6.819 / 6.869: Advances in Computer Vision

Basics of Image Processing II:

Multi-scale representations (pyramids, gabor filters, 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

Filtering, multi-scale, band pass filtering

From M. Lewicky
Model of Retinal and LGN cells

- **On center cell**
  - Light on center only
  - Ganglion cell fires rapidly
  - Ganglion cell does not fire

- **Off center cell**
  - Light on surround only
  - Cell does not fire
  - Cell fires rapidly

- **On center cell**
  - No light on center or surround
  - Cell does not fire

- **Off center cell**
  - Light on center and surround
  - Weak response (low frequency firing)
No light

Response (baseline)
Light

Photoreceptors

Bipolar cells

Ganglion cell

On center cell

Response

Increase
On center cell

Response
Light

Response  Decrease
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
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

Spatial frequency (cycles/image)

amplitude

Log Fourier power spectrum
Subtle expression
Leonardo da Vinci’s Mona Lisa

Smile

Smile

No smile

Low
Spatial frequency
High

Margaret Livingstone
Image Pyramids

Idea: Represent N x N image as a “pyramid” of 1 x 1, 2 x 2, 4 x 4, ..., 2^k x 2^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

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

Solution: filter the image, *then* subsample
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.
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, measures 257 by 257 pixels and each higher level array is roughly half the dimensions of its predecessor. Thus, level 5 measures just 9 by 9 pixels.

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

- 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.
What will the Laplacien pyramid show?
Laplacien pyramid

Residual
Laplacien pyramid: Use for Image blending
Image blending

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

Multi scale orientation
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.
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.
What is a Gabor filter?

a two-dimensional gaussian modulated by a sinusoid.

Human visual system contains Gabor-filter cells

\[ y = \sin(4x) \exp\left(\frac{-x^2}{2}\right) \]
Gabor filters: Orientation x spatial frequency filtering
The image through a Gaussian window

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

\[ e^{-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}} \]

to obtain a Gabor function or Gabor wavelet:

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

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

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

\[ h(x,y;x_0,y_0) = e^{-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}} \]
Even (cosinus)  
Odd (sinus)

Pairs of oriented filters can measure local oriented energy, identify contours
Gabor wavelet:

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

Tuning filter orientation:

\[ x' = \cos(\alpha)x + \sin(\alpha)y \]

\[ y' = -\sin(\alpha)x + \cos(\alpha)y \]
Gabor Filters: a bank of them
Gabor Filter Bank

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

or = [4 4 4 4];
4 spatial scales with 4 orientations each

or = [12 6 3 2];

Gabor/createBankGabor.m
Convolution in 2-d

\[
* \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} =
\]

Which filters will be most useful?
Energy in each sub-band representation

Sum of all values
16 Gabor filters, 4 x 4 spatial grid -> 256 features
Representation of GIST in spectral domain (visualization)
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.

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

% Parameters:
param.imageSize = 128;
param.orientationsPerScale = 8;
param.numberOfBlocks = 4;
param.fc_prefilt = 4;

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

% Visualization
figure
subplot(211)
imshow(img)
title('Input image')
subplot(212)
title('Descriptor')
showGist(gist, param)
```

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

```matlab
gist = LMgist(D, N IMAGES, 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:


preciate if you cite:
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 $4 \times 4$ to $128 \times 128$ pixels), and the second row shows the images upsampled to $256 \times 256$ 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

Filter Kernels

Image

Coarsest scale

Finest scale

Simoncelli, Freeman, Adelson
The steerable pyramid

Decomposition

subband

Simoncelli, Freeman, Adelson
The steerable pyramid
The steerable pyramid

Representation in pixel space

Alternative representation

Low-pass residual
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.

\[ G^1_\theta = \cos(\theta)G^1_0 + \sin(\theta)G^1_{90} \]

Filter Set:

- 0°
- 90°
- Synthesized 30°

Response:

Raw Image

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.

Freeman & Adelson 92
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.
Changes of Meaning with Scale

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

I see an angry guy

It’s a woman!

Up Close
Multiscale subband decomposition + Human Contrast Sensitivity Function

Contrast sensitivity

Spatial frequency (cycles/degree)

Low sensitivity
Contrast Sensitivity Function

Maximum sensitivity
~ 6 cycles / degree of visual angle
10 deg.

20 cm

1 meter

6 c/d * 10 deg/i
8               11                17                 25               38                 57                85                128  c/i

Spatial Frequency

6 c/d * 2 deg/i

6 meters

20 cm

2 deg.
Perception of hybrid images

A man or a woman?

Male dominance

Female dominance

A man or a woman?
It's a woman!

Spatial Frequency

1 meter

10 deg.

20 cm
I see an angry guy.

6 meters

2 deg.

20 cm
Perception of hybrid images

1. Frequency cut = 16 cycles/image
2. Frequency cut = 36 cycles/image

Hybrid/demo.m
Frequency Gap

Hybrid/demo.m

a) Gain vs. frequency (cycles/image)
b) Gain vs. frequency (cycles/image)
Alignment
Alignment
Alignment
Changing expression
Changing expression
Changing expression

Sad ← Surprised

Published in New Scientist, March 31, 2007
Subband decomposition of hybrid
Perceptual Grouping

Shadows
Perceptual Grouping

Reorganization of the shadows as a motorcycle
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

This text can only be read by somebody close enough to the screen.

This text can be read at a distance.

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

In scale space

Approaching: Gaining information

Receding: Loosing information

Far

Close

Low spatial frequency (blurry)

Full spatial frequency range (sharp)
Hybrid Images & Distance

1 meter

4 meters
Which image do people see?

Overlap from 3 meters

Cycles/Degree
Which image do people see?
Which image do people see?

Overlap from 3 meters

Cycles/Degree

Meters
Which image do people see?

Overlap from 3 meters

Cycles/Degree

Meters
Which image do people see?

Overlap from 2.5 meters

Cycles/Degree

Meters
Which image do people see?

Overlap from 2 meters

Cycles/Degree

Meters
Which image do people see?

Overlap from 1.5 meters

[Graph showing overlap from 1.5 meters]
Which image do people see?

Overlap from 1 meters

Cycles/Degree

Meters
Which image do people see?

Overlap from 0.5 meters

Cycles/Degree

Meters
Which image do people see?

Overlap from 0.5 meters

Cycles/Degree

Meters
Experiment 1: Hybrid Faces
Changing expression with viewing distance

Neutral 12 deg. → Surprised 1 deg.
Experiment 1: Hybrid Faces
Matching Task

Distance: 3 meters

Hybrid parameters: 30 cycles/image in LSF, 55 c/i in HSF
Experiment 1: Hybrid Faces

Static condition

Distance: 3 meters

Distance: 2.5 meters

N=8

% HSF responses
Experiment 1: Hybrid Faces
Static condition

High agreement among observers

% HSF Responses vs Distance (meters)

N=8
Walking toward

Walking away

Control

N=8

50-50%
Walking toward

Walking away

Control

Its 50/50 right now

Hysteresis

N=8

50-50%

% HSF responses

69%
84%
93%
Experiment 2: Hybrid Faces

Object ‘motion’

Hold space bar to start image moving

1 meter
Far: Zooming toward

Close: Zooming away

Control

(simulated) meters

N=12

Hysteresis

24%

50-50 %

% HSF responses
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.