Recent Progress in Recognizing and Organizing Images

Daniel Huttenlocher
John P. and Rilla Neafsey Professor
Cornell University

Overview

- State of the art in object recognition
  - Categories rather than specific objects
  - Shared datasets and evaluations
  - Image classification
  - Object localization
- Recent developments in large image collections
  - Image gist and scene-level matching
  - 3D organization of site-specific photos

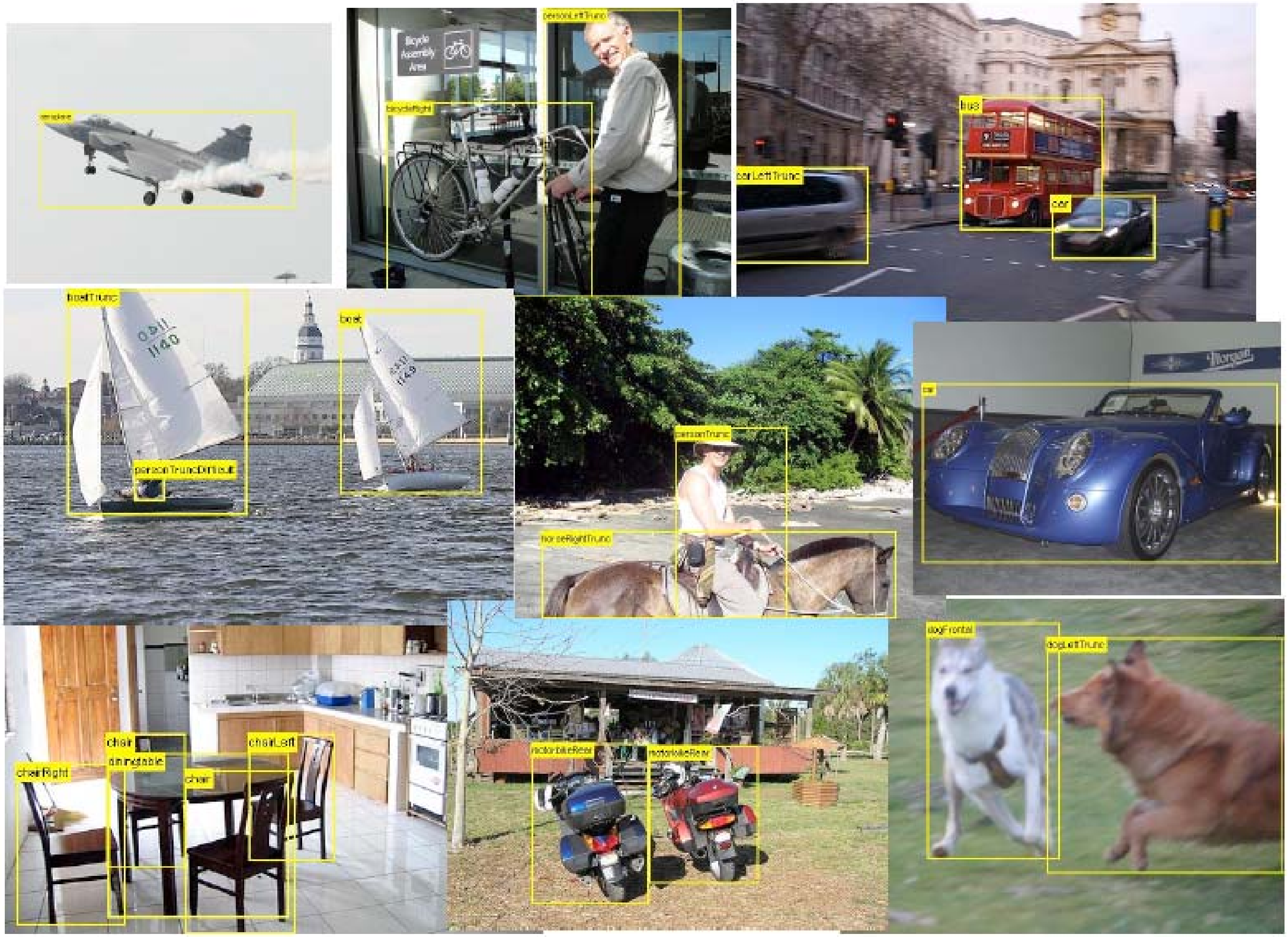
Object Recognition

- Past decade of research has focused largely on recognizing object categories
  - E.g., car, bus, motorcycle as opposed to earlier work on specific car or specific model of car
- Extensive use and development of machine learning techniques
- Moderate-scale datasets largely derived from Web photo sharing sites
  - E.g., PASCAL VOC: 20 categories, 10K images, 25K instances, hand-labeled ground truth, annual competition

Object Recognition Tasks

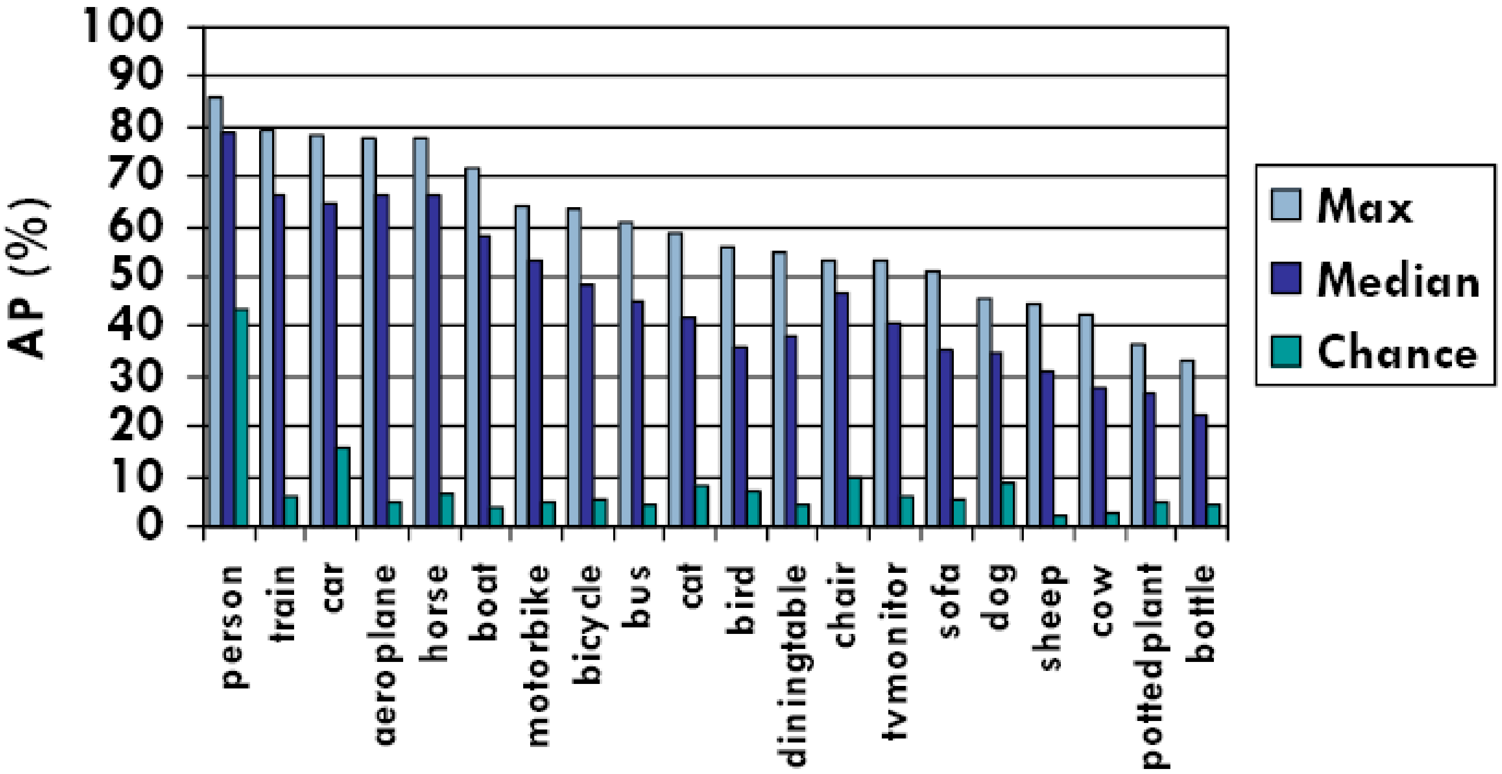
- Classification (ready for commercialization*)
  - Binary determination of whether or not an image contains at least one instance of a given category
- Localization (more research needed)
  - Specification of where each instance of a given category is in an image
    - E.g., a bounding box
- Classification easier
  - What is classification without localization and when is that useful?

PASCAL VOC Example Images



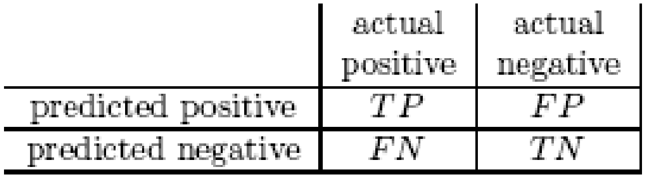
PASCAL VOC 2007 Results

- Classification task (yes/no, chance high)
  - Some categories much harder than others



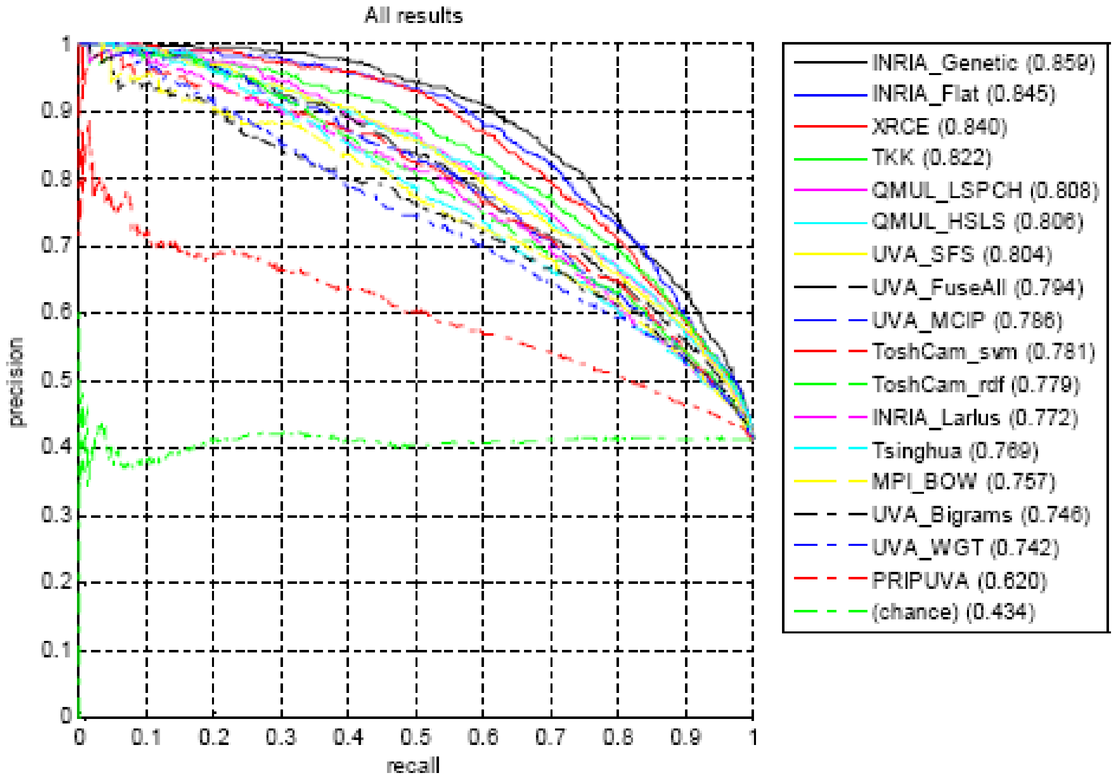
Measuring State of Art

- Precision vs. recall curve, and mean average precision
  - Adopted from TREC text retrieval evaluations
- In contrast with ROC (detection vs. false alarm rate) curve, and area under curve
  - Address problems with highly skewed data
- Precision = TP/(TP+FP)
- Detection rate
  = recall = TP/(TP+FN)
- False alarm rate = FP/(FP+TN)



	actual positive	actual negative
predicted positive	TP	FP
predicted negative	FN	TN

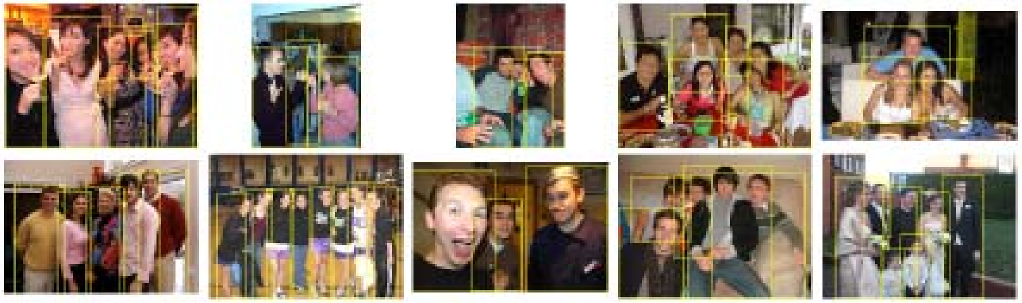
Best Categorization: Person



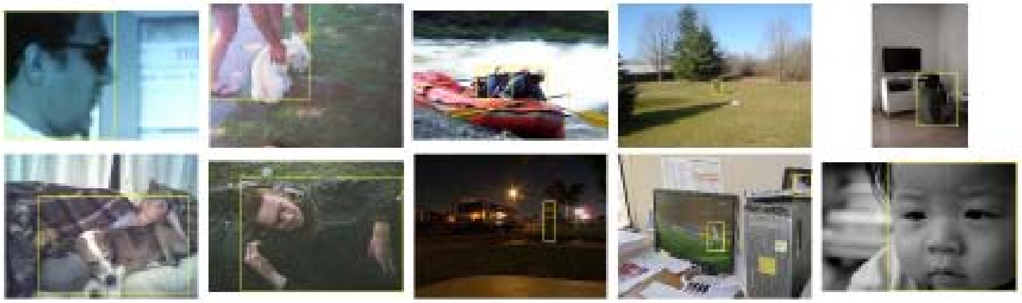
- More research emphasis on category? Predictable scene context? High baseline?

Person Categorization Examples

- Top ranked in-class



- Bottom ranked in-class



- Top ranked non-class
  - Context?

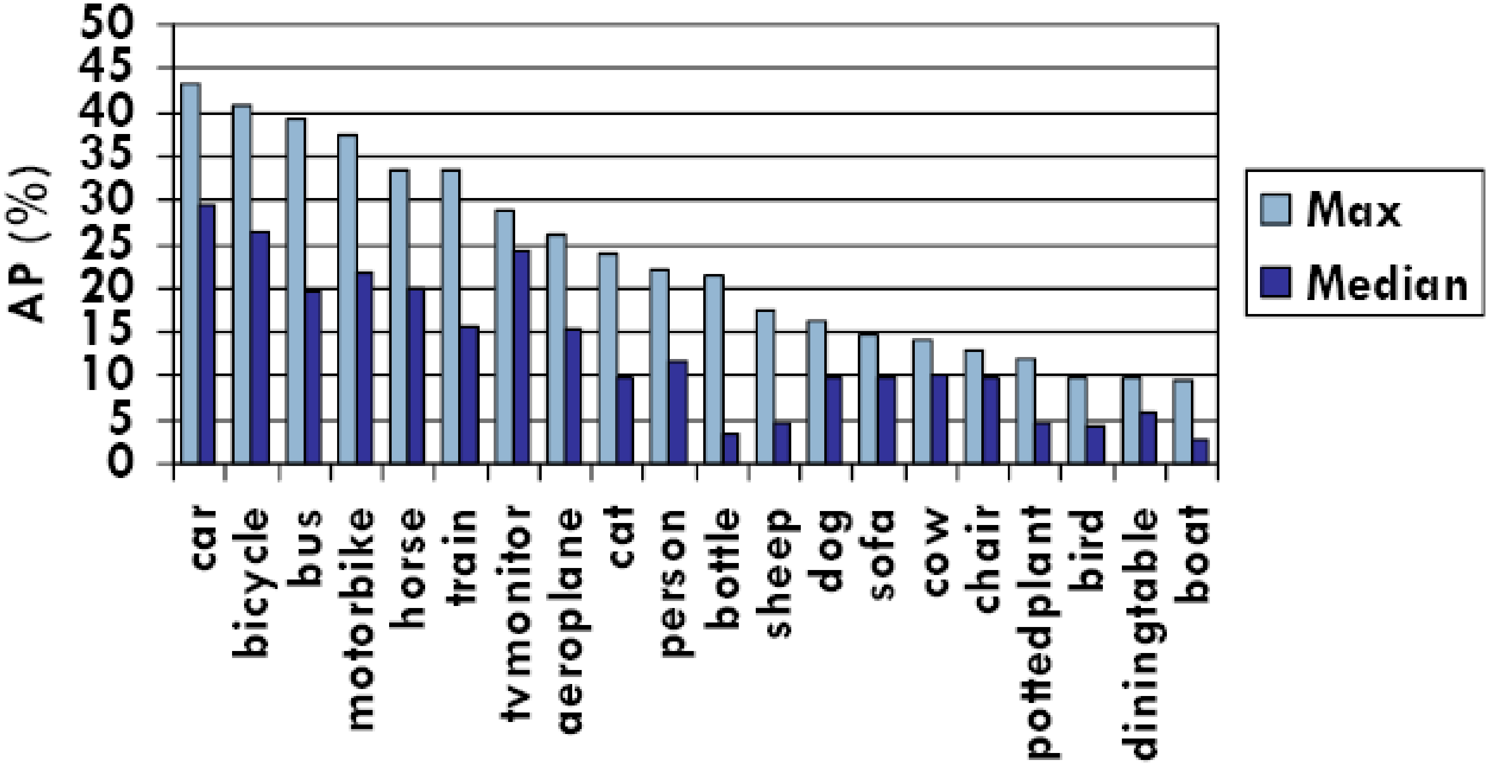
Ground truth boxes

Best Categorization Methods

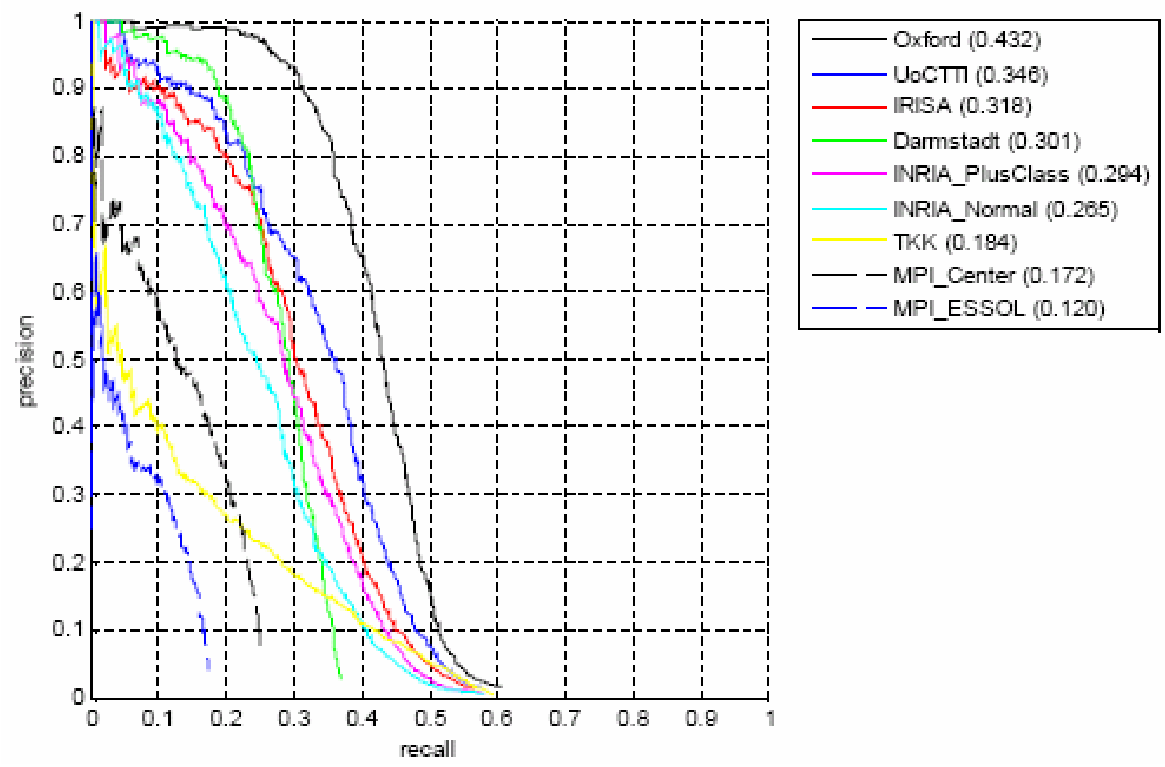
- Bag of words type models
  - Adapted from text classification literature
  - Words not well defined in case of images – local properties
    - “Visual words”
- Interest operators – spatial derivatives
  - Edgels, Laplacian of Gaussian, Harris corners
- Local features – describe interest regions
  - SIFT descriptors, textons, color histograms, pairs of adjacent segments

PASCAL VOC 2007 Results

- Localization task (chance essentially zero)
  - Correct location: intersection over union of bounding boxes ≥.5

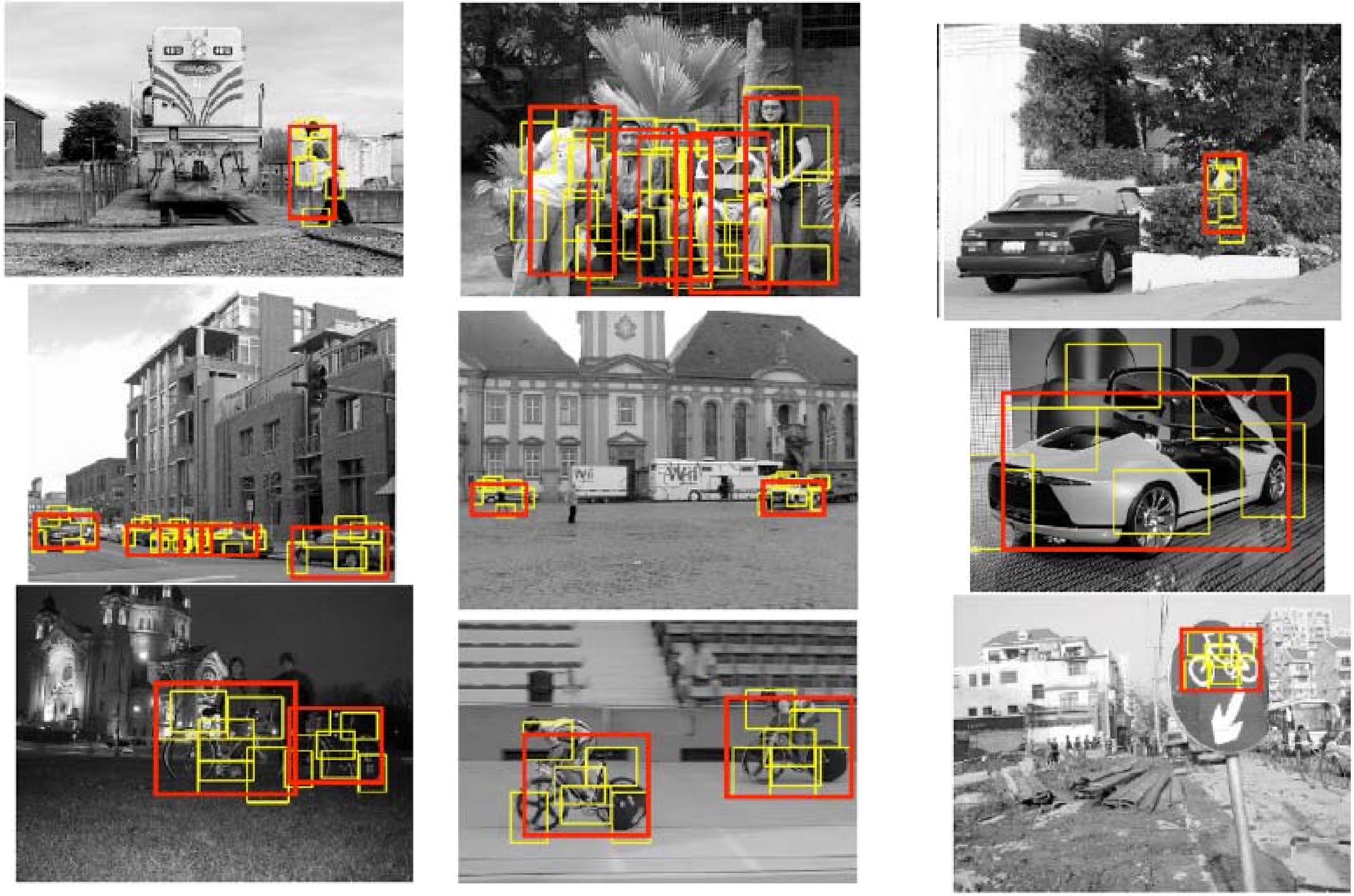


Best Localization: Car



- Again category with lots of research work, also relatively predictable context

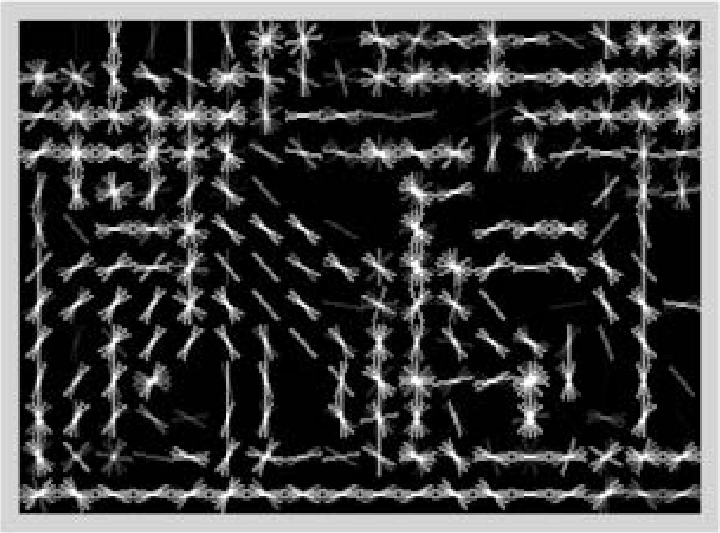
Localization Examples (UofC-TTI)



Best Localization Methods

- Sliding window style classifiers
  - SVM, AdaBoost
  - Flexible spatial template: “star model” of SVM’s
- Separate classifiers by viewpoint
- Use of context in classifiers
- Use of segmentation
- Local features – similar to those used for classification
  - HoG, SIFT, local histograms of derivatives

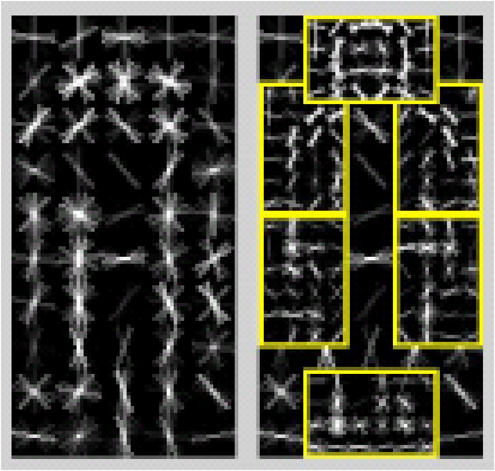
HoG Features



- Image partitioned into 8x8 blocks
- In each block compute histogram of gradient orientations
- Can be done at multiple spatial scales

Flexible Spatial Template (UofC-TTI)

- Hierarchical model [Felzenszwalb et al 08]
  - Coarse template with fine-scale part templates connected by springs (deformable)
  - Learned automatically from examples labeled with bounding boxes



Recap of Recognition

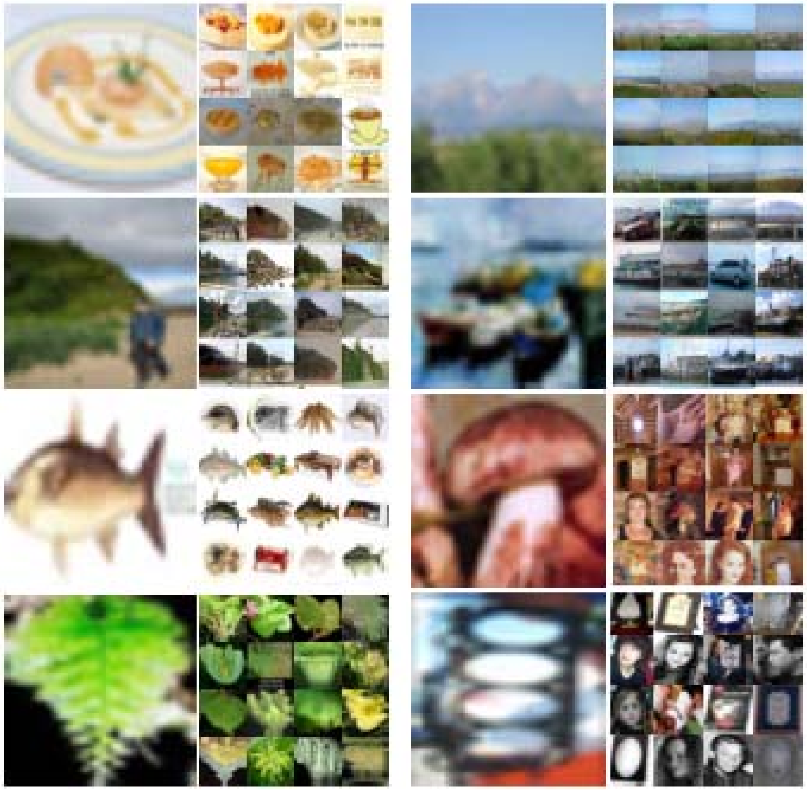
- Object category recognition
  - Classification methods to determine presence or absence of a category in an image
  - Localization methods to determine where each instance of a category is in an image
  - Former more mature and work better than latter; can be useful for image sorting/ranking
- Learning techniques
  - Operate on hundreds or thousands of training instances per category

Image Gist Classification

- Automatically group photos together based on scene characteristics (ready for commercialization*)
  - Gist of scene [Oliva & Torralba 01]
    - Low resolution thumbnails (e.g., 32x32)
    - Oriented filters at various spatial scales
  - Simple techniques become useful for large collections
    - 80M images [Torralba et al 07]
- Capture gross and some fine characteristics
  - Indoor, outdoor, urban, suburban, rural, mountain, water, farmland

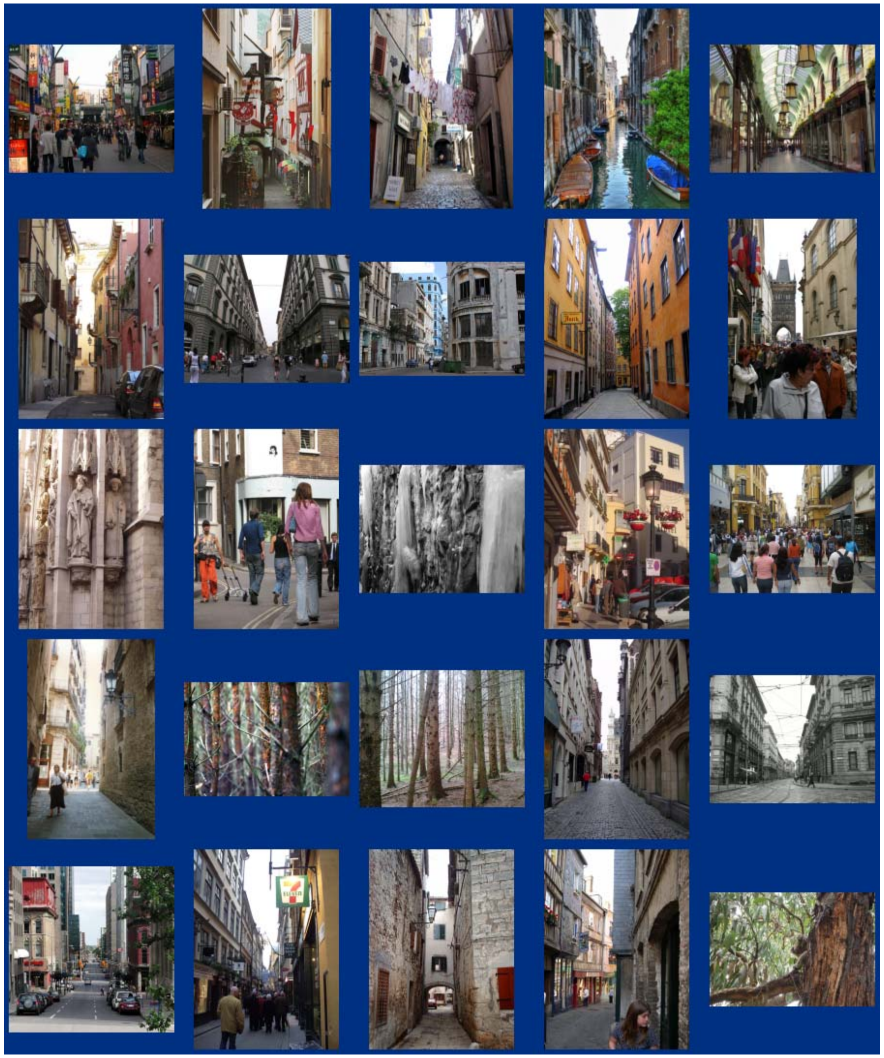
Scene Gist, 80M Tiny Images

- Image and 16 closest by gist (left)
- Average images by label plus gist (right)



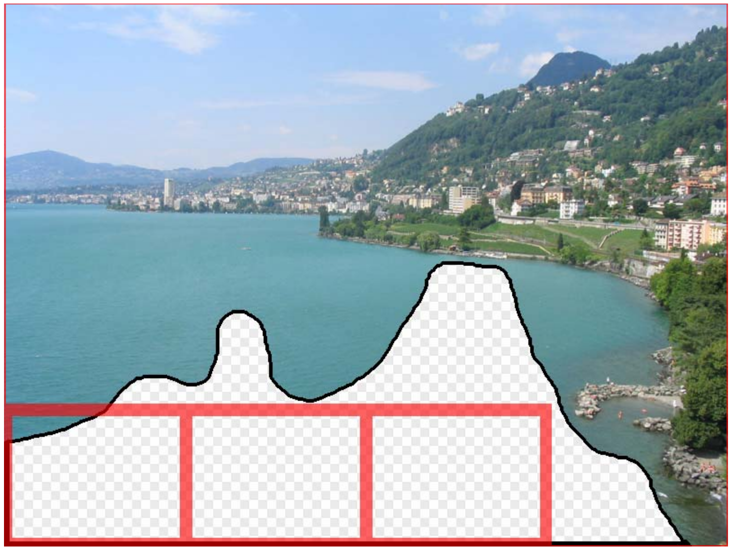
Gist for Scene Matching: How Good?

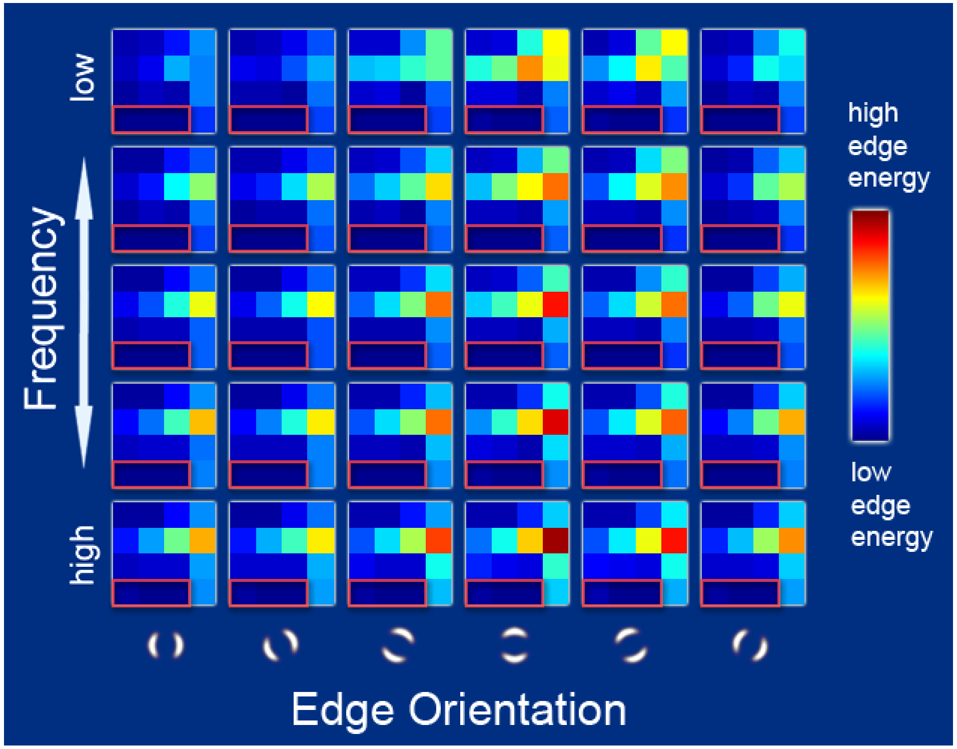
- Closest outdoor scenes by gist
  - 2.1M images from online travel photo groups (no labels)



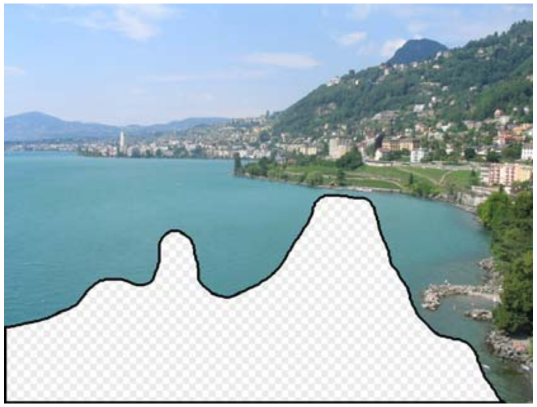
Collection Matching for Filling-In

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

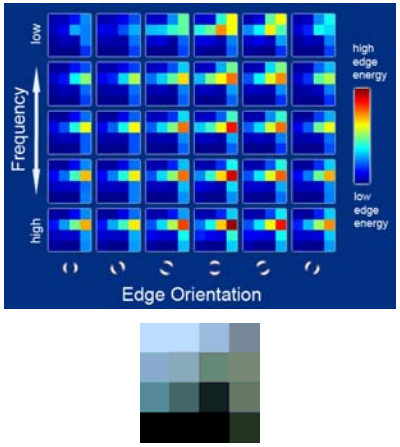




Filling-In Algorithm



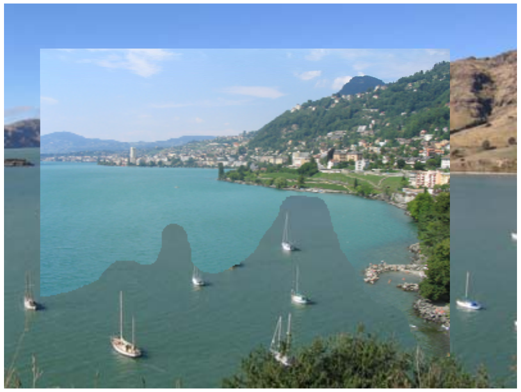
Input image



Scene Descriptor

Image Collection …

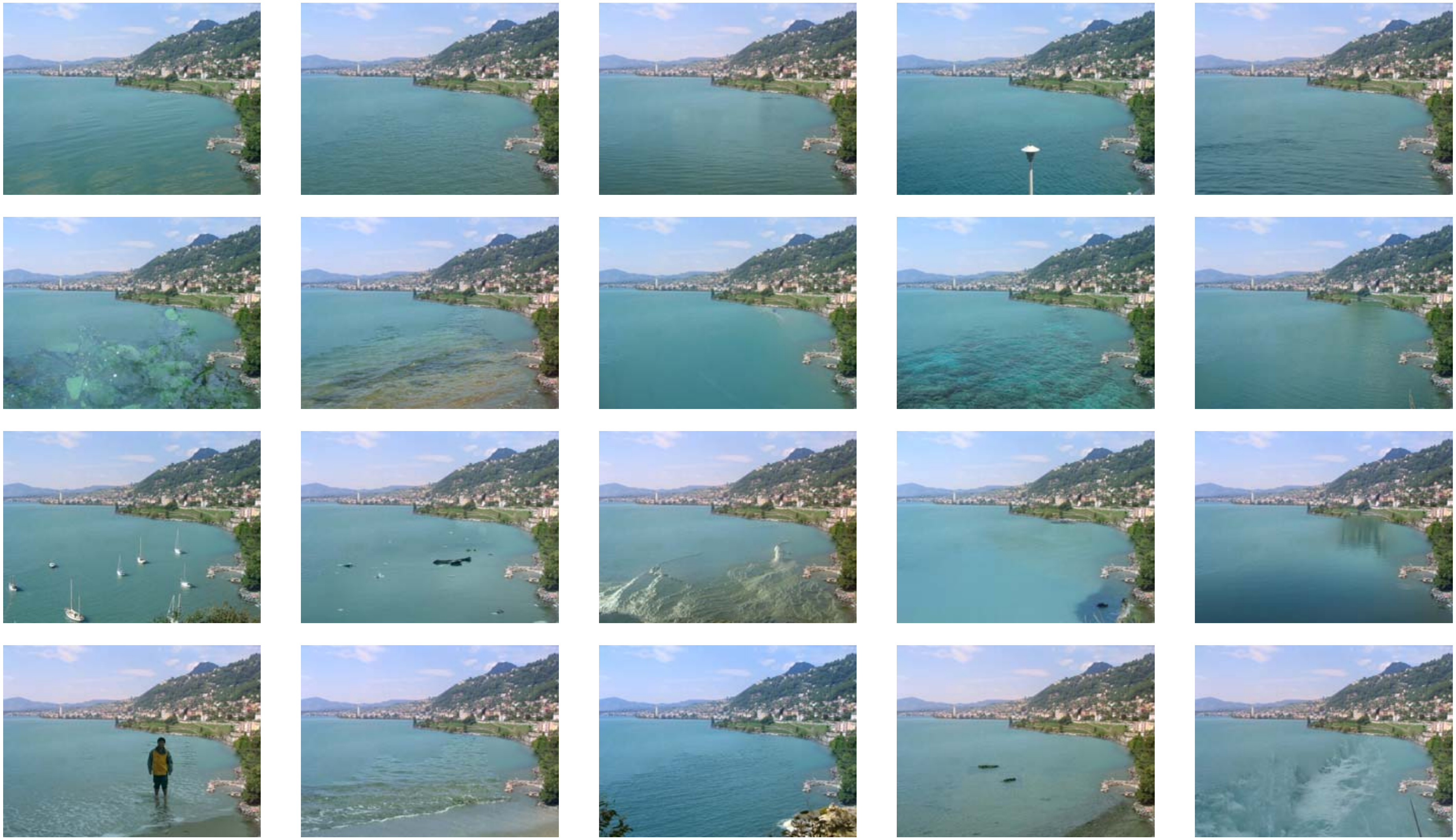
200 matches …



Context match/blend

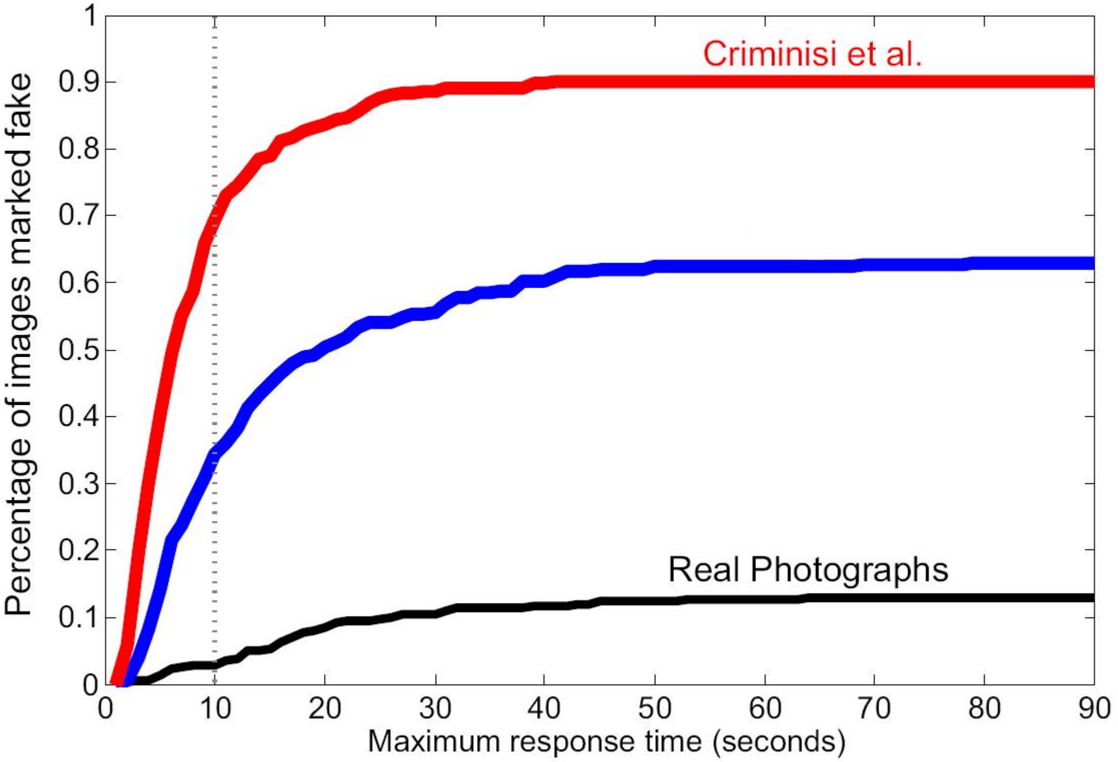
20 completions

Top 20 Results for Filling-In



Better than Local Filling-In Methods

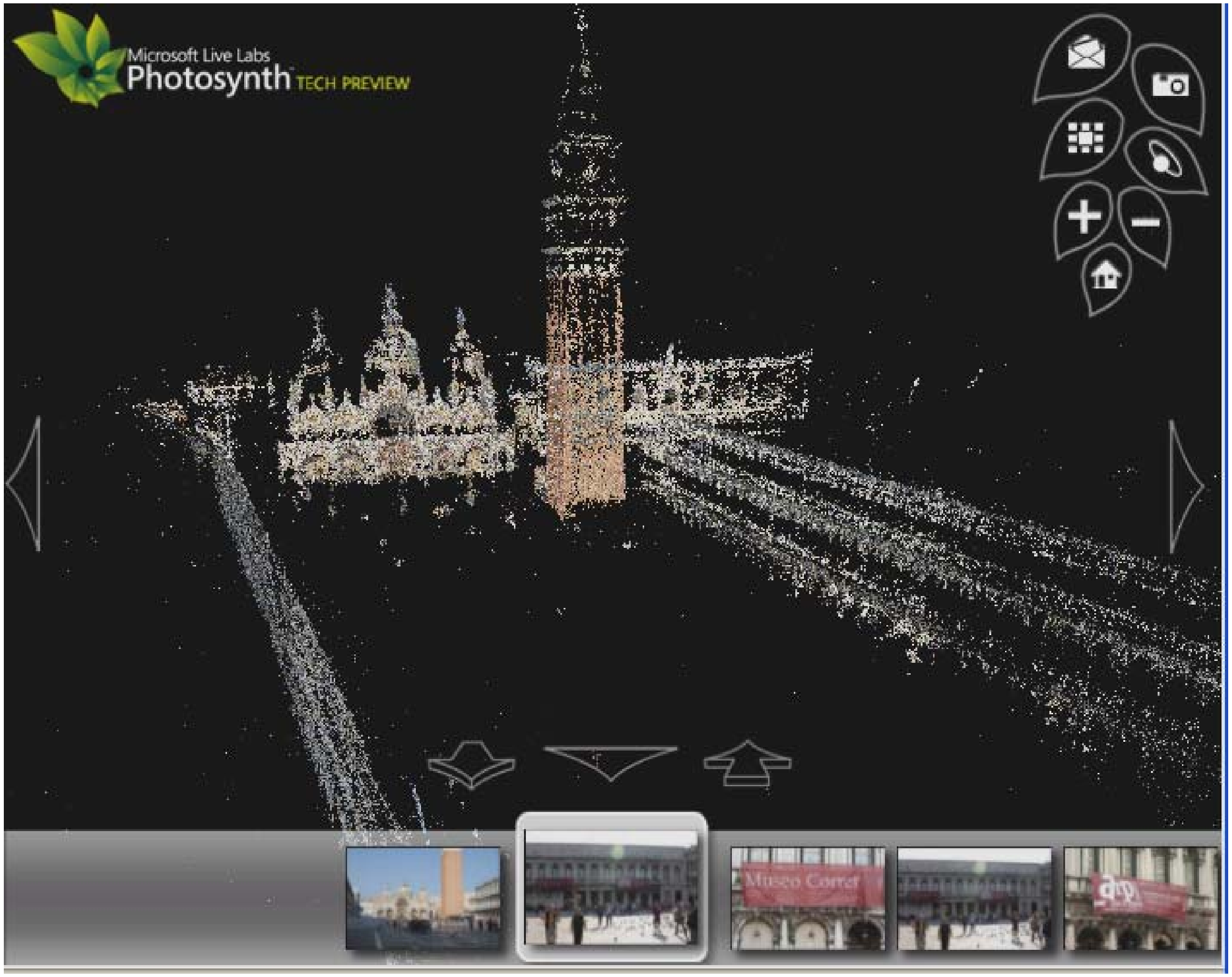
- Judgments of real vs. fake (20 subjects)



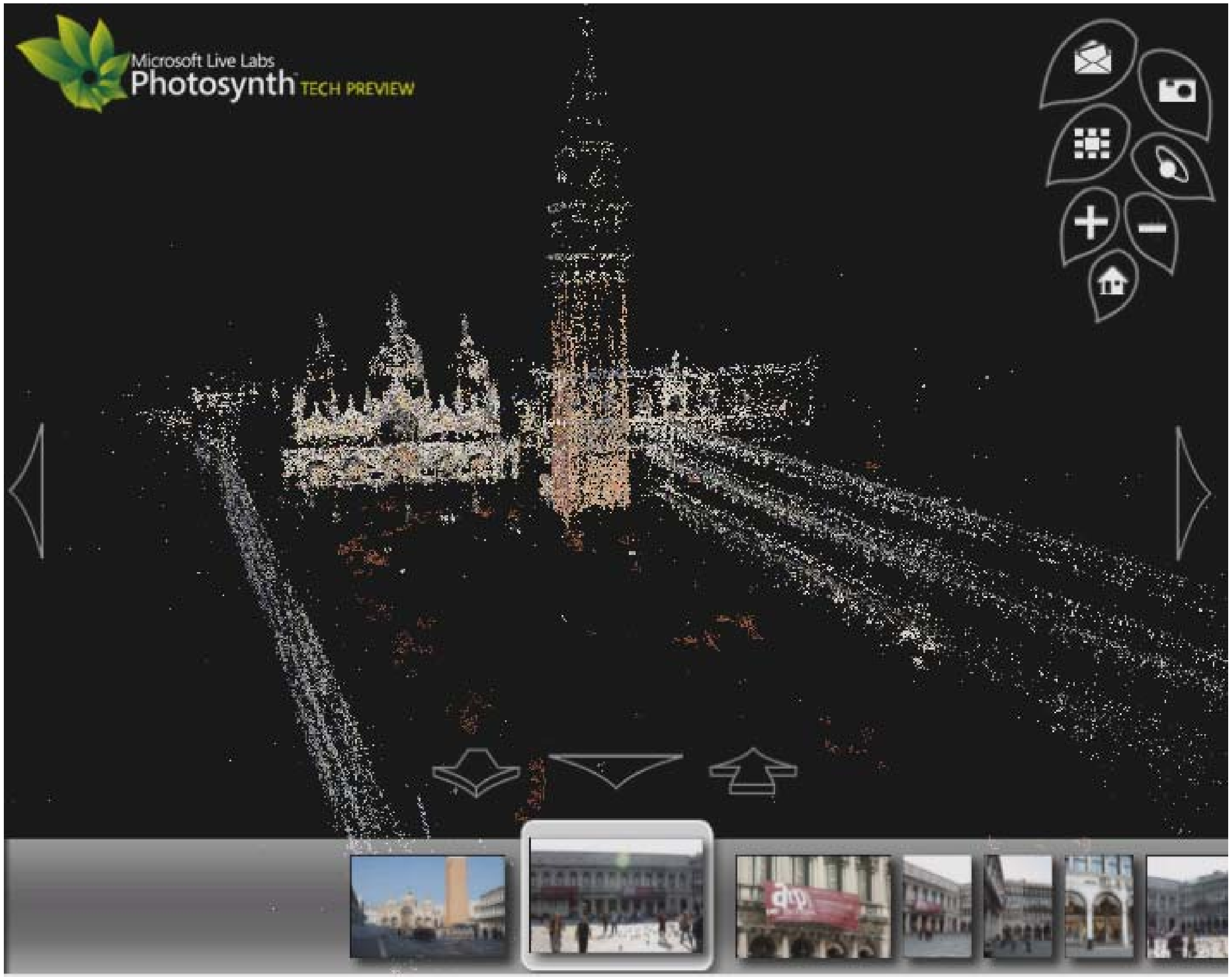
3D Organization of Photos

- Navigating set of uncalibrated photos of a scene [Snavely et al 06]
  - 3D scene structure, image alignment and camera locations
- Microsoft Photosynth (about to be available commercially)
  - Interactive viewer for collections of photos of a single scene
    - Taken with different cameras, at different times, in unknown locations
  - http://labs.live.com/photosynth/

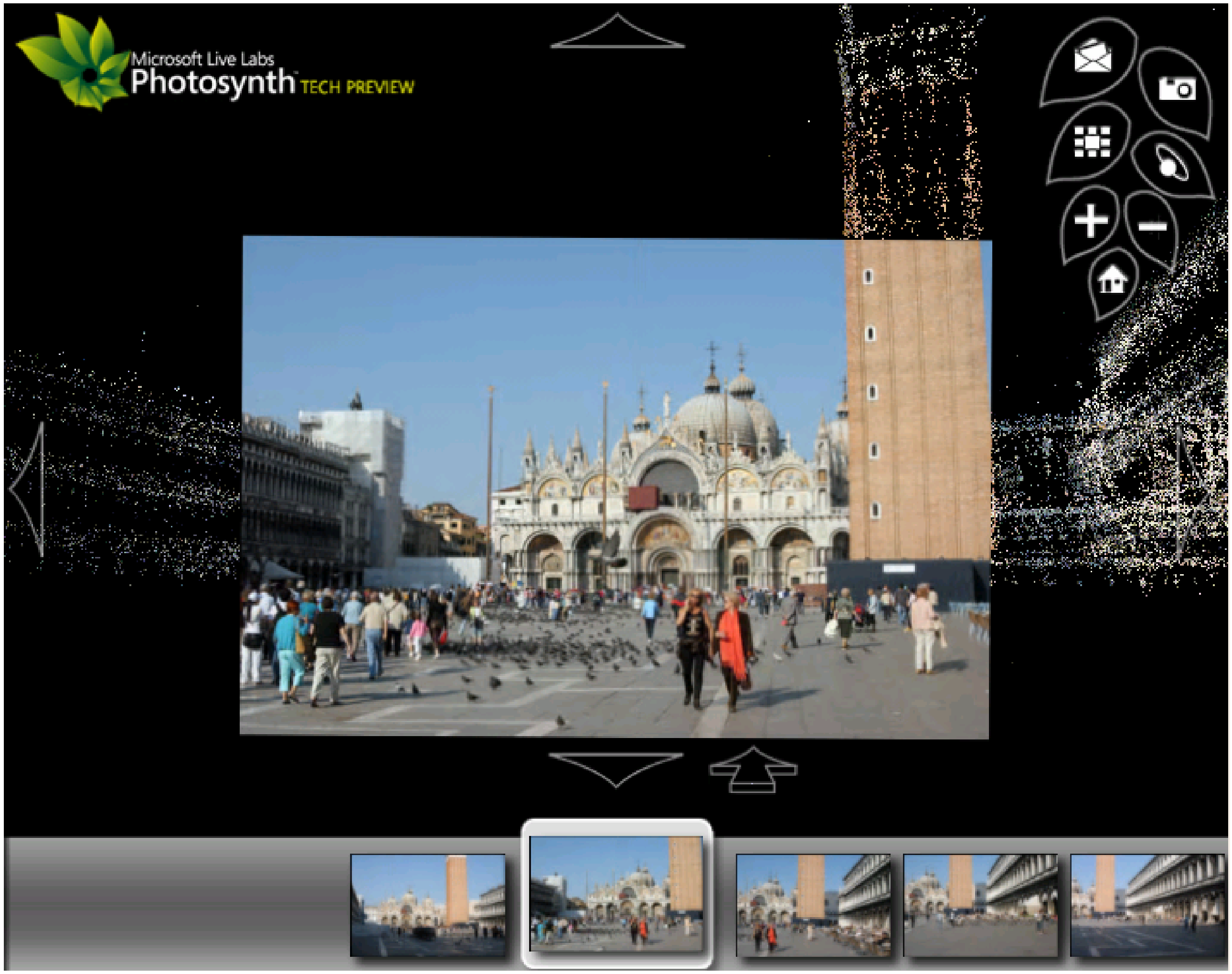
Photosynth: 3D Structure



Photosynth: Camera Locations



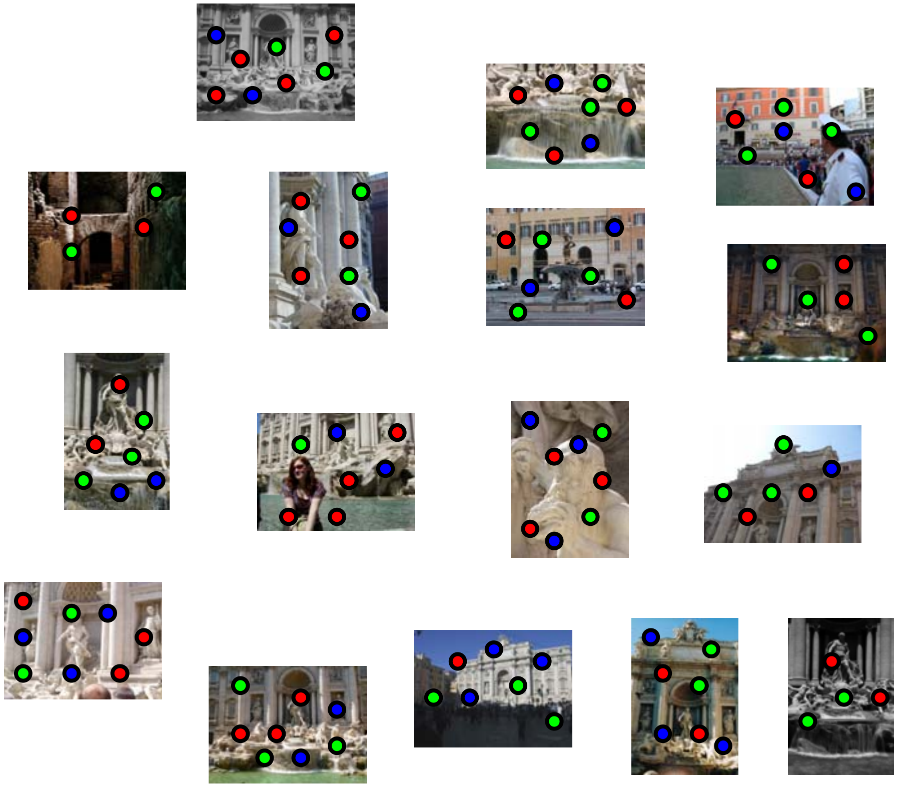
Photosynth: Photo Organization



Photosynth Video

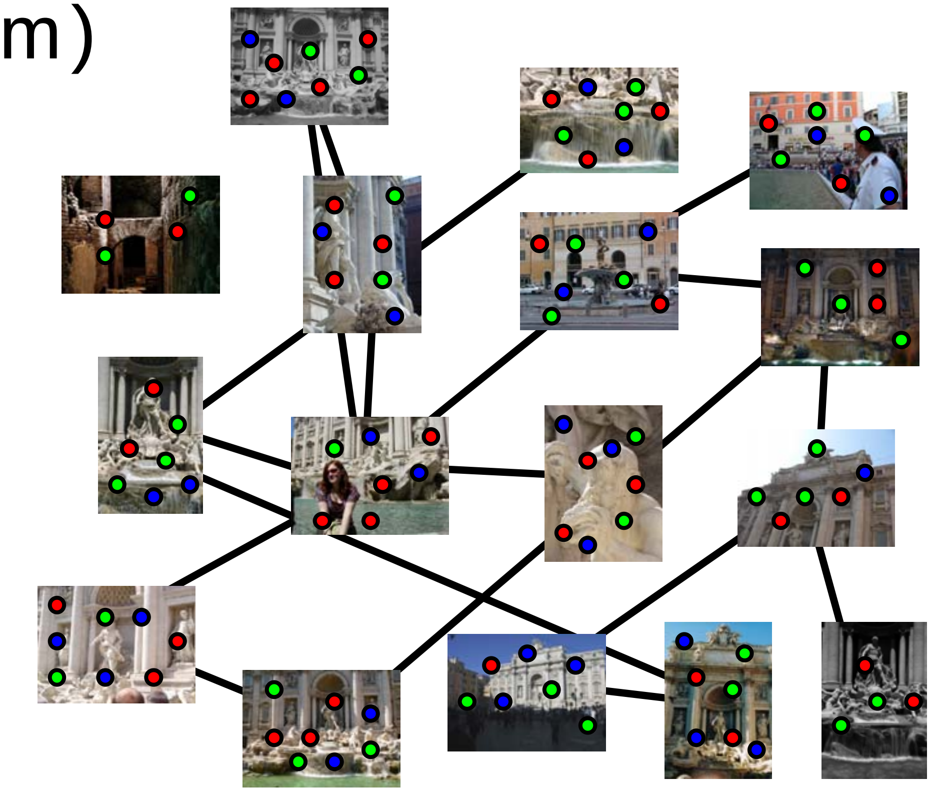
Detect Local Features

- SIFT feature detection

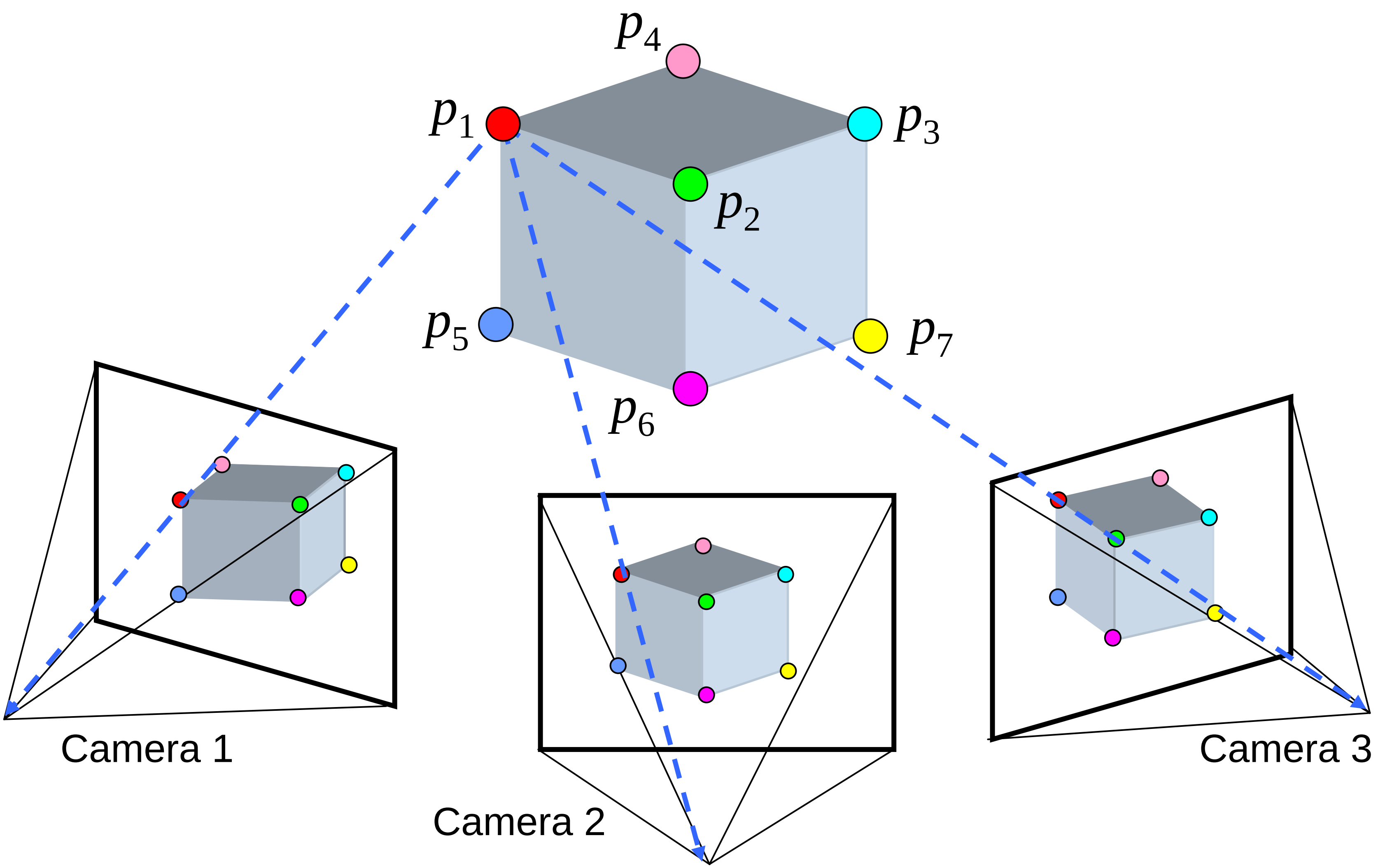


Find Correspondences

- Form graph from all pairs using RANSAC to estimate fundamental matrix (camera transform)



3D Point Cloud: Structure From Motion



Summary

- Substantial recent advances in object category recognition – classification of images as containing given category
  - Small number of categories (tens)
- New techniques for organizing large collections of photos based on coarse 2D or 3D properties
  - Useful in applications such as filling-in missing data in images
  - Organizing photos of a given scene