Combining Detection, Recognition and Segmentation

Presented by:

Ira Kemelmacher-Shlizerman

Sharon Alpert

For the Advanced Topics in Computer Vision course
Spring 2006

Words Matching in Ascii

Find the word sleeping in the string:

“I have never taken any exercise except sleeping and resting”

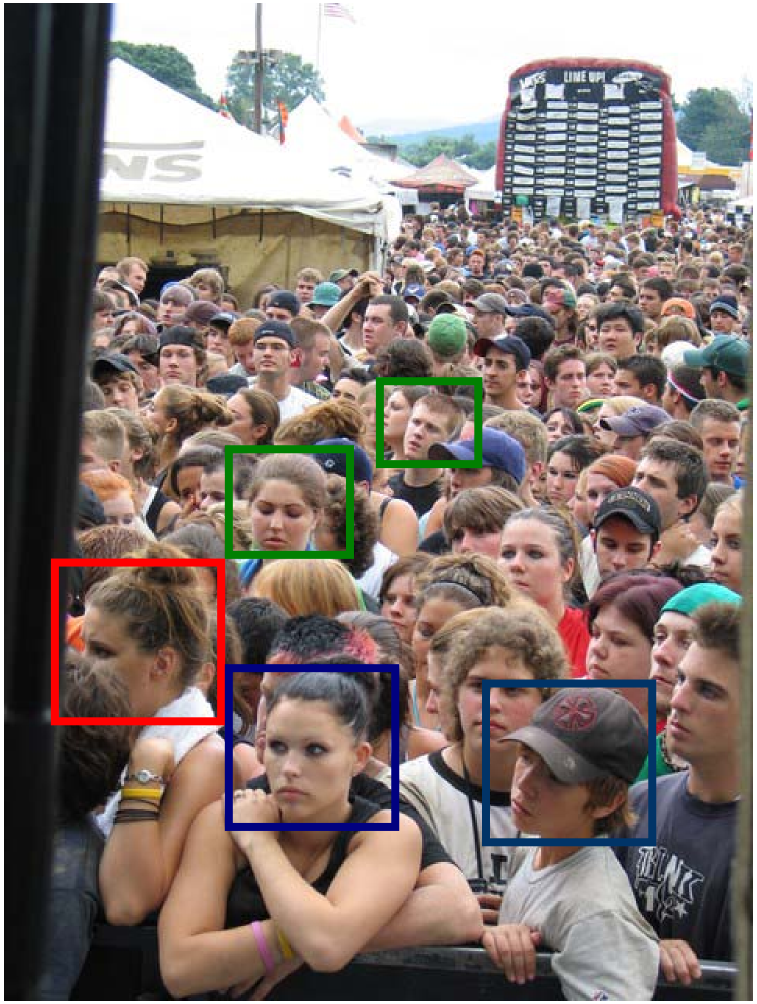
Mark Twain

Simply go over the string and match each substring of length 8.

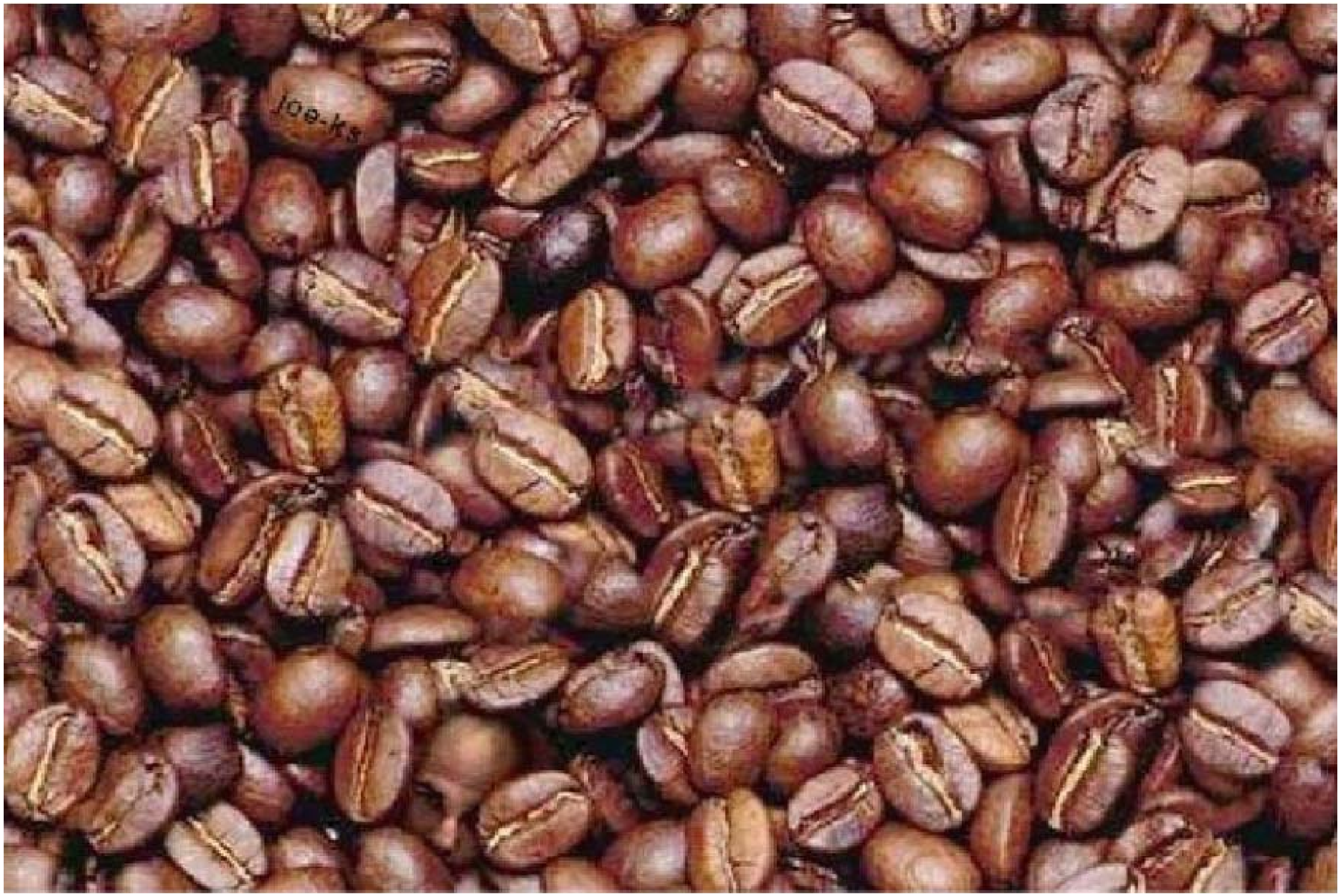
But what if we want to detect a face?

Detection/recognition challenges (partial list)

- Partial occlusions.
- Variation in pose.
- intra-class variations
- Lighting, scale, deformations, background clutter etc.



Try to find the face in this image

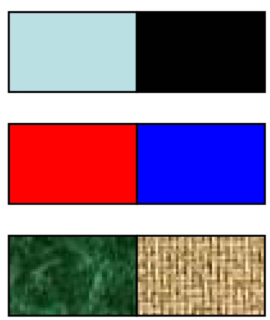


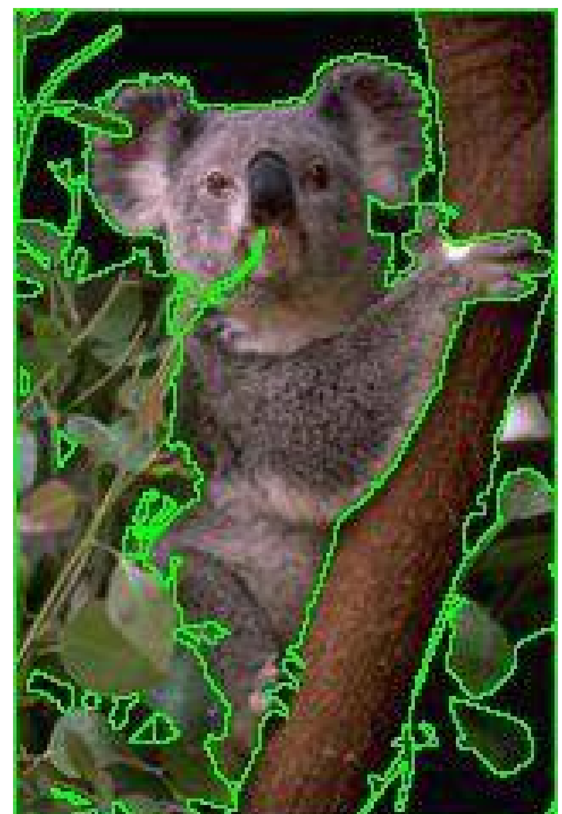
Segmentation

Group homogeneous areas

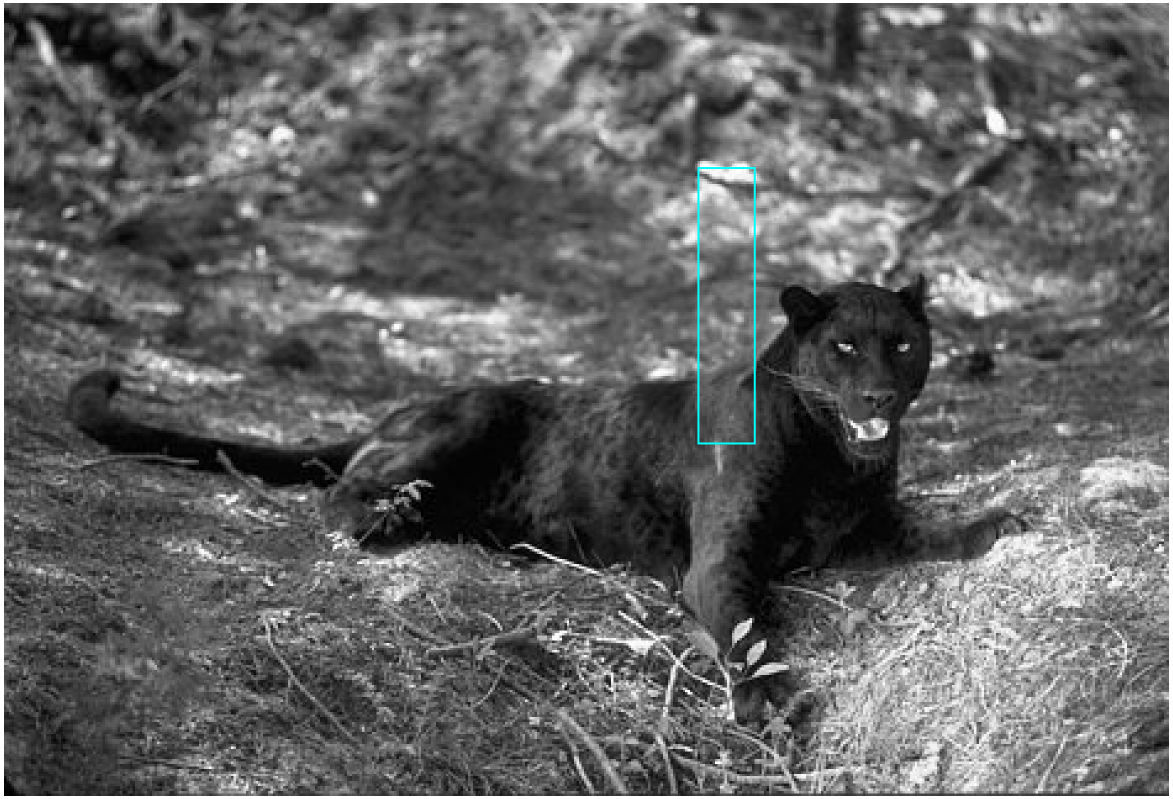
Use low-level cues:

Intensity
Color
Texture





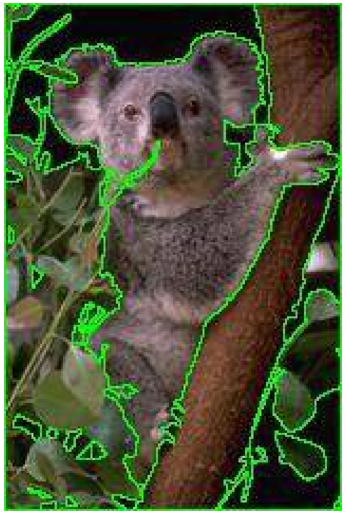
Segmentation - where is the edge?



Low-level local cues are not enough to get meaningful segmentation

Segmentation & Recognition

- Segmentation ⇄ recognition



Segmentation helps recognition

Recognized objects constrain segmentation

Labeling & Segmentation

Knowing the image type or the labeling

Can constraint the segmentation

Labeling & Recognition

Likely: CAR **Unlikely: TOASTER**

We will talk about

- Recognition/Detection

  Scene categories recognition

  S. Lazebnik et al.

  Given an explicit image of the model

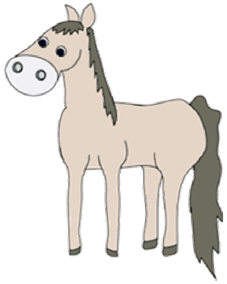
  V. Ferrari et al.

- Segmentation with Recognition

  E. Borenstein and S. Ullman

  E. Borenstein , E. Sharon and S. Ullman

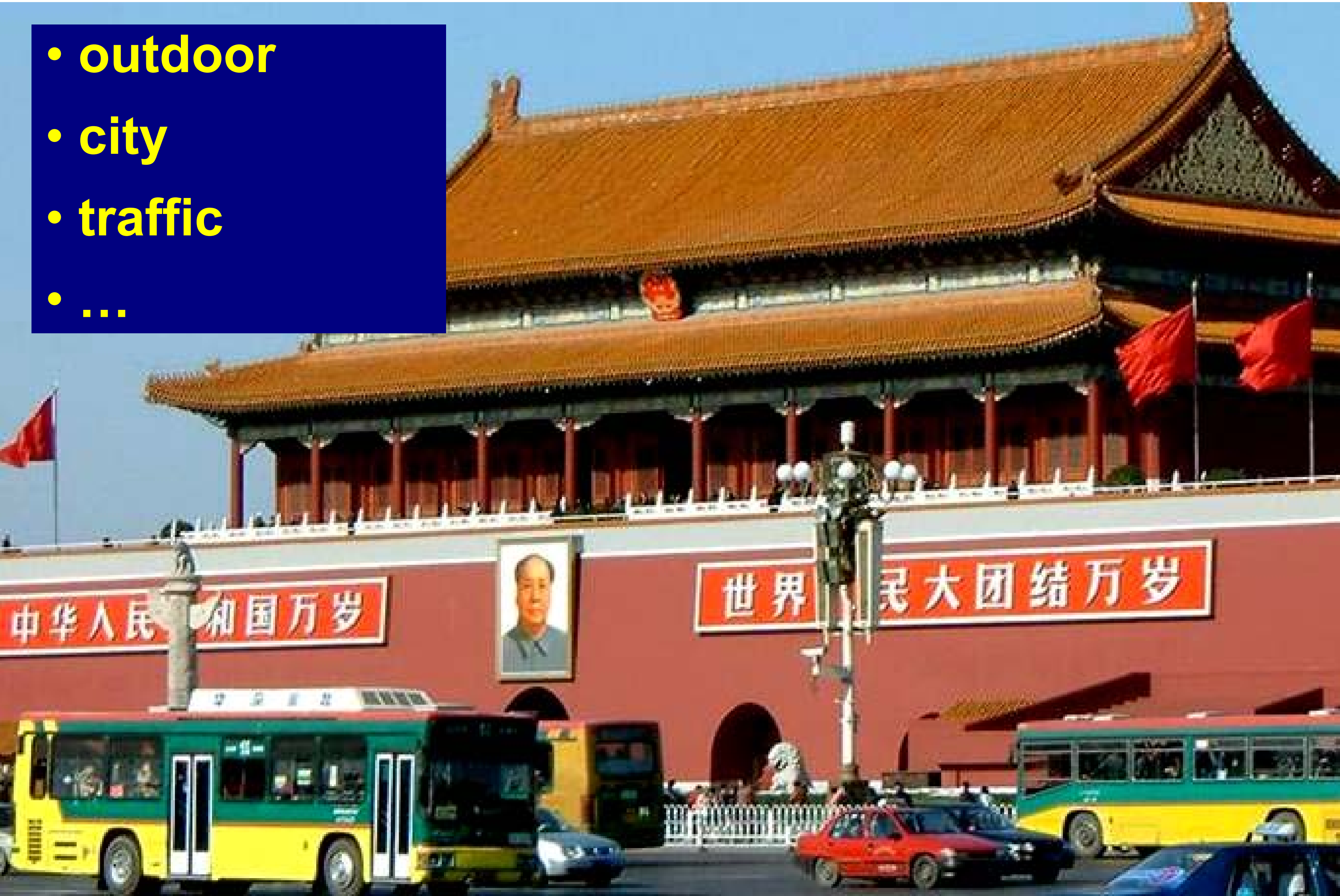
  A. Levin and Y. Weiss



(From Iccv 2005 tutorial)

Scene and context categorization

- **outdoor**
- **city**
- **traffic**
- **…**



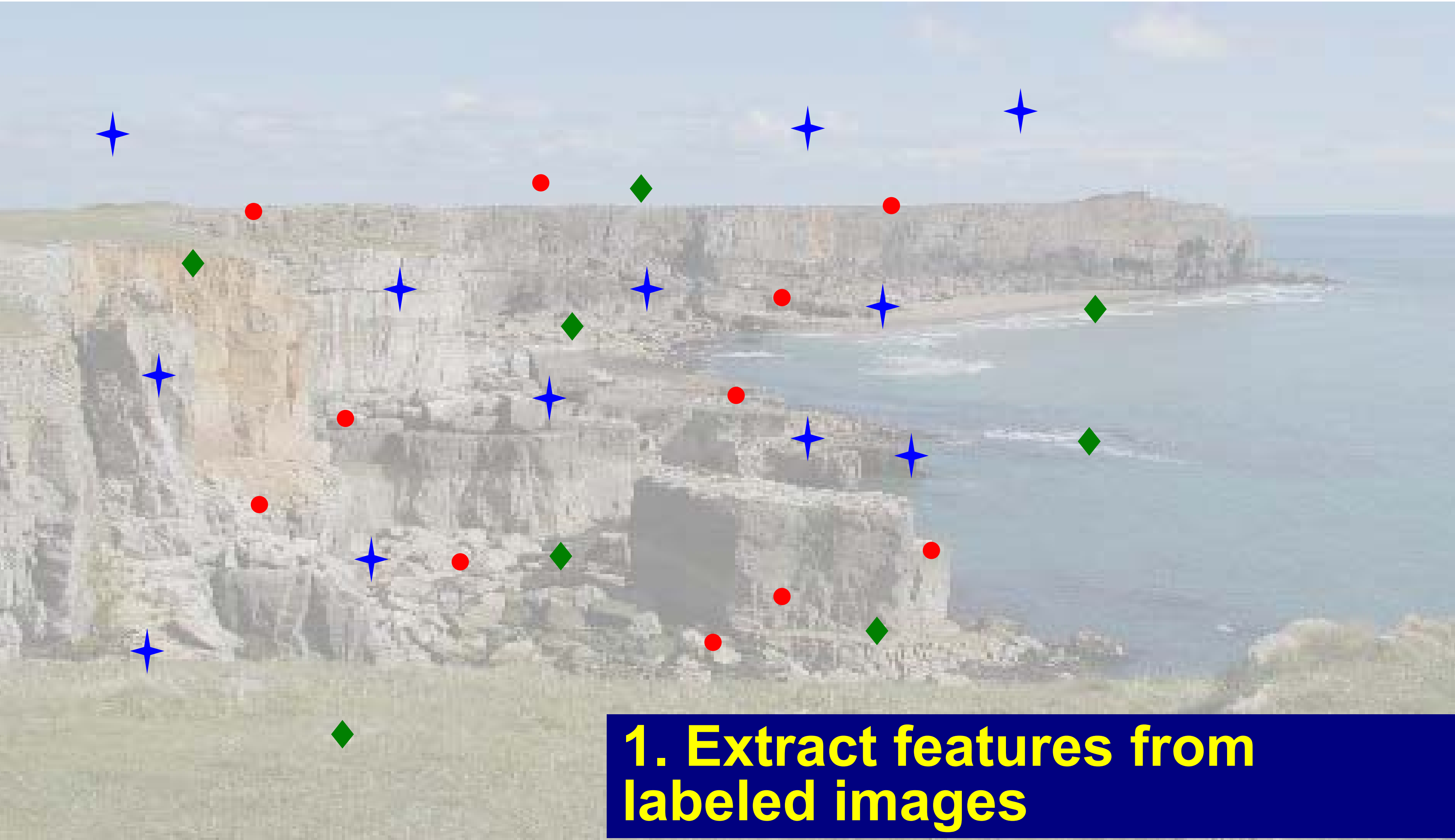
Scene and context categorization

The idea

- Represent images as "bags of features"
- Spatial relations between features
- **Training:**
  1. Get features from *labeled* images
  2. Histogram the features
  3. Use the histograms to train a classifier
- **Testing**
  1. Use classifier to label new image

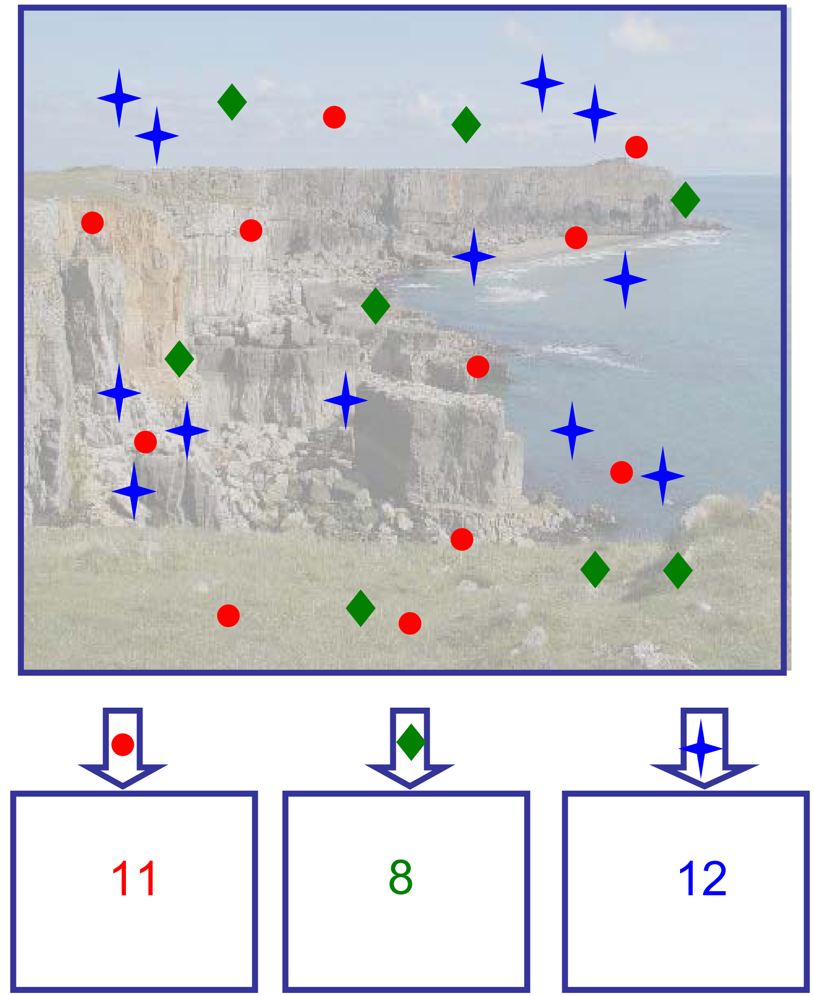
Training: Extract features

Extract features



Count the features

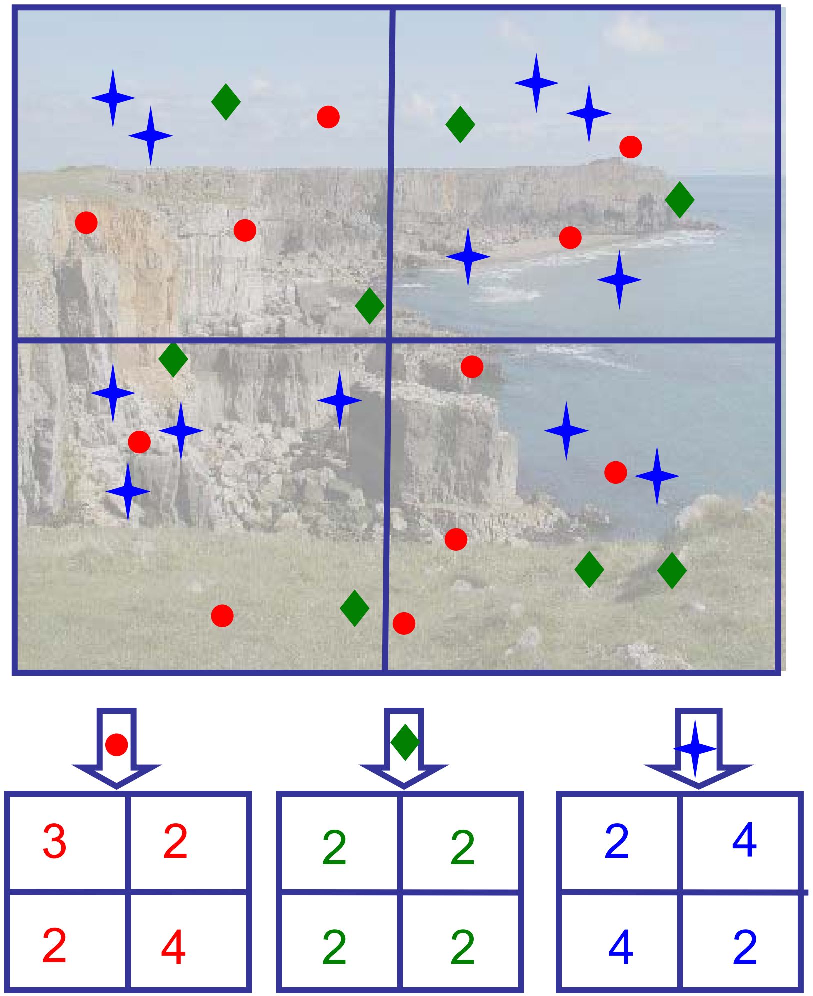
- Level 0



Multi-scale bag of features

Lazebnik *etal*, CVPR06

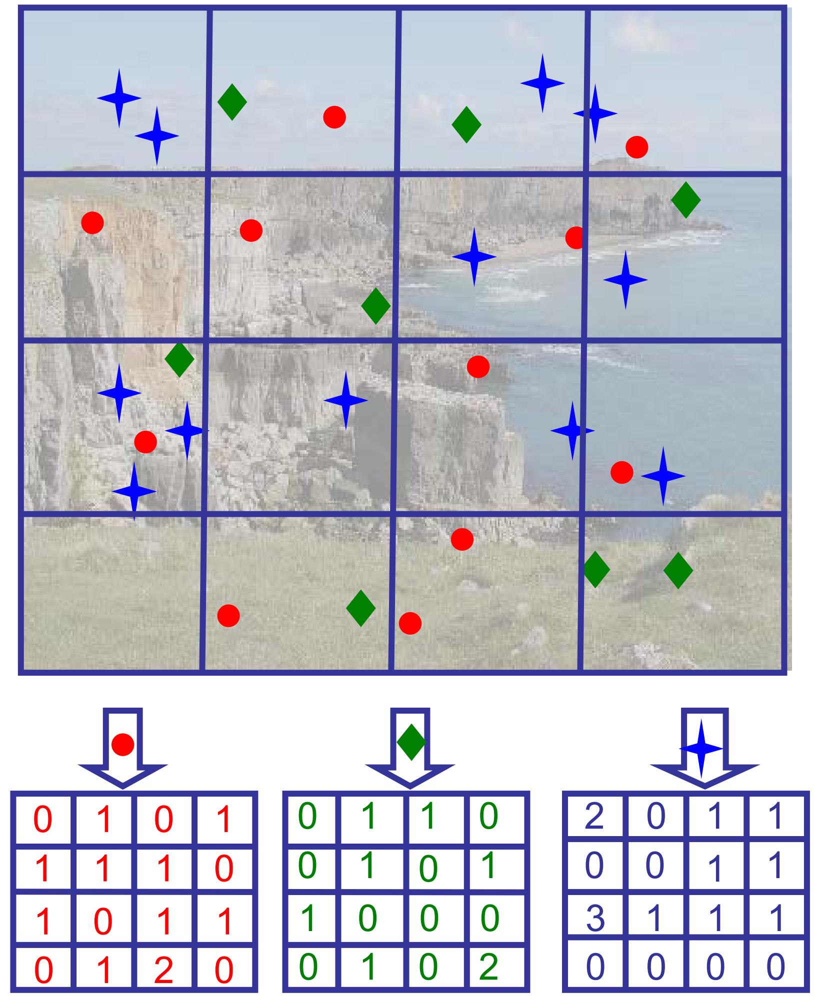
- Level 0
- Level 1



S. Lazebnik, C. Schmid, and J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, CVPR 2006, to appear.

Multi-scale bag of features

- Level 0
- Level 1
- Level 2



S. Lazebnik, C. Schmid, and J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, CVPR 2006, to appear.

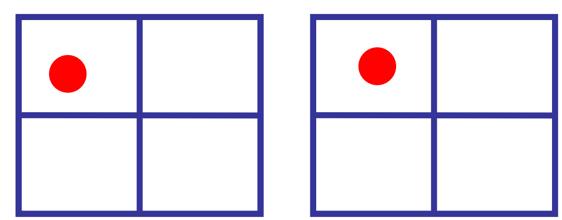
Matching between 2 images

- For each feature type
- For each pyramid level L = 0,1,2

$$\mathcal{I}^{\ell} = \min\left(\underset{\text{Image 1}}{\begin{bmatrix} 3 & 2 \\ 2 & 4 \end{bmatrix}} , \underset{\text{Image 2}}{\begin{bmatrix} 0 & 4 \\ 8 & 1 \end{bmatrix}} \right) = \begin{bmatrix} 0 & 2 \\ 2 & 1 \end{bmatrix}$$

Feature
Level = 1

- Note: coarser level matches include the finer level ones



S. Lazebnik, C. Schmid, and J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, CVPR 2006, to appear.

Matching between 2 images

Weight

$$\kappa^L(X,Y) \;=\; \mathcal{I}^L + \sum_{\ell=0}^{L-1} \frac{1}{2^{L-\ell}}\left(\mathcal{I}^\ell - \mathcal{I}^{\ell+1}\right)$$

Classification function

NEW matches

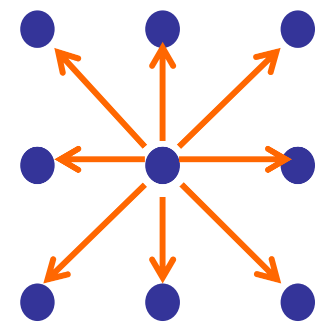
- Weight = penalize matches in larger cells.
- The resulting kernel is Mercer kernel.
- Use for the classification

S. Lazebnik, C. Schmid, and J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, CVPR 2006, to appear.

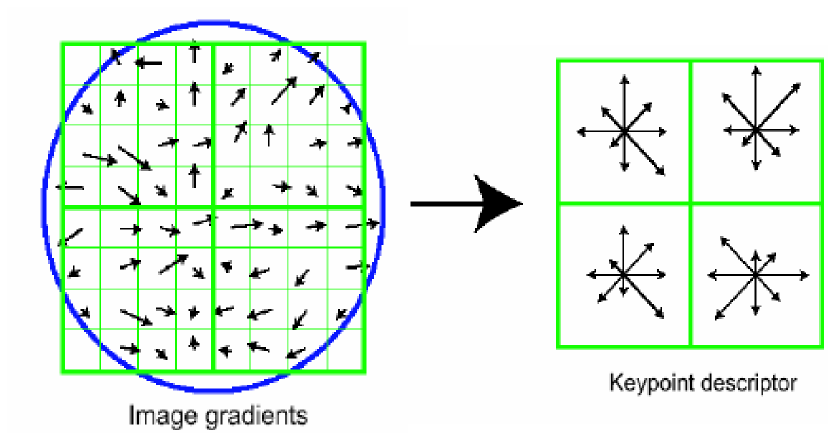
Types of Features

Oriented edge points

8 directions * 2 scales = 16 types



SIFT descriptors

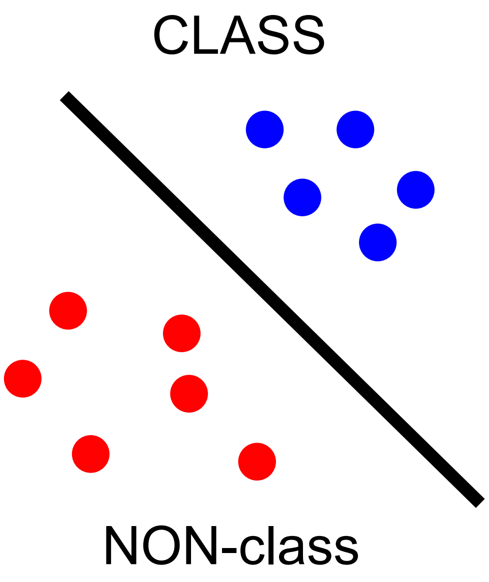


Cluster features to a vocabulary = 200 types

S. Lazebnik, C. Schmid, and J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, CVPR 2006, to appear.

Classification

- By SVM (support vector machine)
- Train classifier by applying the kernel on labeled images
- Test image is assigned the label of the classifier with the highest response



S. Lazebnik, C. Schmid, and J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, CVPR 2006, to appear.

Main contribution

Multi-scale representation to the "bag of features" approach

Considers spatial relation between features

Caltech 101: **64.6%** using **multiscale**, L=2
41.2% with L=0

53.9% State-of-the-art (Zhang etal.)

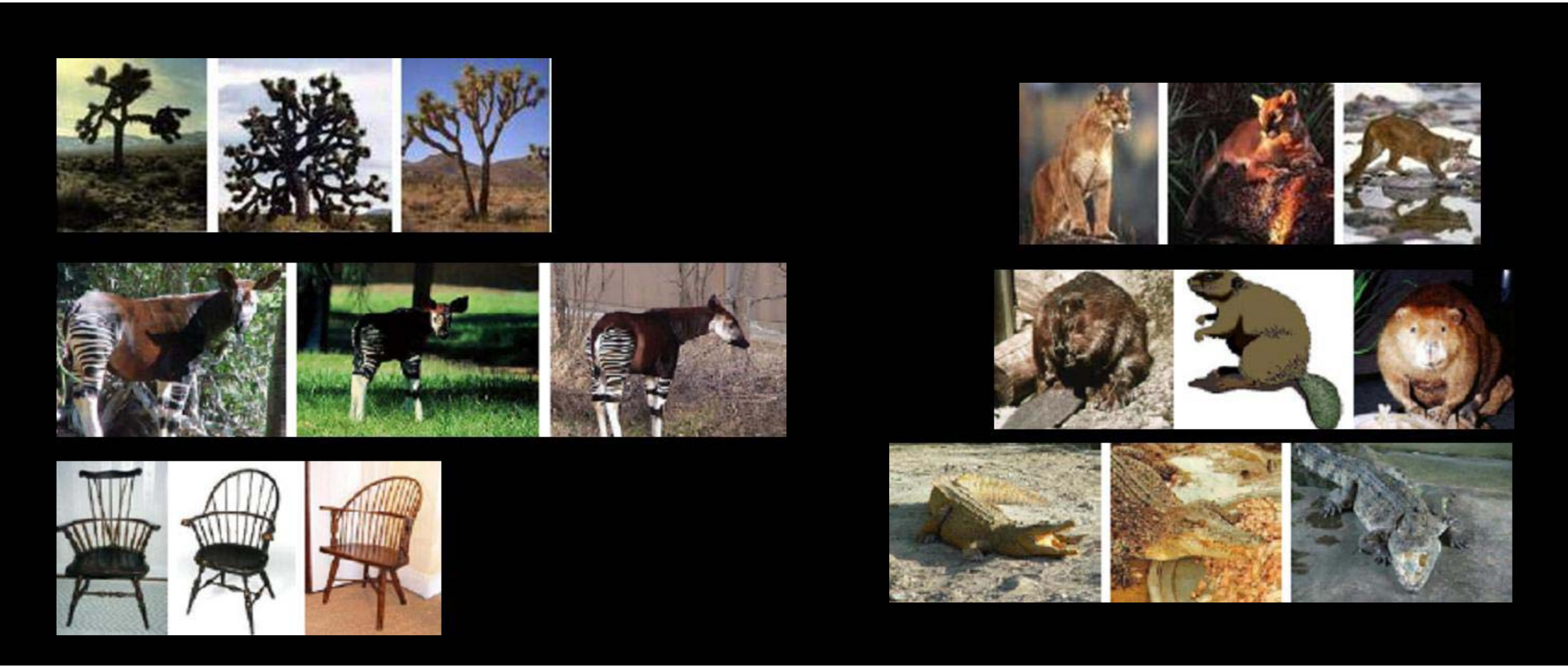
S. Lazebnik, C. Schmid, and J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, CVPR 2006, to appear.

Results

S. Lazebnik, C. Schmid, and J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, CVPR 2006, to appear.

Performance

High performance | Poor performance



Coherent scenes, little clutter | Textureless animals, camouflage

S. Lazebnik, C. Schmid, and J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, CVPR 2006, to appear.

Recognize this **particular** model:

in this input image

Recognition

Model image

Find the location of the model

+ get segmentation

V. Ferrari, T. Tuytelaars, and L. Van Gool. *Simultaneous Object Recognition and Segmentation by Image Exploration.* In Proc. ECCV, 2004.

Why is it hard?

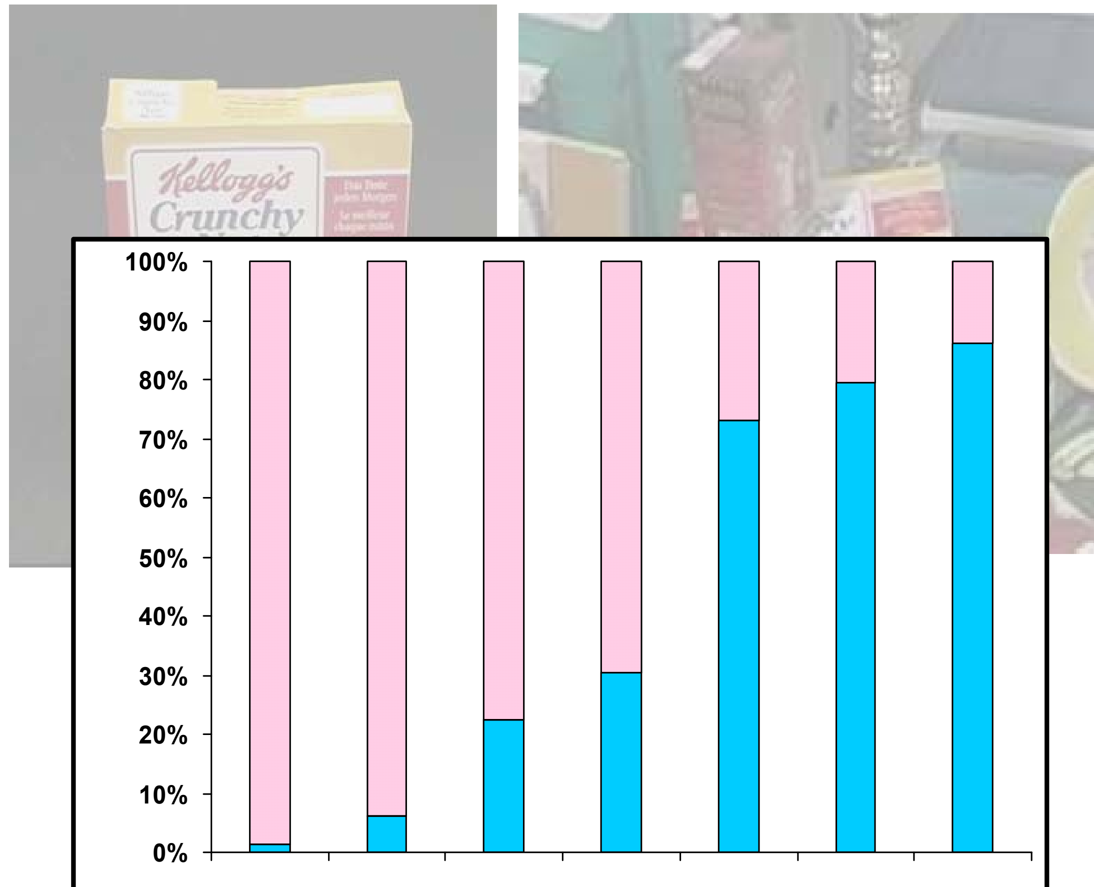
- Clutter
- Occlusions
- Scale
- Viewpoint
- Non-rigid deformations

Matching solely by affine invariant features is not robust

The Idea

Ferrari etal., ECCV 2004

1. Find initial set of invariant features
2. Expansion - Look around them to construct more matching features
3. Contraction - Leave the correct, and remove mismatches
4. Iteratively construct more and more matches, increasingly farther from the initial ones.

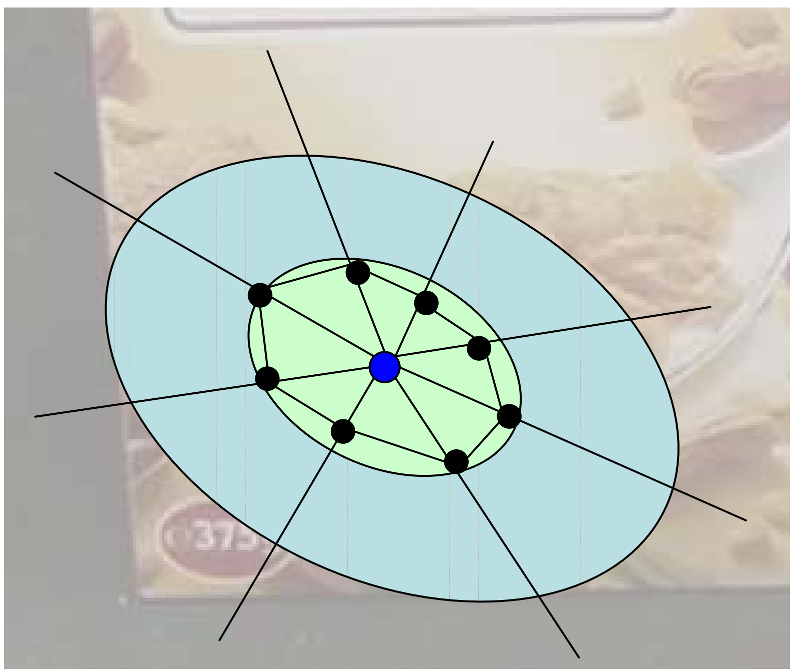


Different stages of the alg

V. Ferrari, T. Tuytelaars, and L. Van Gool. *Simultaneous Object Recognition and Segmentation by Image Exploration.* In Proc. ECCV, 2004.

Soft Matching

- Initial feature extraction in model and image



1. Choose point by non-max suppression
2. Transfer rays
3. Local extremum along the rays

$$f(t) = \frac{|I(t) - I_0|}{\max\left(\frac{\int_0^t |I(t) - I_0|\,dt}{t}, d\right)}$$

4. Calculate moments to get internal ellipse
5. Double ellipse size

T. Tuytelaars and L. Van Gool *Wide Baseline Stereo based on Local, Affinely invariant Regions* In Brit. Mach. Vis. Conf., pp. 412-422, 2000.

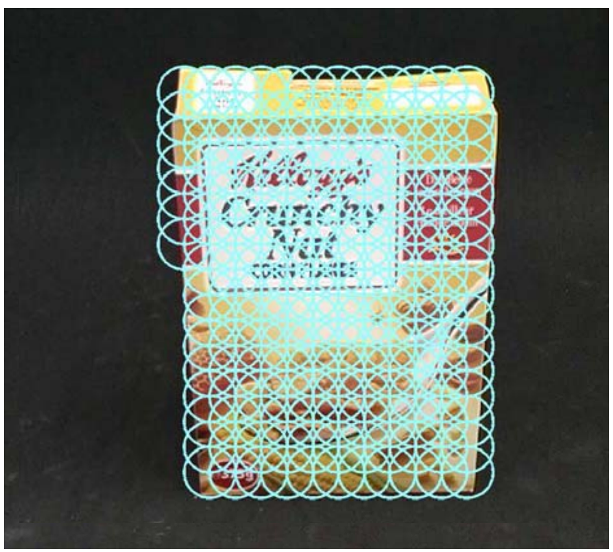
Initial Matching-cont.

- Every test region matched to 0-3 model regions by thresholding the:
  - Mahalanobis distance (on color moments)
  - Similarity measure (NCC on gray levels + Euclidean distance in RGB space)
  - Geometric refinement – find affine transf. that maximized the similarity

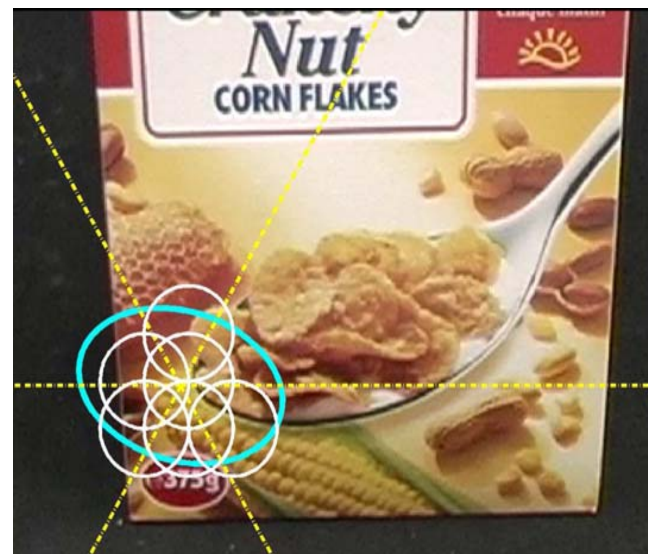
V. Ferrari, T. Tuytelaars, and L. Van Gool. *Simultaneous Object Recognition and Segmentation by Image Exploration.* In Proc. ECCV, 2004.

Early expansion

- Coverage of the model image



•Each feature in model gives support for a part of the covering



V. Ferrari, T. Tuytelaars, and L. Van Gool. *Simultaneous Object Recognition and Segmentation by Image Exploration.* In Proc. ECCV, 2004.

Early expansion

1. Map a region from the model to image using the transf. defined by the initial matching.

2. Refine the transformation.

3. Keep the matching with the best similarity (which also above a detection threshold).

4. Discard all matches that did not succeed in propagating any region.

V. Ferrari, T. Tuytelaars, and L. Van Gool. *Simultaneous Object Recognition and Segmentation by Image Exploration.* In Proc. ECCV, 2004.

Result of the expansion

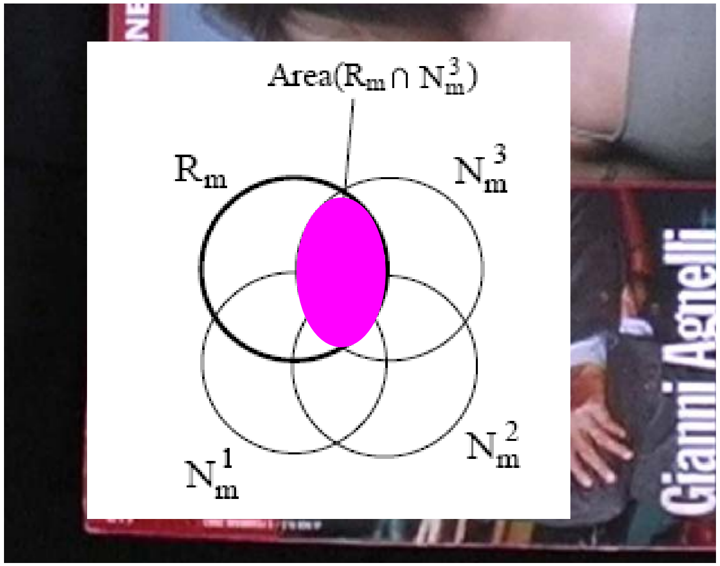
Model Image

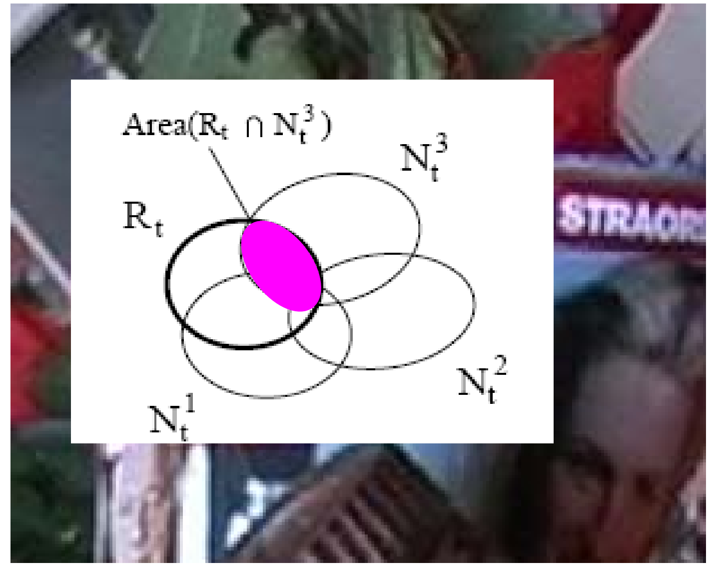
Test Image

V. Ferrari, T. Tuytelaars, and L. Van Gool. *Simultaneous Object Recognition and Segmentation by Image Exploration.* In Proc. ECCV, 2004.

Early Contraction

Model Image





Test Image

Correct matching = same intersections in test & model

Match (R_m, R_t) is removed if:

$$\sum_{\{N_m^i\}} \left| \frac{\text{Area}(R_m \bigcap N_m^i)}{\text{Area}(R_m)} - \frac{\text{Area}(R_t \bigcap N_t^i)}{\text{Area}(R_t)} \right| > t_s$$

V. Ferrari, T. Tuytelaars, and L. Van Gool. *Simultaneous Object Recognition and Segmentation by Image Exploration.* In Proc. ECCV, 2004.

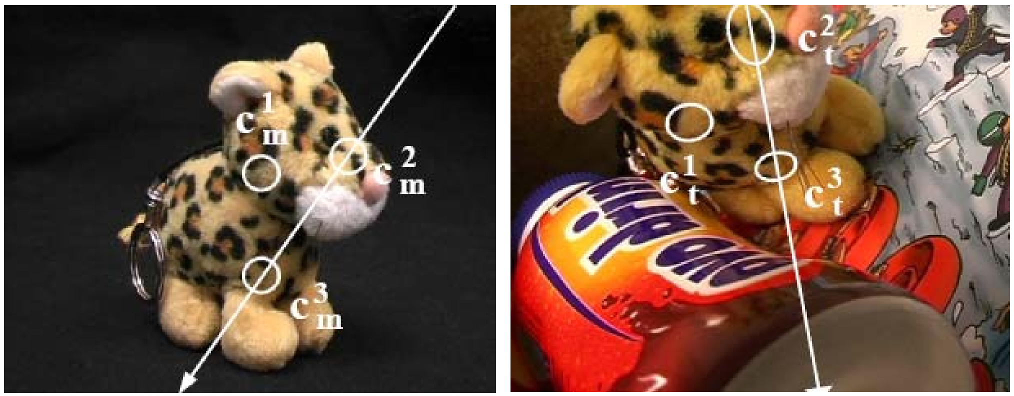
Main expansion

- Matches from pervious step added to support group.
- Follow similar steps as early expansion.
- Refinement is applied after we picked a new matching region.

V. Ferrari, T. Tuytelaars, and L. Van Gool. *Simultaneous Object Recognition and Segmentation by Image Exploration.* In Proc. ECCV, 2004.

Main Contraction

Sidedness Constraint



$$\mathit{side}(R_m^1, R_m^2, R_m^3) = \begin{cases} -1 & \text{right side} \\ 1 & \text{left side} \end{cases}$$

$$\mathrm{err}_{\mathrm{topo}}(R^i) = \frac{1}{v} \sum_{R^j, R^k \in \Gamma \setminus R^i, j>k} \left| \mathrm{side}(R_m^i, R_m^j, R_m^k) - \mathrm{side}(R_t^i, R_t^j, R_t^k) \right|$$

- Test each region R^i with every two other regions R^j, R^k
- Correct match will give 0

V. Ferrari, T. Tuytelaars, and L. Van Gool. *Simultaneous Object Recognition and Segmentation by Image Exploration.* In Proc. ECCV, 2004.

Main contraction –cont.

The filtering algorithm:

1. (Re-)compute $\mathrm{err}_{\mathrm{tot}}(R^i)$ for all $R^i \in \Gamma$.

$$\mathrm{err}_{\mathrm{tot}}(R^i) = \mathrm{err}_{\mathrm{topo}}(R^i) + (t_2 - \mathrm{sim}(R^i_m, R^i_t))$$

sidedness similarity

2. Find the worst match R^w, with $w = \arg\max_i \mathrm{err}_{\mathrm{tot}}(R^i)$

3. If $\mathrm{err}_{\mathrm{tot}}(R^w) > 0$, remove R^w: $\Gamma \leftarrow (\Gamma \backslash R^w)$, and iterate to 1, else stop.

Finally, iterate between contraction & expansion...

V. Ferrari, T. Tuytelaars, and L. Van Gool. *Simultaneous Object Recognition and Segmentation by Image Exploration.* In Proc. ECCV, 2004.

Illustration: Soft matches

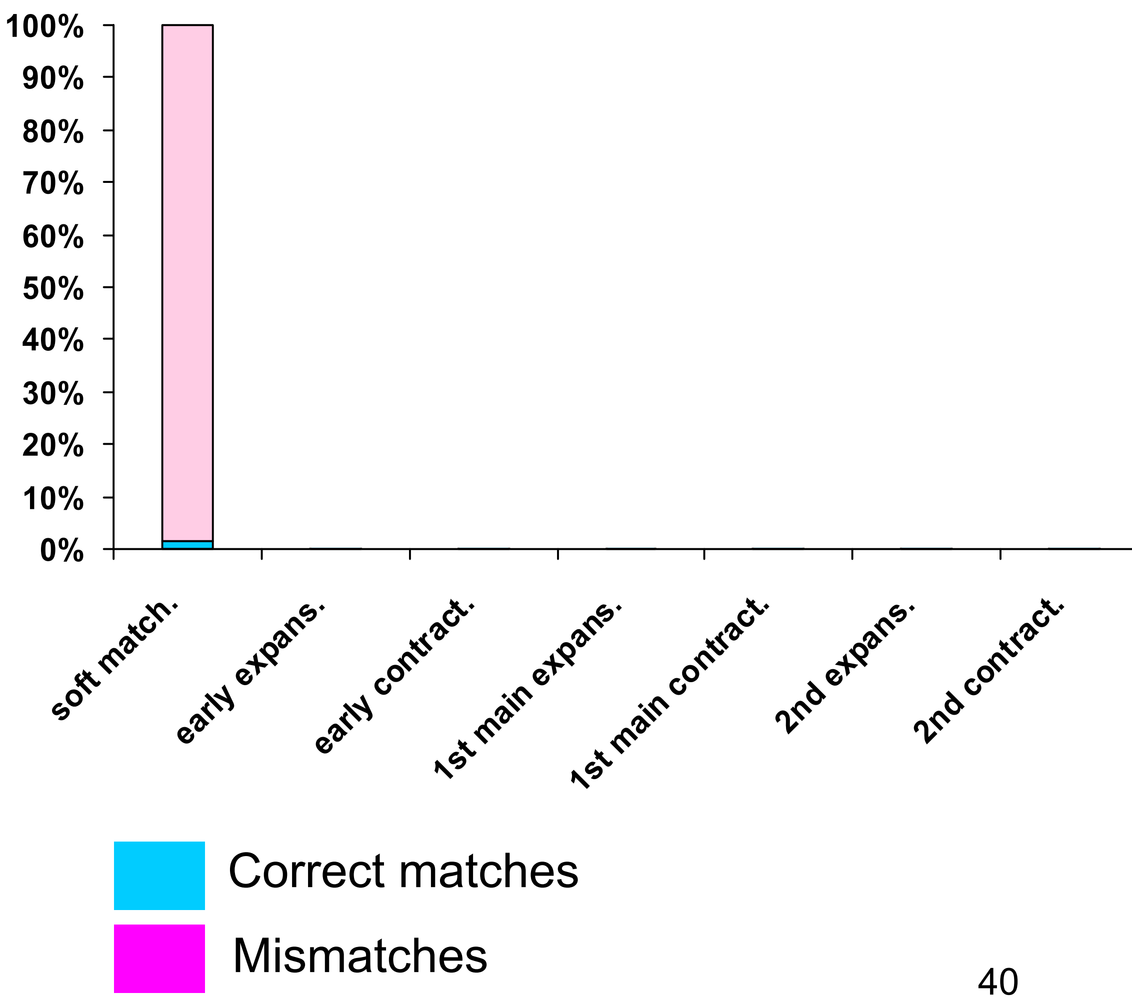


Illustration: Early expansion

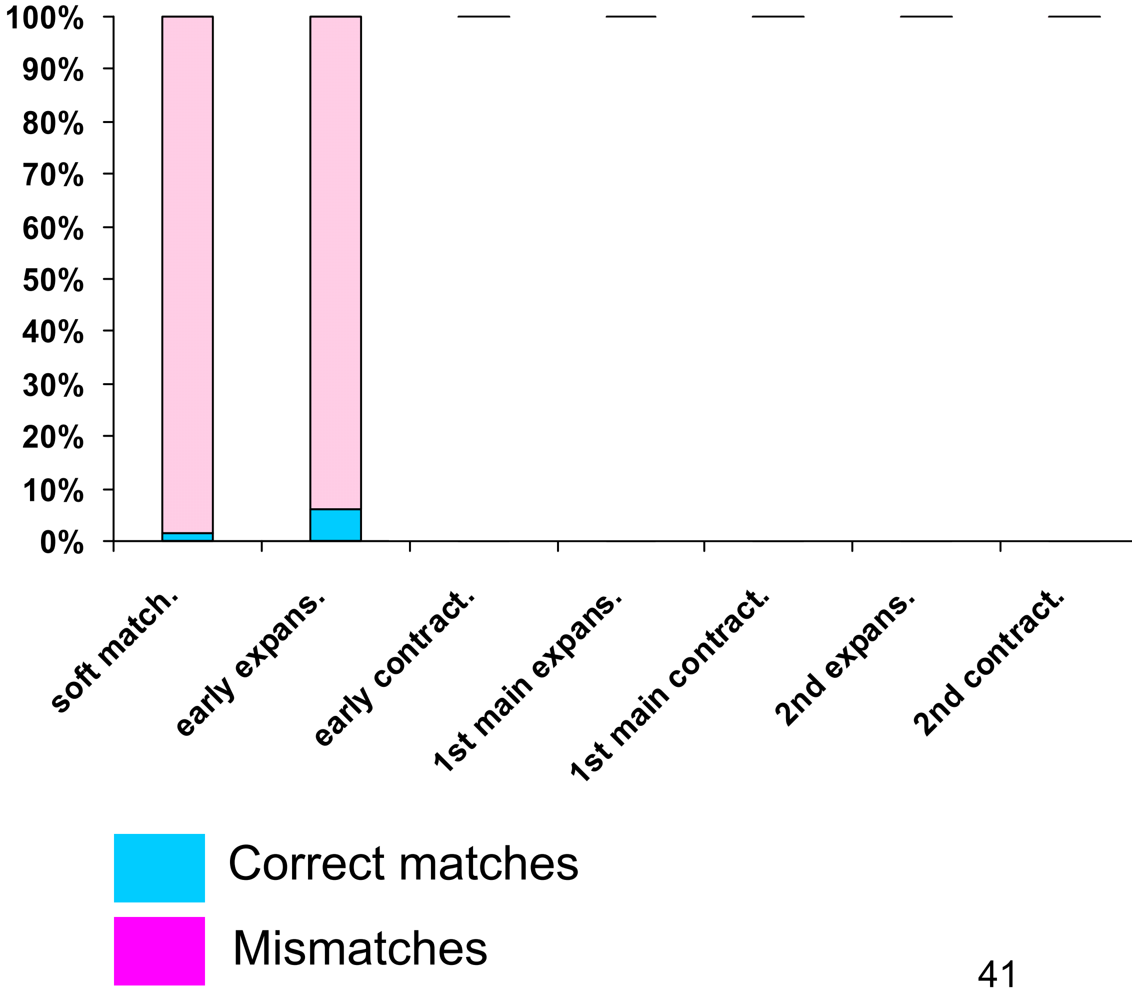


Illustration: Early contraction

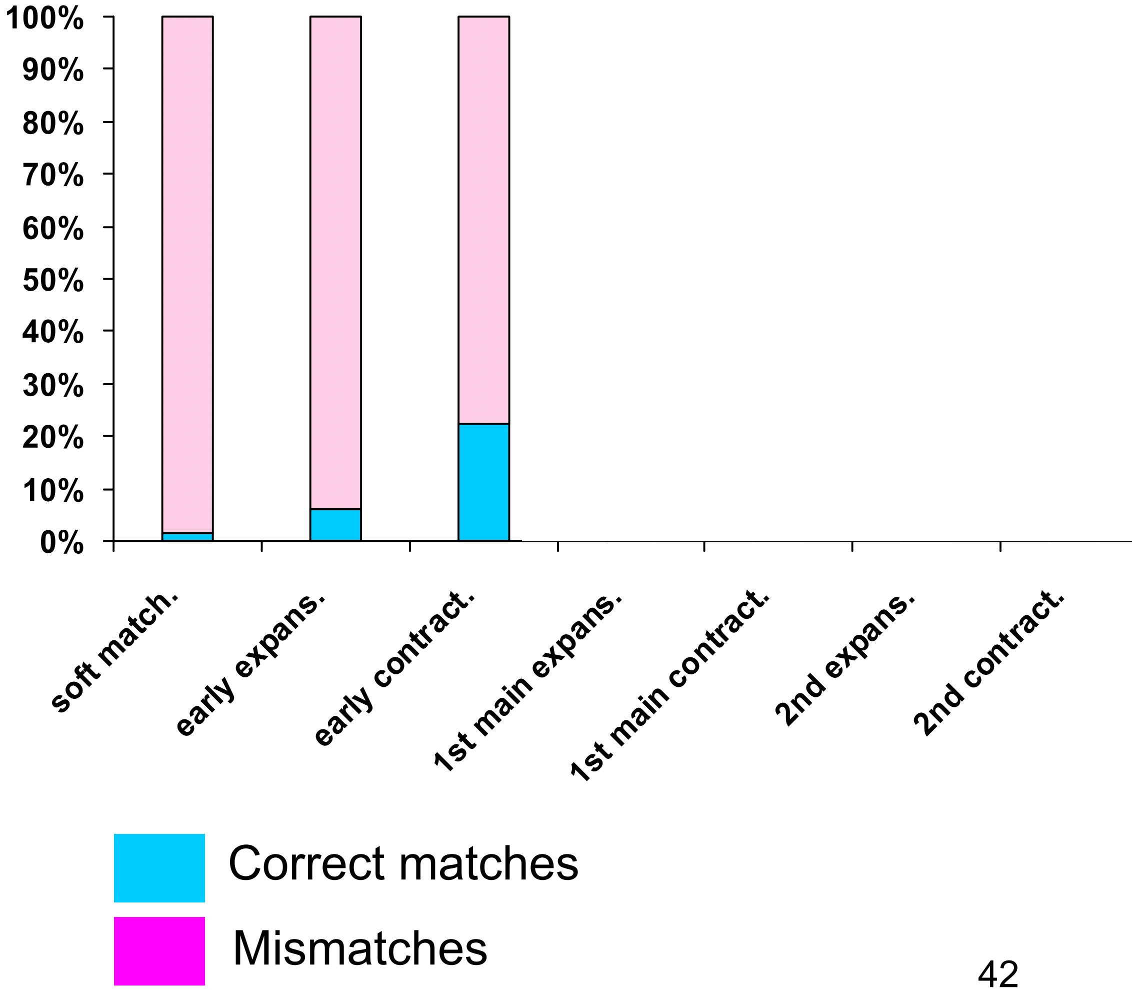


Illustration: 1st Main Expansion

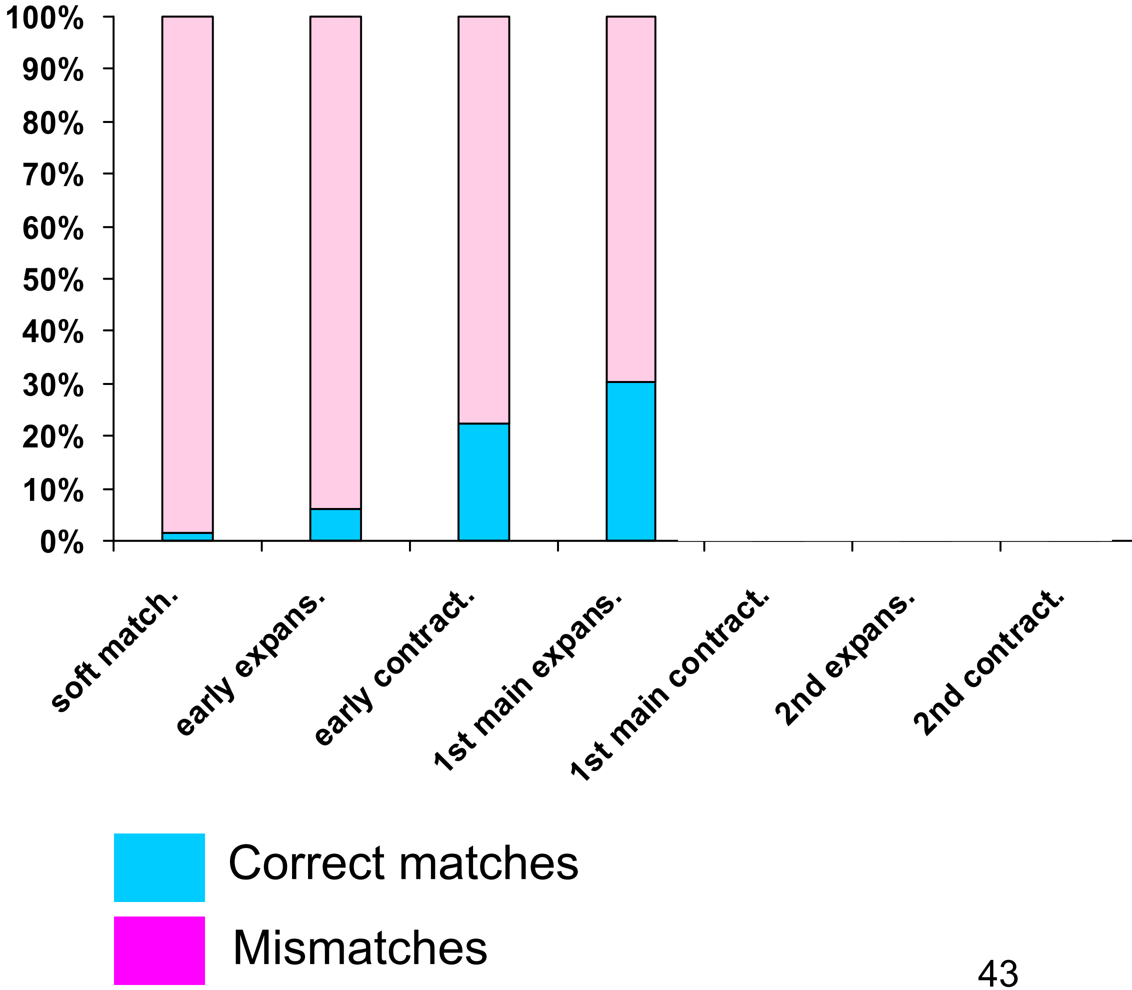


Illustration: 1st Main Contraction

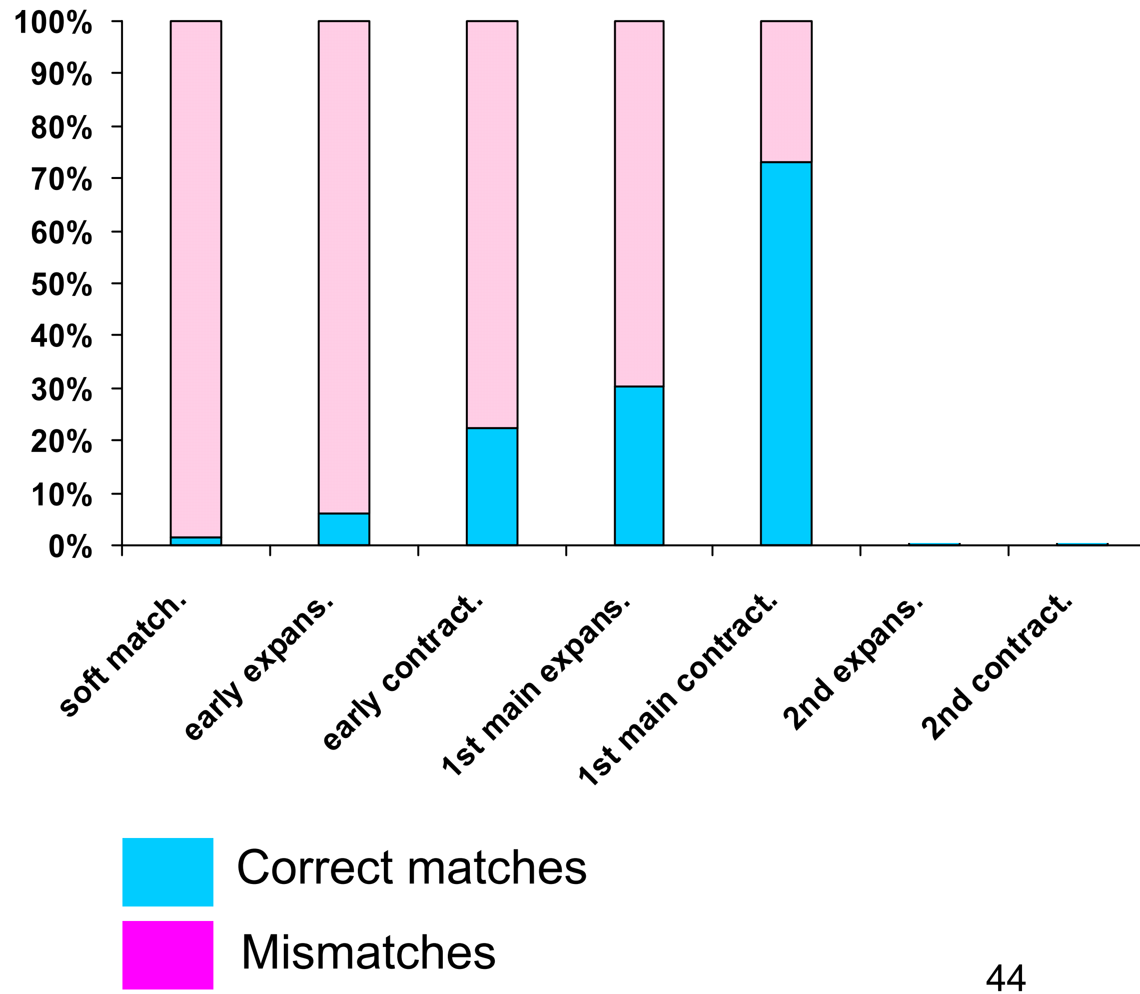


Illustration: 2^{nd} Main Expansion

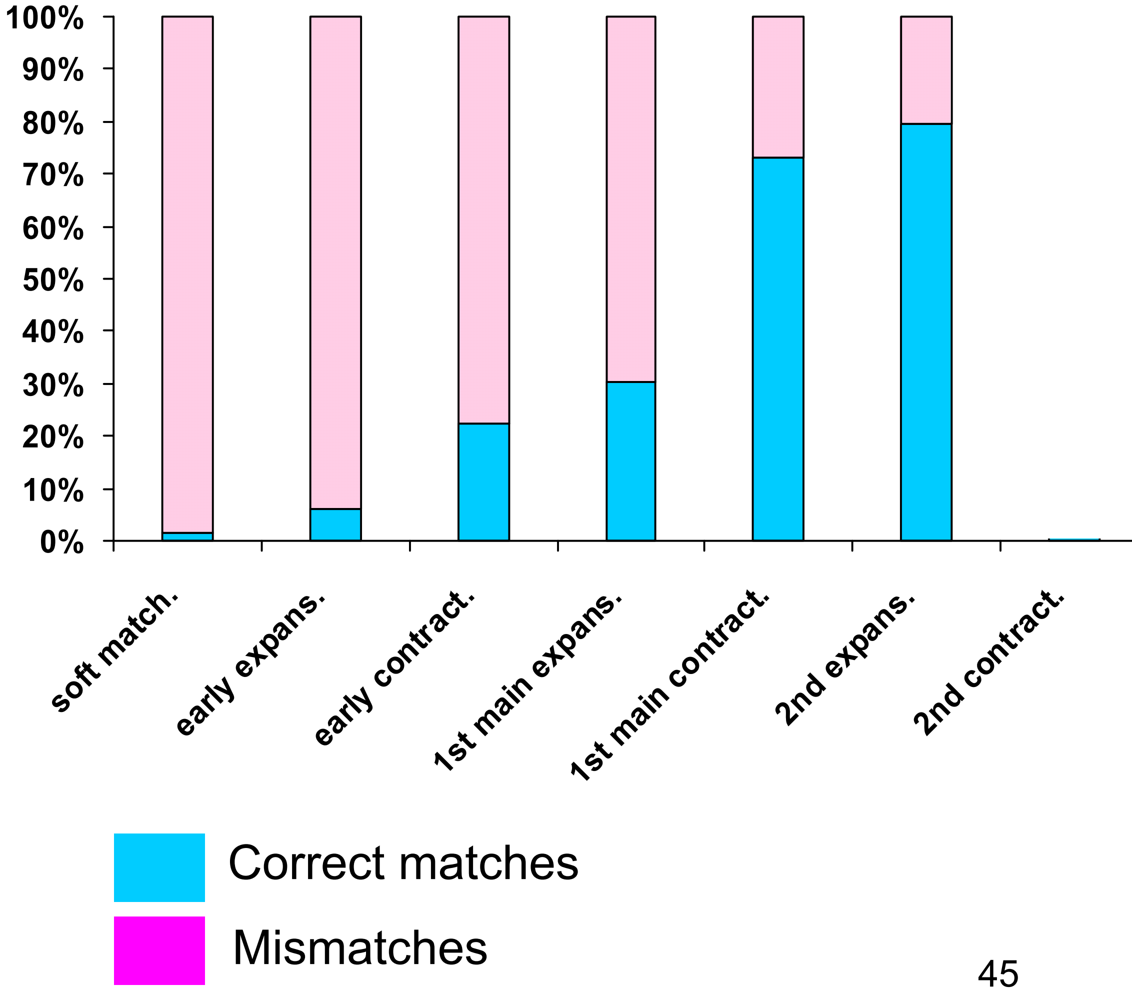
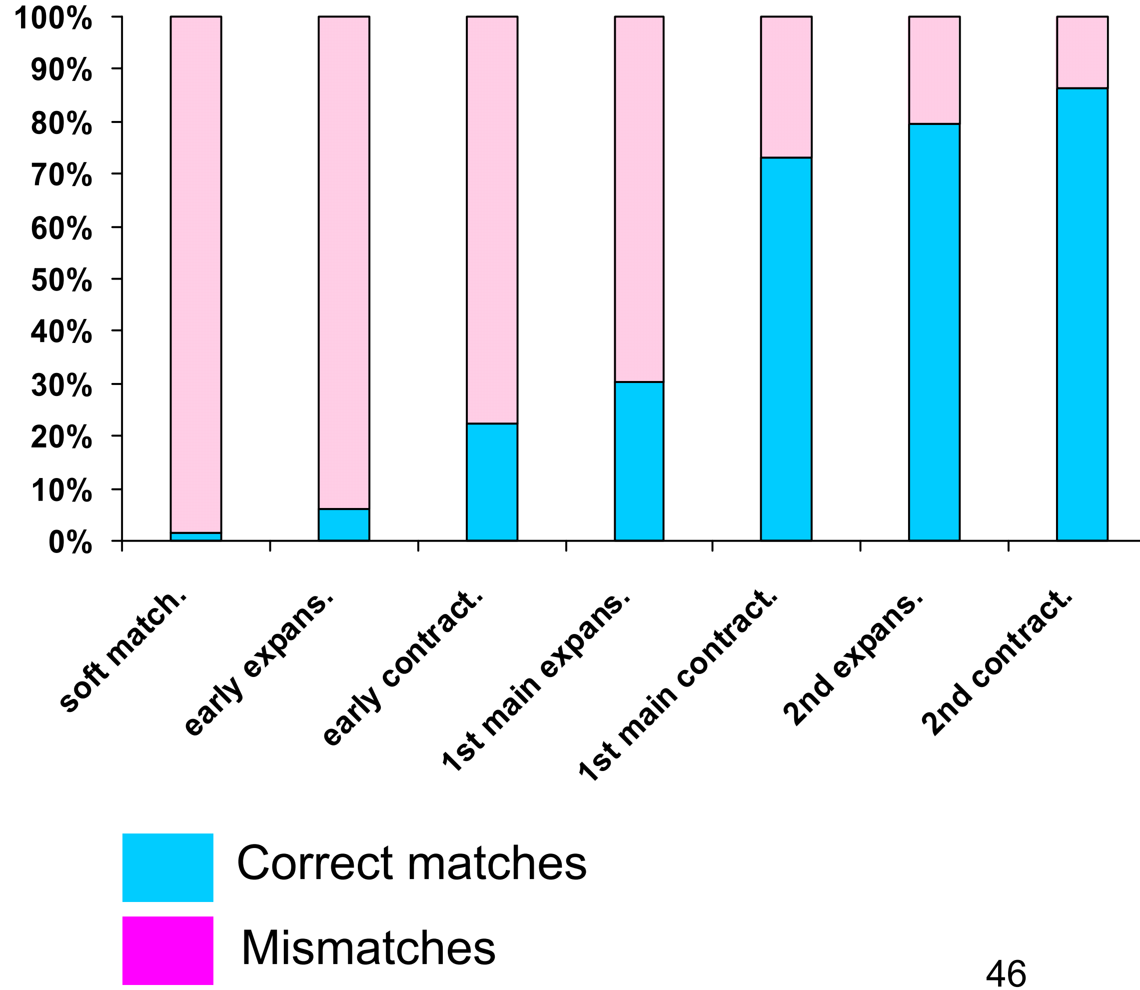


Illustration: 2nd Main Contraction



Results


Chi
EDIZIONE STRAORDINARIA
Michelle Hunziker
'E ora cerco
un amore vero'
Gianni Agnelli
Addio a un uomo
entrato nella storia

Results

Results


ovo drink


Chi
EDIZIONE STRAORDINARIA
Michelle Hunziker
'E ora cerco
un amore vero'
Gianni Agnelli
Addio a un uomo
entrato nella storia


ovo drink

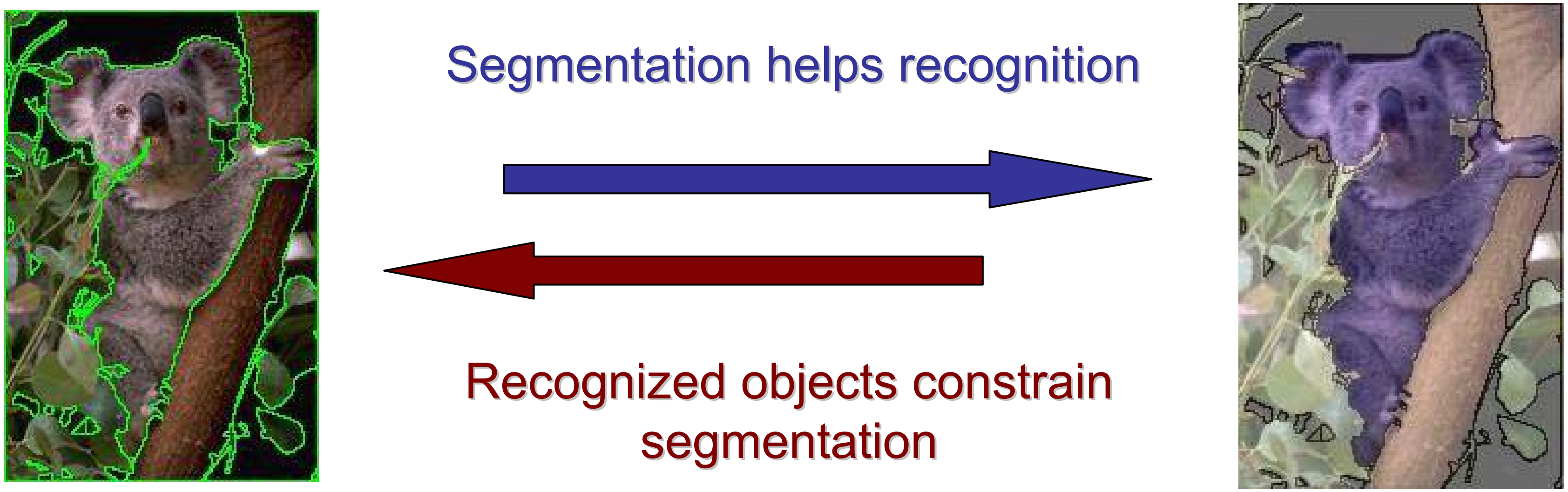
Results


Chi
EDIZIONE STRAORDINARIA
Michelle Hunziker
'E ora cerco un amore vero'
Gianni Agnelli
Addio a un uomo entrato nella storia
Martina Stella
Un successo tra cinema e tv


Hunziker
E ora cerco
amore vero

Segmentation & Recognition

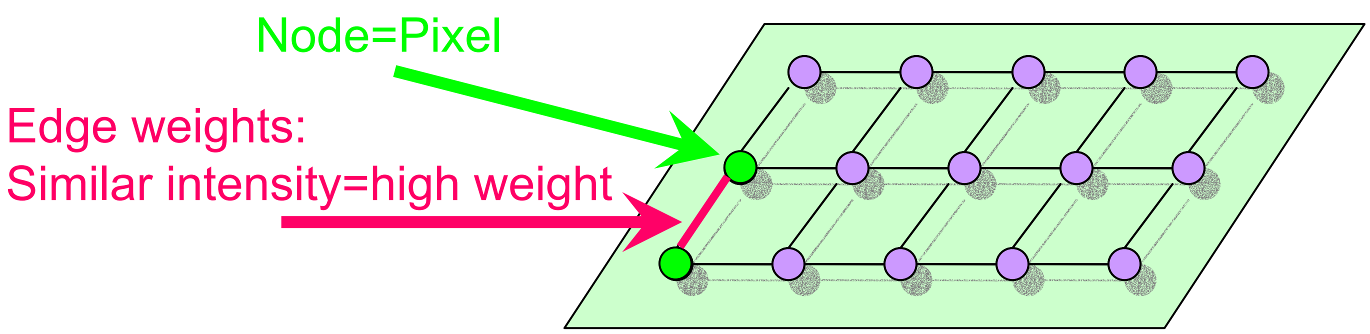
- Segmentation ⇄ recognition



Bottom – Up segmentation

- **Goal aggregate pixels with similar low level features.**

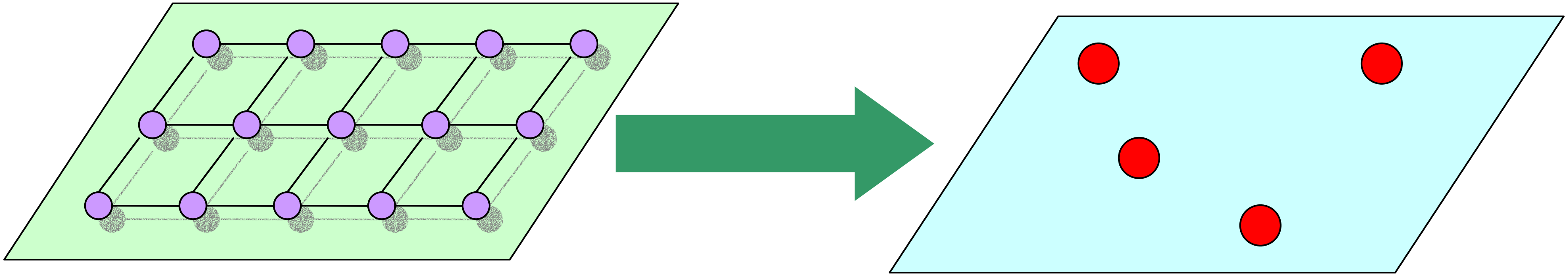
- E. Sharon *et al.* method : construct a 4-connected graph



Eitan Sharon, Achi Brandt, and Ronen Basri, Segmentation and boundary detection using multiscale intensity measurements., CVPR, 2001, pp. 469–476.

Bottom – Up segmentation

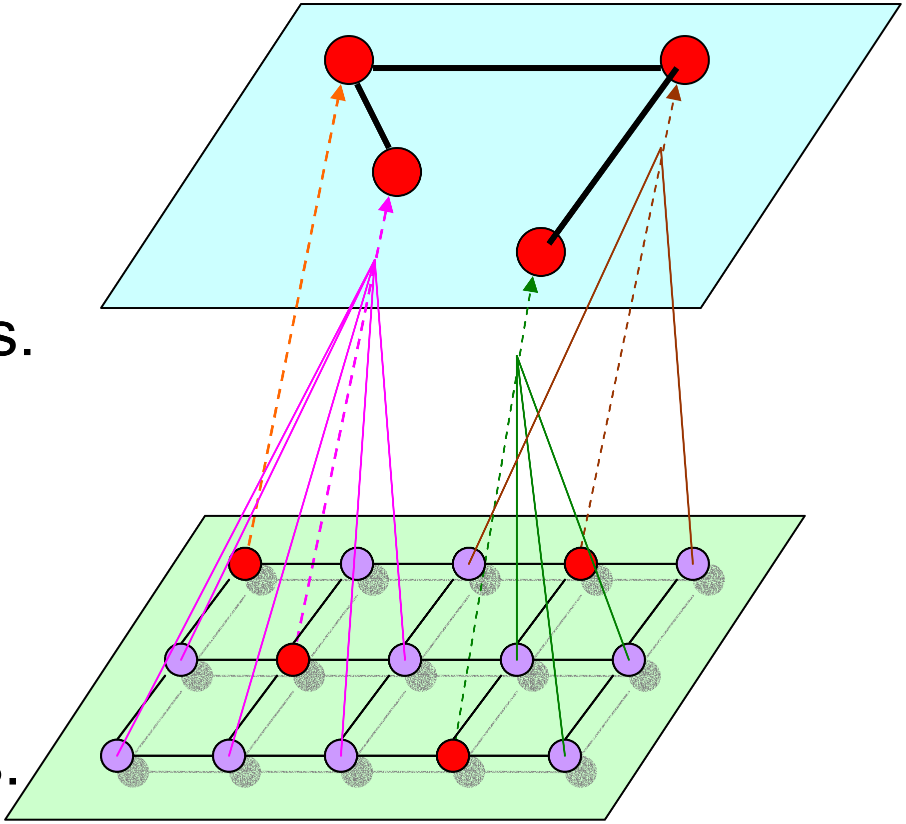
- Apply a fast multi level solver called *Algebraic MultiGrid* (AMG), for graph coarsening.



Eitan Sharon, Achi Brandt, and Ronen Basri, Segmentation and boundary detection using multiscale intensity measurements., CVPR, 2001, pp. 469–476.

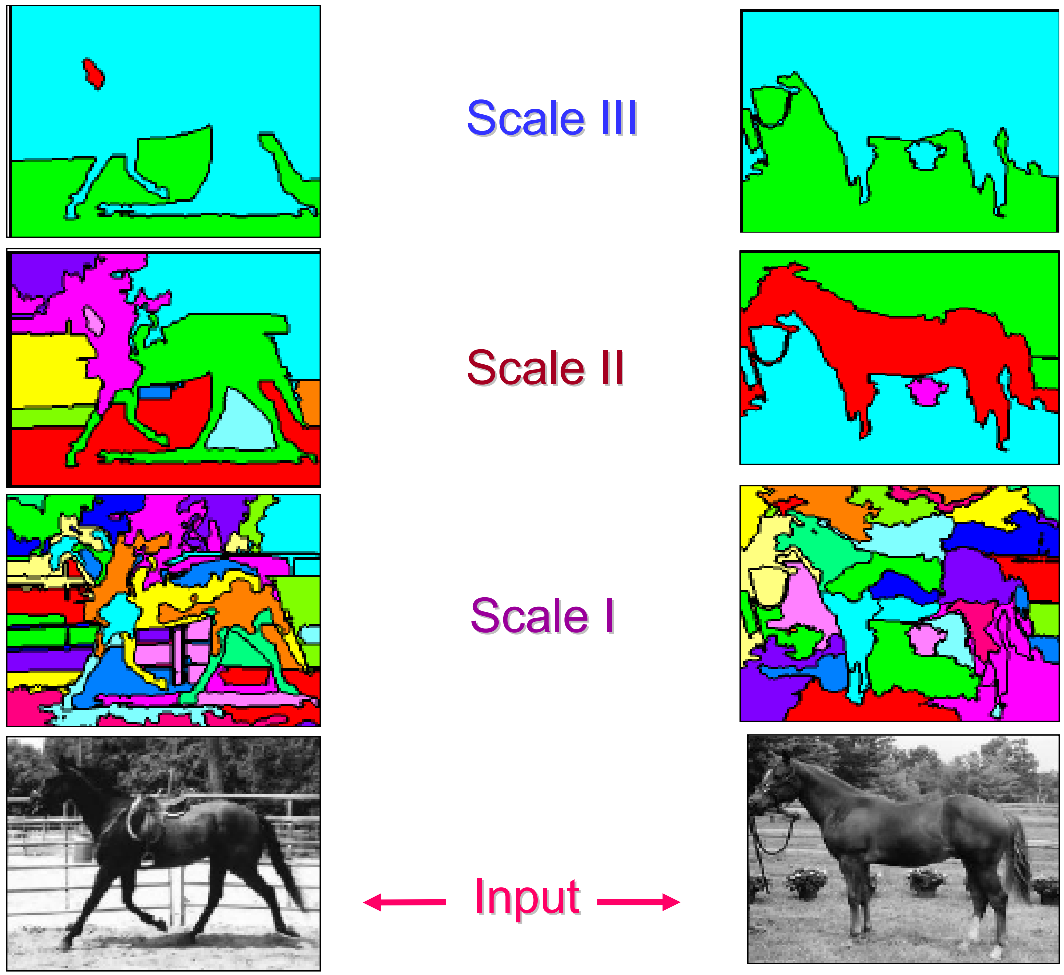
Bottom – Up segmentation

1. Calculate coarse scale node.
2. Create inter scale weights and calculate new features.
3. Update coarse scale weights by averaging the fine scale weights.
4. Repeat the above process. Segments emerges as node at some coarse scale.



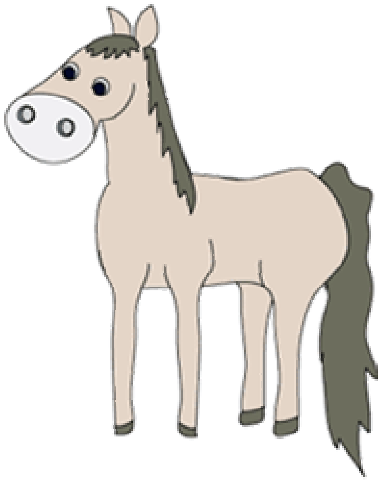
Eitan Sharon, Achi Brandt, and Ronen Basri, Segmentation and boundary detection using multiscale intensity measurements., CVPR, 2001, pp. 469–476.

Difficulties in bottom-up segmentation



How to constrain the segmentation?

Think on an horse for example:



Shape is an important cue

Introducing shape Into the segmentation

- Instead of Bottom up use Top-down

Training images

Learn how to segment this type of images

Result: Binary segmentation Background / foreground

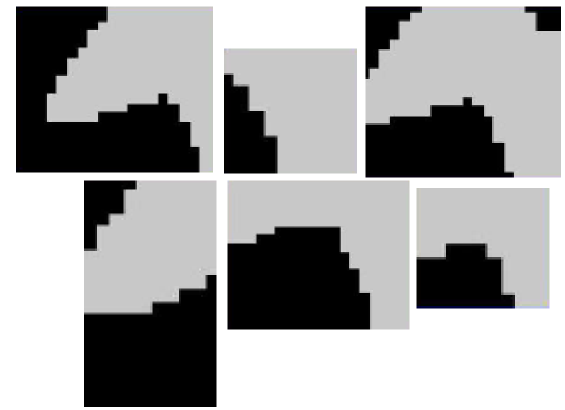
Top-Down - Borenstein & Ullman.

Randomly collect a large set of candid. fragments

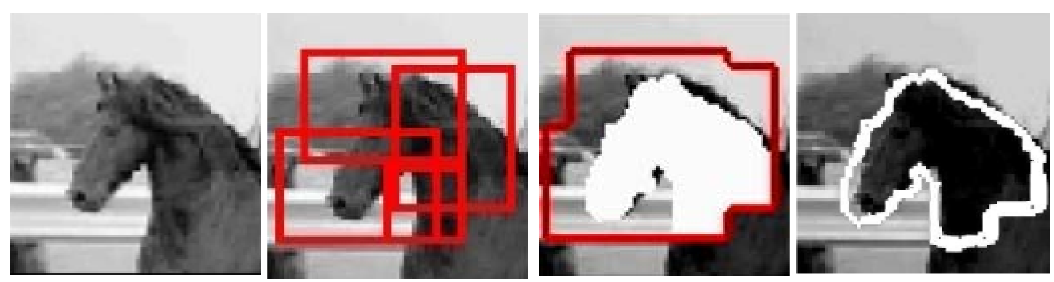
class non-class

Select subset of informative fragments

Figure-Background Labeling



Detection and segmentation

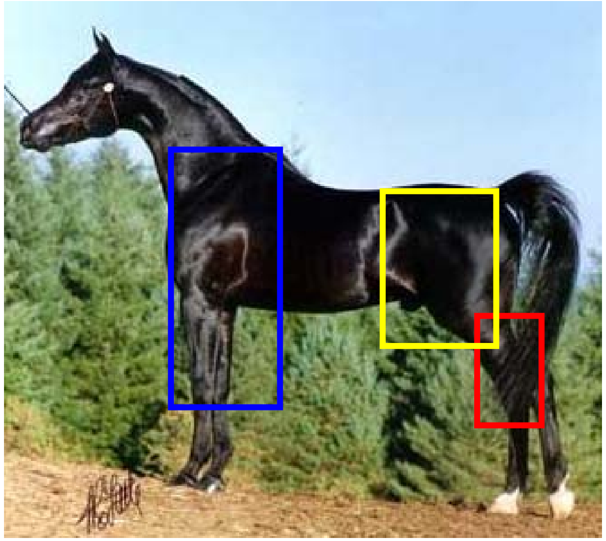


E. Borenstein, S. Ullman. Learning to Segment. ECCV (3) 2004: 315-328

Selecting Informative Fragments

- Informative = likely to be detected in class compared with non-class images

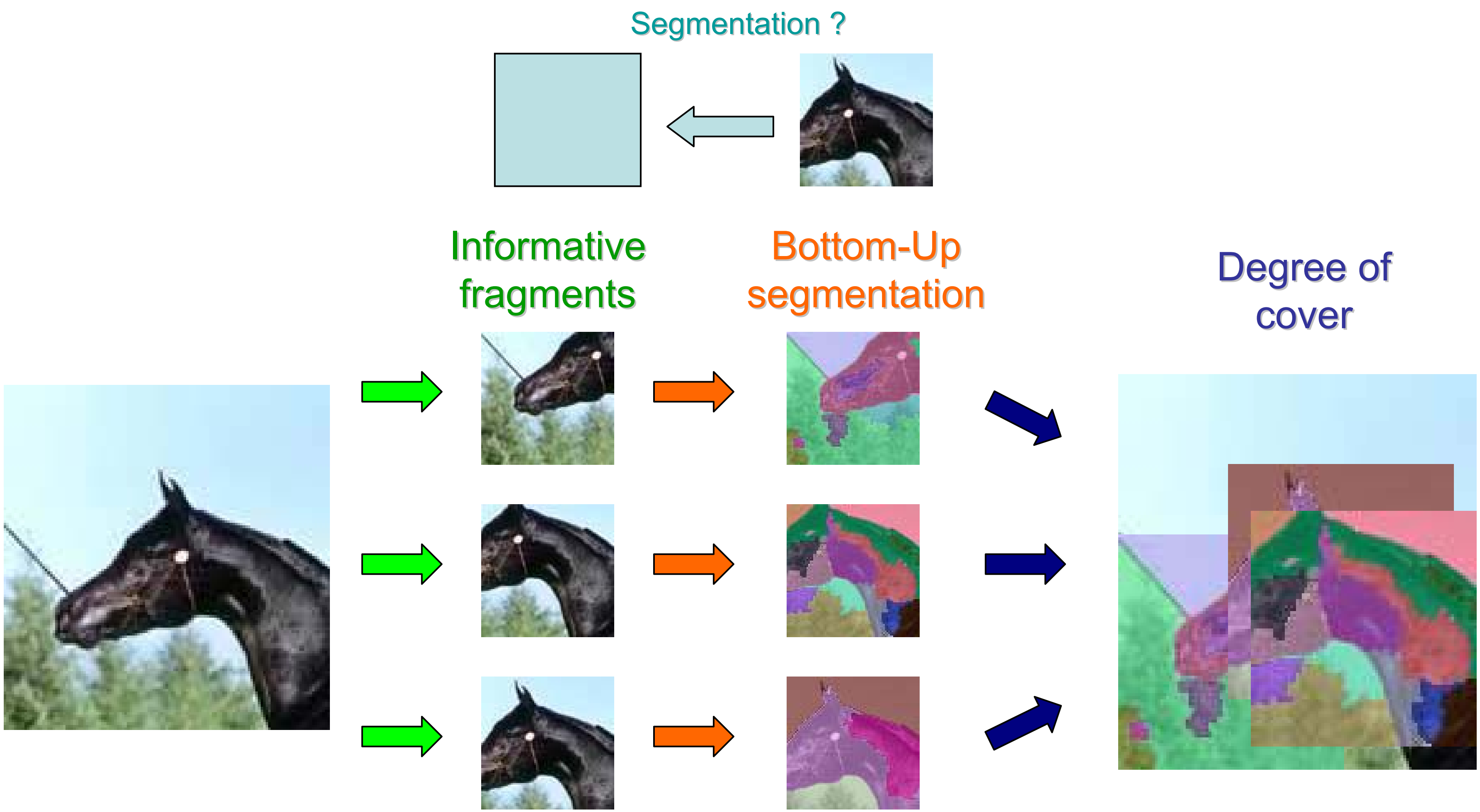
$$f_j = \arg\max_f \left(I(F^s \cup f; C) - I(F^s; C) \right)$$



- Fragments are added to maximize the gain in mutual information
- Highly overlapping
- Well-distributed over the figure

E. Borenstein, S. Ullman. Learning to Segment. ECCV (3) 2004: 315-328

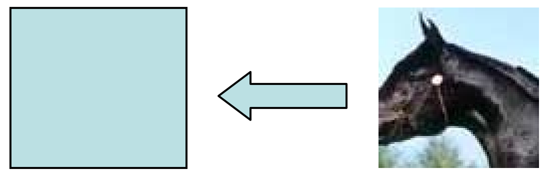
Segmenting Informative Fragments

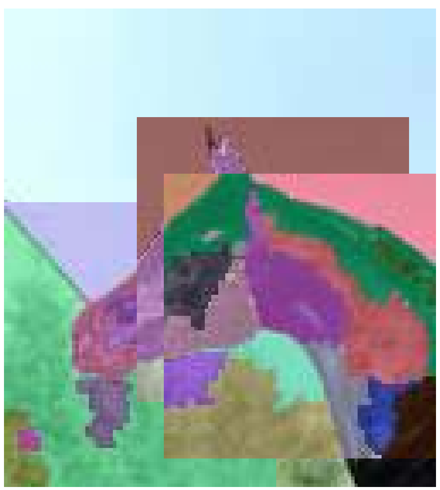


E. Borenstein, S. Ullman. Learning to Segment. ECCV (3) 2004: 315-328

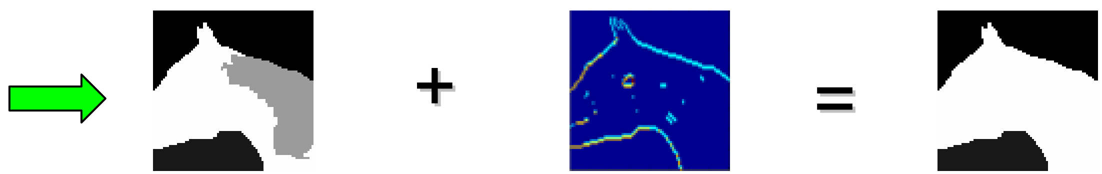
Segmenting Informative Fragments

Segmentation ?





Likelihood determined by
the degree of cover



Overlap with the average
Edges D(x,y) of fragment
in all the training images.

$$P = \arg\max_{P(\bar{r})} \left(\sum_{(x,y)\in P(\bar{r})} D(x,y) + \lambda \sum_{(x,y)\in \partial P(\bar{r})} D(x,y) \right)$$

E. Borenstein, S. Ullman. Learning to Segment. ECCV (3) 2004: 315-328

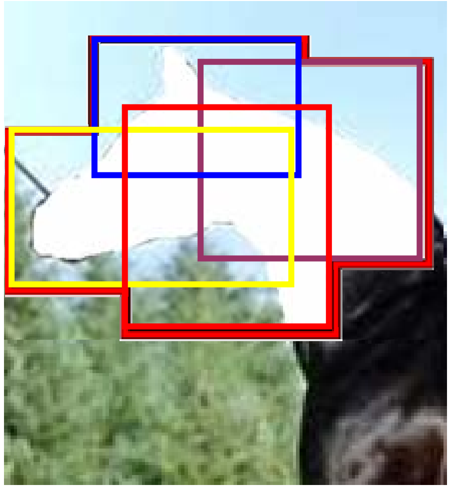
Segmenting a test image

- Each fragment votes for the classification of pixels it covers

$$\text{weight} = \frac{\text{detection rate}}{\text{false alarms rate}}$$

- Count the number of votes for figure and background

- Check consistency test between the received labeling and labeling of each fragment using NCC

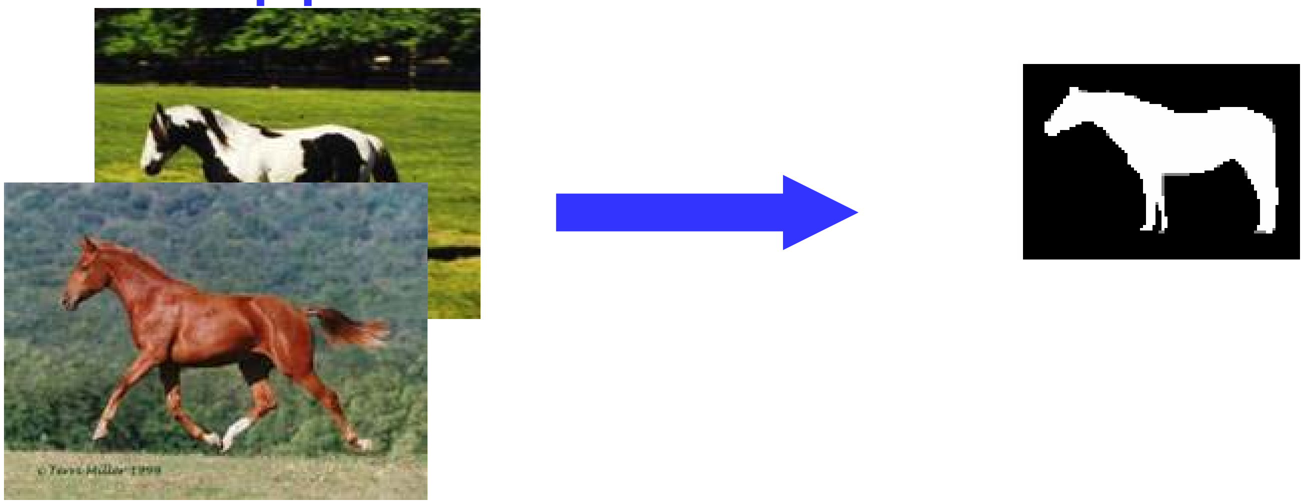


E. Borenstein, S. Ullman. Learning to Segment. ECCV (3) 2004: 315-328

Top-Down - Borenstein & Ullman.

Main contributions

- Top down approach



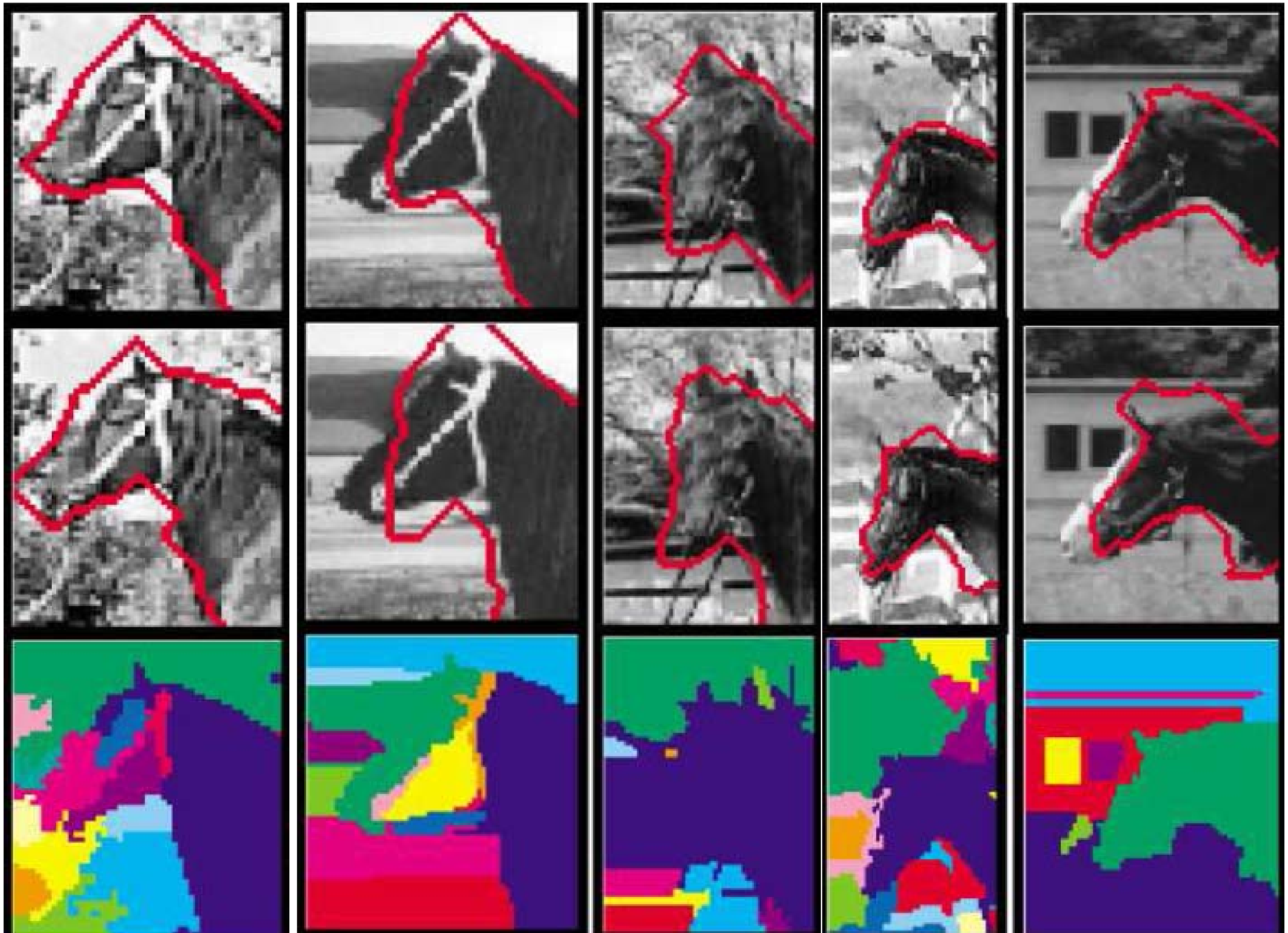
- Automatic fragments labeling

Results

This Alg.
figure-ground
labeling

Manual
figure-ground
labeling

Bottom-up
seg.



E. Borenstein, S. Ullman. Learning to Segment. ECCV (3) 2004: 315-328

Zoom-in to the results

Top-down

Bottom-up

The shape of the horse is captured

The boundaries are not exact

The boundaries are exact

We did not get the horse as one piece

Need to make a combination of
Top-Down & Bottom-Up approaches

Bottom up & Top down segmentations

Training set

Test Image

The “correct” segmentation is some how a compromise between Top-down & bottom up.

How to combine?

- Approach 1 – Do separate segmentations are combine them together.

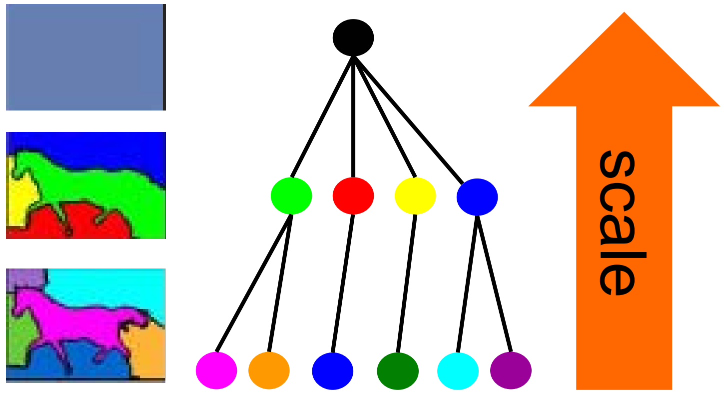
  E. Borenstein, E. Sharon, S. Ullman (2004)

- Approach 2 – Combine them in one pass.

  A. Levin and Y. Weiss (2006)

Combining Top-Down and Bottom-Up Segmentation- Borenstein *et al.*

- Use the hierarchy formed by the bottom-up segmentation.
- Each segment at each scale gets figure/background label.



E. Borenstein, E. Sharon, S. Ullman, Combining Top-Down and Bottom-Up Segmentation, Proceedings IEEE workshop on Perceptual Organization in Computer Vision, IEEE CVPR Washington, DC, June 2004.

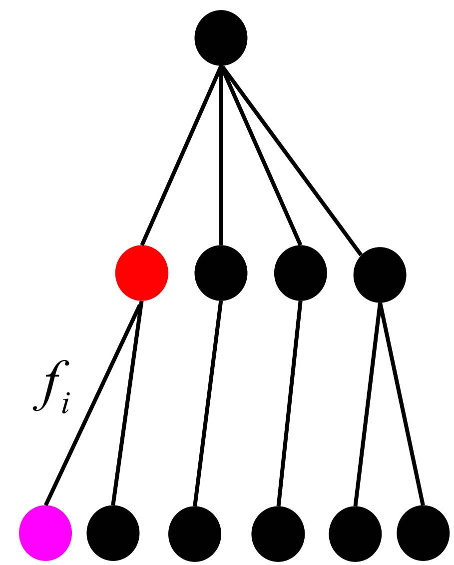
Combining Top-Down and Bottom-Up segmentation

- Each father-descendent cost depends on their labels.

$$TCost(s_1,\ldots,s_N) = \sum_i f_i(s_i, s_i^-)$$

Total Tree cost

father-descendent cost



$$s_i, s_i^- \in \{1, -1\}$$

Descendent label

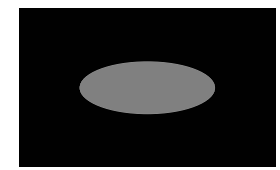
Father label

E. Borenstein, E. Sharon, S. Ullman, Combining Top-Down and Bottom-Up Segmentation, Proceedings IEEE workshop on Perceptual Organization in Computer Vision, IEEE CVPR Washington, DC, June 2004.

Combining Top-Down and Bottom-Up : father-descendent cost

- Each father-descendent energy has two terms:

Image



$$f_i(s_i, s_i^-) = t_i(s_i) + b_i(s_i, s_i^-)$$

Top–down term

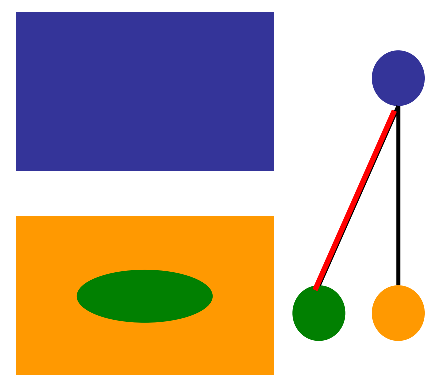
Sum of SSD from the Top-Down segmentation (only on leafs)

Bottom-up term

Descendent more salient - penalty for different father-descendent labels decreases.

$$\lambda |S_i| (1 - h_i)(s_i - s_i^-)^2$$

segmentation tree



E. Borenstein, E. Sharon, S. Ullman, Combining Top-Down and Bottom-Up Segmentation, Proceedings IEEE workshop on Perceptual Organization in Computer Vision, IEEE CVPR Washington, DC, June 2004.

Combining Top-Down and Bottom-Up Segmentation

- Since the cost function factorizes:

$$TCost(s_1,\ldots,s_N) = \sum_i f_i(s_i, s_i^-)$$

Total Tree cost

father-descendent cost

- Sum product algorithm is used (similar to BP)

E. Borenstein, E. Sharon, S. Ullman, Combining Top-Down and Bottom-Up Segmentation, Proceedings IEEE workshop on Perceptual Organization in Computer Vision, IEEE CVPR Washington, DC, June 2004.

Combining Top-Down and Bottom-Up Segmentation

- Requires only one pass in each direction.

- A local computation in each segments results in:

$$Tcost_{s_i}(1) = \min(Tcost(s_1,\ldots,s_{i-1},1,s_{i+1},\ldots,s_N))$$

$$Tcost_{s_i}(-1) = \min(Tcost(s_1,\ldots,s_{i-1},-1,s_{i+1},\ldots,s_N))$$

E. Borenstein, E. Sharon, S. Ullman, Combining Top-Down and Bottom-Up Segmentation, Proceedings IEEE workshop on Perceptual Organization in Computer Vision, IEEE CVPR Washington, DC, June 2004.

Combining Top-Down and Bottom-Up Segmentation

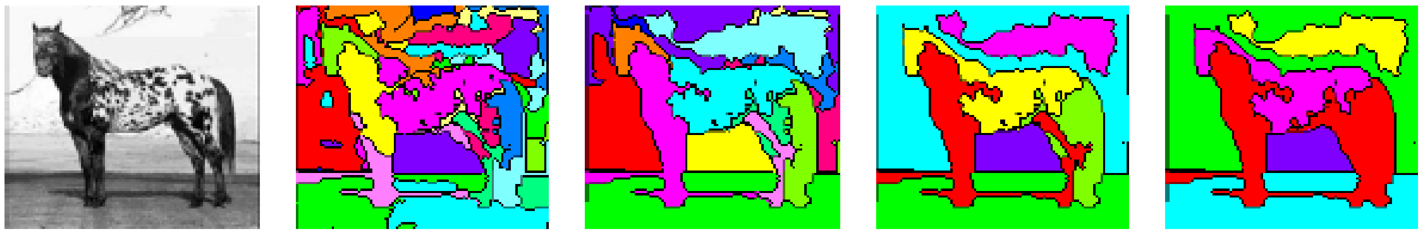
- Confidence map of the segmentation is given by:

$$\frac{\left| m_{S_i}(1) - m_{S_i}(-1) \right|}{|S|}$$

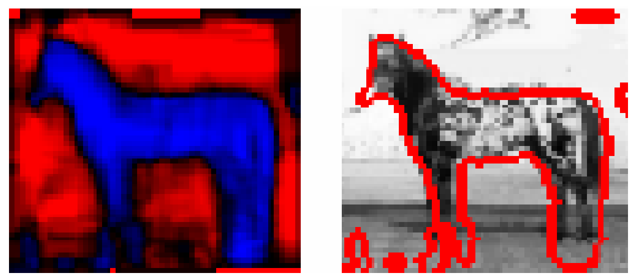
E. Borenstein, E. Sharon, S. Ullman, Combining Top-Down and Bottom-Up Segmentation, Proceedings IEEE workshop on Perceptual Organization in Computer Vision, IEEE CVPR Washington, DC, June 2004.

Results

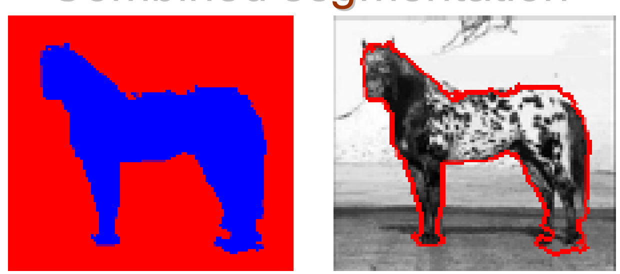
Bottom-Up segmentation



Top-Down segmentation



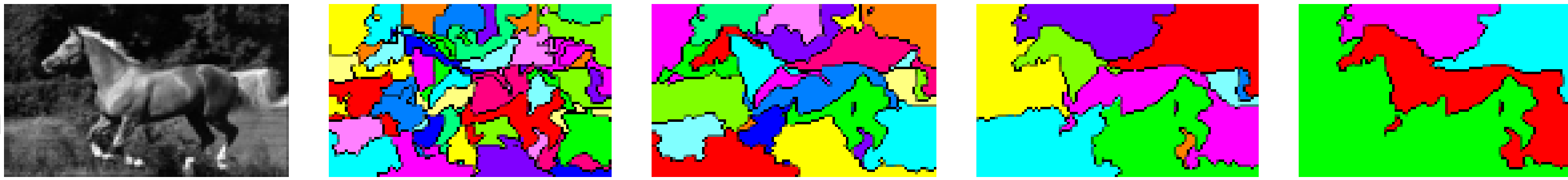
Combined segmentation



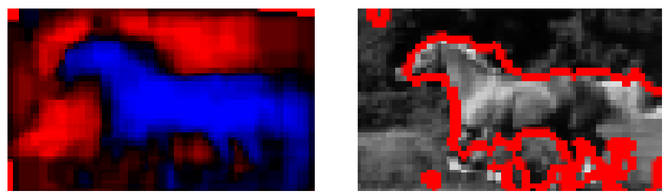
E. Borenstein, E. Sharon, S. Ullman, Combining Top-Down and Bottom-Up Segmentation, Proceedings IEEE workshop on Perceptual Organization in Computer Vision, IEEE CVPR Washington, DC, June 2004.

Results

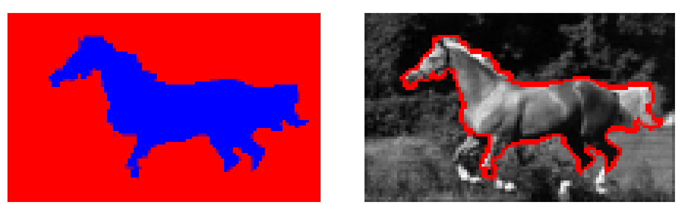
Bottom-Up segmentation



Top-Down segmentation



Combined segmentation



E. Borenstein, E. Sharon, S. Ullman, Combining Top-Down and Bottom-Up Segmentation, Proceedings IEEE workshop on Perceptual Organization in Computer Vision, IEEE CVPR Washington, DC, June 2004.

Combining Top-Down and Bottom-Up Segmentation- A. Levin *et al*.

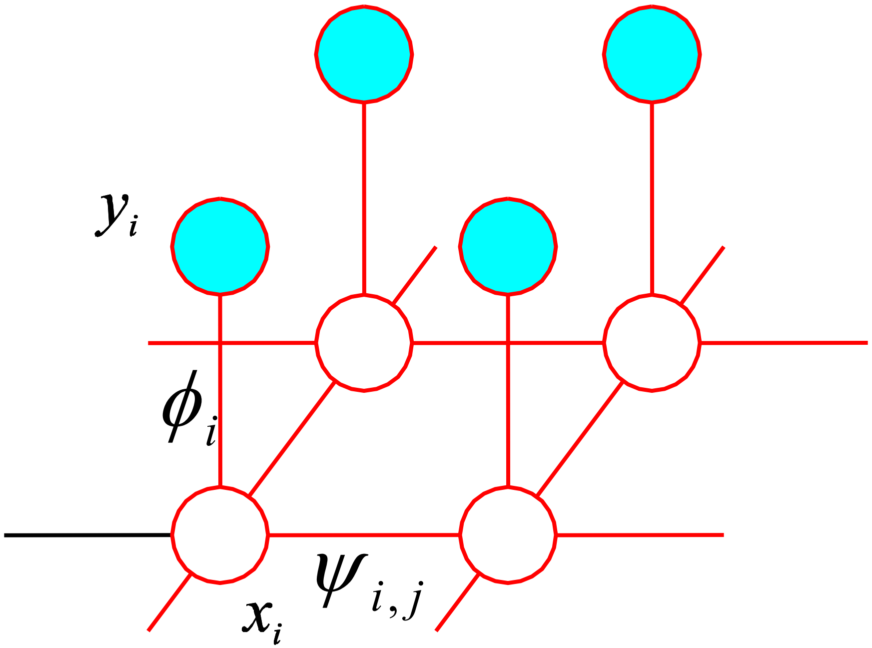
- Combine in one segmentation process.

- Use only few fragments.

Markov Random Field

- Joint distribution of labels and local features

$$p(x, y) = \frac{1}{Z}\prod_{i,j}\psi_{i,j}(x_i, x_j)\prod_i \phi_i(x_i, y_i)$$

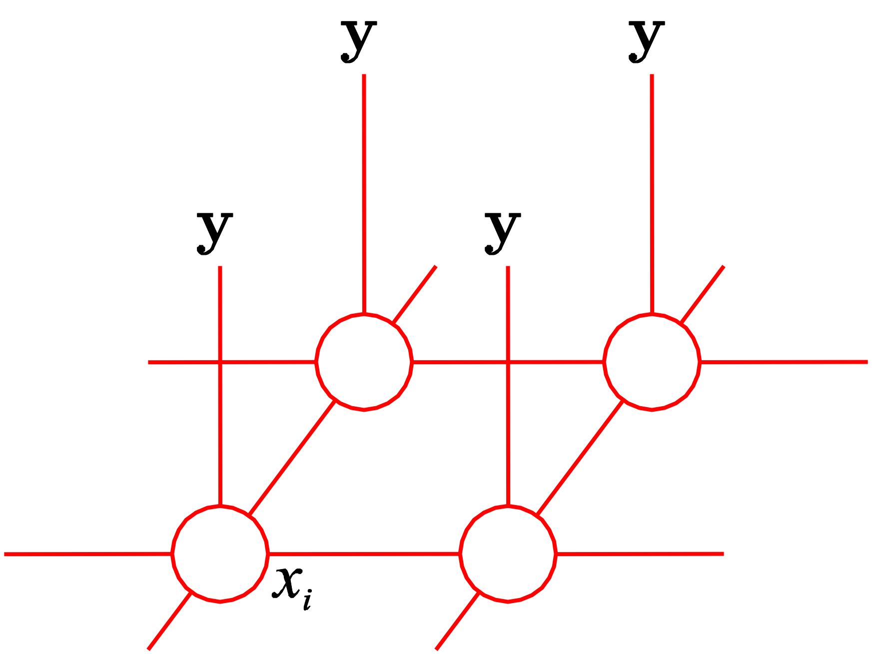


One hidden -- one observation

Conditional Random Field

- Conditional distribution of labels given whole image

$$p(x \mid \mathrm{y}) = \frac{1}{Z(\mathrm{y})} \prod_{i,j} \psi_{i,j}(x_i, x_j; \mathrm{y}) \prod_{i} \phi_i(x_i; \mathrm{y})$$



•Models the conditional and not the joint probability.

•CRF globally conditioned on the observation **Y** – **long range dependences**

Lafferty *et al*. (2001), Kumar and Hebert (2003)

Log notation

$$p(x \mid \mathrm{y}) = \frac{1}{Z(\mathrm{y})} \prod_{i,j} \psi_{i,j}(x_i, x_j; \mathrm{y}) \prod_i \phi_i(x_i; \mathrm{y})$$

Take the log:

$$\overline{\psi}_{i,j}(x_i, x_j; \mathrm{y}) = -\log(\psi_{i,j}(x_i, x_j; \mathrm{y}))$$

$$\overline{\phi}_i(x_i; \mathrm{y}) = -\log(\phi_i(x_i; \mathrm{y}))$$

Easier to work with summations

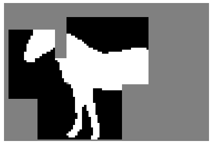
$$p(x \mid \mathrm{y}) = \frac{1}{\overline{Z}(\mathrm{y})} e^{\boxed{-\sum_{i,j} \overline{\psi}_{i,j}(x_i, x_j; \mathrm{y}) + \sum_{i,j} \overline{\phi}_i(x_i; \mathrm{y})}} \quad = E(x; \mathrm{y})$$

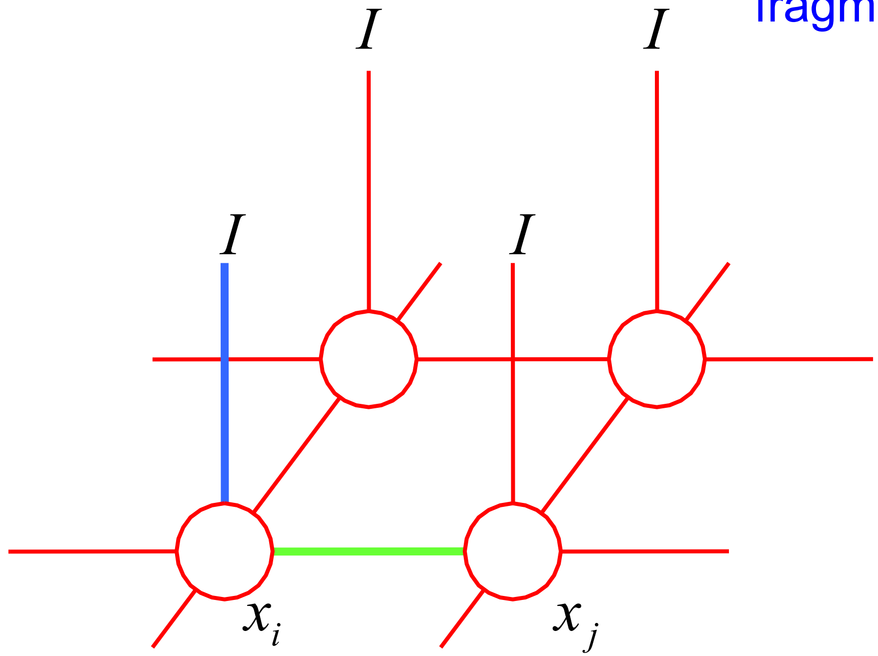
How this helps the segmentation?

$$E(x;I) = \nu \sum_{i,j} w_{ij} |x(i) - x(j)| + \sum_{k} \lambda_k |x - x_{F_k,I}|$$

Low-level term
(Intensity)

Local energy term
derived from image
fragments





A. Levin and Y. Weiss. Learning to Combine Bottom-Up and Top-Down Segmentation. Proc. of the European Conference on Computer Vision (ECCV), Graz, Austria, May 2006.

Learning

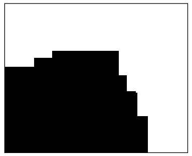
$$E(x;I) = \nu \sum_{i,j} w_{ij} |x(i) - x(j)| + \sum_{k} \lambda_k |x - x_{F_k,I}|$$

Labeling discontinuities == image discontinuities

$$w_{ij} = \frac{1}{1 + \sigma d_{ij}^2}$$ d_{ij}^2 - RGB distance

We want to learn:

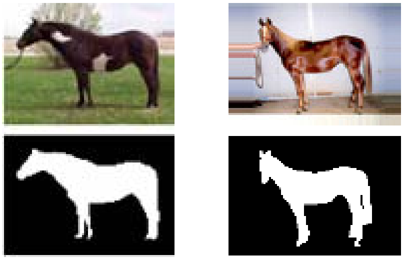
1. ν - power of the low level term

2. $\vec{\lambda}, \vec{F} = \lambda_1$  $+ \lambda_2$  $+ \ldots$

A. Levin and Y. Weiss. Learning to Combine Bottom-Up and Top-Down Segmentation. Proc. of the European Conference on Computer Vision (ECCV), Graz, Austria, May 2006.

Learning

Given a set of segmented training images



$\ell^1(\vec{\lambda},\nu;\vec{F}) + \ell^2(\vec{\lambda},\nu;\vec{F})$

Switch the roll of hidden & observation to get likelihood of the labels *x* conditioned on the image *I.*

Find the parameters that maximize the sum of the log-likelihood :

$$\ell(\vec{\lambda},\nu;\vec{F}) = \sum_t \ell^t(\vec{\lambda},\nu;\vec{F})$$

Where:

$$\ell^t(\vec{\lambda},\nu;\vec{F}) = \log p(x_t \mid I_t;\vec{\lambda},\nu;\vec{F}) =$$

$$-E(x_t;I_t,\vec{\lambda},\nu;\vec{F}) - \log Z(I_t,\vec{\lambda},\nu;\vec{F})$$

A. Levin and Y. Weiss. Learning to Combine Bottom-Up and Top-Down Segmentation. Proc. of the European Conference on Computer Vision (ECCV), Graz, Austria, May 2006.

Learning

- The CRF log likelihood is *convex* with respect to the weighting parameters λ_k, ν
  Lafferty *et al*. (2001)

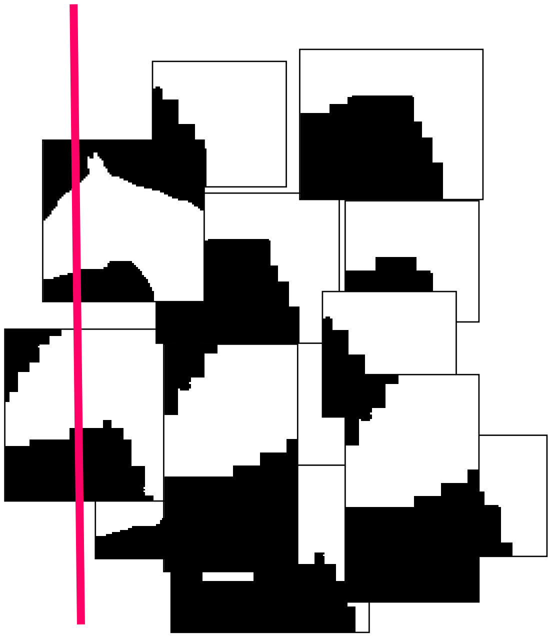
- On the other hand - exact computation of derivatives and *Z*(*I*) is in general intractable.

- Use approximations

A. Levin and Y. Weiss. Learning to Combine Bottom-Up and Top-Down Segmentation. Proc. of the European Conference on Computer Vision (ECCV), Graz, Austria, May 2006.

Learning- selection fragments

Out of a pool of fragments at random sizes & locations we want to select a few informative ones.



Straightforward computation of the likelihood improvement is not practical since each iteration will require inference for each $\vec{\lambda}, \vec{F}$.

How to do it efficiently ?

A. Levin and Y. Weiss. Learning to Combine Bottom-Up and Top-Down Segmentation. Proc. of the European Conference on Computer Vision (ECCV), Graz, Austria, May 2006.

Learning- selection fragments

Recall the energy term: $$E(x;I) = v\sum_{i,j}\left|x(i)-x(j)\right| + \sum_k \lambda_k \left|x - x_{F_k,I}\right|$$

If no fragments are used: $$E_0(x;I) = v\sum_{i,j}\left|x(i)-x(j)\right|$$

If we add one fragment: $$E_1(x;I) = v\sum_{i,j}\left|x(i)-x(j)\right| + \lambda_1 \left|x - x_{F_1,I}\right|$$

Linear approximation: $$E_1(x;I)\,|_{\lambda_1=0} \approx E_0(x;I) + \lambda_1 \frac{\partial E_1(x;I)}{\partial \lambda_1}$$

Energy change by adding F_1

A. Levin and Y. Weiss. Learning to Combine Bottom-Up and Top-Down Segmentation. Proc. of the European Conference on Computer Vision (ECCV), Graz, Austria, May 2006.

Learning- selection fragments

Start without any Fragments $$E_0(x;I) = v\sum_{i,j}\left|x(i)-x(j)\right|$$

If we add one fragment: $$E_1(x;I) = v\sum_{i,j}\left|x(i)-x(j)\right| + \lambda_1\left|x - x_{F_1,I}\right|$$

$$E_1(x;I)\,|_{\lambda_1=0} \approx E_0(x;I) + \lambda_1 \frac{\partial E_1(x;I)}{\partial \lambda_1}$$

Energy change by adding F1

Choose the fragment with large change.

Learning- selection fragments

- Continue in a greedy approach.
- Add a predefined number of fragments.
- After all the fragments are added, improve the accuracy of λ_k.

A. Levin and Y. Weiss. Learning to Combine Bottom-Up and Top-Down Segmentation. Proc. of the European Conference on Computer Vision (ECCV), Graz, Austria, May 2006.

Segmentation

- Given a new image the segmentation is the assignment x that maximizes the probability:

$$p(x \mid I_{new}) = \frac{1}{Z(I_{new})} e^{-E(x;I_{new})}$$

- Where $E(x;I_{new})$ is the learned energy term.

A. Levin and Y. Weiss. Learning to Combine Bottom-Up and Top-Down Segmentation. Proc. of the European Conference on Computer Vision (ECCV), Graz, Austria, May 2006.

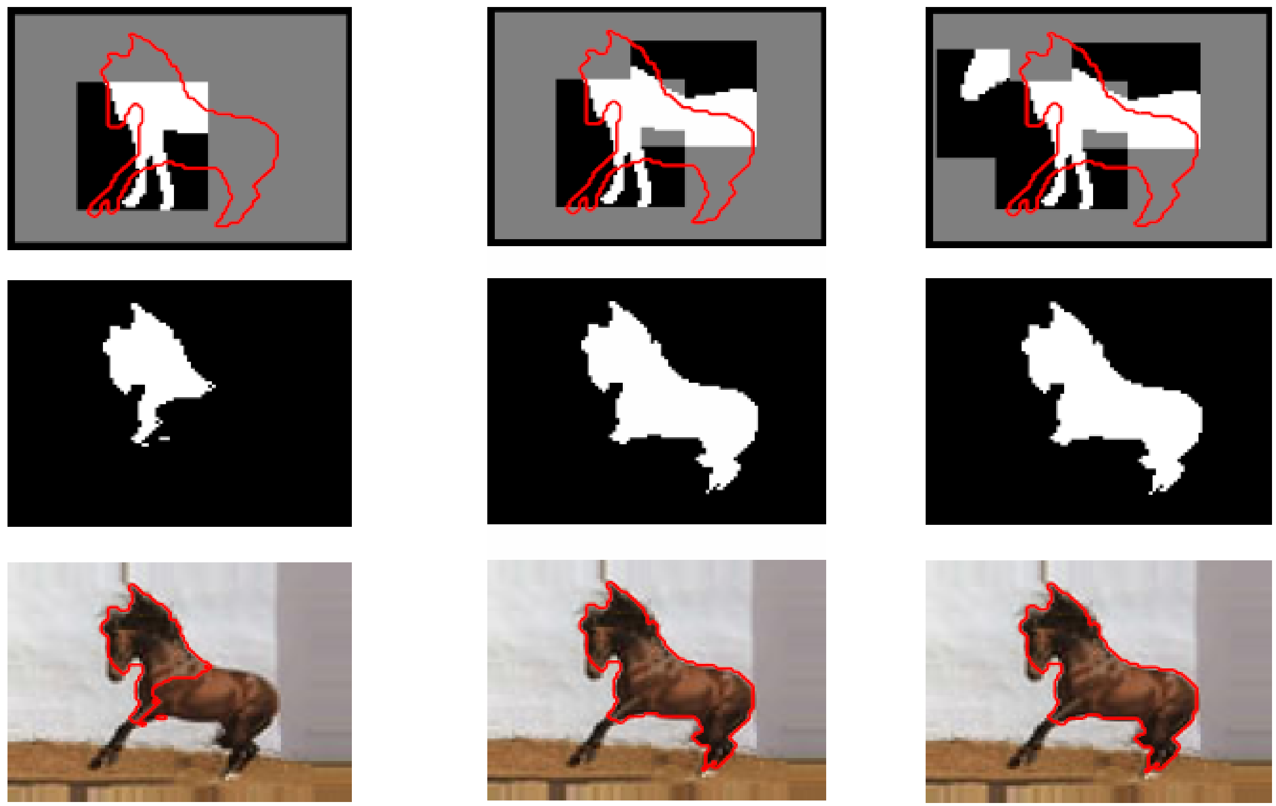
Results

One fragment | Two fragments | Three fragments

Fragments used:

MAP segmentation

Overlaid segmentation



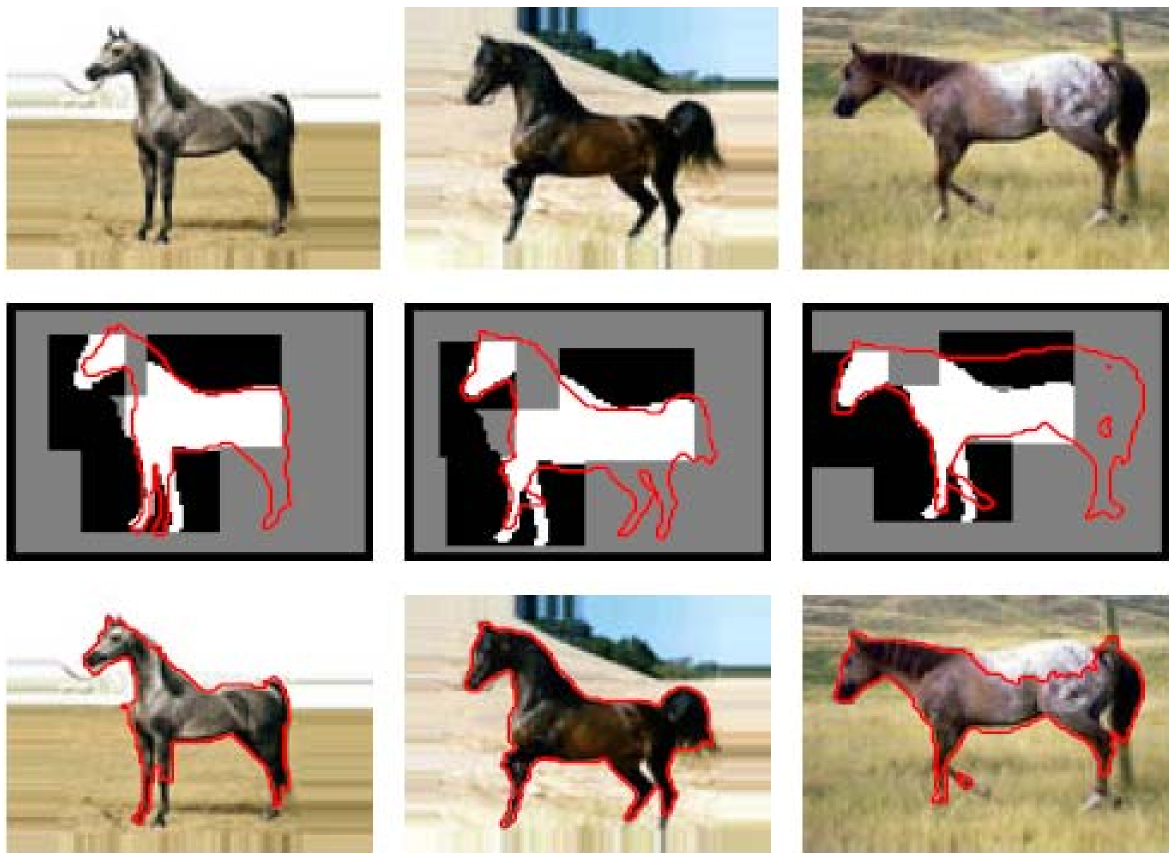
A. Levin and Y. Weiss. Learning to Combine Bottom-Up and Top-Down Segmentation. Proc. of the European Conference on Computer Vision (ECCV), Graz, Austria, May 2006.

Results

Original image

Fragments used:

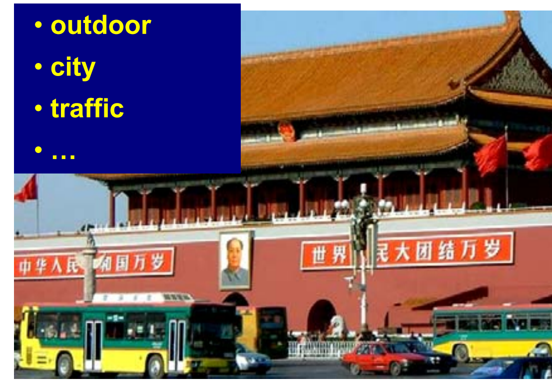
Overlaid segmentation



A. Levin and Y. Weiss. Learning to Combine Bottom-Up and Top-Down Segmentation. Proc. of the European Conference on Computer Vision (ECCV), Graz, Austria, May 2006.

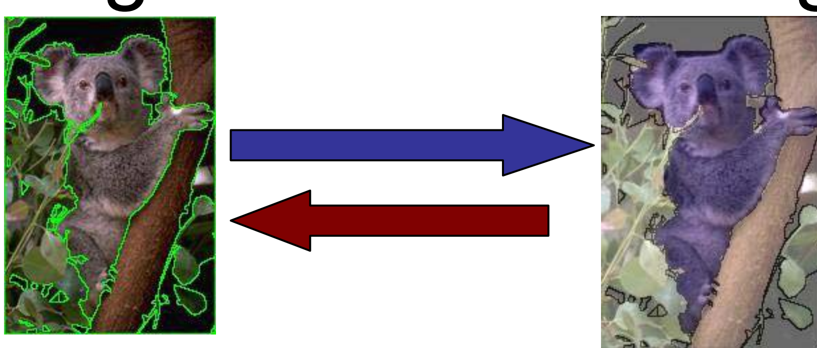
Summary

- Scene Labeling



- Recognition given explicit model

- Segmentation & recognition



Thanks!