VisuaTexture (in human and machine)

Somewhere in Cinque Terre, May 2005

CS194: Image Manipulation, Comp. Vision, and Comp. Photo
Alexei Efros, UC Berkeley, Spring 2020
What is Texture?

• Texture depicts spatially repeating patterns
• Many natural phenomena are textures

radishes  rocks  yogurt
Texture as “stuff”
Texture and Material

http://www-cvr.ai.uiuc.edu/ponce_grp/data/texture_database/samples/
Texture Synthesis

- Goal of Texture Synthesis: create new samples of a given texture
- Many applications: virtual environments, hole-filling, texturing surfaces
The Challenge

- Need to model the whole spectrum: from repeated to stochastic texture
Efros & Leung Algorithm

- Assuming Markov property, compute $P(p|N(p))$
  - Building explicit probability tables infeasible
  - Instead, we search the input image for all similar neighborhoods — that’s our pdf for $p$
  - To sample from this pdf, just pick one match at random
Some Details

• Growing is in “onion skin” order
  – Within each “layer”, pixels with most neighbors are synthesized first
  – If no close match can be found, the pixel is not synthesized until the end

• Using *Gaussian-weighted* SSD is very important
  – to make sure the new pixel agrees with its closest neighbors
  – Approximates reduction to a smaller neighborhood window if data is too sparse
Neighborhood Window

input
Varying Window Size

Increasing window size
Synthesis Results

classical canvas

raffia weave
More Results

white bread

brick wall
Homage to Shannon
Hole Filling
Extrapolation
Summary

• The Efros & Leung algorithm
  – Very simple
  – Surprisingly good results
  – Synthesis is easier than analysis!
  – …but very slow
**Image Quilting** [Efros & Freeman]

- **Observation:** neighbor pixels are highly correlated

**Idea:** unit of synthesis = block

- Exactly the same but now we want $P(B | N(B))$
- Much faster: synthesize all pixels in a block at once
- Not the same as multi-scale!
Random placement of blocks

Neighboring blocks constrained by overlap

Minimal error boundary cut
Minimal error boundary

overlapping blocks

vertical boundary

\[ \text{overlap error} = 2 \]

\[ \text{min. error boundary} \]
Our Philosophy

• The “Corrupt Professor’s Algorithm”:  
  – Plagiarize as much of the source image as you can  
  – Then try to cover up the evidence

• Rationale:  
  – Texture blocks are by definition correct samples of texture so problem only connecting them together
Failures
(Chernobyl Harvest)
input image

Wei & Levoy

Portilla & Simoncelli

Xu, Guo & Shum

Our algorithm
Application: Texture Transfer

- Try to explain one object with bits and pieces of another object:
Texture Transfer

Constraint

Texture sample
Texture Transfer

• Take the texture from one image and “paint” it onto another object

Same as texture synthesis, except an additional constraint:
  1. Consistency of texture
  2. Similarity to the image being “explained”
Image Analogies

Aaron Hertzmann\textsuperscript{1,2}
Chuck Jacobs\textsuperscript{2}
Nuria Oliver\textsuperscript{2}
Brian Curless\textsuperscript{3}
David Salesin\textsuperscript{2,3}

\textsuperscript{1}New York University
\textsuperscript{2}Microsoft Research
\textsuperscript{3}University of Washington
Image Analogies

A

A’

B

B’
Blur Filter

Unfiltered source \((A)\)

Filtered source \((A')\)

Unfiltered target \((B)\)

Filtered target \((B')\)
Edge Filter

Unfiltered source (A)  Filtered source (A')
Unfiltered target (B)  Filtered target (B')
Artistic Filters
Image Analogies

Goal: Process an image by example

A : A' :: B

Hertzmann et al. SIGGRAPH 2001
Colorization

Unfiltered source \( (A) \)

Filtered source \( (A') \)

Unfiltered target \( (B) \)

Filtered target \( (B') \)
Texture-by-numbers
Super-resolution

A

A'
Super-resolution (result!)
Texture Analysis

Compare textures and decide if they’re made of the same “stuff”.

True (infinite) texture

ANALYSIS

“Same” or “different”
When are two textures similar?
Béla Julesz, father of texture
Human vision is sensitive to the difference of some types of elements and appears to be “numb” on other types of differences.
Search Experiment I

The subject is told to detect a target element in a number of background elements. In this example, the detection time is independent of the number of background elements.
In this example, the detection time is proportional to the number of background elements, and thus suggests that the subject is doing element-by-element scrutiny.
Julesz then conjectured the following axiom:

Human vision operates in two distinct modes:

1. Preattentive vision
   parallel, instantaneous (~100--200ms), without scrutiny,
   independent of the number of patterns, covering a large visual field.

2. Attentive vision
   serial search by focal attention in 50ms steps limited to small aperture.
Examples

Pre-attentive vision is sensitive to size/width, orientation changes
Examples

Sensitive to number of terminators

Left: fore-back
Right: back-fore

See previous examples
For cross and terminators
Heuristic (Axiom) II

Textons are the fundamental elements in preattentive vision, including

1. Elongated blobs
   rectangles, ellipses, line segments with attributes
   color, orientation, width, length, flicker rate.
2. Terminators
   ends of line segments.
3. Crossings of line segments.

But it is worth noting that Julesz’s conclusions are largely based on ensemble of artificial texture patterns. It was infeasible to synthesize natural textures for controlled experiments at that time.
Julesz Conjecture

Textures cannot be spontaneously discriminated if they have the same first-order and second-order statistics and differ only in their third-order or higher-order statistics.

(later proved wrong)
1st Order Statistics

5% white

20% white
2nd Order Statistics

10% white
Single Cell Recording

Microelectrode

Electrical response (action potentials)

mV

Time

© Stephen E. Palmer, 2002
Retinal Receptive Fields

Receptive field structure in-ganglion-cells:
On-center Off-surround

Stimulus condition  Electrical response

Response

Time

© Stephen E. Palmer, 2002
Receptive field structure in **ganglion cells**: On-center Off-surround

Stimulus condition

Electrical response
Receptive field structure in ganglion cells: On-center Off-surround
Retinal Receptive Fields

Receptive field structure in **ganglion cells**: On-center Off-surround

Stimulus condition  Electrical response

© Stephen E. Palmer, 2002
Retinal Receptive Fields

Receptive field structure in ganglion cells: On-center Off-surround

Stimulus condition

Electrical response

Response

Time

© Stephen E. Palmer, 2002
Receptive field structure in ganglion cells: On-center Off-surround
RF of On-center Off-surround cells

Neural Response

- Center
- Surround
- On
- Off

Receptive Field

Response Profile

Firing Rate

Horizontal Position

on-center

off-surround

© Stephen E. Palmer, 2002
Retinal Receptive Fields

RF of Off-center On-surround cells

Neural Response

Surround

Center

On      Off

Receptive Field

Response Profile

Firing Rate

on-surround

off-center

Horizontal Position

© Stephen E. Palmer, 2002
Retinal Receptive Fields

- Horizontal cell
- Bipolar cells
- Retinal ganglion cells
- Amacrine cell
- Dendro-dendritic synapse
- Receptors

Surround
- Center
- Surround
Retinal Receptive Fields

Receptive field structure in bipolar cells

- Action potential
- Graded potential
- Stimulus
- Ganglion cell
- Bipolar cell
- Photoreceptor

© Stephen E. Palmer, 2002
Retinal Receptive Fields

Receptive field structure in bipolar cells

A. WIRING DIAGRAM

[Diagram showing light from receptors to horizontal cells and then to bipolar cells via direct and indirect paths]

B. RECEPTIVE FIELD PROFILES

[Graph showing direct excitatory component (D) and indirect inhibitory component (I) resulting in D + I]

© Stephen E. Palmer, 2002
The receptive field of a retinal ganglion cell can be modeled as a “Difference of Gaussians”

\[ G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \]
The receptive field of a receptor is simply the area of the visual field from which light strikes that receptor. For any other cell in the visual system, the receptive field is determined by which receptors connect to the cell in question.
Anatomy of Pathway to Visual Cortex

- Left visual field
- Right visual field
- Nasal
- Temporal
- Optic chiasm
- Pulvinar nucleus
- Lateral geniculate nucleus
- Superior colliculus
- Optic radiation
- Primary visual cortex

© Stephen E. Palmer, 2002
Visual Cortex

Cortical Area V1

aka:
Primary visual cortex
Striate cortex
Brodman’s area 17

Eye Optic nerve Temporal visual cortex Thalamus LGN Parietal visual cortex Dorsal Stream Striate cortex (V1) Extrastriate cortex Ventral Stream
Cortical Receptive Fields

Single-cell recording from visual cortex

David Hubel & Thorston Wiesel
Cortical Receptive Fields

Single-cell recording from visual cortex

https://www.youtube.com/watch?v=IOHayh06LJ4
Cortical Receptive Fields

Three classes of cells in V1

Simple cells

Complex cells

Hypercomplex cells
Orientation Selectivity in V1

Stimulus: on off
Cortical Receptive Fields

Simple Cells: “Line Detectors”

A. Light Line Detector

B. Dark Line Detector

Firing Rate vs. Horizontal Position

© Stephen E. Palmer, 2002
Cortical Receptive Fields

Simple Cells: “Edge Detectors”

C. Dark-to-light Edge Detector
D. Light-to-dark Edge Detector

© Stephen E. Palmer, 2002
Cortical Receptive Fields

Constructing a line detector

Retina

LGN

Cortical Area V1

Receptive Fields

Center-Surround Cells

Simple Cell

© Stephen E. Palmer, 2002
The 1D Gaussian and its derivatives

\[ G_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \]

\[ G'_\sigma(x) = \frac{d}{dx} G_\sigma(x) = -\frac{1}{\sigma} \left( \frac{x}{\sigma} \right) G_\sigma(x) \]

\[ G''_\sigma(x) = \frac{d^2}{dx^2} G_\sigma(x) = \frac{1}{\sigma^2} \left( \frac{x^2}{\sigma^2} - 1 \right) G_\sigma(x) \]

\( G'_\sigma(x) \)'s maxima/minima occur at \( G''_\sigma(x) \)'s zeros. And, we can see that \( G'_\sigma(x) \) is an odd symmetric function and \( G''_\sigma(x) \) is an even symmetric function.
Oriented Gaussian Derivatives in 2D

\[ f_1(x, y) = G''_{\sigma_1}(x)G_{\sigma_2}(y) \]  
\[ f_2(x, y) = G'''_{\sigma_1}(x)G_{\sigma_2}(y) \]  

(10.4)  
(10.5)

We also consider rotated versions of these Gaussian derivative functions.

\[ \text{Rot}_\theta f_1 = G''_{\sigma_1}(u)G_{\sigma_2}(v) \]  
\[ \text{Rot}_\theta f_2 = G'''_{\sigma_1}(u)G_{\sigma_2}(v) \]  

(10.6)  
(10.7)

where we set

\[
\begin{pmatrix}
  u \\
  v
\end{pmatrix} = \begin{pmatrix}
  \cos \theta & -\sin \theta \\
  \sin \theta & \cos \theta
\end{pmatrix} \begin{pmatrix}
  x \\
  y
\end{pmatrix}
\]

These are useful when we convolve with 2D images, e.g. to detect edges at different orientations.
Oriented Gaussian First and Second Derivatives
Receptive fields of complex cells

Stimulus: on  off
Cortical Receptive Fields

Constructing a Complex Cell

Retina → Cortical Area V1

Receptive Fields → Simple Cells → Complex Cell
Modeling hypercolumns

- Elongated directional Gaussian derivatives
- Gabor filters could be used instead
- Multiple orientations, scales
Mapping from Retina to V1
Overcomplete representation: filter banks

LM Filter Bank

Code for filter banks: www.robots.ox.ac.uk/~vgg/research/texclass/filters.html
Showing magnitude of responses

Kristen Grauman
What does it capture?

\[ v = F \times \text{Patch} \quad \text{(where F is filter matrix)} \]

**Fig. 3.** Image reconstruction. Two example image patches (*left*), were reconstructed (*right*) from spatial filter responses at their center. Original image patches masked by a Gaussian (*middle*) are shown for comparison.
How can we represent texture?

• Idea 1: Record simple statistics (e.g., mean, std.) of absolute filter responses
Can you match the texture to the response?

Filters

A

B

C

Mean abs responses
How can we represent texture?

• Can be thought of as an single “orientation histogram”

• Idea 2: Marginal histograms of filter responses
  – one histogram per filter
Multi-scale filter decomposition

Filter bank

Input image
Filter response histograms
Start with a noise image as output

Main loop:

- Match pixel histogram of output image to input
- Decompose input and output images using multi-scale filter bank (Steerable Pyramid)
- Match subband histograms of input and output pyramids
- Reconstruct input and output images (collapse the pyramids)
Figure 7: (Left pair) Inhomogeneous input texture produces blotchy synthetic texture. (Right pair) Homogenous input.

Figure 8: Examples of failures: wood grain and red coral.

Figure 9: More failures: hay and marble.
Marginal statistics are not enough

Neighboring filter responses are highly correlated
  • an edge at low-res will cause an edge at high-res

Let’s match 2nd order statistics too!

Match joint histograms of pairs of filter responses at adjacent spatial locations, orientations, and scales.

Optimize using repeated projections onto statistical constraint surfaces.
How can we represent texture?

• Marginal filter response histograms don’t talk to each other (in a direct way)

• Idea 3: Histograms of joint responses (textons)
We can form a feature vector from the list of responses at each pixel.
Textons (Malik et al, IJCV 2001)

- Cluster vectors of filter responses
Textons (cont.)
Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based essentially on the messages that reach us via our eyes. For a long time it was thought that the retinal image was transmitted point by point to visual centers in the brain; the cerebral cortex was a movie screen, so to speak, upon which the image in the eye was projected. Through the discoveries of Hubel and Wiesel we now know that behind the origin of the visual perception in the brain there is a considerably more complicated course of events. By following the visual impulses along their path to the various cell layers of the optical cortex, Hubel and Wiesel have been able to demonstrate that the message about the image falling on the retina undergoes a step-wise analysis in a system of nerve cells stored in columns. In this system each cell has its specific function and is responsible for a specific detail in the pattern of the retinal image.

China is forecasting a trade surplus of $90bn (£51bn) to $100bn this year, a threefold increase on 2004’s $32bn. The Commerce Ministry said the surplus would be created by a predicted 30% jump in exports to $750bn, compared with a 18% rise in imports to $660bn. This is likely to annoy the US, which has long argued that China’s exports are unfairly helped by a deliberately undervalued yuan. Beijing agrees the surplus is too high, but says the yuan is only one factor. Bank of China governor Zhou Xiaochuan said the country also needed to do more to boost domestic demand so more goods stayed within the country. China has already increased the value of the yuan against the dollar by 2.1% in July and permitted it to trade within a narrow band, but the US wants the yuan to be allowed to trade freely. However, Beijing has made it clear that it will take its time and tread carefully before allowing the yuan to rise further in value.
Patch Features
dictionary formation
Clustering (usually k-means)

Vector quantization

Slide credit: Josef Sivic
Image representation

frequency

codewords

.....
learning

feature detection & representation

image representation

codewords dictionary

category models (and/or) classifiers

category decision

recognition
Object Detection can be very fast

- On a task of judging animal vs no animal, humans can make mostly correct saccades in 150 ms (Kirchner & Thorpe, 2006)
  - Comparable to synaptic delay in the retina, LGN, V1, V2, V4, IT pathway.
  - Doesn’t rule out feedback but shows feed forward only is very powerful

- Detection and categorization are practically simultaneous (Grill-Spector & Kanwisher, 2005)
Scene Classification (Renninger & Malik)

beach

mountain

forest

city

street

farm

kitchen

livingroom

bedroom

bathroom
Texton Histogram Matching

(a) Texton pattern

(b) New image

(c) Label = bedroom

Universal textons

\( \chi^2 = 5.67 \)

Universal textons

\( \chi^2 = 4.17 \times 10^3 \)

Label = beach
Discrimination of Basic Categories

% correct

- texture model
Discrimination of Basic Categories

% correct

street  bedroom  mountain  farm  bathroom  city  beach  kitchen  forest  livingroom

* texture model

chance
Discrimination of Basic Categories

* texture model

37 ms

% correct

street
bedroom
mountain
farm
bathroom
city
beach
kitchen
forest
livingroom

chance
Discrimination of Basic Categories

% correct

* texture model

50 ms

street  bedroom  mountain  farm  bathroom  city  beach  kitchen  forest  livingroom

chance
Discrimination of Basic Categories

% correct

* texture model

69 ms

street  bedroom  mountain  farm  bathroom  city  beach  kitchen  forest  livingroom

chance
Discrimination of Basic Categories

% correct

* texture model  ▲ 37 ms  ● 50 ms  □ 69 ms

street  bedroom  mountain  farm  bathroom  city  beach  kitchen  forest  livingroom

chance
Scene Recognition using Texture
Why these filters?

Wavelet-like receptive fields emerge from a network that learns sparse codes for natural images.

Bruno A. Olshausen and David J. Field

\[ E = -[\text{preserve information}] - \lambda [\text{sparseness of } a_i], \tag{2} \]

where $\lambda$ is a positive constant that determines the importance of the second term relative to the first. The first term measures how well the code describes the image, and we choose this to be the mean square of the error between the actual image and the reconstructed image:

\[ [\text{preserve information}] = -\sum_{x,y} \left[ I(x, y) - \sum_i a_i \phi_i(x, y) \right]^2. \tag{3} \]
Learned filters

\[ a. \]