Single Image Super-resolution
Types of Super-resolution

Single-image (Hallucination)
Super-resolution Goals

• 1) Produce a detailed, realistic output image.
• 2) Be faithful to the low resolution input image.
Bicubic Upsampling

1) Produce a detailed, realistic output image.
2) Be faithful to the low resolution input image.
Best Scene Match

1) Produce a detailed, realistic output image.
2) Be faithful to the low-resolution input image.
Typical Super-resolution Method

1) Build some statistical model of the visual world.

2) Coerce an upsampled image to obey those statistics.

Methods can be divided based on the statistical model – either parametric or non-parametric (data-driven).
Bicubic Upsampling
Example-Based Super-Resolution. Freeman, Jones, and Pasztor. 2000

Input patch

Closest image patches from database

Corresponding high-resolution patches from database
Example-Based Super-Resolution.
Freeman, Jones, and Pasztor. 2000

Bicubic

Super-resolution
Super-resolution from Internet-scale Scene Matching

Libin (Geoffrey) Sun, James Hays
Brown University
Problem Statement
single image super-resolution

\[ y = D(x \otimes k) + n \]

We want:
- more pixels
- sharp edges
- correct textures
...

85 x 128 \rightarrow \text{Extremely ill-posed} \rightarrow 680 x 1024
Why is it hard?

- mathematically ill-posed
- vision-hard
Previous Work

- Freeman et al., 2000
- Baker & Kanade, 2002
- Sun et al., 2003
- Yang et al., 2008
- Glasner et al., 2009
- Sun & Tappen, 2010
- HaCohen et al., 2010

Bicubic Interpolation vs. Ideal Super-resolution Algorithm

Realism of Output
- Fewer Artifacts
- More Artifacts

Details Synthesized by Algorithm
- edges
- textures
- objects

Our Method

- Compact Parametric Models
  - Tappen et al., 2003
  - Fattal, 2007
  - Sun et al., 2008

- Data-driven Patch-based Methods
  - Freeman et al., 2000
  - Baker & Kanade, 2002
  - Sun et al., 2003
  - Yang et al., 2008
  - Glasner et al., 2009
  - Sun & Tappen, 2010
  - HaCohen et al., 2010
HaCohen et al, ICCP 2010

texture database with 13 categories, 106 images
- material/texture recognition is hard
- requires human intervention
- edge handling
- limited categories
4000 natural images, 160,000 low/high segment pairs
- hard to establish 'correct' segment correspondences
Self-similarity Based Methods

[Glasner et al, 2009]
[Freedman & Fattal, 2010]
Overview and Contributions

- The first to use scene matches for SR, at extremely low-res
- Scene match statistics favored over internal statistics
- Competitive results, insertion of details, texture transitions
Scale of Training Data

- Freeman et al: 6 images
- Sun et al: 16 images
- Yang et al: 30 images
- Sun & Tappen: 4000 images
- HaCohen et al: 106 images
- Ours: 6.3 Million images
Scene Matching: Image-level Context

- Input
- Scene matching from large database
- Segment level correspondence
- Patch-based detail synthesis
- Output

Optimization
Scene Matching

**Image restoration/inpainting**
[Hays & Efros, SIGGRAPH 2007]
[Dale *et al*, ICCV 2009]
[Johnson *et al*, TVGC 2010]

**Geolocation**
[Hays & Efros, CVPR 2008]

**Image similarity**
[Shrivastava *et al*, SIGGRAPH ASIA 2011]

**Object recognition**
[Russell *et al*, NIPS 2007]
[Torralba *et al*, CVPR 2008]

**Image-based rendering**
[Sivic *et al*, CVPR 2008]

**Scene parsing**
[Liu *et al*, CVPR 2009]

**Event prediction**
[Yuen & Torralba, ECCV 2010]
Scene Matching

Features:
- Gist
- color/texton
- histogram
- sparse BoW
- geometric
- context

6.3 million images

Example Scene Matches

Input (low-res)
How Useful are the Scene Matches?

Expressiveness and Predictive Power [Zontak & Irani 2011]

1. Internal Database (all scales)

Input image (ground truth)

2. Internal Database (limited)

3. External Database [Zontak & Irani 2011]

BSD training set

4. External Database [Ours]

Scene Matches
Expressiveness

`high-res (ground truth)`

How close is the nearest neighbor?

Database

High-res patches

Expressiveness (5x5 patches, DC removed, grayscale)

- red: internal(all scales)
- red: internal(limited)
- blue: 20 BSD img
- black: 20 scene matches

RMSE per Patch

Mean Gradient Magnitude per Patch
**Predictive Error**

**low-res (observed)**

**error in estimated HR patch?**

**Database**

**low/high patch pairs**

Retrieve kNN patches + Estimate high-res

**high-res (ground truth)**

Prediction Error (5x5 patches, DC removed, grayscale)

- **internal(all scales)**
- **internal(limited)**
- **20 BSD img**
- **20 scene matches**
Segmentation: Region-level Context
Segmentation: Region-level Context

- 1000 textons learned per image/scene
- Color histograms
Optimization Framework

input → scene matching from large database → segment level correspondence → patch-based detail synthesis → optimization → output
Optimization Framework

Greedy selection of pixel candidates

\[ E(I^h) = E(I^h | I^l) + \beta_1 E_h(I^h) + \beta_2 E_e(I^h) \]

[Sun & Tappen 2010]
Bicubic ×8

Ours

Sun & Tappen, CVPR 2010

Glasner et al, ICCV 2009
Bicubic $\times 8$

Ours

Sun & Tappen, CVPR 2010

Glasner et al, ICCV 2009
Failure Modes: Bad Scene Match

Input image

Top Scene Matches
Failure Modes: Bad Texture Transfer

Input image

Top Scene Matches
Evaluation

Perceptual Studies, similar to [Liu et al, 2009]
- 20 test scenes
- Binary comparison: 'higher quality'
- 22 participants

User preference for super-resolution results

Proportion of votes received

<table>
<thead>
<tr>
<th></th>
<th>Our algorithm</th>
<th>Tie</th>
<th>Sun and Tappen [31]</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Good&quot; scene matches</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Bad&quot; scene matches</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

- The first to use scene matches for SR, at extremely low-res

- Scene match statistics favored over internal statistics

- Competitive results, insertion of details, texture transitions
Thank you! And Questions?
Expressiveness 5x5 vs 9x9

Expressiveness (5x5 patches, DC removed, grayscale)

Expressiveness (9x9 patches, DC removed, grayscale)

RMSE per Patch

Mean Gradient Magnitude per Patch
Predictive Error 5x5 vs 9x9

Prediction Error (5x5 patches, DC removed, grayscale)

Prediction Error (9x9 patches, DC removed, grayscale)
Predictive Uncertainty 5x5 vs 9x9
Optimization Framework

[Sun 2010]

\[ E(I^h) = E(I^h|I^l) + \beta_1 E_h(I^h) + \beta_2 E_e(I^h) \]

\[ E(I^h|I^l) = |D(k * I^h) - I^l|^2 \]

\[ E_h(I^h) = -\frac{1}{\lambda} \sum_{p \in \text{pixels}} \log \left( \sum_{i \in \text{candidates}} \exp(-\lambda d_i(p)) \right) \]

\[ E_e(I^h) = -\sum_{p \in \text{pixels}} p_b(p) \sum_k \alpha_k \log \left( 1 + \frac{1}{2}(f_k * I^h(p))^2 \right) \]