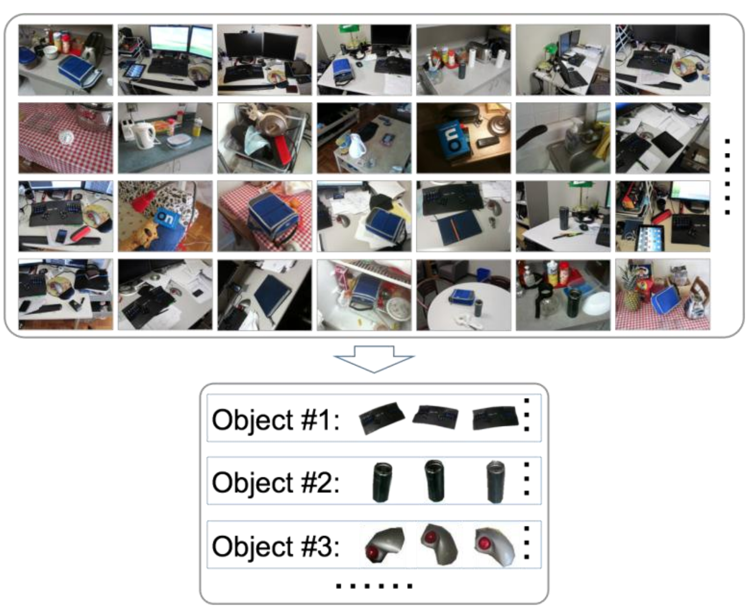
Unsupervised discovery of category and object models

Martial Hebert

The task

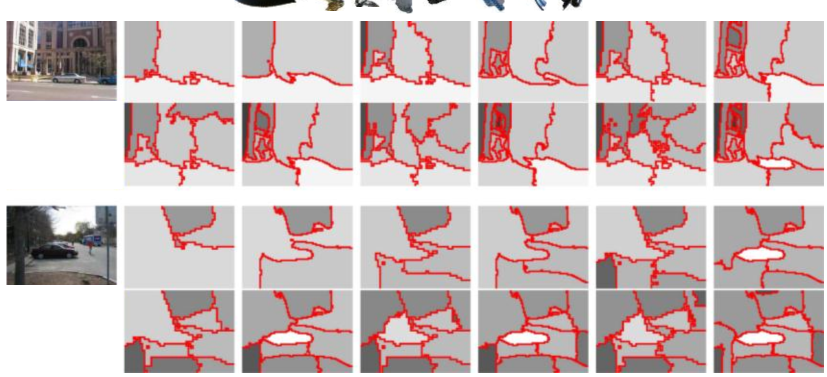


Common ingredients

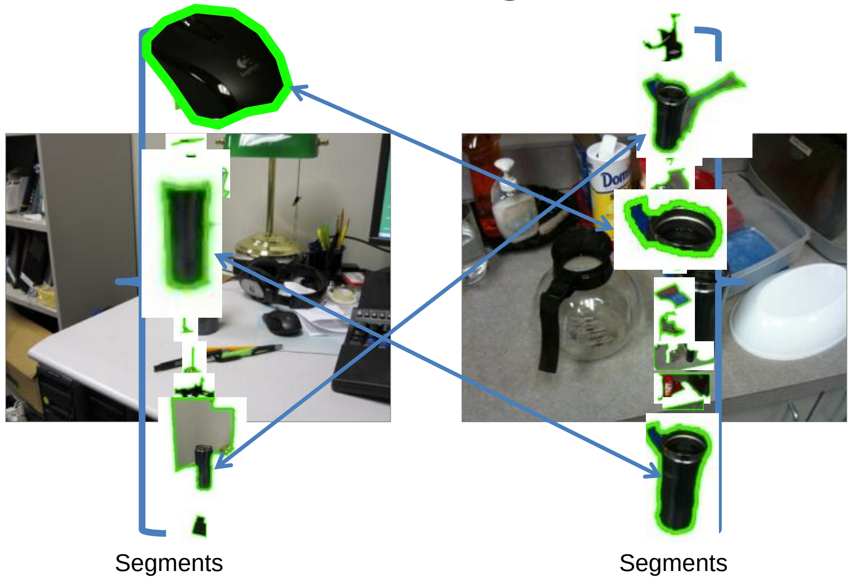
1. Generate candidate segments
2. Estimate similarity between candidate segments
3. Prune resulting (implicit) graph
4. Extract subgraphs corresponding to objects

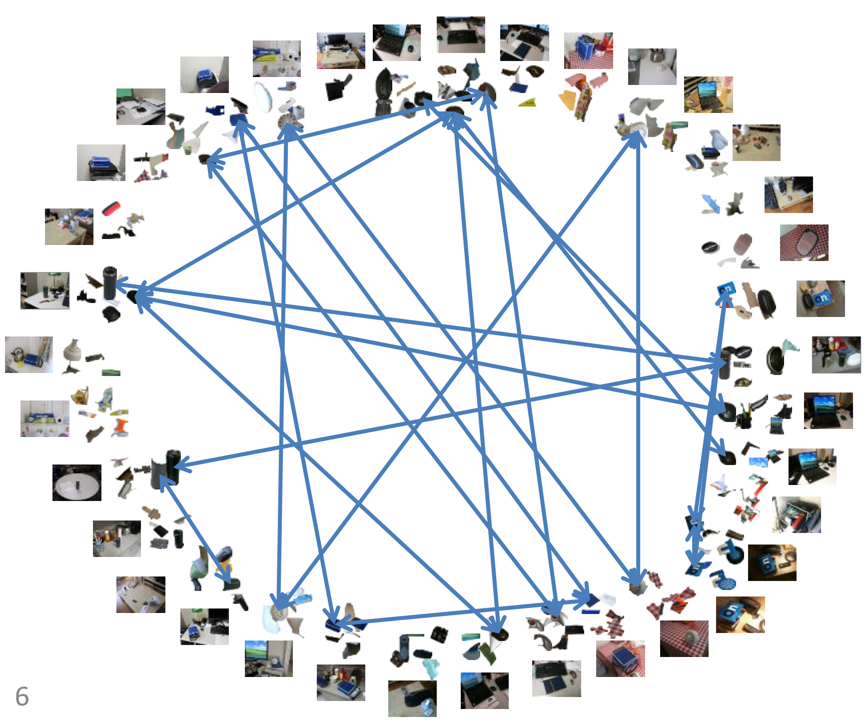
1. Generate candidate segments

- Regions from multiple segmentations



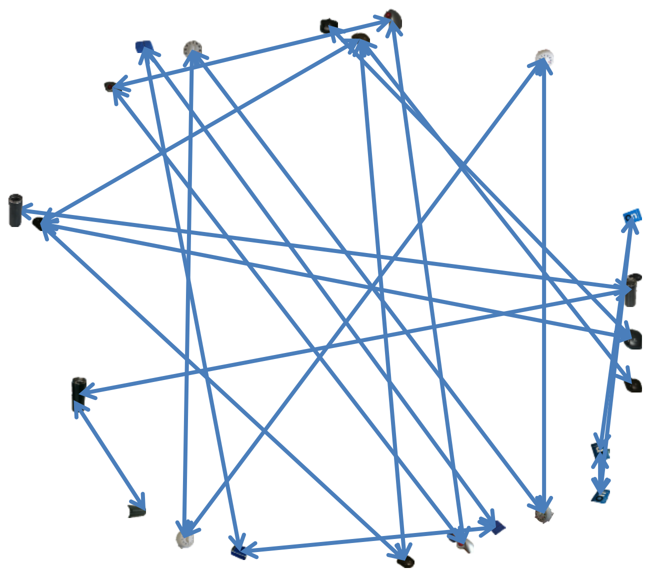
2. Estimate similarity between candidate segments



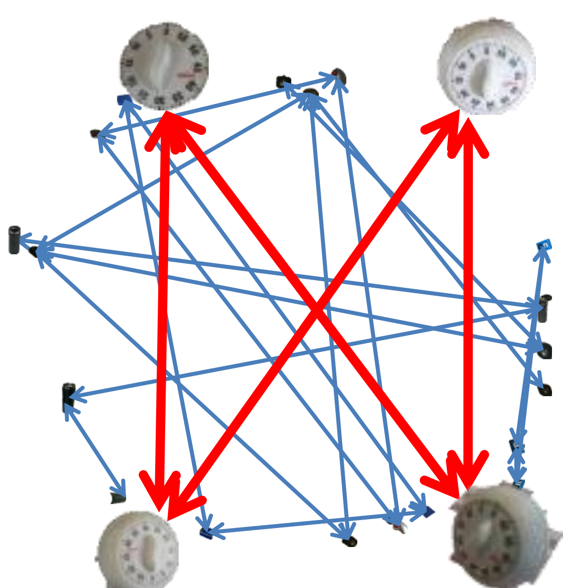


6

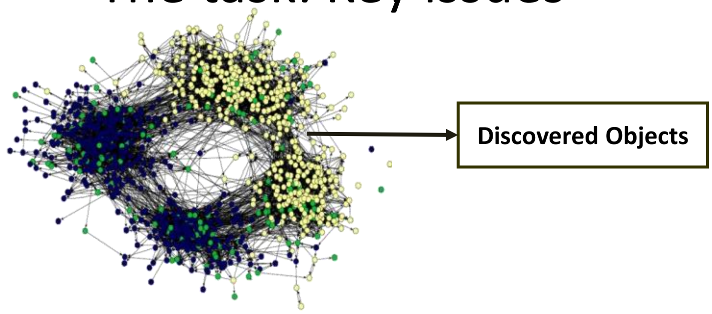
3. Prune graph



4. Extract subgraphs corresponding to objects

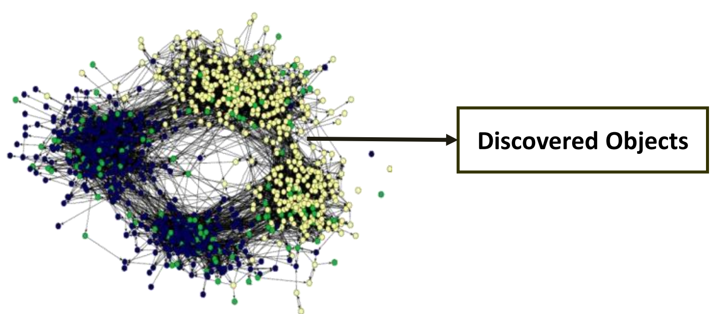


The task: Key issues



- In reality: Very large graph
  - Full pairwise comparison intractable for real problems
  - Most of the graph irrelevant to any category
- Questions:
  - How to cluster segments?
  - How to define robust affinities?
  - How to select/prune candidate segments and edges to reduce graph?

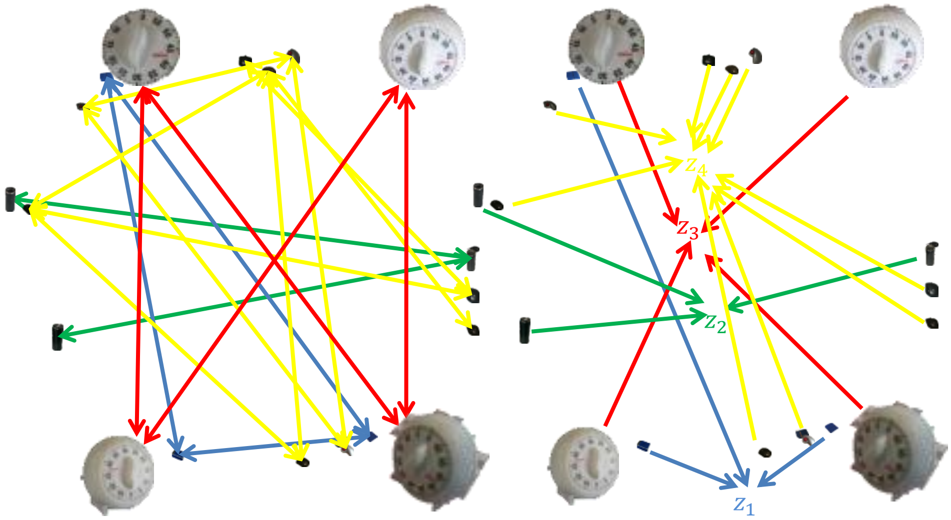
Outline



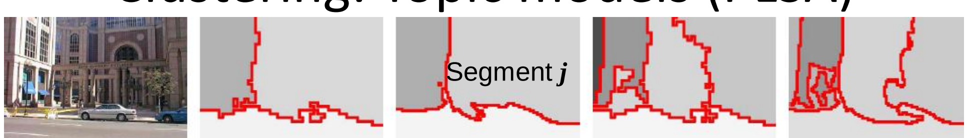
- How to cluster segments?
  - Topic model
  - Graph/spectral clustering
  - Graph analysis
- How to define robust affinities?
  - Robust affinities from contextual information
- How to select/prune candidate segments and edges to reduce graph?
  - Select segments using learned objectness
  - Prune graph edges using domain constraints

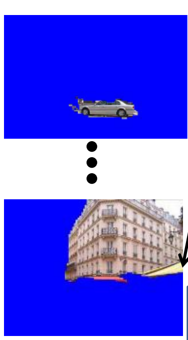
Clustering: Topic models

- Avoid comparing segments explicitly by estimating directly correspondence between segments and unknown (latent) models.

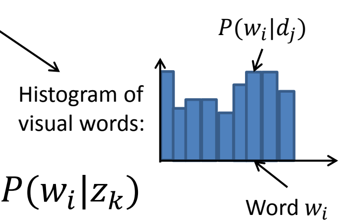


Clustering: Topic models (PLSA)

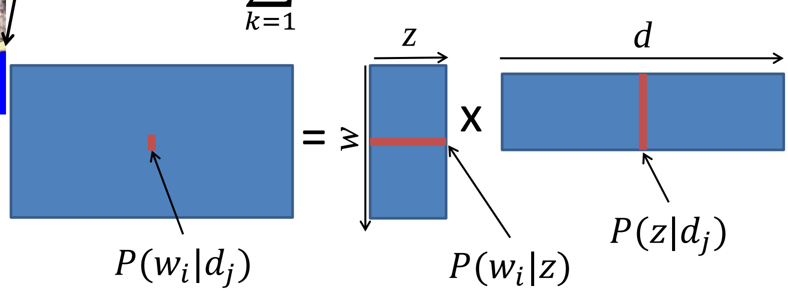




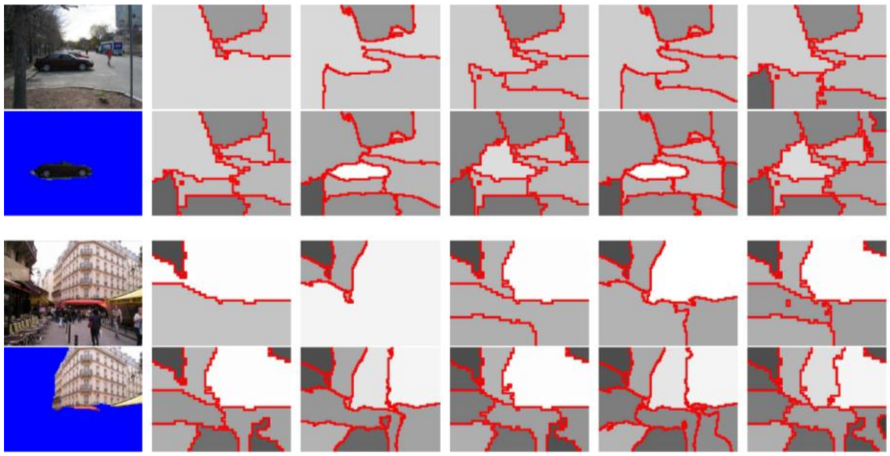
K "topics" z_k= objects



$$P(w_i|d_j) = \sum_{k=1}^{K} P(z_k|d_j)P(w_i|z_k)$$

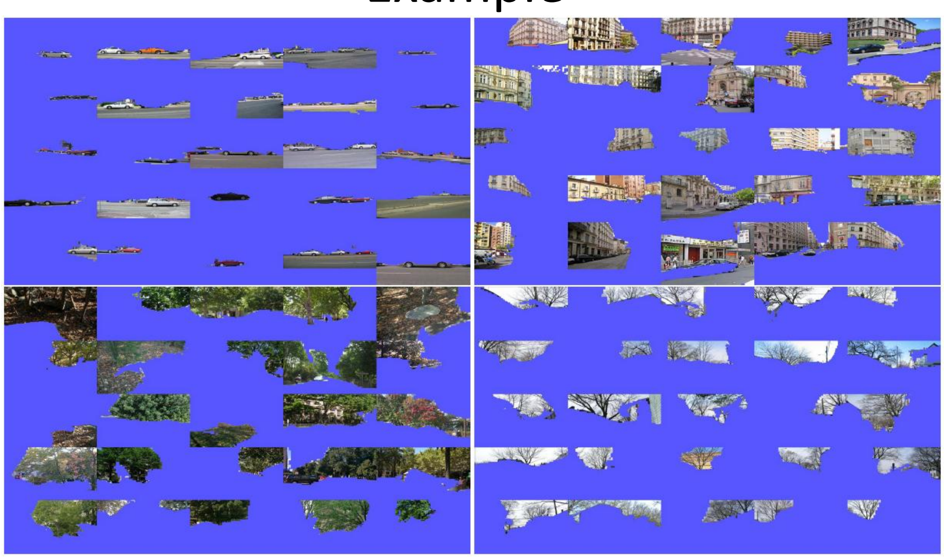


From topics to segments



- Segment score = estimated from difference (KLD) between actual distribution of words in image segment and predicted distribution $P(w_i|z)$
- Select segment with highest score

Example



From LabelMe dataset…..

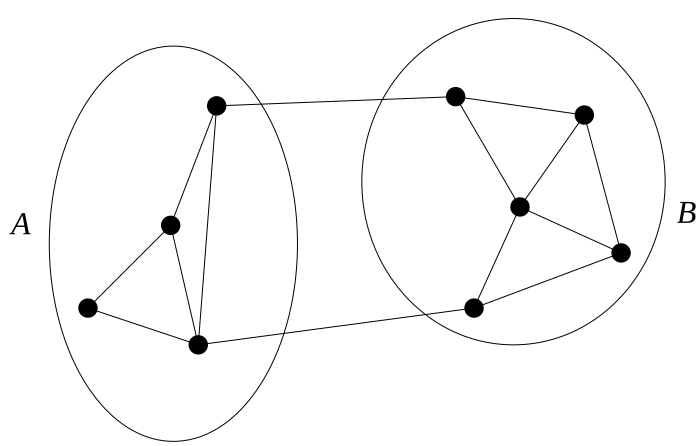
Clustering: Topics models

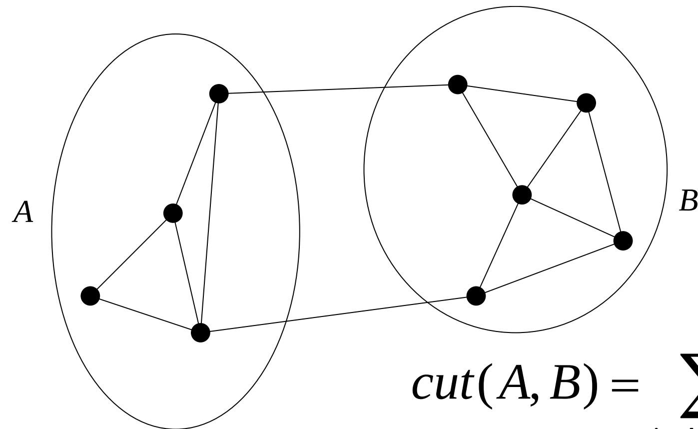
- Extension to LDA by parameterizing and sampling from $p(z|d)$
- Fewer parameters, reduced overfitting
- More general forms (NMF)

- Fixed K
- Similarity matrix

B.C. Russell, A.A. Efros, J. Sivic, W.T. Freeman, A. Zisserman. Using multiple segmentations to discover objects and their extent in image collections. CVPR 2006.
T. Tuytelaars , C.H. Lampert, M.B. Blaschko, W. Buntine. Unsupervised Object Discovery: A Comparison. IJCV 2010.

Clustering: Graph and spectral techniques





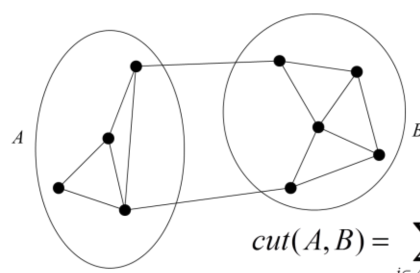
$$cut(A,B) = \sum_{i \in A,\, j \in B} s_{ij}$$

$$vol(A) = \sum_{i \in A,\, j \in A} s_{ij}$$

$$Min\ cut(A,B)\left(\frac{1}{vol(A)} + \frac{1}{vol(B)}\right)$$

Clustering: Graph and spectral techniques Example: Normalized cuts

- N segments
- S = N x N similarity matrix
- D = degree matrix ($D_{ii} = \sum_j s_{ij}$)



$$cut(A,B) = \sum_{i \in A,\, j \in B} s_{ij}$$

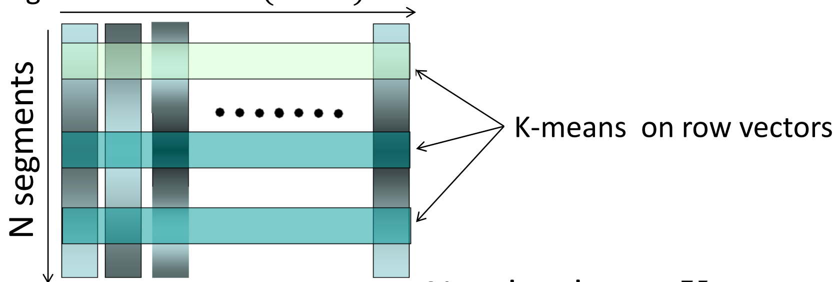
$$vol(A) = \sum_{i \in A,\, j \in A} s_{ij}$$

$$Min\ cut(A,B)\left(\frac{1}{vol(A)} + \frac{1}{vol(B)}\right)$$

Relaxed problem:
Indicator vector of (A,B) approximated by 2nd principal eigenvector of $D^{-\frac{1}{2}}(D-S)D^{-\frac{1}{2}}$

Clustering: Spectral techniques
Example: Normalized cuts

Eigenvectors of $D^{-\frac{1}{2}}(D-S)D^{-\frac{1}{2}}$

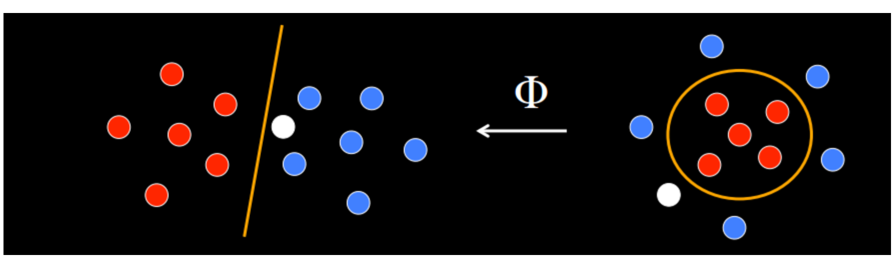


- N segments
- S = N x N similarity matrix
- D = degree matrix ($D_{ii} = \sum_j S_{ij}$)

- Need to know K
- Scalability issues

Y.J. Lee, K. Grauman. Foreground Focus: Unsupervised Learning from Partially Matching Images. IJCV. 2009.

Discriminative clustering



$$\min_{\substack{w \in \mathbb{R}^{K \times d},\\ b \in \mathbb{R}^K \\ y \in \{0,1\}^{N \times K} \\ y 1_K = 1_N}} \frac{1}{N} \sum_{n=1}^{N} \ell(y_n, w\phi(x_n) + b) + \frac{\lambda}{2K} \|w\|_F^2$$

http://www.di.ens.fr/willow/events/cvml2013/materials/slides/thursday/coseg13.pdf

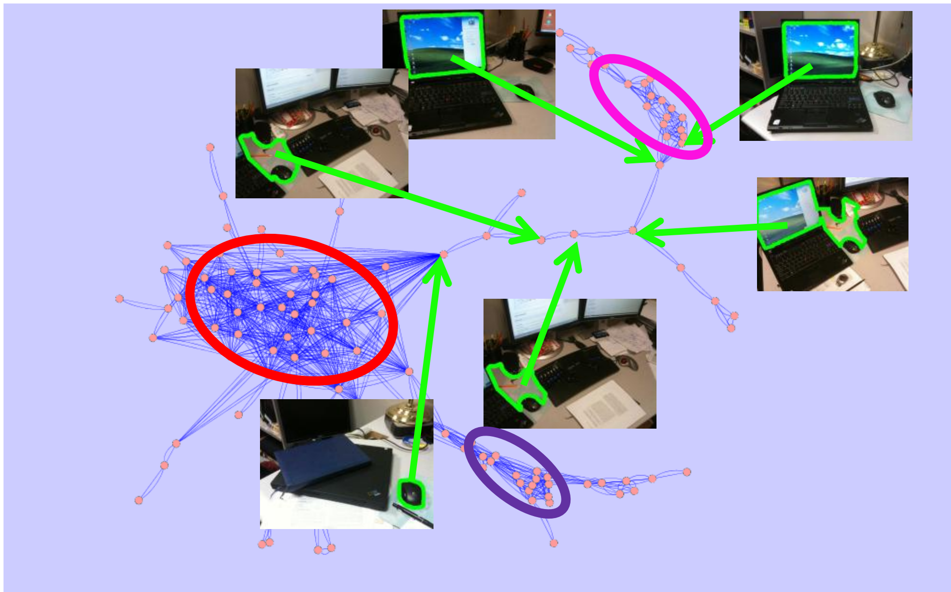
Issues and limitations

- Requires complete graph (or at least computing all the affinities for all the edges)
  - Alternative: Retain edges with high affinity only; compute subset of edges only → Sparse graph
- Requires fixing the number of clusters K in advance
  - Alternative: Automatic stopping criterion

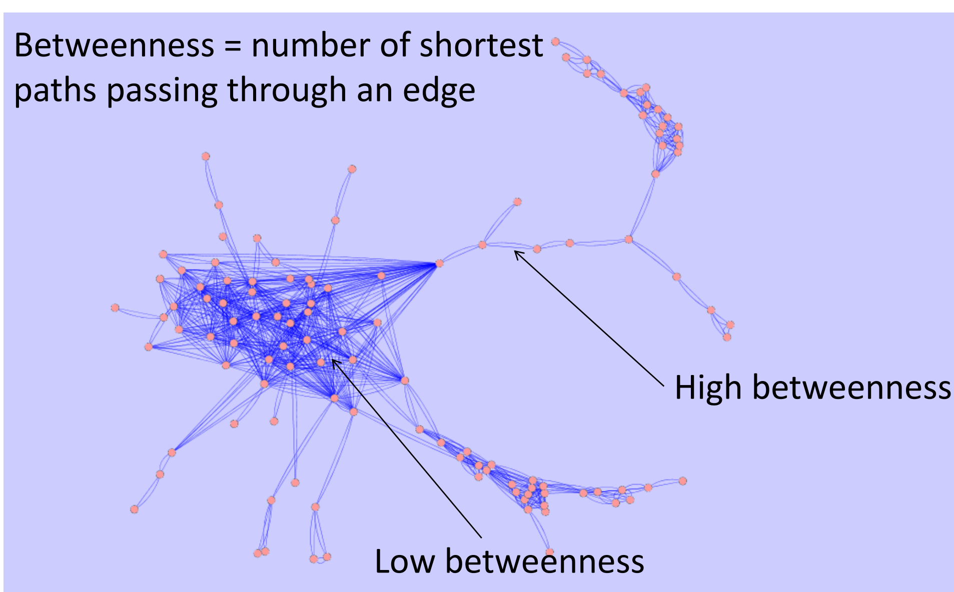
Clustering: Graph analysis

- Idea:
  - Don't try to explain the whole graph by a coverage of clusters. Instead extract a few reliable "influential" subgraphs by analogy with network analysis.
  - Don't fix the number of clusters K in advance
- Examples:
  - Link analysis [G. Kim, C. Faloutsos, M. Hebert. Unsupervised modeling of object categories using link analysis techniques. CVPR 2008.]
  - Graph sampling [N. Payet, S. Todorovic. From a set of shapes to object discovery. ECCV 2010.]
  - Community discovery [H. Kang, T. Kanade, M. Hebert. Discovering object instances from scenes of daily living. ICCV 2011.]

Example graph

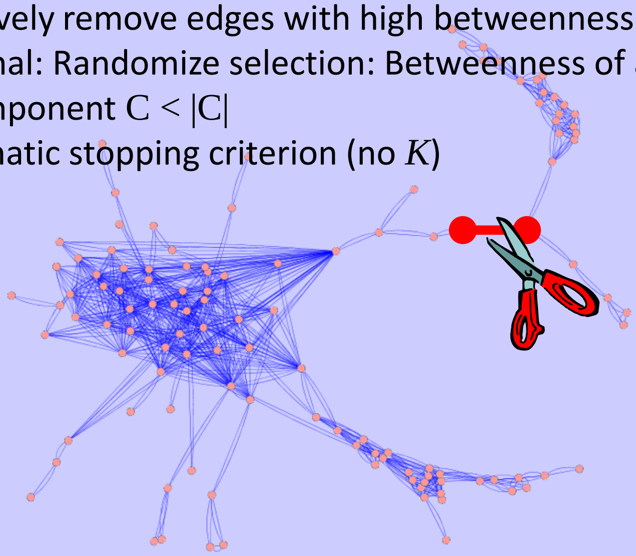


Community discovery

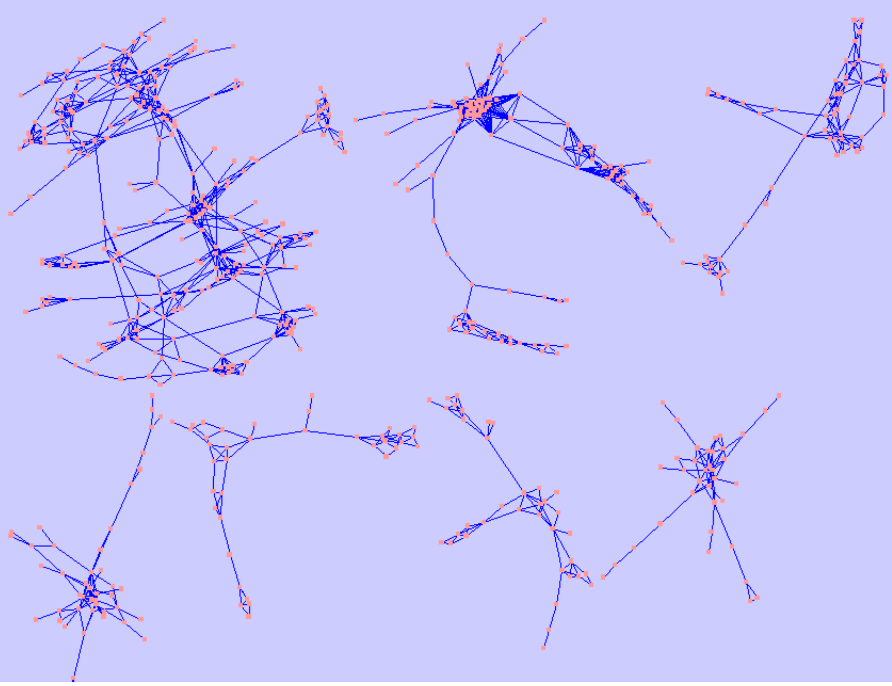


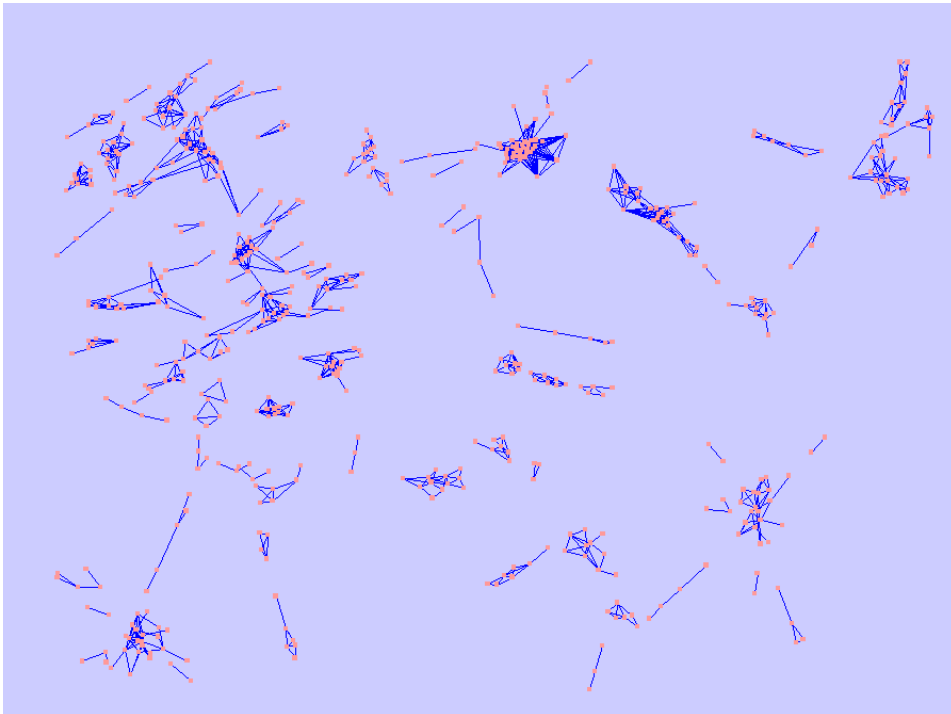
Community discovery

Iterativelyremove edges with high betweenness
Optional: Randomize selection: Betweenness of any edge of component $C < |C|$
Automatic stopping criterion (no K)

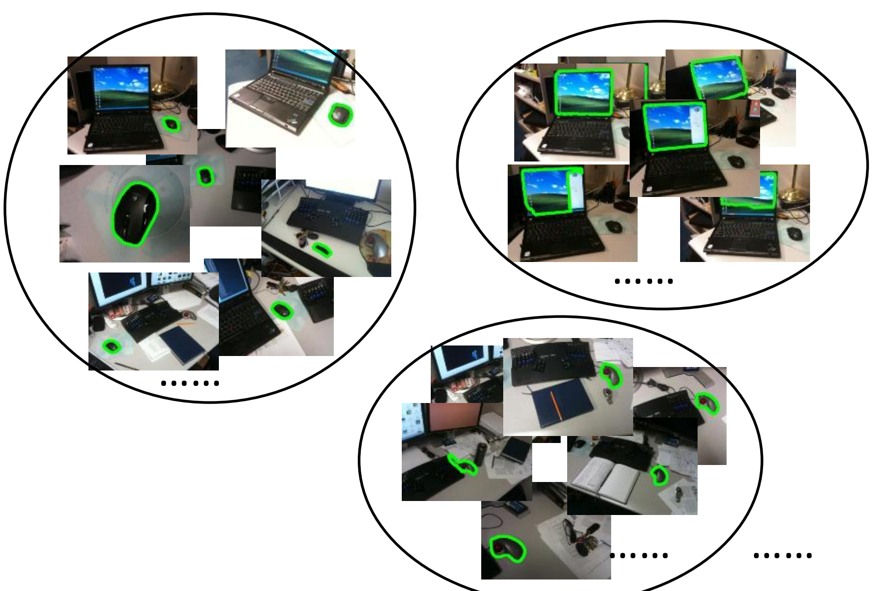


[Freeman 1979; Tyler et.al. 2003]



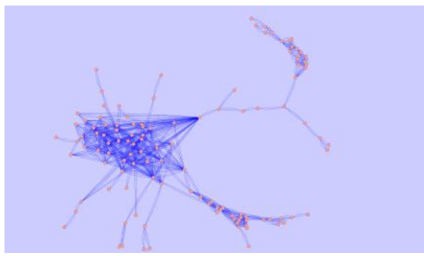


Example: Groups of similar segments



Community discovery

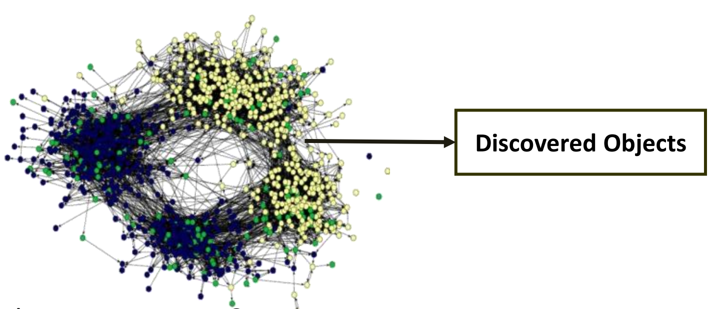
- Appropriate for graphs following power law
- Empirically, graphs obtained by thresholding appearance/shape similarity follow that model



Average degree = 3.1
Density = 16%

- Efficient with respect to graph size

J. Leskovec, K.J. Lang, and M. Mahoney. Empirical comparison of algorithms for network community detection. Intern. Conf. on World wide web, WWW '10. 2010.
J.R. Tyler, D.M. Wilkinson, B.A. Huberman. Automated discovery of community structure within organizations. Intern. Conf. on Communities and Technologies.. 2003.

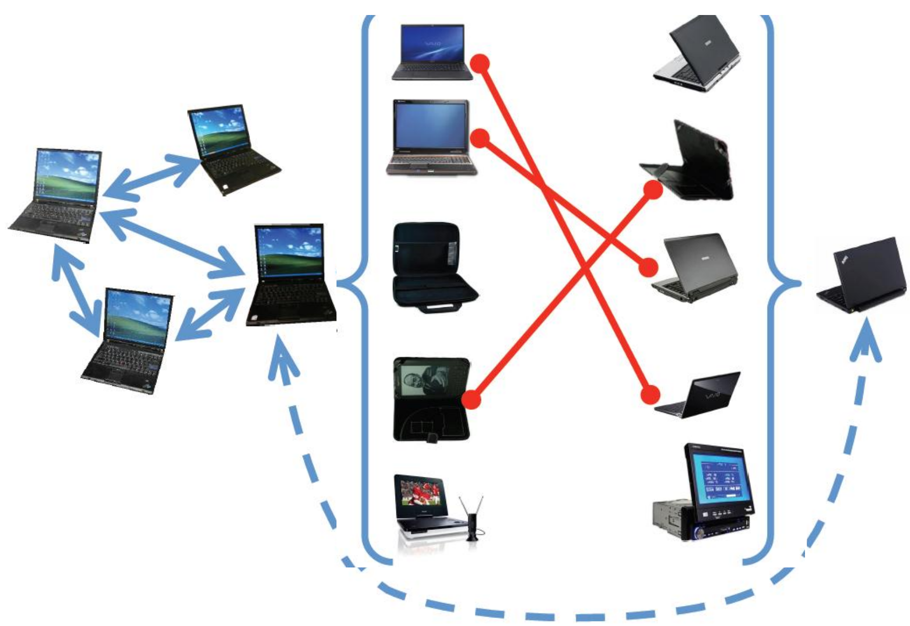


- How to cluster segments?
  - Topic model
  - Graph/spectral clustering
  - Graph analysis
- How to define robust affinities?
  - Robust affinities from contextual information
- How to select/prune candidate segments and edges to reduce graph?
  - Select segments using learned objectness
  - Prune graph edges using domain constraints

Affinity definition: Learn from known categories

- Idea:
  - Define affinity between regions in a way that maximizes efficiency of discovery by
    - Discarding irrelevant graph edges
    - Adding/reinforcing edges that could not be found from appearance alone

The missing link problem

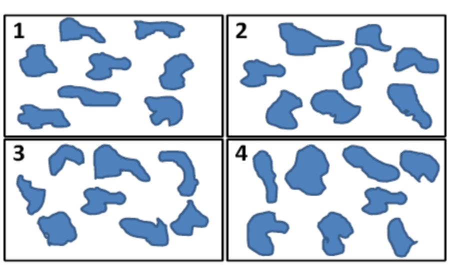


Affinity definition: Learn from known categories

- Idea:
  - Define affinity between regions in a way that maximizes efficiency of discovery by
    - Discarding irrelevant graph edges
    - Adding/reinforcing edges that could not be found from appearance alone

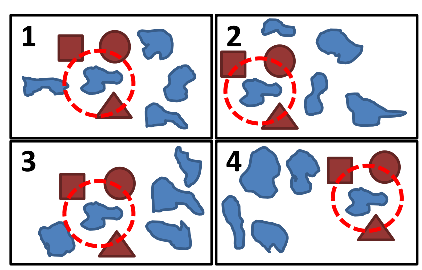
- Examples:
  - Learn similarity metric from known categories [C. Galleguillos, B. McFee, S. Belongie, G. Lanckriet. From region similarity to category discovery. CVPR 2011.]
  - Use known categories as context to unfamiliar categories
  - Completing missing links by using prior data

Robust affinities: Context from known categories



Previous methods:
Cluster images/regions based on their appearance

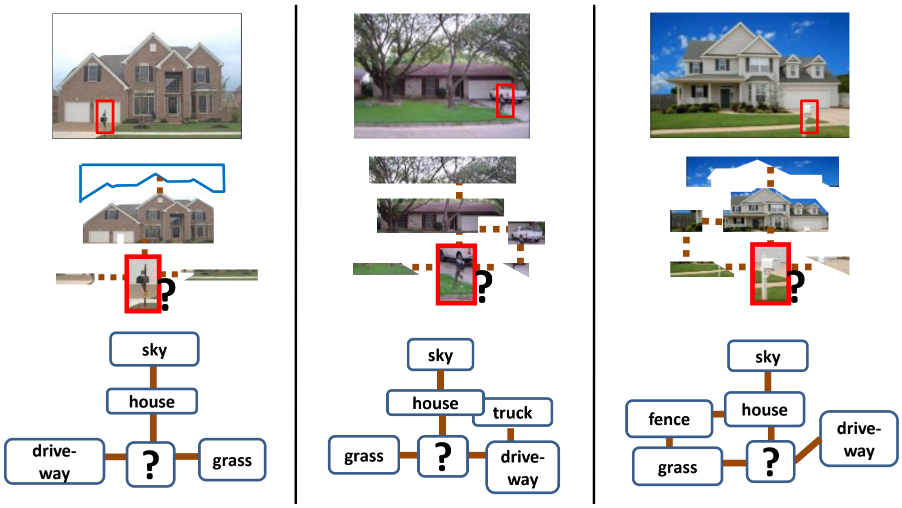
[Sivic et al. 2005, Russell et al. 2006, Liu et al. 2006, Lee et al. 2008, Kim et al. 2008,…]



Insight:
Let *familiar* objects serve as context for *unfamiliar* objects

[Y. Lee and K. Grauman. Object-graphs for context-aware category discovery. CVPR 2010.

Intuition: context-aware discovery



[Lee & Grauman, Object-Graphs, CVPR 2010, TPAMI 2011]

General idea

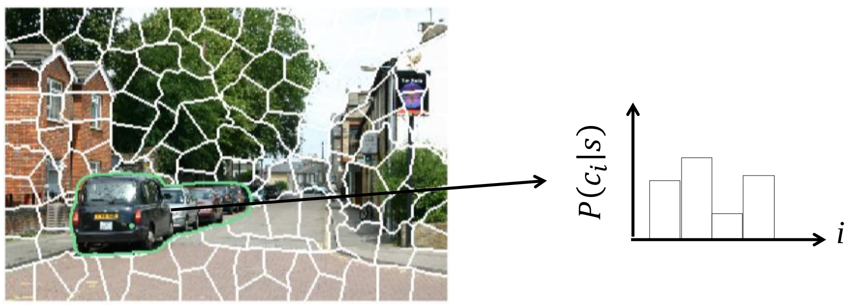
An unknown region within an image

Model the topology of category predictions relative to the unknown (unfamiliar) region

Example from Y. Lee

Assumptions

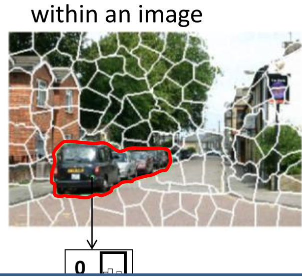
- N known categories
- Classifier produces posterior $P(c_i|s)$ for each segment s and category c_i



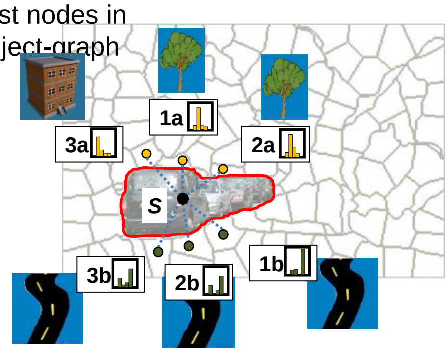
Example from Y. Lee

Object-graphs

An unknown region within an image



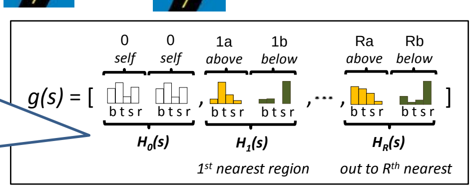
Closest nodes in its object-graph



Average posteriors of i nearest regions:

$$H_i(s) = [\sum_{j=1}^{i} P(m_{j_a}(s)|\ c_1), \ldots, \sum_{j=1}^{i} P(m_{j_a}(s)|\ c_N),$$
$$\sum_{j=1}^{i} P(m_{j_b}(s)|\ c_1), \ldots, \sum_{j=1}^{i} P(m_{j_b}(s)|\ c_N)]$$

$m_{j_a}(s), m_{j_b}(s)$: j'th superpixel above, below



Example from Y. Lee

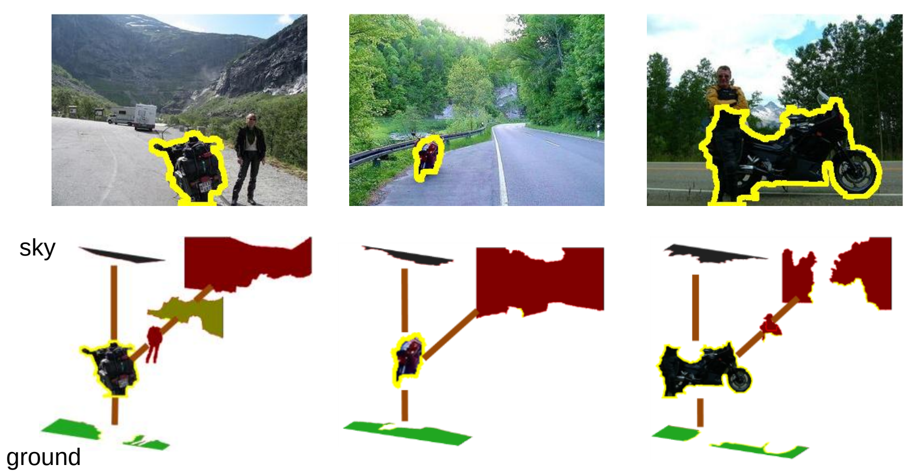
Example object-graphs

unknown building sky road

- Colors indicate the predicted known category (max posterior)

Example from Y. Lee

Example 3D object-graphs



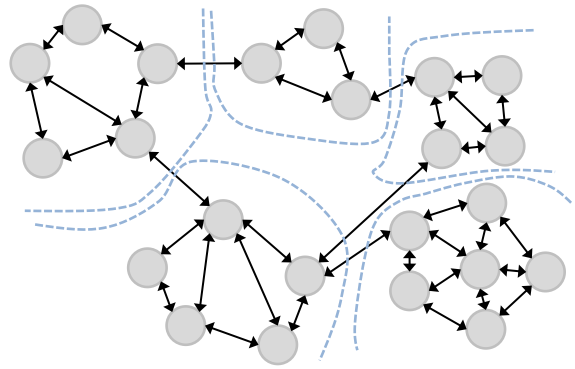
unknown tree road sky grass

- Colors indicate the predicted known category (max posterior)

Example from Y. Lee

Clusters from region-region affinities

Unknown Regions



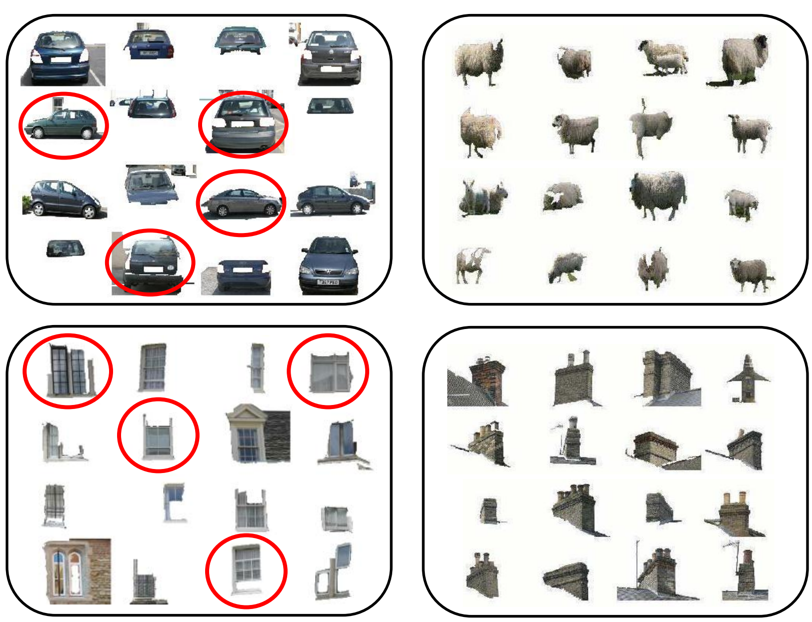
$$K(s_i, s_j) = K_{app}(s_i, s_j) + K_{obj-graph}(s_i, s_j)$$

$$\sum_u K_{\chi^2}(a_u(s_i), a_u(s_j)) \qquad K_{\chi^2}(g(s_i), g(s_j))$$

Appearance channel

Example from Y. Lee

Examples



Example from Y. Lee

Observations

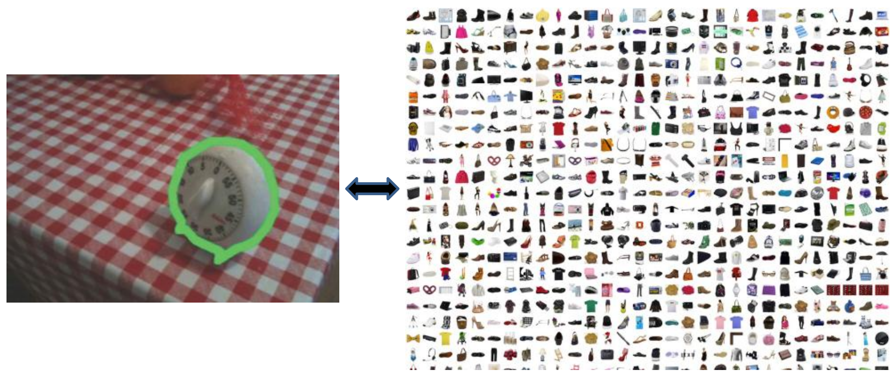
- Added term on context from known categories allows for grouping of segments with distinct appearance
- How to select the segments to cluster (“unknown segments”)?
  - Don’t want to disturb segments with high confidence from know categories
  - Need to limit the number of segments for the clustering

Non-parametric approach

- Classifiers may not be available for known categories
- (Large) prior data may be available
- Can we use it directly instead of trained classifiers

Assumption

- Large prior dataset
- (Approximate) matching based on appearance features

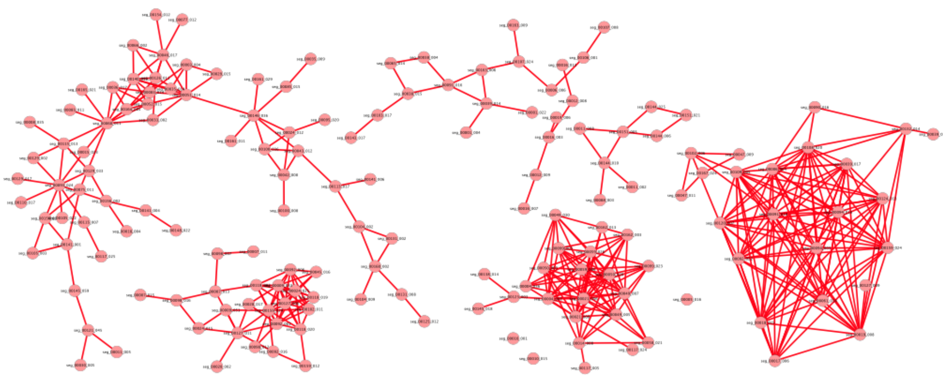


H. Kang, M. Hebert, A. A. Efros, T. Kanade. Connecting Missing Links: Object Discovery from Sparse Observations. ECCV 2012.

- Add term data term to affinity:

$$c_d(s_i, s_j) = \frac{\Phi(s_i) \cap \Phi(s_j)}{K}$$

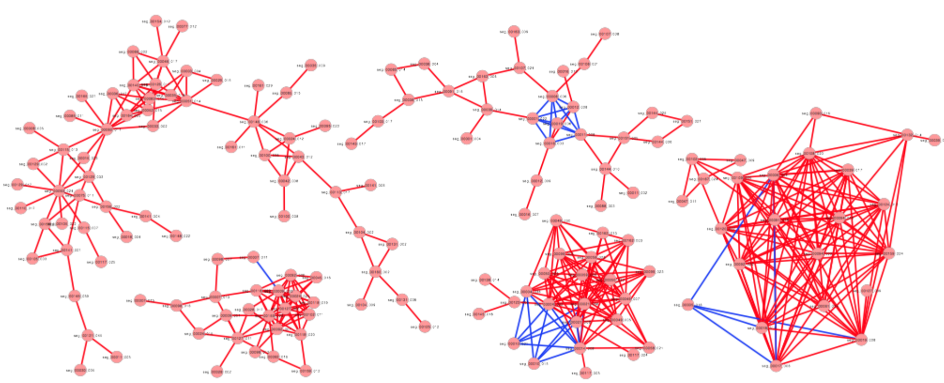
Adding data-driven similarity links



Segment Matching Links

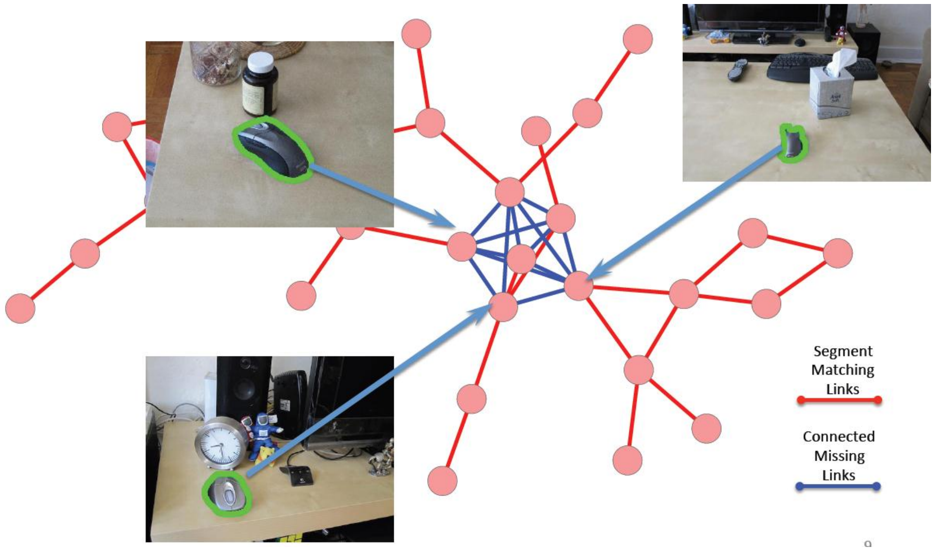
Connected Missing Links

Adding data-driven similarity links



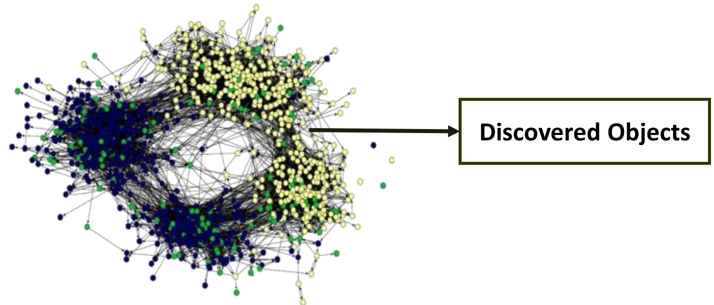
Segment Matching Links

Connected Missing Links



Comments

- Domain-specific
- Better suited to instance discovery

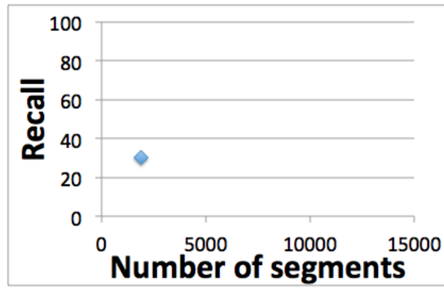


- How to cluster segments?
  - Topic model
  - Graph/spectral clustering
  - Graph analysis
- How to define robust affinities?
  - Robust affinities from contextual information
- How to select/prune candidate segments and edges to reduce graph?
  - Use known categories
  - Select segments using learned objectness
  - Prune graph edges using domain constraints

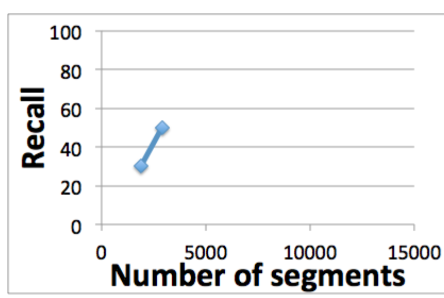
Why controlling the number of segments (if possible!)?

- More candidate segments lead to better recall (more choices) but lower precision (more choices!)
- Chicken-and-egg problem: If we knew which regions are likely to be object we could start with better candidates……but that's the point of discovery…
- General idea: Independent guess at likelihood of being an object segment

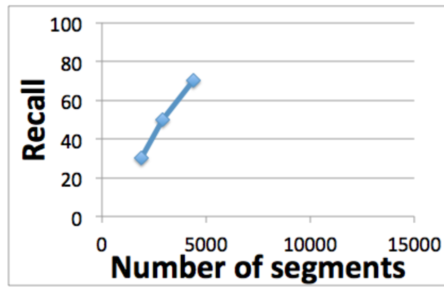
Increase object segmentation recall



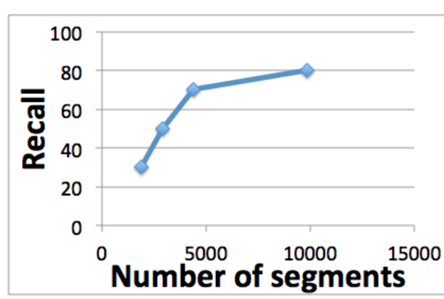
Increase object segmentation recall



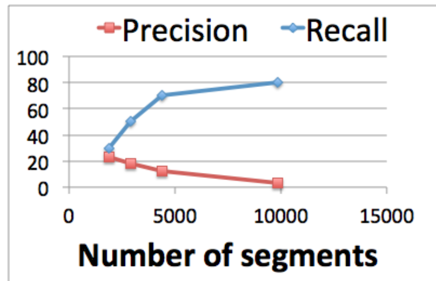
Increase object segmentation recall



Increase object segmentation recall



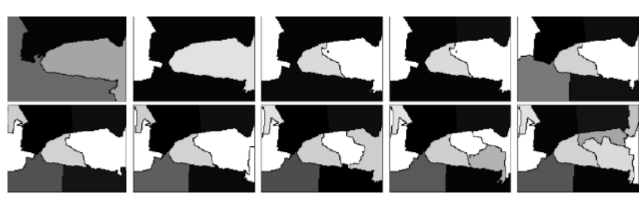
Increase object segmentation recall



Special case: Known categories

- Use the entropy of the distribution of class posterior over known categories as criterion to select segments

$$-\sum_i P(c_i|s)\log P(c_i|s)$$



Known = sky, road

Example from Y. Lee

Segment selection: Learned objectness

- Idea:
  - Reduce the size of the graph by selecting only "object-like" segments based on a separately learned objectness measure

  → Reduces the number of nodes

Objectness/Object saliency

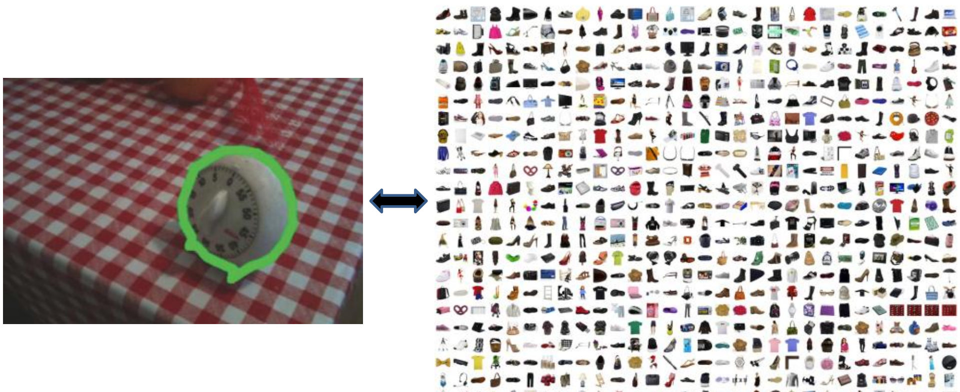
- Parametric: Model explicitly learned from training data
  - Region objectness [Endres, Hoiem, ECCV 2010]

  - Box objectness [Alexe, Deselaers, Ferrari, CVPR 2010, PAMI 2012]

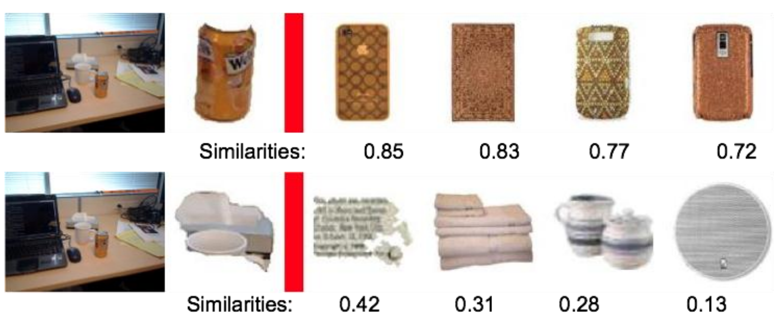
  - Saliency [Zhu et al. CVPR 2012, Feng et al. ICCV2011]

Non-parametric objectness

- Non-parametric: Consistency of matches with a large repository of prior images
- Object regions tend to match more consistent regions from the dataset



Criterion 1: Average similarity



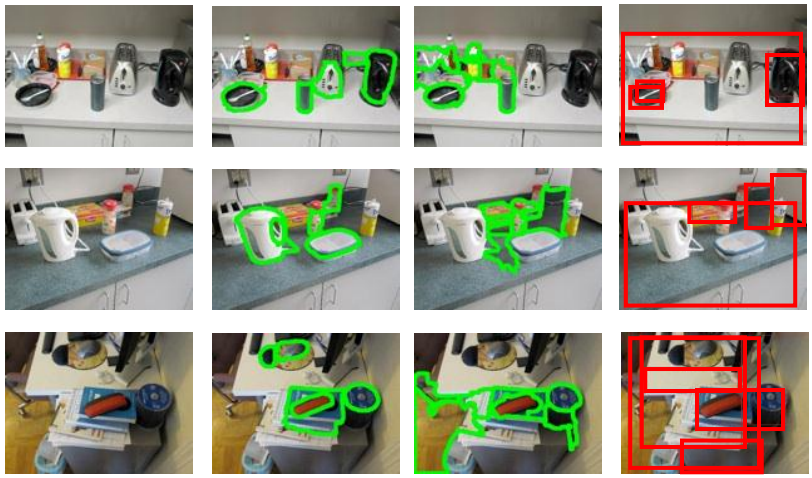
$$a(R) = \frac{1}{K}\sum_{j=1}^{K} S_a(R, R_j^e).$$

Criterion 2: Consistency Visual + Metadata (categories)

$$m(R) = median_{j=1}^{K}(median_{k=1,k\neq j}^{K}(S_e(R_j^e, R_k^e))).$$

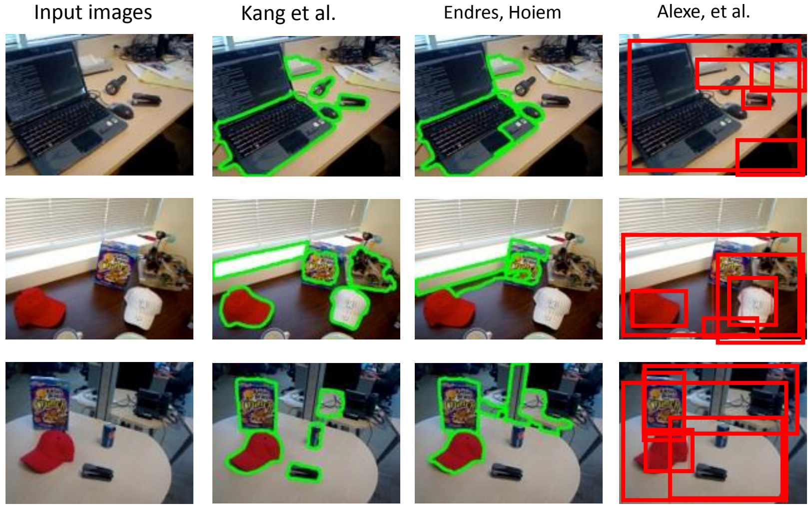
Qualitative comparisons: CMU ADL

Input images | Kang et al. | Endres, Hoiem | Alexe et al.

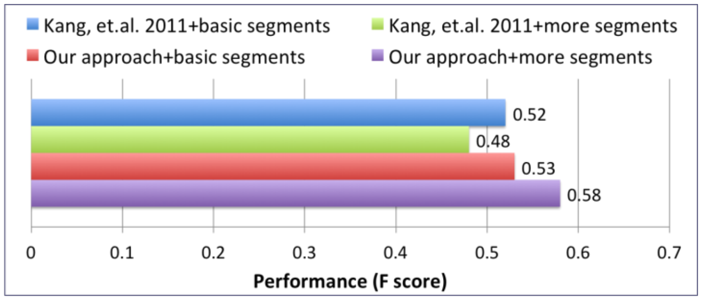


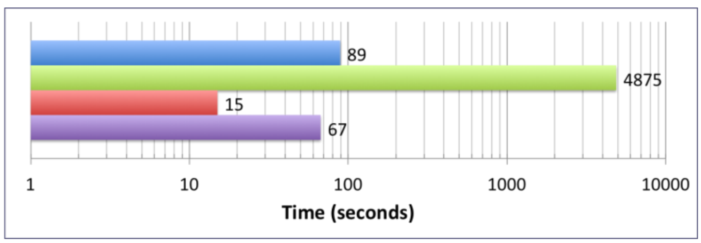
Qualitative comparisons: UW RGB-D

Input images | Kang et al. | Endres, Hoiem | Alexe, et al.

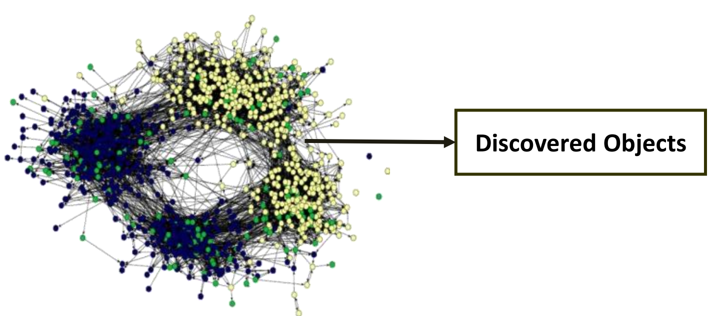


Object discovery with segments of high objectness





Basic segments: 4390 segments; More segments: ~17,000 segments



- How to cluster segments?
  - Topic model
  - Graph/spectral clustering
  - Graph analysis
- How to define robust affinities?
  - Robust affinities from contextual information
- How to select/prune candidate segments and edges to reduce graph?
  - Select segments using learned objectness
  - Prune graph edges using domain constraints

Pruning the graph: Using domain constraints

- Idea:
  - Reduce the size of the graph by enforcing domain constraints early

  → Reduces the number of edges

Pruning the graph: Using domain constraints

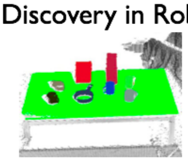
General Graph / Machine Learning tools

Obj. Discovery in CV

[Sivic *et al*, 2005] [Russell *et al*, 2006]
[Endres *et al*, 2009] [Kang *et al*, 2011]

Specific Purpose Systems

Obj. Discovery in Robotics



[Morwald *et al*, 2010] [Marton *et al*, 2010]
[Herbst *et al*, 2011] [Mishra *et al*, 2011]

Slide from A. Collet

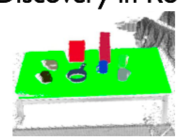
General Graph / Machine Learning tools

Specific Purpose Systems

Obj. Discovery in CV

[Sivic et al, 2005] [Russell et al, 2006]
[Endres et al, 2009] [Kang et al, 2011]

Obj. Discovery in Robotics



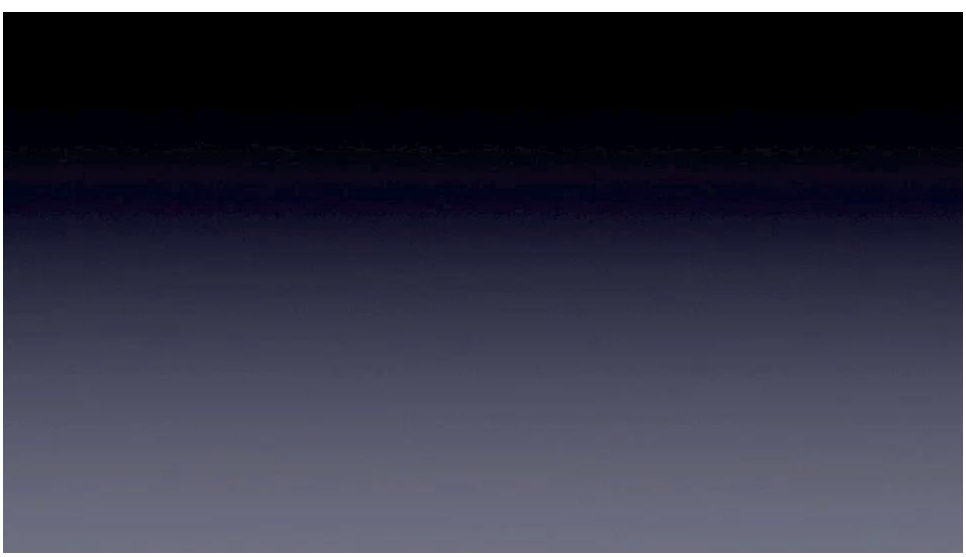
[Morwald et al, 2010] [Marton et al, 2010]
[Herbst et al, 2011] [Mishra et al, 2011]

\- Domain Knowledge +

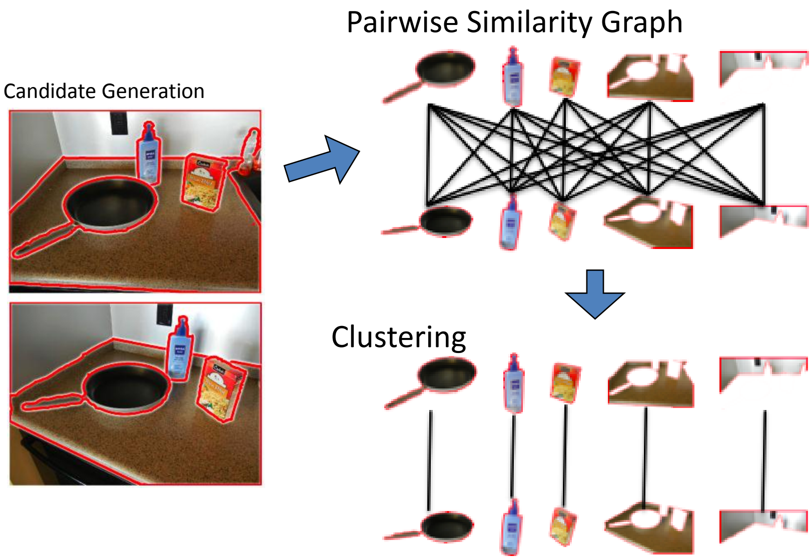
Insight: Formalize the concept of *"*Domain Knowledge"

Result: Common Formulation for Object Discovery

Slide from A. Collet



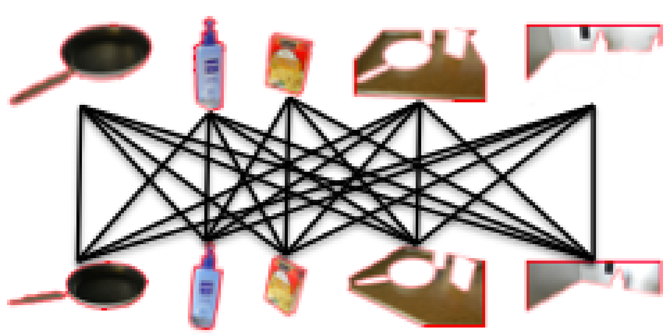
Graph-based discovery



Slide from A. Collet

Similarity graphs

-No domain knowledge-

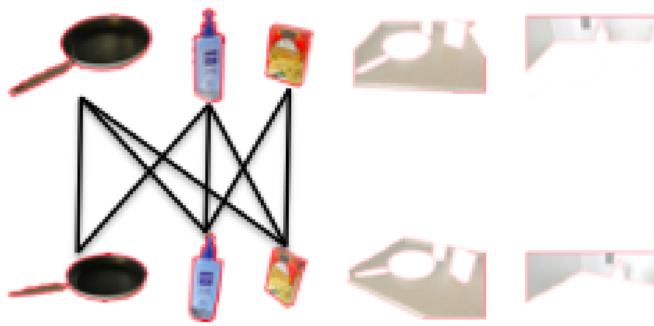


72

Slide from A. Collet

Constrained similarity graphs

"Objects on support surfaces"

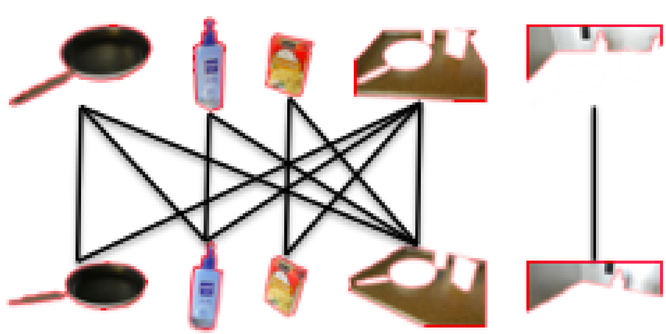


73

Slide from A. Collet

Constrained similarity graphs

"Scene is static" (3D overlap)



74

Slide from A. Collet

Constrained similarity graphs

"Objects are planar"

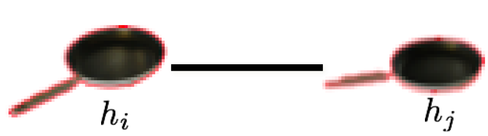
Slide from A. Collet

Constraints

- Bernoulli distribution: *Measurable* yes/no question about nodes or edges with probability of success *p*

$$\Theta^{\mathrm{n}} : h_i \mapsto \{0,1\} \qquad P(\Theta^{\mathrm{n}}(h_i) = 1) = P_{\Theta^{\mathrm{n}}}(h_i) = p$$

$$\Theta^{\mathrm{e}} : h_i, h_j \mapsto \{0,1\} \quad P(\Theta^{\mathrm{e}}(h_i, h_j) = 1) = P_{\Theta^{\mathrm{e}}}(h_i, h_j) = p$$



Slide from A. Collet

Examples of Constraints

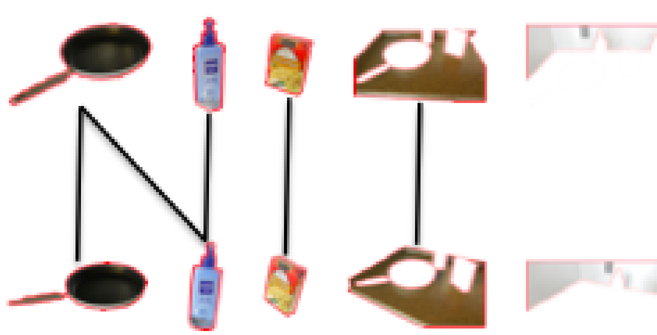
- "are candidates h_i and h_j similar in appearance?"
  - Measure $s(h_i, h_j) \in [0, 1]$
    $\Theta^{\mathrm{e}}_{\mathrm{app}} = 1$ with $p = s(h_i, h_j)$

- "is candidate h_i on a support surface?"
  - Detect support
  - Check if h_i is supported
    $\Theta^{\mathrm{n}}_{\mathrm{support}} = 1$ with $p = 1$

Slide from A. Collet

Constraint Expressions

"Scene is static" AND
("Objects supported" OR "Objects are planar")



Slide from A. Collet

The Logic of Constraints

- Negation

$$\neg\Theta^{\mathrm{n}}: \quad P_{\neg\Theta^{\mathrm{n}}}(h_i) = 1 - p$$

- Conjunction

$$\Theta_a^{\mathrm{n}} \wedge \Theta_b^{\mathrm{n}}: \quad P_{\Theta_a^{\mathrm{n}} \wedge \Theta_b^{\mathrm{n}}}(h_i) = P_{\Theta_a^{\mathrm{n}}}(h_i) P_{\Theta_b^{\mathrm{n}}}(h_i)$$

- Disjunction

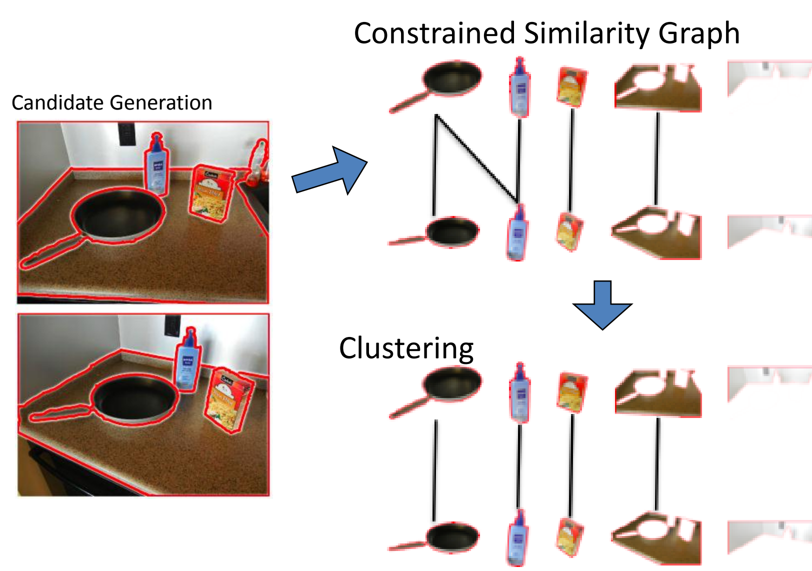
$$\Theta_a^{\mathrm{n}} \vee \Theta_b^{\mathrm{n}}: \quad P_{\Theta_a^{\mathrm{n}} \vee \Theta_b^{\mathrm{n}}}(h_i) = 1 - P_{\neg\Theta_a^{\mathrm{n}} \wedge \neg\Theta_b^{\mathrm{n}}}(h_i)$$

"Scene is static" AND
("Objects lie on tables" OR "Objects are planar")

$$\Theta_{\mathrm{static}} \wedge (\Theta_{\mathrm{tables}} \vee \Theta_{\mathrm{planar}})$$

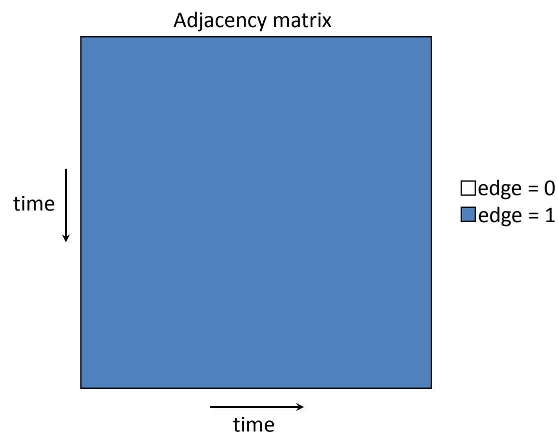
Slide from A. Collet

Discovery with domain knowledge



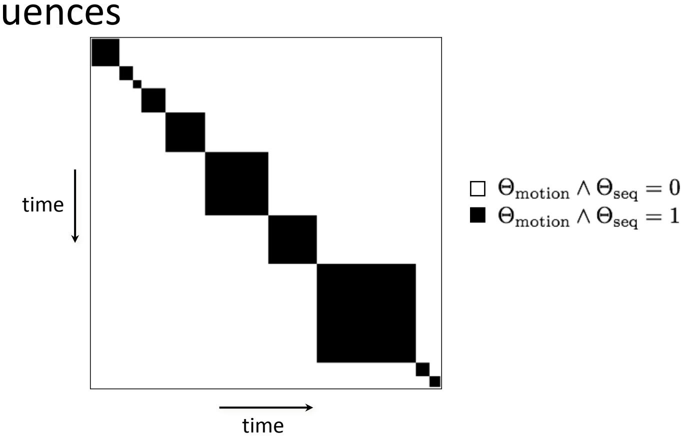
Slide from A. Collet

- No constraints: $O(n^2)$ pairwise similarities

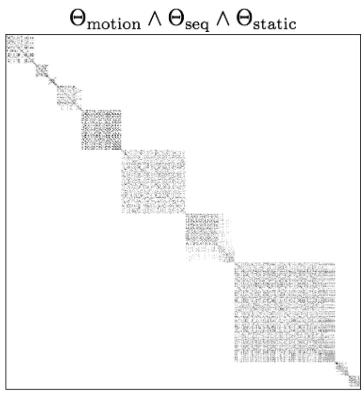


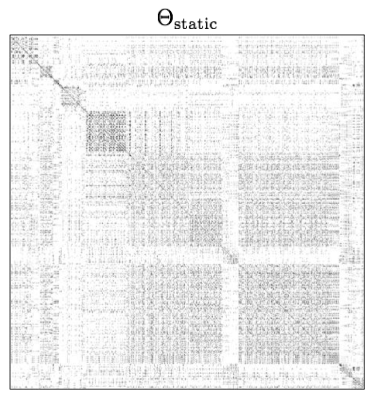
Motion and sequencing

- Θ_{motion} : only sample data stream if there is motion
- Θ_{seq} : data stream is ordered, forms short sequences

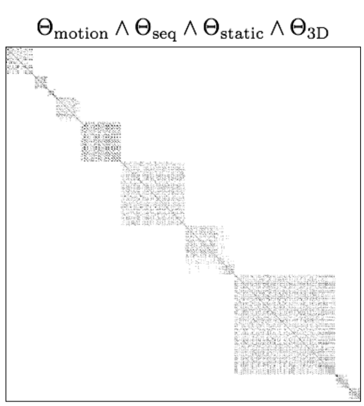


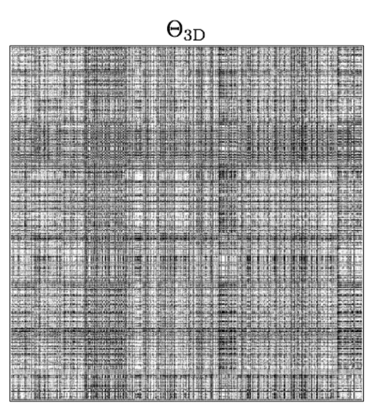
- Θ_{static} : 3D overlap between candidates within sequence





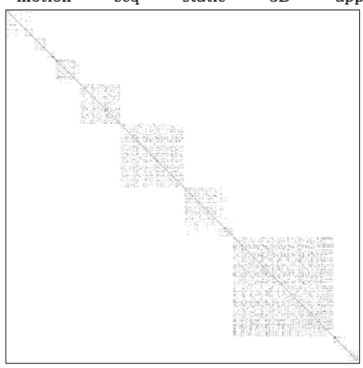
Θ_{3D} : Shape similarity between candidates



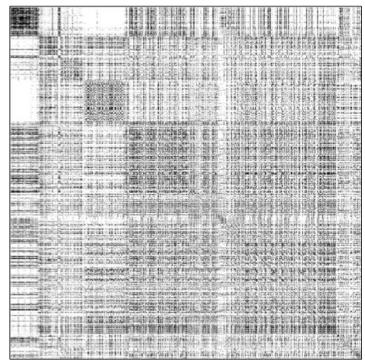


Θ_{app} : Appearance similarity between candidates

$\Theta_{\text{motion}} \wedge \Theta_{\text{seq}} \wedge \Theta_{\text{static}} \wedge \Theta_{\text{3D}} \wedge \Theta_{\text{app}}$

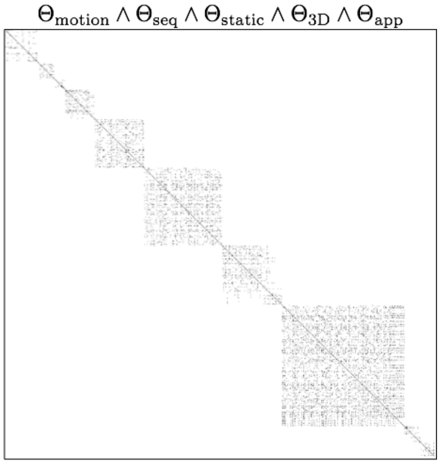


Θ_{app}



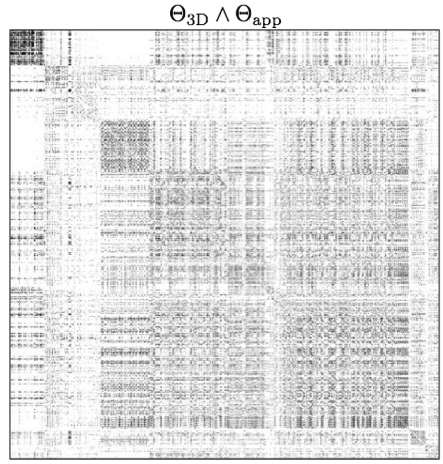
Impact of constraints

$\Theta_{\text{motion}} \wedge \Theta_{\text{seq}} \wedge \Theta_{\text{static}} \wedge \Theta_{\text{3D}} \wedge \Theta_{\text{app}}$



Visual Similarity + Metadata

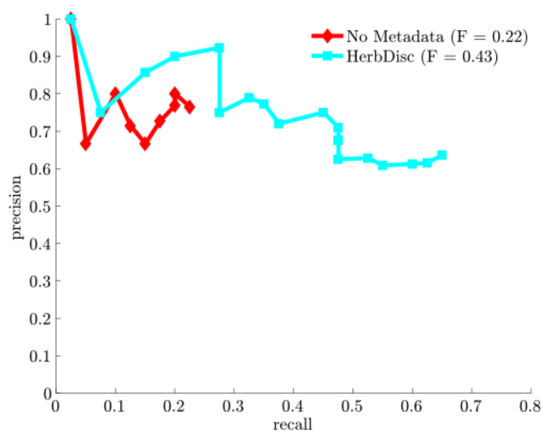
$\Theta_{\text{3D}} \wedge \Theta_{\text{app}}$

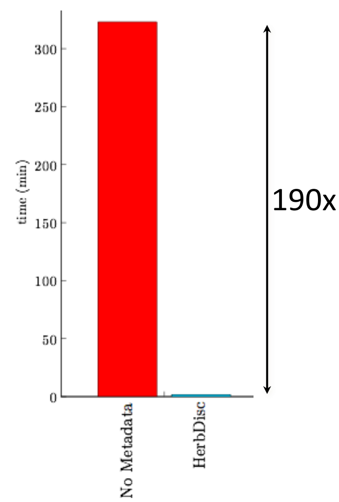


Visual Similarity

Robotic Object Discovery

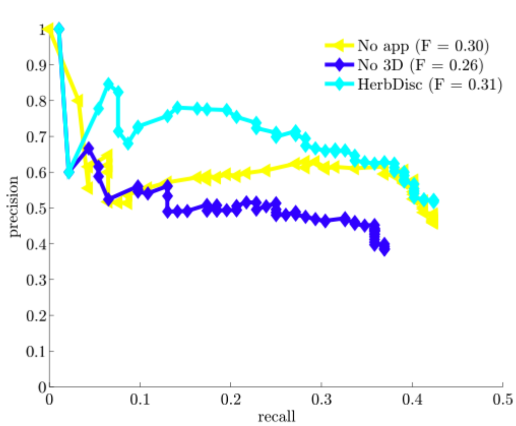
Example evaluation (small)





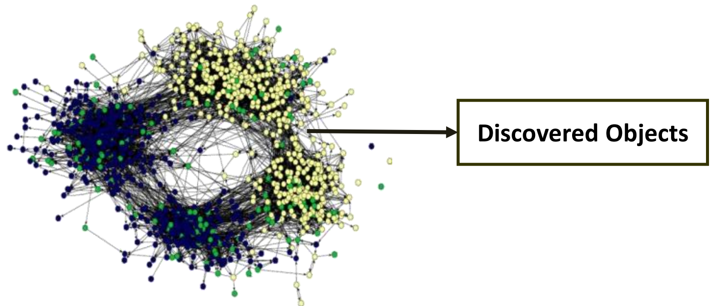
Example from A. Collet

Example evaluation (large)



- 521234 frames
- Discovered 864 objects
  - 283 correct
  - 166 valid
  - 415 invalid
- Processing time: 18 m 34 s
  - 58682 candidates
  - 431121 edges

88



- How to cluster segments?
  - Topic models
  - Graph/spectral clustering
  - Graph analysis
- How to define robust affinities?
  - Robust affinities from contextual information
- How to select/prune candidate segments and edges to reduce graph?
  - Select segments using learned objectness
  - Prune graph edges using domain constraints