Probabilistic Graphical Model Workshop:
Sparsity, Structure and High-dimensionality

The Institute of Statistical Mathematics,
Tokyo, Japan
March 23-25, 2016.

MAP Estimation of Markov Random Fields

With Some Applications in Medical Imaging

Hiroshi Ishikawa

Department of Computer Science & Engineering

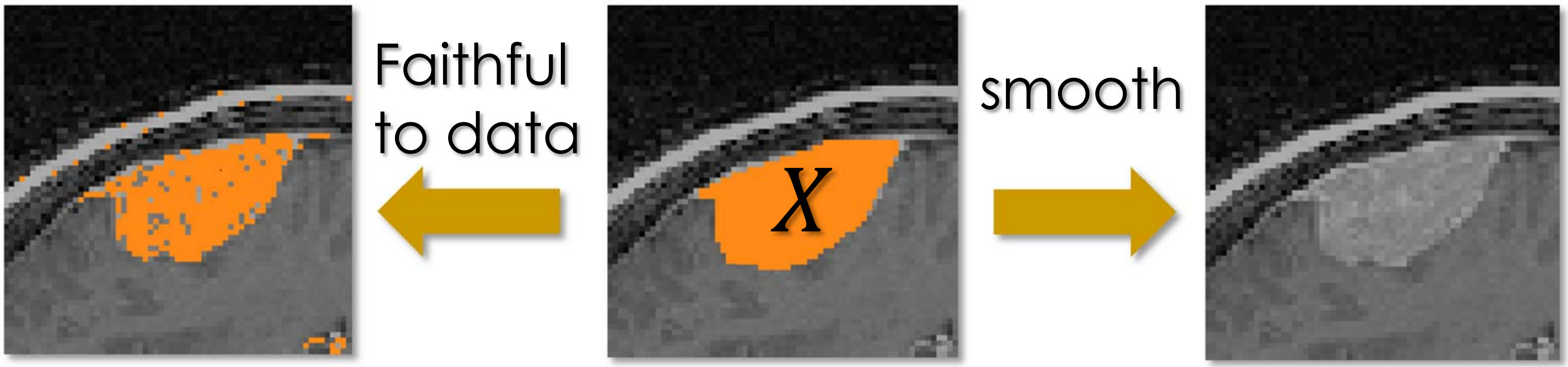
Waseda University

Outline

- Graph cuts
  - Uses minimum-cut algorithm for energy minimization
- Higher-order energy minimization
  - Binary labels: reduce to first order
    - Add variable
    - Some with ELC
  - Then use graph cuts
- Segmentation with shape prior
  - Shape restriction enables better segmentation
    - Realized by higher-order energy potential
    - Pulmonary Artery-Vein Segmentation
    - Coronary Lumen and Plaque Segmentation

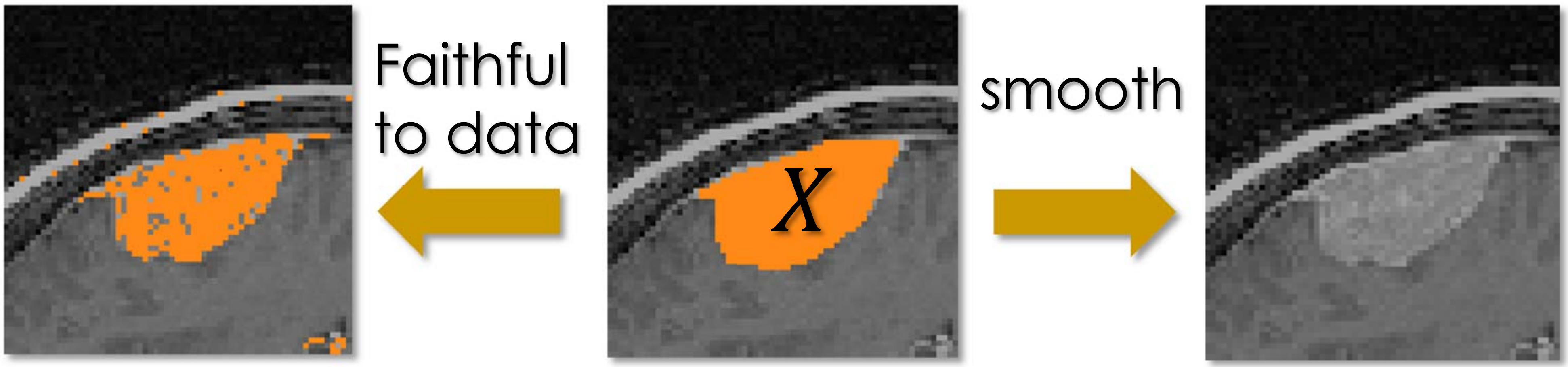
Segmentation Trade-off

- X : assigns $X_v = 0$ or 1 to each point v
- Faithfulness to Data at each point
  - More likely to be foreground or background?
- Prior model / regularization
  - Assume smoothness



Optimization

- X : assigns $X_v = 0$ or 1 to each point v
- Define a function $E(X)$ so that
  - Better segmentation X has lower $E(X)$
- Sum of conflicting functions
  - Data
  - Smoothness

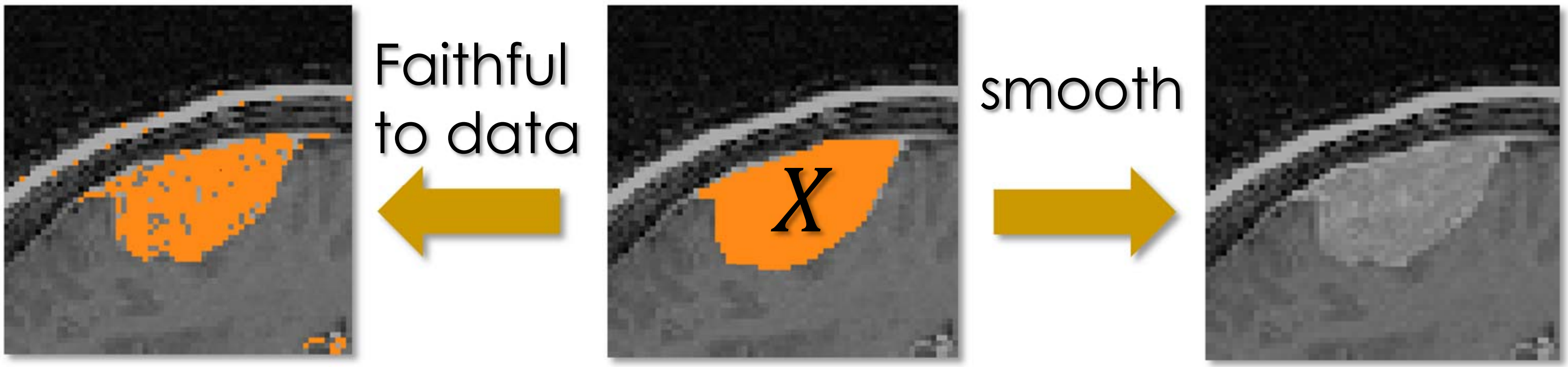


Example

- Consider

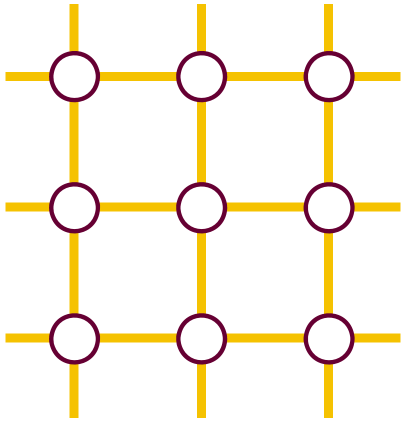
$$E(X) = \sum_{v \in V} g_v(X_v) + \sum_{(u,v) \in E} h_{uv}(X_u, X_v)$$

- $g_v(X_v)$ low when X_v matches the color
- $h_{uv}(X_u, X_v)$ low when smooth



Markov Random Field

- Graph $G = (V, E)$
  - V : each $v \in V$ has an L-valued random variable X_v
  - E : represents dependence
- $\mathscr{C}$: set of cliques in V
- Probability distribution of MRF



$$P(X) = \frac{1}{Z} \prod_{C \in \mathscr{C}} q_C(X_C) = \frac{p(X)}{Z}$$

$$Z = \sum_{X \in \mathscr{X}} \prod_{C \in \mathscr{C}} q_C(X_C) = \sum_{X \in \mathscr{X}} p(X)$$

$$X_C = (X_v)_{v \in C}$$

Markov Random Field

- Rewrite the MRF probability distribution

$$P(X) = \frac{1}{Z} \underbrace{\prod_{C \in \mathscr{C}} q_C(X_C)}_{p(X)} = \frac{p(X)}{Z}$$

as

$$p(X) = e^{-E(X)}$$

$$E(X) = \sum_{C \in \mathscr{C}} f_C(X_C)$$

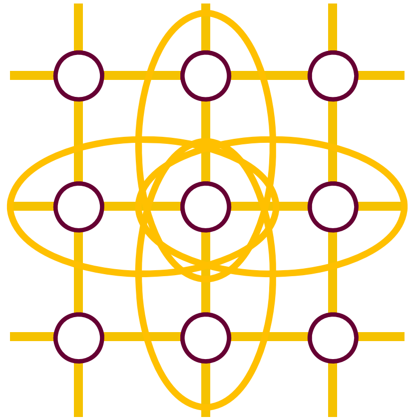
Energy

Potential

First order MRF

- The simplest (interesting) MRF

Geman & Geman 1984; Besag 1986



- Model neighborhood relation

$$\mathscr{C} \approx V \cup E$$

$$p(X) = \prod_{v \in V} \psi_v(X_v) \prod_{(u,v) \in E} \phi_{uv}(X_u, X_v)$$

- Energy

$$E(X) = \sum_{C \in \mathscr{C}} f_C(X_C) = \sum_{v \in V} g_v(X_v) + \sum_{(u,v) \in E} h_{uv}(X_u, X_v)$$

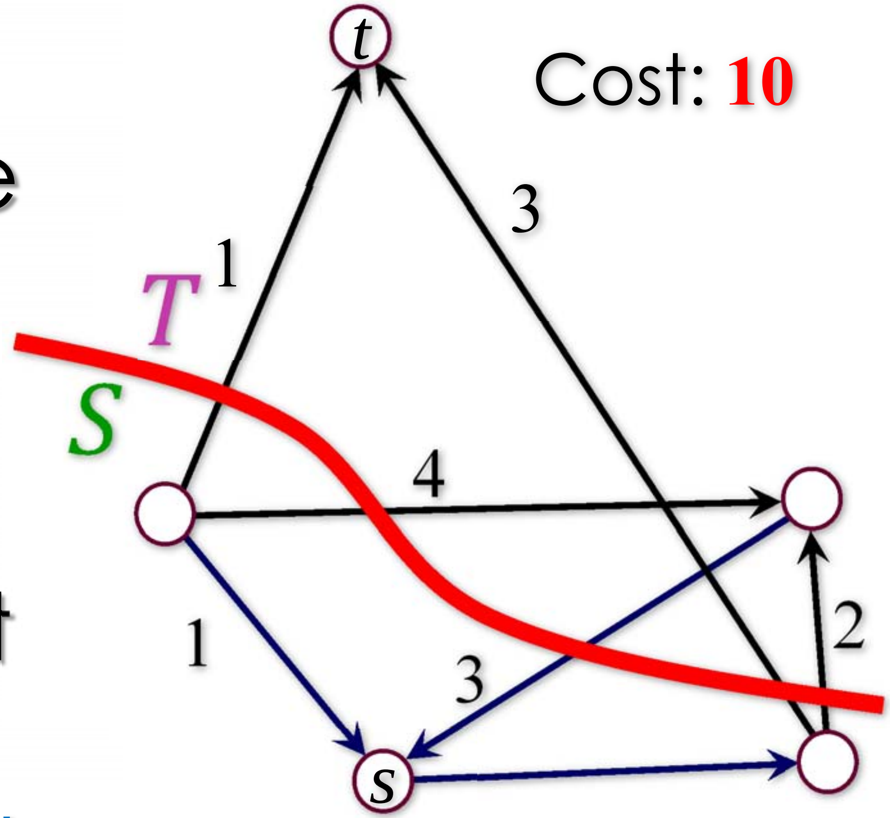
Energy Minimization

$$E(X) = \sum_{v \in V} g_v(X_v) + \sum_{(u,v) \in E} h_{uv}(X_u, X_v)$$

- Problem: Find the X that minimizes $E(X)$
- Possible X
  - Combinations of labeling the sites (pixels)
  - If V is 64×64 and $X: V \rightarrow \{0,1\}$, $\quad 2^{4096} > 10^{1233}$
- NP-Hard in general
- Old method: Monte Carlo
- For *some* form of energy, Graph-cut algorithms can globally minimize

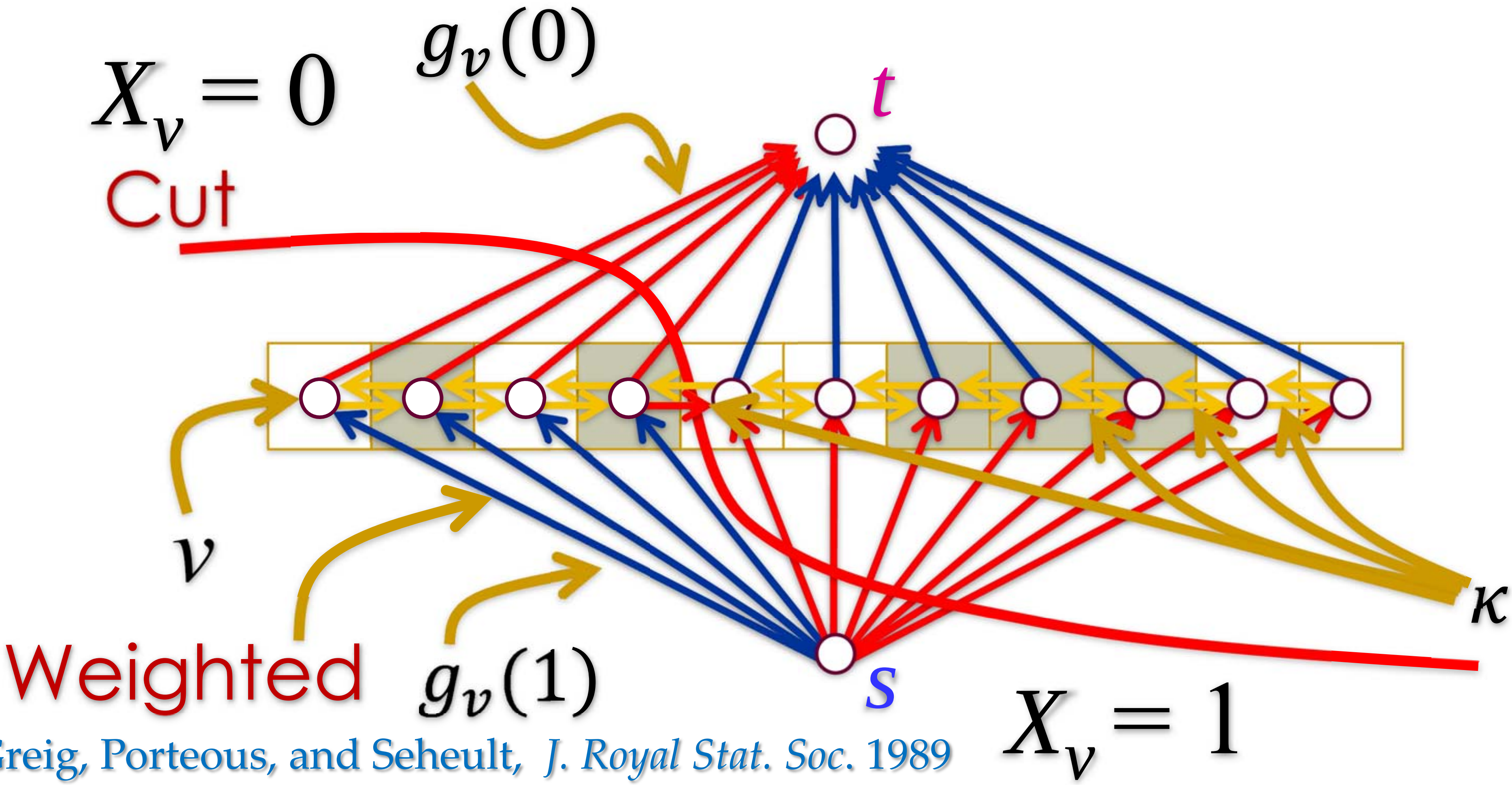
Graphs and cuts

- Choose $s, t \in \mathcal{V}$
- Cut: partition $\mathcal{V} = S \cup T$
  $S \cap T = \emptyset, s \in S, t \in T$
- Cost of cut: sum of the weights of the edges going from S to T
- Minimum cut: the cut with the minimum cost
- When all weights ≥ 0, the minimum cut can be found efficiently



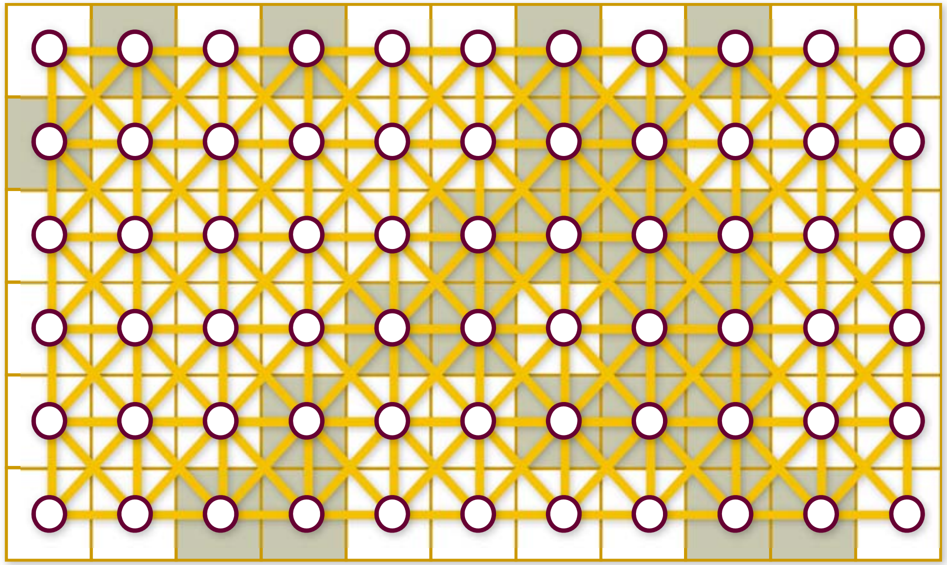
Graph Cuts (Binary case)

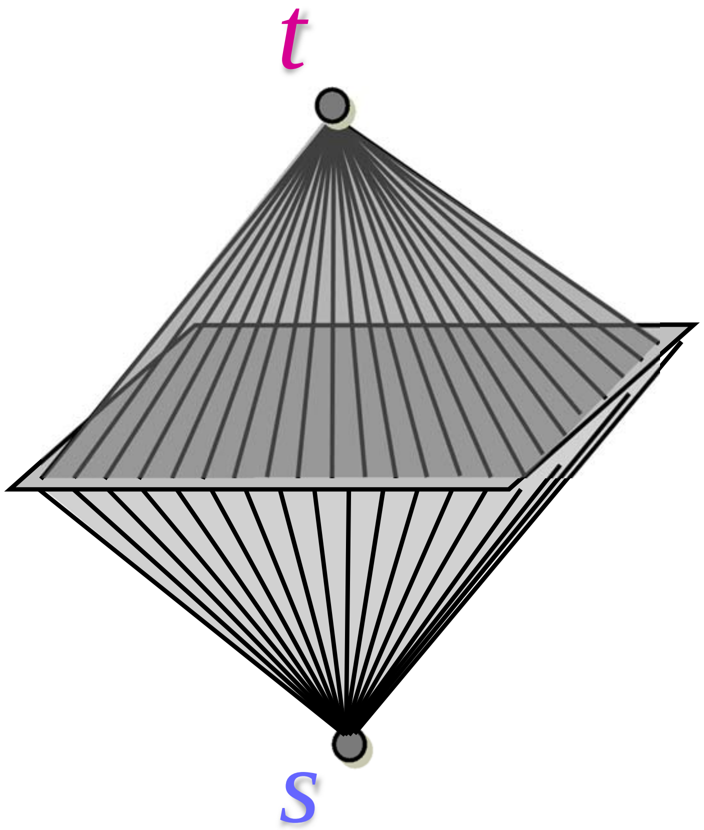
$$E(X) = \sum_{v \in V} g_v(X_v) + \sum_{(u,v) \in E} \kappa |X_u - X_v|$$



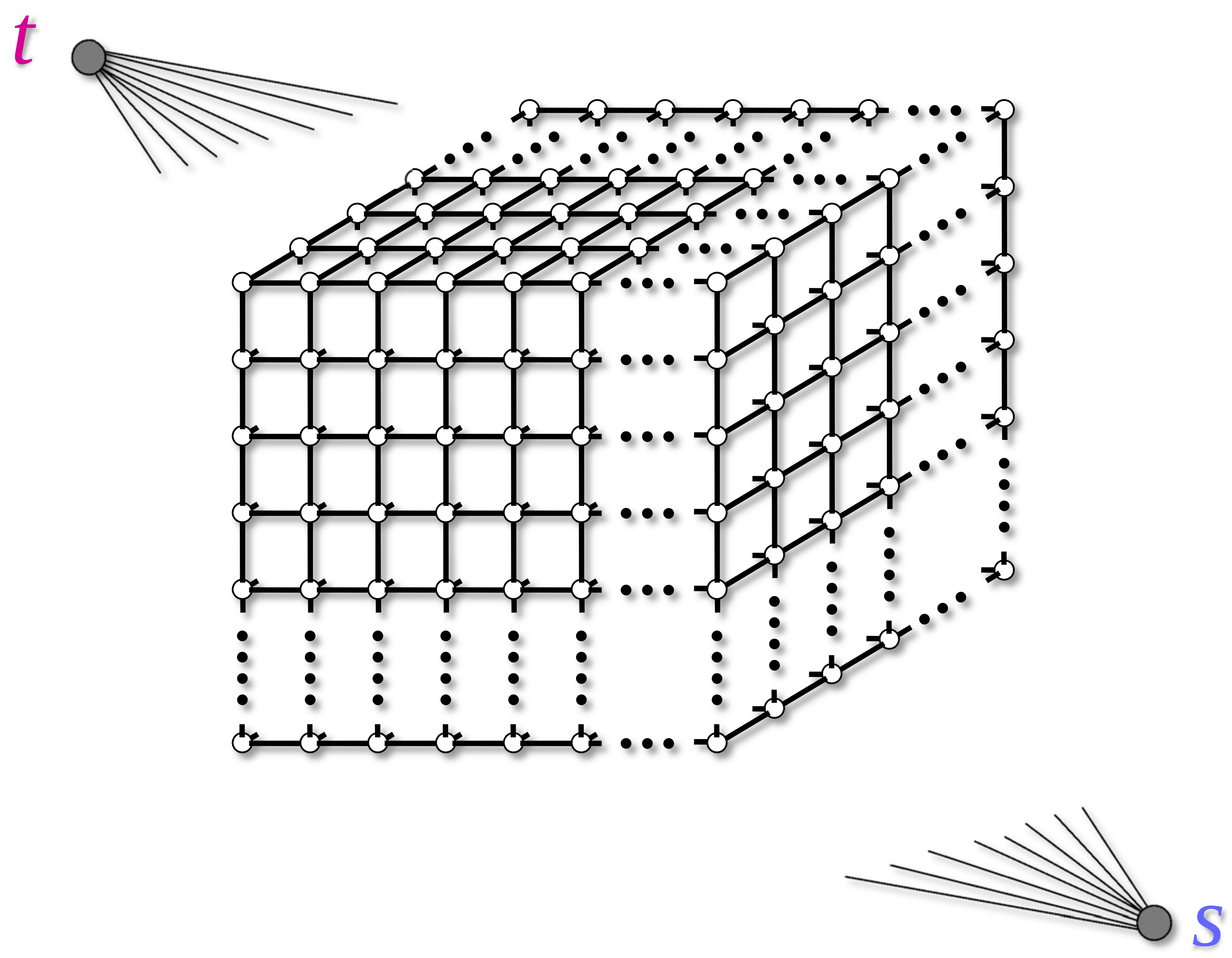
Greig, Porteous, and Seheult, *J. Royal Stat. Soc.* 1989

Image Plane Graph



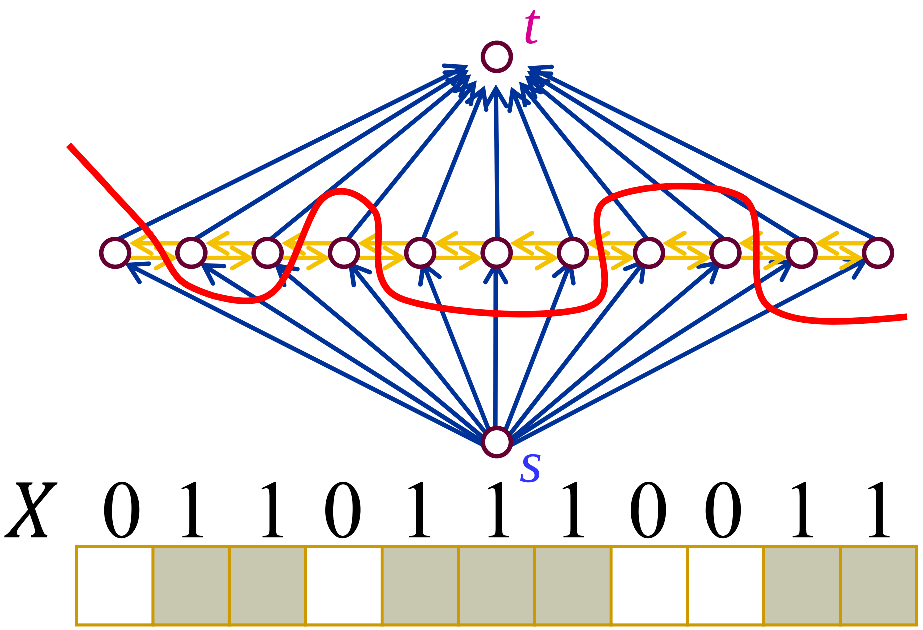


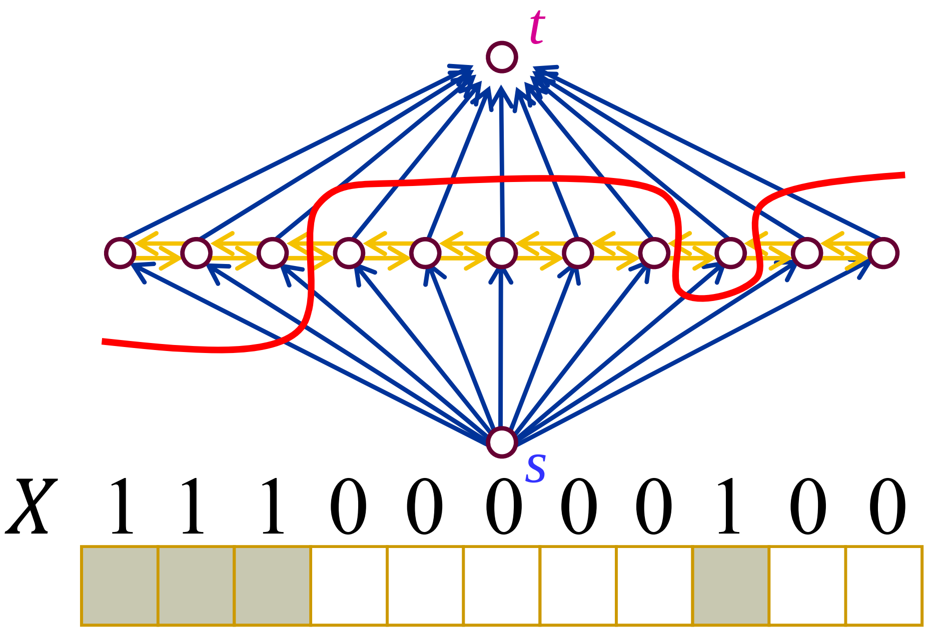
3D Case



Graph Cuts (Binary case)

- 1:1 correspondence between X and cuts





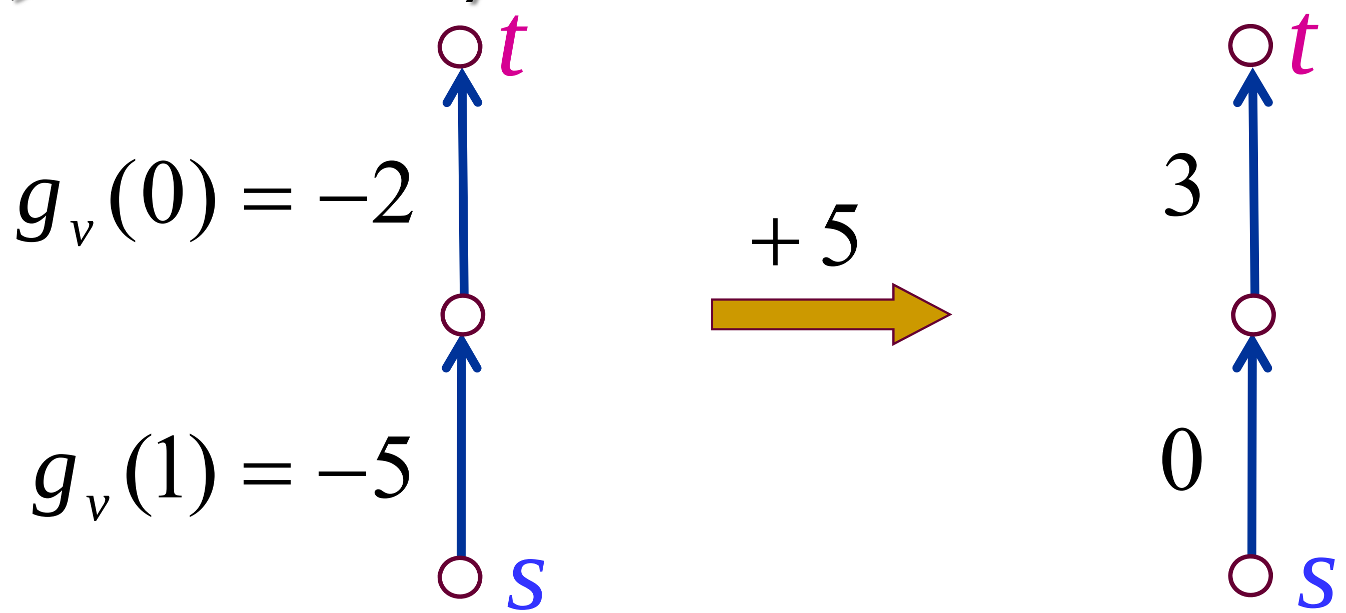
- Energy = Cut cost
- Minimum cut → Energy minimization
- The weight must be ≥ 0

Graph Cuts（Binary case)

- The edge weight must be ≥ 0
  - What energy can be minimized?

$$E(X) = \sum_{v \in V} g_v(X_v) + \sum_{(u,v) \in E} h_{uv}(X_u, X_v)$$

- $g_v(x)$ is arbitrary



Graph Cuts（Binary case)

- The edge weight must be ≥ 0
  - What energy can be minimized?

$$E(X) = \sum_{v \in V} g_v(X_v) + \sum_{(u,v) \in E} h_{uv}(X_u, X_v)$$

- $g_v(x)$ is arbitrary
- What about $h_{uv}(X_u, X_v)$?

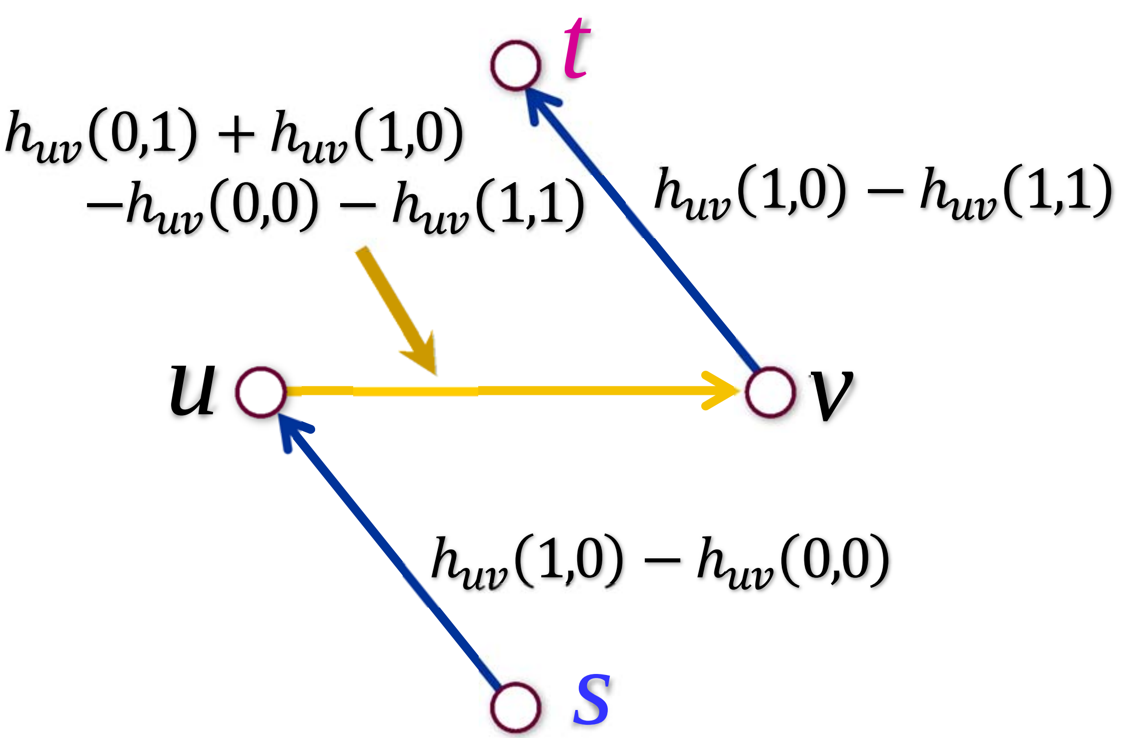
Submodularity condition

$$h_{uv}(0,0) + h_{uv}(1,1) \leq h_{uv}(0,1) + h_{uv}(1,0)$$

Graph Cuts（Binary case)

$$E(X) = \sum_{v \in V} g_v(X_v) + \sum_{(u,v) \in E} h_{uv}(X_u, X_v)$$

$$h_{uv}(0,0) + h_{uv}(1,1) \le h_{uv}(0,1) + h_{uv}(1,0)$$



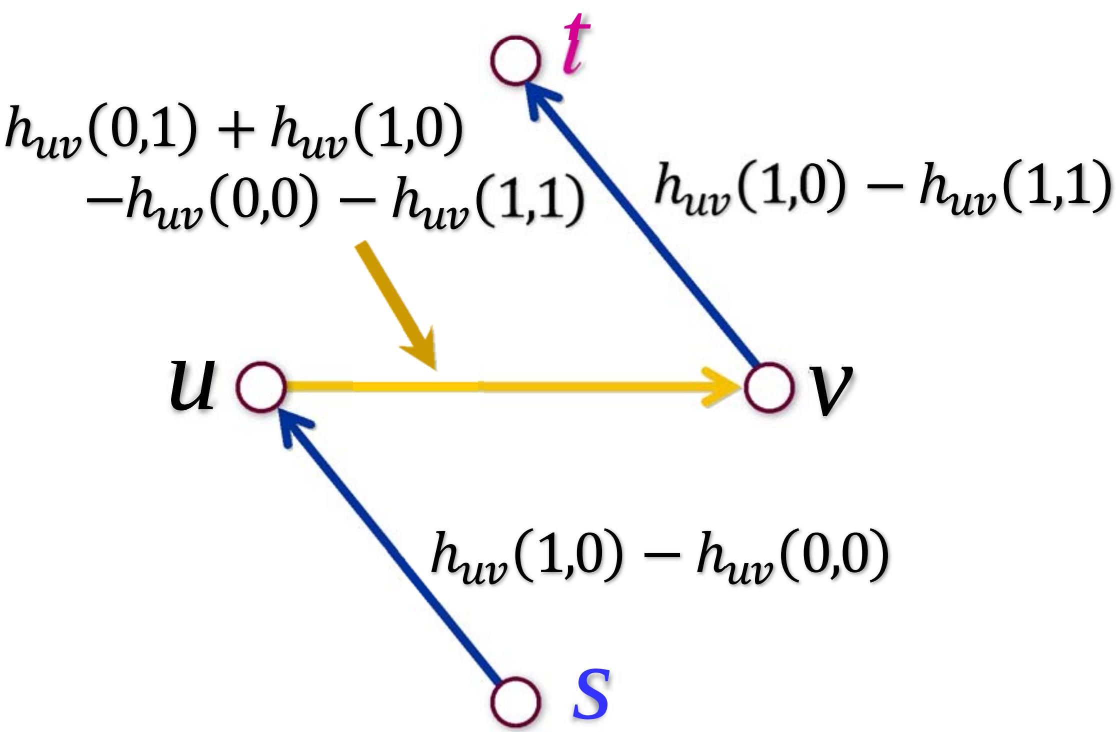
u	v	
0	0	$h_{uv}(1,0) - h_{uv}(1,1)$
0	1	$h_{uv}(0,1) + h_{uv}(1,0) - h_{uv}(0,0) - h_{uv}(1,1)$
1	0	$2h_{uv}(1,0) - h_{uv}(0,0) - h_{uv}(1,1)$
1	1	$h_{uv}(1,0) - h_{uv}(0,0)$

Graph Cuts（Binary case）

$$E(X)=\sum_{v\in V} g_v(X_v)+\sum_{(u,v)\in E} h_{uv}(X_u,X_v)$$

$$h_{uv}(0,0)+h_{uv}(1,1)\le h_{uv}(0,1)+h_{uv}(1,0)$$

Add the same value to all 4 cases

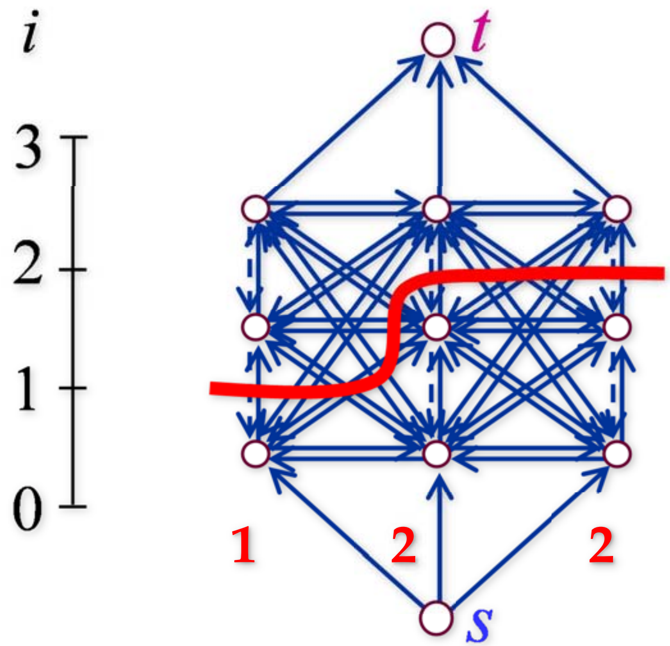


u	v	
0	0	$h_{uv}(0,0)$
0	1	$h_{uv}(0,1)$
1	0	$h_{uv}(1,0)$
1	1	$h_{uv}(1,1)$

Graph Cuts（Multi-label case）
（> 2 labels）

$$E(X) = \sum_{v \in V} g_v(X_v) + \sum_{(u,v) \in E} h_{uv}(X_u, X_v)$$

- If the labels has linear order $L = \{l_0, l_1, \ldots, l_k\}$
- Globally minimizable $\Leftrightarrow$ $h_{uv}(l_i, l_j)$ is a convex function of $i - j$



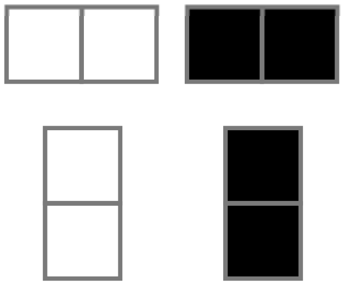
Ishikawa, *IEEE TPAMI* 2003

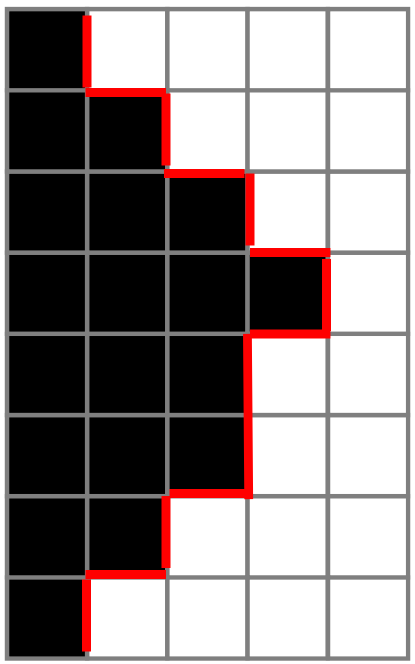
Outline

- Graph cuts
  - Uses minimum-cut algorithm for energy minimization
- Higher-order energy minimization
  - Binary labels: reduce to first order
    - Add variable
    - Some with ELC
  - Then use graph cuts
- Segmentation with shape prior
  - Shape restriction enables better segmentation
    - Realized by higher-order energy potential
    - Pulmonary Artery-Vein Segmentation
    - Coronary Lumen and Plaque Segmentation

First-order (pairwise) energy

Good (Low Energy)

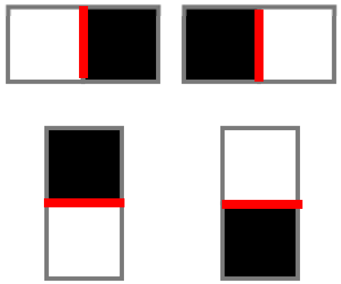


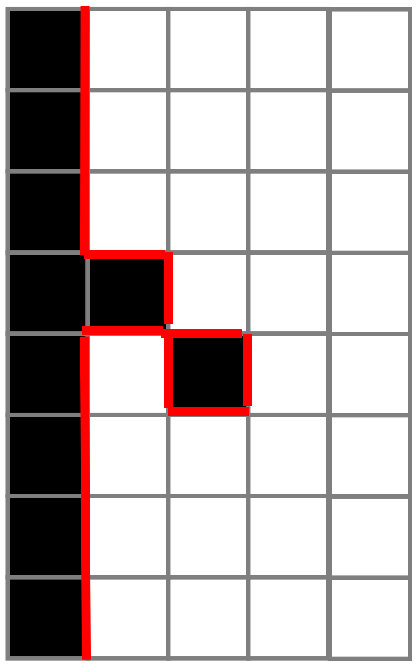


53 Good

14 Bad

Bad (High Energy)



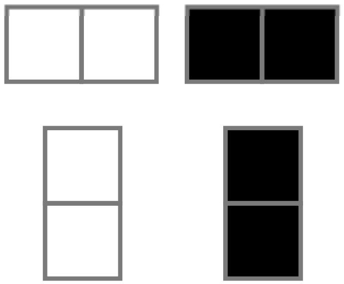


53 Good

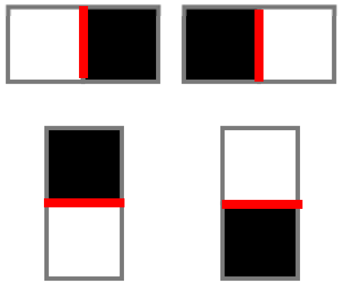
14 Bad

Higher-order energy

Good (Low Energy)



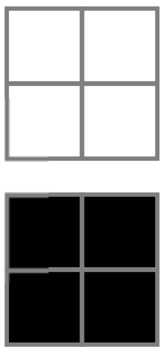
Bad (High Energy)



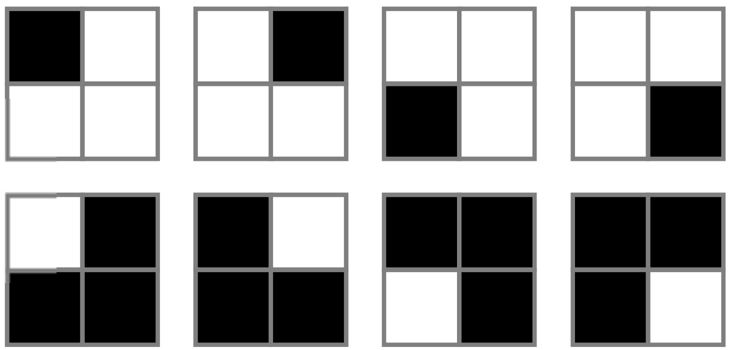
Better (Lower Energy)

Worse (Higher Energy)

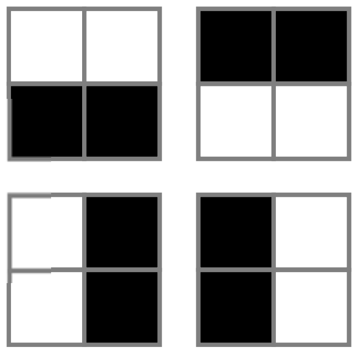
A



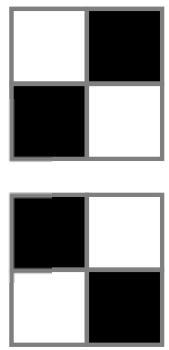
B



C

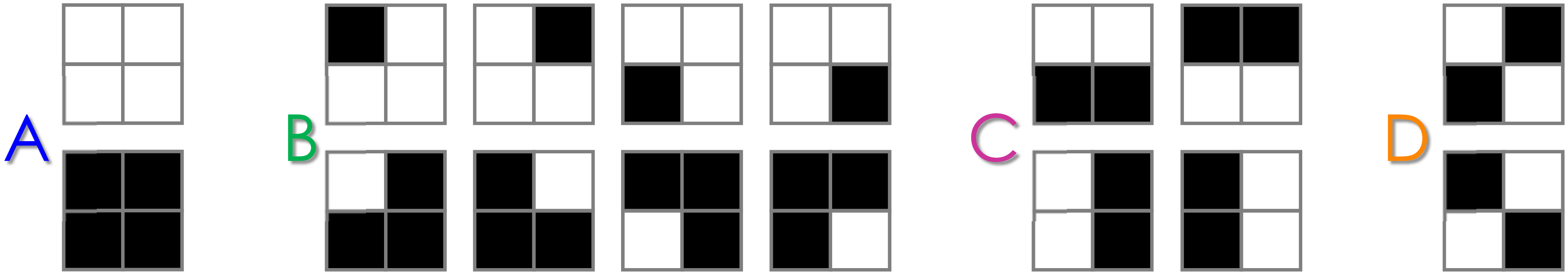


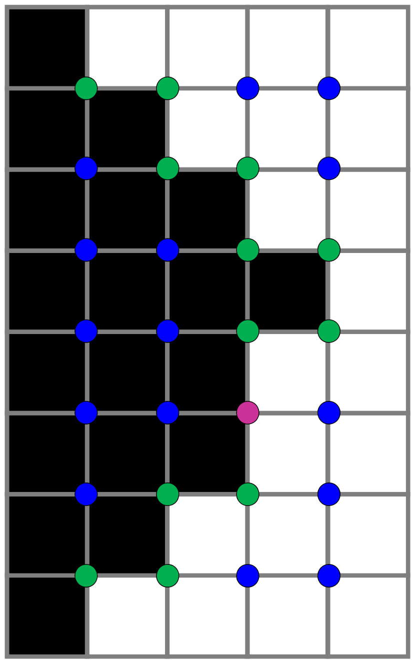
D



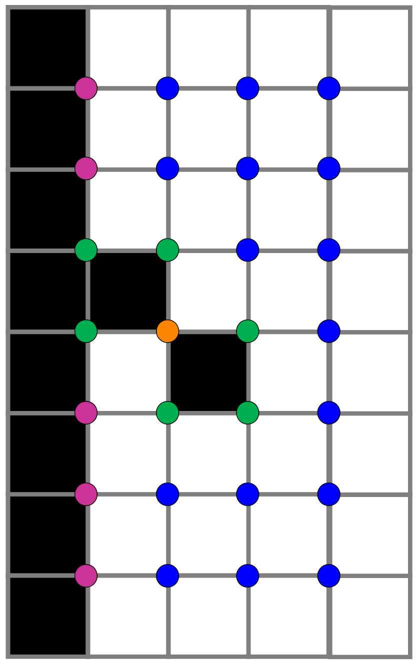
Higher-order energy

Better (Lower Energy) Worse (Higher Energy)





A: 15
B: 12
C: 1
D: 0



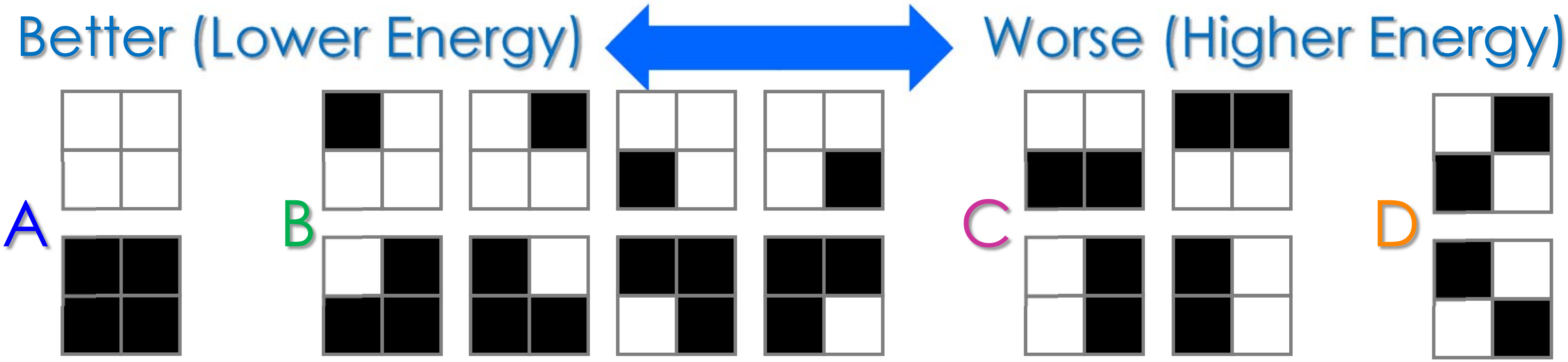
A: 16
B: 6
C: 5
D: 1

Higher-order energy

$$E(X) = \sum_{C \in \mathcal{C}} f_C(X_C)$$

$$= \sum_{v \in V} g_v(X_v) + \sum_{(u,v) \in E} h_{uv}(X_u, X_v)$$

$$+ \sum_{(u,v,s,t)} k_{uvst}\,(X_u, X_v, X_s, X_t)$$



Reducing higher-order energy

- Convert any higher-order binary energy

$$E(X) = E(X_1, \dots, X_n) = \sum_{C \in \mathscr{C}} f_C(X_C)$$

into an equivalent first-order energy

$$\downarrow$$

$$\tilde{E}(X) = \tilde{E}(X_1, \dots, X_n, \dots, X_m) = \sum g_v(X_v) + \sum h_{uv}(X_u, X_v)$$

- Adds variables

Ishikawa CVPR 2009, PAMI 2011

Functions of binary variables

- Pseudo-Boolean function (PBF)
  - Function of binary (0 or 1) variables
  - Can always write it uniquely as a polynomial
- One variable x: $E_0(1-x) + E_1 x$
- Two variables x, y:

$$E_{00}(1-x)(1-y) + E_{01}(1-x)y + E_{10}x(1-y) + E_{11}xy$$

- Three variables x, y, z:

$$E_{000}(1-x)(1-y)(1-z) + E_{001}(1-x)(1-y)z + \cdots + E_{111}xyz$$

2^{nd}-Order (Cubic) Case

- Reduce cubic PBF into quadratic one using

$$xyz = \max_{w \in \{0,1\}} w(x+y+z-2)$$

$x\ y\ z$

$0\ 0\ 0 \qquad 0 = \max_{w \in \{0,1\}} w(-2) = \max\{0,-2\} = 0$

$0\ 0\ 1 \qquad 0 = \max_{w \in \{0,1\}} w(-1) = \max\{0,-1\} = 0$

$0\ 1\ 1 \qquad 0 = \max_{w \in \{0,1\}} w \cdot 0 = 0$

$1\ 1\ 1 \qquad 1 = \max_{w \in \{0,1\}} w \cdot 1 = \max\{0,1\} = 1$

2^{nd}-Order (Cubic) Case

- Reduce cubic PBF into quadratic one using

$$xyz = \max_{w \in \{0,1\}} w(x + y + z - 2)$$

If $a < 0$, $axyz = \min_{w \in \{0,1\}} aw(x + y + z - 2)$

Thus $\min_{x,y,z \in \{0,1\}} axyz = \min_{x,y,z,w \in \{0,1\}} aw(x + y + z - 2)$

So, in a minimization problem, we can substitute $axyz$ by $aw(x + y + z - 2)$

cubic quadratic

Higher-Order Case

$$xyz = \max_{w \in \{0,1\}} w(x + y + z - 2)$$

$$xyzt = \max_{w \in \{0,1\}} w(x + y + z + t - 3)$$

$$xyztu = \max_{w \in \{0,1\}} w(x + y + z + t + u - 4)$$

$$x_1 \cdots x_d = \max_{w \in \{0,1\}} w(x_1 + \cdots + x_d - (d - 1))$$

$$\min a x_1 \cdots x_d = \min a w(x_1 + \cdots + x_d - (d - 1))$$

if $a < 0$

Positive Case

$$x_1 \cdots x_d = \min_{w_1,\ldots,w_{n_d}} \left\{ \sum_{i=1}^{n_d} w_i \left(k_i^d (-S_1 + 2i) - 1 \right) \right\} + S_2$$

quadratic

$$n_d = \left\lfloor \frac{d-1}{2} \right\rfloor \qquad k_i^d = \begin{cases} 1 & d \text{ is odd and } i = n_d \\ 2 & \text{otherwise} \end{cases}$$

$$S_1 = \sum_{i=1}^{d} x_i \qquad S_2 = \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} x_i x_j$$

Elementary symmetric polynomials

Thus,

$$\min a x_1 \cdots x_d = \min a \left\{ \sum_{i=1}^{n_d} w_i \left(k_i^d (-S_1 + 2i) - 1 \right) \right\} + a S_2$$

if $a > 0$

Experiment: 4th deg. FoE denoising

Original

Noise added

4th degree

Discussion

- # of added variables increases exponentially with the degree
- In general, this is inevitable
  - Degree d $\Rightarrow$
    can take 2^d different values in general
- But we don't have to have the same exact function
  - All we want is the minimizer

Reduce Without Adding Variables

- Convert many higher-order binary energies

$$E(X) = E(X_1, \dots, X_n) = \sum_{C \in \mathscr{C}} f_C(X_C)$$

into an equivalent first-order energy

$$\tilde{E}(X) = \tilde{E}(X_1, \dots, X_n \qquad) = \sum g_v(X_v) + \sum h_{uv}(X_u, X_v)$$

- Without adding variables
- How?

Example: a cubic term

- $\varphi(x, y, z)$: cubic (3rd degree)

$$\varphi(0,0,0) = a,\ \varphi(1,0,0) = b,\ \varphi(0,1,0) = c,\ \varphi(1,1,0) = d,$$
$$\varphi(0,0,1) = e,\ \varphi(1,0,1) = f,\ \varphi(0,1,1) = g,\ \varphi(1,1,1) = h$$

- $\varphi(x,y,z) = a(1-x)(1-y)(1-z) + bx(1-y)(1-z)$
  $+ c(1-x)y(1-z) + dxy(1-z) + e(1-x)(1-y)z$
  $+ fx(1-y)z + g(1-x)yz + hxyz$

- xyz coefficient: $s = -a + b + c - d + e - f - g + h$

- Define new function:

$$\varphi'(x, y, z) = \begin{cases} \varphi(x, y, z) & when\ (x, y, z) \neq (0,0,0) \\ \varphi(0,0,0) + s & when\ (x, y, z) = (0,0,0) \end{cases}$$

i.e., value is added s only when $(x, y, z) = (0,0,0)$

Example: a cubic term

- Define new function:

$$\varphi'(x,y,z) = \begin{cases} \varphi(x,y,z) & when\ (x,y,z) \neq (0,0,0) \\ \varphi(0,0,0) + s & when\ (x,y,z) = (0,0,0) \end{cases}$$

  i.e., value is added s only when $(x,y,z) = (0,0,0)$

- New xyz coefficient (replace a with $a + s$):
  $s' = -(a+s) + b + c - d + e - f - g + h = s - s = 0$

- So φ' is now quadratic (2nd degree)

- Similarly, we can reduce the degree by changing one of the 8 possible values

- But φ and φ' are different functions!

When can we do this?

$\varphi'(0,0,0) = a + s$, $\varphi'(1,0,0) = b$, $\varphi'(0,1,0) = c$, $\varphi'(1,1,0) = d$,
$\varphi'(0,0,1) = e$, $\varphi'(1,0,1) = f$, $\varphi'(0,1,1) = g$, $\varphi'(1,1,1) = h$

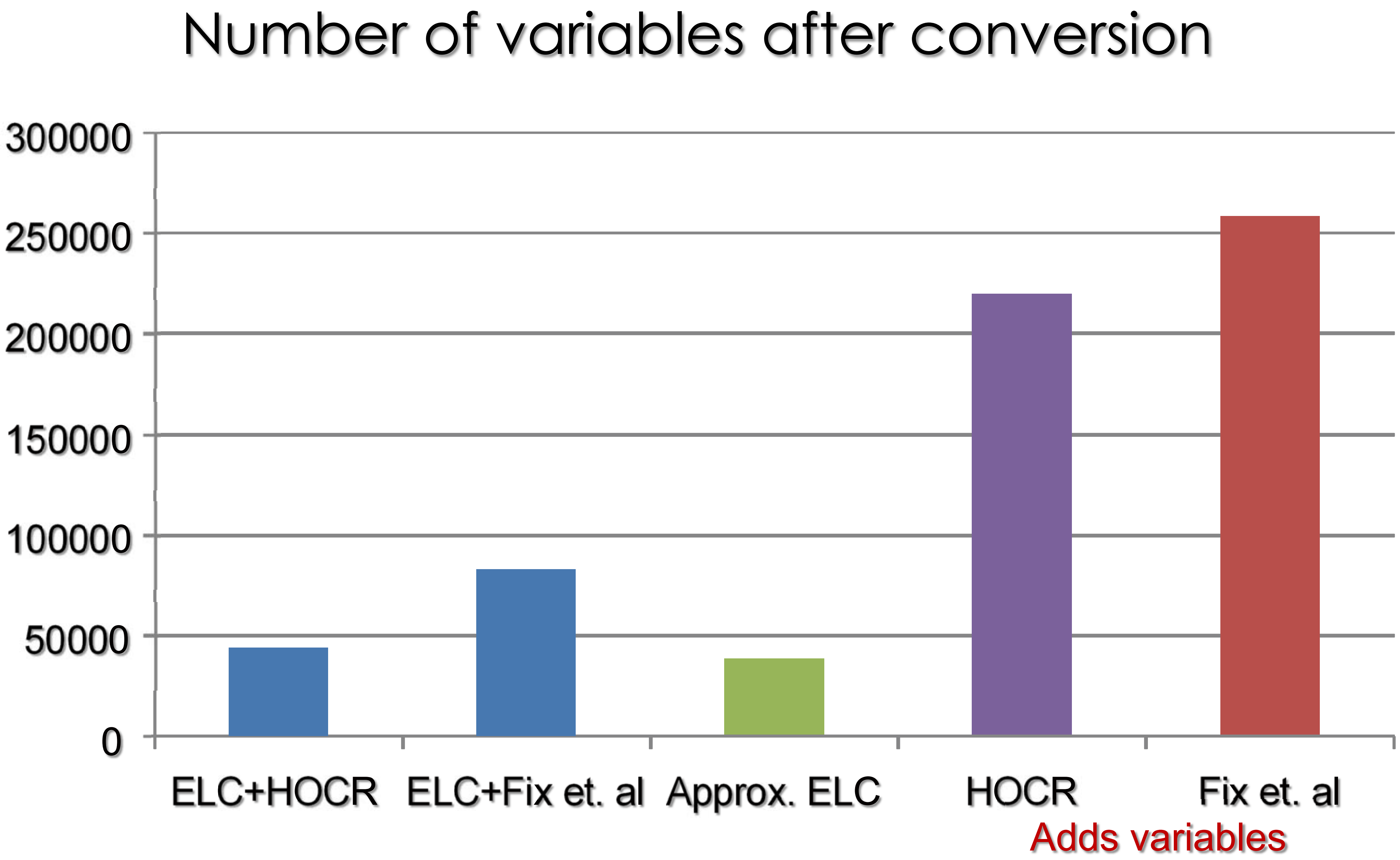
- Different only when $(x, y, z) = (0,0,0)$
- Suppose φ is a potential in $E(X) = \sum f_C(X_C)$
  - x, y, z are three of the variables in X
- If
  - $s > 0$, and
  - For minimizer X of $E(X)$, $(x, y, z) \neq (0,0,0)$,

then, we can replace φ with φ' without changing the minimizer

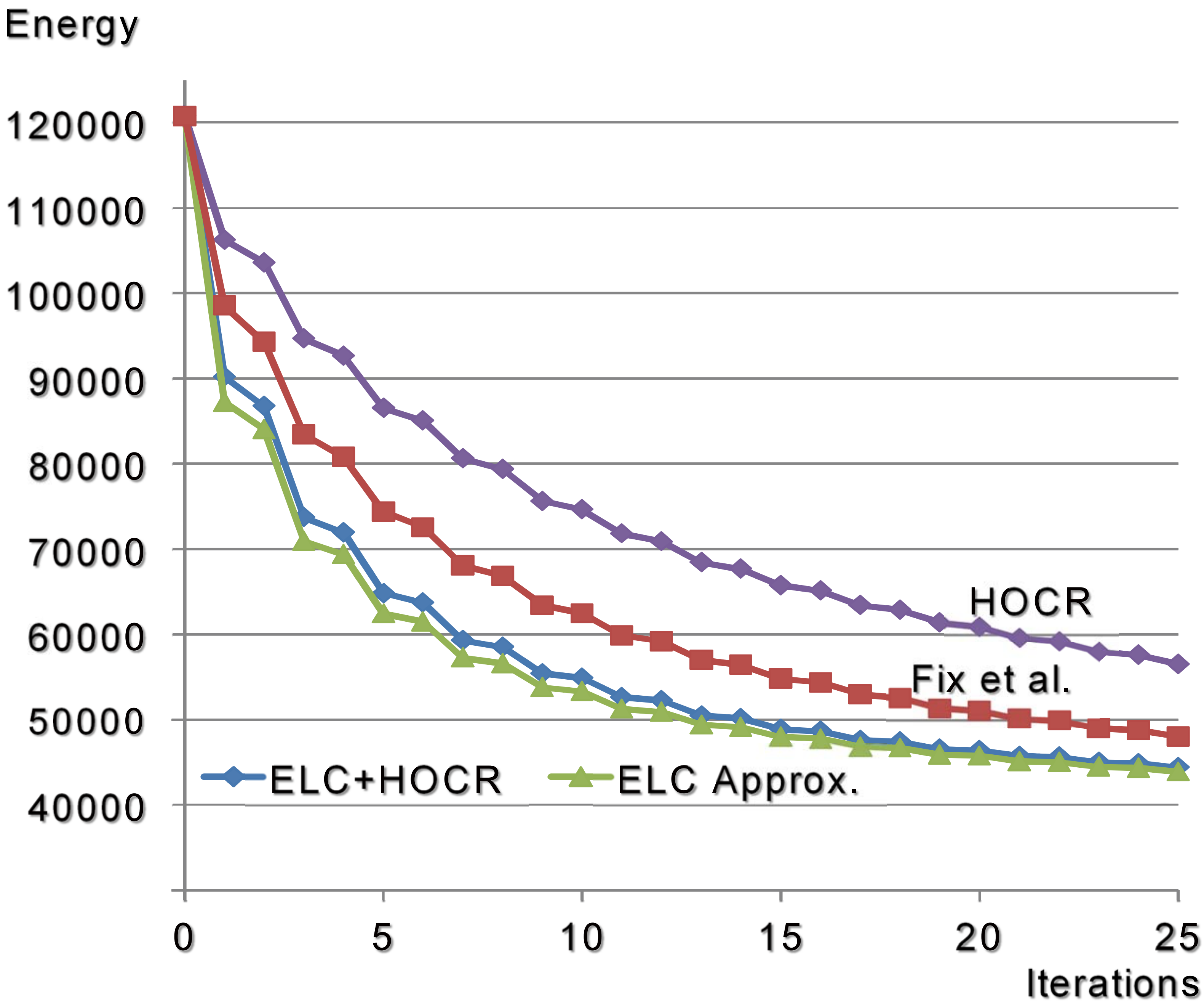
ELC

- Excludable Local Configuration (ELC)
  - Configuration = assignment of 0 or 1 to vars
  - A (usually) locally-testable sufficient condition for local configuration (x, y, z) to be not part of global minimizer
  - "Excludable as a part of global minimizer"
- ELC may not exist
- May take time to find
- Approximation
  - Just use the local configuration (x, y, z) with the largest value $\varphi(x, y, z)$

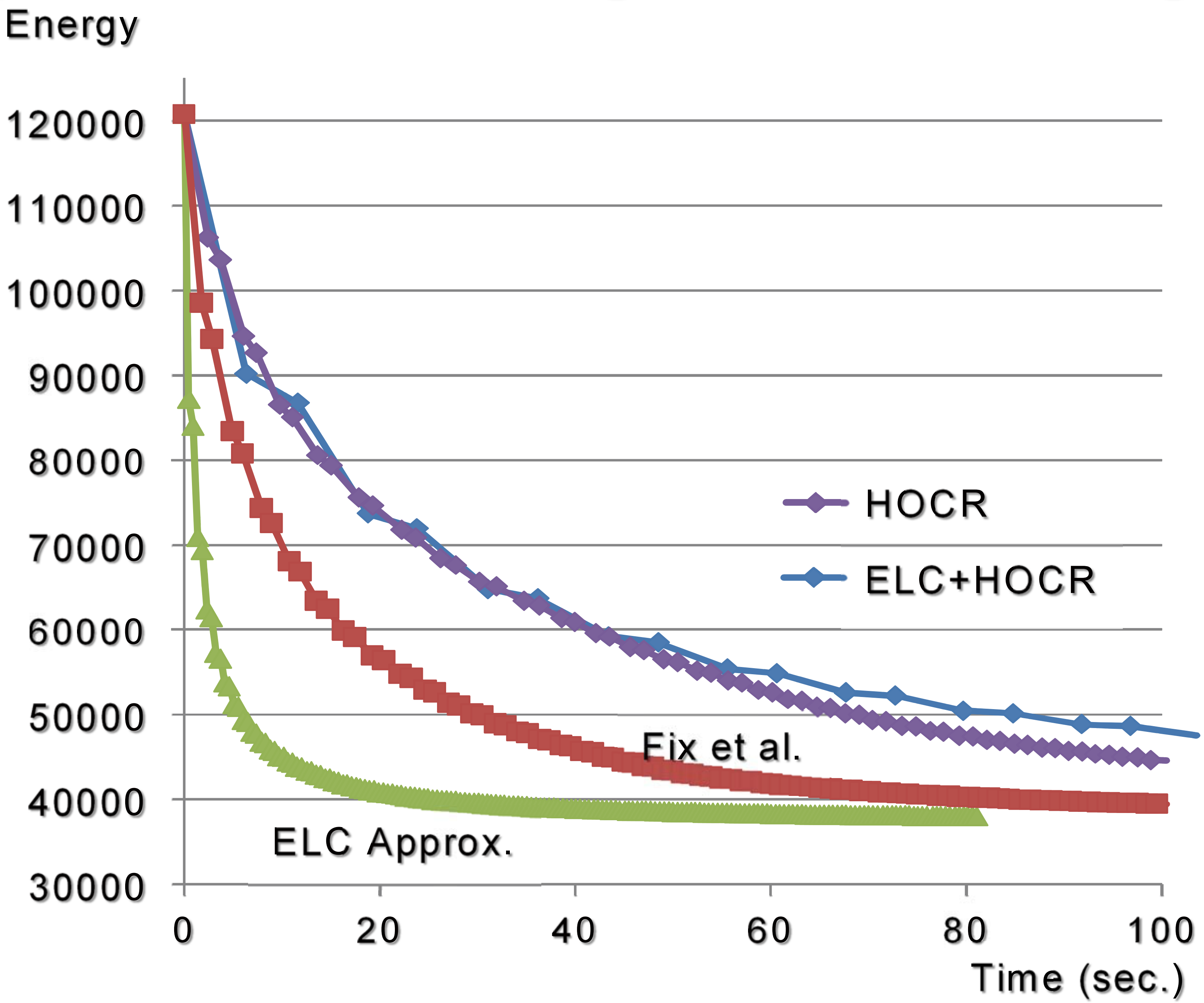
Experiment: 4^{th} deg. FoE denoising



Experiment: 4th deg. FoE denoising



Experiment: 4th deg. FoE denoising



Experiment: 4^{th} deg. FoE denoising

- ELC exists:
  - 3^{rd} degree: 96.12%
  - 4^{th} degree: 99.60%
- Approximation:
  - Guessed configuration is in fact an ELC
    - 3^{rd} degree: 88% of the time
    - 4^{th} degree: 97% of the time
  - Even if it is not an ELC, it is not part of minimizer
    - 3^{rd} degree: 99.98%
    - 4^{th} degree: 99.997%
  - 99.99988% of (labeled) pixels correctly labeled

Outline

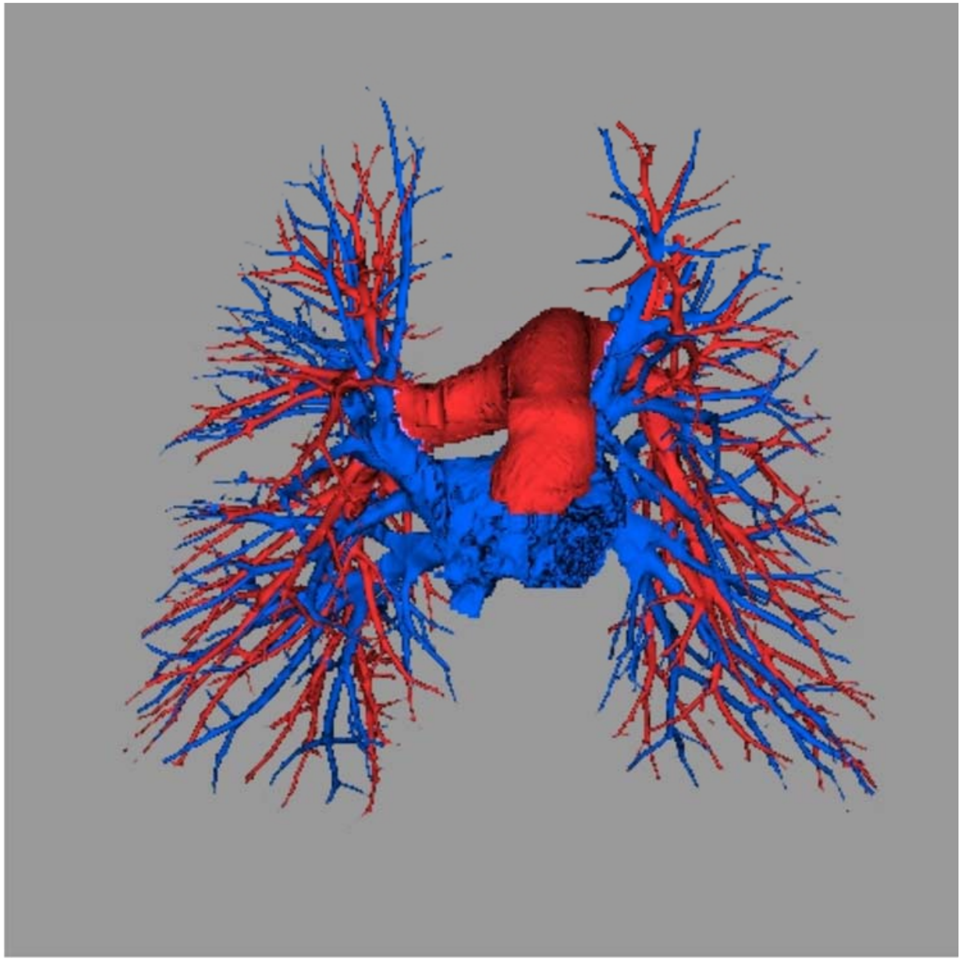
- Graph cuts
  - Uses minimum-cut algorithm for energy minimization
- Higher-order energy minimization
  - Binary labels: reduce to first order
    - Add variable
    - Some with ELC
  - Then use graph cuts
- Segmentation with shape prior
  - Shape restriction enables better segmentation
    - Realized by higher-order energy potential
    - Pulmonary Artery-Vein Segmentation
    - Coronary Lumen and Plaque Segmentation

Higher-order energy: special form

- Special form of binary-valued higher-order energy
  - $f_C(X_C) = f_C(x_1, \cdots, x_k) = -\alpha x_1 \cdots x_k$
    - Takes $-\alpha < 0$ only when all variables are 1, Takes 0 otherwise
  - $f_C(X_C) = f_C(x_1, \cdots, x_k) = -\alpha(1 - x_1) \cdots (1 - x_k)$
    - Takes $-\alpha < 0$ only when all variables are 0, Takes 0 otherwise
- A function of this form is converted to a submodular first-order function

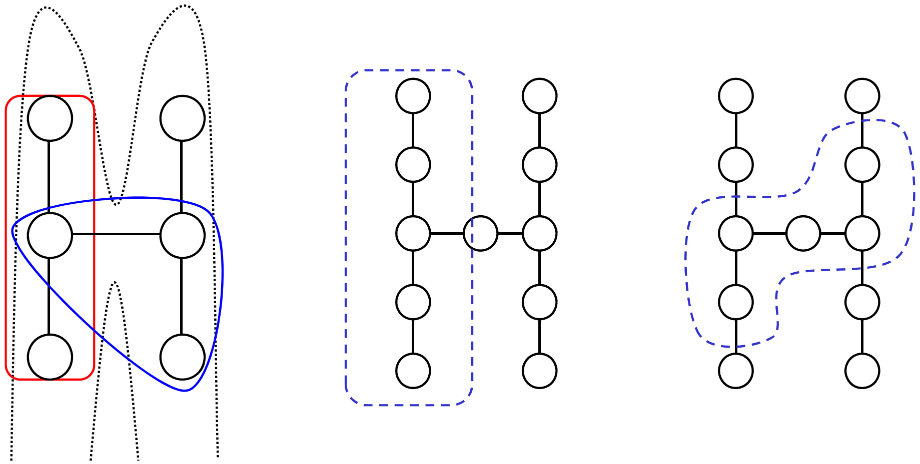
Kohli, Kumar, and Torr, *TPAMI* 2009; Ishikawa *TPAMI* 2011

Fully-automatic Pulmonary Artery-Vein Segmentation



Kitamura, Li, Ito, Ishikawa, *IJCV2015*

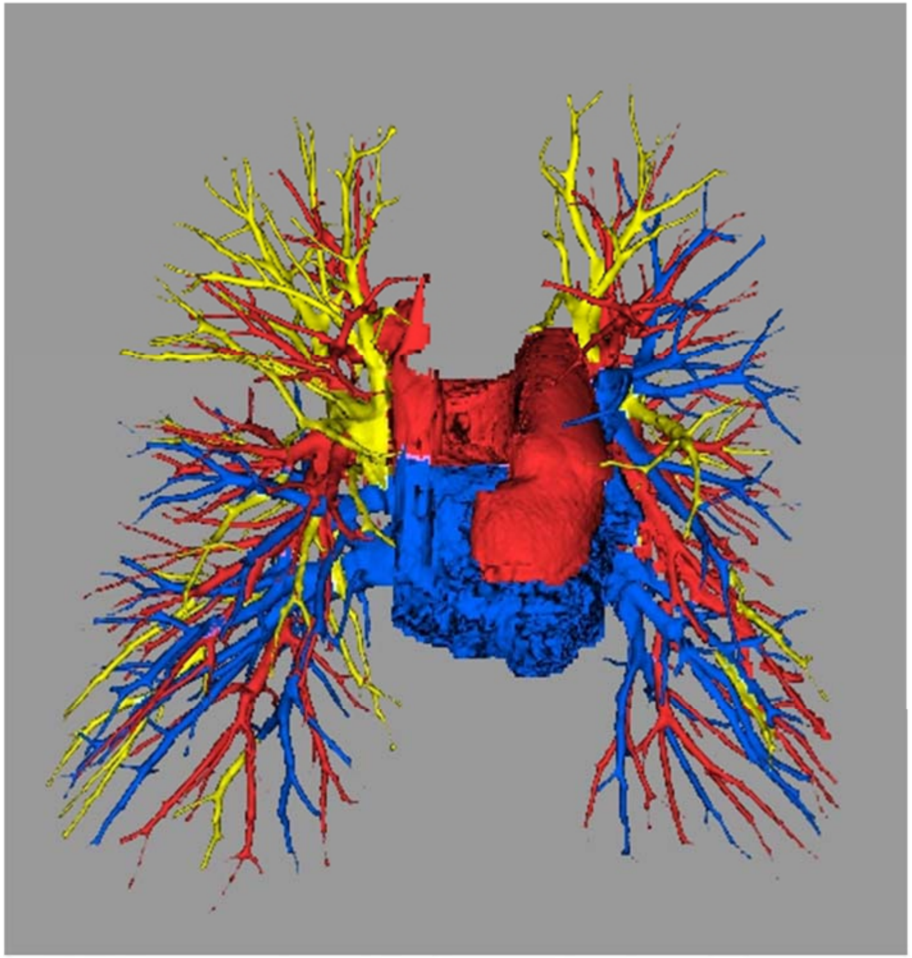
Data-Dependent Clique Potential



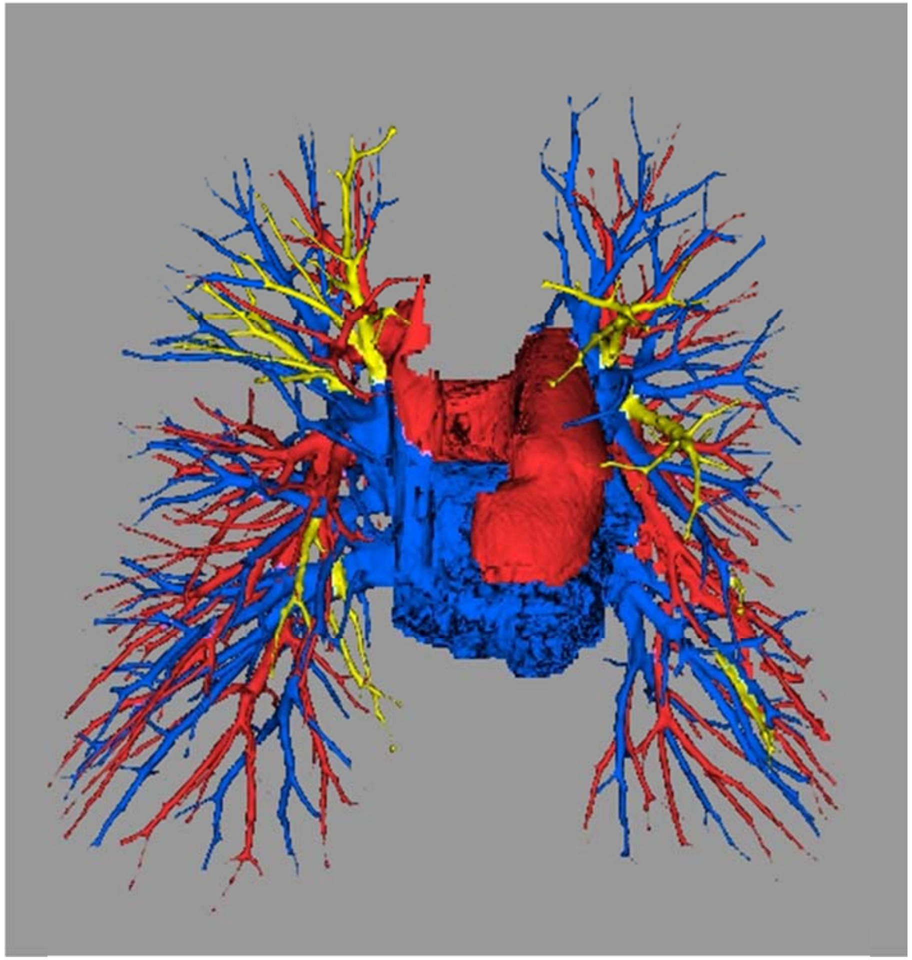
- Arrange higher-order terms adaptively with data
- Straighter cliques has lower energy

Segmentation Results

- Difference by the higher-order terms



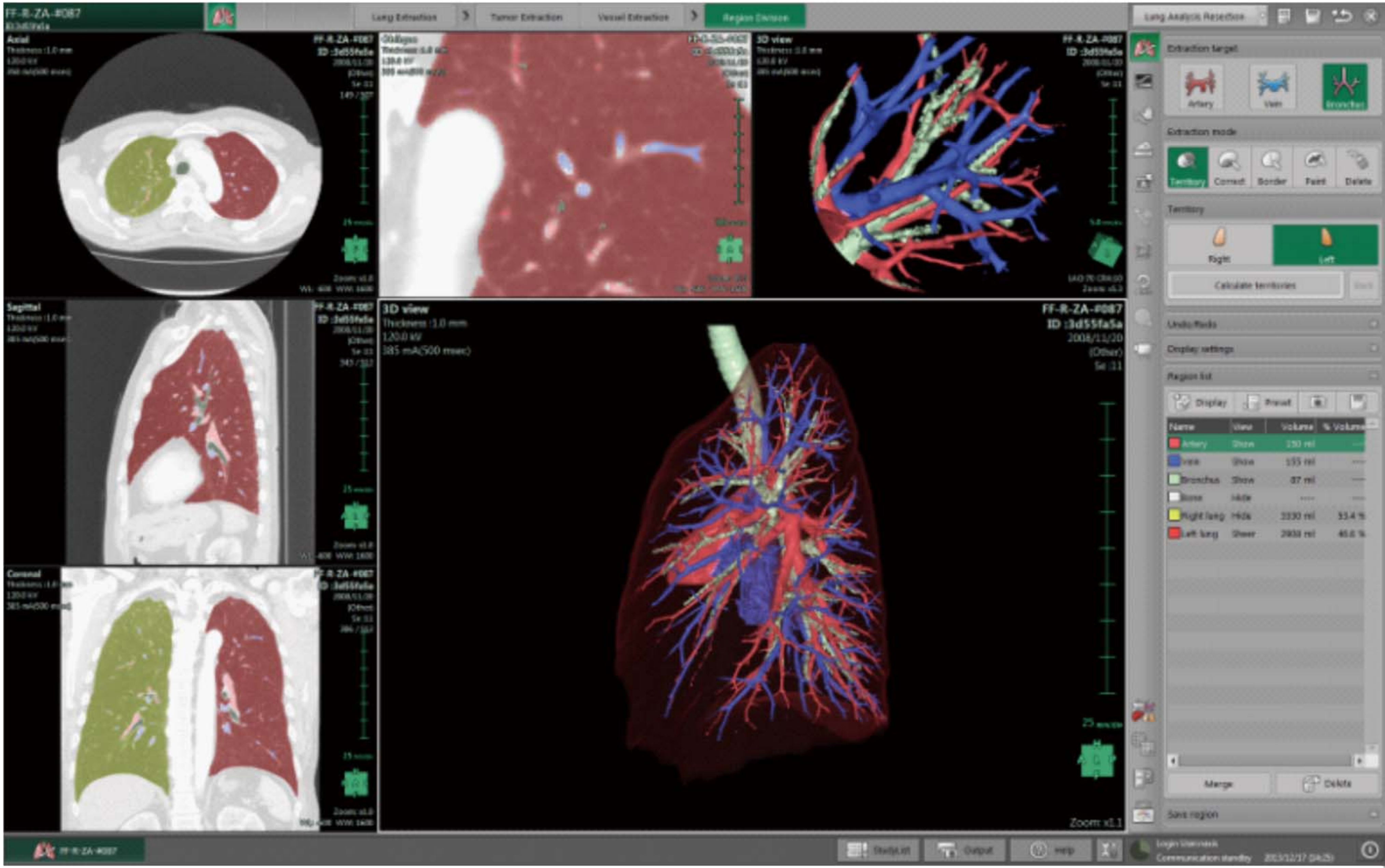
Without the Higher-Order Terms



With the Higher-Order Terms

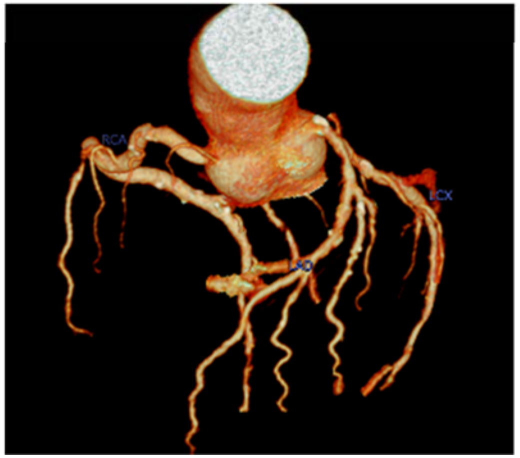
Red: Correctly labeled as artery Blue: Correctly labeled as vein
Yellow: Incorrectly labeled

Pre-Surgery Simulator

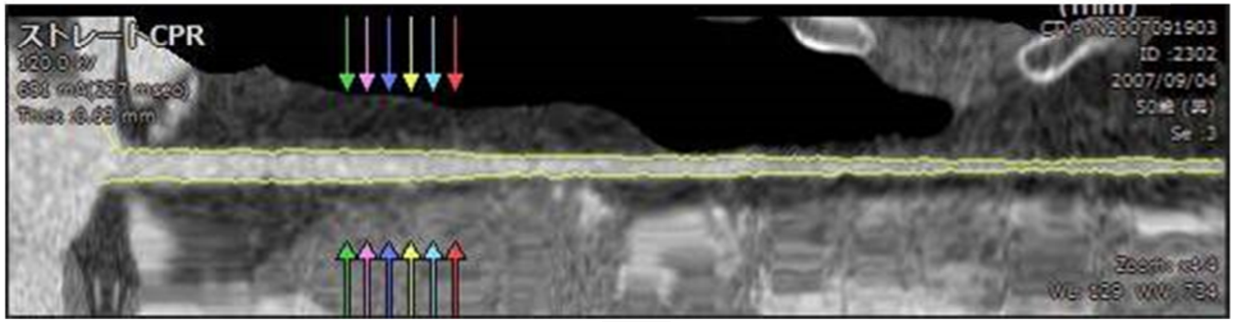


Coronary Lumen and Plaque Segmentation from CTA

Center line detection

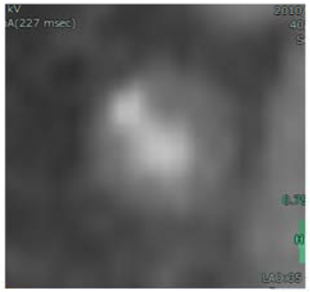


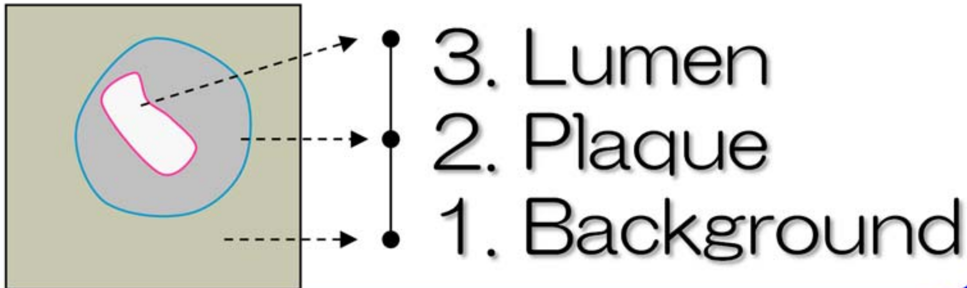
Reconstruction along the center line (3D)



Linearly-orderd 3-class labels

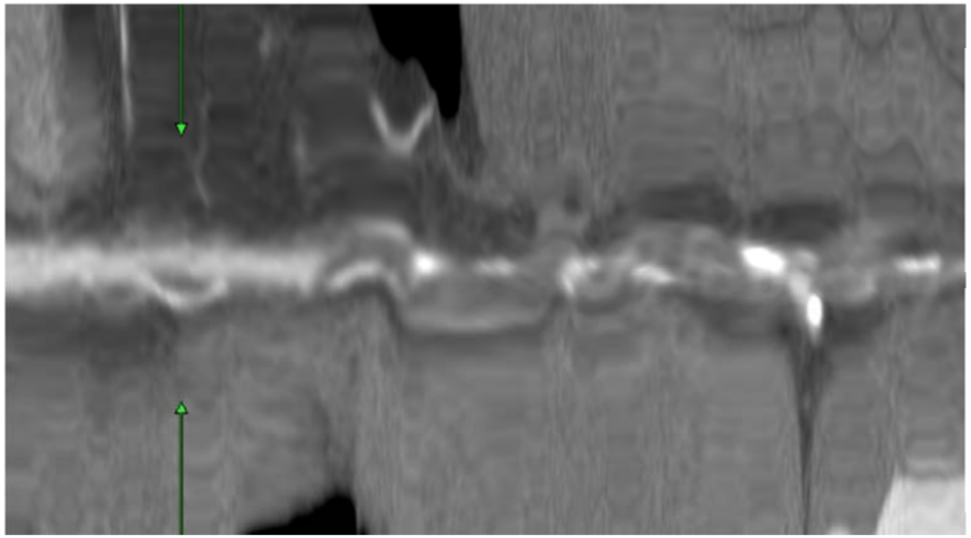
$$\min E(y) = \sum_{a \in V} \theta_a(y_a) + \sum_{a \in V, b \in N_a} \theta_{ab}(y_a, y_b), \quad y_a \in L = \{1, 2, 3\}$$





Coronary Lumen and Plaque Segmentation from CTA

- Difficult with first-order energy

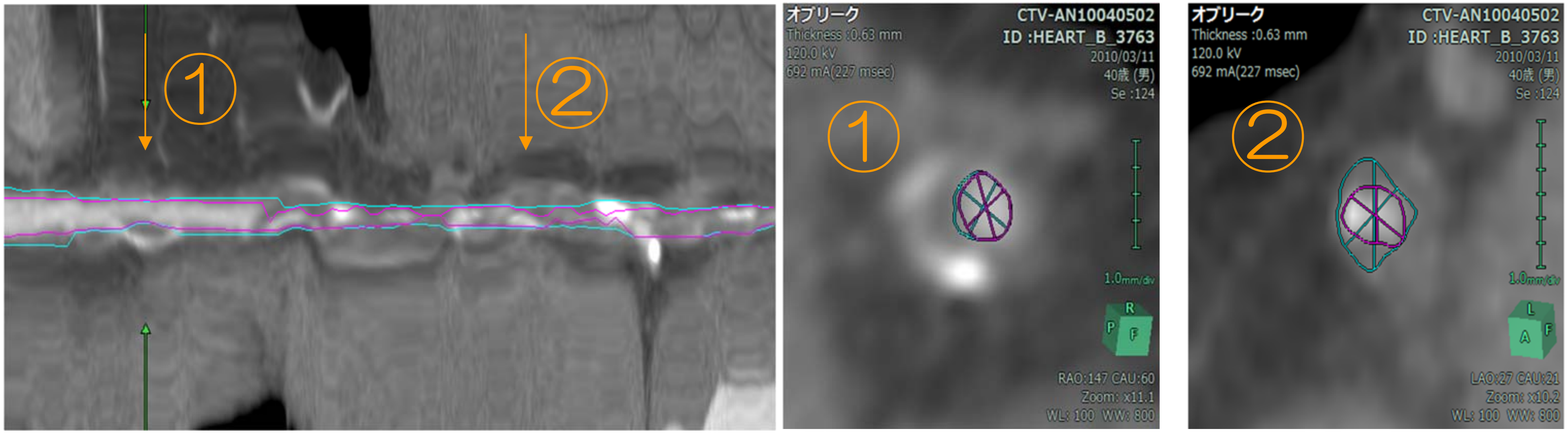


① Hard plaque

Strong edges attract the contour

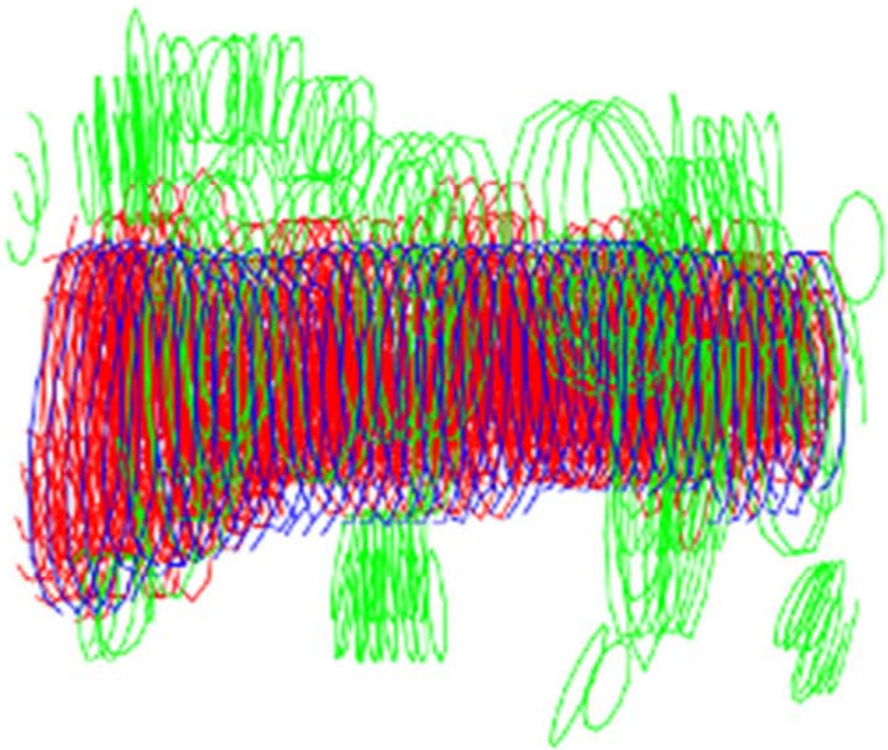
② Soft plaque

Low contrast hard to discern



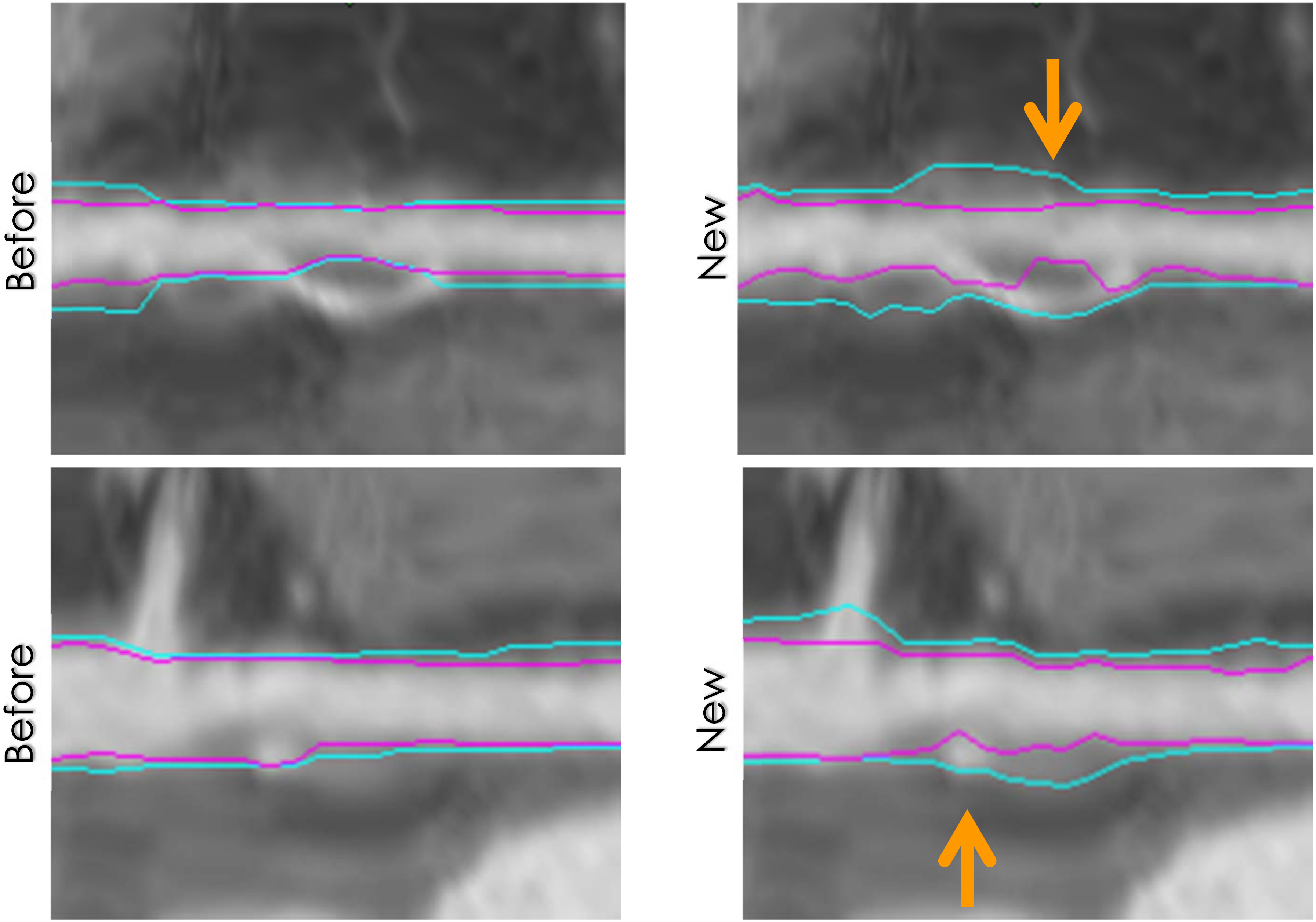
Higher-Order Shape Restriction

- Blood vessel cross section is round
- Generate circular candidate shapes
- Give lower energy for each shape when
  - All voxels outside the shape are labeled background
  - All voxels inside the shape are labeled foreground

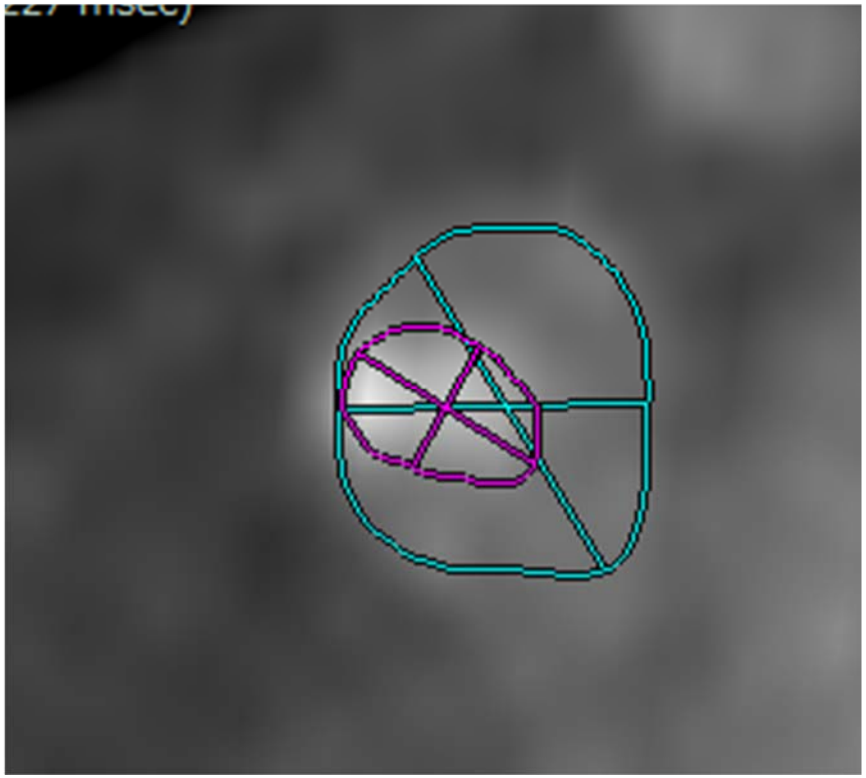
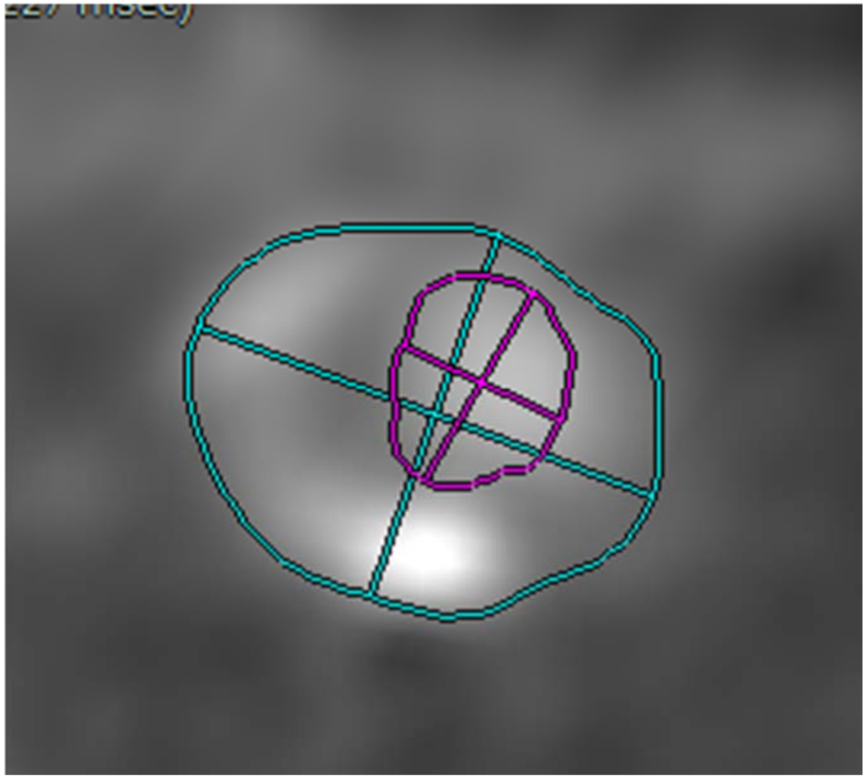


Red: Realized shape restriction
(after energy minimization)
Green: Unrealized shape restriction
Blue: Segmentation result

Segmentation Results



Segmentation Results

Currently Top-ranked in the Rotterdam Coronary Artery Stenoses Detection and Quantification Evaluation Framework

Conclusion

- Graph cuts
  - Uses minimum-cut algorithm for energy minimization
- Higher-order energy minimization
  - Binary labels: reduce to first order
    - Add variable
    - Some with ELC
  - Then use graph cuts
- Segmentation with shape prior
  - Shape restriction enables better segmentation
    - Realized by higher-order energy potential
    - Pulmonary Artery-Vein Segmentation
    - Coronary Lumen and Plaque Segmentation

Excludable Local Configuration

- For any binary energy
$$E(X) = E(x_1, \dots, x_n) = \sum_{C \in \mathcal{C}} \alpha_C x_C$$
and $C \in \mathcal{C}$, denote
$$E_C(X) = \sum_{D \in \mathcal{C},\, C \cap D \neq \emptyset} \alpha_D x_D$$

- **Definition** A value assignment $u \in \mathbb{B}^C$ is said to be an excludable local configuration (ELC) if there exists another assignment $v \in \mathbb{B}^C$ such that
$$\max_{X \in \mathbb{B}^n,\ X|_C = v} E_C(X) < \min_{X \in \mathbb{B}^n,\ X|_C = u} E_C(X)$$
where $X|_C \in \mathbb{B}^C$ is the restriction of $X \in \mathbb{B}^n$ to C

Excludable Local Configuration

- **Theorem** If $u \in \mathbb{B}^C$ is an ELC, then

$$\min_{X \in \mathbb{B}^n} E(X) < \min_{X \in \mathbb{B}^n,\ X|_C = u} E(X)$$

  In other words, no minimizer X of the energy takes the local configuration u on C.

$$E_C(X) = \sum_{D \in \mathcal{C},\ C \cap D \neq \emptyset} \alpha_D x_D$$

- So, if an ELC $u \in \mathbb{B}^C$ can be found, adding

$$\psi(v) = \begin{cases} 0 & when\ v \neq u \\ s > 0 & when\ v = u \end{cases}$$

  to the energy does not change the minimizer

Details

- Usable ELC depends on:
  - Sign of the coefficient of the highest-degree term
  - Parity of the ELC
- xyz coefficient

$$s = -a + b + c - d + e - f - g + h$$