Lect 6: Graph Partition by Swendsen-Wang Cuts

Topics

1. Graph partition and labeling
2. Clustering with bottom-up edge probabilities
3. Moving in the partition space
   Reversible and detailed balance
4. Swendsen-Wang Cuts
5. Examples
   segmentation, stereo, and motion etc.

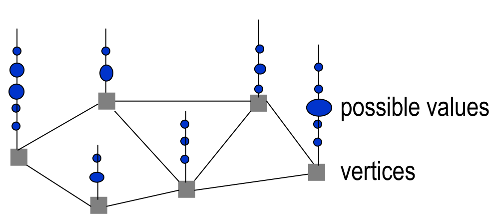
Graph partition

A graph G=<V, E> is partitioned into an unknown number of sub-graphs

$$X=(V_1,V_2,...,V_L),\quad V=\cup_{i=1}^{L}V_i,\quad V_i\cap V_j=\emptyset,\ i\neq j.$$

Each subgraph is often assigned a lable or color, thus the partition problem is augmented into a graph coloring/labeling problem.

E.g. image segmentation (label), stereo (disparity), motion (velocity), …



Traversing the partition space

Let Ω_L be the partition space of all possible L-way partitions of V

$$\Omega_L = \{X = (V_1, \ldots, V_L)\}$$

The partition space is denoted by

$$\Omega = \cup_{L=1}^{|V|} \Omega_L$$

How do we design a Markov chain to traverse the partition space?

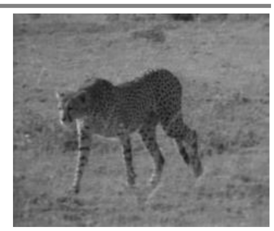
This is a challenging question facing many tasks,

e.g. how can the split-merge operator be made reversible in image segmentation?

So what? You need to have algorithms that is capable of global optimality independent of initial solutions !!

An example of segmentation

For speed, one can group adjacent pixels of constant intensity into super-pixels to reduce the graph. (This step has risk, more proper treatment should be using multi-level Swendsen-Wang cut, which can freeze and open the atomic regions.)



input image

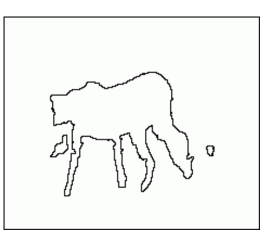
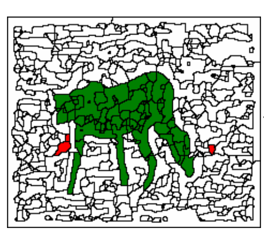
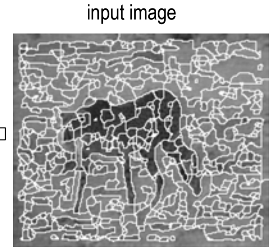


image segmentation result

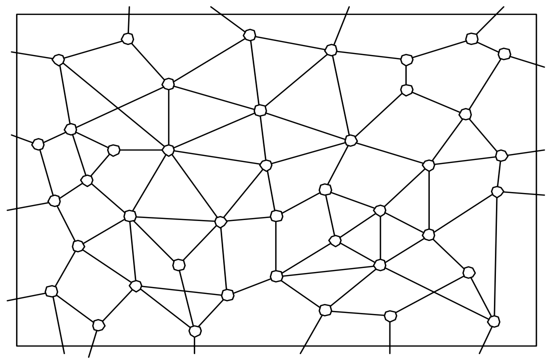


graph partition (labeling)



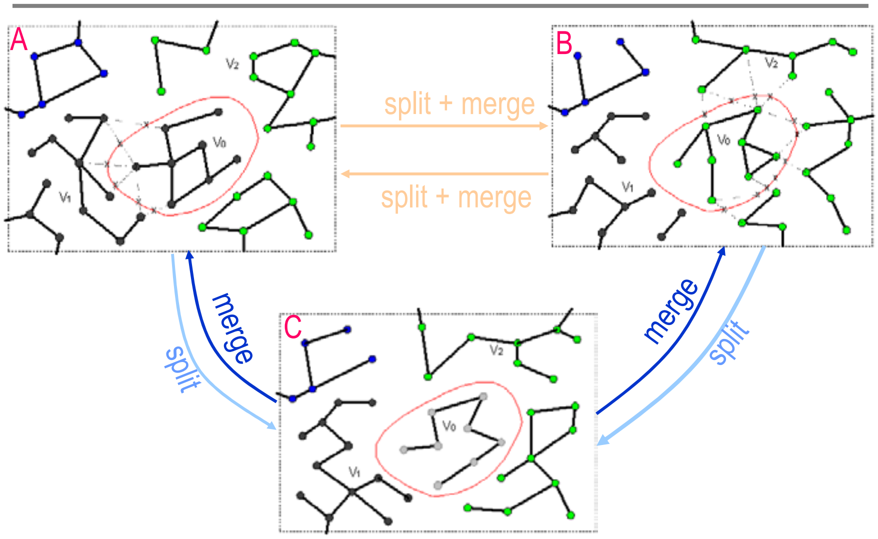
over-segmentation with atomic regions adjacency graph

An example of adjacency graph



Each node could be a pixel, atomic region (super-pixel), feature points,…

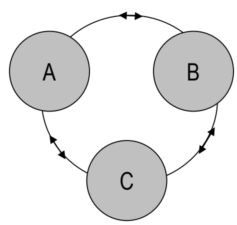
Moves in the partition space



The key issues

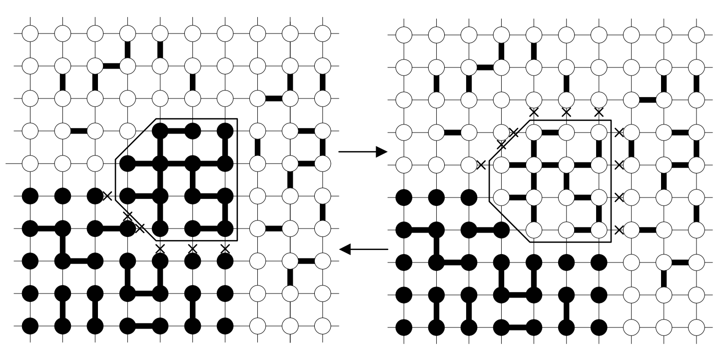
1. How do we select a subset of nodes V_o to flip?
2. How do we maintaining reversibility?
3. How do we ensure detailed balance?

What is the probabilities (ratio) for selecting V_o at two reversible states?



Swendsen-Wang revisited

SW (1987) uses data augmentation and selects Vo by Bernoulli bonds.



Each edge is augmented a bond variable $u_{st} \sim \text{Bernoulli}(\rho\ 1(x_s=x_t)),\ \rho=1-e^{-\beta}$.

Limitations of SW

Although it works well in Ising/Potts models, SW has the following problems that need to be resolved for many vision applications.

1. It is limited to Ising / Potts models.

2. The number of labels L is fixed and known.

3. The forming of clusters is not informed by data, thus it slows down in the presence of an external field (data).

From SW to SW-cut

SW cuts (Barbu and Zhu, ICCV 03, CVPR04) extended SW in three aspects.

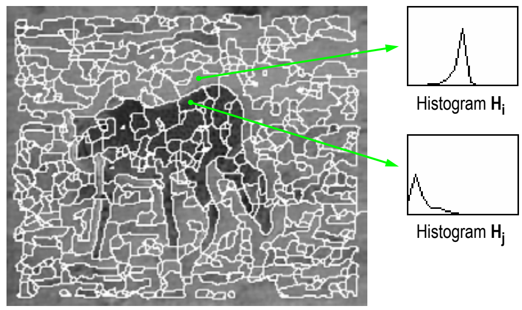
1. Generalize SW to arbitrary probabilities on graphs with variable L.
   It can also be made into a generalized Gibbs sampler which flips a CCP at each step with simple weights on the conditional probabilities.

2. Using discriminative models (data-driven) for the edge probabilities
   The edge probability approaches the marginal posterior probability for how likely two sites s and t belong to the same color (object surface)

3. Hierarchical coloring in a multi-resolution pyramid representation.

Step 1. Computing discriminative edge probability

The edge probability is decided by local features.



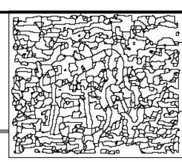
$$q_{st} = e^{-KL(H_i,H_j)/T}$$

It approaches the marginal posterior as the number of features increase.

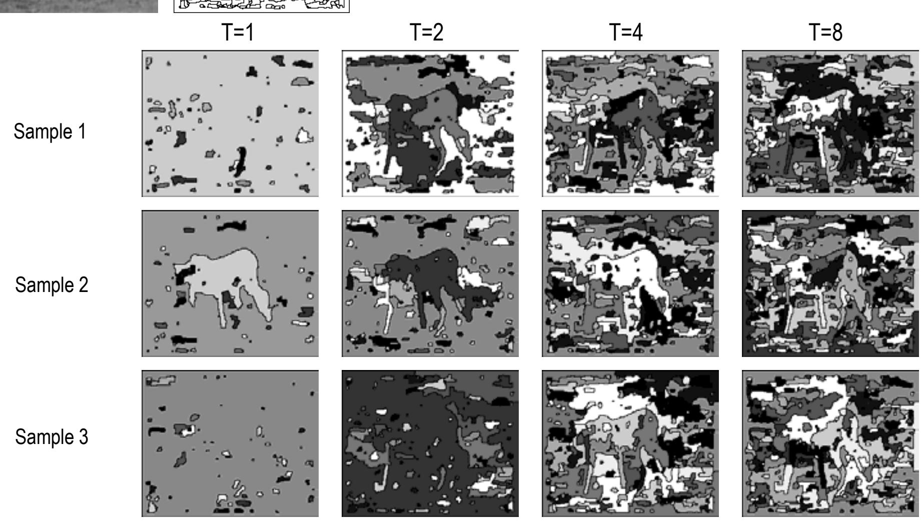
$$q_{st} = q(x_s = x_t | F_s, F_t) \rightarrow p(x_s = x_t | \mathrm{I})$$

$p(x_s = x_t | \mathrm{I})$ is a marginal probability of $\pi(\mathrm{X} \mid \mathrm{I})$

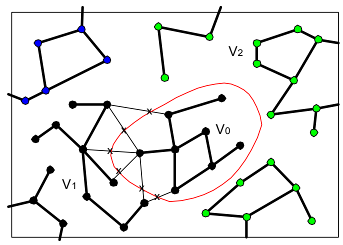
1. Konishi et al 01, Ren et al 04
2. Adaboost, Shapire 00



Clusters (connected components) by flipping the edge probabilities independently



Step 2. Computing SW-Cuts

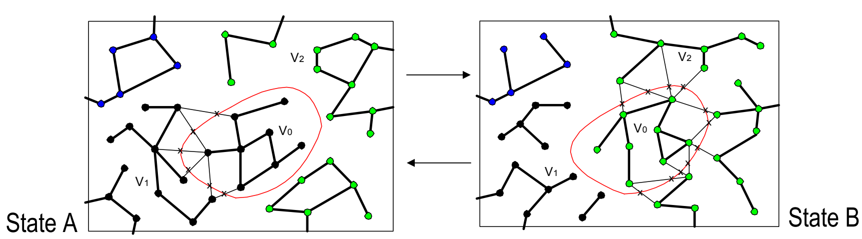


Definition: A *Swendsen-Wang cut* is the set of edges between a cluster (CCP) and other sites of the same color.

$$\mathrm{Cut}(V_0, V_1) = \{<s,t>: s \in V_0, t \in V_1, C_s = C_t\}$$

This is the set of dashed edges marked with crosses.
They must be turned off for V_0 being a CCP.

Probability ratio for a pair of SW-cuts



Theorem. The probability ratio for selecting CCP V_0 at states A and B is

$$\frac{q(V_o|B)}{q(V_o|A)} = \frac{\prod_{<s,t>\in \mathrm{Cut}(V_o,V_2)}(1-q_{st})}{\prod_{<s,t>\in \mathrm{Cut}(V_o,V_1)}(1-q_{st})}$$

(Barbu and Zhu, 2003)

Outline of the proof

We compute the proposal probability ratio:

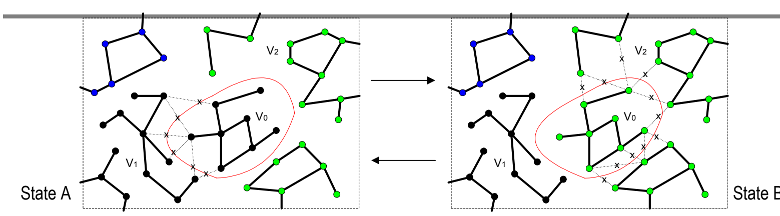
$$\frac{q(B \to A)}{q(A \to B)} = \frac{\sum\limits_{CP \in \Omega_{CP}(B)} q(l|V_o, B)q(V_o|CP)q(CP|B)}{\sum\limits_{CP \in \Omega_{CP}(A)} q(l'|V_o, A)q(V_o|CP)q(CP|A)}$$

All configurations of edges that take state A to B must have all edges of the cut Cut(V_o,V_1-V_o) turned off.

$$q(CP|A) = \prod_{e \in C(V_o, V_l - V_o)} (1 - q_e) \prod_{e \in E_{off}(A,CP) - C(V_o, V_l - V_o)} (1 - q_e) \prod_{e \in E_{on}(A,CP)} q_e$$

$$\frac{q(B \to A)}{q(A \to B)} = \frac{q(l|V_o, B) \prod\limits_{e \in C(V_o, V_l - V_o)} (1 - q_e) \sum\limits_{CP \in \Omega_{CP}(B)} q(V_o|CP) \prod\limits_{e \in E_{off}(A,CP) - C(V_o, V_l - V_o)} (1 - q_e) \prod\limits_{e \in E_{on}(A,CP)} q_e}{q(l'|V_o, A) \prod\limits_{e \in C(V_o, V_l - V_o)} (1 - q_e) \sum\limits_{CP \in \Omega_{CP}(A)} q(V_o|CP) \prod\limits_{e \in E_{off}(A,CP) - C(V_o, V_l - V_o)} (1 - q_e) \prod\limits_{e \in E_{on}(A,CP)} q_e}$$

Outline of the proof



Cancellation of the sums occurs because of the symmetry between states A and B:

- Any CP that takes state A to B is also a CP that takes state B to A

$$\Omega_{CP}(A) = \Omega_{CP}(B)$$

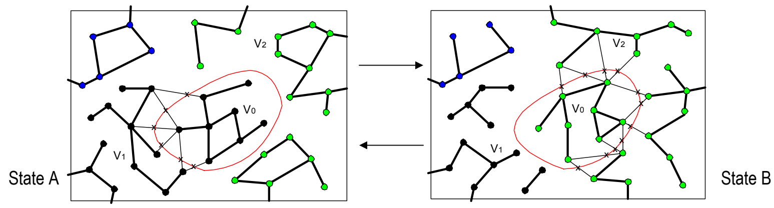
- Any configuration of “on” edges in state A appears in state B and vice versa

$$E_{on}(A, CP) = E_{on}(B, CP)$$

- Any configuration of “off” edges in state A appears in state B

$$E_{off}(A, CP) - C(V_o, V_l - V_o) = E_{off}(B, CP) - C(V_o, V_{l'} - V_o)$$

Step 3. The Metropolis-Hasting Step

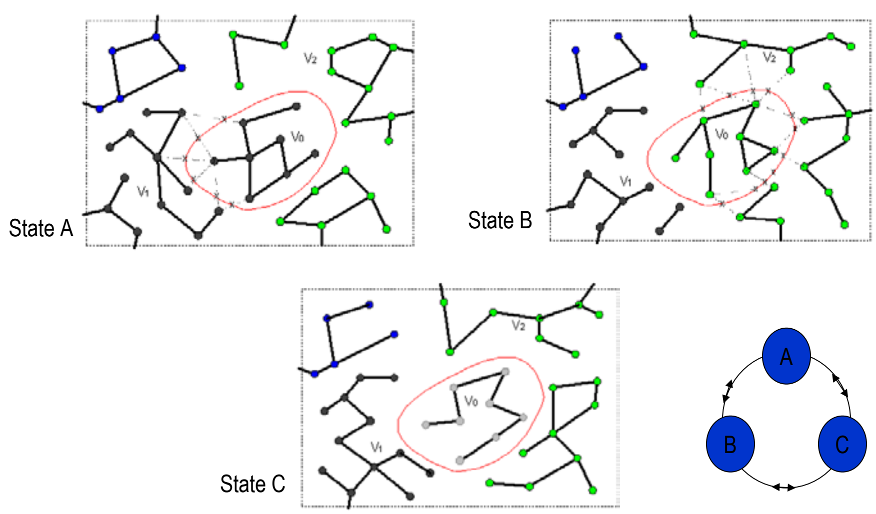


Theorem. The acceptance probability for flipping V_0 is

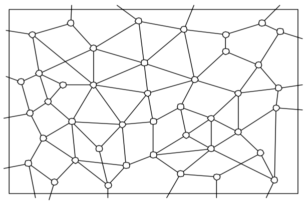
$$
\begin{aligned}
\alpha(A \to B) &= \min\left(1, \frac{q(B \to A)}{q(A \to B)} \cdot \frac{\pi(B)}{\pi(A)}\right) \\
&= \min\left(1, \frac{\prod_{<s,t> \in \mathrm{Cut}(V_o, V_2)}(1 - q_{st})}{\prod_{<s,t> \in \mathrm{Cut}(V_o, V_1)}(1 - q_{st})} \cdot \frac{q(\ell = 1 | V_o, B)}{q(\ell = 2 | V_o, A)} \cdot \frac{\pi(B)}{\pi(A)}\right)
\end{aligned}
$$

results in an ergodic and reversible Markov Chain.

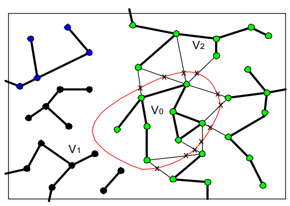
Same conclusion when multiple paths exist



The Swendsen-Wang Cuts algorithm



The initial graph G_o



State B

Swendsen-Wang Cuts: SWC

Input: G_o=<V, E_o>, discriminative probabilities q_e, $e \in E_o$, and generative posterior probability p(W|I).

Output: Samples X~p(X|I).

1. Initialize a graph partition $\pi : G = \cup_{l=1}^{n} G_l$
2. Repeat, for current state A= π
3. Repeat for each subgraph G_l=<V_l, E_l>, l=1,2,...,n in A
4. For $e \in E_l$ turn e="on" with probability q_e.
5. Partition G_l into n_l connected components: g_{li}=<V_{li}, E_{li}>, i=1,...,n_l
6. Collect all the connected components in CP={V_{li}: l=1,...,n, i=1,...,n_l}.
7. Select a connected component $V_0 \in$CP at random
8. Propose to reassign V_0 to a subgraph $G_{l'}$, l' follows a probability q(l'|V_0,A)
9. Accept the move with probability α(A→B).

Acceptance probability can be made always 1

$$\alpha(A \to B) = \min(1, \frac{\prod_{<s,t> \in \mathrm{Cut}(V_o,V_2)}(1-q_{st})}{\prod_{<s,t> \in \mathrm{Cut}(V_o,V_1)}(1-q_{st})} \cdot \frac{q(\ell=1|V_o,B)}{q(\ell=2|V_o,A)} \cdot \frac{\pi(B)}{\pi(A)}) = 1$$

If we select the label probability as

$$q(\ell=1|V_o,B) = \prod_{<s,t> \in \mathrm{Cut}(V_o,V_1)}(1-q_{st})\pi(A)$$

$$q(\ell=2|V_o,A) = \prod_{<s,t> \in \mathrm{Cut}(V_o,V_2)}(1-q_{st})\pi(B)$$

Remark: zero rejection rate may not necessarily be an optimal design.

Another generalized Gibbs sampler

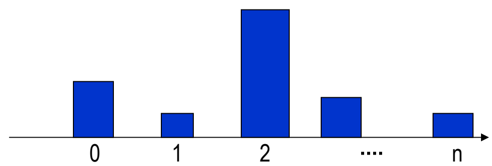
We denote the probabilities on the SW-cuts Cut(V_0, V_k) by weights

$$\omega_k = \prod_{<s,t> \in \mathrm{Cut}(V_o, V_k)} (1 - q_{st}), \quad k = 1, 2, ..., L$$

$$\omega_0 = 1$$

Flip the label of a CCP according to a condition probability weighted by the SW-weights

$$q(\ell = k | V_o) = \omega_k \cdot \pi(\ell(V_o) = k), \quad k = 0, 1, 2, ..., L$$



SW comes as a special case

Consider the reversible moves between states A and B by Metroplis-Hastings:

the proposal probability ratio is:

$$\frac{q(A \to B)}{q(B \to A)} = \frac{(1-q_o)^{|Cut(V_0,V_1)|}}{(1-q_o)^{|Cut(V_0,V_2)|}} = (1-q_o)^{|Cut(V_0,V_1)|-|Cut(V_0,V_2)|}$$

the probability ratio of the two states is:

$$\frac{\pi(A)}{\pi(B)} = \frac{\exp^{-\beta \cdot |Cut(V_0,V_2)|}}{\exp^{-\beta \cdot |Cut(V_0,V_1)|}} = \exp^{\beta \cdot (|Cut(V_0,V_1)|-|Cut(V_0,V_2)|)}$$

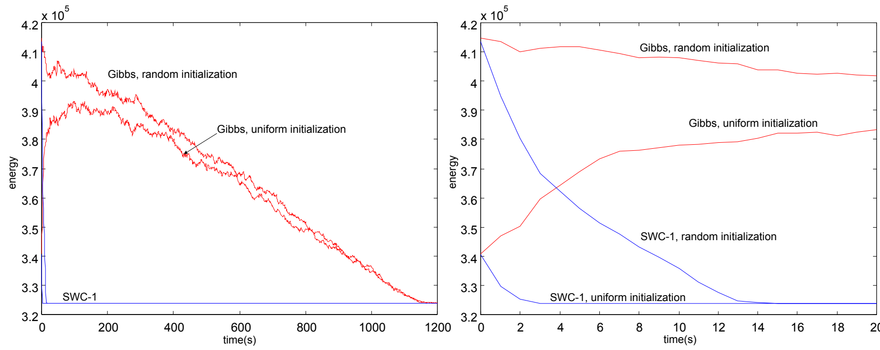
$$\alpha(A \to B) = \min(1, \frac{q(B \to A)}{q(A \to B)} \cdot \frac{\pi(B)}{\pi(A)}) = (\frac{e^{-\beta}}{1-q_o})^{|Cut(V_o,V_1)|-|Cut(V_o,V_2)|}$$

If we choose

$$q_o = 1 - e^{-\beta}$$

Then the acceptance probability is always 1.

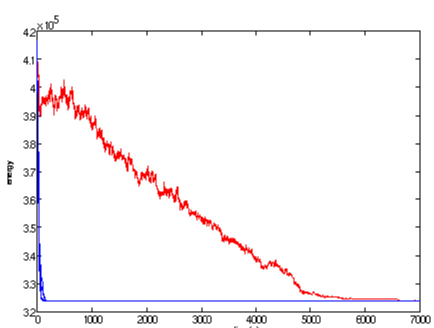
Comparison with Gibbs sampler in CPU time



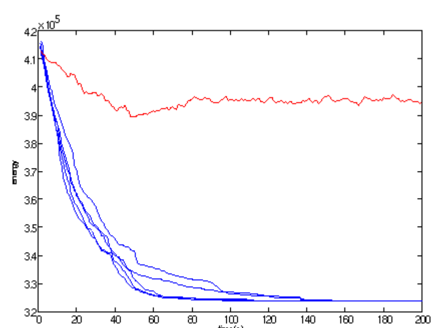
Convergence comparison of SWC-1 and the Gibbs sampler on the cheetah image, starting from a random state or from the state where all nodes have label 0. Right – zoom in view of the first 20 seconds.

Convergence comparison: in seconds

Another example

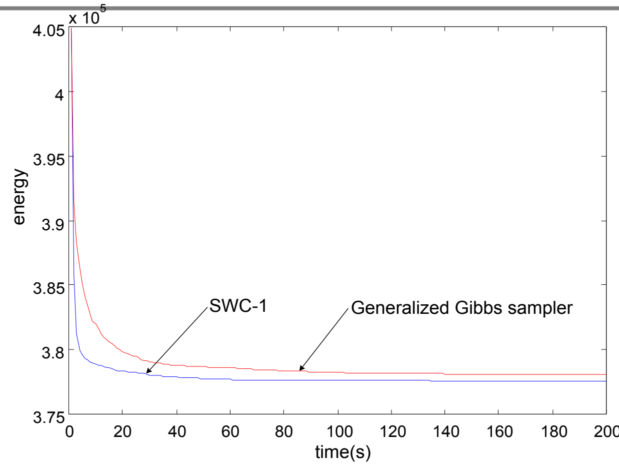


7000 seconds



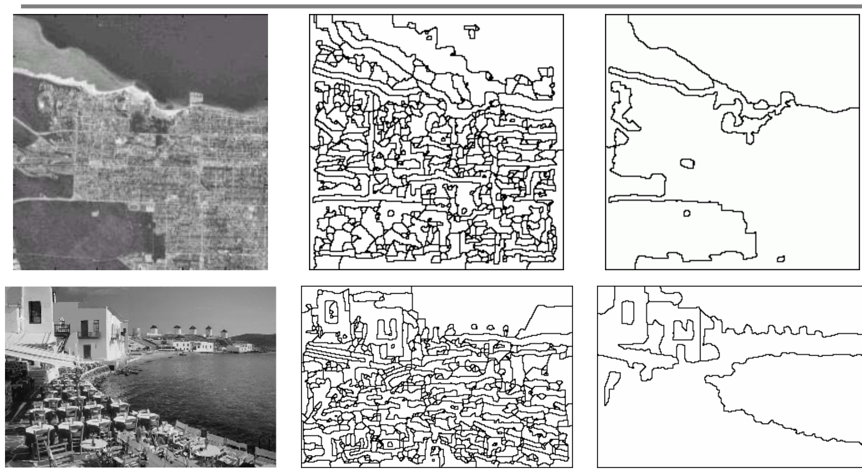
zoom-in view of the first 200 seconds

Comparison



starting from a random state.

Examples of segmentation

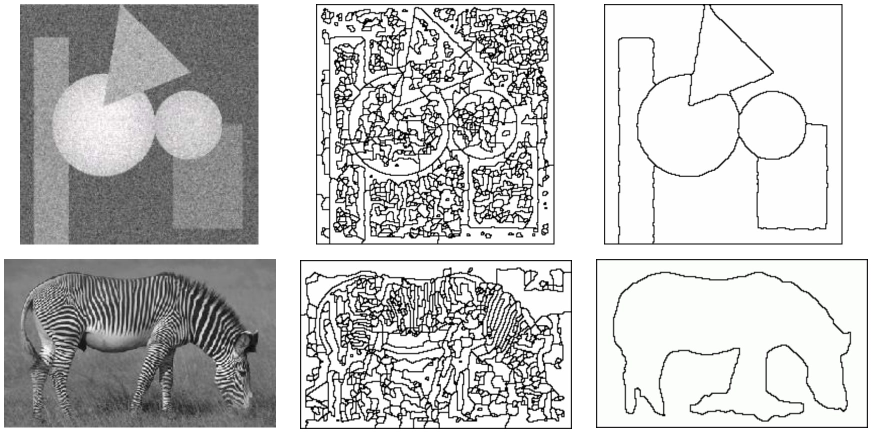


a. input image

b. over-segmentation with atomic regions

c. segmentation result

Examples of segmentation

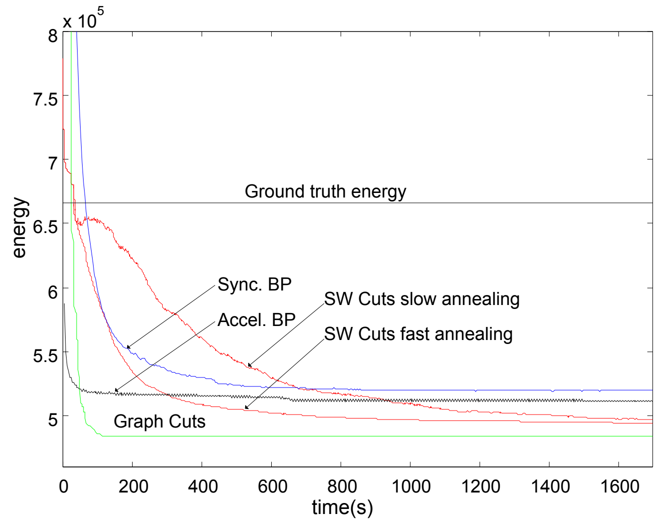


a. input image | b. over-segmentation with atomic regions | c. segmentation

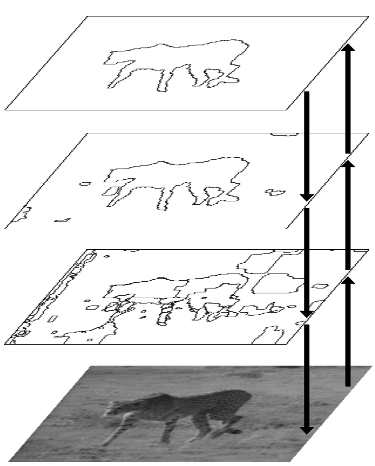
Examples on Stereo Reconstruction

Left image | Ground truth | Segmentation result

Performance comparison with Graph Cuts and Belief propagation on a special (simplified) energy



Hierarchical partition and segmentation



X^2 – Level 2: Intensity regions are grouped into moving objects O_i with motion parameters θ_i

X^1 – Level 1: Atomic regions are grouped into intensity regions R_{ij} of coherent motion with intensity models H_{ij}

X^0 – Level 0: Pixels are grouped into atomic regions r_{ijk} of relatively constant motion and intensity

- motion parameters (u_{ijk}, v_{ijk})
- intensity histogram h_{ijk}

Motion segmentation examples

Input sequence

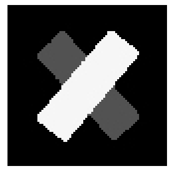
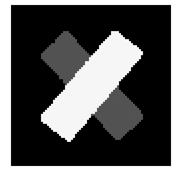


Image Segmentation



Motion Segmentation

Input sequence

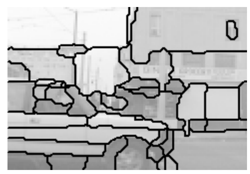
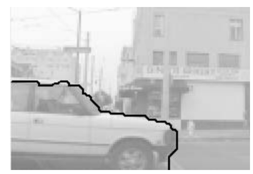
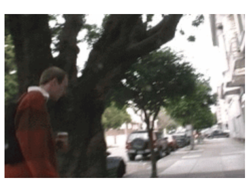


Image Segmentation



Motion Segmentation

Motion segmentation examples



Input sequence

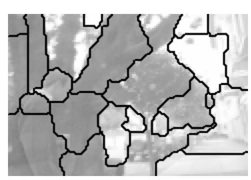
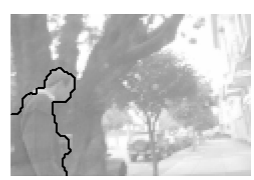


Image Segmentation



Motion Segmentation

Input sequence

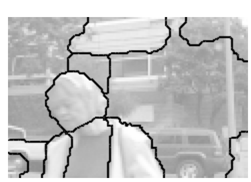
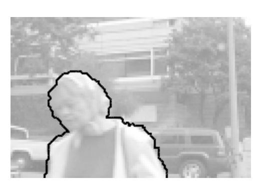


Image Segmentation



Motion Segmentation

Summary

- Generally applicable – allows usage of complex models beyond the scope of the specialized algorithms
- Computationally efficient – performance comparable with the specialized algorithms
- Reversible and ergodic – theoretically guaranteed to eventually find the global optimum