Video & Texture Synthesis

© A.A. Efros

CS194: Intro to Computer Vision & Comp. Photography
Alexei Efros, UC Berkeley, Fall 2021
Michel Gondry train video

http://www.youtube.com/watch?v=0S43lwBF0uM
Let’s predict weather:
- Given today’s weather only, we want to know tomorrow’s
- Suppose weather can only be {Sunny, Cloudy, Raining}

The “Weather Channel” algorithm:
- Over a long period of time, record:
  - How often S followed by R
  - How often S followed by S
  - Etc.
- Compute percentages for each state:
  - $P(R|S)$, $P(S|S)$, etc.
- Predict the state with highest probability!
- It’s a Markov Chain
Markov Chain

What if we know today and yesterday’s weather?

\[
\begin{pmatrix}
0.3 & 0.6 & 0.1 \\
0.4 & 0.3 & 0.3 \\
0.2 & 0.4 & 0.4
\end{pmatrix}
\]
[Shannon, ’48] proposed a way to generate English-looking text using N-grams:

- Assume a generalized Markov model
- Use a large text to compute prob. distributions of each letter given N-1 previous letters
- Starting from a seed repeatedly sample this Markov chain to generate new letters
- Also works for whole words

WE NEED TO EAT CAKE
Results (using \texttt{alt.singles} corpus):

- “As I've commented before, really relating to someone involves standing next to impossible.”
- “One morning I shot an elephant in my arms and kissed him.”
- “I spent an interesting evening recently with a grain of salt”
Still photos
Video clips
Video textures

![American flag]

![Candle flame]

![Swinging girl]

![ELD tree]
Problem statement

video clip  →  video texture
Our approach

- How do we find good transitions?
Finding good transitions

- Compute $L_2$ distance $D_{i,j}$ between all frames

Similar frames make good transitions
Markov chain representation

Similar frames make good transitions
Transition costs

- Transition from $i$ to $j$ if successor of $i$ is similar to $j$
  - Cost function: $C_{i\rightarrow j} = D_{i+1, j}$
Transition probabilities

• Probability for transition $P_{i \rightarrow j}$ inversely related to cost:

  $$P_{i \rightarrow j} \sim \exp \left( - \frac{C_{i \rightarrow j}}{\sigma^2} \right)$$

high $\sigma$  low $\sigma$
Preserving dynamics
Preserving dynamics
Preserving dynamics

- Cost for transition $i \rightarrow j$

$$C_{i \rightarrow j} = \sum_{k = -N}^{N-1} w_k D_{i+k+1, j+k}$$
Preserving dynamics – effect

- Cost for transition $i \rightarrow j$

\[ C_{i \rightarrow j} = \sum_{k = -N}^{N-1} w_k D_{i+k+1, j+k} \]
Dead ends

- No good transition at the end of sequence
Future cost

- Propagate future transition costs backward
- Iteratively compute new cost

\[ F_{i \rightarrow j} = C_{i \rightarrow j} + \alpha \min_k F_{j \rightarrow k} \]
Future cost

- Propagate future transition costs backward
- Iteratively compute new cost

\[ F_{i \rightarrow j} = C_{i \rightarrow j} + \alpha \min_k F_{j \rightarrow k} \]
Future cost

- Propagate future transition costs backward
- Iteratively compute new cost

\[ F_{i \rightarrow j} = C_{i \rightarrow j} + \alpha \min_k F_{j \rightarrow k} \]
Future cost

- Propagate future transition costs backward
- Iteratively compute new cost

\[ F_{i \rightarrow j} = C_{i \rightarrow j} + \alpha \min_k F_{j \rightarrow k} \]
Future cost

- Propagate future transition costs backward
- Iteratively compute new cost
  \[ F_{i \rightarrow j} = C_{i \rightarrow j} + \alpha \min_k F_{j \rightarrow k} \]
- Q-learning
Final result
Finding good loops

• Alternative to random transitions
• Precompute set of loops up front
Video portrait

• c.f. Harry Potter
Region-based analysis

- Divide video up into regions
- Generate a video texture for each region
User-controlled video textures

User controls video textures

slow  variable  fast

User selects target frame range
Video-based animation

• Like sprites computer games
• Extract sprites from real video
• Interactively control desired motion

©1985 Nintendo of America Inc.
Video sprite extraction

blue screen matting and velocity estimation
Video sprite control

- Augmented transition cost:

\[ C_{i\rightarrow j} = \alpha C_{i\rightarrow j} + \beta \text{ angle} \]

- Similarity term

- Control term

- Vector to mouse pointer

- Velocity vector
Video sprite control

- Need future cost computation
- Precompute future costs for a few angles.
- Switch between precomputed angles according to user input
- [GIT-GVU-00-11]
Interactive fish
Summary / Discussion

• Some things are relatively easy
Discussion

• Some are hard
“Amateur” by Lasse Gjertsen

http://www.youtube.com/watch?v=JzqumbhfxRo

similar idea:
http://www.youtube.com/watch?v=MsBMG-p1HDM&feature=share&list=PLFFFD733D0FF425290
Hyperlapse Videos

https://www.youtube.com/watch?v=Wt_Y04xn84M
“Do As I Do” (ICCV 2003)

https://youtu.be/UMJcpLIAwKg
Texture

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

radishes  rocks  yogurt
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
Varying Window Size

Increasing window size
Synthesis Results

french canvas

rafia 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!
Input texture

Random placement of blocks

Neighboring blocks constrained by overlap

Minimal error boundary cut
Minimal error boundary

overlapping blocks

vertical boundary

overlap error

2

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)
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$^{1,2}$
Chuck Jacobs$^2$
Nuria Oliver$^2$
Brian Curless$^3$
David Salesin$^{2,3}$

$^1$New York University
$^2$Microsoft Research
$^3$University of Washington
Image Analogies

A

A'

B

B'
Image Analogies

Goal: Process an image by example

Hertzmann et al. SIGGRAPH 2001
Non-parametric sampling

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
Texture-by-numbers
Sky
Sea
Rock
Sand
Output

Layout

Photo

SPADE, Park et al., CVPR 2019
Minimize Difference

GAN loss

Minimize Difference

pix2pix (Isola et al., CVPR 2017), pix2pixHD (Wang et al., CVPR 2018), SPADE (Park et al., CVPR 2019)
Results of pix2pixHD (Wang et al., CVPR 2018) on the Cityscapes dataset
Results of pix2pixHD (Wang et al., CVPR 2018) on the COCO-stuff dataset
Results of pix2pixHD (Wang et al., CVPR 2018) on the COCO-stuff dataset
Problem with standard approaches

cconv"normalization"
Problem with standard approaches
My simple fix

sky

$G$


**SPADE** (SPatially Adaptive Denormalization. Park et al., CVPR2019)
GAUGAN: SEMANTIC IMAGE SYNTHESIS WITH SPATIALLY ADAPTIVE NORMALIZATION

Taesung Park
University of California Berkeley

Chris Hebert
Gavriil Klimov
NVIDIA
Parametric Texture Synthesis

- Come up with a parametric generative model of the “infinite texture”
Multi-scale filter decomposition (steerable pyramid)
Step 1: Convolve with filterbank
Step 2: match per-channel histograms
Step 3: collapse pyramid and repeat!
Texture Synthesis

Images with equal model response

Slide by Portilla & Simoncelli (2000)
Figure 7: (Left pair) Inhomogeneous input texture produces blotchy synthetic texture. (Right pair) Homogeneous input.

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

Figure 9: More failures: hay and marble.
Simoncelli & Portilla ’98+

- 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 2\textsuperscript{nd} order statistics too!


\textbf{Figure 1.} Textures with matching marginal statistics.
Simoncelli & Portilla ’98+

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

• Optimize using repeated projections onto statistical constraint surfaces
Texture Synthesis

Slide by Portilla & Simoncelli (2000)
CNN as features

Convolutional Neural Network

Gatys et al. (NIPS 2015)
CNN - Multiscale Filter Bank

<table>
<thead>
<tr>
<th>Layer</th>
<th># Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>conv1_1</td>
<td>64</td>
</tr>
<tr>
<td>pool1</td>
<td>64</td>
</tr>
<tr>
<td>pool2</td>
<td>128</td>
</tr>
<tr>
<td>pool3</td>
<td>256</td>
</tr>
<tr>
<td>pool4</td>
<td>512</td>
</tr>
</tbody>
</table>
CNN - Texture Features

\[ F = [\bar{f}_1, \bar{f}_2, \bar{f}_3, \ldots, \bar{f}_N]^T \]

\[ G = FF^T \]

\[ \begin{pmatrix}
\langle \bar{f}_1, \bar{f}_1 \rangle & \cdots & \langle \bar{f}_1, \bar{f}_N \rangle \\
\langle \bar{f}_2, \bar{f}_1 \rangle & \ddots & \vdots \\
\vdots & \ddots & \ddots \\
\langle \bar{f}_N, \bar{f}_1 \rangle & \cdots & \langle \bar{f}_N, \bar{f}_N \rangle
\end{pmatrix} = \begin{pmatrix}
\sum_k F_{ik} F_{jk}
\end{pmatrix} \]
CNN - Texture Features

Gram Matrices

- 512 features
- 256 features
- 128 features
- 64 features
- 64 features
Texture Synthesis
Texture Synthesis

$$E_L = \sum (G^L - G^L)^2$$

$$G_0^L = \sum_{l=0}^{L} E_{l}^L$$

$$\mathcal{L}(\tilde{x}, \hat{x}) = \sum_{l=0}^{L} w_l E_l$$
Texture Synthesis
Texture Synthesis
Texture Synthesis

\[ E_L = \sum (\hat{G}^L - G^L)^2 \]

\[ \hat{G}_0 = \sum L F^L \]

\[ \frac{\partial E_L}{\partial F^L} = \frac{\partial E_L}{\partial F^{L-1}} \]

\[ L(x, x') = \sum_{l=0}^{L} w_l E_l \]

\[ \frac{\partial L}{\partial y} \]
Texture Synthesis
Texture Synthesis
Test Julesz’ Conjecture
Object Recognition is just Texture Recognition
Object Recognition is just Texture Recognition

Convolutional Neural Network

image $X$

“Shetland Sheepdog”

label $Y$

Gatys et al, 2017
A Neural Algorithm of Artistic Style

Gatys, Ecker, Bethge (arXiv 2015)
Van Gogh (1889)
CNN - Texture Synthesis

Gatys et al. (NIPS 2015)
Artistic Style Transfer
Artistic Style Transfer
Artistic Style Transfer
Artistic Style Transfer
Artistic Style Transfer

$$E_L = \sum (G^L - G^s)^2$$

$$G_0 = \sum \bar{F}_k \bar{F}_k^T$$

$$\mathcal{L}_{style} = \sum_{l} w_l E_l$$

Input

Feature maps

Van Gogh’s "Starry Night"
Artistic Style Transfer

\[ L_c = \sum (G^c - G^t)^2 \]
\[ G_t^c = \sum_i F^c_i F^t_i \]

\[ L_{\text{content}} = \sum (\hat{F}^c - F^c)^2 \]

\[ L_{\text{style}} = \sum \lambda L_t \]

Input

\[ \text{Feature maps} \]

\[ \text{Input} \]
Artistic Style Transfer

\[ \mathcal{L}_{\text{total}} = \alpha \mathcal{L}_{\text{content}} + \beta \mathcal{L}_{\text{style}} \]

\[ \mathcal{L}_{\text{content}} = \sum (F^l - F^l)^2 \]

\[ \mathcal{L}_{\text{style}} = \sum_{l} w_{l} E_{l} \]
Artistic Style Transfer
Artistic Style Transfer
Artistic Style Transfer
Artistic Style Transfer
Artistic Style Transfer
Relative Weighting of Content and Style

1e-4

1e-2

1e-3

1e-1
Different Reconstruction Layers

Conv2_2

Conv4_2
Different Reconstruction Layers

Conv2_2

Conv4_2
Different Reconstruction Layers

Original  Conv2_2  Conv4_2
General Style Transfer
General Style Transfer
Conjecture: GANs might be learning the “right” features to match for natural images.