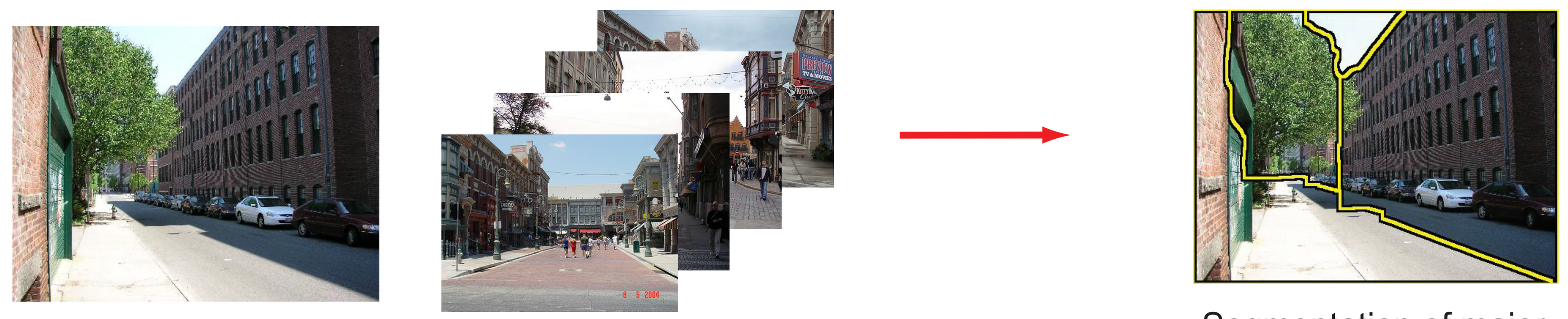
Segmenting scenes by matching image composites

Bryan C. Russell[1] Alexei A. Efros[2,1] Josef Sivic[1] William T. Freeman[3] Andrew Zisserman[4,1]
[1]INRIA/ENS [2]CMU [3]CSAIL MIT [4]Oxford University

Goal: unsupervised scene segmentation

Input Output



Input image
Unlabeled database of 100K images
Segmentation of major scene components

Approach: use lots of similar images to drive image segmentation

Extremes of matching

Image to match

Match individual pixels

Database Top pixel matches

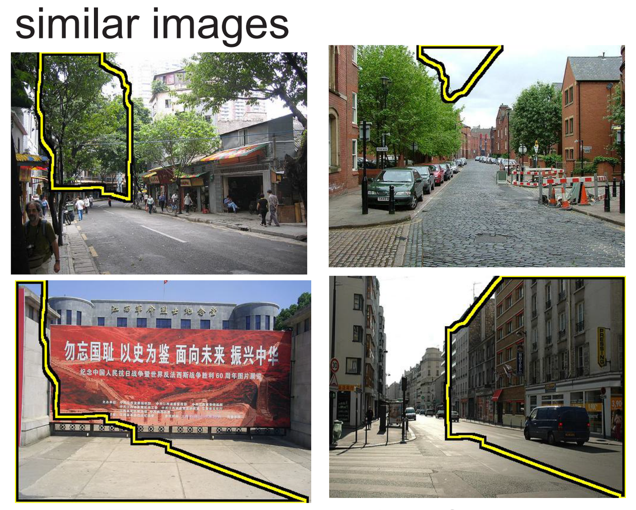
- Each pixel matches almost perfectly
- No semantic correspondences

Match entire image

Top matches from 100K images using gist

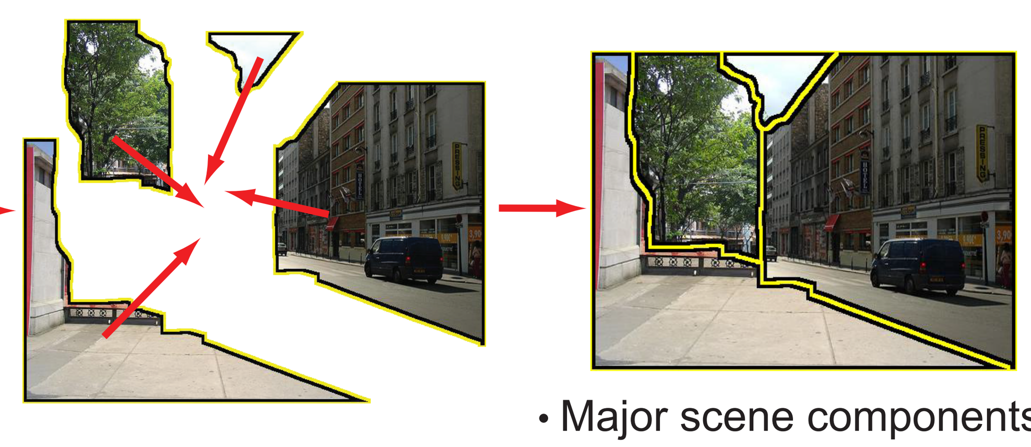
- Images with similar objects are retrieved
- Not all scene components align well

Main idea: match regions from similar images



Top region matches from 5K globally similar images

Composite new explanation of image



- Major scene components are aligned with input image

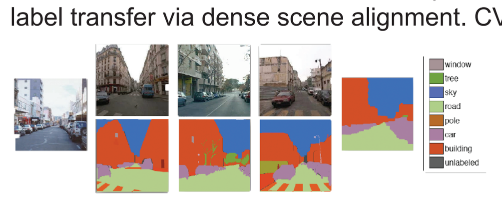
Previous work using globally aligned images

Improve photographs via compositing

- A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, M. Cohen. Interactive Digital Photomontage. ACM SIGGRAPH, 2004.

Supervised object detection

- B. C. Russell, A. Torralba, C. Liu, R. Fergus, W. T. Freeman. Object Recognition by Scene Alignment. NIPS, 2007.
- C. Liu, J. Yuen, and A. Torralba. Nonparametric scene parsing: label transfer via dense scene alignment. CVPR, 2009.

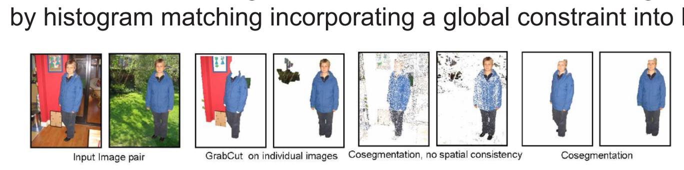


Inpainting and object pop-out

- Oliver Whyte, Josef Sivic, and Andrew Zisserman. Get Out of my Picture! Internet-based Inpainting. BMVC, 2009.
- H. Kang, A. Efros, T. Kanade, M. Hebert. Image Composition for Object Pop-out. 3dRR workshop, ICCV, 2009.

Co-segmenting same objects

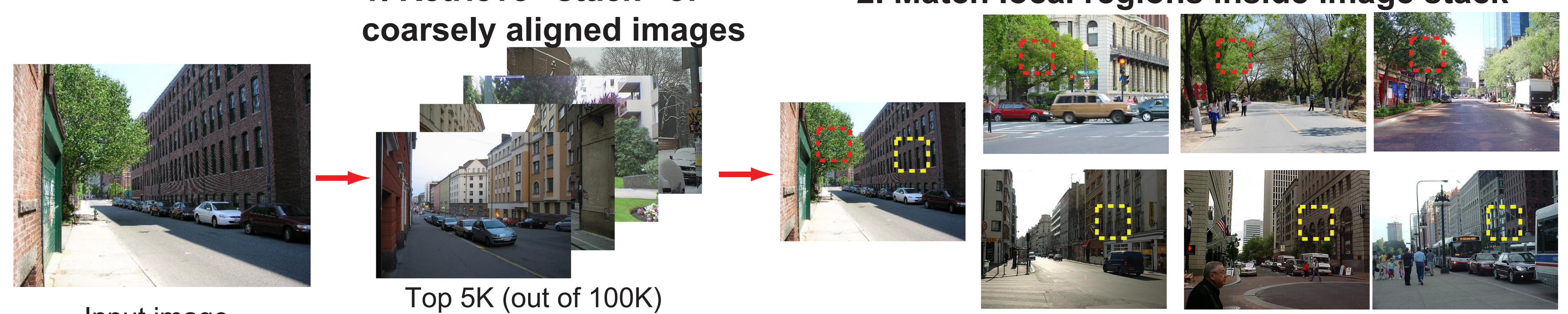
- C. Rother, V. Kolmogorov, T. Minka, and A. Blake. Cosegmentation of image pairs by histogram matching incorporating a global constraint into MRFs. CVPR, 2006.



Overall approach

1. Retrieve "stack" of coarsely aligned images

2. Match local regions inside image stack



Input image
Top 5K (out of 100K) matching images using gist
Top local region matches

3. Analyze statistics of local region matching

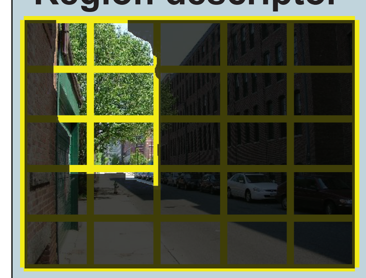
3a. Data-driven image grouping: Find pixels that match to similar images

3b. Data-driven boundary detection: Find pixels that match to different images

4. Perform MRF-based segmentation

Optimize over matches while respecting boundaries

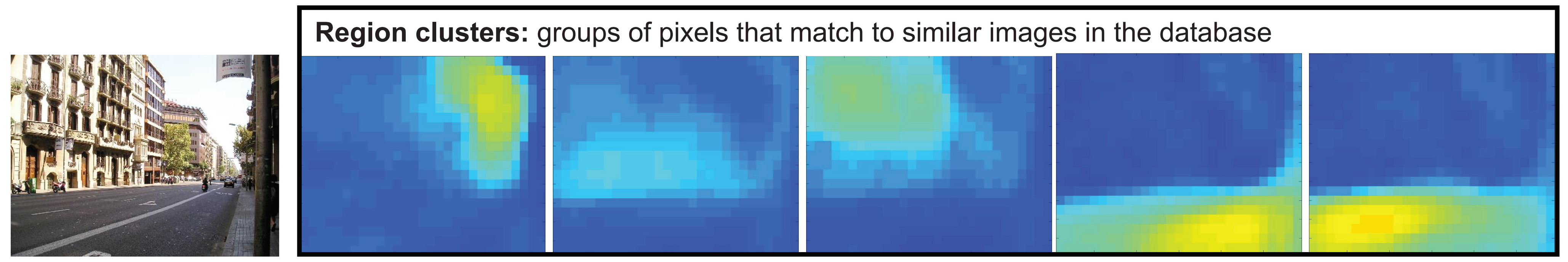
Region descriptor



- Accumulate Gabor and color channels over spatial bins modulated by spatial mask
- Compare descriptors using L1 norm

3a. Data-driven image grouping

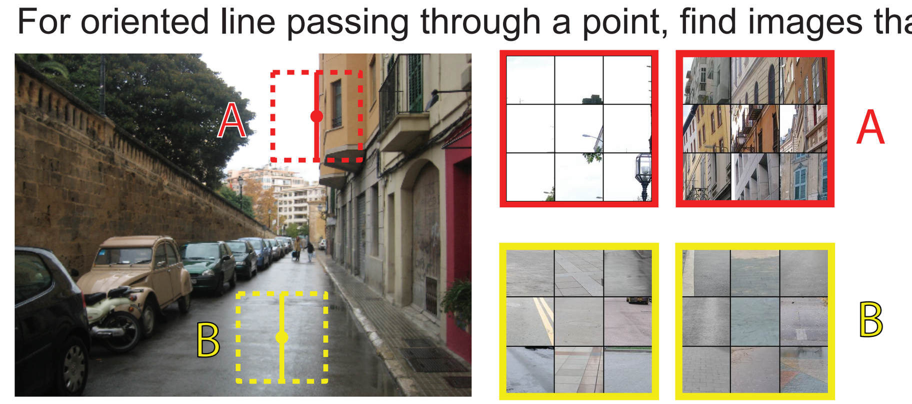
Region clusters: groups of pixels that match to similar images in the database



1. Form binary vectors (length 5K) that indicate top 1K nearest neighbor matches
2. Cluster binary vectors (5 clusters)
3. Measure similarity to cluster center as dot-product

3b. Data-driven boundary detection

For oriented line passing through a point, find images that match the appearance on each side of the line



Left-side matches Right-side matches

Occlusion boundary
Different image sets are matched

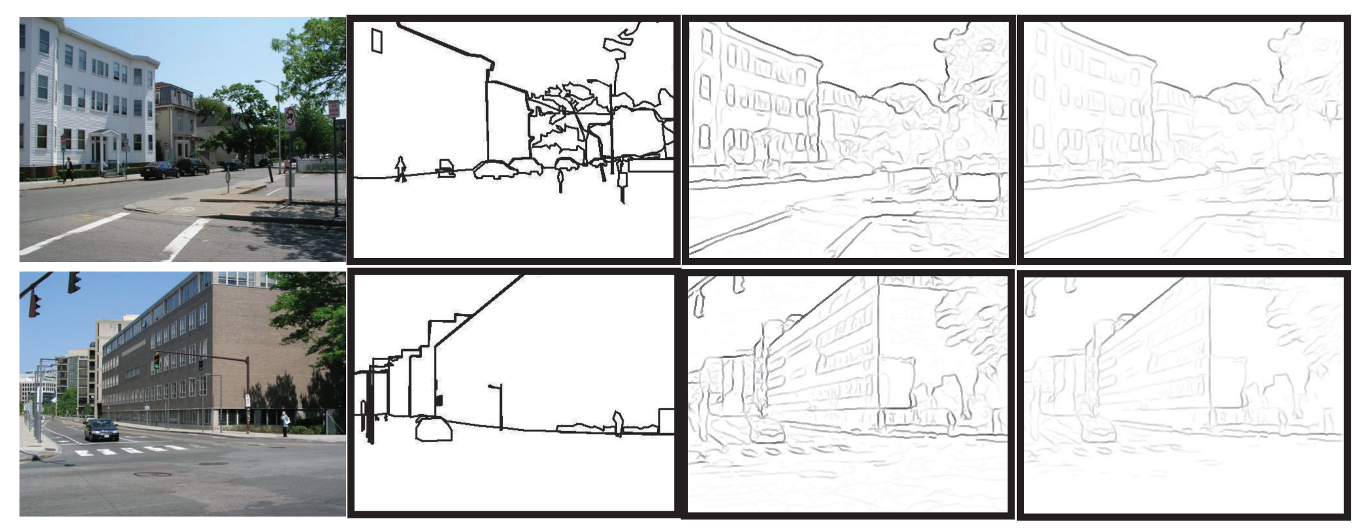
Inside object
Similar image sets are matched

- For each point, consider 8 orientated lines
- Measure how well sets match with Spearman's rank correlation

High correlation → inside object

Boundary outputs

- Compare with probability of boundary (PB) [Martin et al. '04]
- Notice that we suppress contours interior to the major scene components



Input | Ground truth LabelMe boundaries | Probability of boundary (PB) | Ours (1-rank correlation)*PB

4. MRF formulation

$$\min_{\mathbf{x}} \sum_i \phi_i(x_i, \mathbf{y}_i) + \sum_{(i,j)} \psi_{i,j}(x_i, x_j)$$

Assign each pixel to a region cluster or as "outlier"

$$\phi_i(x_i = k, \mathbf{y}_i) = \begin{cases} -s(\mathbf{c}_k, \mathbf{y}_i) & k \in \{1, \dots, K\} \\ \gamma & k = 0 \end{cases}$$

← Similarity to cluster center
← Outlier state

$$\psi_{i,j}(x_i, x_j) = (\alpha + \beta f(i,j))\, \delta[x_i = x_j]$$

Smoothing term
Data-driven boundary (1-rank correlation)*PB

Optimize with GraphCuts using alpha-beta swaps

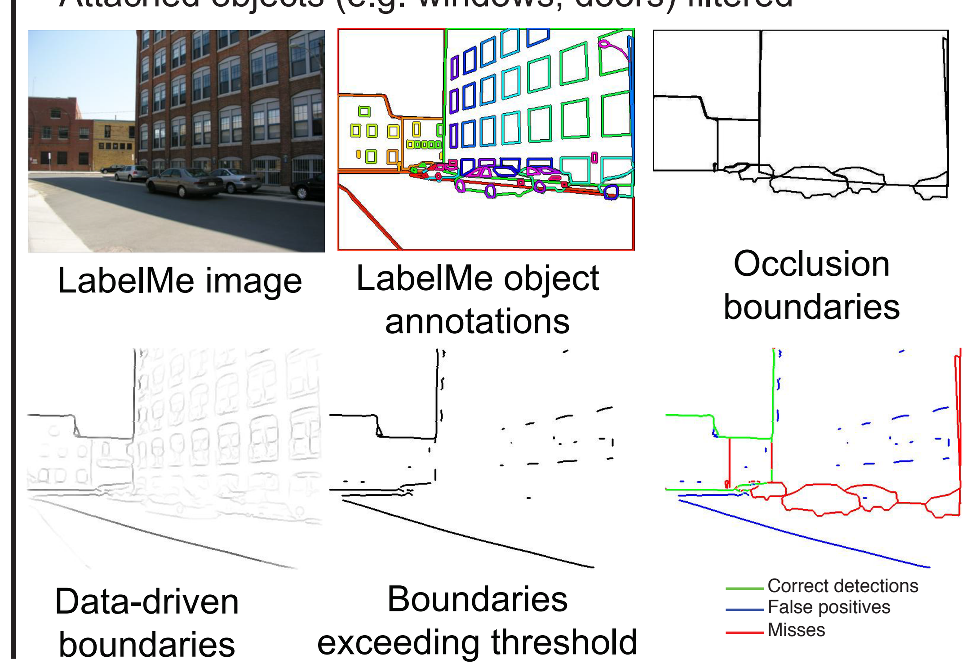
System outputs

Input image with output segmetation overlaid
Image composites of top-matching regions
Top gist matches on entire image

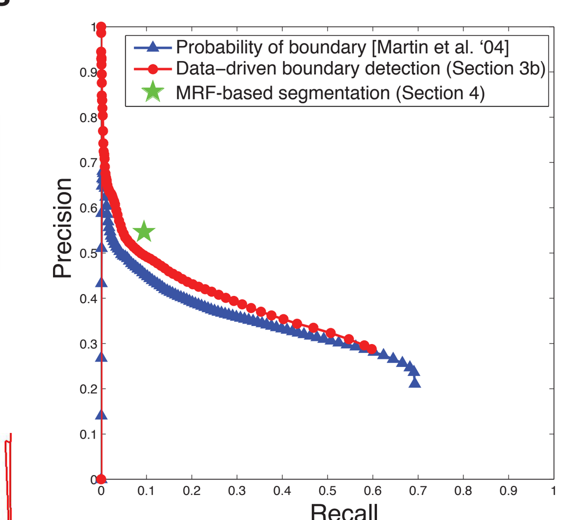
Boundary detection evaluation

Database of principal occluding and contact boundaries

- 100 street scene images from LabelMe database
- Attached objects (e.g. windows, doors) filtered



LabelMe image
LabelMe object annotations
Occlusion boundaries
Data-driven boundaries
Boundaries exceeding threshold



Precision at 0.09 recall:
- 0.45 - Probability of boundary
- 0.50 - Data-driven boundary
- 0.55 - MRF-based segmentation