Image Segmentation: beyond Graph Cuts

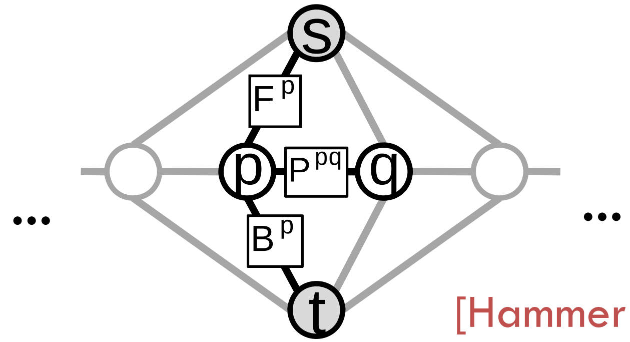
Microsoft Research Cambridge

Victor Lempitsky

Graph cut segmentation

Graph cut segmentation [Boykov&Jolly 01]:

$$E(\mathbf{x}) = \sum_{p \in \mathcal{V}} F^p \cdot x_p + \sum_{p \in \mathcal{V}} B^p \cdot (1 - x_p) + \sum_{p,q \in \mathcal{E}} P^{pq} \cdot |x_p - x_q|$$



[Hammer 64]

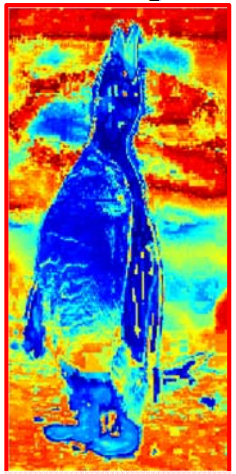
[Greig, Porteous, Seheult 89]

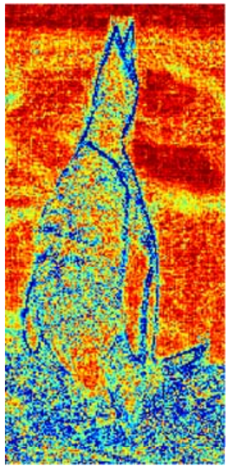
Alternative notation:

$$E(\mathbf{x}) = \sum_{p \in \mathcal{B}} U^p \cdot x_p + \sum_{\{p,q\} \in \mathcal{E}} V^{pq} \cdot |x_p - x_q|, \quad x_p \in \{0, 1\}$$

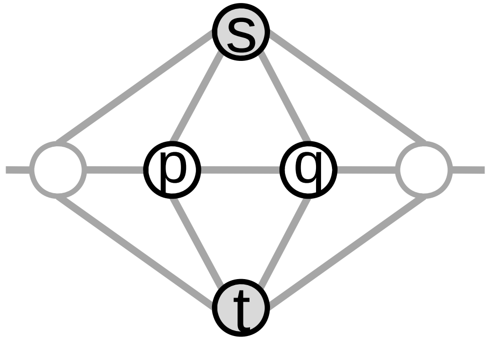
Example: Interactive segmentation

$$E(\mathbf{x}) = \sum_{p \in \mathcal{B}} U^p \cdot x_p + \sum_{\{p,q\} \in \mathcal{E}} V^{pq} \cdot |x_p - x_q|, \quad x_p \in \{0, 1\}$$



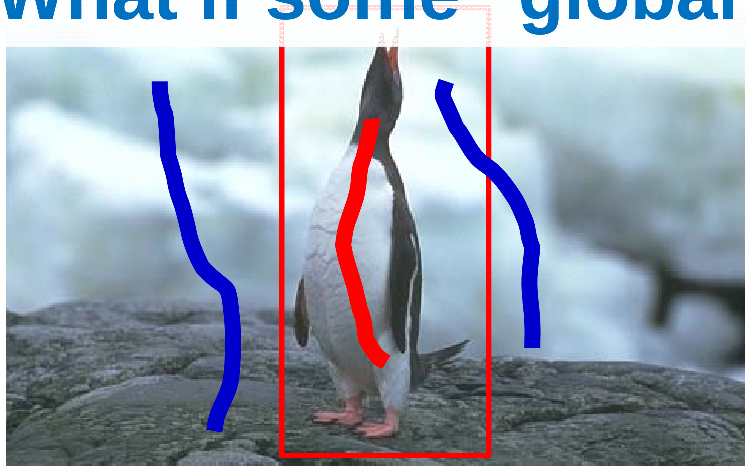


...



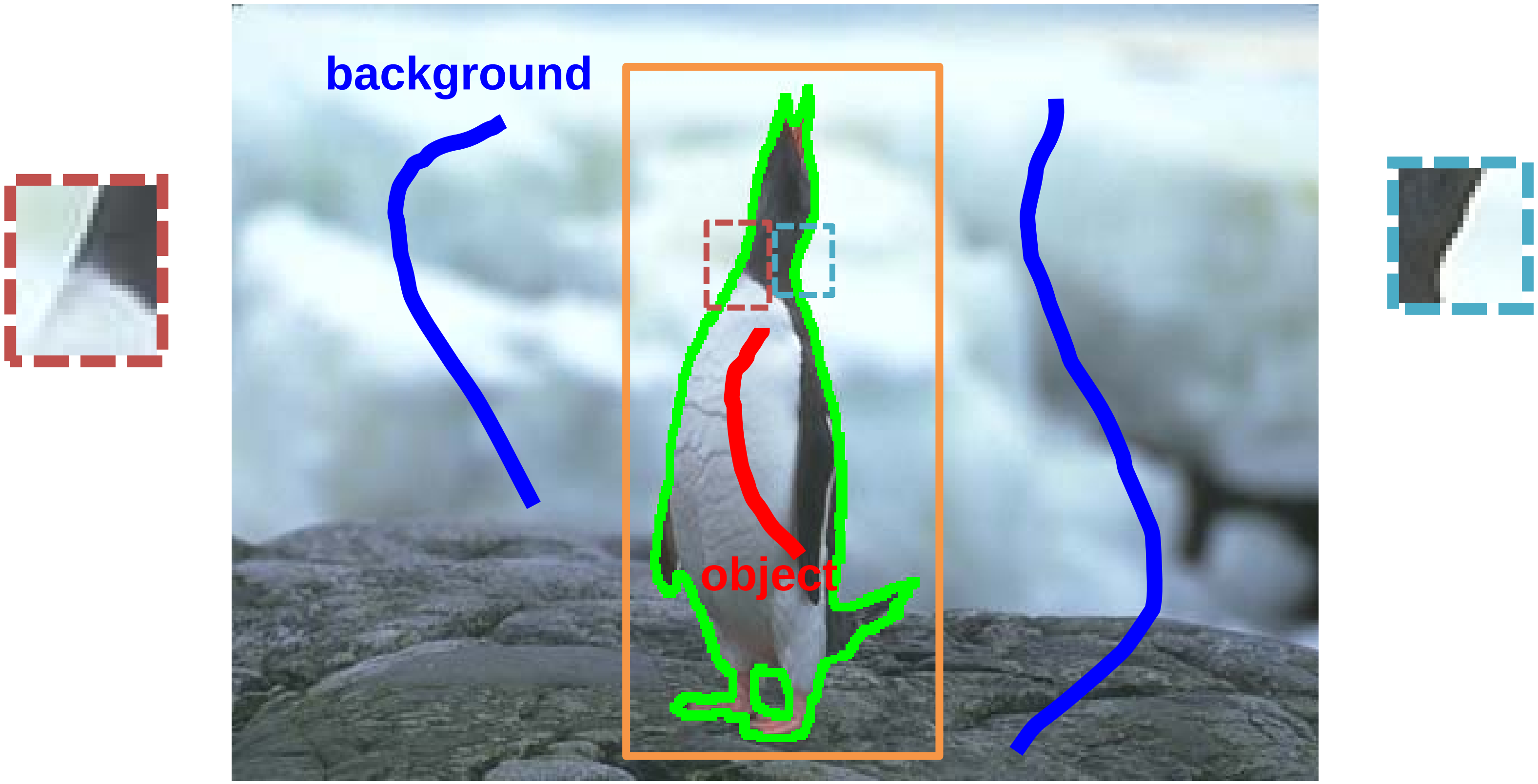
...

What if some "global" cues are also available?



- Integrating **local** cues
- ...but getting **global** solutions
- Many application scenarios...

Image segmentation: the problem



- Prior knowledge (“compactness”)
- Low-level cues (e.g. edge cues)
- High-level knowledge (e.g. “Penguin on a rock”)
- User input

Why image segmentation?

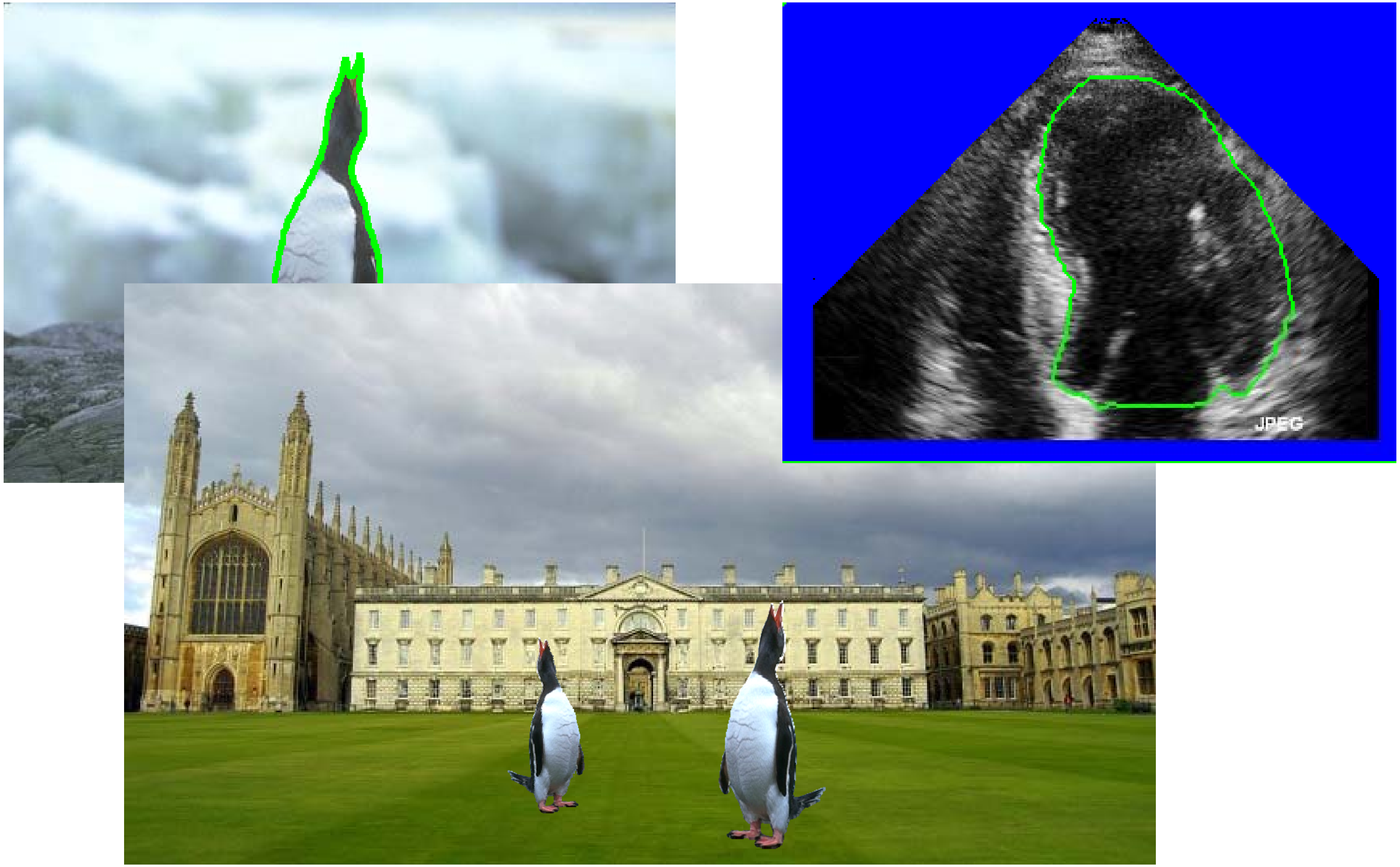


Image segmentation: the story

Since long Ago:

rule-based methods, such as thresholding or region growing (magic wand)

Since 1988 [Kass, Witkin, Terzopoulos]:

energy optimization via local curve evolution

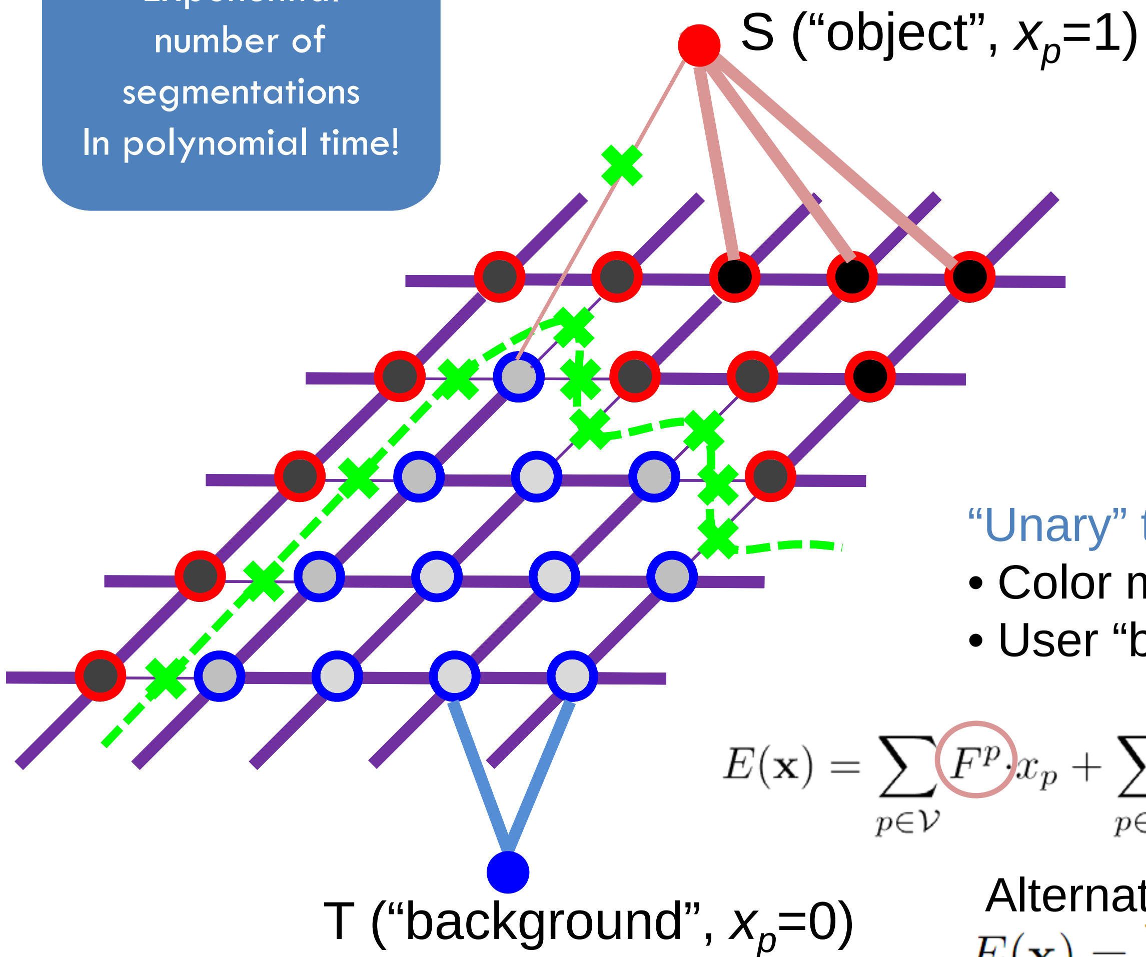
Since 2001[Boykov, Jolly]:

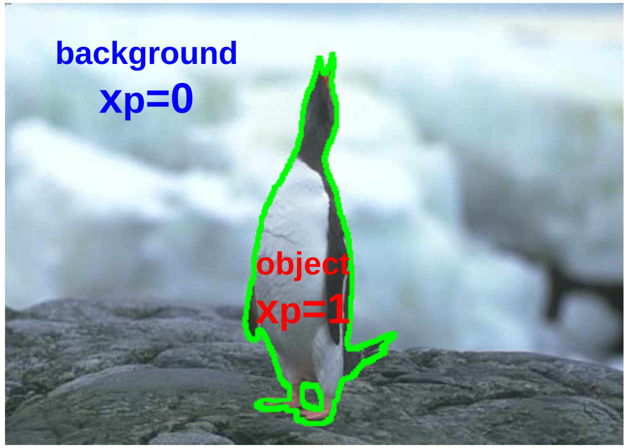
global energy optimization via **graph cuts** *(st-mincut)*

Graph cut segmentation

[Boykov and Jolly, 2001]

Exponential number of segmentations
In polynomial time!





"Unary" terms:
- Color models
- User "brushes"

"Pairwise" terms:
- Ising prior
- Edge cues

$$E(\mathbf{x}) = \sum_{p \in \mathcal{V}} F^p . x_p + \sum_{p \in \mathcal{V}} B^p . (1 - x_p) + \sum_{p,q \in \mathcal{E}} P^{pq} . |x_p - x_q|$$

Alternative notation:

$$E(\mathbf{x}) = \sum_{p \in \mathcal{B}} U^p . x_p + \sum_{\{p,q\} \in \mathcal{E}} V^{pq} . |x_p - x_q|$$

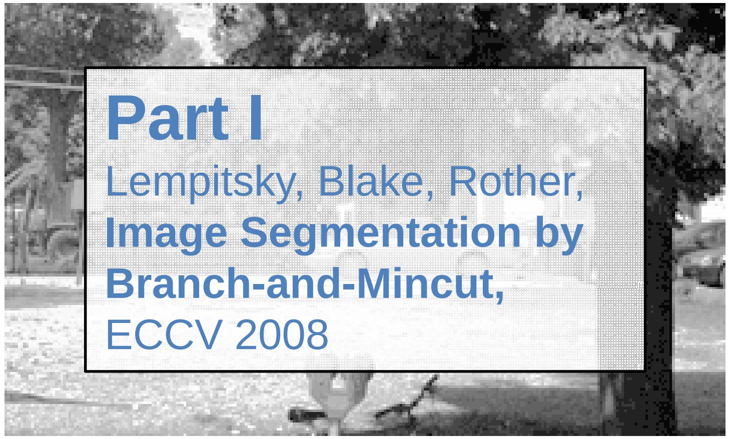
Why go beyond?

We **can** express:

1. Edge cues & Ising prior (via P^{pq})
2. Brushes (set F^p or B^p to ∞)
3. Color likelihoods (via F^p and B^p)

$$E(\mathbf{x}) = \sum_{p \in \mathcal{V}} F^p \cdot x_p + \sum_{p \in \mathcal{V}} B^p \cdot (1 - x_p) + \sum_{p,q \in \mathcal{E}} P^{pq} \cdot |x_p - x_q|$$

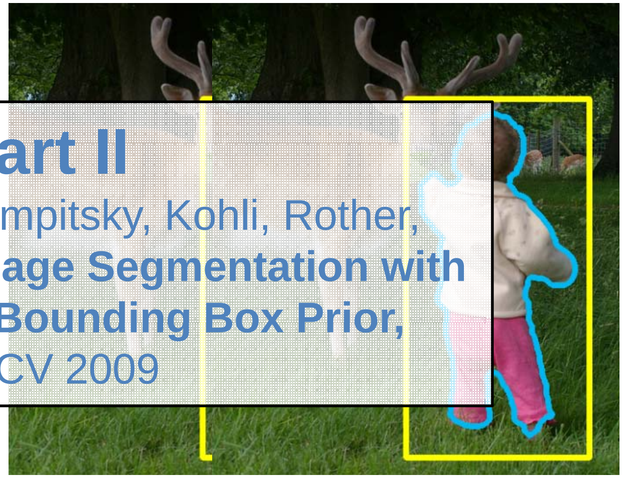
but



Part I

Lempitsky, Blake, Rother,
Image Segmentation by Branch-and-Mincut,
ECCV 2008

How to segment a **car** in the image?



Part II

Lempitsky, Kohli, Rother,
Image Segmentation with A Bounding Box Prior,
ICCV 2009

How to ensure **tightness** of the bounding bo

Still want non-local and efficient optimization!

Image Segmentation by Branch-and-Mincut

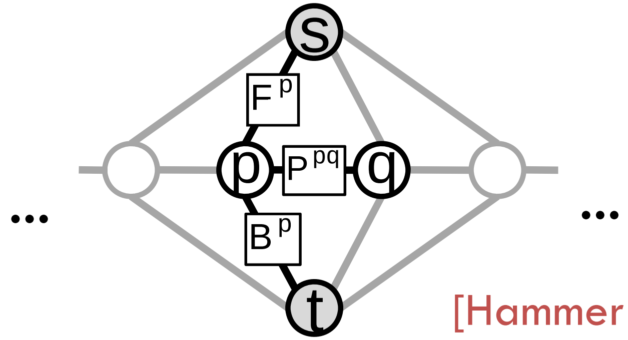
ECCV 2008

Microsoft Research Cambridge

Victor Lempitsky
Andrew Blake
Carsten Rother

An example

image from UIUC car dataset

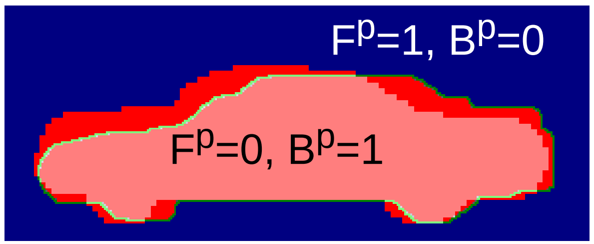


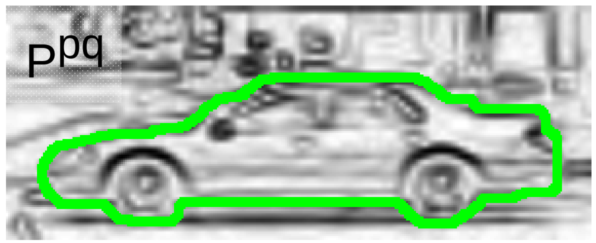
[Hammer 64]

[Greig, Porteous, Seheult 89]

Standard "graph cut" segmentation energy [Boykov, Jolly 01]:

$$E(\mathbf{x}) = \sum_{p \in \mathcal{V}} F^p \cdot x_p + \sum_{p \in \mathcal{V}} B^p \cdot (1 - x_p) + \sum_{p,q \in \mathcal{E}} P^{pq} \cdot |x_p - x_q|$$

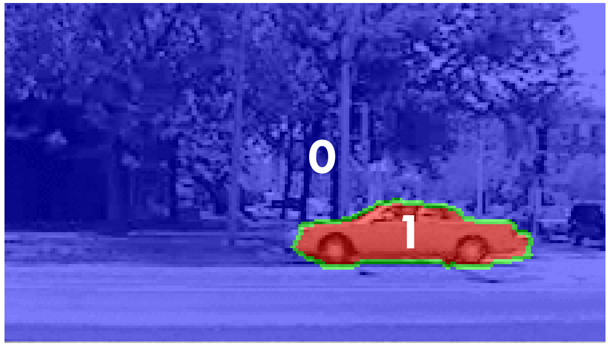




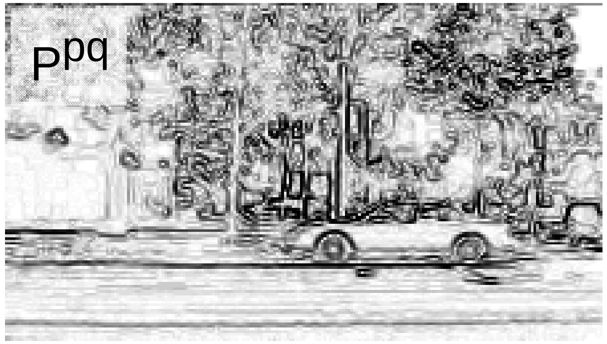
[Freedman, Zhang 05], [Ali, Farag, El-Baz 07],...

A harder example

image from UIUC car dataset



Optimal x



$$E(\mathbf{x}\quad) = \sum_{p\in\mathcal{V}} F^p \quad \cdot x_p + \sum_{p\in\mathcal{V}} B^p \quad \cdot (1 - x_p) + \sum_{p,q\in\mathcal{E}} P^{pq} \cdot |x_p - x_q|$$

Optimal ω

→ **min**
x, ω

Energy optimization

$\omega -$ *global* parameter ($\omega \in \Omega_0$)

$$E(\mathbf{x} \quad) = C \quad + \sum_{p \in \mathcal{V}} F^p \quad \cdot x_p + \sum_{p \in \mathcal{V}} B^p \quad \cdot (1 - x_p) + \sum_{p,q \in \mathcal{E}} P^{pq} \quad \cdot |x_p - x_q|$$

constant unary potentials (costs) pairwise potentials (costs) ≥ 0

....shape priors, color distribution/intensity priors (*Chan-Vese, GrabCut*)...

Optimization options:

- Choose reasonable ω, solve for X
  [Freedman, Zhang 05], [Pawan Kumar, Torr, Zisserman'*ObjCut* 05], [Ali, Farag, El-Baz 07] ....
- Alternate between X and ω (EM)
  [Rother, Kolmogorov, Blake' *GrabCut* 04], [Bray, Kohli, Torr'*PoseCut* 06], [Kim, Zabih 03]....
- Optimize continuously
  [Chan, Vese 01], [Leventon, Grimson, Faugeras 00], [Cremers, Osher, Soatto 06], [Wang, Staib 98]...
- Exhaustive search

Our approach

$$E(\mathbf{x},\omega) = C(\omega) + \sum_{p\in\mathcal{V}} F^p(\omega)\cdot x_p + \sum_{p\in\mathcal{V}} B^p(\omega)\cdot(1-x_p) + \sum_{p,q\in\mathcal{E}} P^{pq}(\omega)\cdot|x_p - x_q|$$

along ω dimension

- Low-dimensional (discretized) domain
- Function of the general form

along x dimension

- Extremely large, structured domain
- Specific "graph cut" function

Branch-and-bound

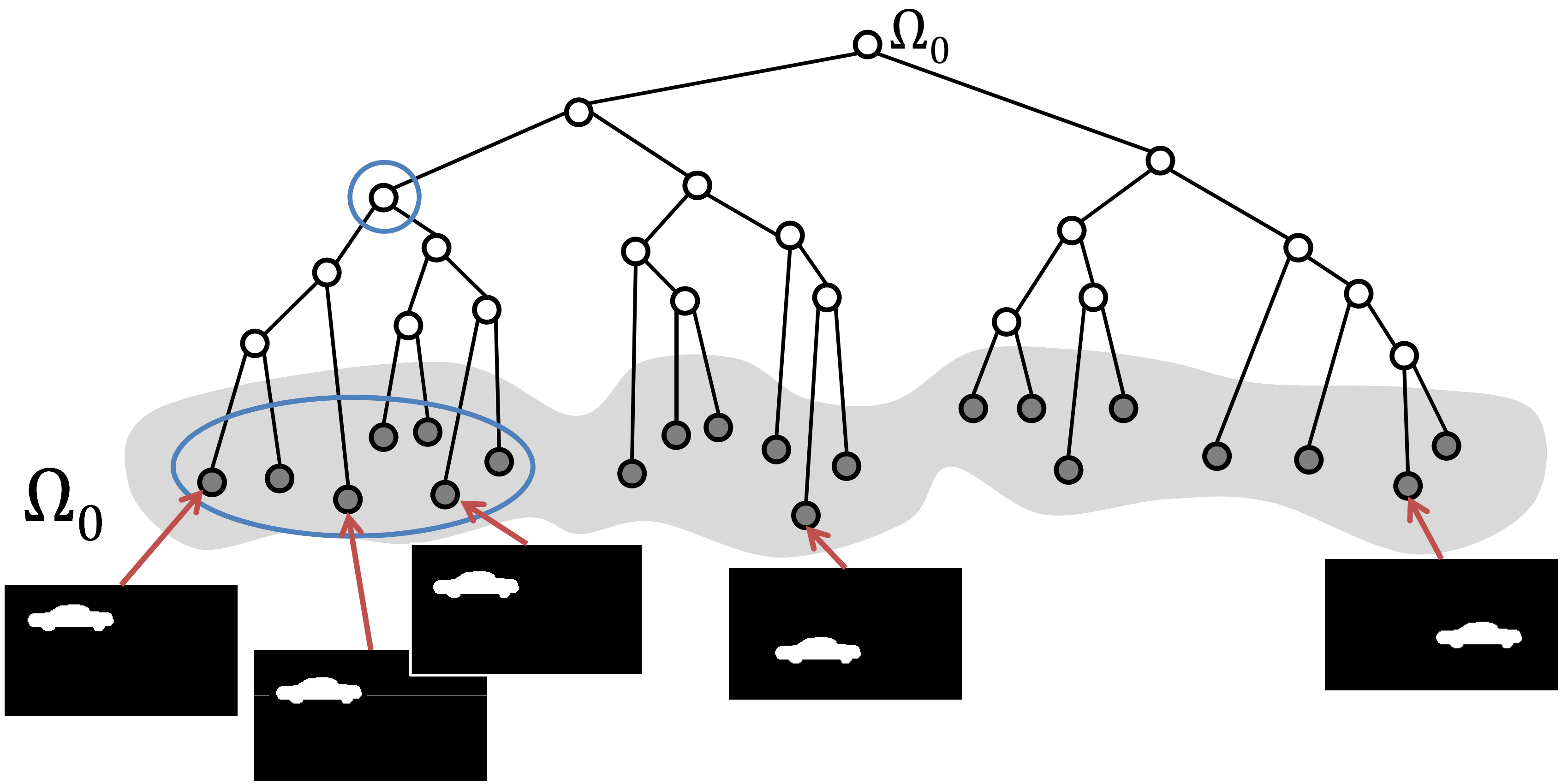
Mincut

Branch-and-Mincut

[Gavrila, Philomin 99], [Lampert, Blaschko, Hofman 08], [Cremers, Schmidt, Barthel 08]

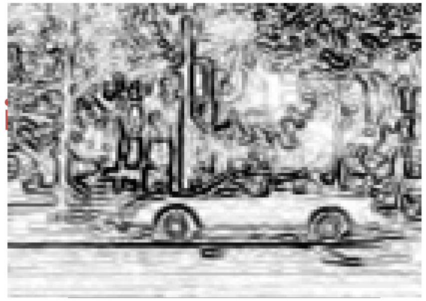
Search tree

$$E(\mathbf{x},\omega) = C(\omega) + \sum_{p\in\mathcal{V}} F^p(\omega)\cdot x_p + \sum_{p\in\mathcal{V}} B^p(\omega)\cdot(1-x_p) + \sum_{p,q\in\mathcal{E}} P^{pq}(\omega)\cdot|x_p - x_q|$$



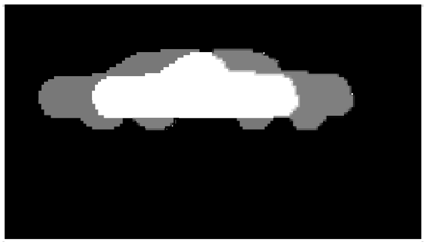
Bounding the energy: an example

$$E(\mathbf{x},\omega) = \sum_{p\in\mathcal{V}} F^p(\omega)\cdot x_p + \sum_{p\in\mathcal{V}} B^p(\omega)\cdot(1-x_p) + \sum_{p,q\in\mathcal{E}} P^{pq}\cdot|x_p - x_q|$$



$$F(\mathbf{x}) = \min_{\omega=\{\omega_1,\omega_2\}} \left[\sum_{p\in\mathcal{V}} F^p(\omega)\cdot x_p + \sum_{p\in\mathcal{V}} B^p(\omega)\cdot(1-x_p) + \sum_{p,q\in\mathcal{E}} P^{pq}\cdot|x_p - x_q| \right]$$

$$\min(a+b) \geq \min a + \min b$$

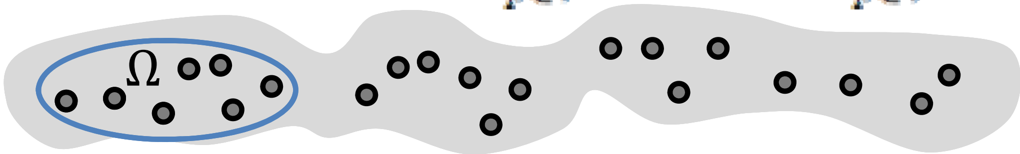


$$\min_{\mathbf{x}\in 2^{\mathcal{V}}} F(\mathbf{x}) \longrightarrow \text{2 mincuts}$$

$$\min_{\mathbf{x}\in 2^{\mathcal{V}}} G(\mathbf{x}) \longrightarrow \text{1 mincut}$$

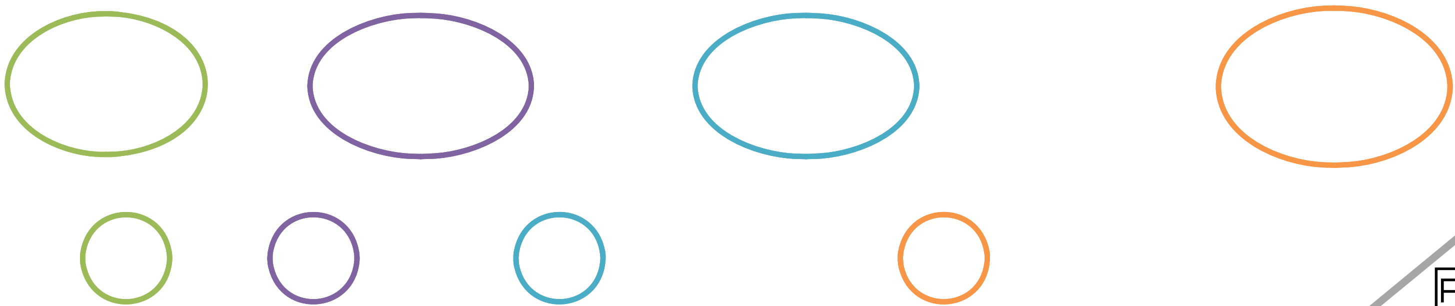
The lower bound

$$E(\mathbf{x},\omega) = C(\omega) + \sum_{p\in\mathcal{V}} F^p(\omega)\cdot x_p + \sum_{p\in\mathcal{V}} B^p(\omega)\cdot(1-x_p) + \sum_{p,q\in\mathcal{E}} P^{pq}(\omega)\cdot|x_p - x_q|$$

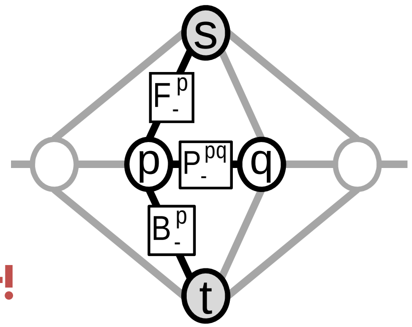


$$\min_{\mathbf{x}\in 2^{\mathcal{V}},\omega\in\Omega} E(\mathbf{x},\omega) =$$

$$\min_{\mathbf{x}\in 2^{\mathcal{V}}} \min_{\omega\in\Omega} \left[C(\omega) + \sum_{p\in\mathcal{V}} F^p(\omega)\cdot x_p + \sum_{p\in\mathcal{V}} B^p(\omega)\cdot(1-x_p) + \sum_{p,q\in\mathcal{E}} P^{pq}(\omega)\cdot|x_p - x_q| \right]$$

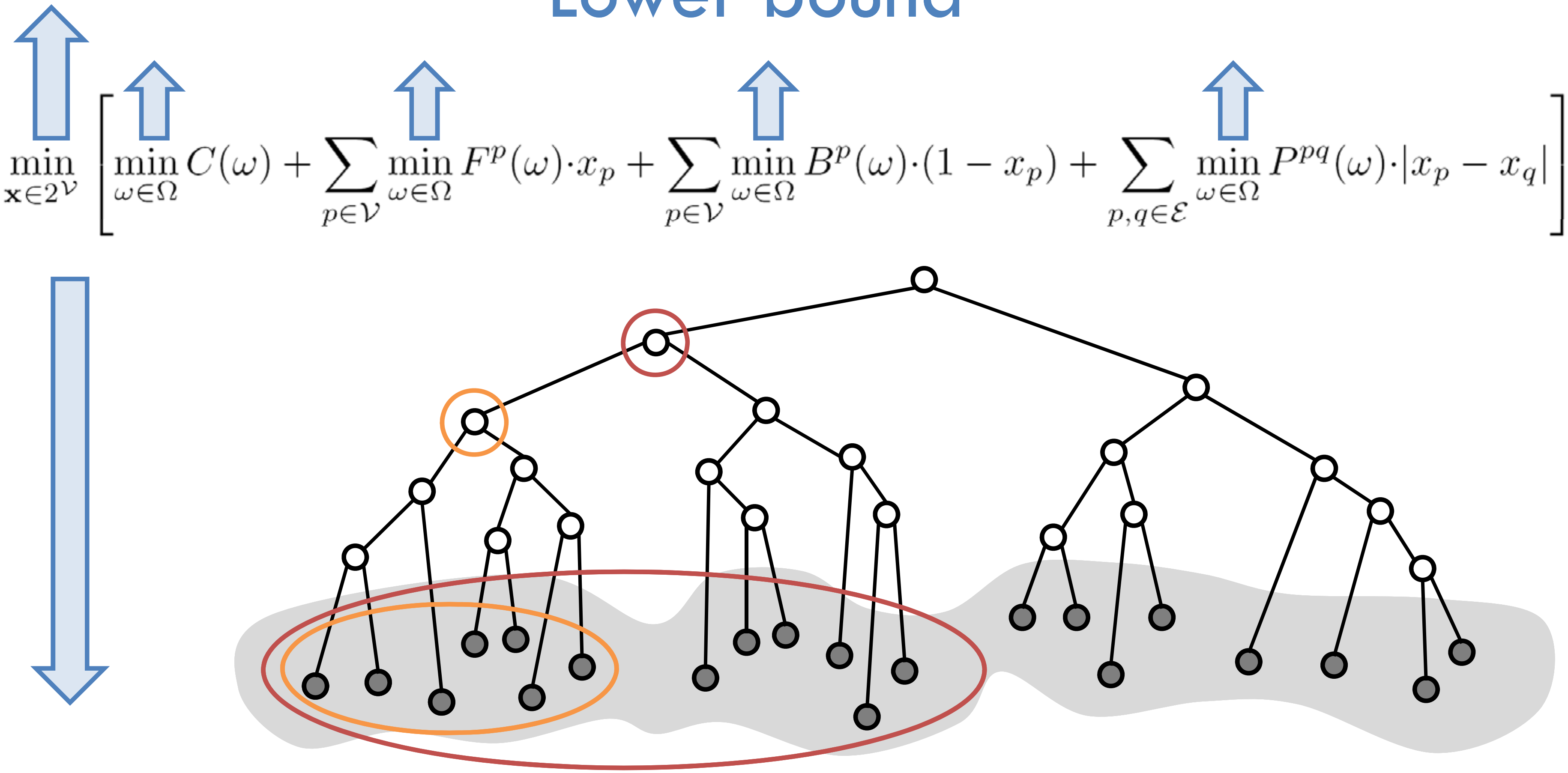


Computable with mincut!



Lower bound

$$\min_{\mathbf{x}\in 2^{\mathcal{V}}} \left[\min_{\omega\in\Omega} C(\omega) + \sum_{p\in\mathcal{V}} \min_{\omega\in\Omega} F^p(\omega)\cdot x_p + \sum_{p\in\mathcal{V}} \min_{\omega\in\Omega} B^p(\omega)\cdot(1-x_p) + \sum_{p,q\in\mathcal{E}} \min_{\omega\in\Omega} P^{pq}(\omega)\cdot|x_p - x_q| \right]$$



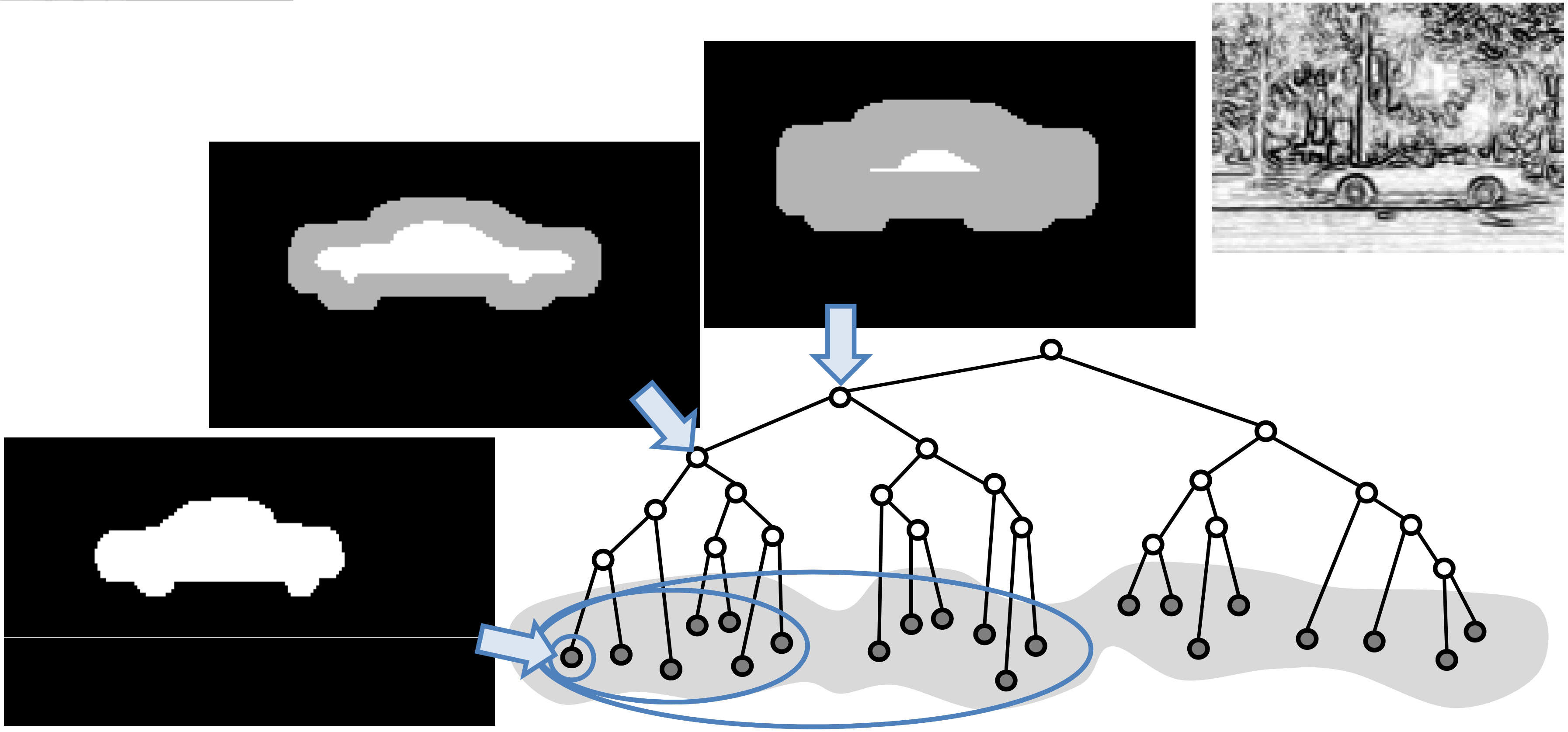
- Monotonic increase towards leaves
- *Tightness* at leaf nodes:

$$L(\{\omega\}) = \min_{\mathbf{x}} E(\mathbf{x},\omega)$$

Lower bound: example

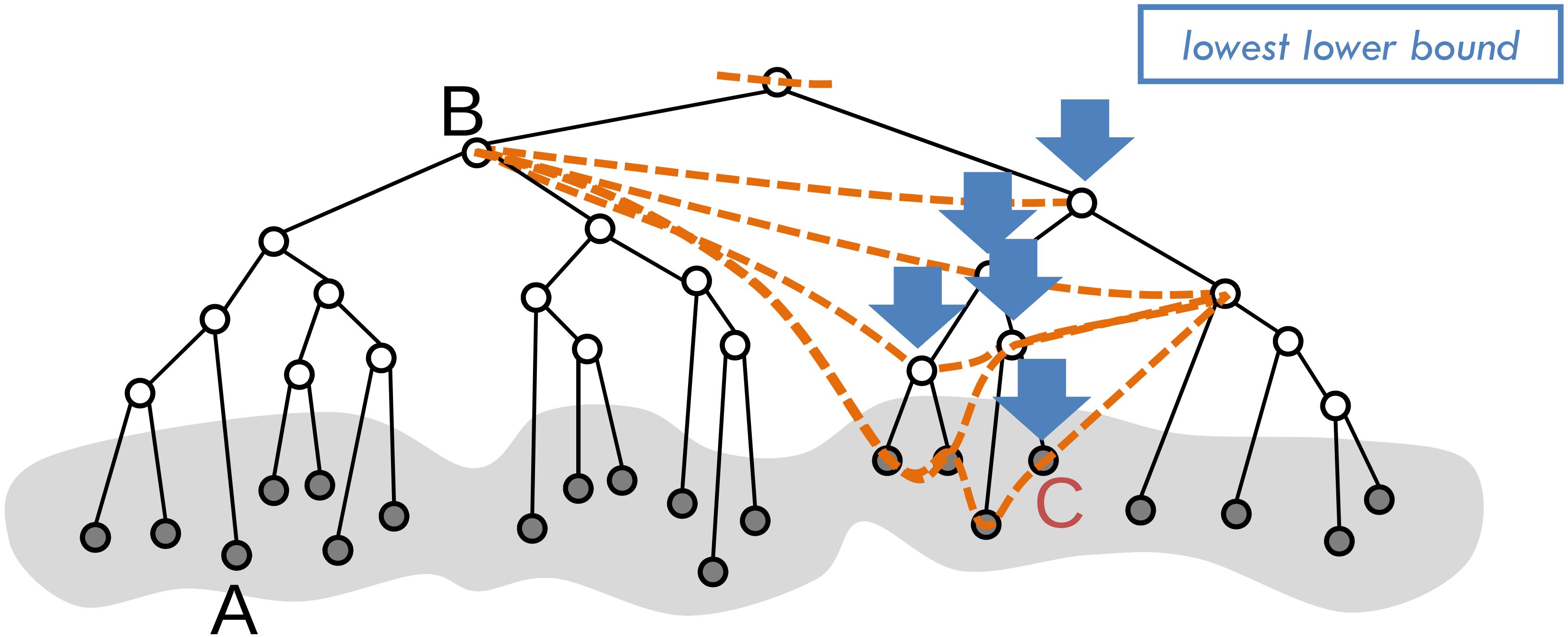
precomputed — computed at runtime

$$\min_{\mathbf{x}\in 2^{\mathcal{V}}} \left[\sum_{p\in\mathcal{V}} \min_{\omega\in\Omega} F^p(\omega)\cdot x_p + \sum_{p\in\mathcal{V}} \min_{\omega\in\Omega} B^p(\omega)\cdot(1-x_p) + \sum_{p,q\in\mathcal{E}} P^{pq}\cdot|x_p - x_q| \right]$$



Branch-and-Bound

Standard *best-first* branch-and-bound search:



$$\min_{\mathbf{x}} E(\mathbf{x}, A) \geq L(B) \geq L(C) = \min_{\mathbf{x}} E(\mathbf{x}, C)$$

Small fraction of nodes is visited

additional speed-up from "reusing" maxflow computations [Kohli,Torr 05]

Results: shape prior

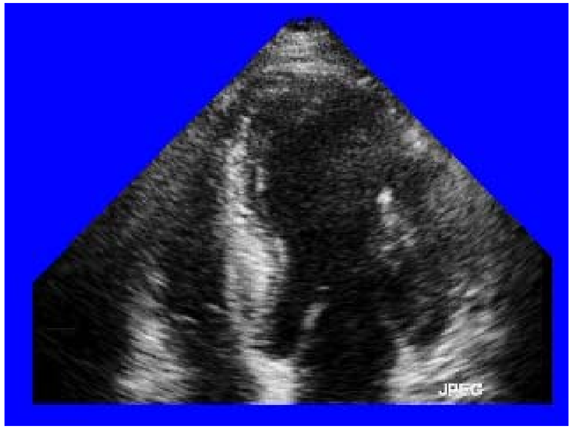
30,000,000 shapes

Exhaustive search: 30,000,000 mincuts
Branch-and-Mincut: 12,000 mincuts

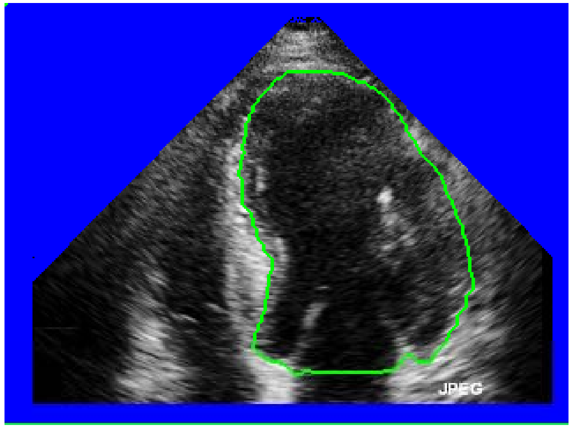
Speed-up: 2500 times
(30 seconds per 312x272 image)

Results: shape prior

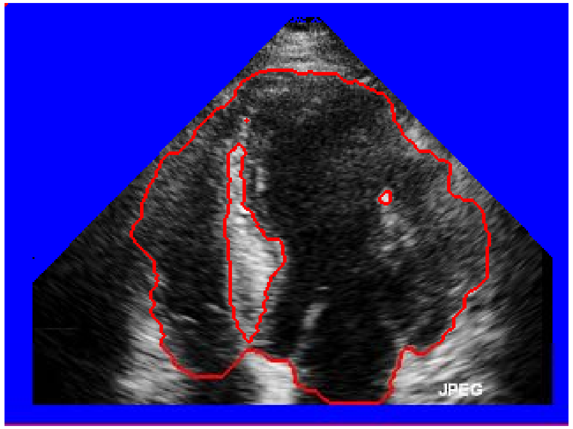
Left ventricle epicardium tracking *(work in progress)*



Original sequence



Our segmentation
Shape prior from other sequences
5,200,000 templates
≈20 seconds per frame
Speed-up 1150



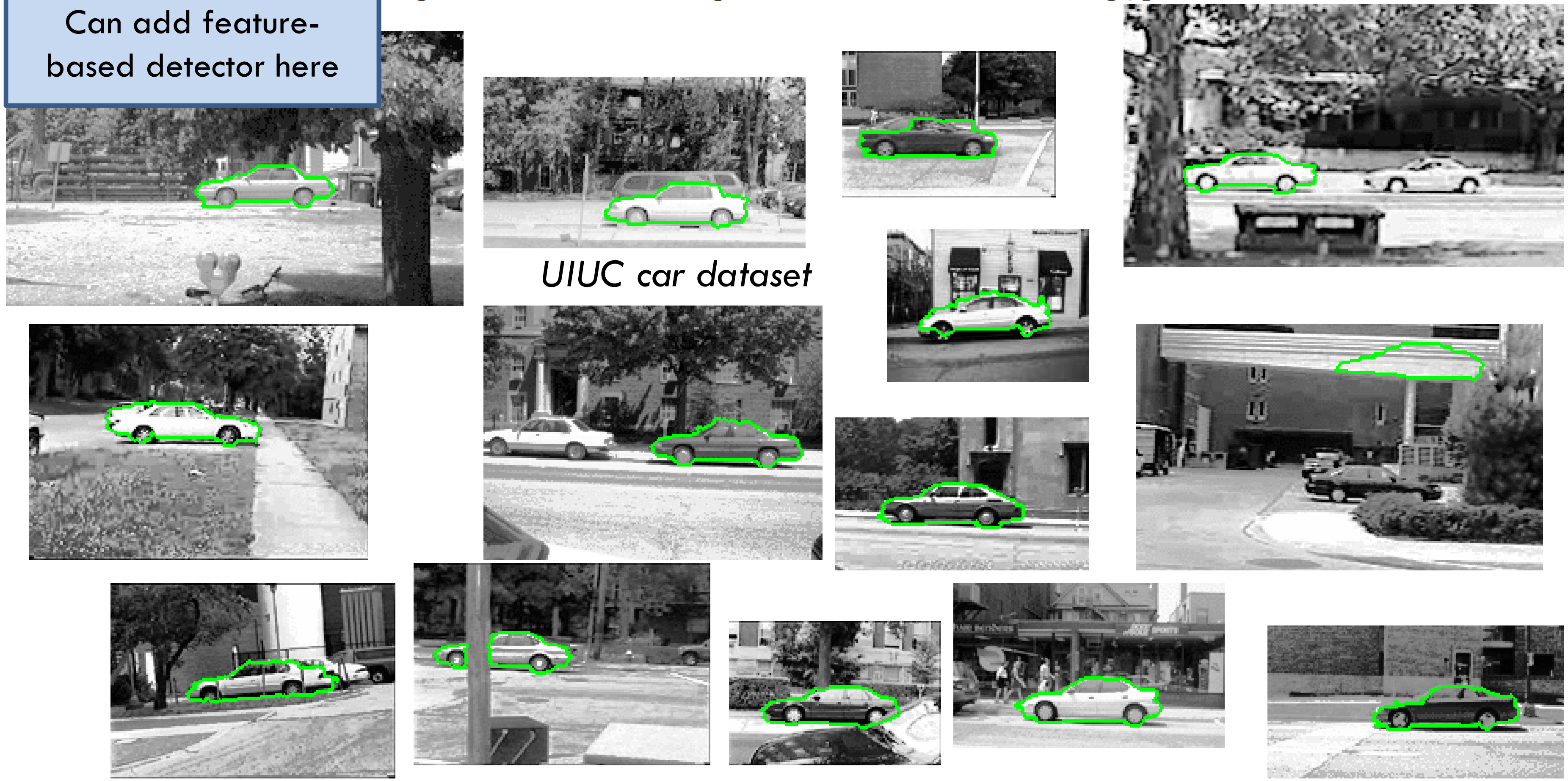
No shape prior

Data courtesy: Dr Harald Becher, Department of Cardiovascular Medicine, University of Oxford

Result: shape prior

$$E(\mathbf{x},\omega) = C(\omega) + \sum_{p\in\mathcal{V}} F^p(\omega)\cdot x_p + \sum_{p\in\mathcal{V}} B^p(\omega)\cdot(1-x_p) + \sum_{p,q\in\mathcal{E}} P^{pq}(\omega)\cdot|x_p - x_q|$$

Can add feature-based detector here

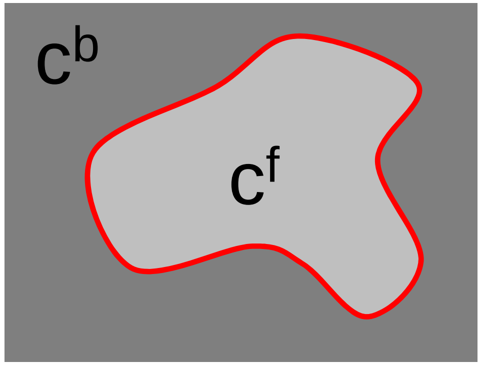


UIUC car dataset

Results: Discrete Chan-Vese functional

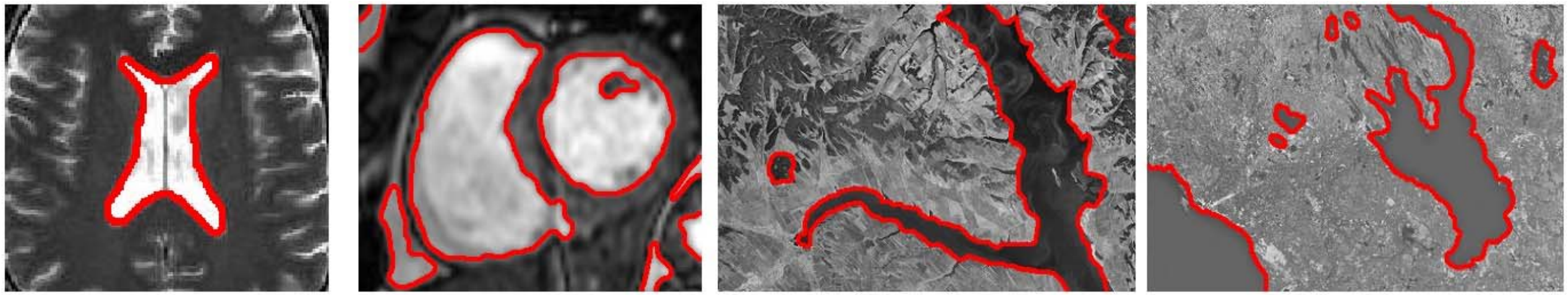
Chan-Vese functional [Chan, Vese 01]:

$$E(\mathbf{x}, c^f, c^b)$$



$\omega = \{c^f, c^b\}$ ∈ [0;255]x[0;255]: quad-tree clustering

Global minima of the discrete Chan-Vese functional:

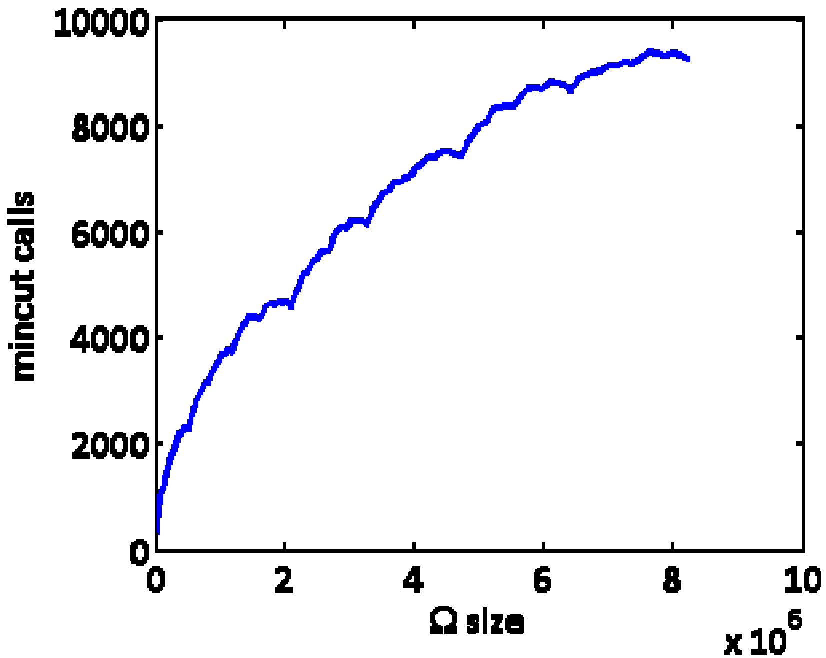


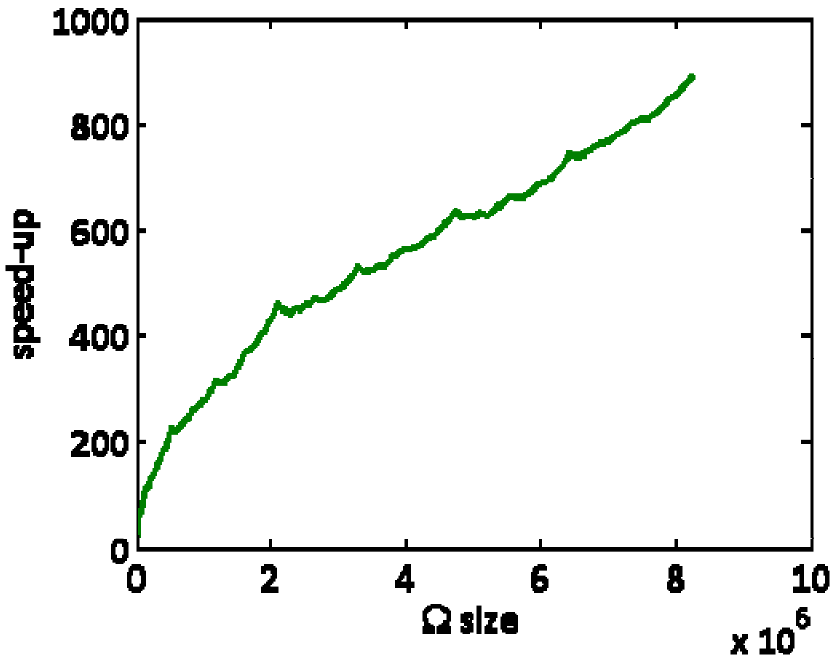
183x162, time=3s, 300x250, time=4s, 385x264, time=16s, 371x255, time=21s,

Speed-up 28-58 times

Performance

Sample Chan-Vese problem:





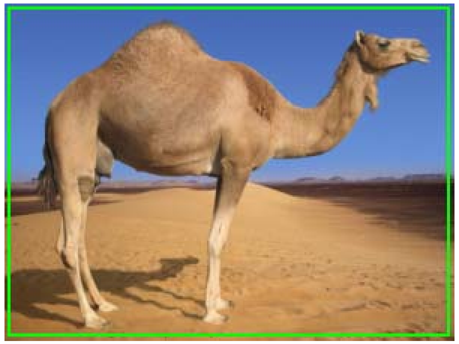
Results: GrabCut

- ω corresponds to color mixtures
- [Rother, Kolmogorov, Blake' *GrabCut* 04] uses EM-like search
- Branch-and-Mincut searches over 65,536 starting points

E = -618

E = -624 (*speed-up 481*)

E = -628



E = -593

E = -584 (*speed-up 141*)

E = -607

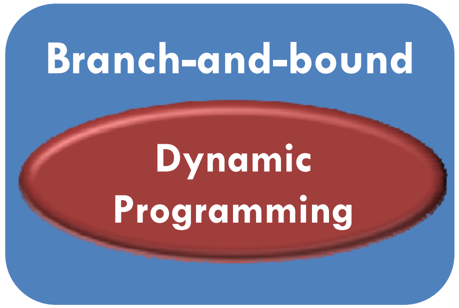
Conclusion

- $$E(\mathbf{x},\omega) = C(\omega) + \sum_{p\in\mathcal{V}} F^p(\omega)\cdot x_p + \sum_{p\in\mathcal{V}} B^p(\omega)\cdot(1-x_p) + \sum_{p,q\in\mathcal{E}} P^{pq}(\omega)\cdot|x_p - x_q|$$

  – good energy to integrate low-level and high-level knowledge in segmentation.

- **Branch-and-Mincut** framework can find its global optimum efficiently in many cases

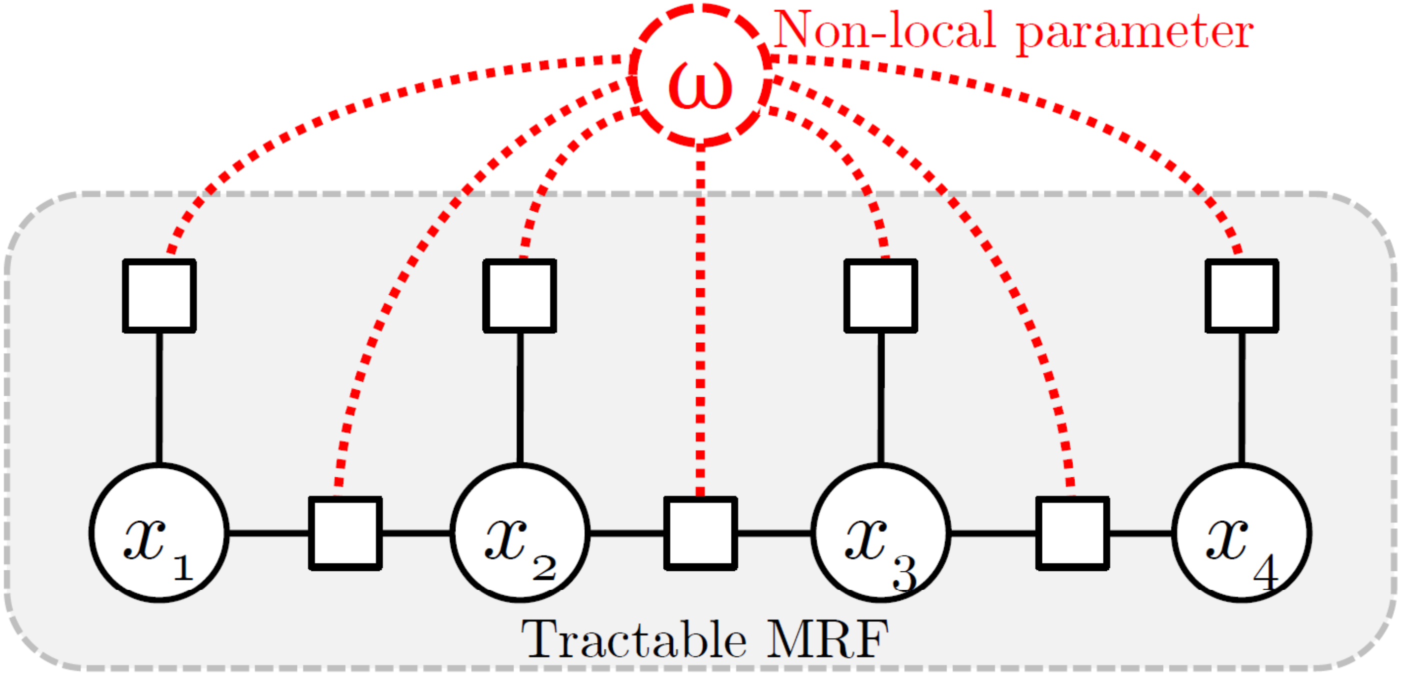
- **Ongoing work:** Branch-and-X algorithms



C++ code at http://research.microsoft.com/~victlem/

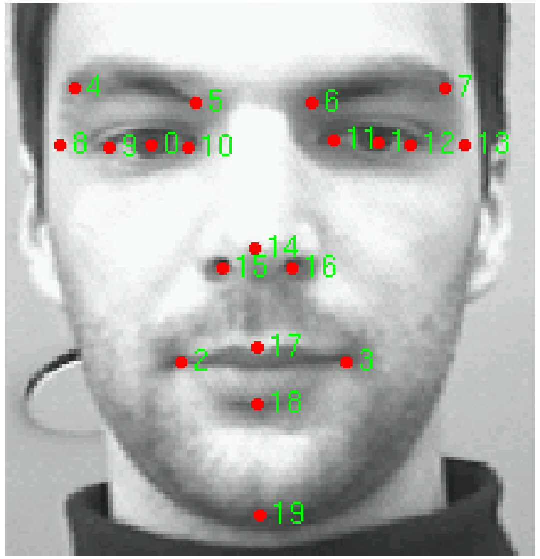
Multi-label Augmented MRFs

$$E(\mathbf{x},\omega) = \sum_{p\in\mathcal{V}} U^p(x_p;\omega) + \sum_{p,q\in\mathcal{E}} P^{pq}(x_p,x_q;\omega)$$



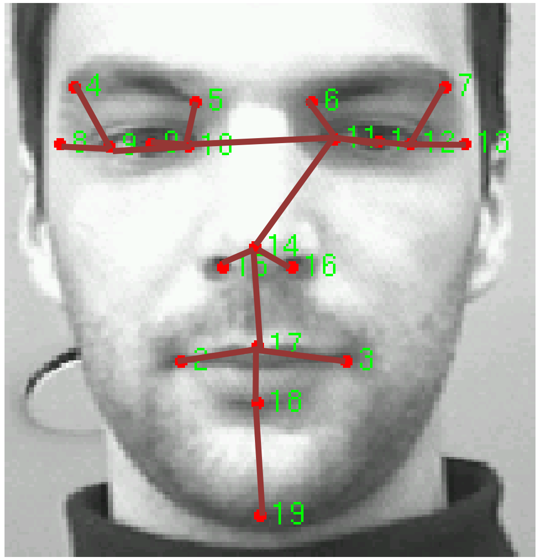
An experiment

BioID dataset: 1520 faces,
800 for training, 720 for testing



22 points
FGNet annotations by
David Cristinacce and
Kola Babalola

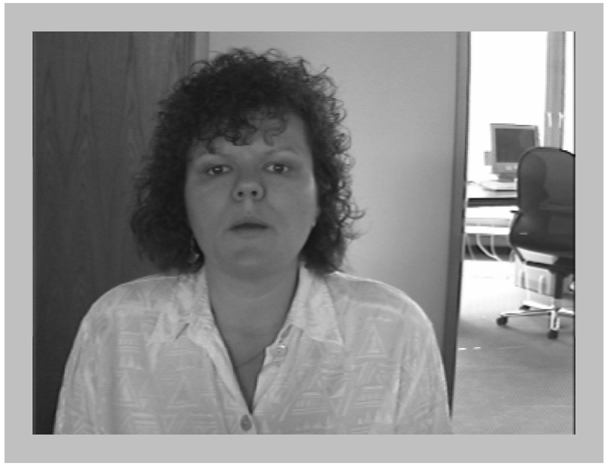
Pictorial structure MRF

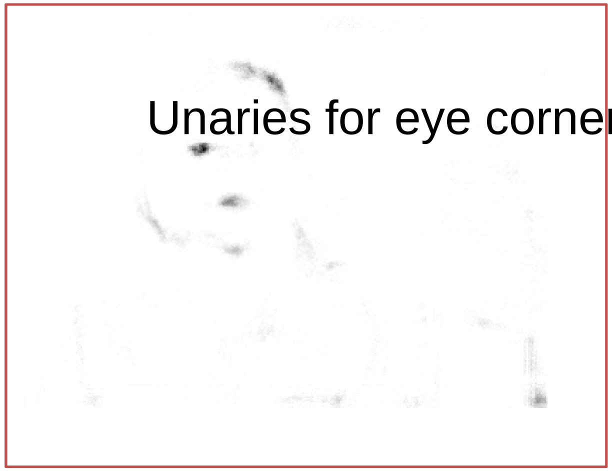


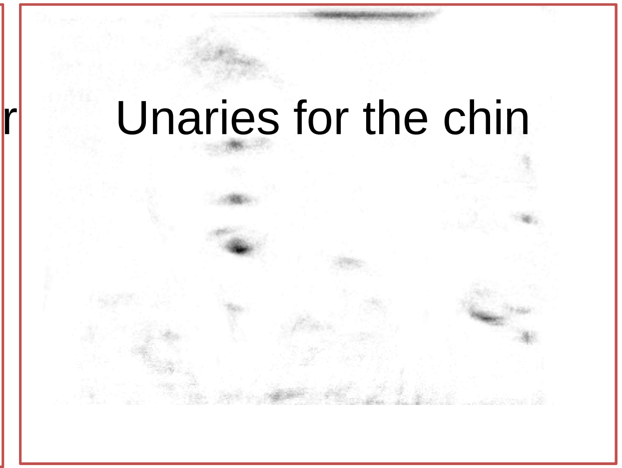
[Felzenszwalb Huttenlocker 05]

$$E(\mathbf{x}) = \sum_{p \in \mathcal{V}} U^p(x_p) + \sum_{p,q \in \mathcal{E}} ||x_p - x_q - l_{pq}||$$

- Tree-structured MRF
- 22 nodes
- Label space – all image locations





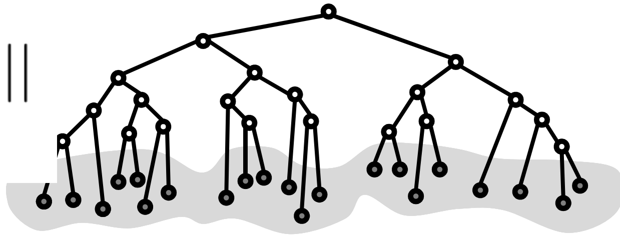


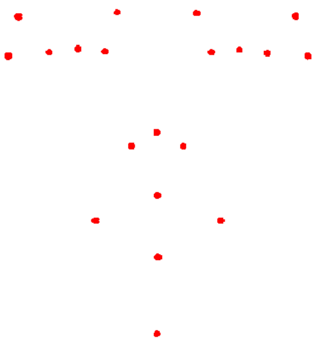
Branch-and-DP

$$E(\mathbf{x},\omega) = \sum_{p\in\mathcal{V}} U^p(x_p) + \sum_{p,q\in\mathcal{E}} ||x_p - x_q - l_{pq}(\omega)||$$

[Felzenszwalb Huttenlocker 05]

“Branch-and-DP” Ω_0

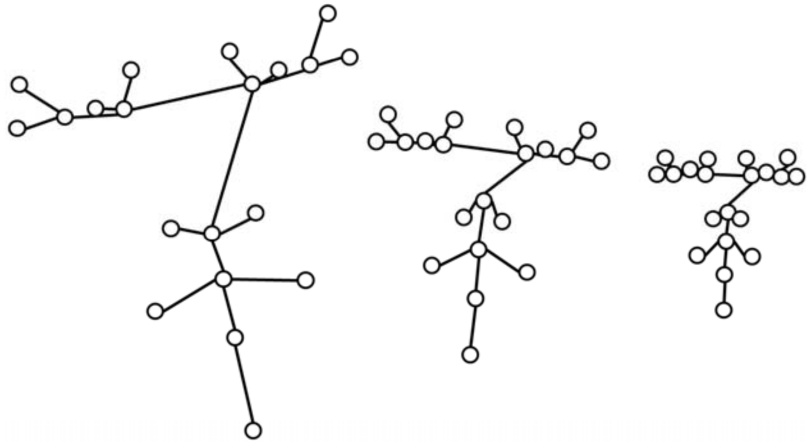




1 configuration

Messages are cheap - O(n)

(distance transforms)



10,000 configurations

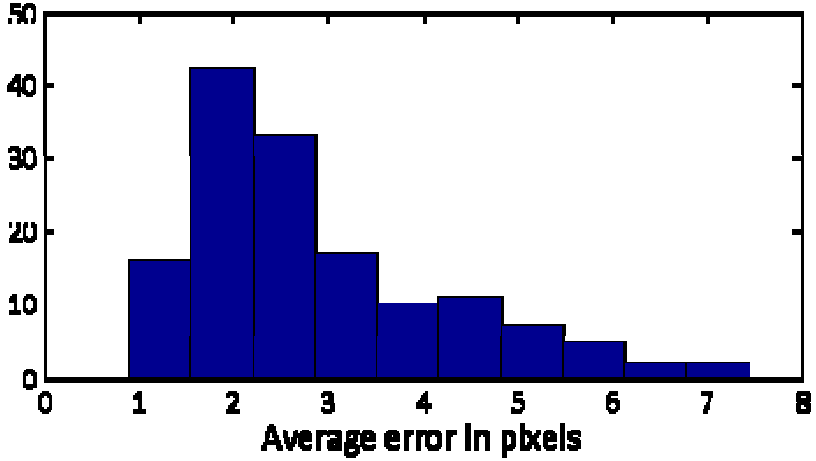
Scales/rotations/deformations

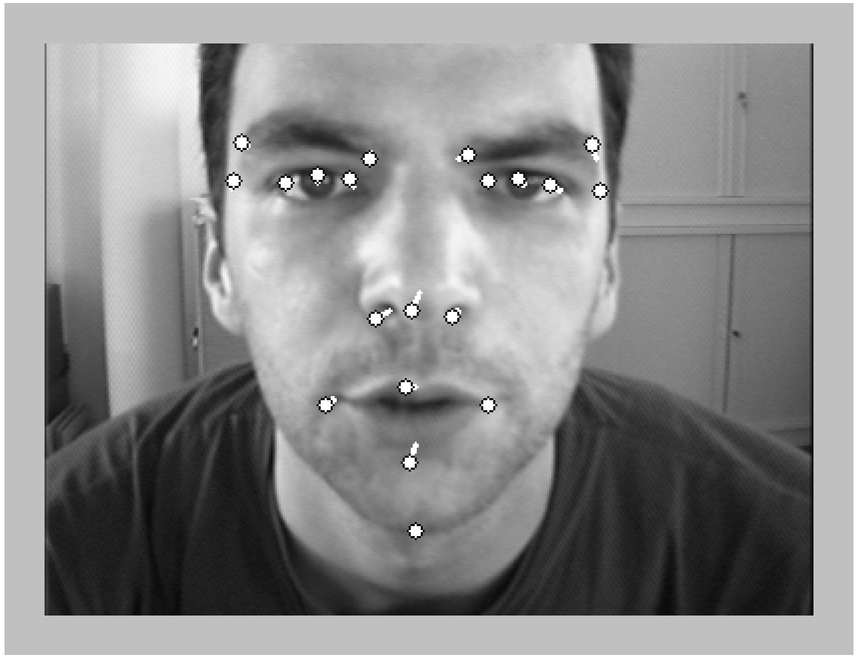
Messages are still cheap - O(n)

(distance transforms + van Herk-Gil-Werman algorithm

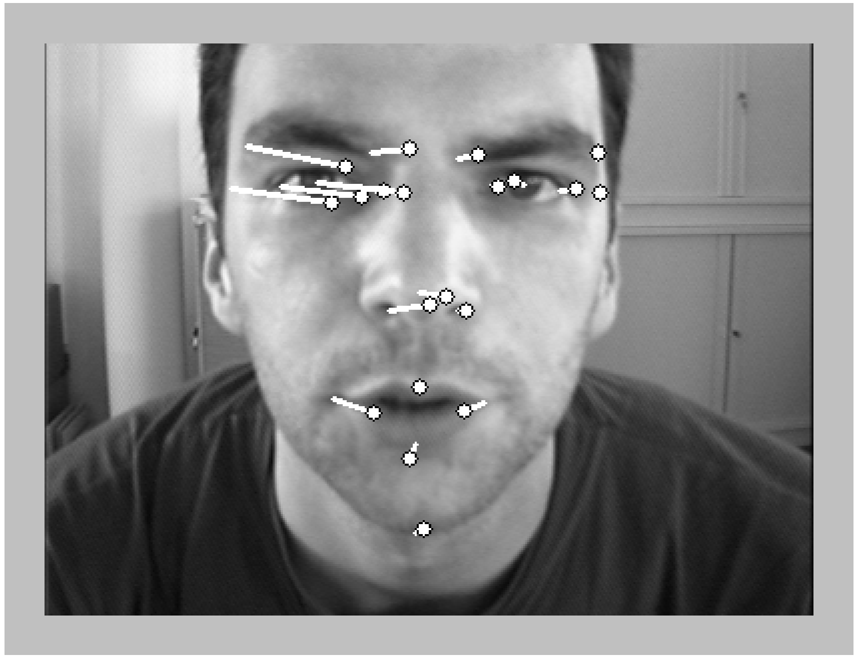
Results

Space size 10,000,
average speed-up 11.5*
(3 minutes for fitting)



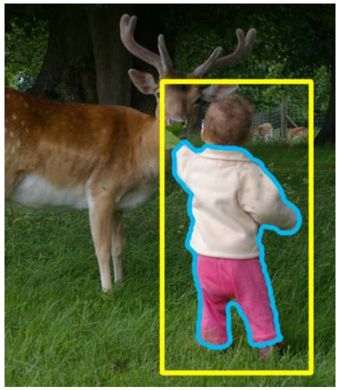


10,000 templates (mean error 2.8)

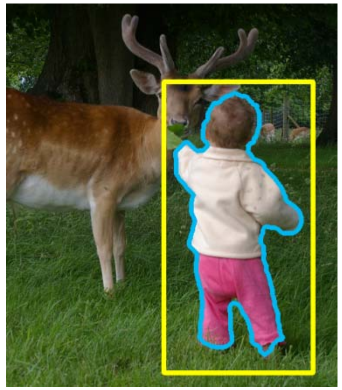


1 template (mean error 4.2)

Image Segmentation with A Bounding Box Prior



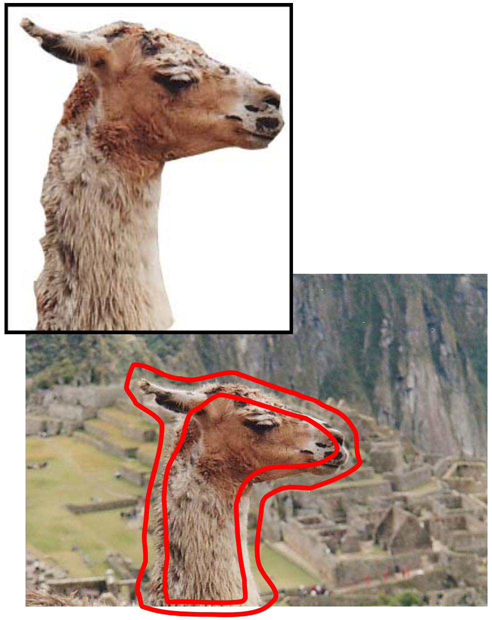
ICCV 2009



Microsoft Research Cambridge

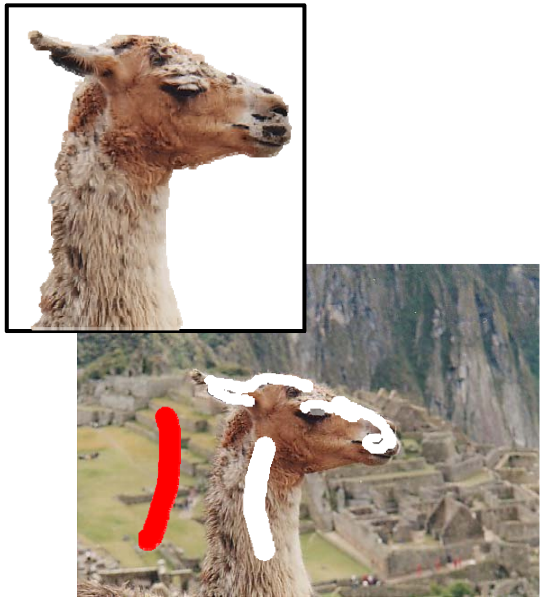
Victor Lempitsky
Pushmeet Kohli
Carsten Rother
Toby Sharp

Motivation



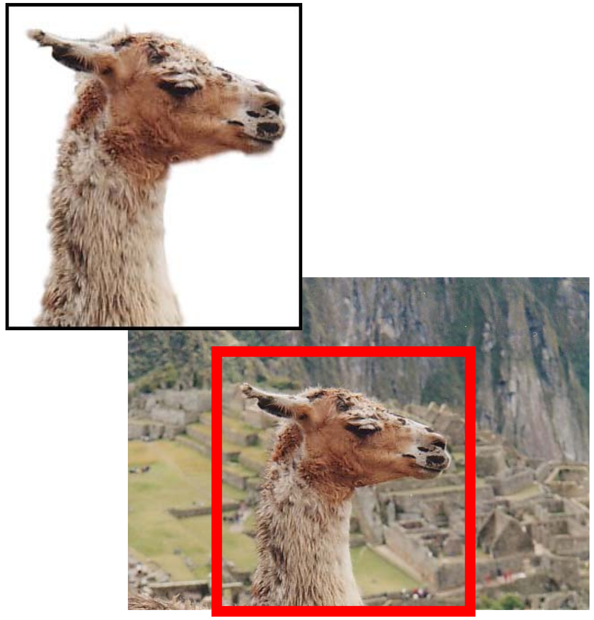
Magnetic Lasso
Mortensen & Barret '95

* Globally optimal
* user intensive



Interactive graph cut
Boykov & Jolly '01

* Globally optimal
* user friendly



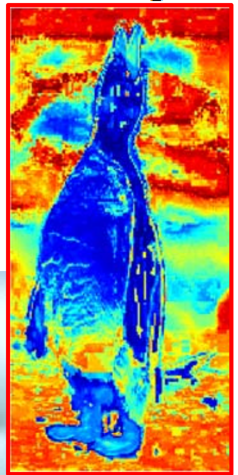
GrabCut
Rother, Kolmogorov & Blake '04

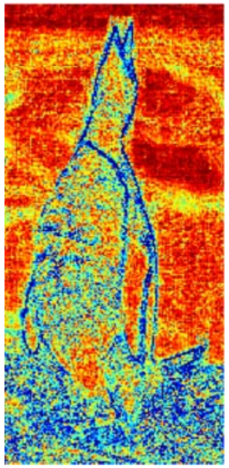
* NP hard (global color model)
* very user friendly

Graph cut systems

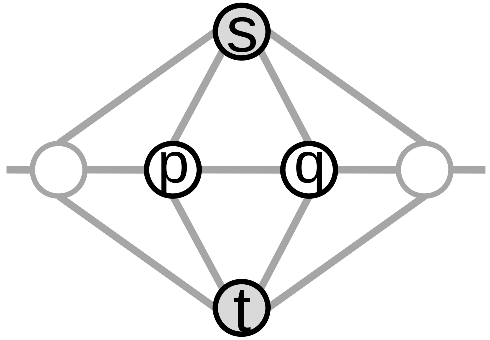
Graph cut segmentation [Boykov&Jolly 01] integrates cues and input via:

$$E(\mathbf{x}) = \sum_{p \in \mathcal{B}} U^p \cdot x_p + \sum_{\{p,q\} \in \mathcal{E}} V^{pq} \cdot |x_p - x_q|, \quad x_p \in \{0, 1\}$$

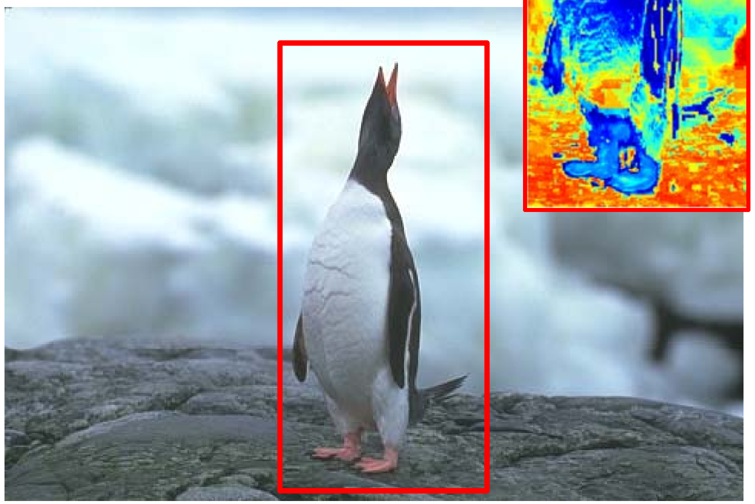




...

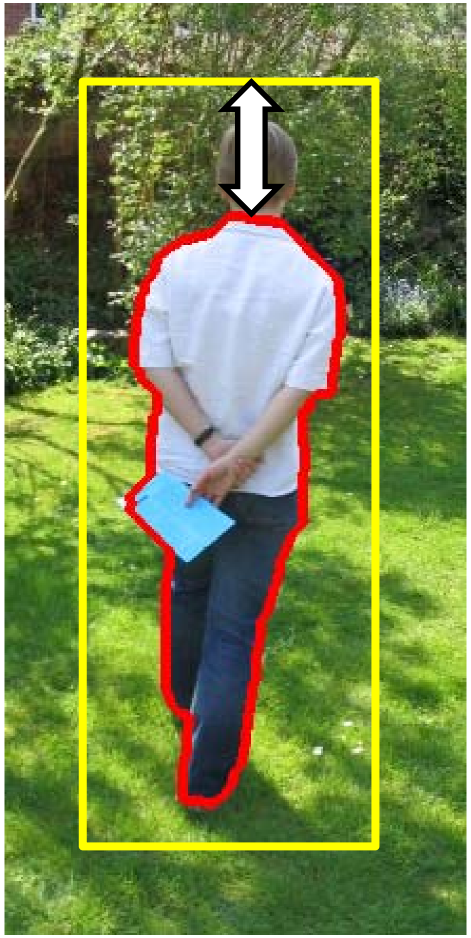


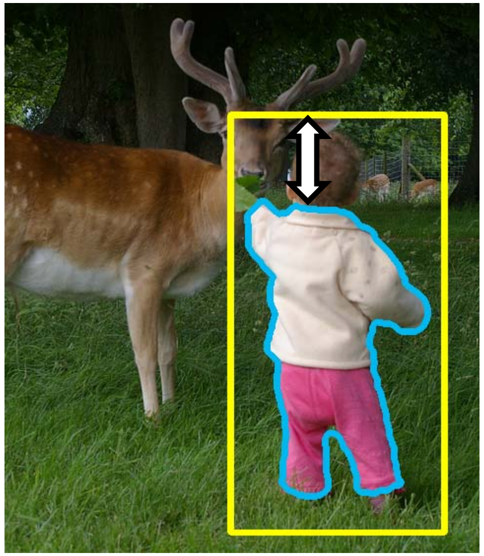
...

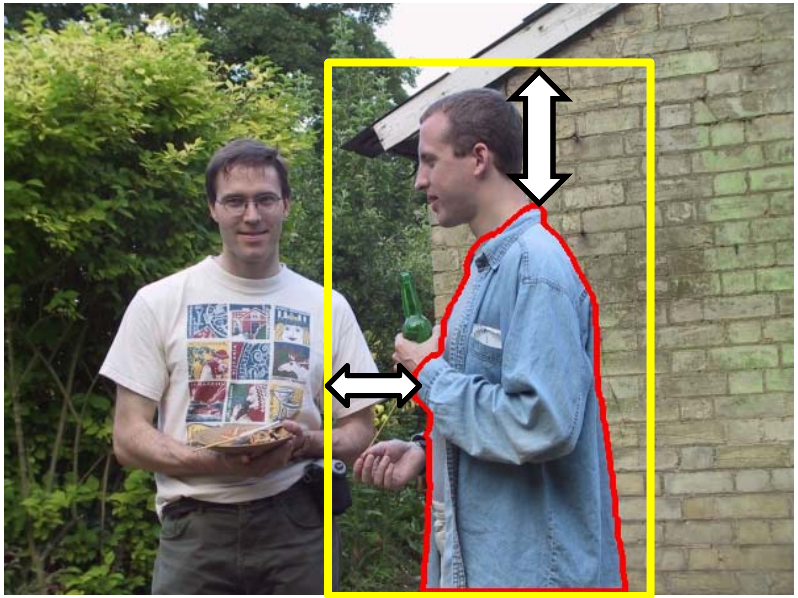


GMMRF&GrabCut [Blake et al. 04, Rother et al.05] adds the Gaussian mixture fitting idea + reiteration

When it is not right straight away…







Solutions:

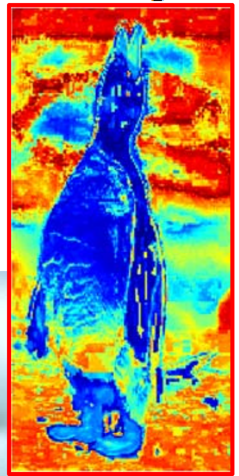
1. More interaction
2. High-level semantic knowledge
3. <u>Or just look at the user input more attentively</u>

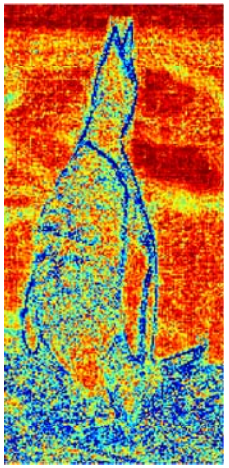
"Tightness" constraint would help!

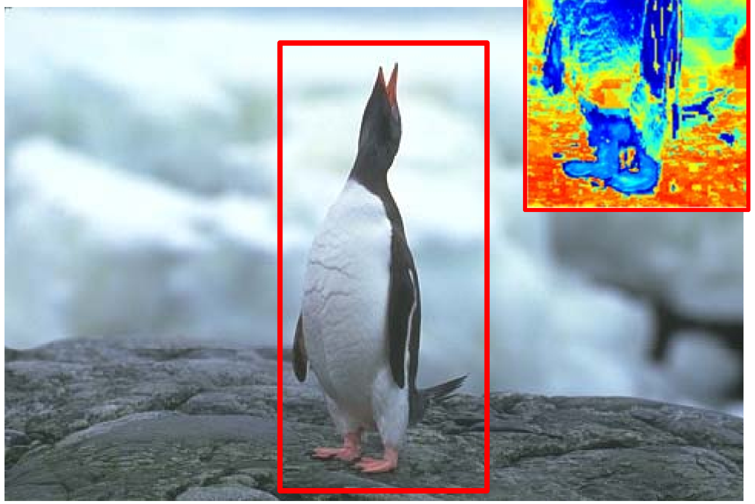
Problem Formulation

Graph cut segmentation [Boykov&Jolly 01] integrates cues and input via:

$$E(\mathbf{x}) = \sum_{p\in\mathcal{B}} U^p \cdot x_p + \sum_{\{p,q\}\in\mathcal{E}} V^{pq} \cdot |x_p - x_q|, \quad x_p \in \{0,1\}$$

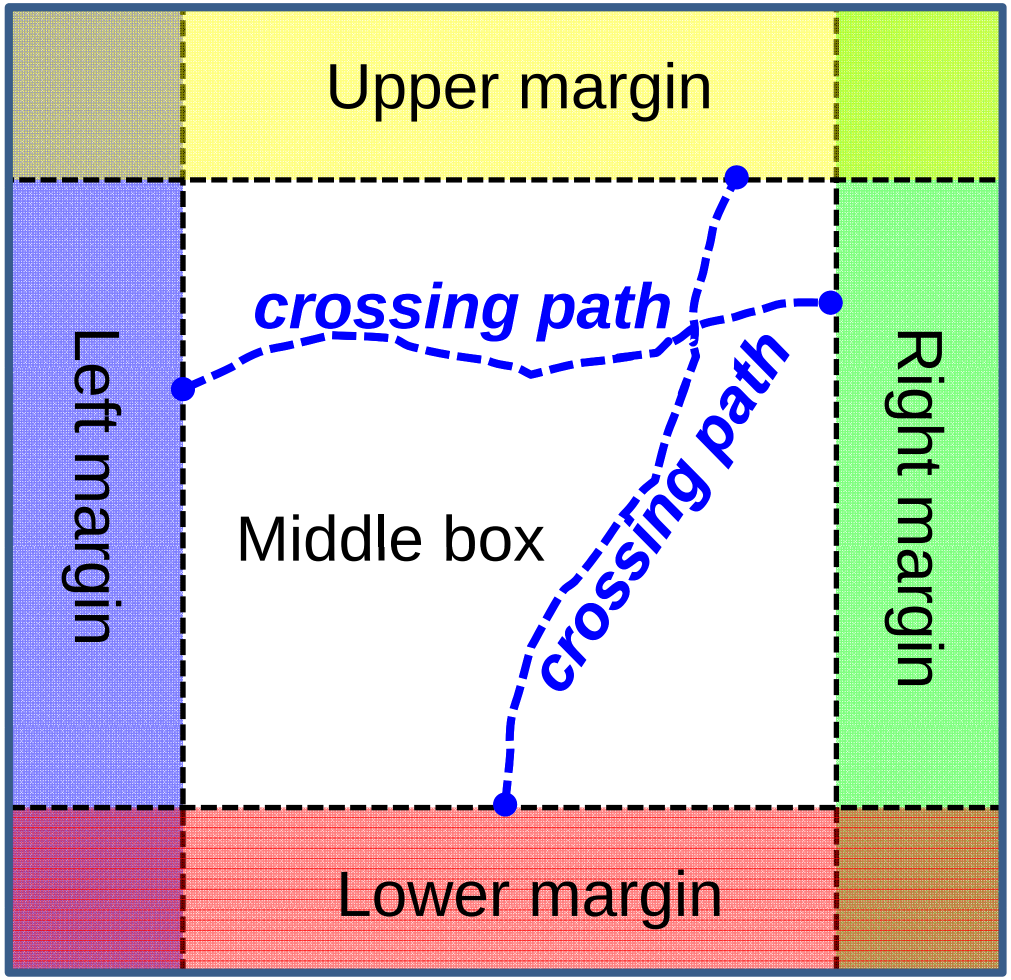




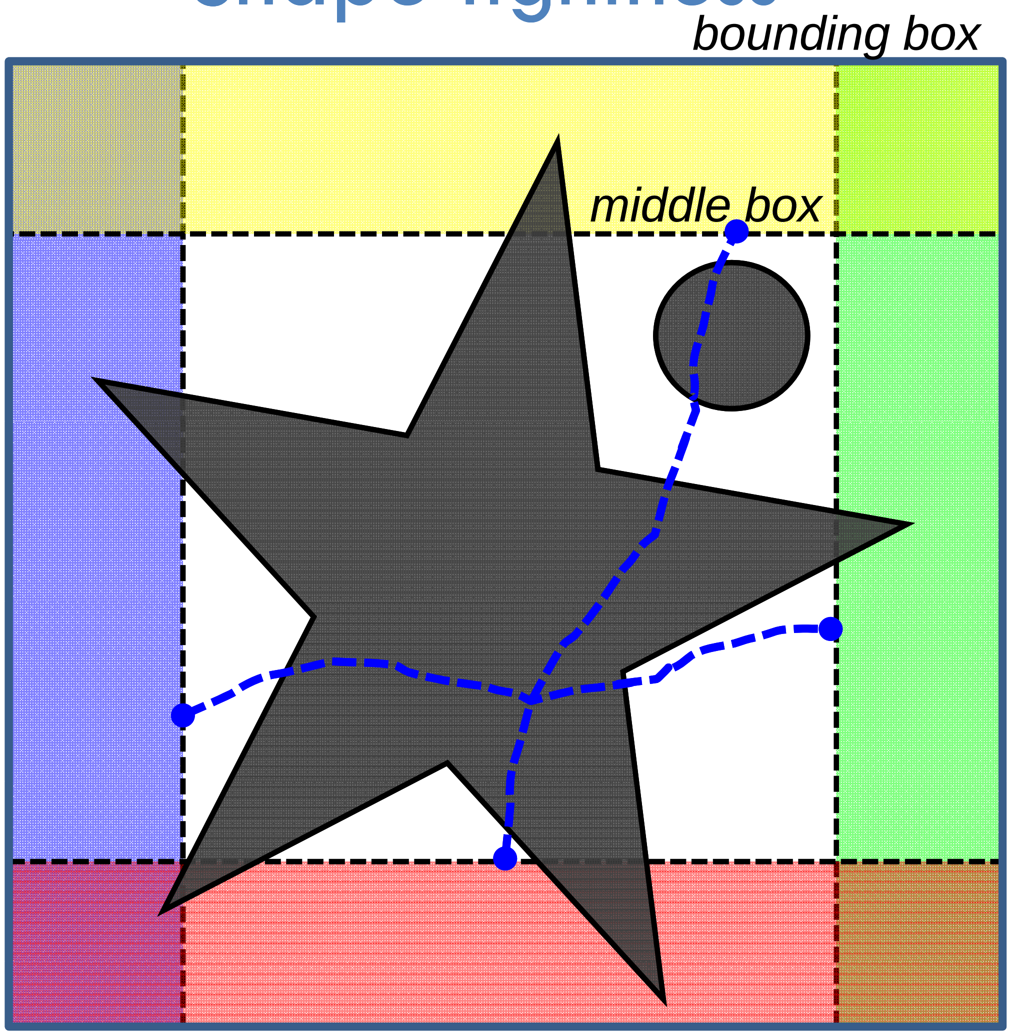


Minimize subject to
"shape is sufficiently tight" ?

Crossing paths



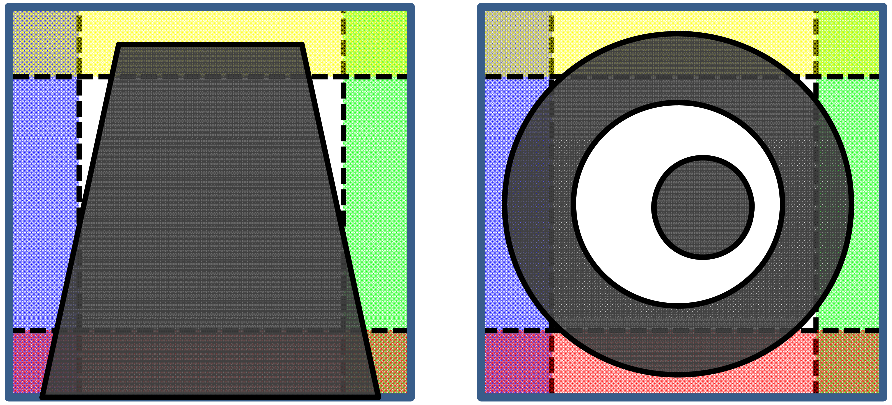
Shape tightness



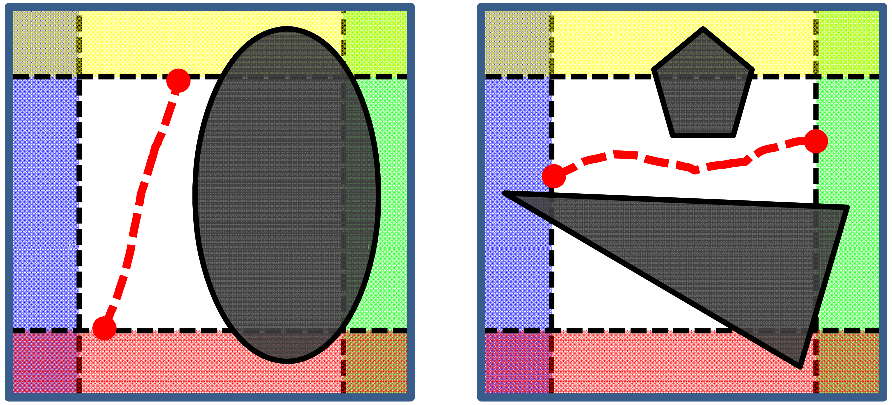
Definition: the shape is tight if it intersects all crossing paths

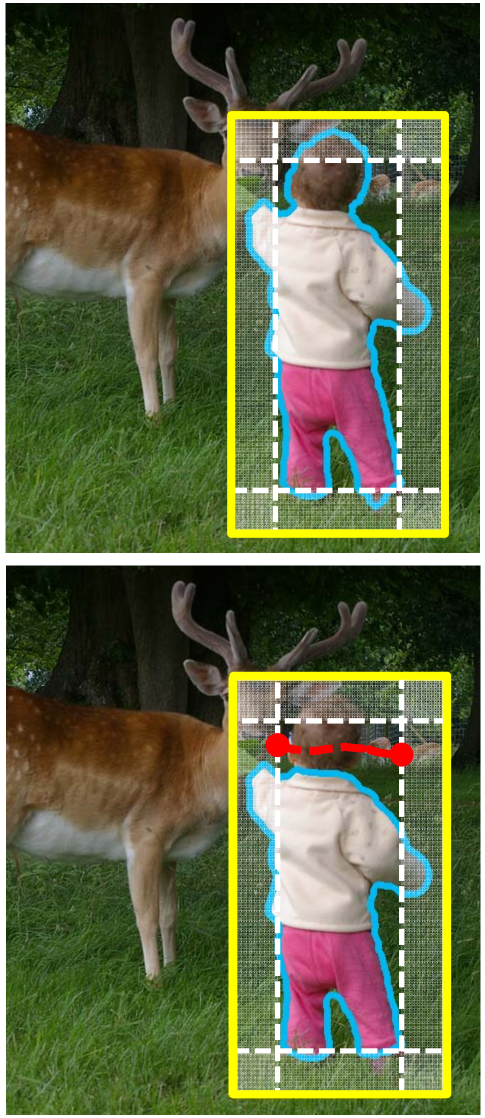
Shape tightness

Tight:



Not tight:





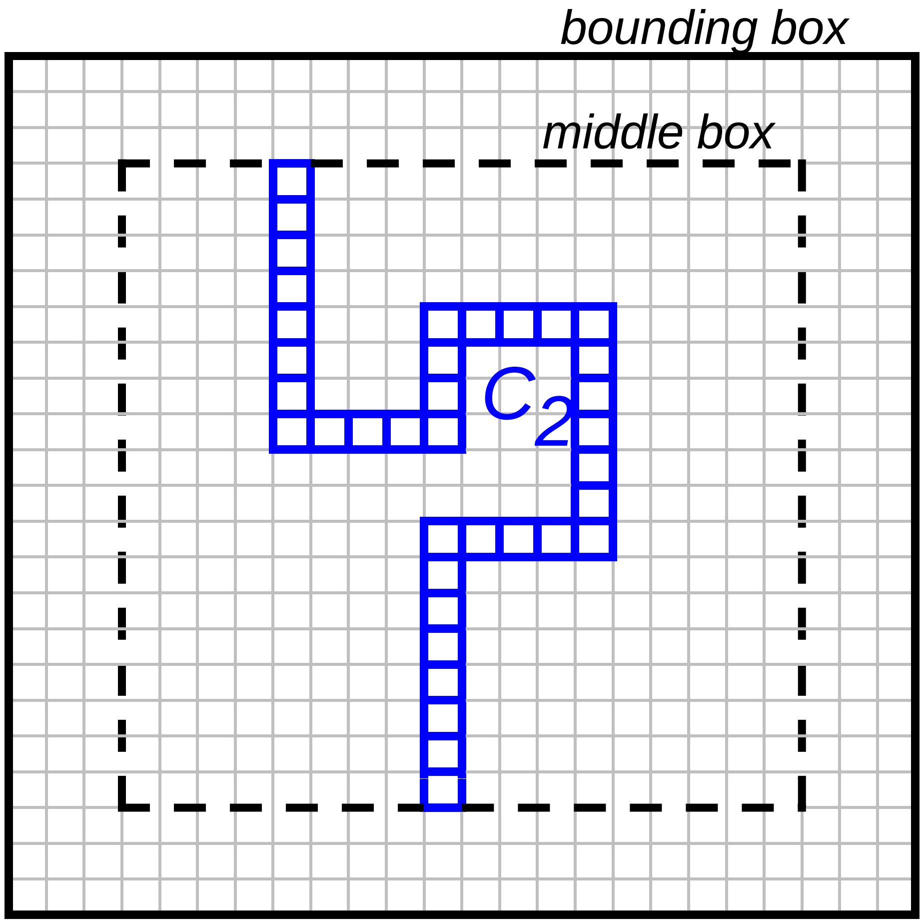
Corollary: the shape is tight *if and only if* the shape has a connected component touching all 4 sides of the middle box

Incorporating Shape Tightness

$$E(\mathbf{x}) = \sum_{p \in \mathcal{B}} U^p \cdot x_p + \sum_{\{p,q\} \in \mathcal{E}} V^{pq} \cdot |x_p - x_q|, \quad x_p \in \{0,1\}$$

$$\forall C \in \Gamma \quad \sum_{p \in C} x_p \geq 1$$

all 4-connected crossing paths



Incorporating Shape Tightness

$$E(\mathbf{x}) = \sum_{p \in \mathcal{B}} U^p \cdot x_p + \sum_{\{p,q\} \in \mathcal{E}} V^{pq} \cdot |x_p - x_q|, \quad x_p \in [0,1]$$

$$\forall C \in \Gamma \quad \sum_{p \in C} x_p \geq 1$$

Trivial to convert to an LP now!

Problems:

1. It's integer (hence **non-convex**) Relax!
2. It has combinatorial number of constraints

Related work:

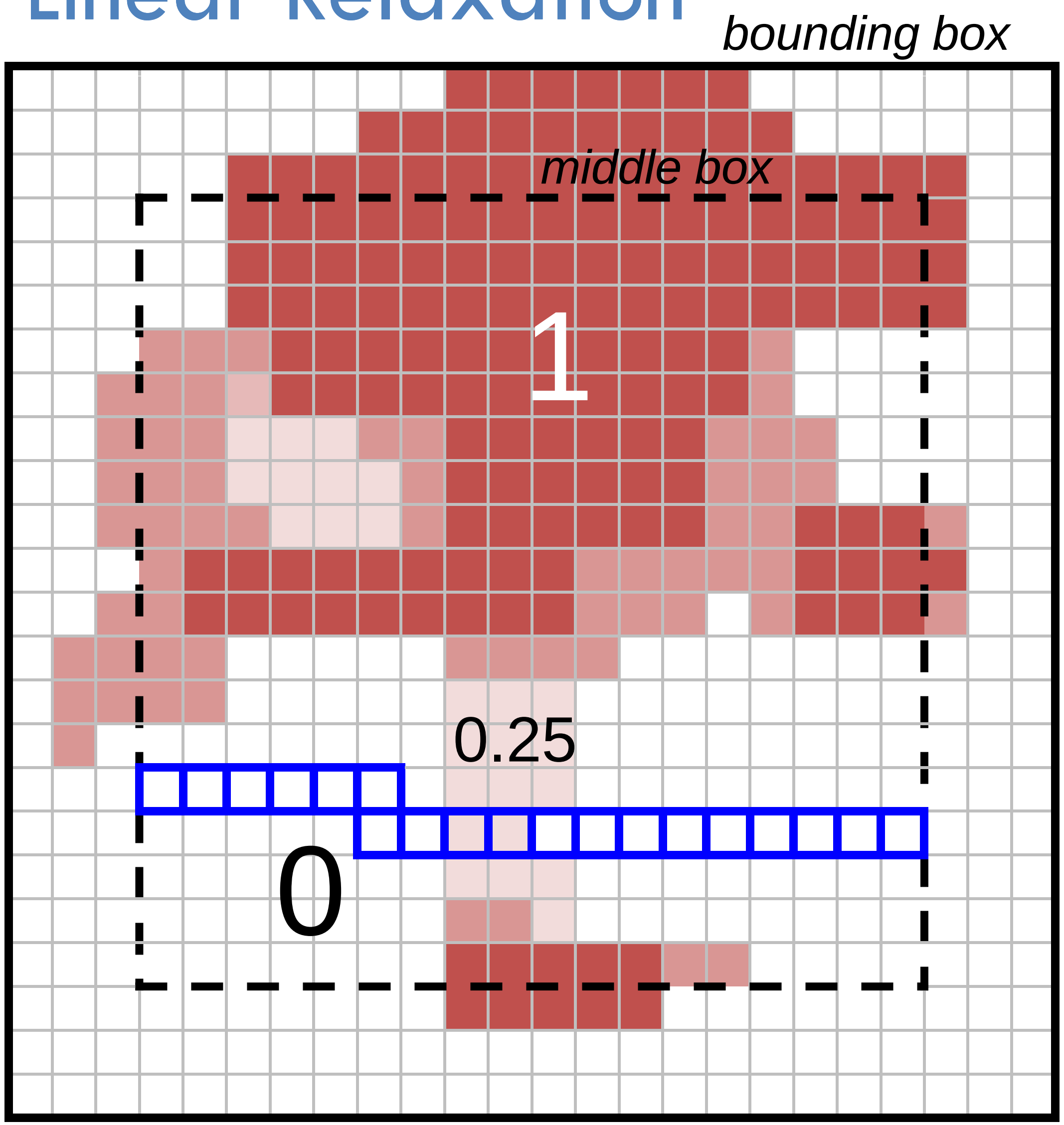
K. Kolev, D. Cremers: ***Integration of Multiview Stereo and Silhouettes via Convex Functionals on Convex Domains.*** ECCV 2008

Solving the Linear Relaxation

$$\forall C \in \Gamma \quad \sum_{p \in C} x_p \geq 1$$

- Cannot enforce all constraints
- Dijkstra can check if all constraints are satisfied
- Dijkstra can find the most violated constraint
- Can switch the constraints on gradually

See also:
S. Nowozin and C. H. Lampert: **Global Connectivity Potentials for Random Field Models.** CVPR 2009



Solving the Linear Relaxation

Start with no constraints

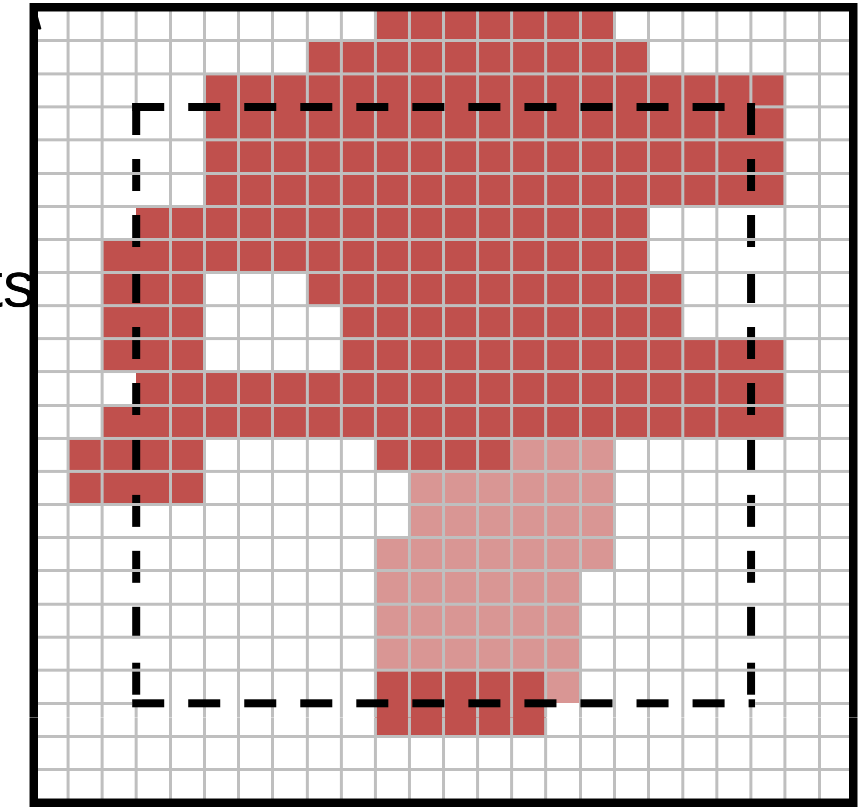
Iterate:

1. Pick a set of violated paths

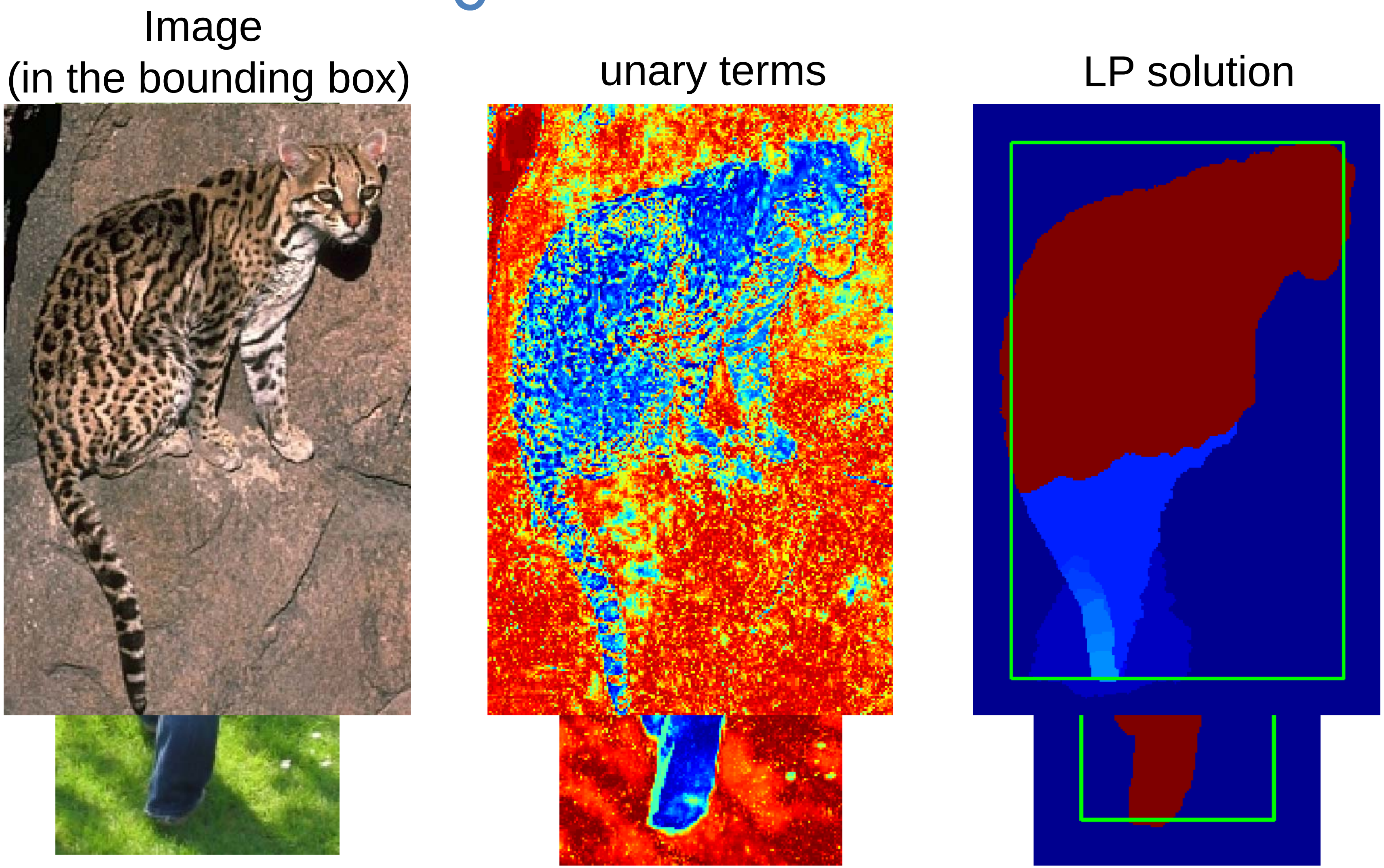
2. Activate respective constraints

3. Rerun linear optimization

Until all constraints satisfied



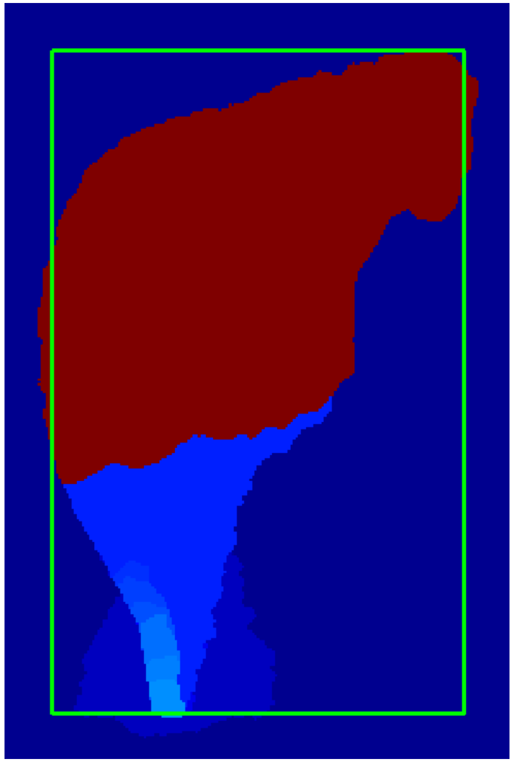
Solving the Linear Relaxation



How to Round?

$$E(\mathbf{x}) = \sum_{p \in \mathcal{B}} U^p \cdot x_p + \sum_{\{p,q\} \in \mathcal{E}} V^{pq} \cdot |x_p - x_q|, \quad x_p \in [0,1]$$

$$\forall C \in \Gamma \quad \sum_{p \in C} x_p \geq 1$$



Previous works e.g. [Kolev&Cremers'08]:
just **threshold** at low enough value

Our work: **use the problem structure** to perform provably better rounding

Pinpointing Algorithm

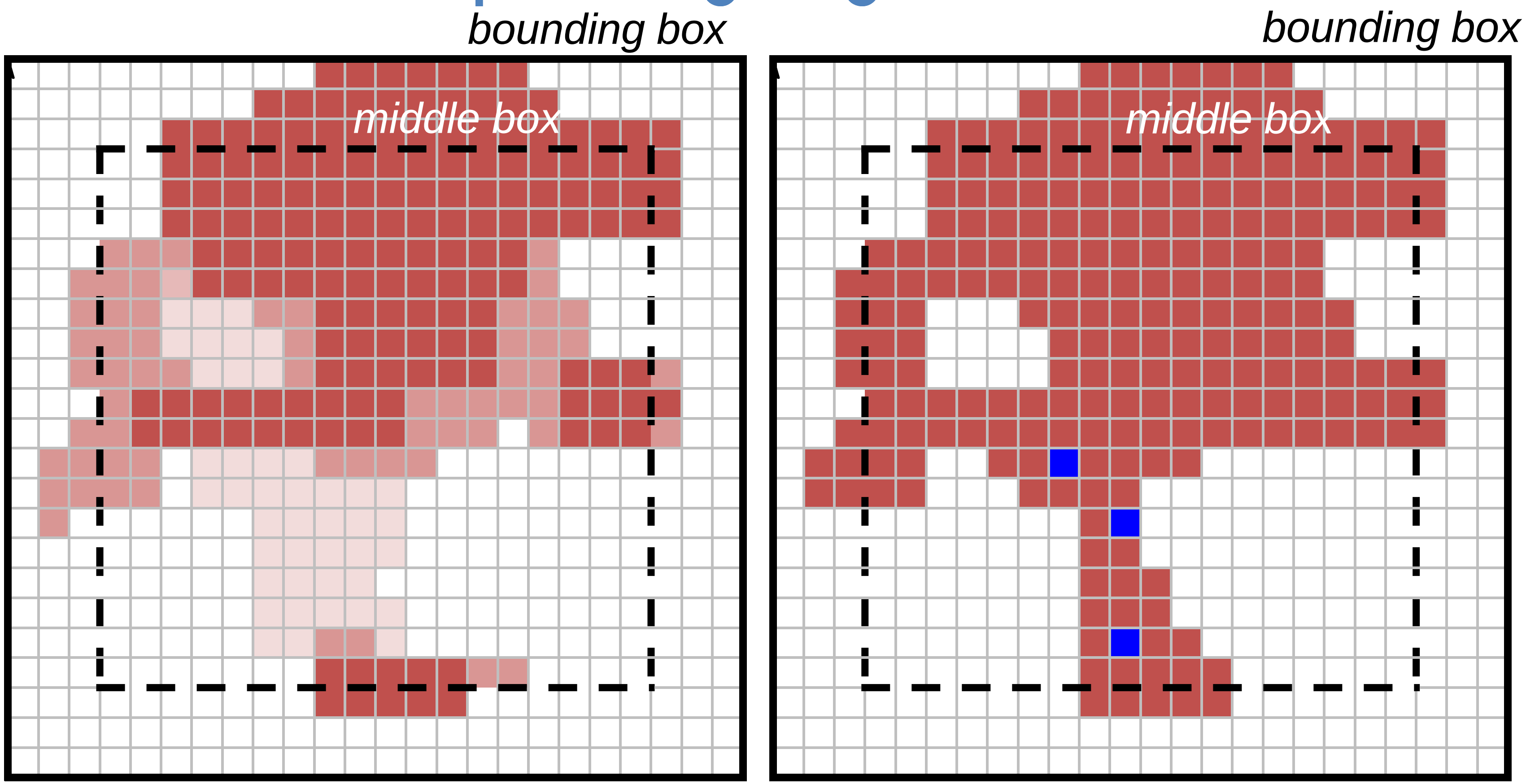
$$E(\mathbf{x}) = \sum_{p \in \mathcal{B}} U^p \cdot x_p + \sum_{\{p,q\} \in \mathcal{E}} V^{pq} \cdot |x_p - x_q|, \quad x_p \in \{0,1\}$$

$$\forall C \in \Gamma \quad \sum_{p \in C} x_p \geq 1 \qquad \forall p \in \Pi \quad x_p = 1$$

"pinpointing" set

Pinpointing algorithm idea:
use the fractional solution to the initial problem
to guide the construction of the pinpointing set

Pinpointing Algorithm

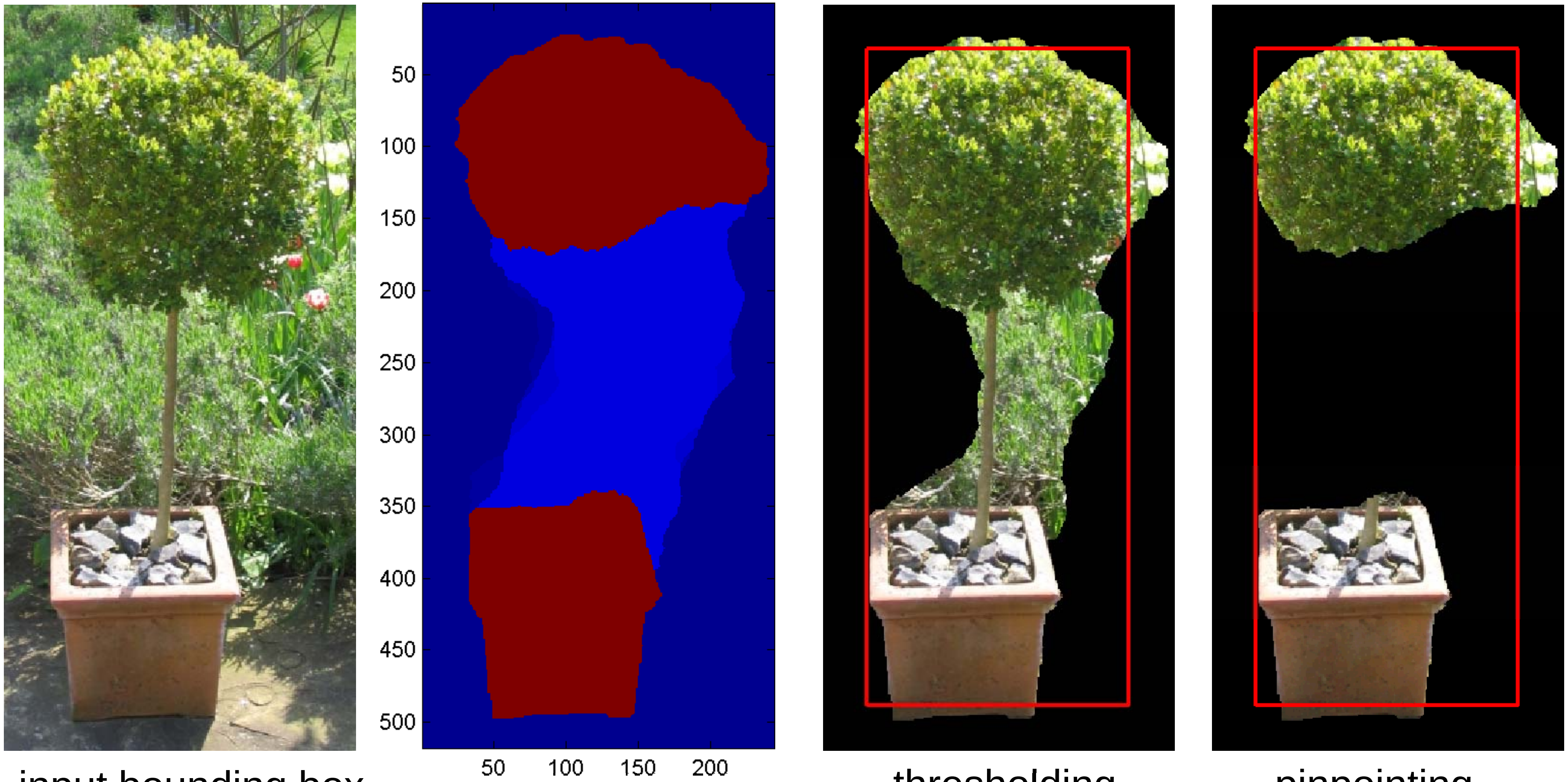


LP solution

Rounded solution

- Use “dynamic graph cut” [Boykov&Jolly’01],[Kohli&Torr’04]

Pinpointing Algorithm



input bounding box

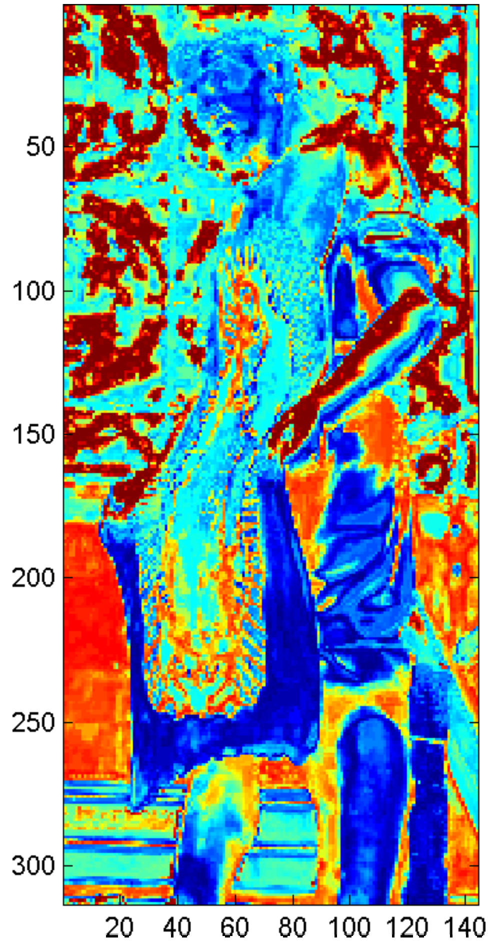
LP solution

thresholding,
E = -0.46

pinpointing,
E = -0.54

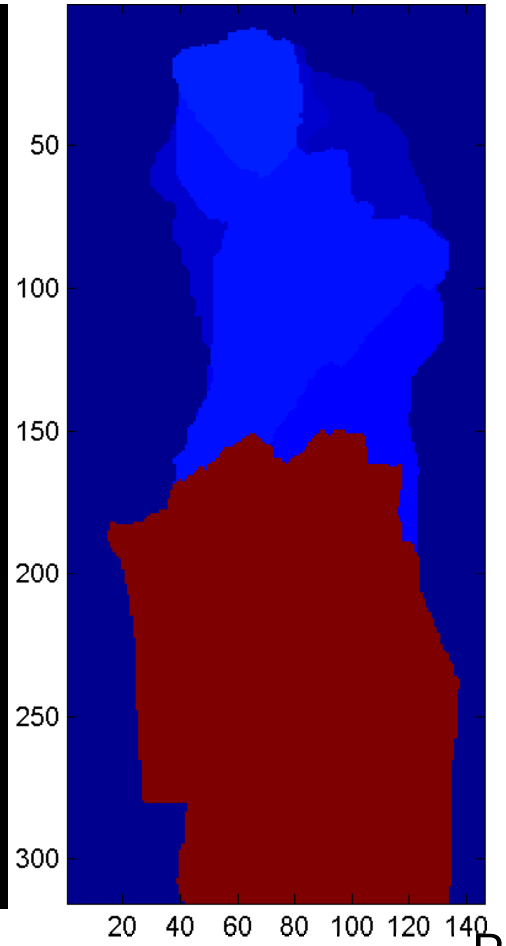
Corollary: pinpointing always gets a lower (or same) energy solution compared to thres

Pinpointing Algorithm



Unary terms

Pinpointing unaries,
E= -16.5



LP solution

Pinpointing LP solution,
E= -17.0

Pinpointing can be used as a fast, standalone heuristics.

Quantitative results

GMMRF [Blake et al. 04]:
1. Fit Gaussian mixtures to get unary terms
2. Optimize the graph cut energy + a bounding box prior

fast

fast

slow

slow

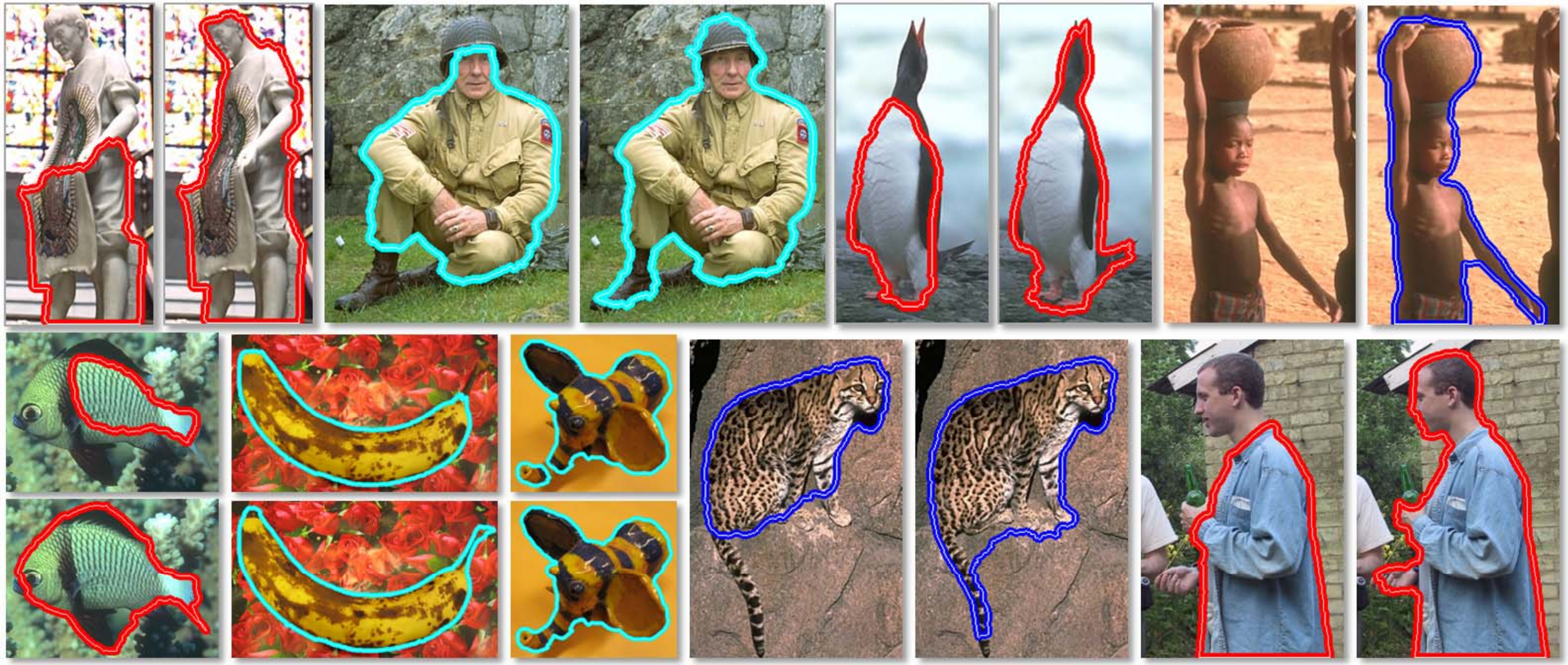
fast

Relative ordering in terms of the obtained energy is the same

Refining color models

GrabCut [Rother et al. 04]: **iterate**

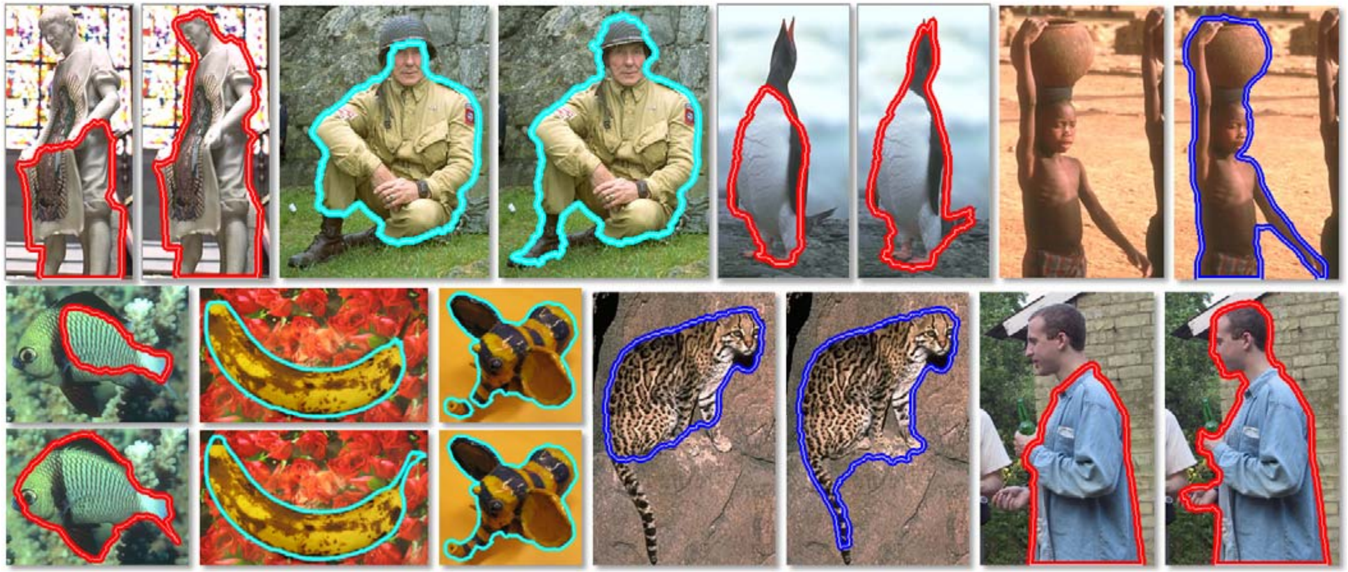
1. Fitting gaussian mixtures
2. Optimizing the graph cut energy + a bounding box prior



The error rate goes down from 5.1% to 3.7%

Conclusion

- Global constraints are powerful
- Approximate optimization is possible



Thank you for your attention!