SCENE COMPLETION USING MILLIONS OF PHOTOGRAPHS

James Hays
Alexei A. Efros
Carnegie Mellon University

Presented by
Harika Sabbella

IMAGE COMPLETION

PREVIOUS APPROACHES

Bertalmio, Sapiro, Caselles, and Ballester.
Image Inpainting. SIGGRAPH 2000.

Efros and Leung. Texture synthesis by non-parametric sampling. ICCV 1999.

Criminisi, Perez, and Toyama. Region filling and object removal by exemplar-based inpainting. IEEE Transactions on Image Processing. 2004.

PREVIOUS APPROACHES (CONTINUED)

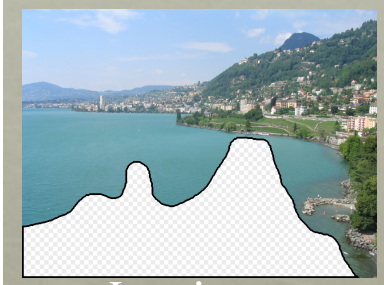
NEW APPROACH

- Two steps:
  - Use content from other images to first match the source image semantically
  - Use local context matching to fill in the hole as precisely as possible

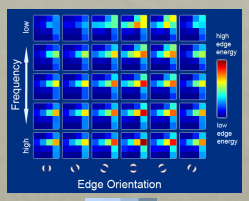
CHALLENGES

- Three main challenges:
  - Computational
  - Difficult to find semantically similar images
  - Data from other images won't contain correct color and illumination

ALGORITHM OVERVIEW



Input image



Scene Descriptor

Image Collection

200 matches

Hays and Efros, SIGGRAPH 2007

Context matching
+ blending

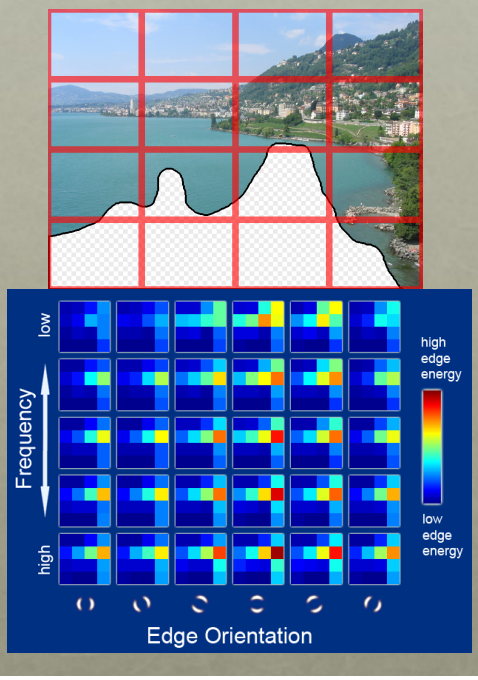
20 completions

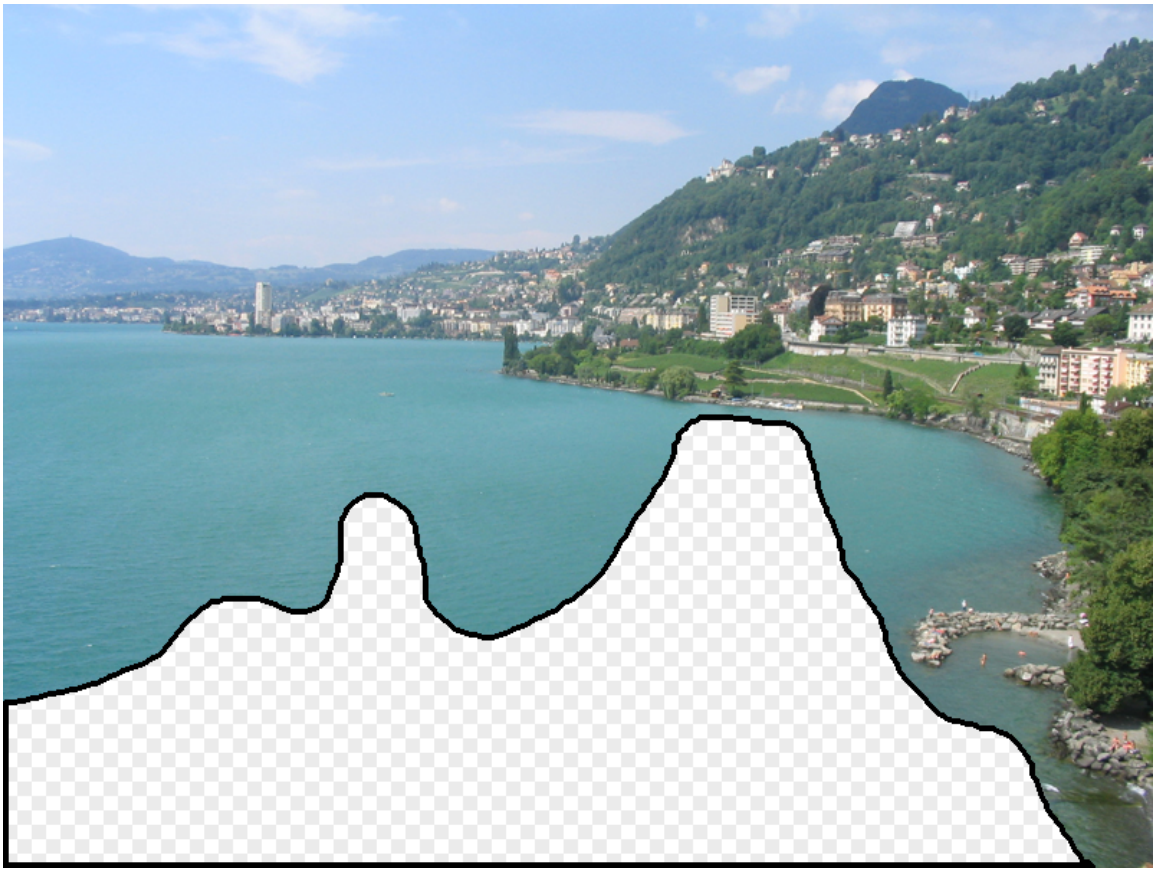
DATA COLLECTION

- 2.3 million images collected and used as input
- Downloading, pre-possessing, and scene matching done using cluster of 15 machines

GIST SCENE DESCRIPTOR

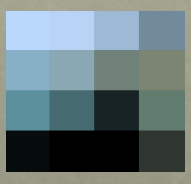
- Low dimensional scene descriptor
- Gist descriptor built from 6 oriented edge responses at 5 scales aggregated to 4x4 spatial resolution



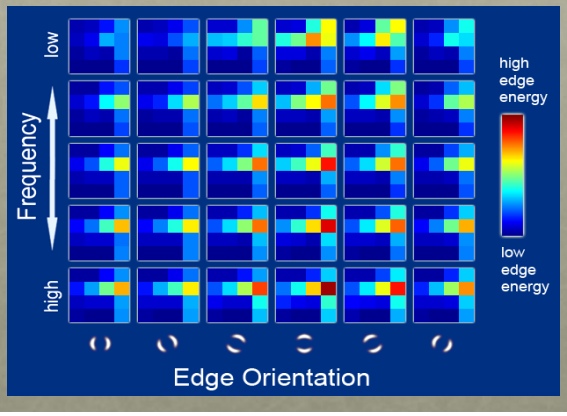


SCENE MATCHING

- Step 1: Create a mask
- Step 2: Compute *gist scores* between gist of the source image (weighted by mask) and 2.3 million images
- Step 3: Compute *color difference scores* between source image and 2.3 million images
- Step 4: Compute final *scene matching distance scores*



+



LOCAL CONTEXT MATCHING

- Minimization of three quantities:
  - Local context matching distance/pixel-wise alignment score
  - Local texture similarity distance
  - Cost of graph cut

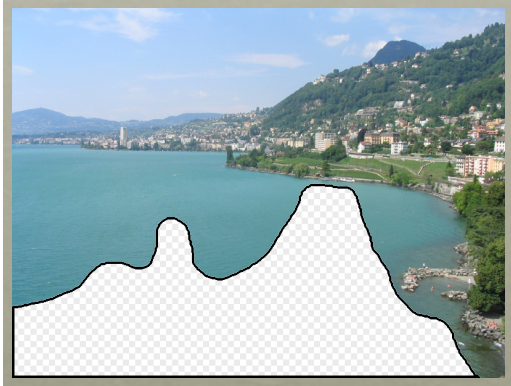
PIXEL-WISE ALIGNMENT SCORE

- Multiply error at each placement by magnitude of translational offset
- Translation and scale with minimum weighted SSD error chosen as best placement for each matching scene

LOCAL TEXTURE SIMILARITY DISTANCE

- Simple texture descriptor
  - 5x5 median filter of image gradient magnitude at each pixel
- Compare SSD of texture descriptors of source image and proposed fill-in region

CONTEXT MATCHING



Hays and Efros, SIGGRAPH 2007

GRAPH CUT SEAM FINDING

- Cost: $C(L) = \sum_{p} C_d(p, L(p)) + \sum_{p,q} C_i(p, q, L(p), L(q))$

  $C_d(p, patch)$

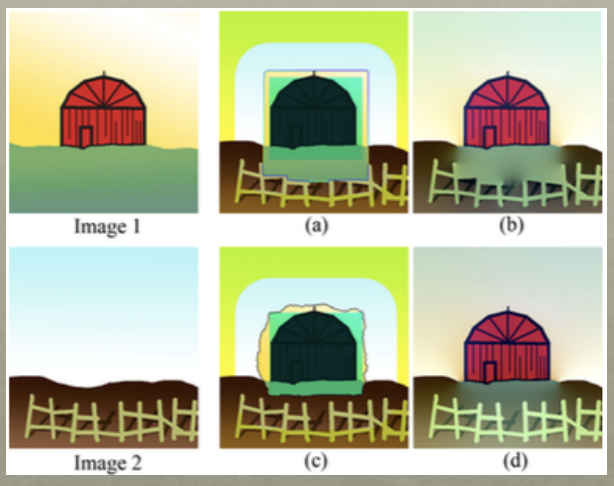
  $C_d(p, exist)$

- For all other pixels (k=0.002)
  - Removing each pixel from the source image increases with distance from the hole $C_d(p, patch) = (k * Dist(p, hole))^3$

GRAPH CUT SEAM FINDING (CONTINUED)

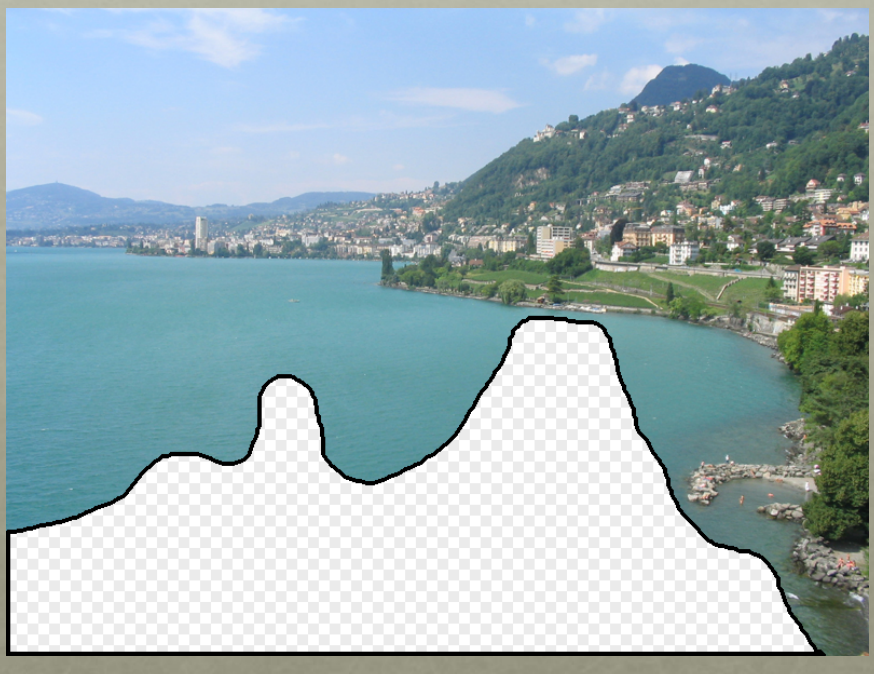
- If L(p) = L(q), cost = 0
- If L(p) != L(q),

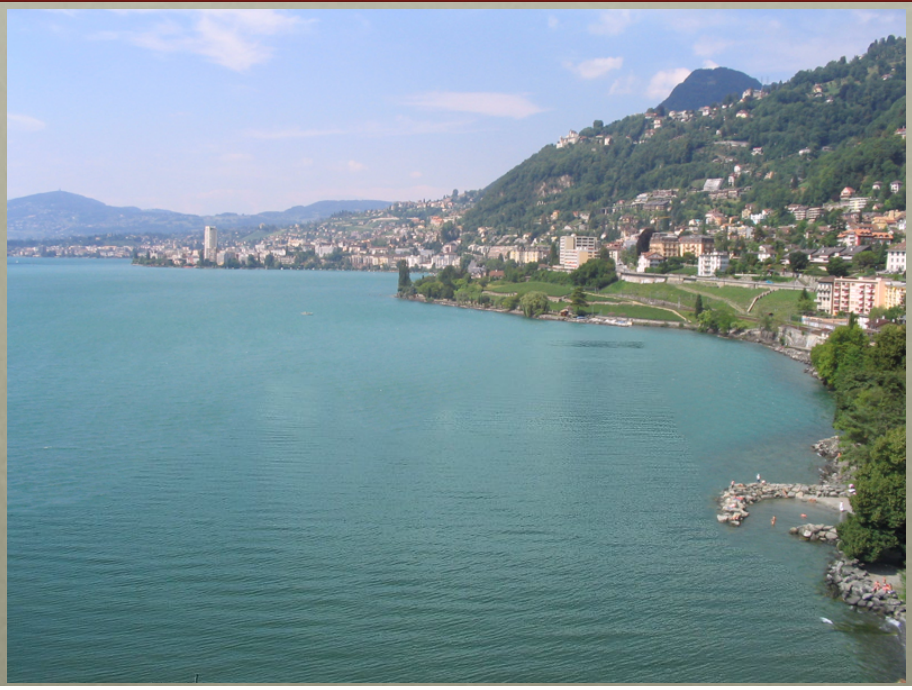
  $C_i(p, q, L(p), L(q)) = \nabla diff(p, q)$

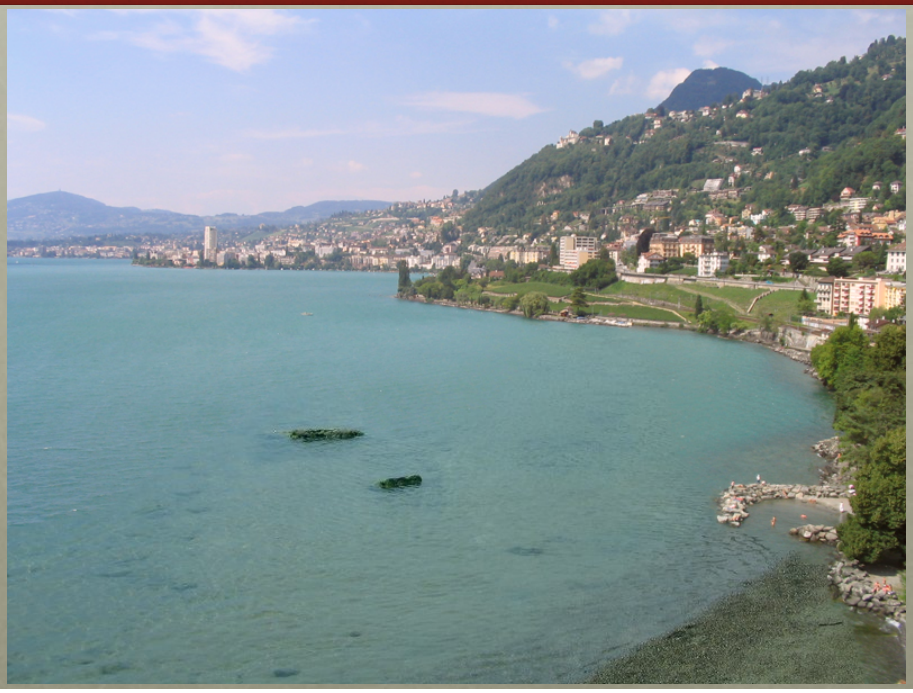


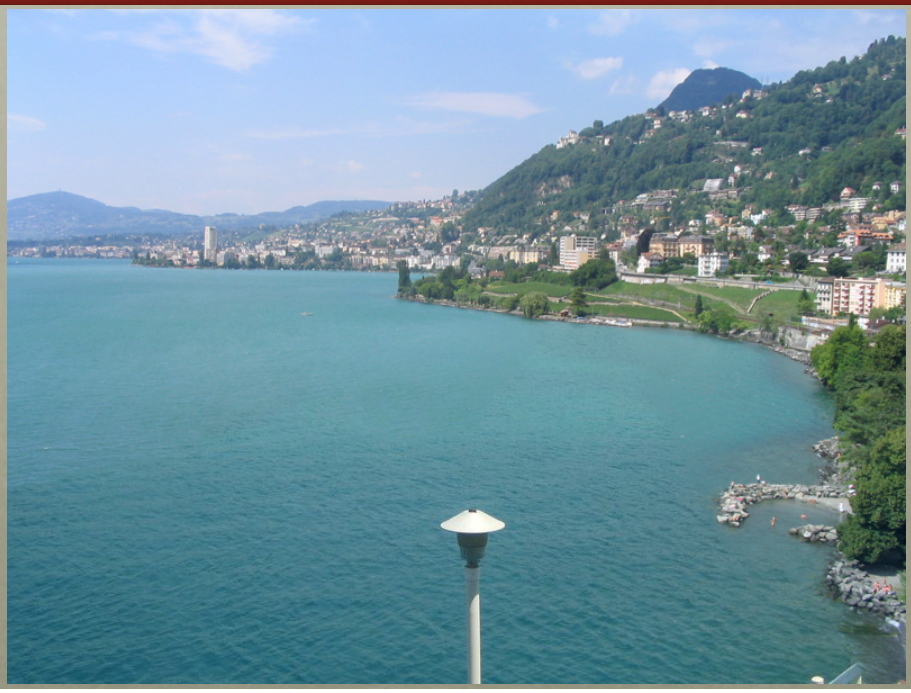
FINAL OUTPUT

Return 20 composites with lowest final scores!

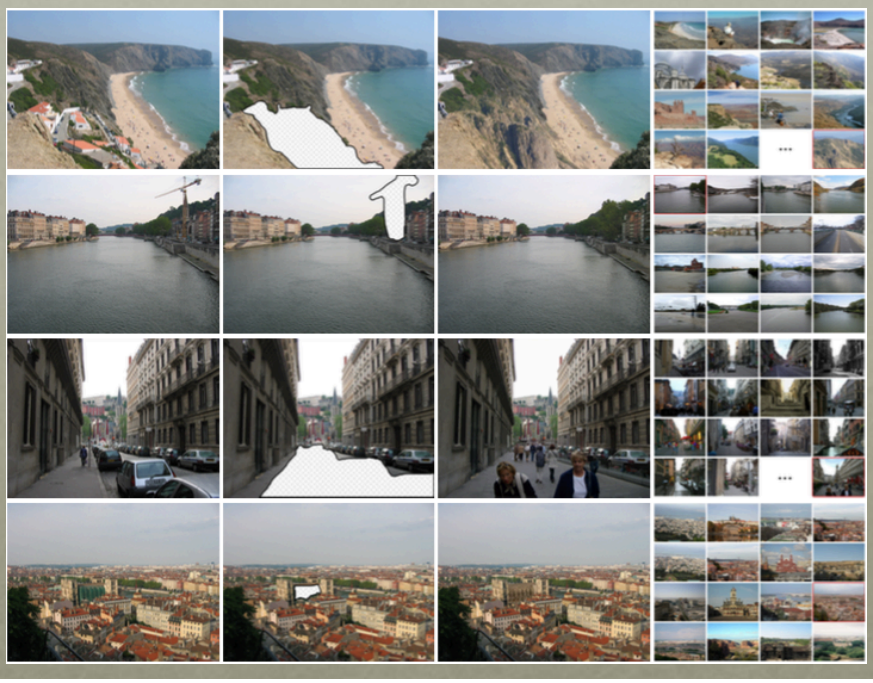








TRIAL RUNS

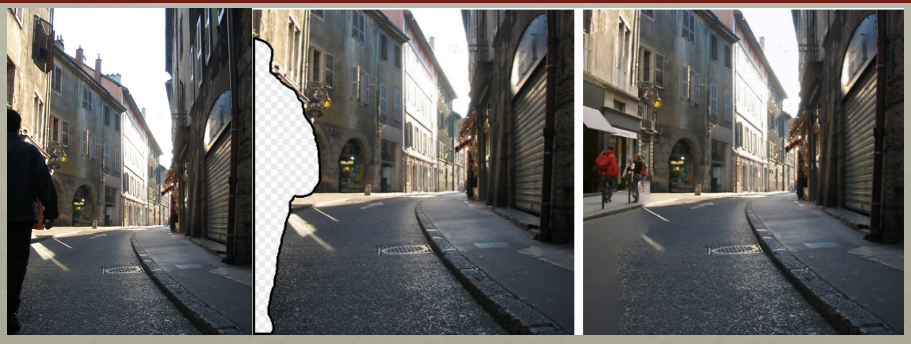


TRIAL RUNS (CONTINUED)

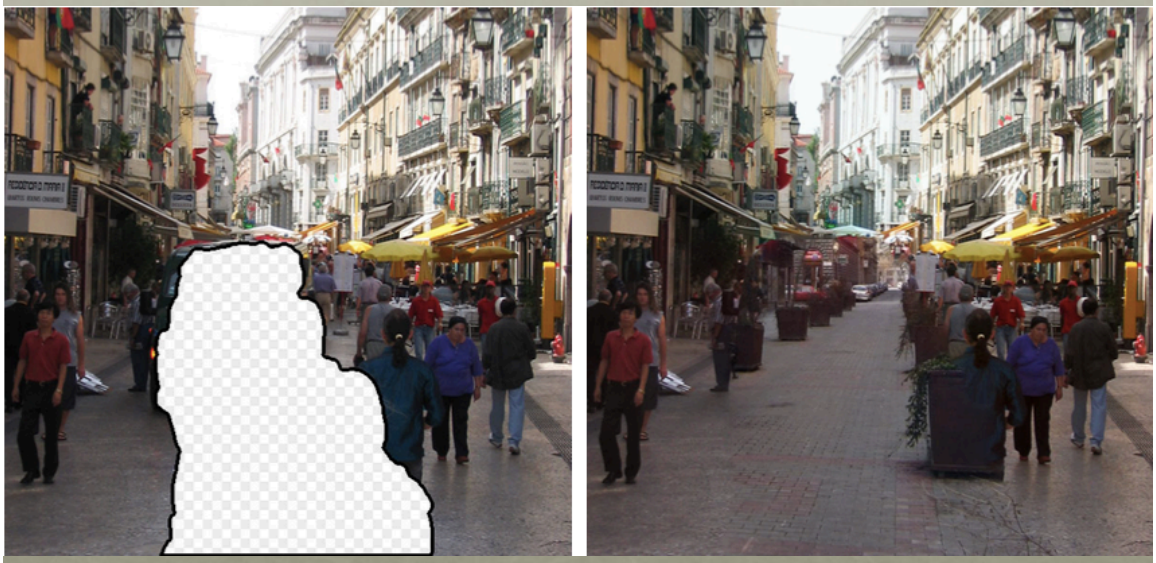
TRIAL RUNS (CONTINUED)

TRIAL RUNS (CONTINUED)

TRIAL RUNS (CONTINUED)



TRIAL RUNS (CONTINUED)



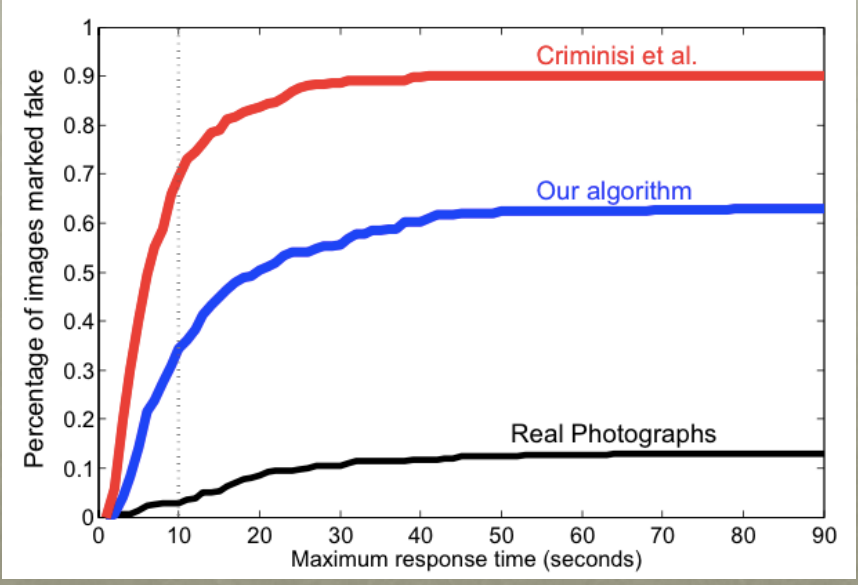
EXPERIMENT

- Study with 51 test images that span different types of collections involving 20 participants
- Generated three versions of each image
  - Original image
  - Result from Criminisi et al.
  - Result from scene matching algorithm

EXPERIMENT RESULTS

- Real images were identified as being real only 87% of the time!!
- Participants classified output of scene completion algorithm as being real 37% of the time vs. 10% of the time for Criminisi et al. method

EXPERIMENT RESULTS (CONTINUED)



DISCUSSION

- Entirely new approach using a large collection of images
- Using more images could improve scene completions
- Hybrid of both single-image techniques and scene matching algorithm likely to be best method for image filling tasks
- Can we actually collect the set of all semantically differentiable scenes?
  - Result of this work shows that this is entirely possible and the number of required images might not be that large!

QUESTIONS