VIDEO INPAINTING & DYNAMIC TEXTURE MODELING

Alasdair Newson³  Andrés Almansa¹²  Matthieu Fradet³
Yann Gousseau²  Patrick Perez³

Challenges in Dynamic Imaging Data - 9-11 June 2015
TGM - Isaac Newton Institute - Cambridge
What is inpainting?
Introduction

What is inpainting?

- Removal and filling of a region in an image or video
- The inpainted region should be visually convincing/pleasing
Introduction

What is inpainting?

- Removal and filling of a region in an image or video
- The inpainted region should be visually convincing/pleasing

What is inpainting useful for?

- Restoring/improving/modifying images/videos
- Post-production of films
Introduction

Video inpainting

Inpainted video
Introduction

Video inpainting

Original video
Introduction

Challenges in *image* inpainting

- Filling-in geometric structures (amodal completion)
  
  [Masnou & Morel 1998], [Bertalmio *et al.* 2000]
Challenges in *image* inpainting

- **Filling-in geometric structures (amodal completion)**
  [Masnou & Morel 1998], [Bertalmio *et al.* 2000]

- **Texture synthesis**
  [Efros & Leung 1999]
Introduction

Challenges in *image* inpainting

- **Filling-in geometric structures (amodal completion)**
  [Masnou & Morel 1998], [Bertalmio *et al.* 2000]

- **Texture synthesis**
  [Efros & Leung 1999]

- **Geometry + conditional texture synthesis**
  [Cao *et al.* 2011]
Additional challenges of *video* inpainting

- *Temporal coherency*
- *Dynamic geometry* (reconstruction of moving objects)
- *Dynamic texture* (water pouring, flowing, flames)
- *Simultaneous foreground/background reconstruction*
- *Extremely long computational times*

Inpainting example (from Wexler *et al.* 2007)
Proposed inpainting algorithm

Exemplar-based inpainting
Similar to [Wexler 2004] (video inpainting) and [Arias et al. 2012] (image inpainting)

Key ingredients
- spatio-temporal patches
- alternated minimization of bi-convex energy
- multi-scale coarse-to-fine model
- accelerated ANN search
- texture-aware distance between patches
- how many scales?
- motion compensation
Proposed inpainting algorithm

Video inpainting notation

\[ W_p: \text{a patch centered at } p \]
Proposed inpainting algorithm

Inpainting Principle

Input: \( u|_D \)
Output: \( u|_H \)
Find \( u|_H \) by minimizing

\[
E(u, \phi) = \sum_{p \in H} \left\| W_p^u - W_{p+\phi(p)}^u \right\|_2^2
\]
Proposed inpainting algorithm

<table>
<thead>
<tr>
<th>Inpainting Principle</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input:</strong> $u</td>
</tr>
<tr>
<td><strong>Output:</strong> $u</td>
</tr>
<tr>
<td>Find $u</td>
</tr>
<tr>
<td>$E(u, \phi) = \sum_{p \in \mathcal{H}}</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>• non-convex energy</td>
</tr>
<tr>
<td>• high dimensionality (dimension $= 5 \times 5 \times 5 \times 3 \approx 500$)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• alternate (convex) minimization w.r.t. $u$ and $\phi$</td>
</tr>
<tr>
<td>• coarse-to-fine processing</td>
</tr>
<tr>
<td>• approximate nearest neighbours</td>
</tr>
<tr>
<td>• fine-level texture features in coarsest level</td>
</tr>
</tbody>
</table>
Proposed inpainting algorithm

**Inpainting Principle**

Input: $u|_D$

Output: $u|_H$

Find $u|_H$ by minimizing

$$E(u, \phi) = \sum_{p \in H} ||W_p^u - W_p^u + \phi(p)||_2^2$$

**Denoising Principle**

Input: noisy $\tilde{u}$

Output: denoised $\hat{u}$

Find $u$ by minimizing

$$E(u, w) = \sum_{p, q} w(p, q)||W_p^\tilde{u} - W_q^u||_2^2 + h \sum_p H(w(p, \cdot))$$

---

‡ P. Arias, G. Facciolo, V. Caselles, G. Sapiro, A Variational Framework for Exemplar-Based Image Inpainting, IJCV 2011
**Proposed inpainting algorithm**

**Inpainting Principle**

Input: $u|_D$

Output: $u|_H$

Find $u|_H$ by minimizing

$$E(u, \phi) = \sum_{p \in \mathcal{H}} ||W_p^u - W_p^u + \phi(p)||^2_2$$

**Algorithm (inspired by † and ‡):**

Alternate Minimization on $u$ and $\phi$:

1. $u^0 \leftarrow$ Initialisation$(u|_D, \mathcal{H})$
2. $\phi^{k+1} \leftarrow$ NearestNeighbourSearch($u^k$)
3. $u^{k+1} \leftarrow$ VideoReconstruction($\phi^{k+1}$)

(Carried out in a Multiresolution scheme)

† Y. Wexler, E. Schechtman, M. Irani, *Space-Time Completion of Video*, PAMI 2007
Proposed inpainting algorithm

**Inpainting Principle**

**Input:** $u|_D$

**Output:** $u|_H$

**Find** $u|_H$ **by minimizing**

$$E(u, \phi) = \sum_{p \in H} ||W^u_p - W^u_{p+\phi(p)}||^2_2$$

- **Target patch**
- **Nearest neighbour patch**
- **Shift map (nearest neighbour offsets)**
- **D** Unoccluded region
- **H** Occlusion

$W^u_p$: a patch centered at $p$

1. **Algorithm (inspired by † and ‡):**
   1. **Alternate Minimization on** $u$ **and** $\phi$:
      1. $u_0 \leftarrow \text{Initialisation}(u|_D, H)$
      2. $\phi_{k+1} \leftarrow \text{NearestNeighbourSearch}(u_k)$
      3. $u_{k+1} \leftarrow \text{VideoReconstruction}(\phi_{k+1})$

( aggregation of patches)

(Carried out in a Multiresolution scheme)
Approximate Nearest Neighbour (ANN) search
Approximate Nearest Neighbour (ANN) search

High dimensionality of problem means NN search is very slow

- Previously used ANN search algorithm (kdTrees) very slow
- We extend the PatchMatch [Barnes et al. 2009] algorithm to spatio-temporal case.
- PatchMatch based on piecewise constancy of the shift map $\phi$


High definition example (1120 × 754)

- 10-50 times speedup with 3D PatchMatch
- 10 times speedup compared to Granados et al.

M. Granados, J. Tompkin, K.I. Kim, O. Grau, J. Kautz, C. Theobalt,
How Not to Be Seen - Object Removal from Videos of Crowded Scenes,

<table>
<thead>
<tr>
<th></th>
<th>One matching pass at full res.</th>
<th>Total timing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beach umbrella (265x68x200)</td>
<td>Duo (960x704x154)</td>
</tr>
<tr>
<td>Wexler (kd-tree)</td>
<td>~1000s</td>
<td>~30s</td>
</tr>
<tr>
<td>Ours (PatchMatch3D)</td>
<td>~50s</td>
<td>~30s</td>
</tr>
<tr>
<td>Granados (graph-cut)</td>
<td>11h</td>
<td>4h</td>
</tr>
<tr>
<td>Ours w/o texture</td>
<td>14mn</td>
<td>4h</td>
</tr>
<tr>
<td>Ours</td>
<td>24mn</td>
<td>6h</td>
</tr>
</tbody>
</table>
Textures in image/video inpainting
Dealing with textures in images and videos

Why do textures pose a problem?

Original image
Dealing with textures in images and videos

Why do textures pose a problem?

Inpainted image
Dealing with textures in images and videos

Why do textures pose a problem?

Incorrect approximate nearest neighbours
Dealing with textures in images and videos

Why do we identify incorrect patches ???
Dealing with textures in images and videos

Why do we identify incorrect patches ???

Imagine we want to find the ANN of a random patch:

\[ x_i \sim \mathcal{N}(\mu, \sigma^2) \quad i = 1 \ldots N \]

Which patch is most similar to \( X \) ?

\[ y_i \sim \mathcal{N}(\mu, \sigma^2) \quad i = 1 \ldots N \]

\[ z_i = \mu \quad i = 1 \ldots N \]
Dealing with textures in images and videos

Why do we identify incorrect patches ???

Imagine we want to find the ANN of a *random* patch:

\[
\begin{align*}
x_i &\sim \mathcal{N}(\mu, \sigma^2) \\
i = 1 \cdots N
\end{align*}
\]

Total patch size = \( N \)

\[ E[d(X, Y)] = 2N\sigma^2 \]
\[ E[d(X, Z)] = N\sigma^2 \]
Dealing with textures in images and videos

Why do we identify incorrect patches???

Imagine we want to find the ANN of a random patch:

\[ x_i \sim \mathcal{N}(\mu, \sigma^2) \quad i = 1 \cdots N \]

Total patch size = \( N \)

\[ y_i \sim \mathcal{N}(\mu, \sigma^2) \quad i = 1 \cdots N \]

\[ z_i = \mu \quad i = 1 \cdots N \]

\[ E[d(X, Y)] = 2N\sigma^2 \]

\[ E[d(X, Z)] = N\sigma^2 \]

On average, \( d(X, Y) \) is \textit{twice} as large as \( d(X, Z) \).

On average, constant patch \( Z \) is preferred!
Dealing with textures in images and videos

Why do we identify incorrect patches ???

Imagine we want to find the ANN of a random patch:

\[ x_i \sim \mathcal{N}(\mu, \sigma^2) \quad i = 1 \cdots N \]

Total patch size = \( N \)

\[ y_i \sim \mathcal{N}(\mu, \sigma^2) \quad i = 1 \cdots N \]

\[ z_i = \mu \quad i = 1 \cdots N \]

On average, \( d(X, Y) \) is twice as large as \( d(X, Z) \).

On average, constant patch \( Z \) is preferred !

Solution ? Change patch distance !
Modified patch distance

We wish to include some information pertaining to the texture.

Idea: include an estimation of the local variance
Modified patch distance

We wish to include some information pertaining to the texture.

Idea: include an estimation of the local variance

Different possibilities were tested. Finally, we chose (inspired by Liu and Caselles 2013\(^\dagger\)):

\[
\text{SSD: } [R, G, B, \alpha g_\nu \ast |\nabla_x I|, \alpha g_\nu \ast |\nabla_y I|]
\]

\(\alpha\): a weighting scalar

\(g_\nu\) in a gaussian kernel of size \(\nu\).

\(^\dagger\) Y. Liu, V. Caselles, Exemplar-Based Image Inpainting Using Multiscale Graph Cuts, IEEE TIP (2013)

\(^\ddagger\) J. Bruna & S. Mallat (2013). Invariant scattering convolution networks. IEEE TPAMI, 35(8), 187286
Modified patch distance

Example of image created by $|\nabla_x l|_\nu$
Modified patch distance

Example of the impact of the modified distance

PatchMatch with regular SSD
Modified patch distance

Example of the impact of the modified distance

PatchMatch with modified SSD
Image example

Inpainting with unmodified patch distance
Image example

Inpainting with “Image Melding” (Darabi et al. 2012)
Image example

Inpainting with modified patch distance
Image example

Original image
Noise example

Inpainting with unmodified patch distance
Noise example

Inpainting with modified patch distance
Noise example

Inpainting with unmodified patch distance
Noise example

Inpainting with modified patch distance
Video example

Original video
Video example

Unmodified patch distance
Video example

Modified patch distance
Local minima, convergence and binary inpainting
The *multi-resolution* scheme is necessary to correctly inpaint structures.
Convergence questions, local minima

The *multi-resolution* scheme is necessary to correctly inpaint structures.

Occluded image  
Result with one pyramid level  
Result with three pyramid levels

Some interesting questions:

- Can we quantify the amount of subsampling needed?
- Can we guarantee convergence to a desirable solution?

Such questions are difficult to answer in general, so we use a simple situation!
Convergence questions, local minima

The *multi-resolution* scheme is necessary to correctly inpaint structures.

- Study (very) simple situation
- Behaviour of the algorithm is easier to study
Binary inpainting

Main theoretical results in simple 1D case:

- Algorithm converges if the occlusion size is less than $2N - 2\sqrt{N} + 1$
  - $N$ is the patch size
- Otherwise, algorithm may be stuck in a local minimum
Binary inpainting

Main theoretical results in simple 1D case:

- Algorithm converges if the occlusion size is less than $2N - 2\sqrt{N} + 1$
  - $N$ is the patch size
- Otherwise, algorithm may be stuck in a local minimum

Verified for simple 2D situations.

Simple structure example. Patch size = $11 \times 11$

Occlusion size = 16    Occlusion size = 17
Inpainting results

Original video
Inpainting results

Our inpainting result
Inpainting results

Original video
Inpainting results

Our inpainting result
Open Issues

- complex motions
- long temporal occlusions
- select the patch size

Input image (occlusion border in red)

Patch size = 3 × 3

Patch size = 5 × 5

Patch size = 7 × 7

Patch size = 9 × 9

Patch size = 11 × 11
More general features to discriminate dynamic textures and shapes

- Motion features
- True multi-scale criterion
- Scattering transform \cite{Bruna& Mallat 2013}

Generative vs. deterministic inpainting:

- Sample from a local conditional gaussian model instead of NN \cite{Raad et al. 2015}
- Accelerated learning/querying of local Gaussian models \cite{Guillemot et al. 2014

\begin{itemize}
  \item \textit{Bruna, J., \& Mallat, S. (2013). Invariant scattering convolution networks. IEEE TPAMI, 35(8), 187286}
  \item \textit{L. Raad, A. Desolneux, J-M. Morel (2014), Locally Gaussian Exemplar-Based Texture Synthesis}
  \item \textit{T. Guillemot, A. Almansa, T. Boubekeur, Covariance Trees for 2D and 3D processing, CVPR 2014.}
\end{itemize}
Thank you for listening!

More videos / paper / source-code:
http://perso.enst.fr/~almansa/video_inpainting/

Come see our poster on

Single-Shot High Dynamic Range Imaging!!

Open PhD Position

Subject  Video Inpainting
Location  Paris and/or Lyon
Supervisors  A. Almansa, Y. Gousseau, S. Masnou