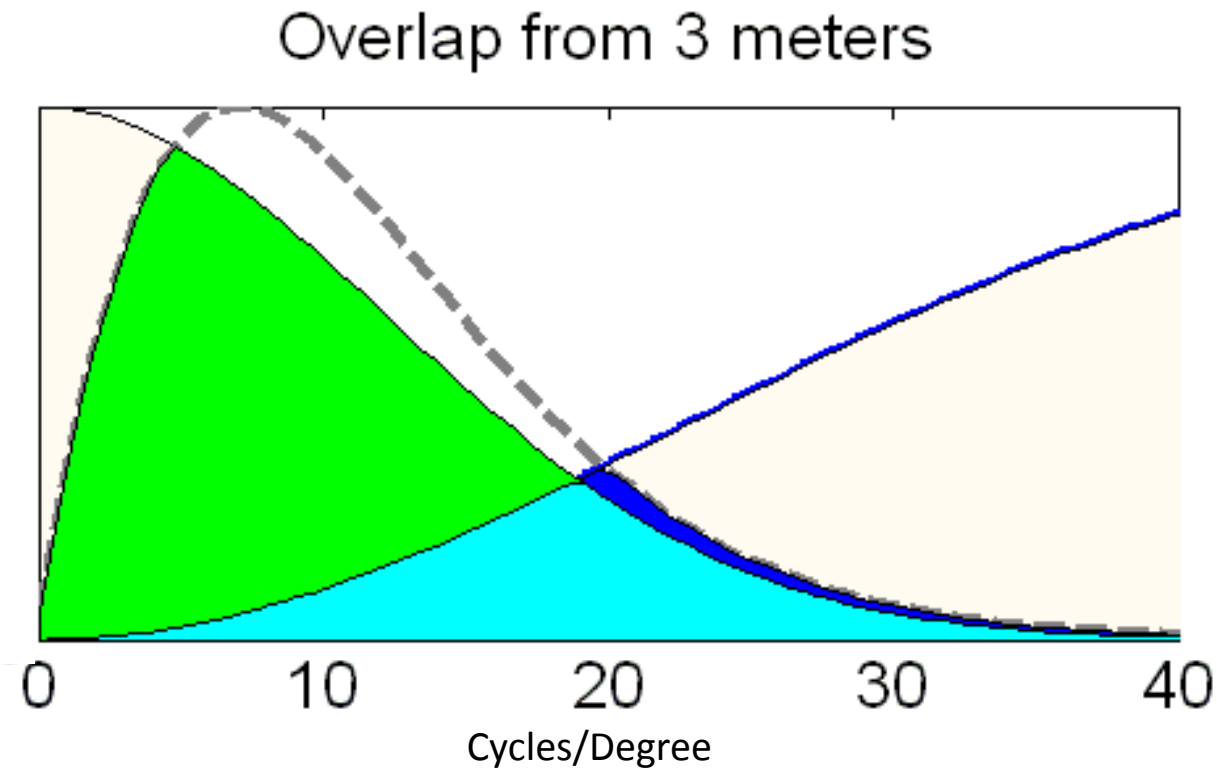
Which image do people see?

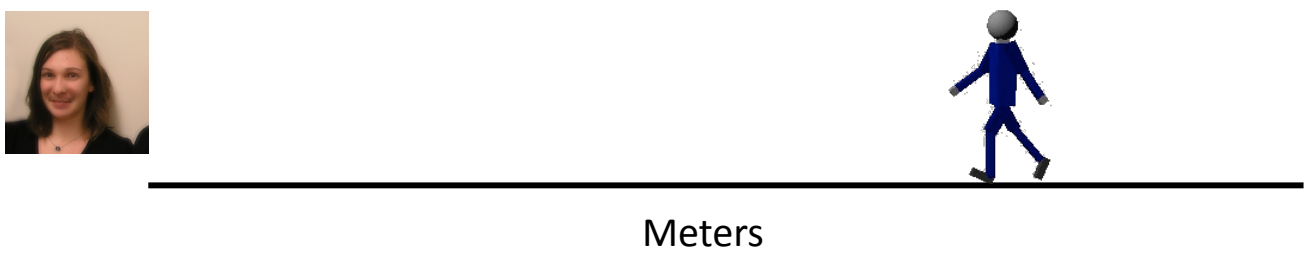


Which image do people see?

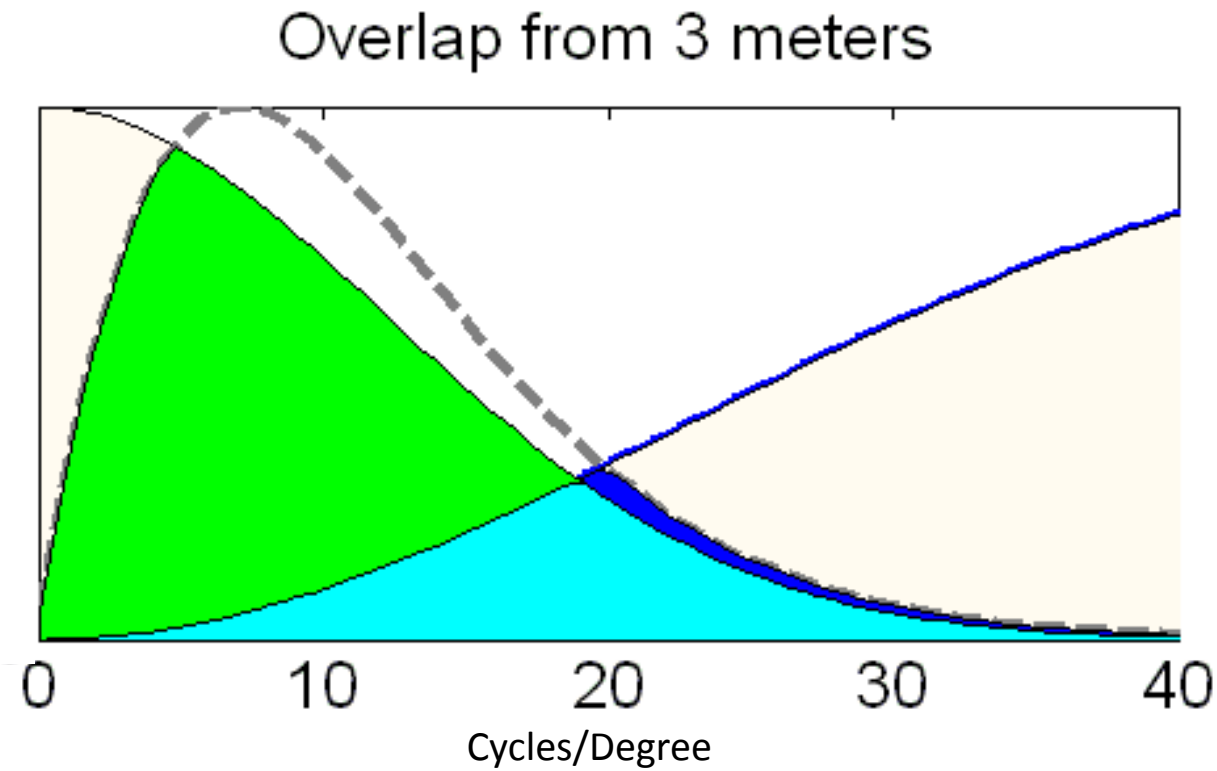


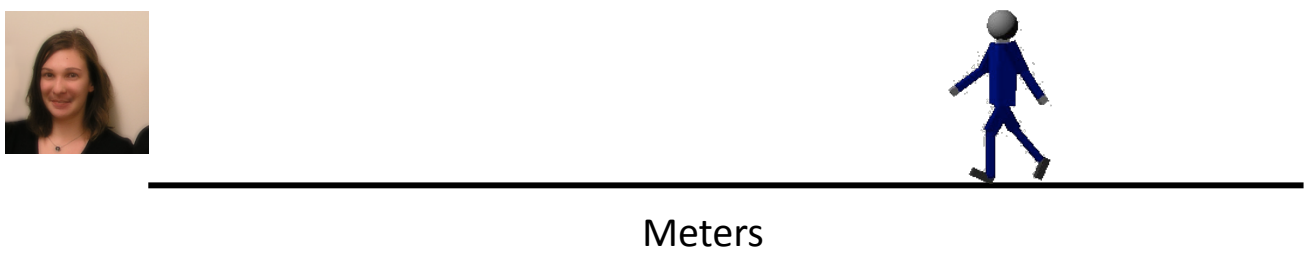
Which image do people see?



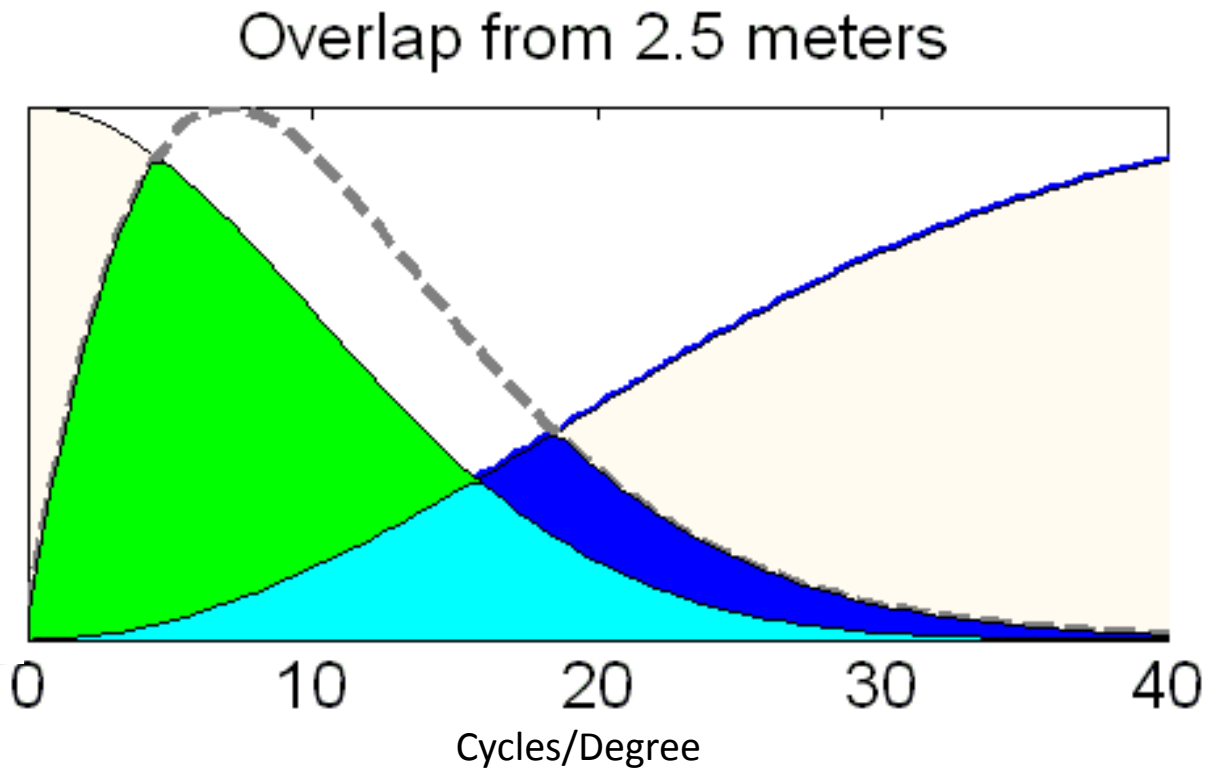


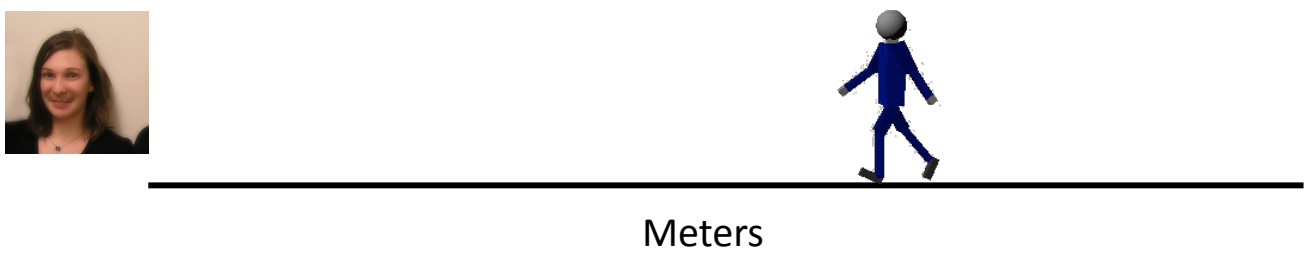
Which image do people see?



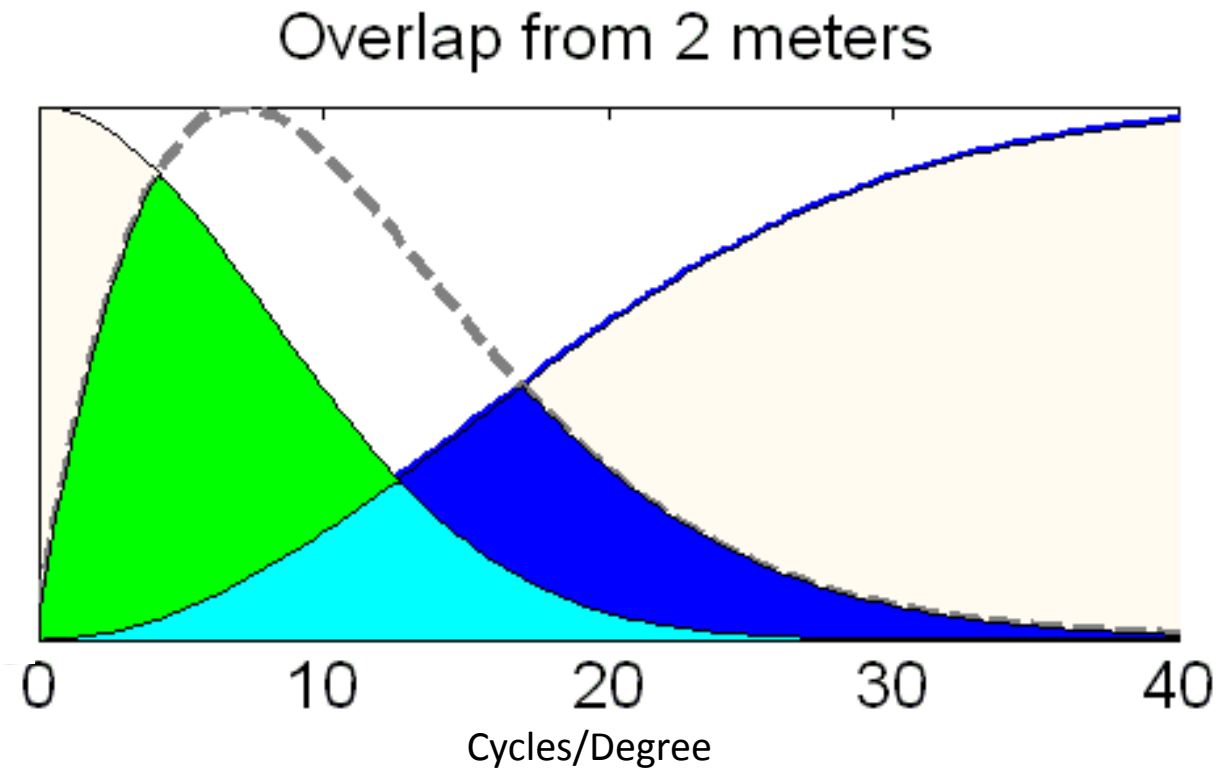


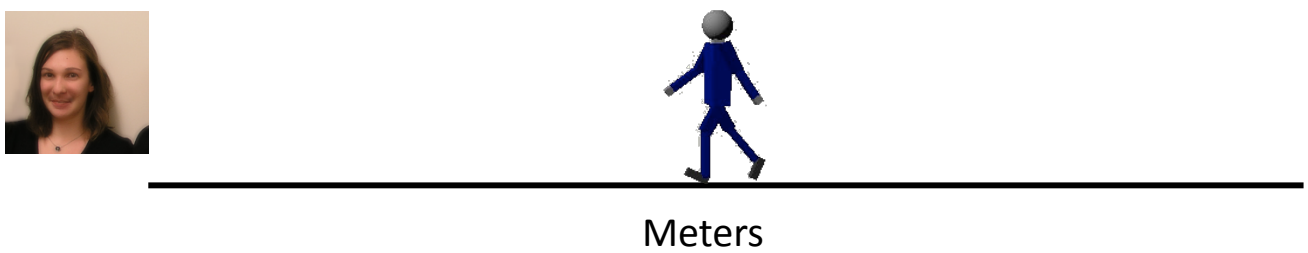
Which image do people see?



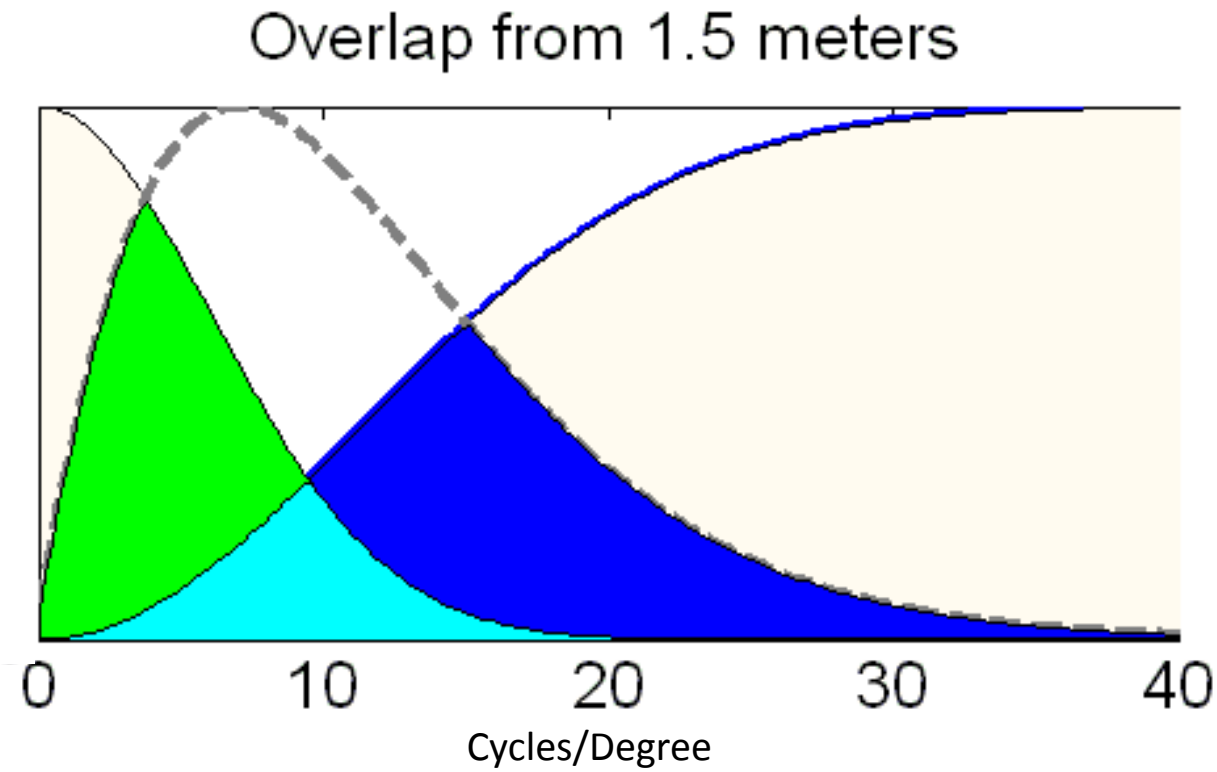


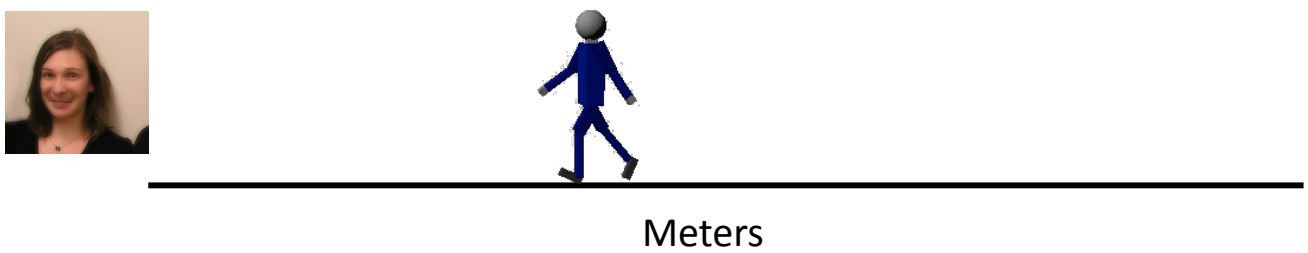
Which image do people see?



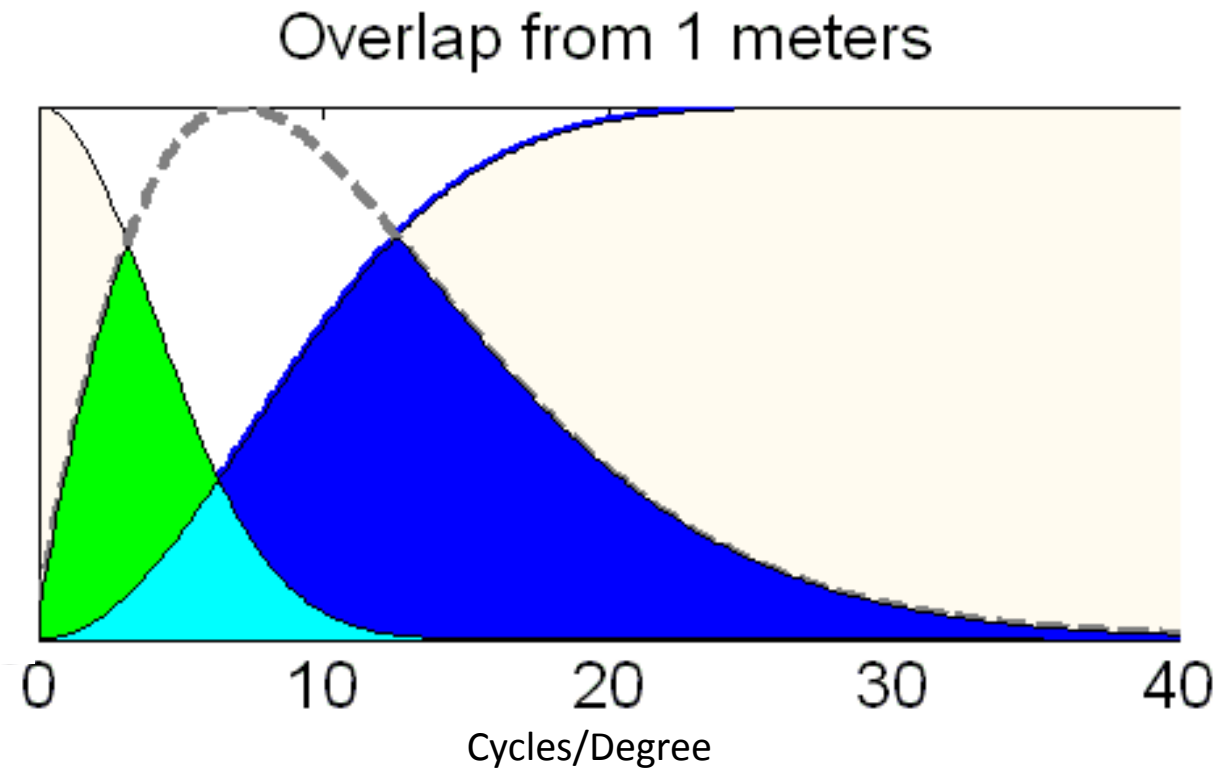


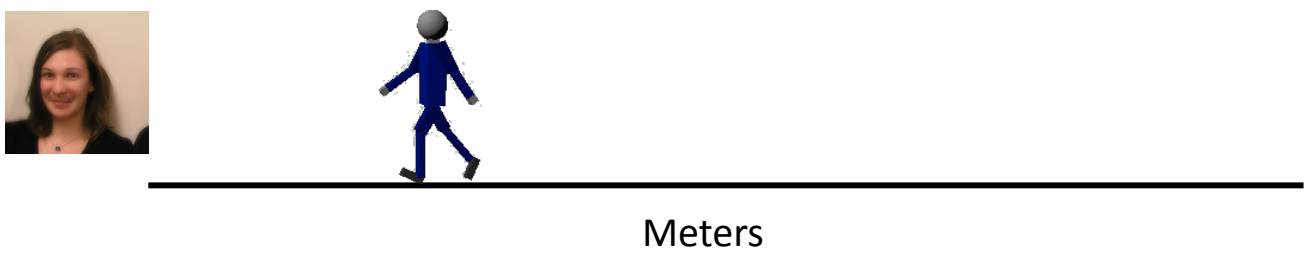
Which image do people see?



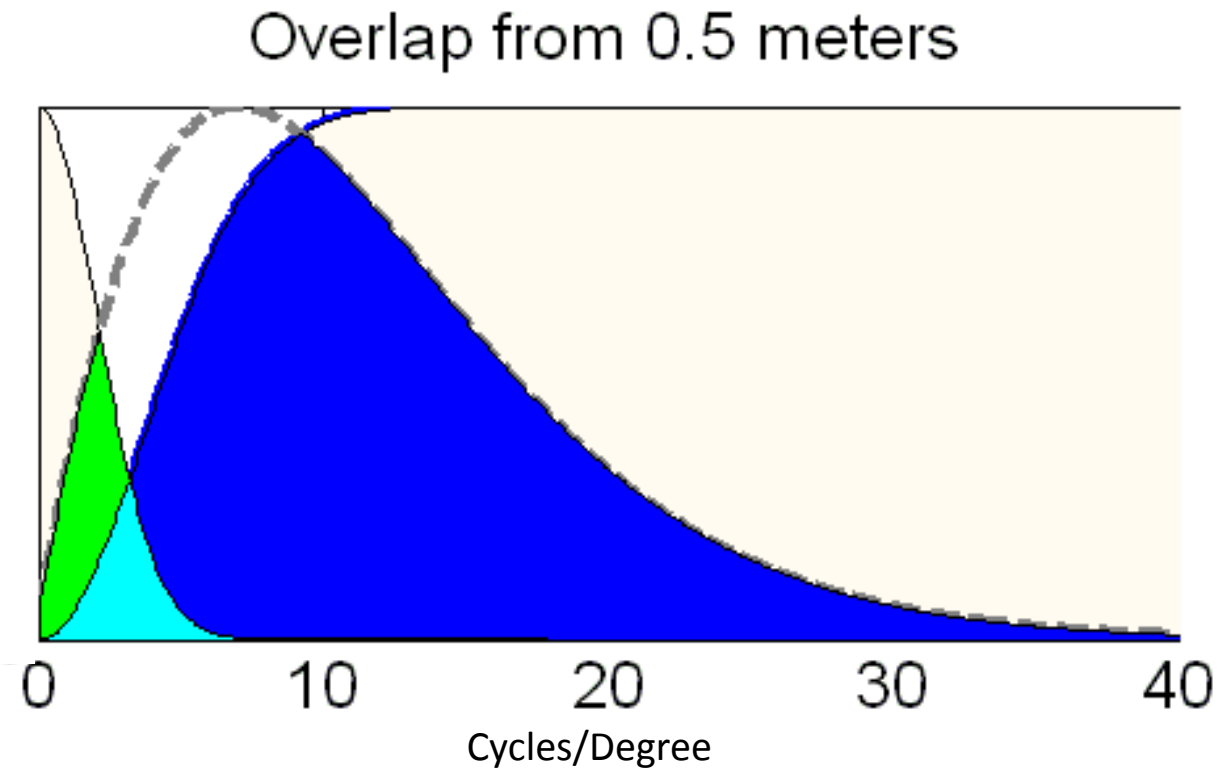


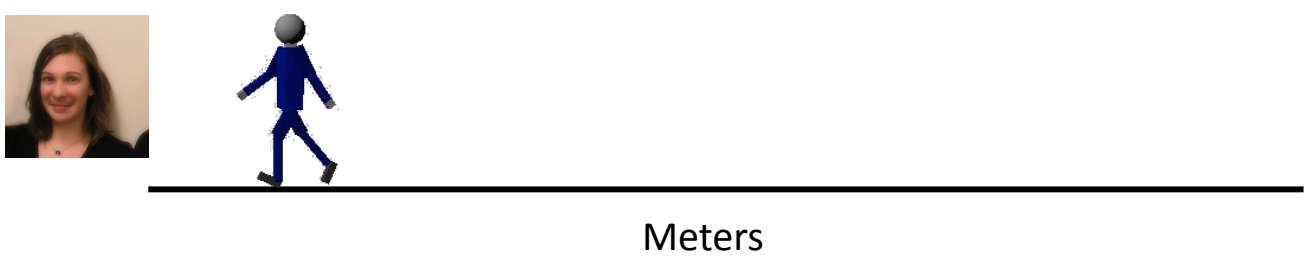
Which image do people see?



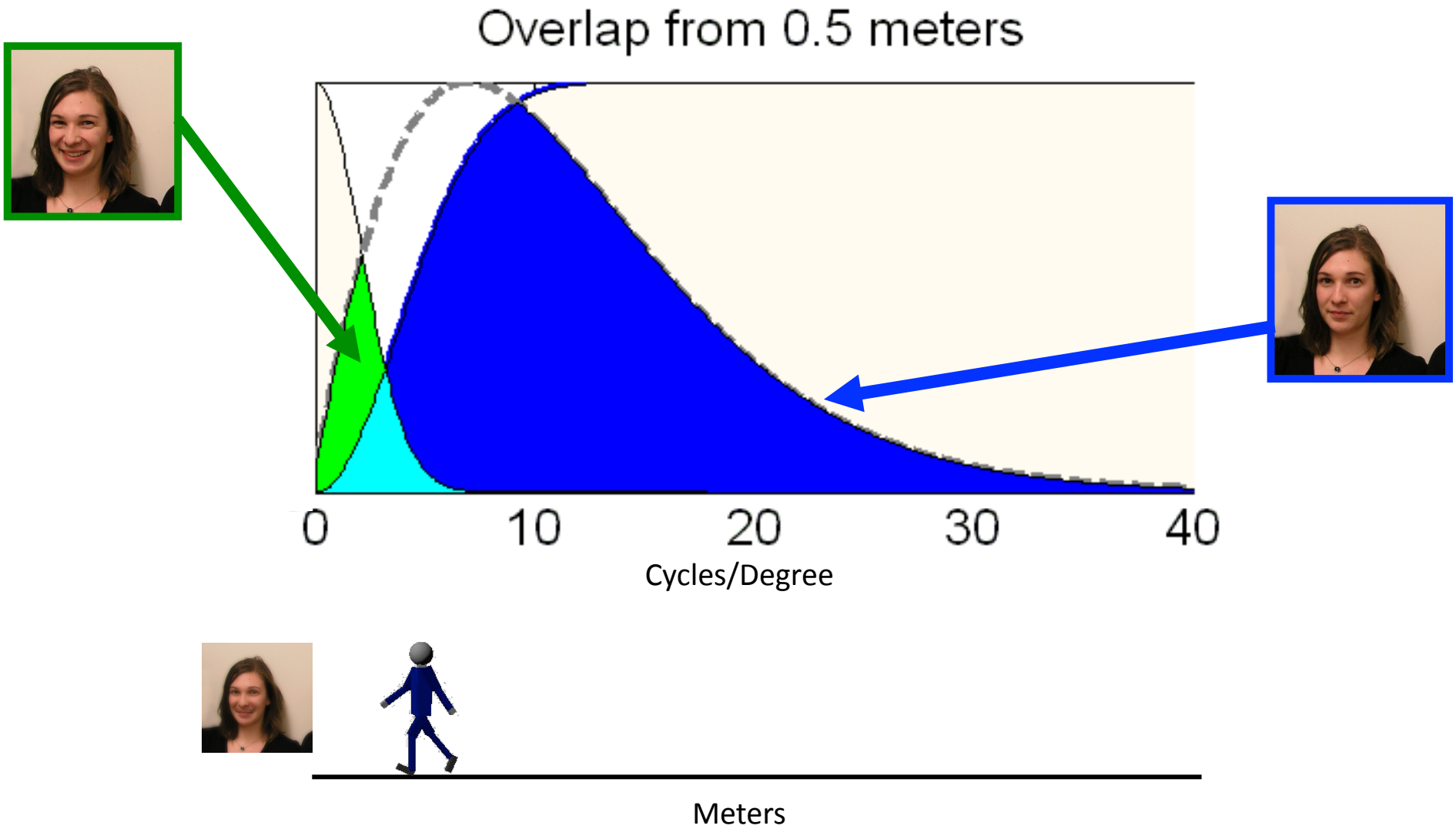


Which image do people see?





Which image do people see?

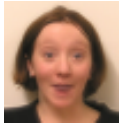


Experiment 1: Hybrid Faces

Changing expression with viewing distance

Neutral ⟷ Surprised

12 deg.



1 deg.

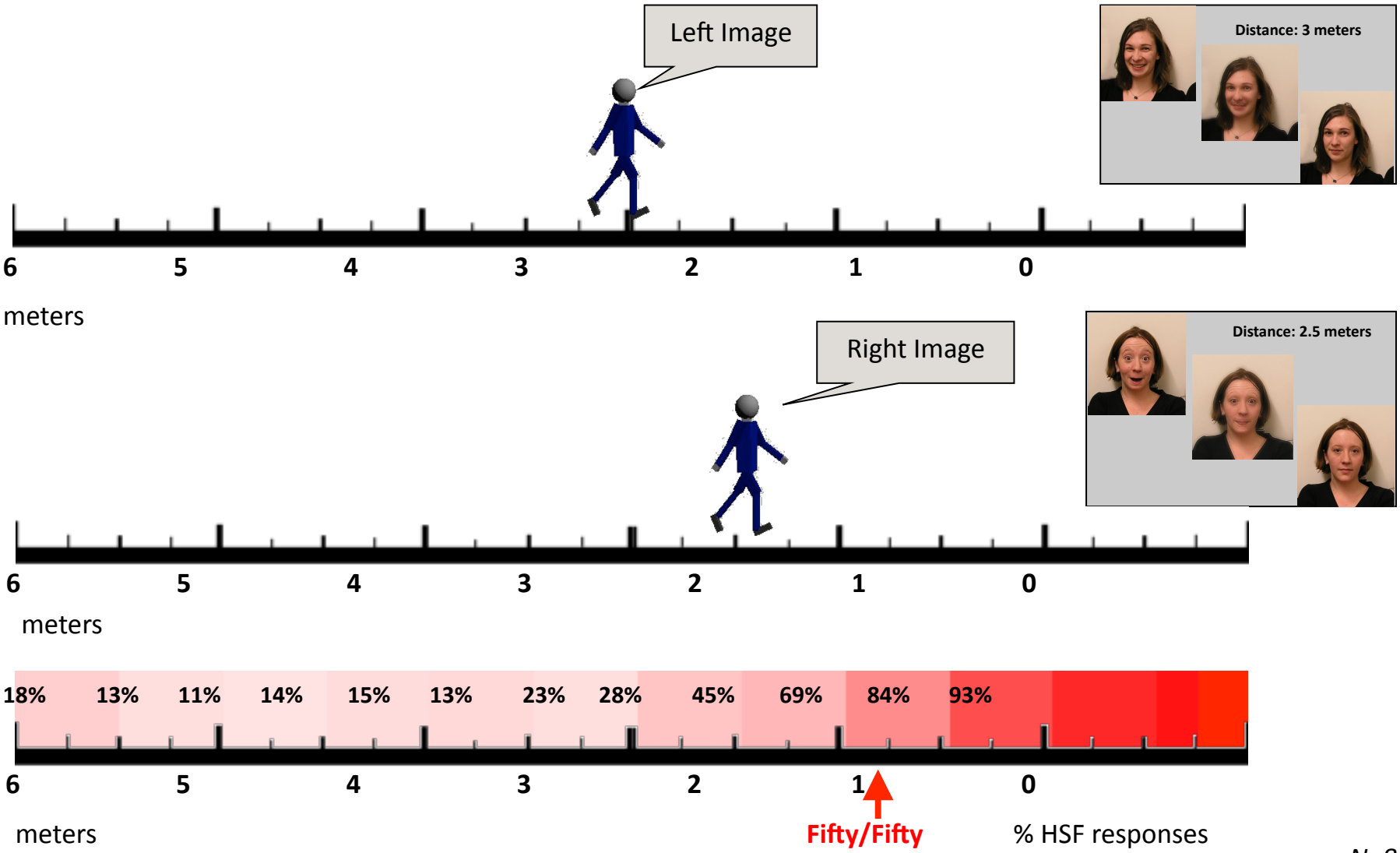
Experiment 1: Hybrid Faces

Matching Task

Hybrid parameters: 30 cycles/image in LSF, 55 c/i in HSF

Experiment 1: Hybrid Faces

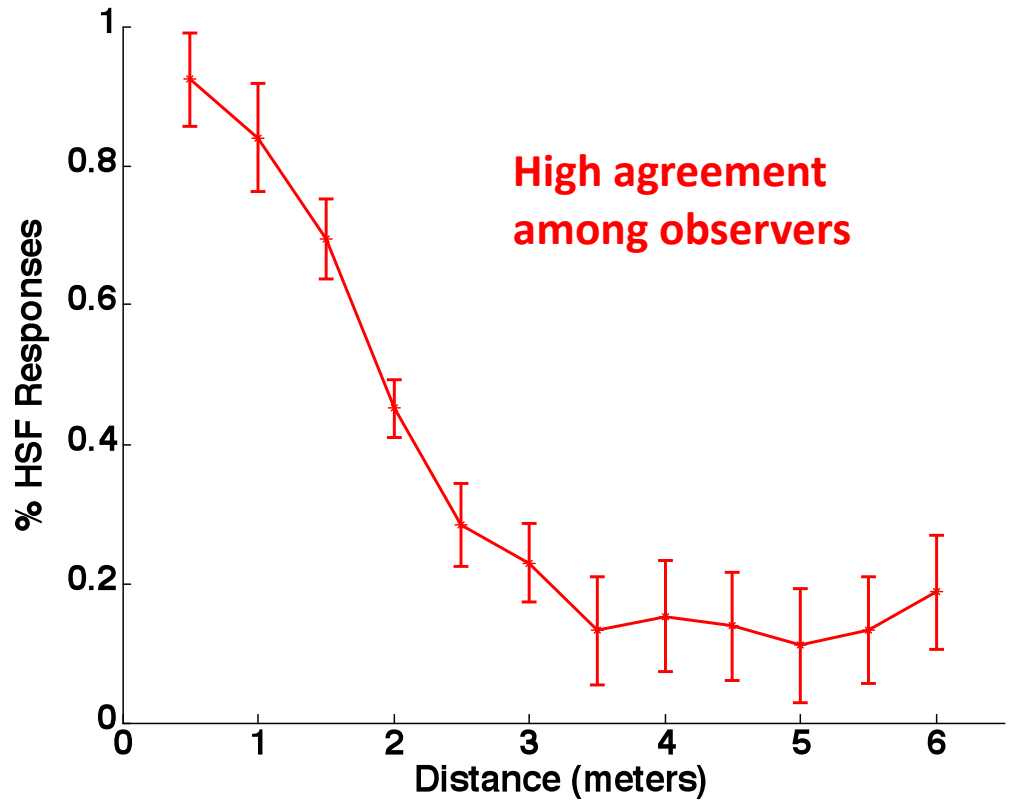
Static condition

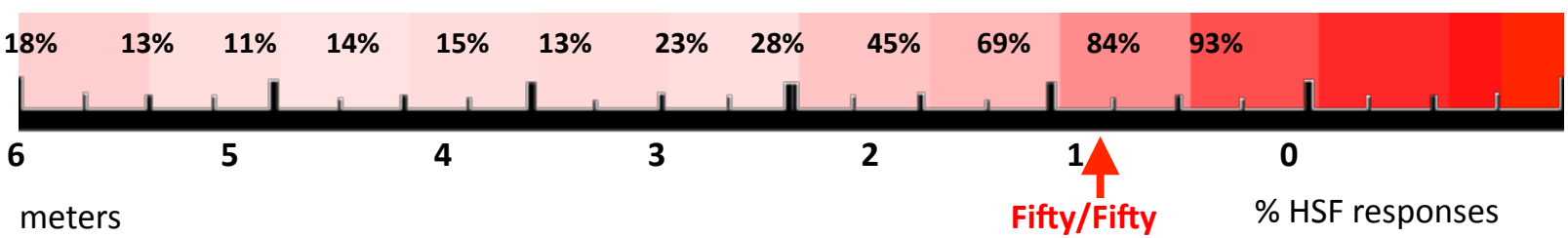


N=8

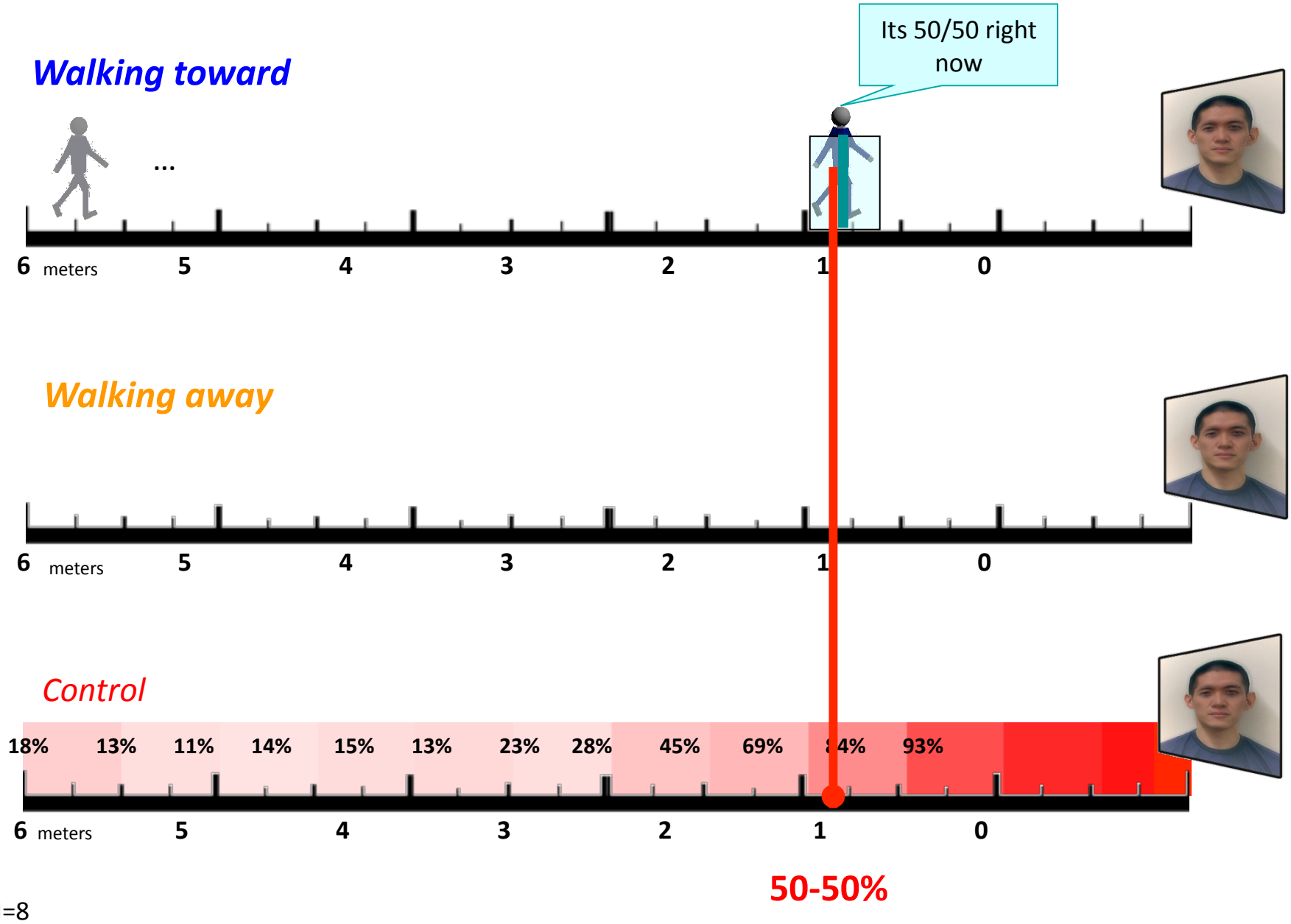
Experiment 1: Hybrid Faces

Static condition

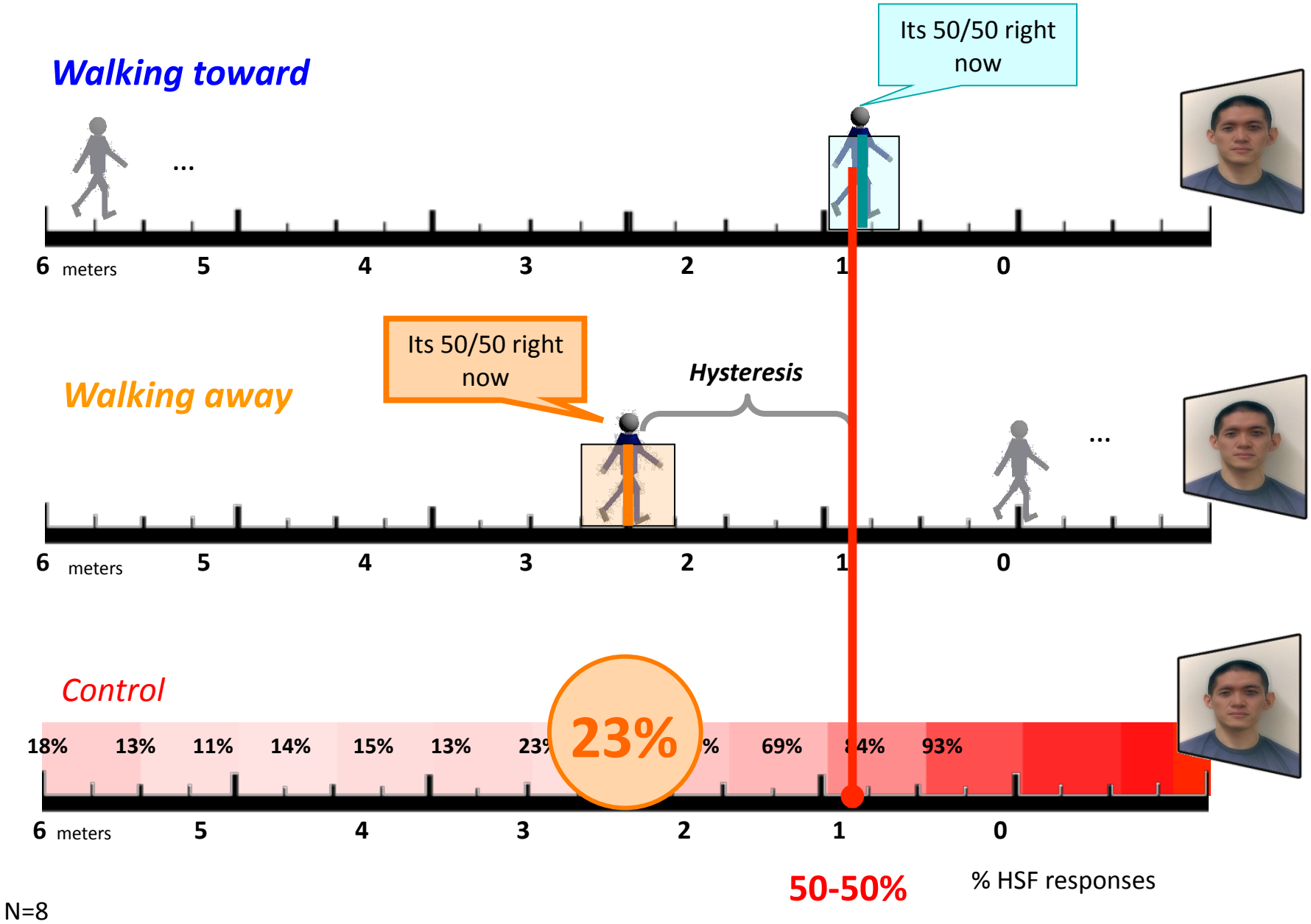




N=8


Its 50/50 right now
Walking toward
...
6 meters 5 4 3 2 1 0
Walking away
6 meters 5 4 3 2 1 0
Control
18% 13% 11% 14% 15% 13% 23% 28% 45% 69% 84% 93%
6 meters 5 4 3 2 1 0
50-50%

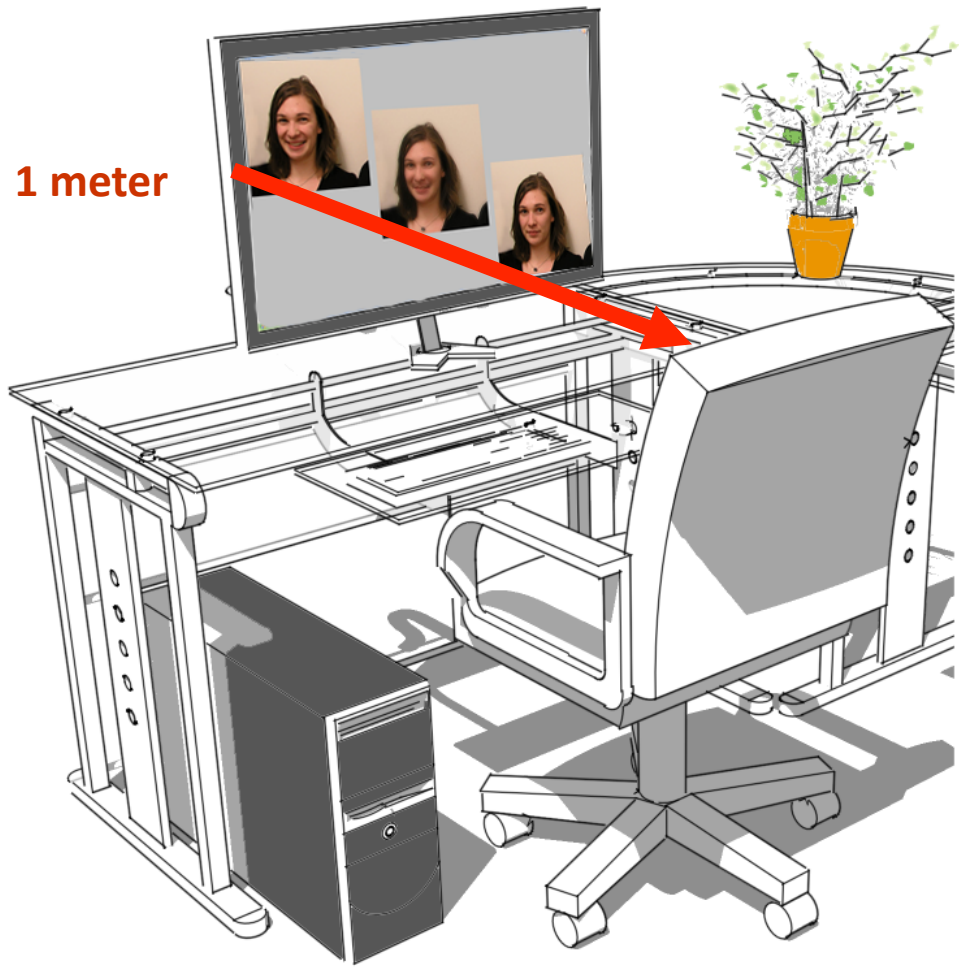
N=8


Walking toward
Its 50/50 right now
...
6 meters
5
4
3
2
1
0
Walking away
Its 50/50 right now
Hysteresis
...
6 meters
5
4
3
2
1
0
Control
18%
13%
11%
14%
15%
13%
23%
23%
%
69%
84%
93%
6 meters
5
4
3
2
1
0
50-50%
% HSF responses

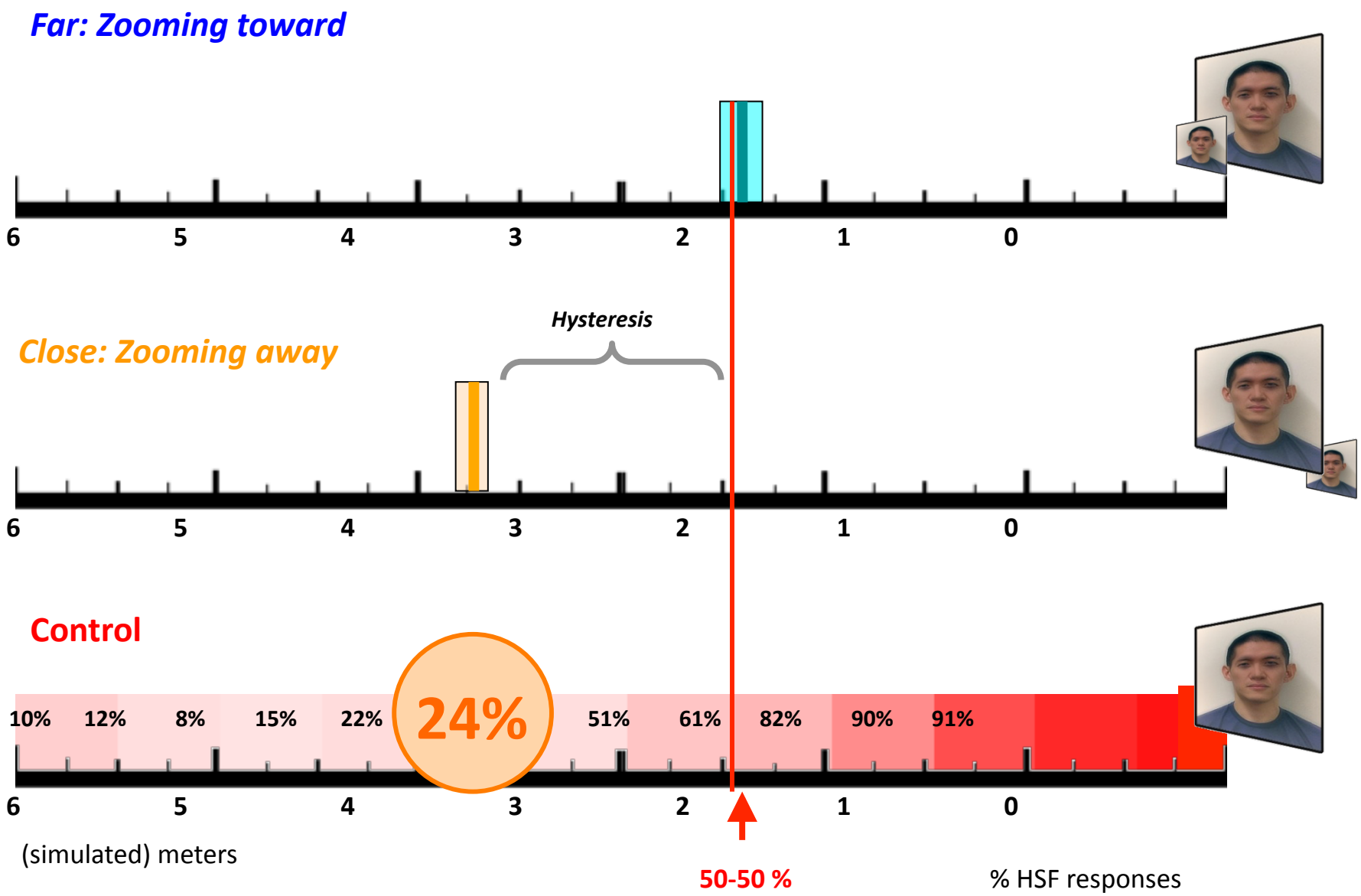
N=8

Experiment 2 : Hybrid Faces

Object ‘motion’



Hold space bar to start image moving


Far: Zooming toward
6
5
4
3
2
1
0
Hysteresis
Close: Zooming away
6
5
4
3
2
1
0
Control
10%
12%
8%
15%
22%
24%
51%
61%
82%
90%
91%
6
5
4
3
2
1
0
(simulated) meters
50-50 %
% HSF responses

N=12

Advances in multi-resolution perception

- We tend to constantly **reinterpret** the representation if we are **gaining** information

- We tend to stick with our first **grouping interpretation** if we are **losing** information. The hysteresis effect is modulated by the strength of perceptual grouping across scales.

The visual system is designed to integrate across multi resolution in a way that provides **the most accurate interpretation** of the world as we move through it.