Subiective continuity: combining perceptual and metric extrapolation

Shay Gepstien and Yosi Keller*

School of Engineering, Bar-Ilan University
Outline

1. Quick review of image inpainting
   - PDE based approach
   - Sampling based approaches
2. Quick review of spectral embedding
3. Image inpainting in the diffusion domain
Subjective continuity: combining perceptual and metric extrapolation

Image inpainting

M. Bertalmeo, G. Sapiro, V. Caselles and C. Ballester.
SIGGRAPH 2000
Heat equation I
Basic idea (over simplified)

\[ I^* = \min_I \int \int |\nabla I|^2 ds, \quad I(\Omega) = I_0 \]

by calculus of variations we get the heat equation

\[ \Delta I = 0, \quad I(\Omega) = I_0 \]

- Bertalmeo et al. used a non-linear heat equation
- Information is propagated outward-in
- Restores smooth image regions and cracks
- Has difficulties handling texture and large holes
Heat equation II
Basic idea (over simplified)

Extensions:

1. Anisotropic Diffusion
2. The U+V model: Bertalmio, Sapiro, Vese, Osher, TIP2003
Heat equation III
Basic idea (over simplified)

The good:
1. a clear objective function to optimize
2. assumes a simplistic manifold structure

The bad:
1. the numerical implementation can only recover smooth functions
2. The algorithm only learns from the boundary (implicit downside)
Subjective continuity: combining perceptual and metric extrapolation

Patch sampling I

Introduced in the seminal work of Efros and Leung CVPR1999
Patch sampling II

Basic idea (in signal processing terms)

The analogue of oversampling an analog signal in order to reconstruct it using a ZOH filter.

Extended by Criminisi, Perez, and Toyama CVPR2003
Patch sampling III

- An anisotropic formulation of Efros’ algorithm: propagates patches along edges
- Automatically maintain the structure during synthesis

Further extension by Michal Irani 2004: multiscale and iterations
When it is small, it induces a weighted majority vote, avoiding blurring in the resulting output. The use of a varying value has significantly improved the convergence of the algorithm. When compared with the original work in [29], the results shown here achieve better convergence. This can be seen from the improved amount of detail in areas which were previously smoothed unnecessarily. The ability to rely on outlier rejection also allows us to use approximate nearest neighbors, as discussed in Section 3.3. The algorithm is summarized in Fig. 5 both graphically and in pseudocode.

3.2 Multiscale Solution

To further enforce global consistency and to speed up convergence, we perform the iterative process in multiple scales using spatio-temporal pyramids. Each pyramid level contains half the resolution in the spatial and in the temporal dimensions. The optimization starts at the coarsest pyramid level and the solution is propagated to finer levels for further refinement. Fig. 6 shows a typical multiscale V-cycle performed by the algorithm. It is worth mentioning that, as each level contains 1/8th of the pixels, both in the hole and in the database, the computational cost of using a pyramid is almost negligible (8/7 of the work). In hard examples, it is sometimes necessary to repeat several such cycles, gradually reducing the pyramid height. This is inspired by multigrid methods [27].

The propagation of the solution from a coarse level to the one above it is done as follows: Let be a location in the finer level and let be its corresponding location in the coarser level. As before, let be the windows around and let be the matching windows in the database. We propagate the locations of onto the finer level to get . Some of these (about , half in each dimension) will overlap and these will be used for the maximum-likelihood step, just as before (except that here there are less windows). This method is better than plain interpolation as the initial guess for the next level will preserve high spatio-temporal frequencies and will not be blurred unnecessarily.

3.3 Algorithm Complexity

We now discuss the computational complexity of the suggested method. Assume we have pixels in the

The good:

1. Able to inpaint large and textured regions
2. Learns from the entire image, can also learn from multiple images
3. Very good results

The bad:

1. No objective function to optimize
Patch sampling V

- Computationally exhaustive
- Does not utilize the manifold structure
Kernel methods

Definition

Given a dataset \( \{ x_i \}_{i=1..n} \):

1. Apply a p.s.d. kernel \( k \) to \( \{ x_i \} \). For instance:
   \[
   w_{ij} = \exp\left(-\frac{d(x_i, x_j)}{\varepsilon}\right), \quad \varepsilon > 0.
   \]

2. Compute the eigenvectors of \( W \):
   \[
   w_{ij} = \sum_{l \geq 0} \lambda_l \psi_l(i) \psi_l(j),
   \]
   The embedding is given by

3. \( \Psi(x_i) : x_i \mapsto (\lambda_1 \psi_1(x_i), \lambda_2 \psi_2(x_i), \ldots) \)

4. \[
   \left\| \Psi_t(x_i) - \Psi_t(z_i) \right\|_{L_2}^2 = \sum_{l=0}^{n-1} \lambda_l^{2t} (\psi_l(x) - \psi_l(z))^2 = w_{ii} + w_{jj} - 2w_{ij} = D(x, z)^2
   \]
Different views of spectral embedding

• Spectral graph partitioning: Shi and Malik *Normalized Cuts and Image Segmentation*, 98

• Random Walk interpretation: Meila and Shi, *A random walks view of spectral segmentation*

• Preserving the infinitesimal geometry, Laplacian eigenmaps, Belkin-Nyogi, NIPS2002

• Diffusion distances, Coifman et. al. PNAS2004

\[ \| \Psi(x_i) - \Psi(z_i) \|_{L^2}^2 = \sum_{y \in \Omega} \frac{(p(x,y) - p(z,y))^2}{\phi_0(y)} \]

Is this is an interesting image?
Is this is an interesting image? No
Is it easy to inpaint?
Subjective continuity: combining perceptual and metric extrapolation

Signal processing in diffusion space
Joint work with Shay Gepstien

- Is this image interesting? No
- Is it easy to inpaint? Yes
- Is it easy to compress?
Signal processing in diffusion space
Joint work with Shay Gepstein

- Is this an interesting image? No
- Is it easy to inpaint? Yes
- Is it easy to compress? No

Looks (metrics) are deceiving!
Constructing our own metrics and embedding (and bases)

For an appropriate metric the embeddings are smooth

LBP texture features + Heat equation
Is the embedding reversible?

The short answer is: no

The embedding is a non-linear operation that is non-invertible.

The embedding captures the intrinsic structure of a signal and not its ambient space manifestation.
Is the embedding reversible?

The short answer is: no

- The embedding is a non-linear operation that is non-invertible.
- The embedding captures the intrinsic structure of a signal and not its ambient space manifestation.
The long answer is: we can try

- For each *interpolated* point \( \{\phi_i(j)\}_{i=1}^{N} \) we can pick the \( K - NN \) points in the learning set.
- We can impose (constrained) spatial smoothness
  - PDE - smoothness over greyscale values
  - Patches - smoothness over *all* patches.
We define the pairwise distance $d(x_i, x_b)$ using the $L_2$ norm

$$W(x_i, x_j) = \exp\left(-\frac{d^2(x_i, x_j)}{\sigma^2}\right)$$

and maximize

$$x^* = \max \arg \sum_{i,j} W(x_i, x_j)$$

Similar to: S. Avidan et al. : Seam Carving, Image Retargeting

Main difference: we only consider $K$ patches per point
Initial results

Source image with hole

patch center

Source

Result

19 / 22
Comparison to previous works

The good:  
1. a globally optimal solution in high dimensional space  
2. the diffusion is applied to low dimensional smooth data  
3. can be applied to different data sources  

The bad: there are a few engineering issues to resolve
Conclusions

- Global optimality goes hand in hand with smoothness
- Image patches can be used as building blocks for image inpainting
- Image patches are part of a manifold: a fact not exploited by current schemes
- Image inpainting is an example of a class of general problems: how to utilize manifolds in signal processing
- Future work: can we apply other signal processing operations and “invert” the embedding?
  - Filtering
  - Compression
- Can we apply it to signals other than images (speech)?
Thanks You!