Joint Optimization of Segmentation and Appearance Models

David Mandle, Sameep Tandon

April 29, 2013
Overview

1. Recap: Image Segmentation
2. Optimization Strategy
3. Experimental
Recap: Image Segmentation Problem

Segment an image into foreground and background

Figure: Left: Input image. Middle: Segmentation by EM (GrabCut). Right: Segmentation by the method covered today
Recall Grid Structured Markov Random Field:

\[ x_i \in \{0, 1\} \text{ corresponding to foreground/background} \]

Observations \( z_i \). Take to be RGB pixel values

Edge potentials \( \Phi(x_i, z_i) \), \( \Psi(x_i, x_j) \)

Latent variables \( x_i \in \{0, 1\} \) corresponding to foreground/background

Observations \( z_i \). Take to be RGB pixel values

Edge potentials \( \Phi(x_i, z_i) \), \( \Psi(x_i, x_j) \)
Recap: Image Segmentation as Energy Optimization

The Graphical Model encodes the following (unnormalized) probability distribution:

\[ P(x, z) = \prod_{i} \Phi(x_i, z_i) \prod_{i,j} \Psi(x_i, x_j) \]
Recap: Image Segmentation as Energy Optimization

Goal: find $x$ to maximize $P(x, z)$ ($z$ is observed)

Taking logs:

$$E(x, z) = \sum_i \phi(x_i, z_i) + \sum_{i, j} \psi(x_i, x_j)$$

 Unary potential $\phi(x_i, z_i)$ encodes how likely it is for a pixel or patch $y_i$ to belong to segmentation $x_i$.

Pairwise potential $\psi(x_i, x_j)$ encodes neighborhood info about pixel/patch segmentation labels.
Recap: Image Segmentation as Energy Optimization

Goal: find $x$ to maximize $P(x, z)$ ($z$ is observed)

- Taking logs:

$$E(x, z) = \sum_i \phi(x_i, z_i) + \sum_{i,j} \psi(x_i, x_j)$$

Unary potential $\phi(x_i, z_i)$ encodes how likely it is for a pixel or patch $y_i$ to belong to segmentation $x_i$.

Pairwise potential $\psi(x_i, x_j)$ encodes neighborhood info about pixel/patch segmentation labels.
Recap: Image Segmentation as Energy Optimization

Goal: find $x$ to maximize $P(x, z)$ ($z$ is observed)

- Taking logs:

$$E(x, z) = \sum_i \phi(x_i, z_i) + \sum_{i,j} \psi(x_i, x_j)$$

- Unary potential $\phi(x_i, z_i)$ encodes how likely it is for a pixel or patch $y_i$ to belong to segmentation $x_i$. 

David Mandle, Sameep Tandon (Stanford)
Recap: Image Segmentation as Energy Optimization

Goal: find $x$ to maximize $P(x, z)$ ($z$ is observed)

- Taking logs:

$$E(x, z) = \sum_i \phi(x_i, z_i) + \sum_{i,j} \psi(x_i, x_j)$$

- Unary potential $\phi(x_i, z_i)$ encodes how likely it is for a pixel or patch $y_i$ to belong to segmentation $x_i$.

- Pairwise potential $\psi(x_i, x_j)$ encodes neighborhood info about pixel/patch segmentation labels.
Recap: GrabCut Model

- **Unary Potentials:** log of Gaussian Mixture Model
  - But to deal with tractability, we assign each $x_i$ to component $k_i$
    \[
    \phi(x_i, k_i, \theta | z_i) = -\log \pi(x_i, k_i) + \log \mathcal{N}(z_i; \mu(k_i), \Sigma(k_i))
    \]
Recap: GrabCut Model

- **Unary Potentials:** log of Gaussian Mixture Model
  - But to deal with tractability, we assign each $x_i$ to component $k_i$
    \[ \phi(x_i, k_i, \theta | z_i) = -\log \pi(x_i, k_i) + \log N(z_i; \mu(k_i), \Sigma(k_i)) \]

- **Pairwise Potentials:**
  \[ \psi(x_i, x_j | z_i, z_j) = [x_i \neq x_j] \exp(-\beta^{-1}\|z_i - z_j\|^2) \]
  where $\beta = 2 \cdot \text{avg}(\|z_i - z_j\|^2)$
Recap: GrabCut Optimization Strategy

GrabCut EM Algorithm

1. Initialize Mixture Models
Recap: GrabCut Optimization Strategy

GrabCut EM Algorithm

1. Initialize Mixture Models
2. Assign GMM components:

\[ k_i = \arg \min_k \phi(x_i, k, \theta | z_i) \]
Recap: GrabCut Optimization Strategy

GrabCut EM Algorithm

1. Initialize Mixture Models

2. Assign GMM components:

   \[ k_i = \arg \min_k \phi(x_i, k_i, \theta|z_i) \]

3. Get GMM parameters:

   \[ \theta = \arg \min_\theta \sum_i \phi(x_i, k_i, \theta|z_i) \]
Recap: GrabCut Optimization Strategy

GrabCut EM Algorithm

1. Initialize Mixture Models
2. Assign GMM components:

\[ k_i = \arg \min_k \phi(x_i, k_i, \theta|z_i) \]

3. Get GMM parameters:

\[ \theta = \arg \min_{\theta} \sum_i \phi(x_i, k_i, \theta|z_i) \]

4. Perform segmentation using reduction to min-cut:

\[ x = \arg \min_x E(x, z; k, \theta) \]
Recap: GrabCut Optimization Strategy

GrabCut EM Algorithm

1. Initialize Mixture Models
2. Assign GMM components:

   \[ k_i = \arg \min_k \phi(x_i, k_i, \theta | z_i) \]

3. Get GMM parameters:

   \[ \theta = \arg \min_{\theta} \sum_i \phi(x_i, k_i, \theta | z_i) \]

4. Perform segmentation using reduction to min-cut:

   \[ x = \arg \min_x E(x, z; k, \theta) \]

5. Iterate from step 2 until converged
New Model

Let’s consider a simpler model. This will be useful soon

- Unary terms: Histograms
New Model

Let’s consider a simpler model. This will be useful soon

- **Unary terms: Histograms**
  - $K$ bins, $b_i$ is bin of pixel $z_i$
Let’s consider a simpler model. This will be useful soon

- **Unary terms: Histograms**
  - \( K \) bins, \( b_i \) is bin of pixel \( z_i \)
  - \( \theta^0, \theta^1 \in [0,1]^K \) represent color models (distributions) over foreground/background
Let’s consider a simpler model. This will be useful soon

- **Unary terms: Histograms**
  - \( K \) bins, \( b_i \) is bin of pixel \( z_i \)
  - \( \theta^0, \theta^1 \in [0, 1]^K \) represent color models (distributions) over foreground/background
  - 
    \[
    \phi(x_i, b_i, \theta) = - \log \theta^{x_i}_{b_i}
    \]
New Model

Let’s consider a simpler model. This will be useful soon

- **Unary terms: Histograms**
  - $K$ bins, $b_i$ is bin of pixel $z_i$
  - $\theta^0, \theta^1 \in [0,1]^K$ represent color models (distributions) over foreground/background

  $$\phi(x_i, b_i, \theta) = -\log \theta_{b_i}^{x_i}$$

- **Pairwise Potentials**

  $$\psi(x_i, x_j) = w_{ij}|x_i - x_j|$$

  We will define $w_{ij}$ later; for now, consider pairwise equal to Grabcut
Let’s consider a simpler model. This will be useful soon

- **Unary terms: Histograms**
  - $K$ bins, $b_i$ is bin of pixel $z_i$
  - $\theta^0, \theta^1 \in [0, 1]^K$ represent color models (distributions) over foreground/background
  
  $$\phi(x_i, b_i, \theta) = -\log \theta^x_{b_i}$$

- **Pairwise Potentials**
  
  $$\psi(x_i, x_j) = w_{ij}|x_i - x_j|$$

  We will define $w_{ij}$ later; for now, consider pairwise equal to Grabcut

- **Total Energy:**

  $$E(x, \theta^0, \theta^1) = \sum_{p \in V} -\log P(z_p|\theta^{x_p}) + \sum_{(p,q) \in N} w_{pq}|x_p - x_q|$$

  $$P(z_p|\theta^{x_p}) = \theta^{x_p}_{b_p}$$
EM under new model

1. Initialize histograms $\theta^0, \theta^1$. 
EM under new model

1. Initialize histograms $\theta^0, \theta^1$.
2. Fix $\theta$. Perform segmentation using reduction to min-cut:
   \[ x = \arg \min_x E(x, \theta^0, \theta^1) \]
EM under new model

1. Initialize histograms $\theta^0, \theta^1$.
2. Fix $\theta$. Perform segmentation using reduction to min-cut:
   \[
   x = \underset{x}{\operatorname{arg\,min}} E(x, \theta^0, \theta^1)
   \]
3. Fix $x$. Compute $\theta^0, \theta^1$ (via standard parameter fitting).
EM under new model

1. Initialize histograms $\theta^0, \theta^1$.
2. Fix $\theta$. Perform segmentation using reduction to min-cut:
   \[ x = \arg \min_x E(x, \theta^0, \theta^1) \]
3. Fix $x$. Compute $\theta^0, \theta^1$ (via standard parameter fitting).
4. Iterate from step 2 until converged
Optimization

Goal: Minimize Energy

\[ \min_x E(x) \]

\[ E(x) = \sum_k h_k(n^1_k) + \sum_{(p,q) \in N} w_{pq} |x_p - x_q| + h(n^1) \]

where \( n^1_k = \sum_{p \in V_k} x_p \) and \( n_1 = \sum_{p \in V} x_p \)

This is hard!

But efficient strategies for optimizing \( E^1(x) \) and \( E^2(x) \) separately

David Mandle, Sameep Tandon (Stanford)
Goal: Minimize Energy

\[ \min_x E(x) \]

\[ E(x) = \sum_k h_k(n_k^1) + \sum_{(p,q) \in N} w_{pq}|x_p - x_q| + h(n_1^1) \]

\[ n_k^1 = \sum_{p \in V_k} x_p \quad \text{and} \quad n_1 = \sum_{p \in V} x_p \]

- This is hard!
Goal: Minimize Energy

$$\min_x E(x)$$

$$E(x) = \sum_k h_k(n^1_k) + \sum_{(p,q)\in N} w_{pq}|x_p - x_q| + h(n^1)$$

where $n^1_k = \sum_{p\in V_k} x_p$ and $n_1 = \sum_{p\in V} x_p$

- This is hard!
- But efficient strategies for optimizing $E^1(x)$ and $E^2(x)$ separately
Consider an optimization of the form

$$\min_x f_1(x) + f_2(x)$$
Consider an optimization of the form

$$\min_x f_1(x) + f_2(x)$$

where optimizing $f(x) = f_1(x) + f_2(x)$ is hard.
Consider an optimization of the form

$$\min_x f_1(x) + f_2(x)$$

where optimizing $f(x) = f_1(x) + f_2(x)$ is hard

But $\min_x f_1(x)$ and $\min_x f_2(x)$ are easy problems
Consider an optimization of the form

$$\min_x f_1(x) + f_2(x)$$

where optimizing $f(x) = f_1(x) + f_2(x)$ is hard.

But $\min_x f_1(x)$ and $\min_x f_2(x)$ are easy problems.

Dual Decomposition idea: Optimize $f_1(x)$ and $f_2(x)$ separately and combine in a principled way.
Optimization via Dual Decomposition

- Original Problem:

\[ \min_x f_1(x) + f_2(x) \]
Optimization via Dual Decomposition

- **Original Problem:**
  \[ \min_x f_1(x) + f_2(x) \]

- **Introduce local variables:**
  \[ \min_{x_1, x_2} f_1(x_1) + f_2(x_2) \]
  \[ \text{s.t} \ x_1 = x_2 \]
Optimization via Dual Decomposition

Original Problem:

\[ \min_x f_1(x) + f_2(x) \]

Introduce local variables:

\[ \min_{x_1,x_2} f_1(x_1) + f_2(x_2) \]

\[ \text{s.t } x_1 = x_2 \]

Equivalent Problem:

\[ \min_{x_1,x_2} f_1(x_1) + f_2(x_2) \]

\[ \text{s.t } x_2 - x_1 = 0 \]
Optimization via Dual Decomposition

- Primal Problem:

\[
\begin{align*}
\min_{x_1, x_2} & \quad f_1(x_1) + f_2(x_2) \\
\text{s.t} & \quad x_2 - x_1 = 0
\end{align*}
\]
Optimization via Dual Decomposition

- **Primal Problem:**

  \[
  \min_{x_1, x_2} f_1(x_1) + f_2(x_2) \\
  \text{s.t.} \quad x_2 - x_1 = 0
  \]

- **Lagrangian Dual:**

  \[
  g(y) = \min_{x_1, x_2} f_1(x_1) + f_2(x_2) + y^T(x_2 - x_1)
  \]
Optimization via Dual Decomposition

- Primal Problem:
  \[
  \min_{x_1, x_2} f_1(x_1) + f_2(x_2) \\
  \text{s.t} \quad x_2 - x_1 = 0
  \]

- Lagrangian Dual:
  \[
  g(y) = \min_{x_1, x_2} f_1(x_1) + f_2(x_2) + y^T(x_2 - x_1)
  \]

- Decompose Lagrangian Dual:
  \[
  g(y) = (\min_{x_1} f_1(x_1) - y^T x_1) + (\min_{x_2} f_2(x_2) + y^T x_2)
  \]
Optimization via Dual Decomposition

\[
g(y) = (\min_{x_1} f_1(x_1) - y^T x_1) + (\min_{x_2} f_2(x_2) + y^T x_2)
\]

- For all \( y \), \( g(y) \) is a lower bound on optimal of the primal problem.
Optimization via Dual Decomposition

\[ g(y) = (\min_{x_1} f_1(x_1) - y^T x_1) + (\min_{x_2} f_2(x_2) + y^T x_2) \]

\[ g_1(y) \]
\[ g_2(y) \]

- For all \( y \), \( g(y) \) is a lower bound on optimal of the primal problem.
- Maximize \( g(y) \) w.r.t. \( y \) to get the tightest bound
Optimization via Dual Decomposition

\[ g(y) = \left( \min_{x_1} f_1(x_1) - y^T x_1 \right) + \left( \min_{x_2} f_2(x_2) + y^T x_2 \right) \]

For all \( y \), \( g(y) \) is a lower bound on optimal of the primal problem.

Maximize \( g(y) \) w.r.t. \( y \) to get the tightest bound

Further, \( g(y) \) is concave in \( y \). Many techniques for concave maximization (subgradient ascent, etc).
Optimization via Dual Decomposition

\[
g(y) = \left( \min_{x_1} f_1(x_1) - y^T x_1 \right) + \left( \min_{x_2} f_2(x_2) + y^T x_2 \right)
\]

- For all \( y \), \( g(y) \) is a lower bound on optimal of the primal problem.
- Maximize \( g(y) \) w.r.t. \( y \) to get the tightest bound.
- Further, \( g(y) \) is concave in \( y \). Many techniques for concave maximization (subgradient ascent, etc).
- Ideally have fast optimization strategies for \( g_1(y) \) and \( g_2(y) \).
Back to our Segmentation problem

- Energy

\[ E(x) = \sum_k h_k(n^1_k) + \sum_{(p,q) \in N} w_{pq}|x_p - x_q| + h(n^1) \]

where \( n^1_k = \sum_{p \in V_k} x_p \) and \( n_1 = \sum_{p \in V} x_p \)
Back to our Segmentation problem

- **Energy**

\[
E(x) = \sum_k h_k(n^1_k) + \sum_{(p,q) \in N} w_{pq}|x_p - x_q| + h(n^1)
\]

where \( n^1_k = \sum_{p \in V_k} x_p \) and \( n^1 = \sum_{p \in V} x_p \)

- **Group terms**

\[
E(x) = \sum_k h_k(n^1_k) + \sum_{(p,q) \in N} w_{pq}|x_p - x_q| + h(n^1)
\]

\( E_1(x) \)

\( E_2(x) \)
Back to our Segmentation problem

- Energy

\[ E(x) = \sum_k h_k(n^1_k) + \sum_{(p,q)\in N} w_{pq}|x_p - x_q| + h(n^1) \]

where \( n^1_k = \sum_{p\in V_k} x_p \) and \( n^1 = \sum_{p\in V} x_p \)

- Group terms

\[ E(x) = \sum_k h_k(n^1_k) + \sum_{(p,q)\in N} w_{pq}|x_p - x_q| + h(n^1) \]

\[ E_1(x) = \underbrace{E_1(x)}_{E_1(x)} + \underbrace{E_2(x)}_{E_2(x)} \]

- \( E(x) = E_1(x) + E_2(x) \)
Optimization via Dual Decomposition

- Dual Decomposition on $E(x) = E^1(x) + E^2(x)$

$$
\Phi(y) = \min_{x_1} (E^1(x_1) - y^T x_1) + \min_{x_2} (E^2(x_2) + y^T x_2)
$$

$\Phi^1(y)$
$\Phi^2(y)$
Optimization via Dual Decomposition

- Dual Decomposition on $E(x) = E^1(x) + E^2(x)$

  $\Phi(y) = \min_{x_1} (E^1(x_1) - y^T x_1) + \min_{x_2} (E^2(x_2) + y^T x_2)$

  $\Phi^1(y)$ \hspace{2cm} $\Phi^2(y)$

- Maximize $\Phi(y)$ w.r.t. $y$ to get the tightest lower bound.
Dual Decomposition on $E(x) = E^1(x) + E^2(x)$

$$\Phi(y) = \min_{x_1}(E^1(x_1) - y^T x_1) + \min_{x_2}(E^2(x_2) + y^T x_2)$$

- Maximize $\Phi(y)$ w.r.t. $y$ to get the tightest lower bound.
- $\Phi^1(y)$ can be computed efficiently via reduction to min s-t cut.
Optimization via Dual Decomposition

- Dual Decomposition on $E(x) = E^1(x) + E^2(x)$

$$
\Phi(y) = \min_{x_1} (E^1(x_1) - y^T x_1) + \min_{x_2} (E^2(x_2) + y^T x_2)
$$

- $\Phi^1(y)$ can be computed efficiently via reduction to min s-t cut.
- $\Phi^2(y)$ via convex minimization (several strategies)

Maximize $\Phi(y)$ w.r.t. $y$ to get the tightest lower bound.
Optimization via Dual Decomposition

- Dual Decomposition on $E(x) = E^1(x) + E^2(x)$
  $$
  \Phi(y) = \min_{x_1} (E^1(x_1) - y^T x_1) + \min_{x_2} (E^2(x_2) + y^T x_2)
  $$
  $\Phi^1(y)$ and $\Phi^2(y)$

- Maximize $\Phi(y)$ w.r.t. $y$ to get the tightest lower bound.
- $\Phi^1(y)$ can be computed efficiently via reduction to min s-t cut.
- $\Phi^2(y)$ via convex minimization (several strategies)
- You could now use your favorite concave maximization strategy for $\Phi(y)$
Optimization via Dual Decomposition

- Turns out can maximize $\Phi(y)$ over $y$ in polynomial time using a parametric max-flow technique.

Sketch:

▶ Theorem: Given $\Phi_1(y)$ and $\Phi_2(y)$ as described, optimal $y$ has form $y = s_1$

▶ Implication: maximize $\Phi(s_1)$ over all possible $s_1$

$\Phi_1(s_1)$ is piecewise-linear concave $\star |V|$ breakpoints computed by parametric max-flow $\star |V|$ breakpoints can be enumerated from $h(\cdot)$

$\Phi(s_1)$ is piecewise-linear concave with $2|V|$ breakpoints, so finding max over $\Phi(s_1)$ is easy – it's one of the breakpoints.

▶ Given the breakpoint, return the segmentation $x$ with minimum energy from set given by PMF.
Turns out can maximize $\Phi(y)$ over $y$ in polynomial time using a parametric max-flow technique.

Sketch:

- Theorem: Given $\Phi_1(y)$ and $\Phi_2(y)$ as described, optimal $y$ has form $y = s_1$
- Implication: maximize $\Phi(s_1)$ over all possible $s$
- $\Phi_1(s_1)$ is piecewise-linear concave $\forall |V|$ breakpoints computed by parametric max-flow
- Parametric max-flow also returns between 2 and $|V| + 1$ solutions segmentations $x$ per breakpoint
- $\Phi_2(s_1)$ is piecewise-linear concave $\forall |V|$ breakpoints can be enumerated from $h(\cdot)$
- $\Phi(s_1)$ is piecewise-linear concave with 2 $|V|$ breakpoints, so finding max over $\Phi(s_1)$ is easy – it's one of the breakpoints.
- Given the breakpoint, return the segmentation $x$ with minimum energy from set given by PMF.
Optimization via Dual Decomposition

- Turns out can maximize \( \Phi(y) \) over \( y \) in polynomial time using a parametric max-flow technique.

Sketch:

- Theorem: Given \( \Phi^1(y) \) and \( \Phi^2(y) \) as described, optimal \( y \) has form \( y = s1 \)
Optimization via Dual Decomposition

- Turns out can maximize $\Phi(y)$ over $y$ in polynomial time using a parametric max-flow technique.

Sketch:
- Theorem: Given $\Phi^1(y)$ and $\Phi^2(y)$ as described, optimal $y$ has form $y = s1$
- Implication: maximize $\Phi(s1)$ over all possible $s$

$\Phi^1(y)$ is piecewise-linear concave
- $|V|$ breakpoints computed by parametric max-flow
- Parametric max-flow also returns between 2 and $|V|+1$ solutions segmentations

$\Phi^2(y)$ is piecewise-linear concave
- $|V|$ breakpoints can be enumerated from $h(\cdot)$

$\Phi(s1)$ is piecewise-linear concave with $2|V|$ breakpoints, so finding max over $\Phi(s1)$ is easy – it's one of the breakpoints.
- Given the breakpoint, return the segmentation $x$ with minimum energy from set given by PMF.
Optimization via Dual Decomposition

- Turns out can maximize $\Phi(y)$ over $y$ in polynomial time using a parametric max-flow technique.
- Sketch:
  - Theorem: Given $\Phi^1(y)$ and $\Phi^2(y)$ as described, optimal $y$ has form $y = s1$
  - Implication: maximize $\Phi(s1)$ over all possible $s$
  - $\Phi_1(s1)$ is piecewise-linear concave
    - $|V|$ breakpoints computed by parametric max-flow
    - Parametric max-flow also returns between 2 and $|V| + 1$ solutions segmentations $x$ per breakpoint
Optimization via Dual Decomposition

- Turns out can maximize $\Phi(y)$ over $y$ in polynomial time using a parametric max-flow technique.

Sketch:
- Theorem: Given $\Phi^1(y)$ and $\Phi^2(y)$ as described, optimal $y$ has form $y = s1$
- Implication: maximize $\Phi(s1)$ over all possible $s$
- $\Phi_1(s1)$ is piecewise-linear concave
  - $|V|$ breakpoints computed by parametric max-flow
  - Parametric max-flow also returns between 2 and $|V| + 1$ solutions segmentations $x$ per breakpoint
- $\Phi_2(s1)$ is piecewise-linear concave
  - $|V|$ breakpoints can be enumerated from $h(\cdot)$
Optimization via Dual Decomposition

- Turns out can maximize $\Phi(y)$ over $y$ in polynomial time using a parametric max-flow technique.

Sketch:

- **Theorem:** Given $\Phi^1(y)$ and $\Phi^2(y)$ as described, optimal $y$ has form $y = s1$

- **Implication:** maximize $\Phi(s1)$ over all possible $s$

- $\Phi_1(s1)$ is piecewise-linear concave
  - $|V|$ breakpoints computed by parametric max-flow
  - Parametric max-flow also returns between 2 and $|V| + 1$ solutions segmentations $x$ per breakpoint

- $\Phi_2(s1)$ is piecewise-linear concave
  - $|V|$ breakpoints can be enumerated from $h(\cdot)$

- $\Phi(s1)$ is piecewise-linear concave with $2|V|$ breakpoints, so finding max over $\Phi(1s)$ is easy – it’s one of the breakpoints.
Optimization via Dual Decomposition

- Turns out can maximize $\Phi(y)$ over $y$ in polynomial time using a parametric max-flow technique.
- Sketch:
  - Theorem: Given $\Phi^1(y)$ and $\Phi^2(y)$ as described, optimal $y$ has form $y = s1$
  - Implication: maximize $\Phi(s1)$ over all possible $s$
  - $\Phi_1(s1)$ is piecewise-linear concave
    - $|V|$ breakpoints computed by parametric max-flow
    - Parametric max-flow also returns between 2 and $|V| + 1$ solutions segmentations $x$ per breakpoint
  - $\Phi_2(s1)$ is piecewise-linear concave
    - $|V|$ breakpoints can be enumerated from $h(\cdot)$
  - $\Phi(s1)$ is piecewise-linear concave with $2|V|$ breakpoints, so finding max over $\Phi(1s)$ is easy – it’s one of the breakpoints.
  - Given the breakpoint, return the segmentation $x$ with minimum energy from set given by PMF.
References

MRF Review Slides:

Dual Decomposition Slides: Petter Strandmark and Fredrik Kahl. Parallel and Distributed Graph Cuts by Dual Decomposition.
http://www.robots.ox.ac.uk/~vgg/rg/slides/parallelgraphcuts.pdf