Visual inpainting

- Complete visual data, given surrounding
  - Visually plausible, at least pleasing
  - Different from texture synthesis (though related)
Applications

- Object removal, concealment of all sorts (including packet loss)

Welcome to King's College, a constituent college of Cambridge University, England. King's is one of the oldest Cambridge colleges, having been founded in 1441 by Henry VI. It is also Cambridge's premier tourist attraction, due above all to its spectacular Perpendicular chapel.
Applications

- Object removal, concealment of all sorts (including packet loss)
Applications

- Image extension
Applications

- Image extension
Brief (partial) history

- 1998: Masnou and Morel. Disocclusion with level lines
- [1999: Efros and Leung. Example-based texture synthesis]
- 2000: Bertalmio et al. Inpainting as PDE
- [2001: Efros and Freeman. Texture synthesis by quilting]
- 2001: Harrison. Inpainting with exampled-based texture synthesis
- 2002: Bornard et al. Greedy patch-based inpainting
- 2003: Criminisi et al. Greedy patch-based inpainting with priorities
- [2005: Buades and Morel. NL means]
- 2008: Komodakis and Tziritas. Iterative example-based inpainting
- [2009: Barnes et al. PatchMatch]
- 2010: Aujol et al. Variational patch-based inpainting
- 2012: Granados et al. HD video inpainting
An exemplar-based inpainting

Dictionary of block's neigh.

Source

Target
An examplar-based inpainting
An exemplar-based inpainting
An examplar-based inpainting
An examplar-based inpainting
An examplar-based inpainting
An examplar-based inpainting
Why (square) patches?

- Powerful to recreate both geometry and texture (even if none)
- Easy to manipulate
- Can be linearly combined in semi-local processing
- Have become variationally friendly
Which patch size?

- Depends on occlusion size (at coarsest level if multiscale)
- Depends on scale of visual “patterns” in image
Video inpainting

- Not a sequence of image inpainting
- Challenges
  - Temporal consistency (inc. at object level)
  - Dynamic plausibility
  - Complexity of dynamic scene
  - Computational cost
- Wexler et al. (2004-2007)
# Patch-based video inpainting

<table>
<thead>
<tr>
<th>Method</th>
<th>Iterative</th>
<th>Generic</th>
<th>Automatic</th>
<th>Slow</th>
<th>Difficult to reproduce</th>
<th>Patwardhan et al. (2005, 2007)</th>
<th>Greedy</th>
<th>Reproducible</th>
<th>Relatively fast</th>
<th>Requires segmentation</th>
<th>No cost optimization</th>
<th>Granados et al. (2012)</th>
<th>Graph-cut</th>
<th>Generic</th>
<th>High resolution</th>
<th>Requires segmentation</th>
<th>Slow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wexler et al. (2004-2007)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patwardhan et al. (2005, 2007)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Granados et al. (2012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Toward fast and generic video inpainting

- Alasdair Newson’s PhD work
  - Co-advised with A. Almansa & Y. Gousseau
- Builds on Wexler *et al.*
  - Alternate patch search and reconstruction
  - Multi-scale

- Distinctive features
  - Fast approximate patch search
  - Greedy inpainting as initialization
  - Dominant motion compensation
  - Texture-sensitive patch comparison
  - Reproducible!
Toward fast and generic video inpainting

- Alasdair Newson’s PhD work
  - Co-advised with A. Almansa & Y. Gousseau
- Builds on Wexler *et al.*
  - Alternate patch search and reconstruction
  - Multi-scale

- Distinctive features
  - Fast approximate patch search
  - Greedy inpainting as initialization
  - Dominant motion compensation
  - Texture-sensitive patch comparison
  - Reproducible!
Notations

- Image: \( u : \Omega \rightarrow \mathbb{R}^3 \)
- Spatio-temporal cuboid centered at \( p = (x, y, t) : \mathcal{N}_p \subset \Omega \)
- Spatio-temporal image patch: \( W_p = [u(q)]_{q \in \mathcal{N}_p} \in \mathbb{R}^{3 \times N} \)
- Hole and data: \( \Omega = \mathcal{H} \cup \mathcal{D} \)
- “Shift map”: \( \phi : \Omega \rightarrow \mathbb{R}^3 \text{ s.t. } \forall p, p + \phi(p) \in \tilde{\mathcal{D}} \)
Patch-based energy

- Compound energy

\[
E(u, \phi) = \sum_{p \in \mathcal{H}} \| W_p - W_{p + \phi(p)} \|^2_2 \\
= \sum_{p \in \mathcal{H}} \sum_{q \in \mathcal{N}_p} \| u(q) - u(q + \phi(p)) \|^2_2
\]

- Alternate minimization
  - Patch matching, given image

\[
\phi(p) = \arg \min_{q \in \tilde{D}} \| W_p - W_q \|^2_2 - p, \forall p \in \Omega \setminus \tilde{D}
\]
  - Image reconstruction over occlusion, given shift map

\[
u(p) = \sum_{q \in \mathcal{N}_p} s^q_p u(p + \phi(q)) \sum_{q \in \mathcal{N}_p} s^q_p, \forall p \in \mathcal{H}
\]

Uniform weights, except for last reconstruction: copy from best match
Coarse-to-fine

- To improve speed and result quality for large occlusions
- Image pyramid, same patch size through it
- Several iterations (e.g., 20 max.) at each level and up-sample to next
- Final reconstruction: with best match only
Coarse-to-fine

- To improve speed and result quality for large occlusions
- Image pyramid, same patch size through it
- Several iterations (e.g., 20 max.) at each level and up-sample to next
- Final reconstruction: with best match only
Initialization

- First filling-in a coarsest resolution: greedy (concentric)
Approximate patch matching

- PatchMatch (Barnes et al. 2009): fast, dense patch correspondences

Key ideas
- Shift map is piece-wise constant
- Randomization

Principle
- Lexicographic passes
- Exploit regularity: past neighbors propose their shifts
- Random proposal as well

Straightforward adaption to 2D+t
First results
First results
First results
# Timings

<table>
<thead>
<tr>
<th></th>
<th>Beach umbrella (265x68x200)</th>
<th>Crossing ladies (170x80x87)</th>
<th>Jumping girl (1120x754x200)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wexler (kd-tree)</td>
<td>~1000s</td>
<td>~1000s</td>
<td>~8000s</td>
</tr>
<tr>
<td>Ours (PatchMatch3D)</td>
<td>~50s</td>
<td>~30s</td>
<td>~150s</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Beach umbrella (265x68x200)</th>
<th>Duo (960x704x154)</th>
<th>Museum (1120x754x200)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Granados (graph-cut)</td>
<td>11h</td>
<td>90h</td>
<td></td>
</tr>
<tr>
<td>Ours w/o texture</td>
<td>14mn</td>
<td>4h</td>
<td>4h</td>
</tr>
<tr>
<td>Ours</td>
<td>24mn</td>
<td>6h</td>
<td>6h</td>
</tr>
</tbody>
</table>
Dealing with texture

- SSD not sensitive enough to fine grain texture
Dealing with texture

- SSD not sensitive enough to fine grain texture
Extended patch

- Include simple texture information (inspired by Liu and Caselles 2013)

\[
T(p) = \frac{1}{\text{card}(\nu)} \sum_{q \in \nu} (|I_x(q)|, |I_y(q)|)
\]

\[
W_{p} = [(u(q), \lambda T(q))]_{q \in \mathcal{N}_p}
\]

- Joint multi-scale pyramid
Texture-aware inpainting
Texture-aware inpainting
Texture-aware inpainting
Texture-aware inpainting
Texture-aware inpainting
Texture-aware inpainting
Dealing with camera motion

- Camera motion, even small, is a problem
- Good match = similar structure with similar apparent movement

\[ W_p = [\begin{array}{cccc}
\end{array}] \]

\[ W_{p+\phi(p)} = [\begin{array}{cccc}
\end{array}] \]

- Simple solution
  - Estimate and compensate dominant motion wrt middle frame
  - Inpaint video
  - Put dominant motion back
Dealing with camera motion
Dealing with camera motion
Dealing with camera motion
Dealing with camera motion
Dealing with camera motion
Dealing with camera motion
Results of complete system
Results of complete system
Results of complete system
Results of complete system

4/8/2014
Inpainting outlooks

- Between *inventing* and *copying*
  - Inpainting in the (spatio-temporal) gradient domain
  - Dictionary learning and sparse coding
  - Revisiting texture analysis/synthesis
  - Structure, texture, chromaticity, reflectance, transparency

- Other *data completion* problems
  - Video and multi-view
  - Unconventional images (plenoptic, multispectral, MRI, etc.)
  - Audio
  - Depth maps, point clouds and meshes