Image Segmentation using Dual Distribution Matching

Tatsunori TANIAI¹
Viet-Quoc PHAM²
Keita TAKAHASHI³
Takeshi NAEMURA¹

¹: The University of Tokyo, Japan
²: Toshiba Corporate R&D Center
³: The University of Electro-Communications, Japan
Image segmentation

Applications to image editing and object recognition
Main idea of our method

Simultaneous distribution matching of Fg and Bg. Feasible for approximate input distributions.
Challenges

Energy function for the resulting segmentation $L$:

$$\mathcal{E}(L) = -\lambda_F B\left(\mathbb{P}_F^L, \mathcal{H}_F\right) - \lambda_B B\left(\mathbb{P}_B^L, \mathcal{H}_B\right) + \lambda_S S(L)$$

- **Fg matching term**
  - Output distribution
  - Input distribution

- **Bg matching term**
  - Output distribution
  - Input distribution

- Estimation of optimal weights $\lambda_F$ and $\lambda_B$
- Optimization
  (approximate solutions to NP-hard problems)
Agenda

• Introduction
• Background and related works
• Proposed method
  • Estimation of weights
  • Optimization
• Experiments
• Conclusion
Background & Related works
Binary labeling & energy optimization

The variety of segmentation methods stems from:

- **Smoothness term:** Penalizes discontinuities btw the neighboring pixels.
  \[ E(L) = \mathcal{A}(L) + \lambda \mathcal{S}(L) \]
  
  - **Appearance term:** Measures consistencies btw the segmentation and input color info.

Def. of \( \mathcal{A}(L), \mathcal{S}(L) \) + Optimization of \( E(L) \)

Binary labeling \( L \)  

Assignment of \( F_g \) and \( B_g \) label

Pixel \( p \)

\[ L_p = F \]

\[ L_p = B \]
Consistency measure: local & global

Classify by formulation schemes of

Local measures
\[ \sum_{p \in P} g_p(L_p) \]
Evaluate each pixel independently.

Global measures
\[ g(L_1, L_2, \ldots, L_n) \]
Evaluate all pixels simultaneously.

Appearance term
\[ A(L) \]
(high order)
## Overview of local & global measures

<table>
<thead>
<tr>
<th>Appearance consistency</th>
<th>Local measures</th>
<th>Global measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{A}(L)$</td>
<td>$\sum_{p \in P} g_p(L_p)$</td>
<td>$g(L_1, L_2, \ldots, L_n)$</td>
</tr>
<tr>
<td></td>
<td>Each pixel independently</td>
<td>All pixels simultaneously</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Optimization</th>
<th>Easy</th>
<th>Hard</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Optimal solutions via graph cuts)</td>
<td>(Approximate solutions)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(But require accurate input distributions)</td>
<td></td>
</tr>
</tbody>
</table>
# Overview of local & global measures

<table>
<thead>
<tr>
<th>Local measures</th>
<th>Global measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ \sum_{p \in P} g_p(L_p) ]</td>
<td>[ g(L_1, L_2, \ldots, L_n) ]</td>
</tr>
<tr>
<td>Each pixel independently</td>
<td>All pixels simultaneously</td>
</tr>
<tr>
<td><strong>Appearance consistency</strong></td>
<td><strong>Optimization</strong></td>
</tr>
<tr>
<td>[ A(L) ]</td>
<td><strong>Easy</strong></td>
</tr>
<tr>
<td>(Optimal solutions via graph cuts)</td>
<td>(Approximate solutions)</td>
</tr>
<tr>
<td><strong>Accuracy</strong></td>
<td><strong>Hard</strong></td>
</tr>
<tr>
<td></td>
<td>(But require accurate input distributions)</td>
</tr>
<tr>
<td><strong>Moderate</strong></td>
<td><strong>High</strong></td>
</tr>
</tbody>
</table>
Methods based on local measures

- Interactive graph cuts [Boykov et al. ICCV'01]
- GrabCut [Rother et al. SIGGRAPH'04]

\[
E(L) = \sum_{p \in P} g_p(L_p) + \sum_{(p,q) \in N} h_{pq}(L_p, L_q)
\]

Unary terms

Pairwise terms

Exactly optimized via graph cuts.

With simple models, the min solution prefers shorter boundaries.

⇒ Shortcut across thin structures.
# Overview of local & global measures

<table>
<thead>
<tr>
<th>Appearance consistency</th>
<th>Local measures</th>
<th>Global measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A(L) )</td>
<td>( \sum_{p \in P} g_p(L_p) )</td>
<td>( g(L_1, L_2, \ldots, L_n) )</td>
</tr>
<tr>
<td></td>
<td>Each pixel independently</td>
<td>All pixels simultaneously</td>
</tr>
<tr>
<td>Optimization</td>
<td><strong>Easy</strong> (Optimal solutions via graph cuts)</td>
<td><strong>Hard</strong> (Approximate solutions)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>Moderate</td>
<td><strong>High</strong> (But require accurate input distributions)</td>
</tr>
</tbody>
</table>
Concept of global measures

Fg distribution matching

\[ P^L_F(z) \approx H_F(z) \]

**INPUT**

Fg distribution

**OUTPUT**

Distribution of segmented region

Output a label

Extract a region consistent with an input dist.

[BMGC method, Ayed et al. CVPR'10] etc.
Issues on global measures

True input distribution

\[ H_F(z) \]

Approximate input dist.

\[ H_F(z) \]

Accurate in detail

but require an accurate input distribution.
Motivation of this research

**Goal** Relax the strict assumption for input.

**Prev** Either $F_g$ matching or $B_g$ matching.

**Ours** Simultaneous matching of $F_g$ and $B_g$.

---

**Local measures**

- Appearance consistency

**Global measures**

- Optimisation

**Accuracy**

- Moderate

**High**

(But require accurate input distributions)
Proposed Method
Proposed dual matching method

**INPUT**

- Fg
  - $\mathcal{H}_F(z)$
- Bg
  - $\mathcal{H}_B(z)$

**OUTPUT**

Output a label

Distributions of segmented regions

- Fg matching: $\mathcal{P}_F^L(z) \approx \mathcal{H}_F(z)$
- Bg matching: $\mathcal{P}_B^L(z) \approx \mathcal{H}_B(z)$

Robust even for approximate input distributions
Proposed new energy function

\[
B(p, q) = \sum_{z \in Z} \sqrt{p(z) \cdot q(z)}
\]

Bhattacharyya coef. Similarity measure of probability distributions. (Takes max 1 if \(p = q\))

\[
E(L) = -\lambda_F B(P_F^L, H_F) - \lambda_B B(P_B^L, H_B) + \lambda_S S(L)
\]

Fg matching term

Bg matching term

More likely to capture the true solution by using dual constraints.

Challenges

- Estimation of optimal weights \(\lambda_F, \lambda_B\)
- Optimization
Estimation of weights
Optimal weight: proportional to area size

\[ \tilde{r}_F^L \mathcal{B}(\mathcal{P}_F^L, \mathcal{H}_F) - \tilde{r}_B^L \mathcal{B}(\mathcal{P}_B^L, \mathcal{H}_B) \]

- $\tilde{r}_F^L$: Fg area size ratio
- $\tilde{r}_B^L$: Bg area size ratio

Intuitive explanation

Simultaneous matching of Fg and Bg distributions is deeply associated with matching of the entire image distribution. Therefore, with such weights, we can implicitly enforce appearance consistency in the sense of the entire image distribution.
Estimation of area size ratios

- **Entire image distribution** (known): $\Omega(z)$
- It can be approximated by input distributions:

$$\tilde{\Omega}(z; \eta) = \eta \mathcal{H}_F(z) + (1 - \eta) \mathcal{H}_B(z)$$

Find the parameter $\eta$ that maximizes the similarity:

$$\eta_F = \arg \max \mathcal{B}(\tilde{\Omega}(\eta), \Omega)$$

which is repeatedly updated in the optimization.
Optimization
Upper bound function such that graph cuts can be applied.

\[
\mathcal{E}(L) \leq \mathcal{F}(L) = \sum_{p \in P} g_p(L_p) + \lambda S(L)
\]

Optimize \( \mathcal{E}(L) \) by iteratively minimizing its upper bound functions.
Overview of optimization process

We approximate Fg & Bg terms by upper bound functions such that graph cuts can be applied. A kind of alternating optimization of Fg and Bg.
Experiments
Common setup

Used GrabCut segmentation database
- consists of 50 test images

Experiment 1
## Evaluation of estimated weights

<table>
<thead>
<tr>
<th>Targets</th>
<th>With estimated weights</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[ \mathcal{A}(L) = -\tilde{r}_F^L \mathcal{B}(\mathcal{P}_F^L, \mathcal{H}_F) - \tilde{r}_B^L \mathcal{B}(\mathcal{P}_B^L, \mathcal{H}_B) ]</td>
</tr>
<tr>
<td>With fixed weights</td>
<td></td>
</tr>
<tr>
<td>[ \mathcal{A}_{fixed}(L) = -0.5 \mathcal{B}(\mathcal{P}_F^L, \mathcal{H}_F) - 0.5 \mathcal{B}(\mathcal{P}_B^L, \mathcal{H}_B) ]</td>
<td></td>
</tr>
</tbody>
</table>

### Input variable

<table>
<thead>
<tr>
<th>Input variable</th>
<th>(-10\text{px})</th>
<th>(0\text{px})</th>
<th>(+10\text{px})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(L) : we shrink and expand (L) around the GT border by -10 px to +10 px.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Fixed

\(\mathcal{H}_F, \mathcal{H}_B\): approximate distributions from trimap.
Purpose of this experiment

\[ F_g = \mathcal{H}_F (z) \]
\[ B_g = \mathcal{H}_B (z) \]

Incomplete distributions

Fg matching term

Bg matching term

Appearance function with proper weights.
Evaluation of weight estimation

Average on 50 images.

Energy func. value

$L = \text{Ground Truth}$

Fg matching

Bg matching

Border shifting width [pixels]
Evaluation of weight estimation

$L =$ Ground Truth

Average on 50 images.

Energy func. value

Ideal form that captures the GT

Border shifting width [pixels]
Experiment 2
Image segmentation

- **Target:**
  - (a) **Proposed** (Dual matching of Fg and Bg)
  - (b) With fixed weights (as $\lambda_F = \lambda_B = 0.5$)
  - (c) BMGC (Fg matching)
  - (d) BMGC (Bg matching)
  - (e) Interactive graph cuts (local measure)

- **Input:** approximate distributions from trimap
- **Inference region:** entire image

Infer all pixel labels using only input distributions

Previous
Our method was most accurate
Estimated weights were effective
Proposed

With fixed weights

Fg matching

Bg matching

Local measurea
Our method is relatively accurate in detail.
Experiment 3
Purpose: global vs local measures

- Local measures
- Global measures

Vary accuracy of input distributions

Segmentation accuracy?
Approximate input from block masks

Masks of 5x5 blocks

Reference rate 50%

Only white is referred

Input distributions are made inaccurate by limiting the reference regions using masks

$$F_g = \mathcal{H}_F (z)$$

$$B_g = \mathcal{H}_B (z)$$
Comparison using block masks

Most accurate at mid to high accuracies
Dual matching vs single matching

Far more robust than Fg or Bg matching
Estimated weights vs fixed weights

With fixed weights

Proposed

Valid at almost all accuracies
Local measures outperform global measures if input distributions are extremely inaccurate.
Conclusion
Contributions

- Proposed a dual matching segmentation method of Fg and Bg distributions with
  - estimated weights
  - a valid optimization method

- Compared global & local measure methods by varying accuracy of input distributions
The presentation was organized as:

**Background & Related works**
- Binary labeling problem & Energy optimization
- Local measures
- Global measures

**Proposed methods**
- Dual matching of $F_g$ and $B_g$ distributions
- Estimation of weights
- Optimization

**Experiments**
- Evaluation of estimated weights
- Image segmentation
- Comparison of local & global measures

Any questions?
I hope you speak sloooooooowly 😊
Experiment Extra-1
Video segmentation

- 176x144 size and 382 frames
- Input: Fg&Bg distributions from prev frames
- Inference region: entire image (did not use any motion information)

Targets
- Proposed
- BMGC (Fg matching)
- Interactive Graph Cuts (local measure)

Infer all pixels using only input distributions
Target
Fg and Bg are dynamic

Fg matching
Prone to flickering

Proposed
Most stable

Local measure
Significant error (shoulder)
Smoothness term

- Interactive graph cuts

\[ S(L) = \left[L_p \neq L_q\right] \left( e^{-\kappa |I_p - I_q|^2} + \epsilon \right) / |p - q| \]

- Proposed

\[ S(L) = \left[L_p \neq L_q\right] \left( \frac{1}{1 + |I_P - I_q|^2} + \epsilon \right) / |p - q| \]

Note: [true] = 1, [false] = 0,

- is a term that is added as an extension
Issue on optimizing our energy

Converge to opposite directions.

\[ \Rightarrow \text{Cannot apply simple iterations.} \]