Swendsen-Wang Cuts

Topics

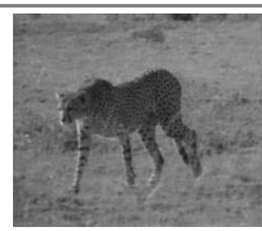
1. Graph partition problem
2. Designing Markov chain in the partition space
3. Data-driven proposal
4. Generalizing SW to arbitrary probability on graphs

Reference: Barbu and Zhu 2003, 2005

Graph partition in computer vision

Image segmentation

- Group pixels based on intensity
- For speed, one can use an over-segmentation by edge detection and edge tracing



input image

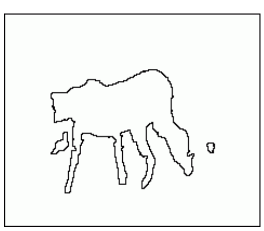
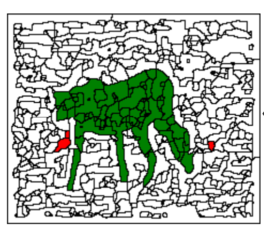
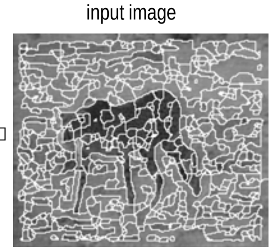


image segmentation result



graph partition (labeling)



over-segmentation with atomic regions adjacency graph

The graph partition problem

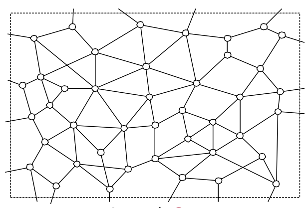
Given:

– A graph G_o=(V,E)

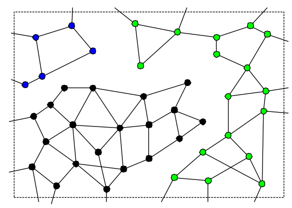
  • Nodes V are image elements
  • Edges E represent spatial relationship or similarity

– A probability p(π(V) |I) or energy E(π(V)) defined on partitions π(V)

Find a partition π(V) that maximizes p(π(V) |I)



A graph G_o

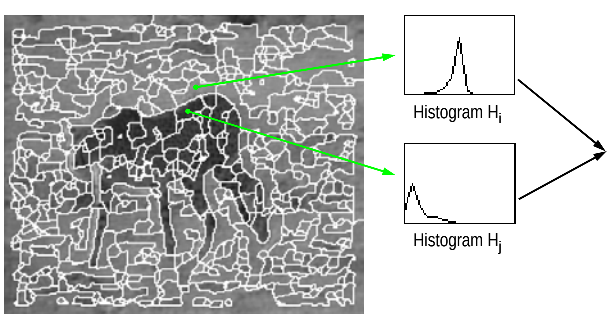


A partition π of the graph

Improving the clustering step

The edge probability q_{ij} is decided by local features F_i,F_j
e.g. image segmentation: KL divergence of histograms

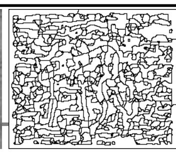
1. Konishi et al 01, Ren et al 04
2. Adaboost, Shapire 00



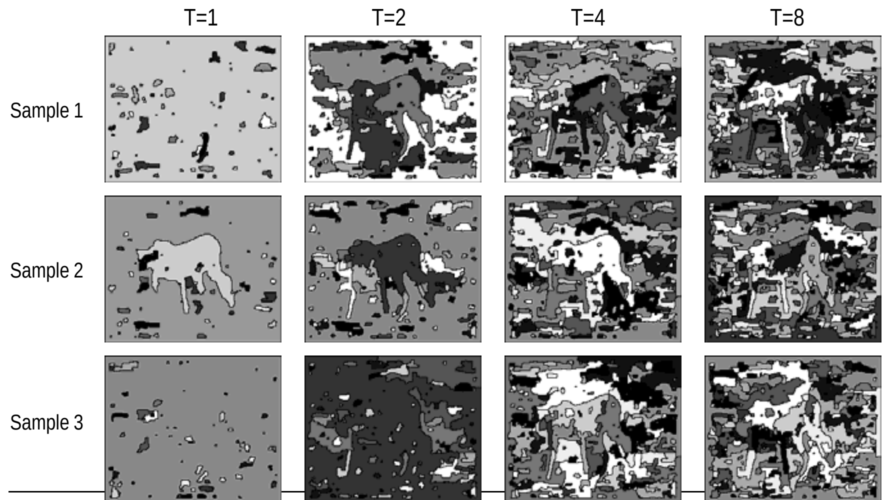
$$q_{ij} = e^{-KL(H_i,H_j)\cdot T}$$

Atomic regions on the input image → Intensity Histograms → Edge weights

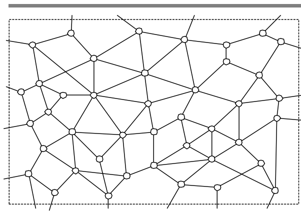
- In general $q_{ij} = q(l_i = l_j | F_i, F_j) \sim p(l_i = l_j | I)$
  $p(l_i = l_j | I)$ is a marginal probability of p(W|I)



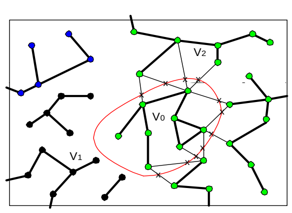
Random connected components



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 W~p(W|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).

SW Cuts: the acceptance probability

Theorem (Metropolis-Hastings) For any proposal probability q(A→B) and probability p(A), if the Markov chain moves by taking samples from q(A → B) which are accepted with probability

$$\alpha(A \to B) = \min(1, \frac{q(B \to A)}{q(A \to B)} \cdot \frac{p(B)}{p(A)})$$

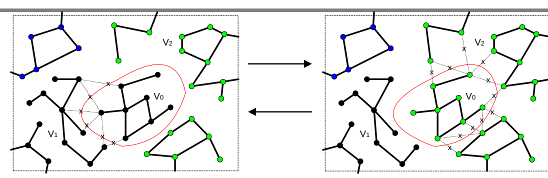
then the Markov chain is reversible with respect to p and has stationary distribution p.

Theorem (Barbu,Zhu '03). The acceptance probability for the Swendsen-Wang Cuts algorithm is

$$\alpha(A \to B) = \min(1, \frac{\prod_{e \in C(V_o, V_{l'} \backslash V_o)} (1-q_e)}{\prod_{e \in C(V_o, V_l \backslash V_o)} (1-q_e)} \cdot \frac{q(l|V_o, B)}{q(l'|V_o, A)} \cdot \frac{p(B)}{p(A)})$$

Outline of the proof

State A



State B

We compute the ratio:

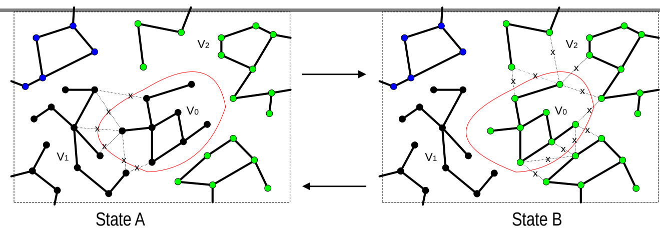
$$\frac{q(B \to A)}{q(A \to B)} = \frac{\sum_{CP \in \Omega_{CP}(B)} q(l|V_o, B) q(V_o|CP) q(CP|B)}{\sum_{CP \in \Omega_{CP}(A)} q(l'|V_o, A) q(V_o|CP) q(CP|A)}$$

All configurations of edges that take state A to B must have all edges of the cut $C(V_0,V_1-V_0)$ turned off.

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

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

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

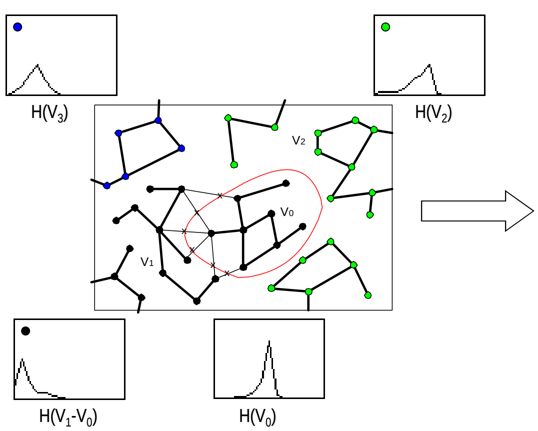
$$E_{on}(A, CP) = E_{on}(B, CP)$$

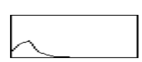
- Modulo the cuts, any configuration of "off" edges in state A appears in state B

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

The reassignment probability

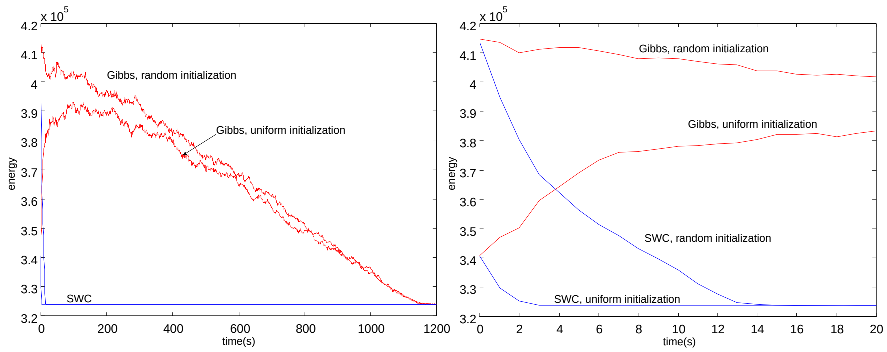
The reassignment probability $q(l\ |V_0,A)$ can also be data-driven.





$$q(l|V_0, A) \propto e^{-KL(H(V_i), H(V_0))}$$

Comparison with the Gibbs sampler



Convergence comparison of SWC 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.

Advantages of the SW Cuts algorithm

Our algorithm bridges the gap between the specialized and generic algorithms:

- 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

A generalized Gibbs sampler

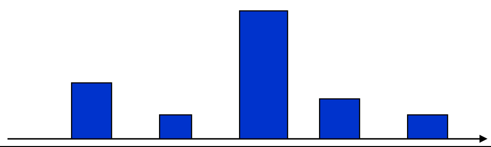
We can obtain acceptance probability

$$\alpha(A \to B) = \min\left(1, \frac{\prod_{e \in C(V_o, V_{l'} - V_o)} (1 - q_e)}{\prod_{e \in C(V_o, V_l - V_o)} (1 - q_e)} \cdot \frac{q(l|V_o, B)}{q(l'|V_o, A)} \cdot \frac{p(B)}{p(A)}\right) = 1$$

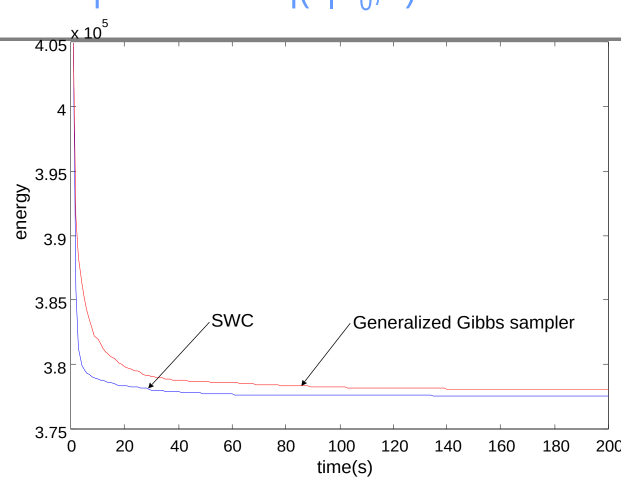
if we select the probability $q(l|V_0,A)$ to reassign V_0 to V_l (obtaining state A_l)

$$q(l|V_o, A) \propto \prod_{e \in C(V_o, V_l - V_o)} (1 - q_e) p(A)$$

Then we basically flip the label of the connected subgraph by a generalized Gibbs sampler.



The importance of $q(l\,|V_0,A)$



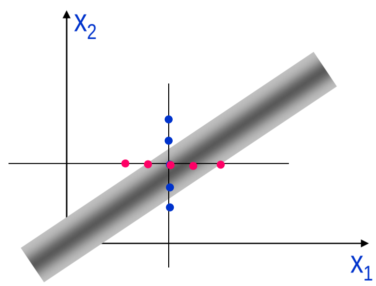
Convergence of SWC with data-driven $q(l\,|V_0,A)$ (blue) and of the generalized Gibbs sampler (red), starting from a random state.

Revisit: A problem with Gibbs sampler

As we discussed in Ch3, in general, Gibbs sampler becomes inefficient when the target probability is defined in a subspace of much lower dimensions due to strong coupling or variables (dimensions).

E.g. for a probability $p(x_1,x_2)$ whose probability mass is focused on a 1D line segment, sampling the two dimensional iteratively is obviously inefficient. i.e. the chain is "jagging".

This is because the two variables are tightly coupled. It is best if we move along the direction of the line.



Revisit: SW can be viewed as a Hit-and-Run

In the low-dimensional space, the Markov chain is not allowed to move in the normal directions (off the manifold) but it can move freely in the tangent directions.

In the SW algorithm, the auxiliary variables U at the current state X probabilistically select the ccp V_o, which corresponds to a randomly selected direction (i.e. the hit direction). The binding probability q_e carries the information from target probability p, as the former is an approximation to the marginal probability of the latter.