

CS4670 / 5670: Computer Vision

Noah Snavely

Modern Object Recognition: The PASCAL Challenge



Visual Object Classes Challenge 2009 (VOC2009)



[click on an image to see the annotation]

More videos involving head-tracking

<http://www.youtube.com/watch?v=ydIPKkjBIMw>

<http://www.youtube.com/watch?v=Jd3-eiid-Uw>

The PASCAL Visual Object Classes Challenge 2009 (VOC2009)

- Twenty object categories (aeroplane to TV/monitor)
- Three challenges:
 - Classification challenge (is there an X in this image?)
 - Detection challenge (draw a box around every X)
 - Segmentation challenge



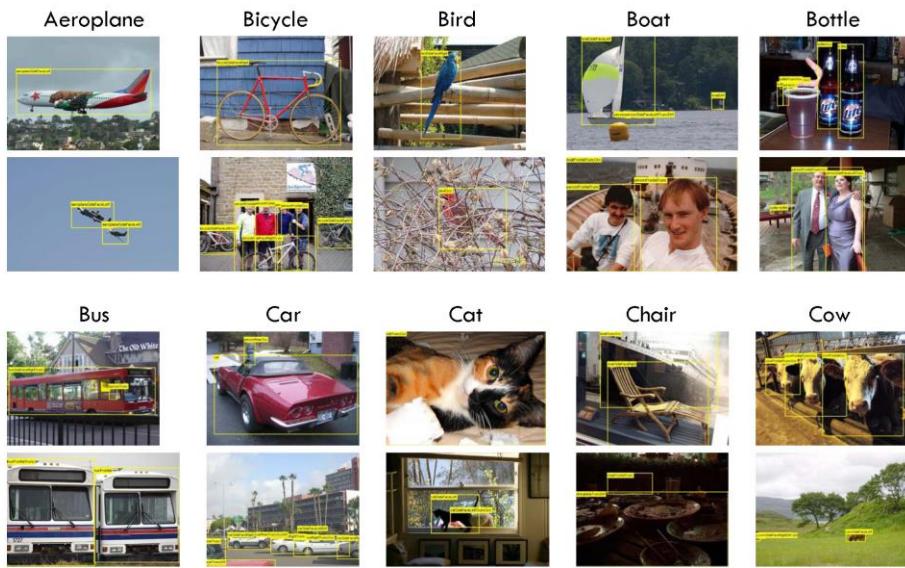
Dataset: Collection

- Images downloaded from **flickr**
 - 500,000 images downloaded and random subset selected for annotation

Dataset: Annotation

- Complete annotation of all objects
- Annotated over web with written guidelines
 - High quality (?)

Examples



Examples

Dining Table



Dog



Horse



Motorbike



Person



Potted Plant



Sheep



Sofa



Train



TV/Monitor



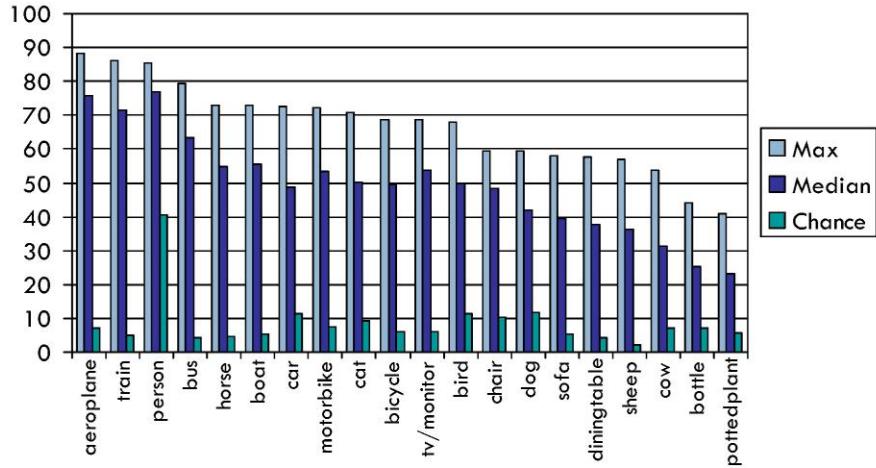
Classification Challenge

- Predict whether at least one object of a given class is present in an image



is there a cat?

AP by Class



- Max AP: 88.1% (aeroplane) ... 40.8% (potted plant)

Precision / Recall for a Category X

- Precision:

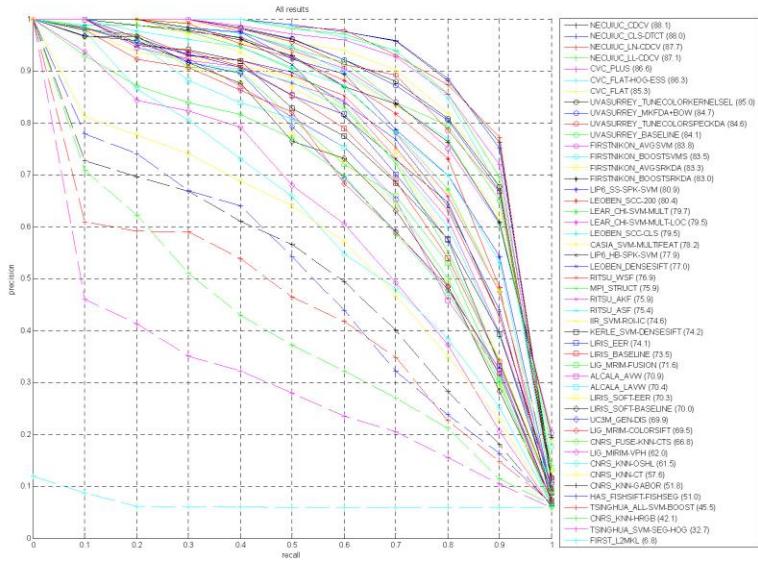
$$\frac{|\{\text{images that contain an } X\} \cap |\{\text{images classified as } X\}|}{|\{\text{images classified as } X\}|}$$

- Recall:

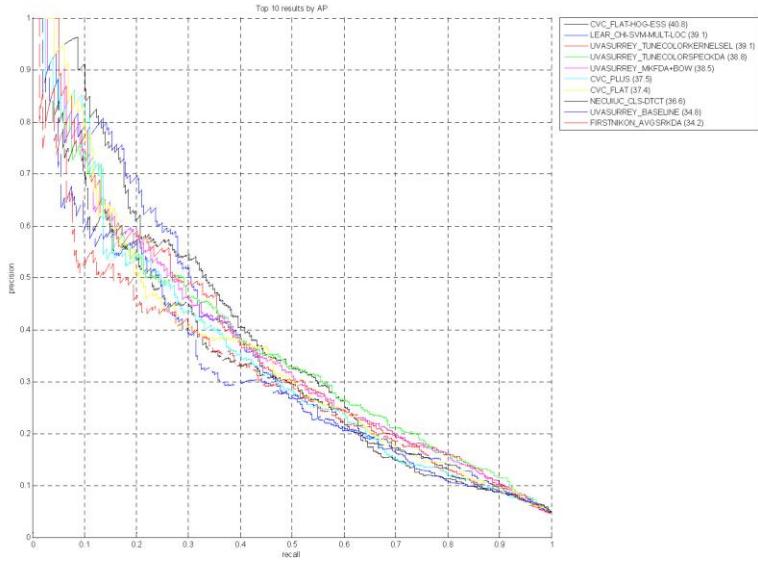
$$\frac{|\{\text{images that contain an } X\} \cap |\{\text{images classified as } X\}|}{|\{\text{images that contain an } X\}|}$$

- In reality, methods give a continuous-valued score for each image / category → PR curve

Precision/Recall: Aeroplane (All)

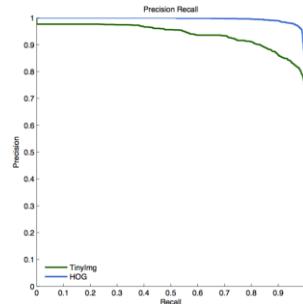


Precision/Recall: Potted plant (Top 10 by AP)



Precision Recall curves

- Related to but different from ROC curves
- Start at (0, 1), higher curves are better



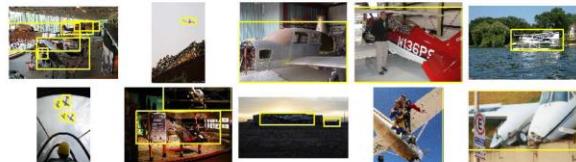
- Average Precision (AP) = area under the curve

Ranked Images: Aeroplane

- Class images:
Highest ranked



- Class images:
Lowest ranked



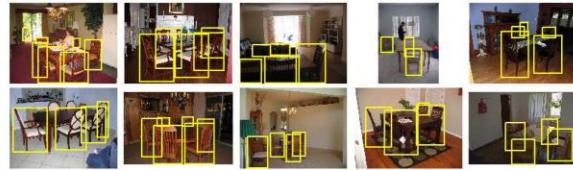
- Non-class images:
Highest ranked



- Context?

Ranked Images: Chair

- Class images:
Highest ranked



- Class images:
Lowest ranked



- Non-class images:
Highest ranked



- Scene context?



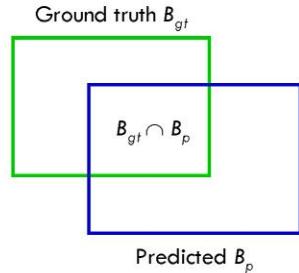
Detection Challenge

- Predict the bounding boxes of all objects of a given class in an image (if any)



Evaluating Bounding Boxes

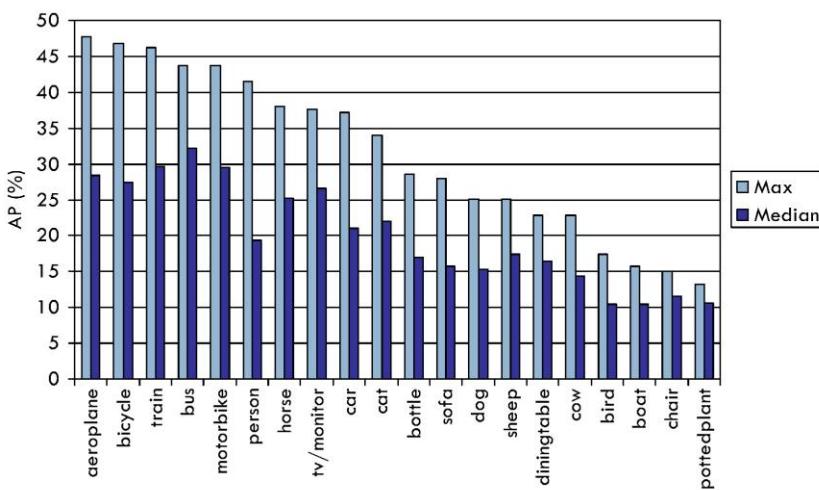
- Area of Overlap (AO) Measure



$$AO(B_{gt}, B_p) = \frac{|B_{gt} \cap B_p|}{|B_{gt} \cup B_p|}$$

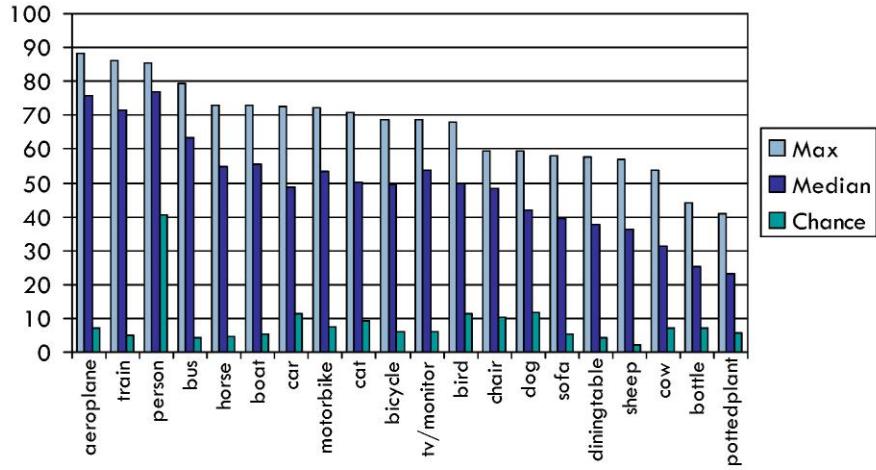
- Need to define a threshold t such that $AO(B_{gt}, B_p)$ implies a correct detection: 50%

AP by Class



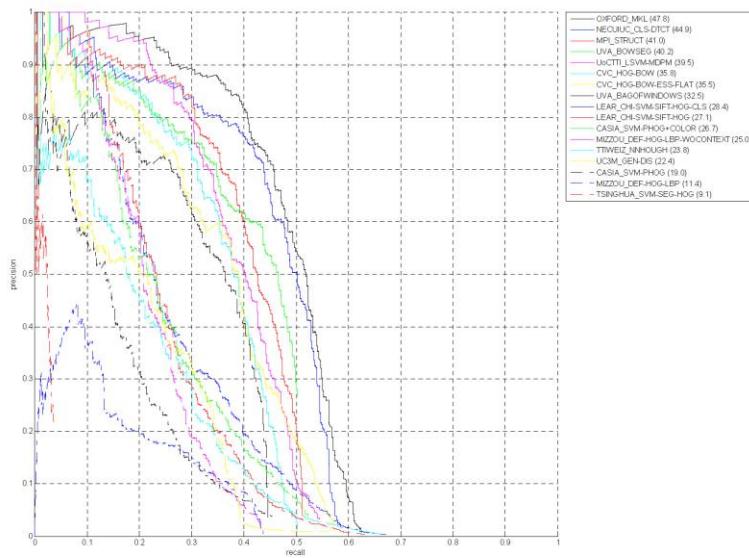
Chance essentially 0

AP by Class Classification Challenge Results (for comparison)

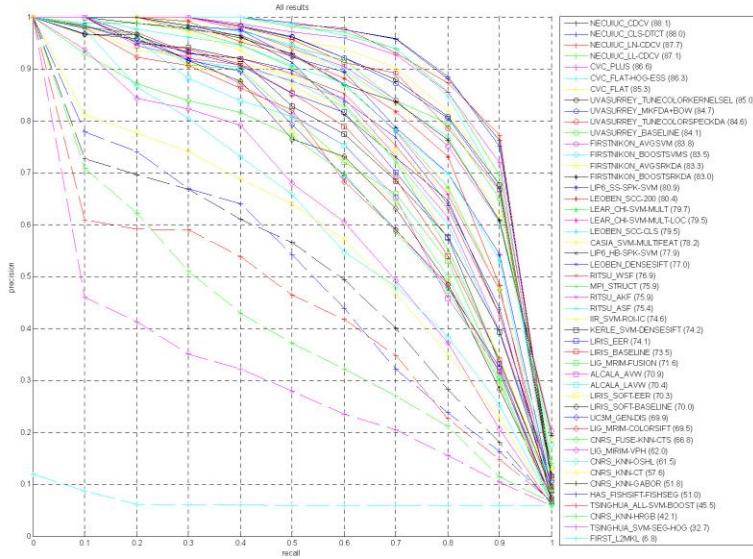


- Max AP: 88.1% (aeroplane) ... 40.8% (potted plant)

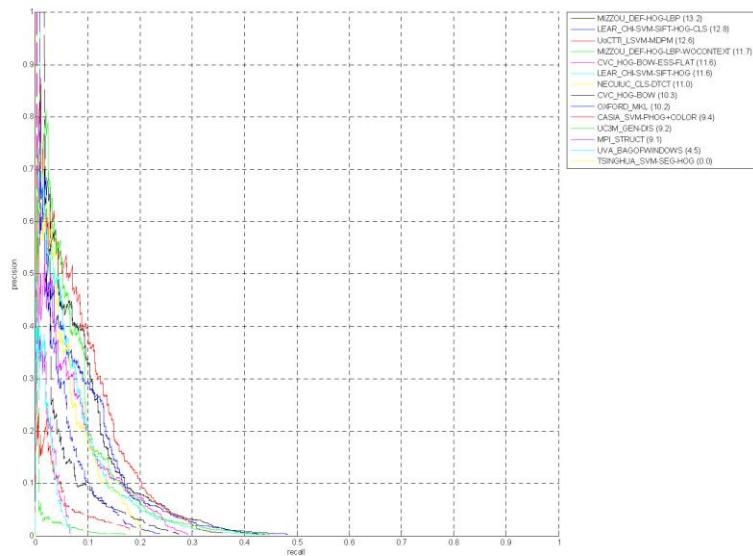
Precision/Recall - Aeroplane



Precision/Recall: Aeroplane (All) (Classification)



Precision/Recall – Potted plant



True Positives - Person

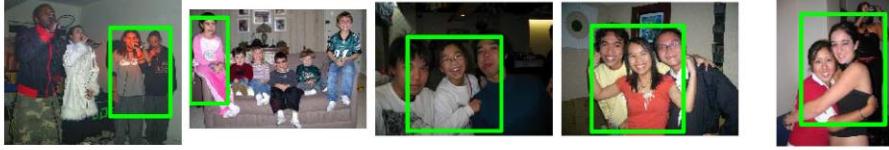
UoCTTI_L SVM-MDPM



MIZZOU_DEF-HOG-LBP



NECUIUC_CLS-DTCT

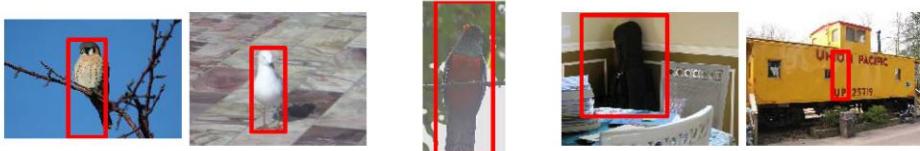


False Positives - Person

UoCTTI_L SVM-MDPM



MIZZOU_DEF-HOG-LBP



NECUIUC_CLS-DTCT



“Near Misses” - Person

UoCTTI_LSVM-MDPM



MIZZOU_DEF-HOG-LBP



NECUIUC_CLS-DTCT



True Positives - Bicycle

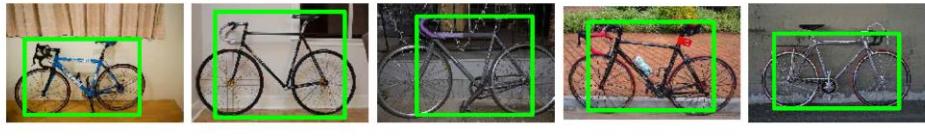
UoCTTI_LSVM-MDPM



OXFORD_MKL



NECUIUC_CLS-DTCT



False Positives - Bicycle

UoCTTI_LSVM-MDPM



OXFORD_MKL



NECUIUC_CLS-DTCT



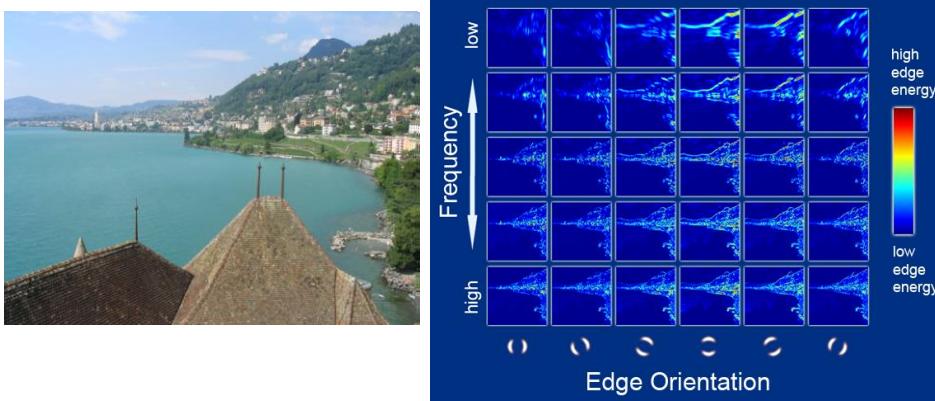
Questions?

Best localization methods

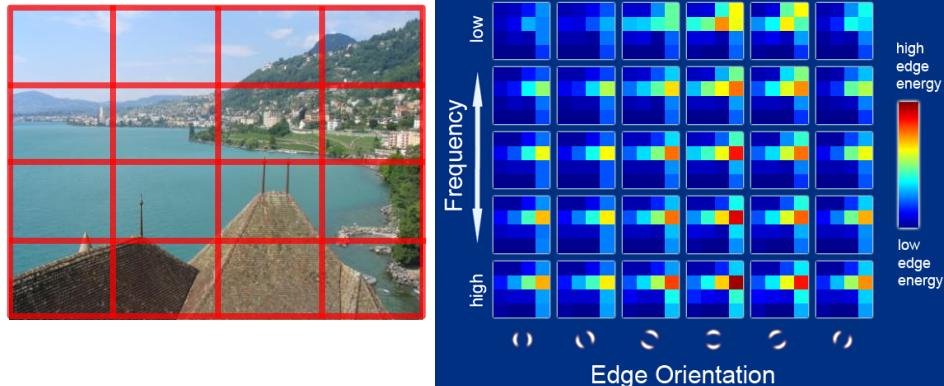
- Sliding window-style classifiers
 - SVM, Adaboost
 - Flexible spatial template: “star model”
 - Deep learning?
- Separate classifiers by viewpoint
- Use of context in classifiers
- Local features
 - HoG, SIFT, local histograms of gradient orientations

GIST descriptors

- Global descriptor computed from entire image [Oliva and Torralba 2001]

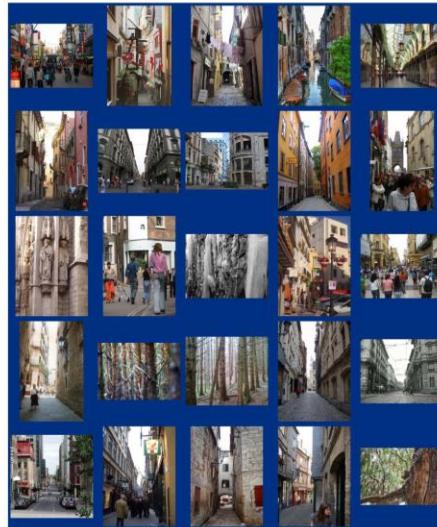


GIST Descriptor



Gist for Scene Matching: How Good?

- Closest outdoor scenes by gist
 - 2.1M images from online travel photo groups (no labels)

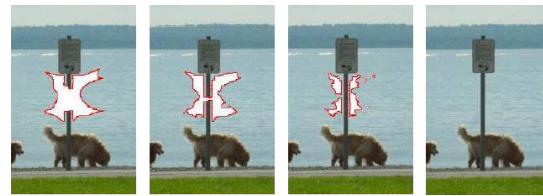




Hays and Efros, SIGGRAPH 2007



Hays and Efros, SIGGRAPH 2007



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

[Hays and Efros, SIGGRAPH 2007](#)



Criminisi et al. result

[Hays and Efros, SIGGRAPH 2007](#)

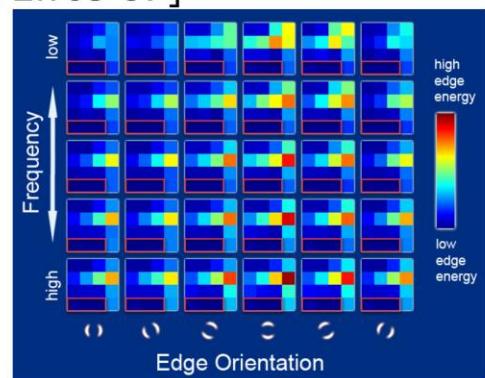
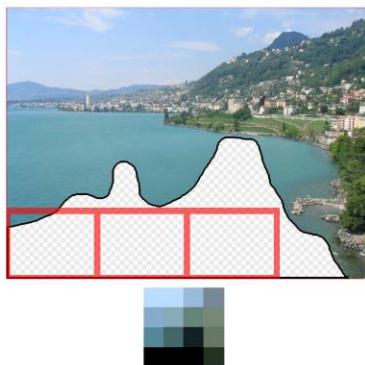


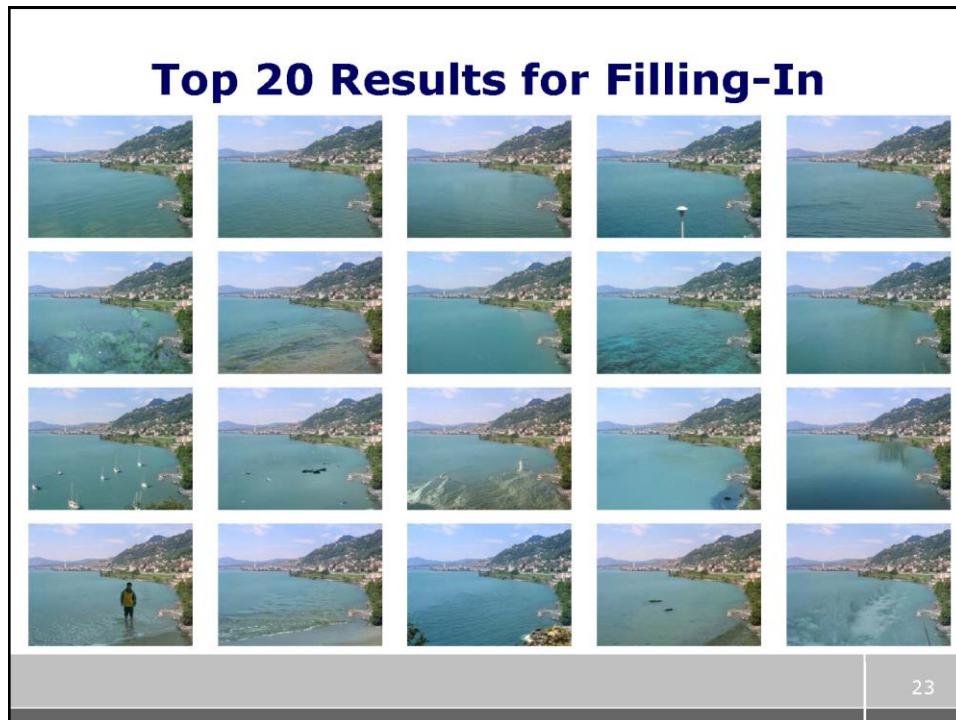
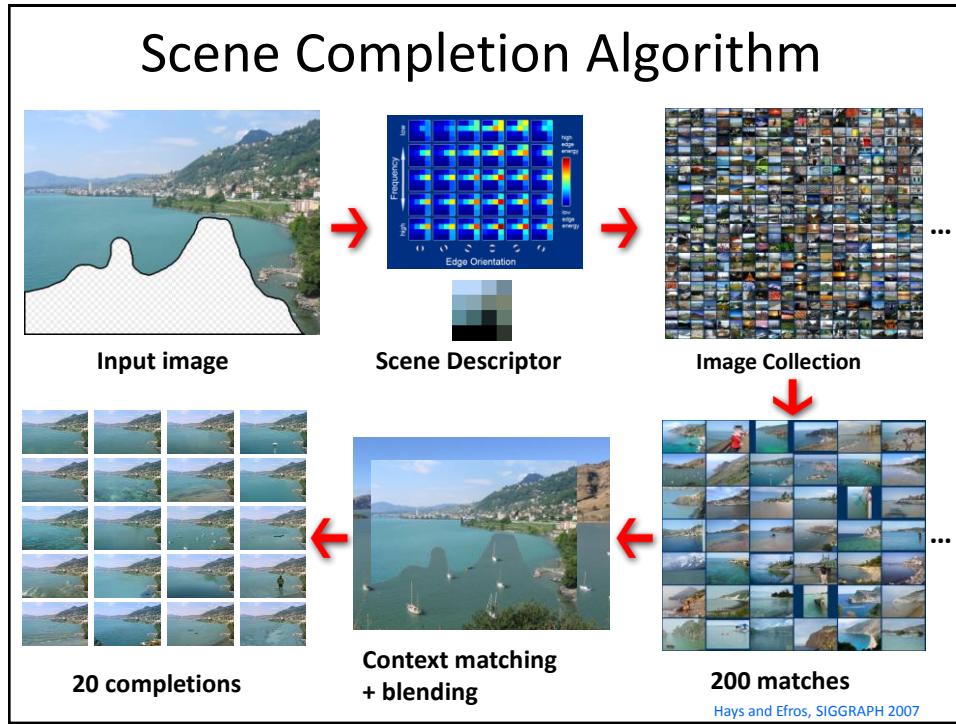
Criminisi et al. result

Hays and Efros, SIGGRAPH 2007

Collection Matching for Filling-In

- Use gist plus color to select closest images from large set (size important) then find best blend [Hays & Efros 07]







Hays and Efros, SIGGRAPH 2007



Hays and Efros, SIGGRAPH 2007



[Hays and Efros, SIGGRAPH 2007](#)



[Hays and Efros, SIGGRAPH 2007](#)



Hays and Efros, SIGGRAPH 2007

Better than Local Filling-In Methods

- Judgments of real vs. fake (20 subjects)

