### From Image to Video Inpainting with Patches



#### Patrick Pérez

JBMAI 2014 - LABRI



## Visual inpainting

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





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







4

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



Image extension







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
- 2004: Wexler et al. Video inpainting (2007, jal version)
- [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







#### Dictionary of block's neigh.

### Source Target















![](_page_10_Picture_1.jpeg)

![](_page_10_Picture_2.jpeg)

![](_page_11_Picture_1.jpeg)

![](_page_11_Picture_2.jpeg)

![](_page_12_Picture_1.jpeg)

![](_page_12_Picture_2.jpeg)

![](_page_13_Picture_1.jpeg)

![](_page_13_Picture_2.jpeg)

# 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

![](_page_14_Picture_5.jpeg)

## Which patche size?

- Depends on occlusion size (at coarsest level if multiscale)
- Depends on scale of visual "patterns" in image

![](_page_15_Picture_3.jpeg)

![](_page_15_Picture_4.jpeg)

technicolor

![](_page_15_Picture_6.jpeg)

# 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)

![](_page_16_Picture_8.jpeg)

technicolor

### Patch-based video inpainting

Wexler et al. (2004-2007)				
Iterative	Generic Automatic	<ul><li>Slow</li><li>Difficult to reproduce</li></ul>		
Patwardhan <i>et al.</i> (2005, 2007)				
Greedy	Reproducible Relatively fast	<ul><li>Requires segmentation</li><li>No cost optimization</li></ul>		
Granados et al. (2012)				
Graph-cut	Generic High resolution	<ul><li>Requires segmentation</li><li>Slow</li></ul>		

![](_page_17_Picture_2.jpeg)

## 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 inialization
  - Dominant motion compensation
  - Texture-sensitive patch comparison
  - Reproducible!

![](_page_18_Picture_12.jpeg)

![](_page_18_Picture_13.jpeg)

## 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 inialization
  - Dominant motion compensation
  - Texture-sensitive patch comparison
  - Reproducible!

![](_page_19_Picture_12.jpeg)

![](_page_19_Picture_13.jpeg)

### Notations

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

![](_page_20_Figure_6.jpeg)

![](_page_20_Picture_7.jpeg)

### 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

$$\phi(p) = \arg\min_{q \in \tilde{\mathcal{D}}} \|W_p - W_q\|_2^2 - p, \ \forall p \in \Omega \setminus \tilde{\mathcal{D}}$$

Image reconstruction over occlusion, given shift map

$$u(p) = \frac{\sum_{q \in \mathcal{N}_p} s_p^q u(p + \phi(q))}{\sum_{q \in \mathcal{N}_p} s_p^q}, \quad \forall p \in \mathcal{H}$$

Uniform weights, except for last reconstruction: copy from best match

technicolor

### 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

![](_page_22_Picture_5.jpeg)

![](_page_22_Picture_6.jpeg)

### 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

![](_page_23_Picture_5.jpeg)

![](_page_23_Picture_6.jpeg)

### Initialization

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

![](_page_24_Picture_2.jpeg)

![](_page_24_Picture_3.jpeg)

## Approximate patch matching

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

![](_page_25_Picture_10.jpeg)

![](_page_25_Picture_11.jpeg)

### First results

![](_page_26_Picture_1.jpeg)

![](_page_26_Picture_2.jpeg)

![](_page_26_Picture_3.jpeg)

![](_page_26_Picture_4.jpeg)

![](_page_26_Picture_5.jpeg)

![](_page_26_Picture_6.jpeg)

![](_page_26_Picture_7.jpeg)

![](_page_26_Picture_8.jpeg)

![](_page_26_Picture_9.jpeg)

technicolor

### First results

![](_page_27_Picture_1.jpeg)

![](_page_27_Picture_2.jpeg)

![](_page_27_Picture_3.jpeg)

### First results

![](_page_28_Picture_1.jpeg)

![](_page_28_Picture_2.jpeg)

![](_page_28_Picture_3.jpeg)

	One matching pass at full res.		
	Beach umbrella (265x68x200)	Crossing ladies (170x80x87)	Jumping girl (1120x754x200)
Wexler (kd-tree)	$\sim 1000s$	~1000s	$\sim 8000s$
Ours (PatchMatch3D)	$\sim 50s$	$\sim 30s$	~150s
	Total timing		
	Beach umbrella (265x68x200)	Duo (960x704x154)	Museum (1120x754x200)
Granados (graph-cut)	11h		90h
Ours w/o texture	14mn	4h	4h
Ours	24mn	6h	6h

![](_page_29_Picture_2.jpeg)

## Dealing with texture

SSD not sensitive enough to fine grain texture

![](_page_30_Picture_2.jpeg)

![](_page_30_Picture_3.jpeg)

## Dealing with texture

SSD not sensitive enough to fine grain texture

![](_page_31_Picture_2.jpeg)

![](_page_31_Picture_3.jpeg)

## Extended patch

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

$$T(p) = \frac{1}{\operatorname{card}(\nu)} \sum_{q \in \nu} (|I_x(q)|, |I_y(q)|)$$
$$W_p = [(u(q), \lambda T(q))]_{q \in \mathcal{N}_p}$$

Joint multi-scale pyramid

![](_page_32_Figure_4.jpeg)

![](_page_33_Picture_1.jpeg)

![](_page_33_Picture_2.jpeg)

![](_page_34_Picture_1.jpeg)

![](_page_34_Picture_2.jpeg)

![](_page_35_Picture_1.jpeg)

![](_page_35_Picture_2.jpeg)

![](_page_35_Picture_3.jpeg)

![](_page_35_Picture_4.jpeg)

![](_page_35_Picture_5.jpeg)

![](_page_36_Picture_1.jpeg)

![](_page_36_Picture_2.jpeg)

![](_page_37_Picture_1.jpeg)

![](_page_37_Picture_2.jpeg)

![](_page_38_Picture_1.jpeg)

![](_page_38_Picture_2.jpeg)

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

![](_page_39_Figure_3.jpeg)

- Simple solution
  - Estimate and compensate dominant motion wrt middle frame
  - Inpaint video
  - Put dominant motion back

![](_page_39_Picture_8.jpeg)

![](_page_40_Picture_1.jpeg)

![](_page_40_Picture_2.jpeg)

![](_page_41_Picture_1.jpeg)

![](_page_41_Picture_2.jpeg)

![](_page_42_Picture_1.jpeg)

![](_page_42_Picture_2.jpeg)

![](_page_42_Picture_3.jpeg)

![](_page_43_Picture_1.jpeg)

technicolor

![](_page_43_Picture_3.jpeg)

![](_page_44_Picture_1.jpeg)

technicolor

![](_page_44_Picture_3.jpeg)

![](_page_45_Picture_1.jpeg)

![](_page_45_Picture_2.jpeg)

![](_page_45_Picture_3.jpeg)

![](_page_46_Picture_1.jpeg)

![](_page_46_Picture_2.jpeg)

![](_page_47_Picture_1.jpeg)

![](_page_47_Picture_2.jpeg)

![](_page_48_Picture_1.jpeg)

![](_page_48_Picture_2.jpeg)

![](_page_49_Picture_1.jpeg)

![](_page_49_Picture_2.jpeg)

### 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

![](_page_50_Picture_11.jpeg)