Image Segmentation using Dual Distribution Matching

Tatsunori Taniai[1]
taniai@nae-lab.org
Viet-Quoc Pham[2]
quocviet.pham@toshiba.co.jp
Keita Takahashi[3]
keita.takahashi@ieee.org
Takeshi Naemura[1]
naemura@nae-lab.org

[1] Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan
[2] Toshiba Corporate Research and Development Center, Kanagawa, Japan
[3] Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo, Japan

Abstract

We propose an image segmentation method that divides an image into foreground and background regions when the approximate color distributions for these regions are given. Our approach was inspired by global consistency measures that directly evaluate the similarity between a given distribution and the distribution of the resulting segmentation, which were recently proposed in order to overcome the limitations of traditional pixel-wise (local) consistency measures. The main feature of our proposal is that it uses two (foreground and background) input distributions, which increases the robustness compared to previous studies. To achieve this, we formulated a new mathematical model that describes the consistencies between the two input distributions and the segmentation, in which weighting parameters for the two distribution matching terms are set to be approximately proportional to the size of the foreground and background areas. We call this dual distribution matching (DDM). We also derived an optimization method that uses graph cuts. Experimental results that show the effectiveness of our method and comparisons between local and global consistency measures are presented.

1 Introduction

This paper addresses the problem of foreground-background image segmentation where only the approximate color distributions of the foreground and background regions are given as the input. For example, when a video sequence is processed, such distributions are given from the previous frames. Our aim is to derive a fundamental algorithm with this primitive setup that can find foreground and background regions that are consistent with the given input distributions. The essential question here is how to measure consistencies between the given distributions and the segmentation.

Local measures are widely adopted [5, 10, 11, 14] by virtue of their simplicity. Each pixel is *individually* evaluated to determine how likely it is to belong to the foreground or background based on its color, which can be formulated as unary terms. Typically, a smoothness

constraint is appended as the pairwise terms penalizing discontinuities between the neighboring pixels. Accordingly, the segmentation problem is expressed as the minimization of an energy function that consists of unary and pairwise terms. Such types of functions can be exactly minimized by using graph cuts [6, 8] in polynomial time. However, local-measure-based methods are subject to the shrinking bias (where the border length tends to be shorter), which often results in shortcutting across thin structures [13].

Recent studies [3, 12, 15] have shown that methods based on *global measures* outperform conventional local-measure-based methods. The global consistency is measured by the similarity between a given distribution and the resulting distribution from the extracted region. However, its optimization is complicated because it cannot be formulated as unary terms; in principle, the similarities cannot be calculated exactly until the extracted region is fixed in the target image. Of the global-measure-based methods, Ayed *et al*. [3] formulated the distribution similarity as the Bhattacharyya coefficient [1, 4], and proposed an efficient graph-cut based optimization method *Bhattacharyya Measure Graph Cut* (*BMGC*) that extracts a region consistent with an input distribution. When the input distribution is sufficiently accurate to represent the object region, the BMGC method outperforms previous methods, including active contour models [2, 7, 15, 16]. However, such accurate distributions are usually unknown, and in reality, approximate distributions, which are practically available, often lead to poor results [12]. Pham *et al*. [12] dealt with this problem by assuming that the distribution of the extracted region is similar to the input distribution (the matching condition) but distinct from that of the complementary region (the complementary condition).

We introduce a new distribution matching method named dual distribution matching (DDM) as another approach to increasing the robustness of global measures. In this method, *the consistencies between two input distributions (the foreground and background distributions) and the resulting segmentation are enforced simultaneously*. We not only combine two (foreground and background) matching terms, we also derive *the optimal weighting parameters* for these terms. Additionally, we derive a minimization method for the energy function of DDM. Our method makes it possible to achieve robust and accurate segmentations even with not-so-accurate input distributions, as revealed in the Experiments section.

2 Dual Matching Model of Foreground and Background

2.1 Formulating the Estimation Model

Binary segmentation is formulated as a problem that involves finding a label $\boldsymbol{L}$ for the set of pixels P, as $\boldsymbol{L} = \{L_p | L_p \in \{F,B\}, \forall p \in P\}$, where p denotes a pixel, and F/B denotes the foreground/background label. The foreground/background region is the set of all pixels with F/B and is denoted as $\mathbf{R}_l^{\boldsymbol{L}} = \{p \in P | L_p = l\}$ $(l = F,B)$. The probability distribution of colors (or intensities) within region $\mathbf{R}_l^{\boldsymbol{L}}$ is written as $\mathcal{P}_l^{\boldsymbol{L}}$ $(l = F,B)$.

Let us assume that only the approximate distributions for both the foreground and background are given as $\mathcal{H}_F \simeq \mathcal{P}_F^{\boldsymbol{L}^*}$ and $\mathcal{H}_B \simeq \mathcal{P}_B^{\boldsymbol{L}^*}$, where $\boldsymbol{L}^*$ is the ground truth of $\boldsymbol{L}$. Here, $\boldsymbol{L}^*$ is inferred as the label that minimizes the following energy function $\mathcal{E}(\boldsymbol{L})$:

$$\mathcal{E}(\boldsymbol{L}) = \underbrace{\lambda_F \mathcal{M}_F(\boldsymbol{L})}_{\text{Foreground Matching}} + \underbrace{\lambda_B \mathcal{M}_B(\boldsymbol{L})}_{\text{Background Matching}} + \underbrace{\lambda_S \mathcal{S}(\boldsymbol{L})}_{\text{Smoothness}}, \tag{1}$$

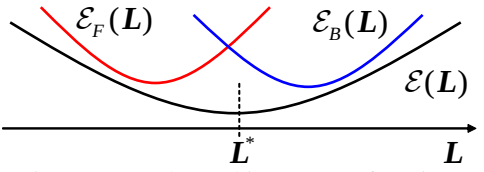


Figure 1: Dual matching energy function.

where $\mathcal{M}_l(\boldsymbol{L})$ is the negative of the distribution similarity measure $\mathcal{B}(,)$:

$$\mathcal{M}_l(\boldsymbol{L}) = -\mathcal{B}\left(\mathcal{P}_l^{\boldsymbol{L}}, \mathcal{H}_l\right) \quad (l = F, B). \tag{2}$$

The $\mathcal{S}(\boldsymbol{L})$ is a smoothness function composed of pairwise discontinuity penalties. This is called *dual distribution matching* or DDM, because both the foreground and background distributions are matched simultaneously.

The distribution $\mathcal{P}_l^{\boldsymbol{L}}$ within region $\mathbf{R}_l^{\boldsymbol{L}}$ is given by the kernel density estimation (KDE) with an arbitrary kernel function K_z:

$$\mathcal{P}_l^{\boldsymbol{L}}(I) = \sum_{p \in \mathbf{R}_l^{\boldsymbol{L}}} K_I(I_p) / |\mathbf{R}_l^{\boldsymbol{L}}|, \tag{3}$$

where $I, I_p \in \mathbb{R}^n$ is a color vector of a pixel p ($n = 3$ with the RGB coordinates), and $|\mathbf{R}| = \sum_{\mathbf{R}} 1$ is the number of pixels within the region $\mathbf{R}$. When the color space is quantized in discrete bins $z \in Z$, and the kernel is the Dirac function $K_z(I_p)$ that takes 1 for $I_p \in z$ and 0 otherwise, $\mathcal{P}_l^{\boldsymbol{L}}(z)$ becomes a histogram. The term $\mathcal{B}(,)$ is the Bhattacharyya coefficient that measures the amount of overlap between two distributions f and g, which takes 1 as the maximum when $f = g$:

$$\mathcal{B}(f, g) = \sum_{z \in Z} \sqrt{f(z) g(z)} \leq 1 \tag{4}$$

With the definitions above, $\mathcal{E}(\boldsymbol{L})$ with $\lambda_B = 0$ or $\lambda_F = 0$, which we define as $\mathcal{E}_F(\boldsymbol{L})$ or $\mathcal{E}_B(\boldsymbol{L})$ respectively, is equivalent to the single distribution matching of the BMGC method [3]. We refer to the BMGC method with $\mathcal{E}_F(\boldsymbol{L})$ or $\mathcal{E}_B(\boldsymbol{L})$ as F-BMGC or B-BMGC. As illustrated in Fig. 1, those methods cannot capture the true solution $\boldsymbol{L}^*$ if the input distribution $\mathcal{H}_F$ or $\mathcal{H}_B$ is inaccurate. In contrast, *our method is more likely to capture the true solution by using both distributions simultaneously*.

2.2 Estimation of Weighting Parameters

The weighting parameters λ_F and λ_B in Eq. (1) should be set so that the global minimum of $\mathcal{E}(\boldsymbol{L})$ captures the true solution $\boldsymbol{L}^*$. We employ a new concept of *matching the entire image distribution* to select the parameters, as it unifies the independent concepts of foreground matching and background matching.

The entire image distribution Ω can be expressed as

$$\Omega(z) = r_F^{\boldsymbol{L}} \mathcal{P}_F^{\boldsymbol{L}}(z) + r_B^{\boldsymbol{L}} \mathcal{P}_B^{\boldsymbol{L}}(z) \tag{5}$$

with an arbitrary label $\boldsymbol{L}$ [1], where $r_l^{\boldsymbol{L}}$ ($l = F, B$) is the ratio of areas defined as $r_l^{\boldsymbol{L}} = |\mathbf{R}_l^{\boldsymbol{L}}| / |\mathbf{R}|$ with the area of the entire image $|\mathbf{R}|$. An approximation of the entire image distribution $\tilde{\Omega}$ is

[1] Note that although the right-hand side of Eq. (5) contains the variable label $\boldsymbol{L}$, the entire image distribution Ω itself is known and constant.

expressed with the input distributions $\mathcal{H}_F$ and $\mathcal{H}_B$ as

$$\tilde{\Omega}(z;\eta) = \eta\,\mathcal{H}_F(z) + (1-\eta)\,\mathcal{H}_B(z),\ \eta \in [0,1]. \tag{6}$$

$\tilde{\Omega}$ coincides with Ω if the input distributions are accurate and η is equal to $|\mathbf{R}_F^{\boldsymbol{L}^*}|/|\mathbf{R}|$.

To maximize the similarity between $\tilde{\Omega}$ and Ω, we seek η_F that minimizes $\mathcal{E}_A(\eta)$:

$$\mathcal{E}_A(\eta) \quad = \quad -\mathcal{B}\left(\Omega(z), \tilde{\Omega}(z;\eta)\right) \tag{7}$$

Here, η_F and $\eta_B = 1-\eta_F$ indicate the contributions $\mathcal{H}_F$ and $\mathcal{H}_B$ have on the entire image distribution. It is also expected that $\eta_F \simeq r_F^{\boldsymbol{L}^*}$ and $\eta_B \simeq r_B^{\boldsymbol{L}^*}$ with the true label $\boldsymbol{L}^*$, because $\Omega(z) = r_F^{\boldsymbol{L}^*}\,\mathcal{P}_F^{\boldsymbol{L}^*}(z) + r_B^{\boldsymbol{L}^*}\,\mathcal{P}_B^{\boldsymbol{L}^*}(z)$ from Eq. (5). Notably, as shown in Appendix A, $\mathcal{E}_A(\eta)$ is upper bounded by a dual matching function:

$$\mathcal{E}_A(\eta) \leq \sqrt{\eta\, r_F^{\boldsymbol{L}}}\,\mathcal{M}_F(\boldsymbol{L}) + \sqrt{(1-\eta)\, r_B^{\boldsymbol{L}}}\,\mathcal{M}_B(\boldsymbol{L}) \tag{8}$$

We rewrite the right-hand side of Eq. (8) as $\mathcal{D}(\boldsymbol{L};\eta)$. This inequality shows that *the energy $\mathcal{E}_A(\eta)$ that measures the correctness of the estimated entire image distribution is closely associated with the energy $\mathcal{D}(\boldsymbol{L};\eta)$ that matches the foreground and background distributions individually*. In summary, we have $\mathcal{E}_A(\eta_F) \leq \mathcal{E}_A(\eta) \leq \mathcal{D}(\boldsymbol{L};\eta)$; the lower bound of $\mathcal{D}(\boldsymbol{L};\eta)$ is minimized to $\mathcal{E}_A(\eta_F)$ when $\eta = \eta_F$, and $\mathcal{D}(\boldsymbol{L};\eta_F)$ is minimized to $\mathcal{E}_A(\eta_F)$ if one of the following two sufficient conditions is satisfied for each $z \in Z$.

$$\mathcal{P}_F^{\boldsymbol{L}}(z) = \mathcal{H}_F(z),\ \mathcal{P}_B^{\boldsymbol{L}}(z) = \mathcal{H}_B(z),\ r_F^{\boldsymbol{L}} = \eta_F \quad \text{(the matching condition)} \tag{9}$$

$$\mathcal{P}_F^{\boldsymbol{L}}(z)\cdot\mathcal{H}_B(z) = \mathcal{P}_B^{\boldsymbol{L}}(z)\cdot\mathcal{H}_F(z) = 0 \quad \text{(the complementary condition)} \tag{10}$$

The matching condition is satisfied when the input distributions ($\mathcal{H}_F$ and $\mathcal{H}_B$) coincide with the distributions computed from the estimated label ($\mathcal{P}_F^{\boldsymbol{L}}$ and $\mathcal{P}_B^{\boldsymbol{L}}$), and η_F and $r_F^{\boldsymbol{L}}$ are the same. The complementary condition is satisfied when the intersection between the estimated foreground distribution and the given background distributions is zero, and vice versa. Both conditions are often used as the constraints for image segmentation [12].

Consequently, we use $\mathcal{D}(\boldsymbol{L};\eta_F)$ as the first two terms in Eq. (1) to obtain:

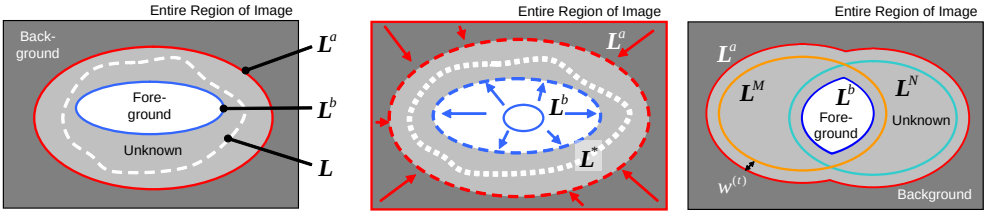
$$\mathcal{E}(\boldsymbol{L}) = \underbrace{\sqrt{\eta_F\, r_F^{\boldsymbol{L}}}\,\mathcal{M}_F(\boldsymbol{L}) + \sqrt{\eta_B\, r_B^{\boldsymbol{L}}}\,\mathcal{M}_B(\boldsymbol{L})}_{\text{appearance term } \mathcal{A}(\boldsymbol{L})} + \underbrace{\lambda\left(\sqrt{\eta_F\, r_F^{\boldsymbol{L}}} + \sqrt{\eta_B\, r_B^{\boldsymbol{L}}}\right)}_{\lambda_S}\mathcal{S}(\boldsymbol{L}), \tag{11}$$

where not only $\mathcal{M}_l(\boldsymbol{L})$ but also the weighting parameters depend on the label $\boldsymbol{L}$. The above definition of λ_S reflects the fact that the absolute size $|\mathcal{A}(\boldsymbol{L})|$ is upper bounded by $\sqrt{\eta_F\, r_F^{\boldsymbol{L}}} + \sqrt{\eta_B\, r_B^{\boldsymbol{L}}}$. The physical meaning of the derived weighting parameters in Eq. (11) is intuitive: *$\mathcal{M}_F(\boldsymbol{L})$ and $\mathcal{M}_B(\boldsymbol{L})$ should be weighted in proportion to the size of the foreground and background areas*. Use of this energy function is equivalent to *not only matching the foreground and background distributions simultaneously, but also inferring the entire image distribution*.

3 Optimizing the Energy Function $\mathcal{E}(\boldsymbol{L})$

3.1 Auxiliary Function

The function $\mathcal{E}(\boldsymbol{L})$ of Eq. (11) is not expressed by the sum of unary and pairwise terms, so it cannot be directly optimized using graph cuts [6, 8]. We derived auxiliary functions (see



(a) Relation between $\boldsymbol{L}$ and auxiliary labels $\boldsymbol{L}^a$ and $\boldsymbol{L}^b$

(b) Updating auxiliary labels around true label $\boldsymbol{L}^*$

(c) Updating auxiliary labels with $\boldsymbol{L}^M$ and $\boldsymbol{L}^N$

Figure 2: Auxiliary labels. Closed lines (including dashed ones) indicate borderlines of labels, where inner area is foreground and outer area is background.

Fig. 1 in [9] for the concept illustration.) that comprise unary terms and pairwise terms to optimize $\mathcal{E}(\boldsymbol{L})$. Our inspiration for this was the BMGC method [3]. An auxiliary function $g(\boldsymbol{L},\boldsymbol{L}^c)$ (note that the parameters are not interchangeable) of $f(\boldsymbol{L})$ satisfies

$$f(\boldsymbol{L}) \leq g(\boldsymbol{L},\boldsymbol{L}^c), \quad f(\boldsymbol{L}) = g(\boldsymbol{L},\boldsymbol{L}), \tag{12}$$

where $\boldsymbol{L}^c$ is a fixed label called an auxiliary label [3] that satisfies $\mathbf{R}_l^{\boldsymbol{L}^c} \supseteq \mathbf{R}_l^{\boldsymbol{L}}$. The function $f(\boldsymbol{L}^{(\tau)})$ does not increase throughout the following iterations:

$$\boldsymbol{L}^{(\tau+1)} = \arg\min_{\boldsymbol{L}} g(\boldsymbol{L},\boldsymbol{L}^{(\tau)}), \tag{13}$$

where the previous result $\boldsymbol{L}^{(\tau)}$ is used as the auxiliary label. This is proved by Eq. (12) and (13) as $f(\boldsymbol{L}^{(\tau+1)}) \leq g(\boldsymbol{L}^{(\tau+1)},\boldsymbol{L}^{(\tau)}) \leq g(\boldsymbol{L}^{(\tau)},\boldsymbol{L}^{(\tau)}) = f(\boldsymbol{L}^{(\tau)})$.

3.2 Giving an Upper Bound Function to Energy Function $\mathcal{E}(\boldsymbol{L})$

According to the BMGC method [3], the function $\mathcal{G}_l(\boldsymbol{L},\boldsymbol{L}^c,\varepsilon)$ $(l = F,B)$ given by Eq. (14) is expressed with unary terms and is also an auxiliary function of $\mathcal{M}_l(\boldsymbol{L})$ when $\varepsilon = 0$.

$$\mathcal{G}_l(\boldsymbol{L},\boldsymbol{L}^c,\varepsilon) = (1-\varepsilon)\sum_{p\in\mathbf{R}_l^{\boldsymbol{L}}} \frac{\mathcal{M}_l(\boldsymbol{L}^c)}{|\mathbf{R}_l^{\boldsymbol{L}^c}|} + \sum_{p\in\mathbf{R}_{\bar{l}}^{\boldsymbol{L}}} \frac{\delta_{L_p^c=l}}{|\mathbf{R}_l^{\boldsymbol{L}^c}|}\left(\mathcal{M}_l(\boldsymbol{L}^c) + \sum_{z\in Z} K_z(I_p)\sqrt{\frac{\mathcal{H}_l(z)}{\mathcal{P}_l^{\boldsymbol{L}^c}(z)}}\right) \tag{14}$$

Here, $\boldsymbol{L}^c$ is a fixed label called an auxiliary label that satisfies $\mathbf{R}_l^{\boldsymbol{L}^c} \supseteq \mathbf{R}_l^{\boldsymbol{L}}$, and $\varepsilon \in [0,1]$. The value $\bar{l}$ is the inverse of l, and $\delta_{(\text{true})} = 1$ and $\delta_{(\text{false})} = 0$.

We incorporate $\mathcal{G}_l(\boldsymbol{L},\boldsymbol{L}^c,\varepsilon)$ into the weighted matching terms in Eq. (11), resulting in an upper bound for each weighted matching term:

$$\sqrt{\eta_F\, r_F^{\boldsymbol{L}}}\,\mathcal{M}_F(\boldsymbol{L}) \leq \sqrt{\eta_F\, r_F^{\boldsymbol{L}^a}}\,\mathcal{G}_F(\boldsymbol{L},\boldsymbol{L}^a,\alpha), \quad \sqrt{\eta_B\, r_B^{\boldsymbol{L}}}\,\mathcal{M}_B(\boldsymbol{L}) \leq \sqrt{\eta_B\, r_B^{\boldsymbol{L}^b}}\,\mathcal{G}_B(\boldsymbol{L},\boldsymbol{L}^b,\beta) \tag{15}$$

with auxiliary labels $\boldsymbol{L}^a$ and $\boldsymbol{L}^b$ that satisfy $\mathbf{R}_F^{\boldsymbol{L}^a} \supseteq \mathbf{R}_F^{\boldsymbol{L}}$ and $\mathbf{R}_B^{\boldsymbol{L}^b} \supseteq \mathbf{R}_B^{\boldsymbol{L}}$ (refer to Fig. 2a for a visualization), and $\alpha,\beta \in [0,1]$. Consequently, the upper bound $\hat{\mathcal{E}}(\boldsymbol{L},\boldsymbol{L}^a,\boldsymbol{L}^b,\alpha,\beta)$ of the entire energy $\mathcal{E}(\boldsymbol{L})$ of Eq. (11) is expressed by unary terms and pairwise terms as

$$\hat{\mathcal{E}}(\boldsymbol{L},\boldsymbol{L}^a,\boldsymbol{L}^b,\alpha,\beta) = \sqrt{\eta_F\, r_F^{\boldsymbol{L}^a}}\,\mathcal{G}_F(\boldsymbol{L},\boldsymbol{L}^a,\alpha) + \sqrt{\eta_B\, r_B^{\boldsymbol{L}^b}}\,\mathcal{G}_B(\boldsymbol{L},\boldsymbol{L}^b,\beta) + \hat{\lambda}_S\mathcal{S}(\boldsymbol{L}) \geq \mathcal{E}(\boldsymbol{L}), \tag{16}$$

$$\text{where} \quad \hat{\lambda}_S = \lambda\left(\sqrt{\eta_F\, r_F^{\boldsymbol{L}^a}} + \sqrt{\eta_B\, r_B^{\boldsymbol{L}^b}}\right) \geq \lambda_S. \tag{17}$$

Process 1 Optimize $\mathcal{E}(\boldsymbol{L})$

initialize

- Initialize auxiliary labels: $L_p^a = F$, $L_p^b = B \quad \forall p \in P$
- Estimate the ratio of areas: $\eta_F = \arg\min \mathcal{E}_A(\eta)$
- Initialize result buffer labels: $\boldsymbol{L}^{a*} = \boldsymbol{L}^{b*} = \boldsymbol{L}^{\text{out}} = \text{null}$
- Obtain a result using a local-measure-based method: $\boldsymbol{L}^{\text{local}} = \text{standard\_graphcut}(\mathcal{H}_F, \mathcal{H}_B)$

for $t = 1$ **to** T **do**

1: Update foreground auxiliary label by **Process 2**: Input:$\boldsymbol{L}^a$, $\boldsymbol{L}^b$, $\boldsymbol{L}^{a*}$,Output:$\boldsymbol{L}^{a*}$
2: Update background auxiliary label by **Process 3**: Input:$\boldsymbol{L}^a$, $\boldsymbol{L}^b$, $\boldsymbol{L}^{b*}$, Output:$\boldsymbol{L}^{b*}$
3: Refine auxiliary labels by **Process 4**: Input:$\boldsymbol{L}^{a*}$, $\boldsymbol{L}^{b*}$, $\boldsymbol{L}^{\text{out}}$, $\boldsymbol{L}^{\text{local}}$, Output:$\boldsymbol{L}^a$, $\boldsymbol{L}^b$
4: Segmentation using Eq. (16): $\boldsymbol{L}^{(t)} = \arg\min \hat{\mathcal{E}}(\boldsymbol{L}, \boldsymbol{L}^a, \boldsymbol{L}^b, \alpha_0, \beta_0)$
5: Compare and update the results: $\boldsymbol{L}^{\text{out}} = \arg\min \mathcal{A}(\boldsymbol{L})$ with $\boldsymbol{L} \in \{\boldsymbol{L}^{(t)}, \boldsymbol{L}^{\text{out}}\}$
6: Refine auxiliary labels by **Process 4**: Input:$\boldsymbol{L}^{a*}$, $\boldsymbol{L}^{b*}$, $\boldsymbol{L}^{\text{out}}$, $\boldsymbol{L}^{\text{local}}$, Output:$\boldsymbol{L}^a$, $\boldsymbol{L}^b$

end for

return $\boldsymbol{L}^{\text{out}}$

Process 2 Update Foreground Auxiliary Label

initialize

- Initialize α and β: $\alpha = \alpha_0$, $\beta = \beta_0$
- Initialize auxiliary labels $\boldsymbol{L}^{(a)}$ and $\boldsymbol{L}^{(b)}$: $\boldsymbol{L}^{(a)} = \boldsymbol{L}^a$, $\boldsymbol{L}^{(b)} = \boldsymbol{L}^b$

repeat

1: Graph cut by Eq. (16): $\boldsymbol{L}^{(\tau)} = \arg\min \hat{\mathcal{E}}(\boldsymbol{L}, \boldsymbol{L}^{(a)}, \boldsymbol{L}^{(b)}, \alpha, \beta)$
2: Update the auxiliary label $\boldsymbol{L}^{(a)}$: $\boldsymbol{L}^{(a)} = \boldsymbol{L}^{(\tau)}$
3: Decrease α: $\alpha = \alpha^\rho \quad (\rho > 1)$

until $\mathcal{A}(\boldsymbol{L}^{(\tau)})$ converges or increases

update Compare and update the result: $\boldsymbol{L}^{a*} = \arg\min \mathcal{A}(\boldsymbol{L})$ with $\boldsymbol{L} \in \{\boldsymbol{L}^{(a)}, \boldsymbol{L}^{a*}\}$

Process 3 Update Background Auxiliary Label

This process is equivalent to **Process 2** except for a few differences: at **repeat** 2: update auxiliary label by $\boldsymbol{L}^{(b)} = \boldsymbol{L}^{(\tau)}$; at **repeat** 3: decrease β by $\beta = \beta^\gamma \quad (\gamma > 1)$; and at **update**, compare and update the result by $\boldsymbol{L}^{b*} = \arg\min \mathcal{A}(\boldsymbol{L})$ with $\boldsymbol{L} \in \{\boldsymbol{L}^{(b)}, \boldsymbol{L}^{b*}\}$.

We can optimize $\mathcal{E}(\boldsymbol{L})$ by repeatedly minimizing $\hat{\mathcal{E}}$. Strictly speaking, $\hat{\mathcal{E}}(\boldsymbol{L}, \boldsymbol{L}^a, \boldsymbol{L}^b, \alpha, \beta)$ does not work as an auxiliary function[2], although $\mathcal{E}(\boldsymbol{L}) = \hat{\mathcal{E}}(\boldsymbol{L}, \boldsymbol{L}^a, \boldsymbol{L}^b, \alpha, \beta)$ is satisfied when $\alpha = \beta = 0$ and $\boldsymbol{L} = \boldsymbol{L}^a = \boldsymbol{L}^b$. However, when the two given auxiliary labels are sufficiently similar, *i.e.* $\boldsymbol{L}^a \simeq \boldsymbol{L}^b$, $\hat{\mathcal{E}}(\boldsymbol{L}, \boldsymbol{L}^a, \boldsymbol{L}^b, \alpha, \beta)$ is expected to approximate $\mathcal{E}(\boldsymbol{L})$. Therefore, we optimize $\mathcal{E}(\boldsymbol{L})$ by alternately updating $\boldsymbol{L}^a$ and $\boldsymbol{L}^b$ to make them converge to the true label $\boldsymbol{L}^*$ from the outer and inner sides (refer to Fig. 2b).

3.3 Optimization Algorithm

Process 1 shows an optimization algorithm of the energy function $\mathcal{E}(\boldsymbol{L})$ of Eq. (11). The process basically consists of "Update Foreground and Background Auxiliary Labels" and "Segmentation" with "Refine Auxiliary Labels" between them. We prepared $\boldsymbol{L}^{a*}$, $\boldsymbol{L}^{b*}$, and

[2] The upper bound $\hat{\mathcal{E}}(\boldsymbol{L}, \boldsymbol{L}^a, \boldsymbol{L}^b, \alpha, \beta)$ of $\mathcal{E}(\boldsymbol{L})$ has two fixed auxiliary labels, $\boldsymbol{L}^a$ and $\boldsymbol{L}^b$, which bind the border of variable label $\boldsymbol{L}$ from both sides (Fig. 2a). For this, in a simple iteration process following Eq. (13): $\boldsymbol{L}^{(\tau+1)} = \arg\min_{\boldsymbol{L}} \hat{\mathcal{E}}(\boldsymbol{L}, \boldsymbol{L}^{(\tau)}, \boldsymbol{L}^{(\tau)}, \alpha, \beta)$, the variable $\boldsymbol{L}$ is fixed at $\boldsymbol{L}^{(\tau)}$; therefore, iterative optimization is not achieved.

Process 4 Refine Auxiliary Labels

Select two labels ($\boldsymbol{L}^M$ and $\boldsymbol{L}^N$) from $\boldsymbol{L}^{a*}$, $\boldsymbol{L}^{b*}$, $\boldsymbol{L}^{\text{out}}$, and $\boldsymbol{L}^{\text{local}}$ that most minimize $\mathcal{A}(\boldsymbol{L})$. As Fig. 2c illustrates, the auxiliary labels $\boldsymbol{L}^a$ and $\boldsymbol{L}^b$ are updated so that $\mathbf{R}_B^{\boldsymbol{L}^a} \supset (\mathbf{R}_B^{\boldsymbol{L}^M} \cup \mathbf{R}_B^{\boldsymbol{L}^N})$ and $\mathbf{R}_F^{\boldsymbol{L}^b} \subset (\mathbf{R}_F^{\boldsymbol{L}^M} \cap \mathbf{R}_F^{\boldsymbol{L}^N})$ with a margin $w^{(t)}$.

$\boldsymbol{L}^{\text{out}}$ to store the results. Also, $\boldsymbol{L}^{\text{local}}$ is a label obtained by using a method based on local appearance measures such as interactive graph cuts [3]. **Process 2** and **Process 3** are used to implement the iteration of Eq. (13) in two ways (foreground and background) in order to reduce $\mathcal{E}(\boldsymbol{L}^{(\tau)})$, which is a direct expansion of the BMGC optimization method. **Process 4** is an implementation of the process of conservatively updating the auxiliary labels $\boldsymbol{L}^a$ and $\boldsymbol{L}^b$. The entire algorithm looks rather complicated since it adaptively selects reasonable labels from the ones produced during the process. This process might be simplified in the future.

4 Experiments

We used the GrabCut database [14], which is composed of 50 test images with their ground truth labels (two values of foreground and background boundaries) and lasso-trimap labels (three values of foreground, background, and unknown boundaries). We used three-dimensional $64 \times 64 \times 64$ uniform histograms in the RGB space for the distribution expression.

4.1 Evaluation of Weighting Parameters

We compared the proposed appearance term $\mathcal{A}(\boldsymbol{L})$ with a fixed weight version where $\mathcal{A}_{\text{fixed}}(\boldsymbol{L}) = 0.5 \cdot \mathcal{M}_F(\boldsymbol{L}) + 0.5 \cdot \mathcal{M}_B(\boldsymbol{L})$ in order to evaluate the proposed weighting parameters in Eq. (11). Approximate input distributions $\mathcal{H}_F$ and $\mathcal{H}_B$ were produced from the foreground and background regions in the lasso-trimap included in the database. We produced a label $\boldsymbol{G}^{(h)}$ for each image by expanding (shrinking if $h < 0$) the foreground of the ground truth label $\boldsymbol{G}$ ($= \boldsymbol{L}^*$) by h pixels. Then, we substituted $\boldsymbol{L} = \boldsymbol{G}^{(h)}$ ($-10 \leq h \leq 10$) into the four energies $\mathcal{M}_F(\boldsymbol{L})$, $\mathcal{M}_B(\boldsymbol{L})$, $\mathcal{A}(\boldsymbol{L})$, and $\mathcal{A}_{\text{fixed}}(\boldsymbol{L})$.

Figure 3 plots these four energies averaged over the 50 images. The input distributions were produced from lasso-trimaps, and therefore, the foreground terms $\mathcal{M}_F(\boldsymbol{L})$ and background terms $\mathcal{M}_B(\boldsymbol{L})$ tend to be minimized by relatively smaller regions than the ground truth. The function $\mathcal{A}_{\text{fixed}}(\boldsymbol{L})$ is excessively influenced by the foreground term. This is because the foreground is smaller than the background for most images in the database, but both terms are equally weighted in $\mathcal{A}_{\text{fixed}}(\boldsymbol{L})$. In contrast, the function $\mathcal{A}(\boldsymbol{L})$ draws an ideal curve that is minimized at $h = 0$ *i.e.* $\boldsymbol{L} = \boldsymbol{L}^*$, because $\mathcal{M}_F(\boldsymbol{L})$ and $\mathcal{M}_B(\boldsymbol{L})$ are weighted nearly in proportion to the size of the foreground and background areas.

4.2 Evaluation of Segmentation Accuracy

Experimental Conditions We compared five methods: (a) DDM (proposed), (b) DDM with $\mathcal{A}_{\text{fixed}}(\boldsymbol{L})$ (fixed weighting parameters), (c) F-BMGC [3] (foreground matching), (d) B-BMGC [3] (background matching), and (e) interactive graph cuts [3] (local measure). The segmentation target was the entire image; the ground truth and lasso-trimap labels were used only for creating the input distributions. Accuracy was measured by the average of error

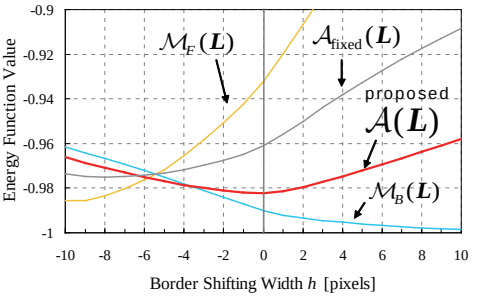


Figure 3: Profiles of $\mathcal{M}_F(\boldsymbol{L})$, $\mathcal{M}_B(\boldsymbol{L})$, $\mathcal{A}(\boldsymbol{L})$, and $\mathcal{A}_{\text{fixed}}(\boldsymbol{L})$, where the border of $\boldsymbol{L}$ is shifted by h pixels from the ground truth. Because the input distributions of $\mathcal{H}_F$ and $\mathcal{H}_B$ are inaccurate, neither the foreground or background term alone captures the true solution, while $\mathcal{A}(\boldsymbol{L})$ with the estimated weighting parameters takes the minimum at $h = 0$ ($\boldsymbol{L} = \boldsymbol{L}^*$).

Method	Results using Parameter 1		Results using Parameter 2	
	EPR (mean±std)	Time [sec]	EPR (mean±std)	Time [sec]
(a) DDM	1.226 ± 0.788 %	2.32	1.345 ± 0.820 %	2.37
(b) DDM (fixed weighting parameters)	1.959 ± 1.279 %	3.05	2.281 ± 1.343 %	2.82
(c) F-BMGC [6]	3.509 ± 2.903 %	1.75	4.635 ± 2.798 %	0.54
(d) B-BMGC [6]	2.032 ± 1.683 %	0.84	2.429 ± 1.974 %	0.49
(e) Interactive graph cuts [6]	1.530 ± 0.958 %	0.23	1.590 ± 1.120 %	0.25

Table 1: Comparison of segmentation accuracy with lasso-trimap distributions, with average accuracy and processing time over 50 images from the GrabCut database [14].

pixel rate (EPR), which is the ratio of the number of misclassified pixels to the number of all pixels. We used two sets of parameters (Parameters 1 and 2) for the smoothness term, which are detailed in Appendix B.

Segmentation with Lasso-Trimap Distributions In this experiment, we used approximate input distributions $\mathcal{H}_F$ and $\mathcal{H}_B$ learned from the foreground and background regions in the lasso-trimap of the database. Table 4.2 lists the EPRs (mean and std), and the processing time per image for each method, and it indicates that our method yielded the best average EPR. When Parameter 1 was used, the proposed method outperformed method (b) for 41 images, method (c) for 46 images, method (d) for 41 images, and method (e) for 39 images, out of the 50 images. Several resulting images are shown in Fig. 4 (see Appendix C for more results). We observe that (a) the proposed method captures well the details of objects thanks to the global appearance measure that evaluates both the foreground and background consistencies, while (e) interactive graph cuts often produces shortcuts through thin structures. It can also be seen that the foregrounds tend to be reduced by (c) F-BMGC and expanded by (d) B-BMGC because the input distributions are taken from the lasso-trimap. (b) DDM with fixed weighting parameters overweights the foreground terms, yielding similar results to (c).

Video Segmentation We applied our method to a video sequence ("carphone") consisting of 176×144 pixels and 382 frames by taking the input distributions from the segmentation result for the previous frame with the first frame manually segmented. The results of the

Figure 4: Segmentation results using lasso-trimap distributions. From left to right, (GT) target image with its ground truth label, and the results of (a) the proposed, (b) fixed weighting parameters, (c) F-BMGC, (d) B-BMGC, and (e) interactive graph cuts.

Figure 5: Video segmentation.

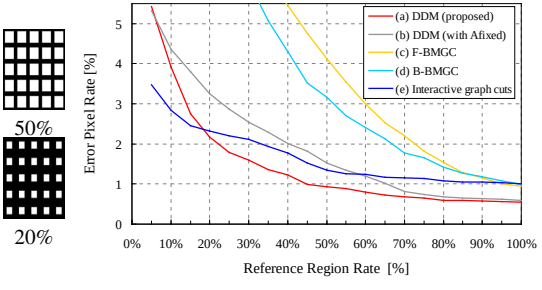


Figure 6: Comparison of consistency measures.

proposed method, F-BMGC, and interactive graph cuts are shown in Fig. 5, from top to bottom (see also Appendix D and the supplementary video). The proposed method produced the most stable results, while F-BMGC was unstable and prone to flickering, and interactive graph cuts produced significant errors around the right shoulder.

Comparative Simulation between Local and Global Consistency Measures Finally, we compared local and global consistency measures while varying the accuracy of the input distributions, which, as far as we know, has yet to be investigated and reported. The input distributions were purposely made inaccurate by limiting the reference region using masks like the ones shown in Fig. 6; we varied the ratio of the masking region (reference rate) from 100% to 5% in 5% increments in order to control the accuracy of the input distributions.

As the graph in Fig. 6 illustrates, (a) our method outperformed the others at high and

medium accuracies (the reference rate of 100%–20%), but (e) interactive graph cuts performed the best at very low accuracies. This is because *global measures enforce the consistency between the input distributions and the segmentation even if they are inaccurate*. We conclude that *global measures are advantageous when combined with highly/moderately accurate distributions*. Meanwhile, the local-measure-based method (e) seems more robust to the inaccuracies of the input distributions. BMGC methods (c) and (d) slightly outperformed (e) interactive graph cuts when the reference rate was 100%–95%, but their performance rapidly worsened as the accuracy of the input distributions decreased. This indicates that using only a single distribution is insufficient unless it is accurate. Also, the proposed method outperformed (b) fixed weighting parameters, which shows the validity of the weighting parameter estimation.

References

[1] F. J. Aherne, N. A. Thacker, and Peter Rockett. The bhattacharyya metric as an absolute similarity measure for frequency coded data. *Kybernetika*, 34(4):363–368, 1998.

[2] Ismail Ben Ayed, Shuo Li, and Ian Ross. A statistical overlap prior for variational image segmentation. *Int. J. Computer Vision*, 85:115–132, 2009.

[3] Ismail Ben Ayed, Hua mei Chen, Kumaradevan Punithakumar, Ian G. Ross, and Shuo Li. Graph cut segmentation with a global constraint: Recovering region distribution via a bound of the bhattacharyya measure. In *Proc. CVPR*, pages 3288–3295, 2010.

[4] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.*, 35:99–109, 1943.

[5] Yuri Boykov and Marie-Pierre Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. In *Proc. ICCV*, pages 105–112, 2001.

[6] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. PAMI*, 26:1124–1137, 2004.

[7] Daniel Freedman and Tao Zhang. Active contours for tracking distributions. *IEEE Trans. Image Processing*, 13(4):518 –526, 2004.

[8] Vladimir Kolmogorov and Ramin Zabih. What energy functions can be minimized via graph cuts? *IEEE Trans. PAMI*, 26(2):147–159, 2004.

[9] Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In *Proc. NIPS*, pages 556–562, 2002.

[10] Yin Li, Jian Sun, Chi-Keung Tan, and Heung-Yeung Shum. Lazy snapping. *ACM Trans. Graph.*, 23:303–308, 2004.

[11] Jiangyu Liu, Jian Sun, and Heung-Yeung Shum. Paint selection. *ACM Trans. Graph.*, 28:1–7, 2009.

[12] Viet-Quoc Pham, Keita Takahashi, and Takeshi Naemura. Foreground-background segmentation using iterated distribution matching. In *Proc. CVPR*, pages 2113–2120, 2011.

[13] Brian L. Price, Bryan Morse, and Scott Cohen. Geodesic graph cut for interactive image segmentation. In *Proc. CVPR*, pages 3288–3295, 2010.

[14] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23:309–314, 2004.

[15] Carsten Rother, Vladimir Kolmogorov, Tom Minka, and Andrew Blake. Cosegmentation of image pairs by histogram matching–incorporating a global constraint into mrfs. In *Proc CVPR*, pages 993–100, 2006.

[16] Tao Zhang and D. Freedman. Improving performance of distribution tracking through background mismatch. *IEEE Trans. PAMI*, 27(2):282 –287, 2005.