Image Segmentation using Dual Distribution Matching (Supplementary)

A A Proof for $\mathcal{E}_A(\eta) \leq \mathcal{D}(\boldsymbol{L};\eta)$

The entire image distribution $\Omega(z)$ and its estimate $\tilde{\Omega}(z;\eta)$ are written as

$$\Omega(z) = \underbrace{r_F^{\boldsymbol{L}} \mathcal{P}_F^{\boldsymbol{L}}(z)}_{f_1} + \underbrace{r_B^{\boldsymbol{L}} \mathcal{P}_B^{\boldsymbol{L}}(z)}_{b_1} \tag{18}$$

$$\tilde{\Omega}(z;\eta) = \underbrace{\eta\, \mathcal{H}_F(z)}_{f_2} + \underbrace{(1-\eta)\, \mathcal{H}_B(z)}_{b_2}. \tag{19}$$

Using these equations, we can derive

$$\mathcal{E}_A(\eta) = -\mathcal{B}\left(\Omega, \tilde{\Omega}(\eta)\right) \tag{20}$$

$$= -\sum_{z\in Z} \sqrt{f_1 f_2 + (f_1 b_2 + b_1 f_2) + b_1 b_2} \tag{21}$$

$$\leq -\sum_{z\in Z} \sqrt{f_1 f_2 + 2\sqrt{f_1 f_2}\sqrt{b_1 b_2} + b_1 b_2} \tag{22}$$

$$= -\sum_{z\in Z} \sqrt{f_1 f_2} - \sum_{z\in Z} \sqrt{b_1 b_2} = \mathcal{D}(\boldsymbol{L};\eta). \tag{23}$$

Therefore, $\mathcal{E}_A(\eta) \leq \mathcal{D}(\boldsymbol{L};\eta)$ is proven. Equation (22) is derived from the relation of the arithmetic and geometric means:

$$\frac{\alpha+\beta}{2} \geq \sqrt{\alpha\beta} \quad \text{(holds equal when} \quad \alpha = \beta). \tag{24}$$

The equality $\mathcal{E}_A(\eta) = \mathcal{D}(\boldsymbol{L};\eta)$ is attained when $f_1 b_2 = b_1 f_2$, *i.e.*

$$r_F^{\boldsymbol{L}} \mathcal{P}_F^{\boldsymbol{L}}(z) \cdot (1-\eta)\, \mathcal{H}_B(z) = r_B^{\boldsymbol{L}} \mathcal{P}_B^{\boldsymbol{L}}(z) \cdot \eta\, \mathcal{H}_F(z), \tag{25}$$

which is satisfied by the conditions in Eq. (9) and (10) when $\eta = \eta_F$.

B Details of the Experimental Conditions

We used a laptop computer with Windows 7 OS, Core i7 2.8GHz CPU (2640M) for mobile and 8GB main memories. The segmentation target was the entire image; the ground truth and lasso-trimap labels were used only for creating the input distributions. 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 [5] (local measure). For method (b), we used the energy function and its upper bound as below

$$\mathcal{E}_{\text{fixed}}(\boldsymbol{L}) = \mathcal{A}_{\text{fixed}}(\boldsymbol{L}) + \lambda \mathcal{S}(\boldsymbol{L})$$

$$\leq 0.5 \cdot \mathcal{G}_F(\boldsymbol{L}, \boldsymbol{L}^a, \alpha) + 0.5 \cdot \mathcal{G}_B(\boldsymbol{L}, \boldsymbol{L}^b, \beta) + \lambda \mathcal{S}(\boldsymbol{L})$$

Table 2: Parameters of Smoothness Term

Method	Parameter 1 (trimap dist.)		Parameter 2 (true dist.)	
	λ	ε	λ	ε
(a) DDM	10^{-3}	8×10^{-4}	2×10^{-4}	10^{-2}
(b) DDM (fixed weighting parameters)	6×10^{-3}	5×10^{-4}	4×10^{-4}	2×10^{-2}
(c) F-BMGC [3]	16×10^{-3}	8×10^{-4}	10^{-4}	5×10^{-2}
(d) B-BMGC [3]	4×10^{-3}	10^{-3}	5×10^{-4}	2×10^{-3}
(e) Interactive graph cuts [5]	50	0	60	2/60

and optimized it similarly to the original DDM method. We used a graph cut algorithm of Boykov and Kolmogorov [6].

For DDM methods (a) and (b), and BMGC methods (c) and (d), we defined $\mathcal{S}(\boldsymbol{L})$ of the smoothness term $\lambda\mathcal{S}(\boldsymbol{L})$ as

$$\mathcal{S}(\boldsymbol{L}) = \sum_{(p,q)\in N} \frac{\delta_{L_p\neq L_q}}{|p-q|}\left(\frac{1}{1+|I_p-I_q|^2}+\varepsilon\right)$$

whereas, for (e) interactive graph cuts, we slightly modified the traditional function by adding an ε term as

$$\mathcal{S}(\boldsymbol{L}) = \sum_{(p,q)\in N} \frac{\delta_{L_p\neq L_q}}{|p-q|}\left(e^{-\kappa|I_p-I_q|^2}+\varepsilon\right),$$

where κ is computed by $\kappa = \left(2E[|I_p-I_q|^2]\right)^{-1}$ for each image, and N is the set of all 8-connected neighboring pixel pairs. We used two sets of parameters of the smoothness terms shown in Tab. 2. Parameter 1 is tuned to maximize the evaluation measure for the input distributions given from lasso-trimaps, and Parameter 2 is tuned for the ground truth distributions.

For the parameters of DDM methods (a) and (b), we used $\alpha_0=\beta_0=0.85$, $\rho=\gamma=1.1$, $T=2$, $w^{(t=1)}=25$, $w^{(t=2)}=5$, and we used the result of (e) for $\boldsymbol{L}^{\text{local}}$.

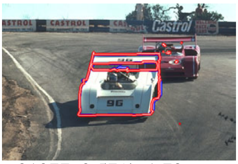
C Segmentation Results with Lasso-Trimap Distributions

All the resulting images obtained with the proposed method using the lasso-trimap distributions and Parameter 1 are presented in the following pages. The caption below each image indicates its file name, EPR, and processing time. Red lines in the figures show the obtained segmentation boundaries, and blue ones show the ground truth boundaries.

D Results of Video Segmentation

See also the supplementary video file carphone_result.mp4 (H.264 format).

The parameters for the proposed method were $\lambda=10^{-2}$, $\varepsilon=2.5\times10^{-3}$, $T=2$, $\{w^{(t)}\}=\{10,2\}$, $\alpha_0=\beta_0=0.85$, and $\rho=\gamma=1.1$; for the F-BMGC method, they were $\lambda=2\times10^{-2}$, and $\varepsilon=2.5\times10^{-3}$; and for interactive graph cuts, $\lambda=1$, and $\varepsilon=5$.

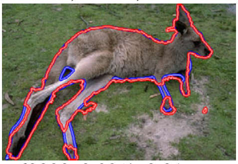


21077, 0.57%, 1.73 sec.

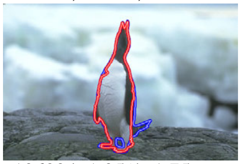

24077, 3.24%, 2.03 sec.


37073, 1.09%, 1.95 sec.


65019, 0.35%, 1.76 sec.

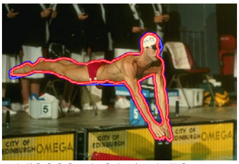

69020, 3.09%, 2.01 sec.


86016, 0.56%, 1.56 sec.


106024, 1.05%, 1.75 sec.


124080, 1.79%, 1.75 sec.


153077, 2.55%, 1.90 sec.

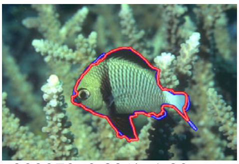


153093, 1.01%, 1.73 sec.

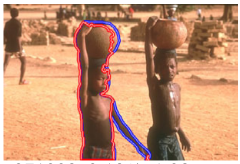

181079, 1.76%, 1.73 sec.


189080, 1.47%, 1.79 sec.

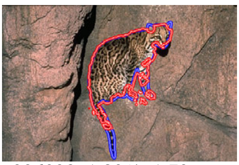

208001, 1.07%, 1.69 sec.


209070, 0.82%, 1.89 sec.


227092, 0.83%, 1.64 sec.

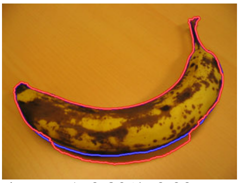

271008, 2.18%, 1.98 sec.


304074, 1.70%, 1.83 sec.


326038, 1.80%, 1.73 sec.


376043, 1.57%, 1.67 sec.

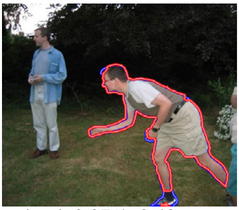

388016, 1.18%, 1.70 sec.

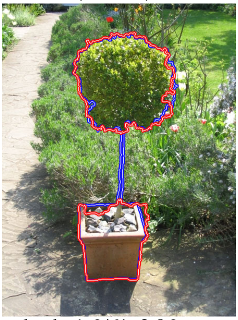

banana1, 2.80%, 3.23 sec.


banana2, 0.65%, 2.96 sec.

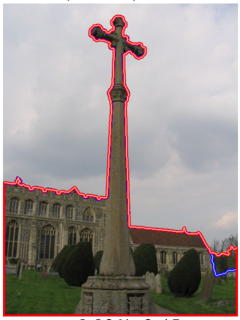

banana3, 1.29%, 3.06 sec.


book, 2.17%, 3.07 sec.


bool, 0.97%, 2.43 sec.


bush, 1.64%, 2.96 sec.

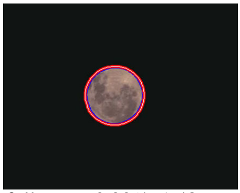

ceramic, 0.72%, 3.37 sec.

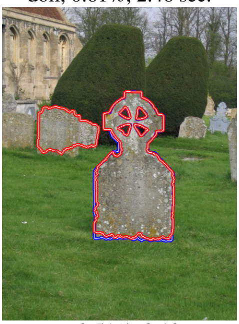

cross, 0.93%, 2.45 sec.


doll, 0.61%, 2.46 sec.


elefant, 0.55%, 3.67 sec.


flower, 0.48%, 2.62 sec.

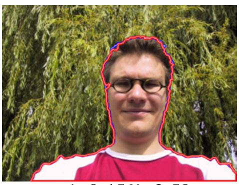

fullmoon, 0.30%, 1.48 sec.

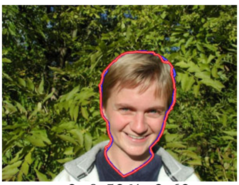

grave, 3.51%, 3.12 sec.


llama, 1.34%, 2.01 sec.

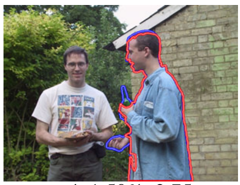

memorial, 1.19%, 2.87 sec.

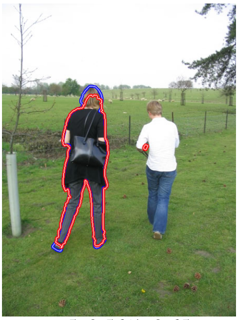

music, 0.66%, 2.89 sec.


person1, 0.45%, 2.59 sec.

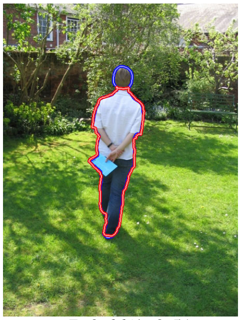

person2, 0.53%, 2.62 sec.

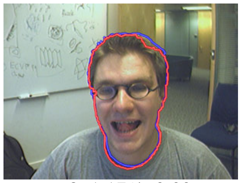

person3, 0.35%, 2.51 sec.


person4, 1.50%, 2.75 sec.

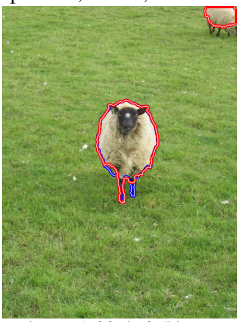

person5, 0.56%, 2.65 sec.


person6, 1.42%, 2.72 sec.


person7, 0.66%, 2.51 sec.

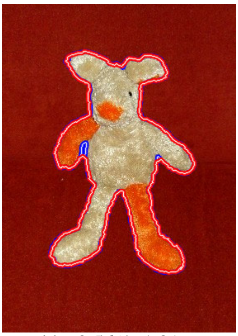

person8, 1.17%, 3.09 sec.

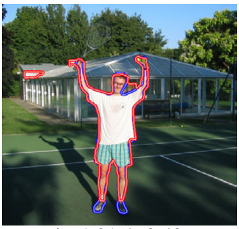

scissors, 1.75%, 2.79 sec.


sheep, 1.09%, 2.54 sec.


stone1, 0.20%, 2.76 sec.


stone2, 0.32%, 2.73 sec.


teddy, 0.56%, 1.31 sec.


tennis, 1.21%, 2.40 sec.