Image and Video Retargeting

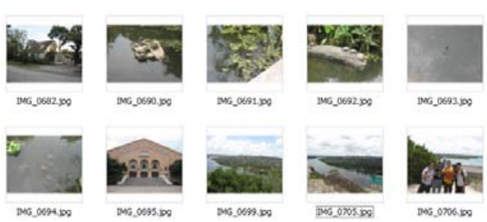
CS 395T: Visual Recognition and Search
Harshdeep Singh

What's coming?

- Content-aware retargeting
- Texture synthesis


One Big Squid


One Big Squid



What can be done?


Resize


Letterbox


Crop

Content-aware Retargeting

Lose the “insignificant” while preserving the “significant”…

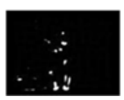
… and not disfiguring the image/video

What is “Significant”?

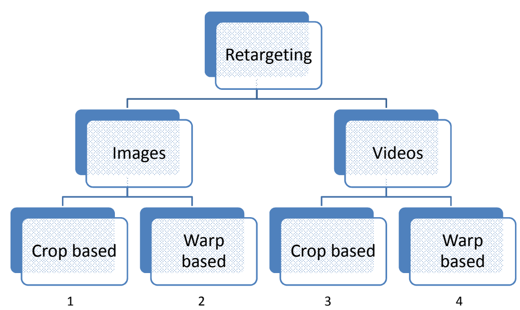
- High energy regions
  - Gradient, edges, entropy, histogram of gradient direction etc
- High motion regions
  - Or high *motion contrast* regions
- Faces
  - Or other known objects like cars
- Text

Saliency Map

Degree of saliency for each position in the image



Retargeting Algorithms



Automatic Thumbnail Cropping

Automatic Thumbnail Cropping and its Effectiveness, Suh et al, 2003

Automatic Thumbnail Cropping

- Problem
  - Find a rectangle in the image that
    - Has a small size
    - Contains most of the salient parts
- Solution (Greedy)
  - Initialize Rc as a small rectangle at the center
  - While cumulative saliency < threshold
    - R = small rectangle around the next most salient point
    - Rc = Rc U R

Automatic Thumbnail Cropping and its Effectiveness, Suh et al, 2003

Automatic Thumbnail Cropping

- Threshold can be adaptively chosen at the point of diminishing returns.
- Finding sum of pixels in a rectangular area is very fast (Integral image/summed area tables)

Automatic Thumbnail Cropping and its Effectiveness, Suh et al, 2003

User Experiments

Automatic Thumbnail Cropping and its Effectiveness, Suh et al, 2003

Seam Carving

Vertical seam – an 8-connected path of pixels from top to bottom, containing one pixel in each row

Horizontal seam – left to right

Remove lowest energy seam iteratively

Energy of a pixel

$$e_1(\mathbf{I}) = |\frac{\partial}{\partial x}\mathbf{I}| + |\frac{\partial}{\partial y}\mathbf{I}|$$

Seam Carving for Content-Aware Image Resizing, Avidan et al, SIGGRAPH 2007

Use of Dynamic Programming

- To find the optimal seam
- To find the optimal order of horizontal and vertical seams to be removed to resize an n x m image to n' x m'.

Seam Carving for Content-Aware Image Resizing, Avidan et al, SIGGRAPH 2007

Works?

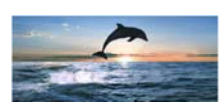
Using IntuImage - http://www.intuimage.com/

Seam Carving for Content-Aware Image Resizing, Avidan et al, SIGGRAPH 2007

Image Enlarging

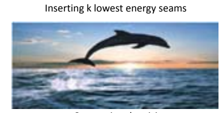
Find k lowest energy seams. Insert a new seam for each of them by averaging with left and right neighbors.

Original



Inserting k lowest energy seams

Repeatedly inserting the same seam



Conventional resizing

Seam Carving for Content-Aware Image Resizing, Avidan et al, SIGGRAPH 2007

Other applications

Content amplification

Scale up the image using standard methods. Apply seam carving to bring back to original dimensions.

Object removal

User marks an object. Remove seams until all marked pixels have been eliminated. Insert new seams.

Seam Carving for Content-Aware Image Resizing, Avidan et al, SIGGRAPH 2007

Cropping vs. Warping

Image: Non-homogeneous Content-driven Video-retargeting, Wolf et al, ICCV 2007

Video Retargeting by Cropping

- Salient region may change from one frame to another
- May need to add camera motion to preserve it
- The resulting video must be *cinematically plausible*. (Avoid zooms, instant camera acceleration etc)
- Works on each *shot* separately

Video Retargeting: Automating Pan and Scan, Liu et al ,ACM Multimedia, 2006

Video Shot Detection

- Shot
  - An unbroken sequence of frames from one camera
- Detecting shot boundaries
  - Pixel differences
  - Histogram comparisons
  - Edge differences
  - Motion vectors

Comparison of video shot boundary detection techniques, Boreczky 1996

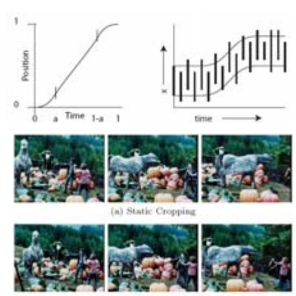
Retargeting one shot

- Crop
  - Salient features stay within the same region throughout the shot
  - A single cropping window for the entire shot
  - No camera motion added

Video Retargeting: Automating Pan and Scan, Liu et al ,ACM Multimedia, 2006

Retargeting video shot

- Virtual Pans
  - Salient region changes during the shot gradually
  - Limited to a single horizontal pan
  - Easy in easy out



(a) Static Cropping

(b) Virtual Pan

Video Retargeting: Automating Pan and Scan, Liu et al ,ACM Multimedia, 2006

Retargeting video shot

- Virtual Cuts
  - Salient region changes abruptly
  - One shot into two
  - One subshot comes from the left part, other from the right

Video Retargeting: Automating Pan and Scan, Liu et al ,ACM Multimedia, 2006

Video Retargeting by Warping

- Warp – maps pixels in the original frame to the retargeted frame
- An unimportant pixel should be mapped close to its neighbors
  - Gets blended with them
- An important pixel should be mapped far from its neighbors
  - Size of regions of important pixels remains the same

Non-homogeneous Content-driven Video-retargeting, Wolf et al, ICCV 2007

Optimize under constraints

1. Each pixel should be at a fixed distance from its left and right neighbors (depending on importance)
2. Each pixel needs to be mapped to a location similar to one of its upper and lower neighbors
3. Mapping of a pixel at time t should be similar to its mapping at t+1
4. Warped locations must fit to the dimensions of the target frame

Non-homogeneous Content-driven Video-retargeting, Wolf et al, ICCV 2007

Benefits over Seam-Carving

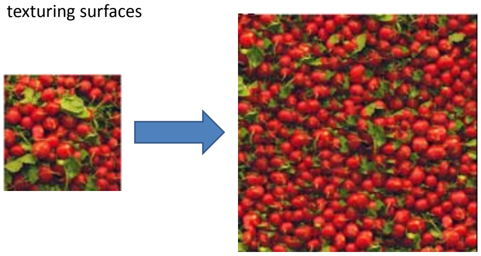
- Maintains temporal coherence in videos
- Causes less deformation under severe down-sizing

Original Wolf et al Seam-Carving

Non-homogeneous Content-driven Video-retargeting, Wolf et al, ICCV 2007

Texture Synthesis

- Goal – Create new samples of a given texture
- Many applications – virtual environments, hole filling, texturing surfaces

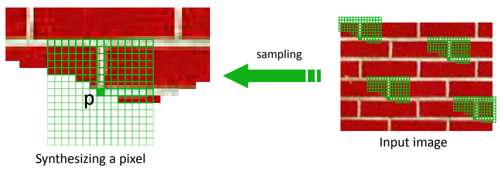


Slide from Kristen, CS 378 Fall 07

Roadmap

- A simple and intuitive algorithm
  - But slow ☹
    - Efros and Leung, 1999
- Acceleration strategies
  - Improving search time with a tree
    - Wei et al, 2000
  - Synthesizing in bigger blocks, using spatial coherence
    - Efros and Freeman, 2001
- Video Textures
  - Schodl at al, 2000
- Using Graphcuts iteratively for image and video textures
  - Kwatra et al, 2003

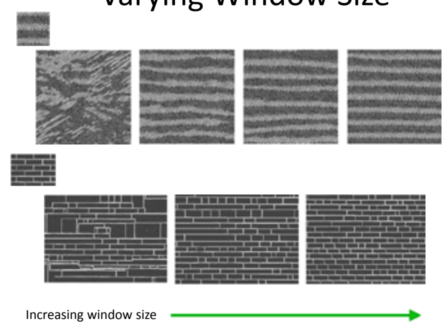
Efros & Leung '99



- Assuming Markov property, compute P(p|N(p))
  - Building explicit probability tables infeasible
  - Instead, let's *search the input image* for all similar neighborhoods — that's our histogram for **p**
  - To synthesize p, just pick one match at random

Slide from Efros SIGGRAPH 2001

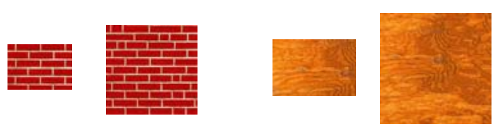
Varying Window Size



Slide from Kristen, CS 378 Fall 07

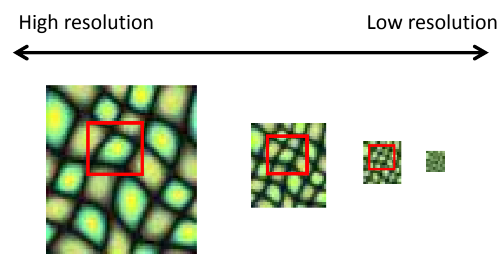
Efros & Leung '99

- The algorithm
  - o Very simple
  - o Surprisingly good results
  - o ...but very slow
- Bottlenecks
  - o Have to search entire input texture to synthesize each pixel
  - o Bigger neighborhood => Slower search



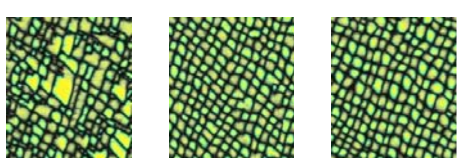
Slide modified from Efros SIGGRAPH 2001

Multi-resolution Pyramid



Fast Texture Synthesis using Tree-structured Vector Quantization , Wei et al,
Slide from Wei, SIGGRAPH 2000

Results

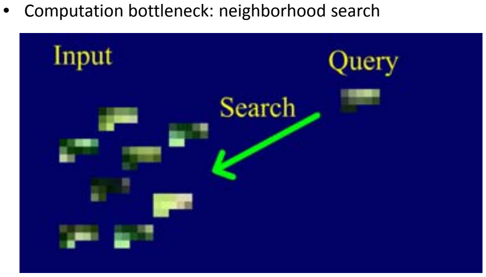


1 level 5×5 | 1 level 11×11 | 3 levels 5×5

Fast Texture Synthesis using Tree-structured Vector Quantization , Wei et al,
Slide modified from Wei, SIGGRAPH 2000

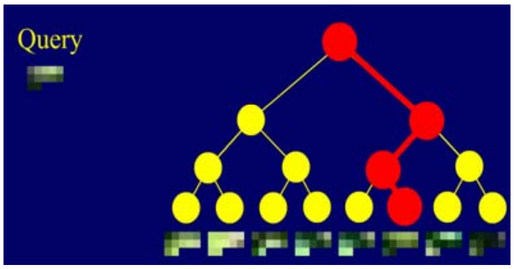
Tree-structured Vector Quantization

- Computation bottleneck: neighborhood search



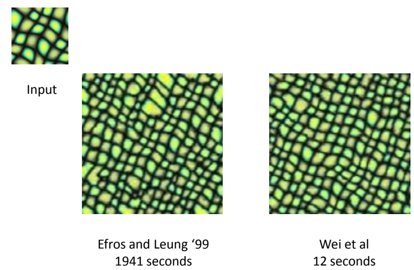
Fast Texture Synthesis using Tree-structured Vector Quantization , Wei et al,
Slide from Wei, SIGGRAPH 2000

Tree-structured Vector Quantization



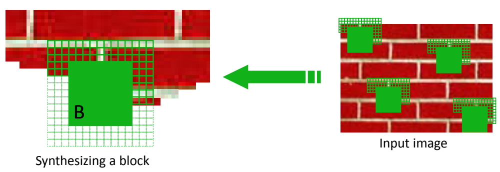
Fast Texture Synthesis using Tree-structured Vector Quantization , Wei et al,
Slide from Wei, SIGGRAPH 2000

Comparison



Fast Texture Synthesis using Tree-structured Vector Quantization , Wei et al,
Slide from Wei, SIGGRAPH 2000

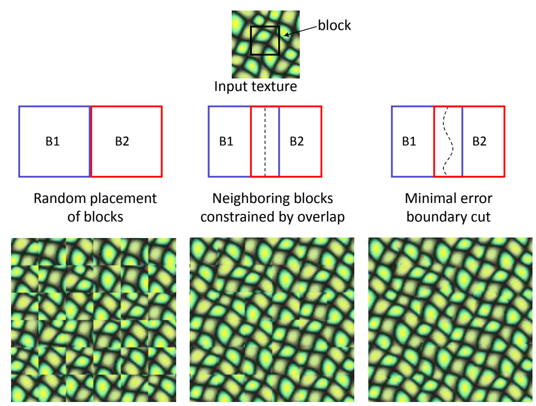
Efros & Leung '99 extended



- Observation: neighbor pixels are highly correlated

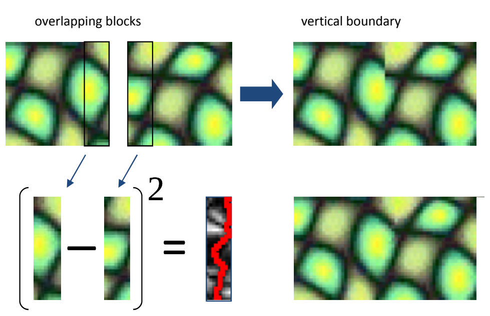
Idea: unit of synthesis = block
- Exactly the same but now we want P(B|N(B))
- Much faster: synthesize all pixels in a block at once

Slide from Efros SIGGRAPH 2001



Slide from Efros SIGGRAPH 2001

Minimal error boundary



Slide from Efros SIGGRAPH 2001

Synthesis Results

Slide from Kristen, CS 378 Fall 07

Failures

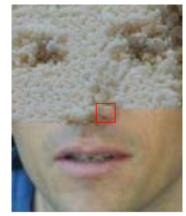
(Chernobyl Harvest)

Slide from Efros SIGGRAPH 2001

Texture Transfer

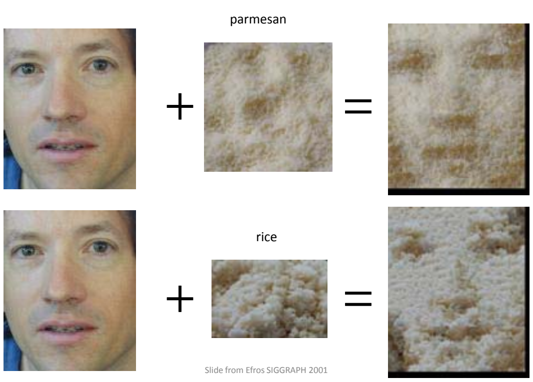
•Take the texture from one object and "paint" it onto another object

- This requires separating texture and shape
- That's HARD, but we can cheat
- Assume we can capture shape by boundary and rough shading



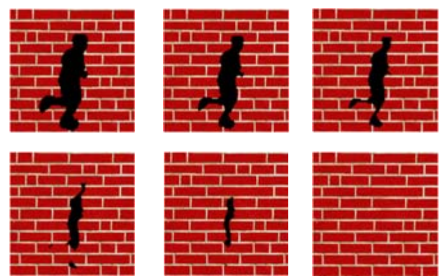
Then, just add another constraint when sampling: similarity to underlying image at that spot

Slide from Efros SIGGRAPH 2001



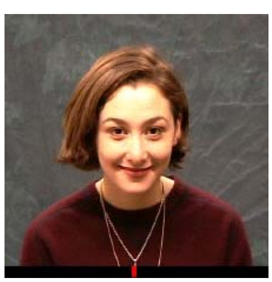
Slide from Efros SIGGRAPH 2001

Hole Filling



Slide from Kristen, CS 378 Fall 07

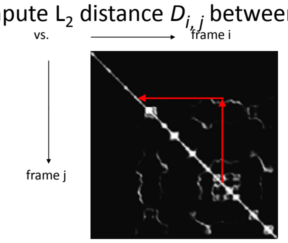
Video textures



Video Textures, Schodl et al, SIGGRAPH 2000

Finding good transitions

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



Similar frames make good transitions

Video Textures, Schodl et al,
Slide form Schodl, SIGGRAPH 2000

Transition probabilities

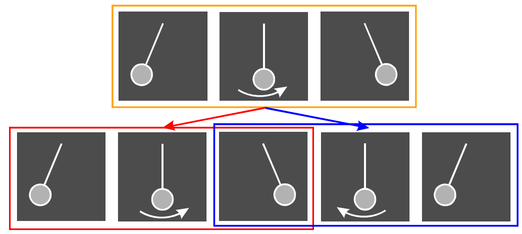
Probability for transition $P_{i\to j}$ inversely related to cost:

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

high σ low σ

Video Textures, Schodl et al,
Slide form Schodl, SIGGRAPH 2000

Preserving dynamics

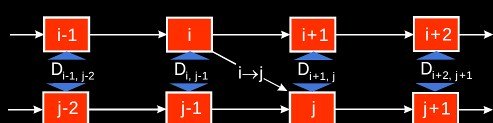


Video Textures, Schodl et al,
Slide form Schodl, SIGGRAPH 2000

Preserving dynamics

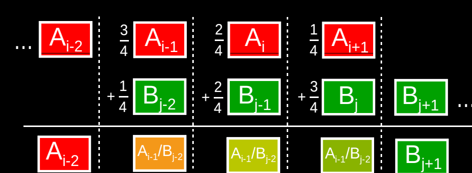
Cost for transition $i\to j$

$$C_{i\to j} = \sum_{k=-N}^{N-1} w_k D_{i+k+1,\, j+k}$$



Video Textures, Schodl et al.
Slide from Schodl, SIGGRAPH 2000

Crossfading

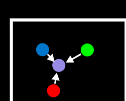


Video Textures, Schodl et al.
Slide from Schodl, SIGGRAPH 2000

Morphing

- Interpolation task:

  $\frac{2}{5}$ A + $\frac{2}{5}$ B + $\frac{1}{5}$ C



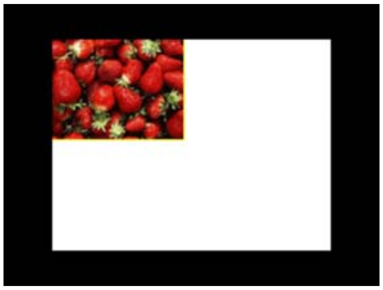
- Compute correspondence between pixels of all frames
- Interpolate pixel position and color in morphed frame
- based on [Shum 2000]

Video Textures, Schodl et al.
Slide from Schodl, SIGGRAPH 2000

Interactive fish

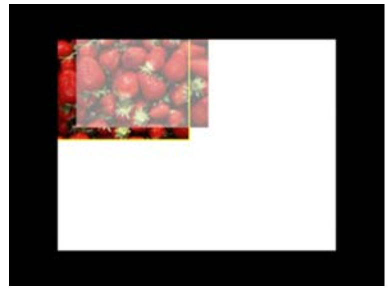
Video Textures, Schodl et al.
Slide from Schodl, SIGGRAPH 2000

Graphcut Textures



Graphcut Textures: Image and Video Synthesis Using Graph Cuts, Kwatra et al, SIGGRAPH 2003

Graphcut Textures



Graphcut Textures: Image and Video Synthesis Using Graph Cuts, Kwatra et al, SIGGRAPH 2003

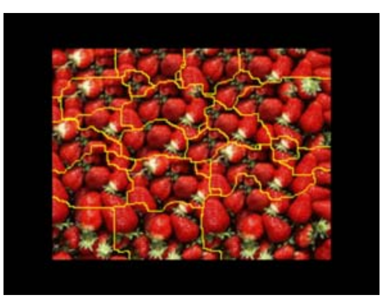
Graphcut Textures

Graphcut Textures: Image and Video Synthesis Using Graph Cuts, Kwatra et al, SIGGRAPH 2003

Graphcut Textures

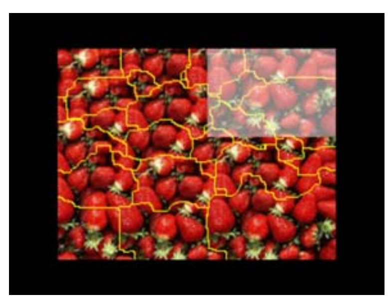
Graphcut Textures: Image and Video Synthesis Using Graph Cuts, Kwatra et al, SIGGRAPH 2003

Graphcut Textures



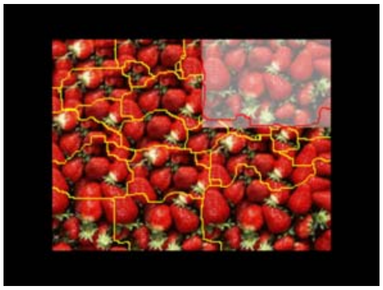
Graphcut Textures: Image and Video Synthesis Using Graph Cuts, Kwatra et al, SIGGRAPH 2003

Graphcut Textures



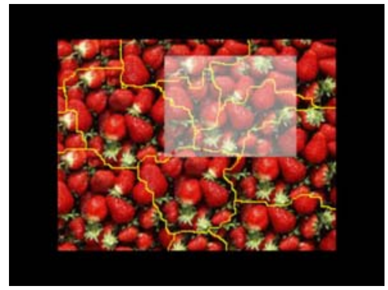
Graphcut Textures: Image and Video Synthesis Using Graph Cuts, Kwatra et al, SIGGRAPH 2003

Graphcut Textures



Graphcut Textures: Image and Video Synthesis Using Graph Cuts, Kwatra et al, SIGGRAPH 2003

Graphcut Textures



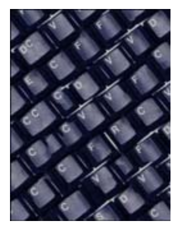
Graphcut Textures: Image and Video Synthesis Using Graph Cuts, Kwatra et al, SIGGRAPH 2003

Graphcut Textures

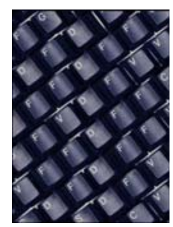
Graphcut Textures: Image and Video Synthesis Using Graph Cuts, Kwatra et al, SIGGRAPH 2003

Result comparison

Original



Efros and Freeman



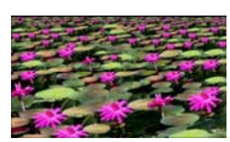
Graphcut

Graphcut Textures: Image and Video Synthesis Using Graph Cuts, Kwatra et al, SIGGRAPH 2003

Rotation and Mirroring

Graphcut Textures: Image and Video Synthesis Using Graph Cuts, Kwatra et al, SIGGRAPH 2003

Hallucinating Perspective



Graphcut Textures: Image and Video Synthesis Using Graph Cuts, Kwatra et al, SIGGRAPH 2003

Interactive Merging and Blending

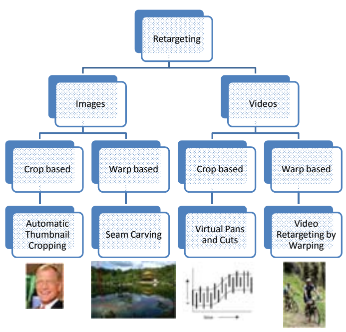
Graphcut Textures: Image and Video Synthesis Using Graph Cuts, Kwatra et al, SIGGRAPH 2003

Video Textures using Graphcut

- Works as well in 3D
- Patch – 3D space-time block of video
- Seam – a 2D surface that sits in 3D
- Transition is determined on a per-pixel basis and not for the entire image
- Does not have to use crossfading or morphing (like Schodl et al), so no blur artifacts

Graphcut Textures: Image and Video Synthesis Using Graph Cuts, Kwatra et al, SIGGRAPH 2003

Recap



Recap

- A simple and intuitive algorithm
  - But slow ☹
    - Efros and Leung, 1999
- Acceleration strategies
  - Improving search time with a tree
    - Wei et al, 2000
  - Synthesizing in bigger blocks, using spatial coherence
    - Efros and Freeman, 2001
- Video Textures
  - Schodl at al, 2000
- Using Graphcuts iteratively for image and video textures
  - Kwatra et al, 2003

Discussion Points

- What other features can be used to define saliency?
- How can multi-size videos be generated and represented efficiently?
- Looking at the content of an image, can we estimate how much we can warp it (or how many seams we can remove) without distorting it much?
- Can we automatically decide whether to use a crop-based or warp-based retargeting?
- What sort of experiments should be carried out to evaluate the results of a retargeting algorithm?
- How and when can texture synthesis be used for image/video compression?
- What all needs to be taken care of while extending hole filling, object removal, expansion etc to videos?
- How can the neighborhood scale in space and time be automatically selected for texture synthesis.