**CO$_3$ for Ultra-fast and Accurate Interactive Image Segmentation**

Yibiao Zhao$^{123}$, Song-Chun Zhu$^{23}$, Siwei Luo$^1$

Beijing Jiaotong University$^1$
Lotus Hill Research Institute$^2$
University of California, Los Angeles$^3$
1. Introduction

Motivation:

This paper presents an ultra-fast and accurate, interactive foreground/background segmentation algorithm for complex images.
1. Introduction

- This paper presents an interactive image segmentation framework which is ultra-fast and accurate. Our framework, termed “CO3”, consists of three components:
  - COupled representation, COnditional model and COnvex inference.
2. Formulation

- This paper presents an interactive image segmentation framework which is ultra-fast and accurate. Our framework, termed “CO3”, consists of three components:
  - COupled representation, COnditional model and COnvex inference.
2. Formulation

- Coupled Image representation

foreground vs. background

on-boundary vs. off-boundary

\[ \Pi = \begin{cases} 
\Lambda = \Lambda^+_R \cup \Lambda^-_R \\
\Lambda = \Lambda^+_B \cup \Lambda^-_B 
\end{cases} \]

\[ s.t. \quad \Lambda^+_B = \partial \Lambda^+_R \]
2. Formulation

- The discriminative probability test introduced is based on the dual region/boundary representation, with the backbone equation being:

\[
E(\Pi) = -\int u \log \frac{P_R(+|I,S)}{P_R(-|I,S)} \, dx - \alpha \int v \log \frac{P_B(+|I)}{P_B(-|I)} \, dx
\]

\[s.t. \quad v = |\nabla u|.

Region probability test \hspace{1cm} Boundary probability test
2. Formulation

The discriminative probability test introduced is based on the dual region/boundary representation, with the backbone equation being:

\[ E(\Pi) = -\int u \log \frac{P_R(\text{+}\mid I,S)}{P_R(\text{-}\mid I,S)} dx - \alpha \int v \log \frac{P_B(\text{+}\mid I)}{P_B(\text{-}\mid I)} dx \]

\[ s.t. \quad v = |\nabla u|. \]

Region characteristic function
\[ u = 1 \text{ (foreground)} \]
\[ u = 0 \text{ (background)} \]

Boundary characteristic function
\[ v = 1 \text{ (on-boundary)} \]
\[ v = 0 \text{ (off-boundary)} \]
2. Formulation

- The discriminative probability test introduced is based on the dual region/boundary representation, with the backbone equation being:

\[
E(\Pi) = -\int u \log \frac{P_R(\pm |I,S)}{P_R(- |I,S)} dx - \alpha \int v \log \frac{P_B(\pm |I)}{P_B(- |I)} dx
\]

s.t. \( v = |\nabla u| \)

**Coupling function**

Region characteristic function
- \( u = 1 \) (foreground)
- \( u = 0 \) (background)

Boundary characteristic function
- \( v = 1 \) (on-boundary)
- \( v = 0 \) (off-boundary)
2. Formulation

- The discriminative probability test introduced is based on the dual region/boundary representation, with the backbone equation being:

\[
E(\Pi) = -\int u \log \frac{P_R(+ | I, S)}{P_R(- | I, S)} dx - \alpha \int v \log \frac{P_B(+ | I)}{P_B(- | I)} dx
\]

\[s.t. \quad v = |\nabla u|.
\]

- The optimization problem is presented as searching for the best answer of two **hypothesis testing** combined:
  - whether a specific pixel belongs to foreground?
  - whether a specific pixel belongs to boundary?
3. Discriminative learning

- This paper presents an interactive image segmentation framework which is ultra-fast and accurate. Our framework, termed “CO3”, consists of three components:
  
  - COupled representation,
  - COnditional model
  - COnvex inference.
3. Discriminative learning

- For a two class problem, we express the probability in an exponential form

\[
P(+ | I) = \frac{1}{Z} \exp \left\{ - \sum_{i} \lambda_i h_i \right\}
\]

- The log-posterior ratio is expressed as a simple linear form,

\[
\log \frac{P(+ | I)}{P(- | I)} = \sum_{i} \lambda_i h_i
\]

- where \( h_i \) is an image feature, which we will refer to as a weak classifier in boosting, and the sufficient statistics in an exponential model. \( \lambda_i \) is its corresponding coefficient.

- Region appearance model.
  - On-line training using Logistic regression version of Adaboost [20] (GentleBoost)

- Boundary appearance model.
  - Off-line training using Logistic regression
3. Discriminative learning

- RGB value
- Histogram of color
- Histogram of gradient
- All features
- EM
- Boosting
3. Discriminative learning

- Incorporating user’s intention

- Foreground pixels (such as red flag) located nearer to foreground scribbles, while the background pixels are closer to background scribbles.

- The spatial prior is defined on the frequency of each label with respect to the Euclidean distance to foreground and background scribbles.

\[
P_R(+(i,j) \mid S) = f(+(i,j) \mid d((i,j),S_+),d((i,j),S_-))
\]
3. Discriminative learning

- the discriminative model
- the prior model
- the model with prior incorporated
4. Convex inference

- This paper presents an interactive image segmentation framework which is ultra-fast and accurate. Our framework, termed “CO3”, consists of three components:
  - COupled representation, COnditional model and COnvex inference.
4. Convex inference

- Revisit the energy function

\[ E(u) = -\int u T_R \, dx - \alpha \int |\nabla u| T_B \, dx, \quad u \in \{0, 1\} \]

- By relaxing the discrete characteristic function \( u \) value into a continues interval \([0, 1]\), the take a simple threshold \( \mu \), we get following theorem:

**Theorem 1:** Call \( u^* \) any minimizer of \( E(u) \) (which is a global minimizer by convexity), then the characteristic functions:

\[
U_R(x) = \begin{cases} 
1 & \text{if } u(x) > \mu, \\
0 & \text{otherwise} \\
\end{cases}, \quad \mu \in [0, 1]
\]

- are also global minimizers of above equation. Minimizing the energy with respect to \( u \) is equivalent to minimize a relaxed formula which takes \( u \) in \([0, 1]\), [25]
4. Convex inference

- Revisit the energy function

\[ E(u) = -\int uT_R dx - \alpha \int |\nabla u|T_B dx, \quad u \in \{0,1\} \]

- By relaxing the discrete characteristic function \( u \) value into a continues interval \([0, 1]\), the take a simple threshold \( \mu \), we get following theorem:

**Theorem 1:** Call \( u^* \) any minimizer of \( E(u) \) (which is a global minimizer by convexity), then the characteristic functions:

\[ U_R(x) = \begin{cases} 
1 & \text{if } u(x) > \mu, \\
0 & \text{otherwise}
\end{cases} \quad \mu \in [0,1] \]

- are also global minimizers of above equation. Minimizing the energy with respect to \( u \) is equivalent to minimize a relaxed formula which takes \( u \) in \([0,1]\), [25]
4. Convex inference

- **Bregman iteration**
  - In order to compute a minimizer of this convex form with relaxed target variable, T. Goldstein et al. [15] introduced a fast and accurate minimization algorithm, Bregman iteration.

  - The key of this algorithm is to enforce “Coupling” of the region term and boundary by solving the following unconstrained problem,

\[
(u^{k+1}, v^{k+1}) = \arg\min_{u \in [0,1], v} |v| T_B + u T_R + \frac{\beta}{2} \left\| v - \nabla u - \varepsilon^k \right\|_2^2
\]

\[
\varepsilon^{k+1} = \varepsilon^k + \nabla u^{k+1} - v^{k+1}
\]

- where \(\varepsilon^k\) is the cumulative error in the iteration \(k\), and it can be mutually cancelled on the next iteration \(k + 1\) by the iteration process.
4. Convex inference

- **Bregman iteration**
  - In order to compute a minimizer of this convex form with relaxed target variable, T. Goldstein et al. [15] introduced a fast and accurate minimization algorithm, Bregman iteration.
  
  - The key of this algorithm is to enforce “Coupling” of the region term and boundary by solving the following unconstrained problem,

    \[
    (u^{k+1}, v^{k+1}) = \underset{u \in [0,1]}{\text{arg min}} T_B + uT_R + \frac{\beta}{2} \left\| v - \nabla u - \epsilon^k \right\|^2_2
    \]

  - **Iteration 1: Minimizing the region function by “Gauss-Seidel” method:**

    \[
    \mu \nabla u = T_R + \mu \text{div}(\epsilon^k - v^k), u \in [0,1]
    \]
4. Convex inference

- **Bregman iteration**
  - In order to compute a minimizer of this convex form with relaxed target variable, T. Goldstein et al. [15] introduced a fast and accurate minimization algorithm, Bregman iteration.

  - The key of this algorithm is to enforce “Coupling” of the region term and boundary by solving the following unconstrained problem,

\[
(u^{k+1}, v^{k+1}) = \arg \min_{u \in [0, 1], v} \| v T_B + u T_R + \frac{\beta}{2} \| v - \nabla u - \varepsilon^k \|_2^2
\]

  - **Iteration 1**: Minimizing the boundary function by “shrinkage” operator:

\[
v_{i,j}^{k+1} = \frac{\nabla u^{k+1} + \varepsilon^k}{|\nabla u^{k+1} + \varepsilon^k|} \max(|\nabla u^{k+1} + \varepsilon^k| - \mu^{-1} T_B, 0)
\]
4. Convex inference

- Solving with multigrid

- The dilemma of high-order features
  subtle localization vs. discriminative power
5. Experimental analysis

- The proposed approach was tested on several challenging images from LHI interactive segmentation benchmark [25], Berkeley segmentation dataset [17], and MSRC dataset [20].

- We use an Intel Core Duo E7300(2.67 GHz) CPU and 2GB RAM PC as the experiment platform.
5. Experimental analysis

- Qualitative evaluation
5. Experimental analysis

- Qualitative evaluation
5. Experimental analysis

- Qualitative evaluation
5. Experimental analysis

- Qualitative evaluation
5. Experimental analysis

- Qualitative evaluation
5. Experimental analysis

- Qualitative evaluation
5. Experimental analysis

- Qualitative evaluation
5. Experimental analysis

- Qualitative evaluation
5. Experimental analysis

- Qualitative evaluation
5. Experimental analysis

- Qualitative evaluation
5. Experimental analysis

- Quantitative evaluation

- Accuracy

Two evaluation criteria of accuracy are applied:

1. Region precision, measures an overlap rate between a result foreground and the corresponding ground truth foreground;
2. Boundary precision, calculates an inverse of Chamfer distance between a result contour and the corresponding ground truth contour.

According to performance comparison in Tab.1, our algorithm outperforms these methods in both region precision and boundary precision by a large margin.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Region precision</td>
<td>0.50</td>
<td>0.56</td>
<td>0.58</td>
<td>0.69</td>
<td>0.73</td>
<td>0.79</td>
</tr>
<tr>
<td>Boundary precision</td>
<td>0.05</td>
<td>0.10</td>
<td>0.11</td>
<td>0.14</td>
<td>0.16</td>
<td>0.21</td>
</tr>
<tr>
<td>Average running time</td>
<td>0.52 s</td>
<td>1.58 s</td>
<td>0.73 s</td>
<td>0.65 s</td>
<td>0.84 s</td>
<td>0.12 s</td>
</tr>
</tbody>
</table>
5. Experimental analysis

- Quantitative evaluation

- Efficiency

![Graph showing efficiency comparison between different methods.](image-url)
5. Experimental analysis

- Quantitative evaluation

- Efficiency

<table>
<thead>
<tr>
<th>Method</th>
<th>Learning</th>
<th>Inference</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Santner et al. [14] (2009)</td>
<td>1.14 s</td>
<td>0.5 s</td>
<td>1.64 s</td>
</tr>
<tr>
<td>Our system</td>
<td>0.015 s</td>
<td>0.016 s</td>
<td>0.031 s</td>
</tr>
</tbody>
</table>

Table 2: The computational cost of our system (with a common CPU) compared with the GPU implementation of Random Forests and TV optimization reported in [14] on a 500×350 image.
5. Experimental analysis

- Algorithm analysis

- Discriminative learning

The contribution of discriminative learning on different image categories of LHI dataset [26]
5. Experimental analysis

- Algorithm analysis

- Spatial prior

![Graph showing precision vs. percentage of scribbles with and without prior]
5. Experimental analysis

- Algorithm analysis

- Inference
  - We set the same input energy terms for these algorithms, and then optimize the energy by different optimization algorithms including Level sets [18], TV optimization of Primal-Dual [4] and Bregman Iteration [15], and discrete graph-based optimization algorithms of Shortest Path (Geodesics) [2], Random Walker [3] and Graph Cuts [1].
5. Experimental analysis

- Algorithm analysis

1a) Graph Cuts
1b) Random Walker
1c) Shortest Path
1d) Level Sets
1e) Primal-Dual
1f) Bregman Iteration (ours)
5. Experimental analysis

- Algorithm analysis

(2a) Graph Cuts
(2b) Random Walker
(2c) Shortest Path
(2d) Level Sets
(2e) Primal-Dual
(2f) Bregman Iteration (ours)
6. Conclusion

- In this paper, we define a coupled representation in the form of a probability ratio test based on both region and boundary information, and combine various discriminative image features in a learning-based conditional model, as well as spatial prior of user scribbles. By relaxing discrete solution of pixel labeling, the energy function can be transformed into a convex form, and thus iteratively solved by a Bregman iteration. We evaluate our algorithm on several datasets, and our system outperforms current approaches with higher precision, and distinct efficiency.

- The project page of CO3 is available at
  
  [http://www.stat.ucla.edu/~ybzho/research/co3/](http://www.stat.ucla.edu/~ybzho/research/co3/).
Thanks for your attention!

ybzhao@ucla.edu