CO3 for Ultra-fast and Accurate Interactive Segmentation

Yibiao Zhao
Beijing Jiaotong University;
Lotus Hill Research Institute;
University of California, Los Angeles.
ybzhao@ucla.edu

Song-Chun Zhu
Lotus Hill Research Institute,
Wuhan, China;
University of California, Los Angeles.
sczhu@stat.ucla.edu

Siwei Luo
School of Computer and Information Technology,
Beijing Jiaotong University,
Beijing, China.
swluo@bjtu.edu.cn

ABSTRACT

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. (i) In representation, we pose the segmentation problem as partitioning an image domain into regions (foreground vs. background) or boundaries (on vs. off) which are dual but simultaneously compete with each other. Then, we formulate segmentation process as a combinatorial posterior ratio test in both the region and boundary partition space. (ii) In modeling, we use discriminative learning methods to train conditional models for both region and boundary based on interactive scribbles. We exploit rich image features at multi-scales, and simultaneously incorporate user's intention behind the interactive scribbles. (iii) In computing, we relax the energy function into an equivalent continuous form which is convex. Then, we adopt the Bregman iteration method to enforce the "coupling" of region and boundary terms with fast global convergence. In addition, a multigrid technique is further introduced, which is a coarse-to-fine mechanism and guarantees both feature discriminativeness and boundary preciseness by adjusting the size of image features gradually.

The proposed interactive system is evaluated on three public datasets: Berkeley segmentation dataset, MSRC dataset and LHI dataset. Compared to five state-of-the-art approaches including Boycov et al.[1], Bai et al.[2], Grady [3], Unger et al.[4] and Couprie et al.[5], our system outperforms those established approaches in both accuracy and efficiency by a large margin and achieves state-of-the-art results.

Categories and Subject Descriptors

I.4.6 [**Image Processing And Computer Vision**]: Segmentation; I.5.5 [**Pattern Recognition**]: Implementation-Interactive systems

MM'10, October 25–29, 2010, Firenze, Italy.

General Terms

Algorithms

Keywords

CO3, interactive multimedia system, image segmentation, discriminative learning, convex optimization

1. INTRODUCTION

Interactive image segmentation has been studied widely in the literature with a broad range of applications in the field of multimedia. The objective of interactive image segmentation is to segment objects of interest as perfectly and yet quickly as possible, while requiring the least amount of user's effort in drawing the interactive scribbles. The main challenges lie in the large variabilities of objects of interest including inhomogeneous region appearance and ambiguous object boundaries in nature images. In this paper, we present an ultra-fast and accurate interactive image segmentation system which consists of three components unified in a general framework as illustrated in Fig.1.

(I) Coupled representation for segmentation. As illustrated in the middle panel of Fig.1, we pose image segmentation problem as partitioning the image domain into a coupled representation of region and boundary. Region and boundary compete each other in a joint solution space. Then, the segmentation process is formulated as a posterior ratio test (foreground vs. background regions and on vs. off boundaries) in the partition space and is described by an energy function.

In the literature, representations for segmentation can be roughly divided into three types: boundary-based, region-based and graph-based methods. Region-based approaches [6, 7] partition the image domain into several mutually exclusive regions by learning a probabilistic model for each region individually. Boundary-based methods [8, 9] detect edges and boundaries by exploiting image gradient features and then trace and connect boundaries to achieve a similar goal of segmentation. From regions to contours, Paragios et al. [10] introduce geodesic weight into the region competition framework [6] to integrate region and boundary information. From contours to regions, Arbelaez et al. [11] show that a more sophisticated boundary model improves segmentation results largely. Graph-based methods [1, 2, 3] explore pair-wise connectivities with local image feature, and define both region and boundary information in a sparse graph structure.

(II) Conditional models for region and boundary. As il-

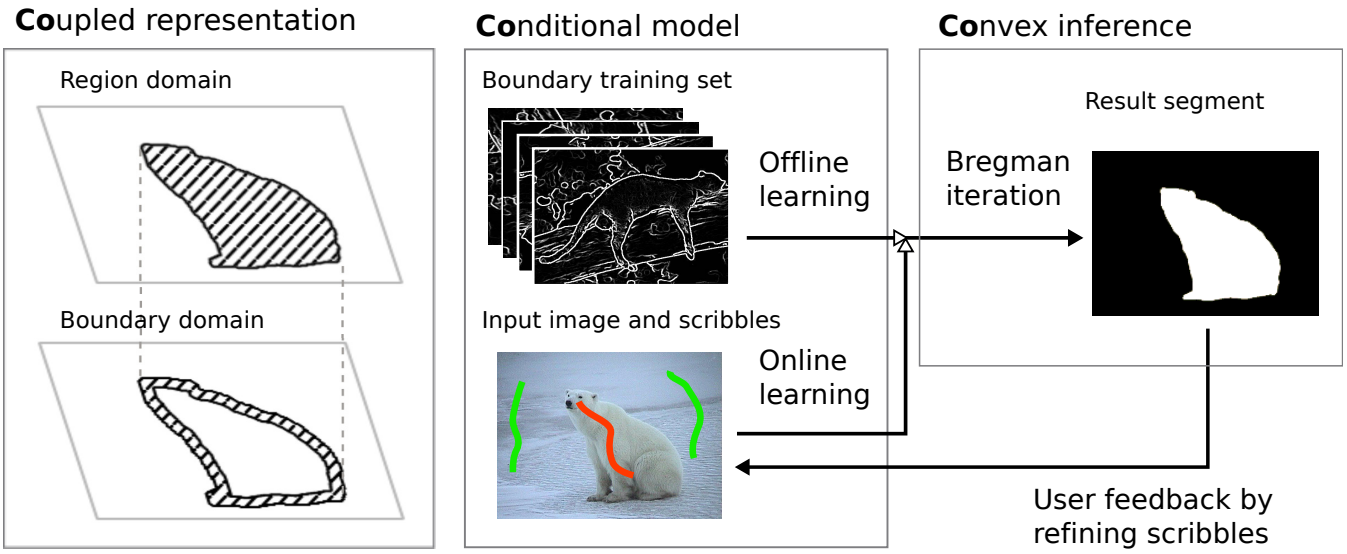


Figure 1: A general framework of CO3.

lustrated in the left panel of Fig.1, we use discriminative learning methods to train conditional models for both region and boundary based on interactive user's scribbles. The discriminative model for region (foreground vs. background) are trained online by exploiting rich image features at multi-scales (such as local gradient histogram and color histogram). The pixels around the user scribbles (foreground and background) are treated as initial tranning samples (positive and negative). The discriminative model for boundary (on vs. off) is trained off-line using a set of images with ground-truth object boundaries manually labelled. In addition, we incorporate user's intention behind scribbles as a prior model in our statistical framework.

In the literature, most state-of-the-art methods [1, 2, 3, 6, 12] define region models by some simple parametric generative forms using color or intensity, which can not capture the large variations in object appearance. Instead, discriminative models directly learn a label posterior to classify different regions in segmentation. SpatialBoost [13] firstly introduces the discriminative learning to this task and induces spatial relation as boosting features. Recently, Santner et al. [14] use the Random Forests to explore rich region features and apply Primal-Dual algorithm to optimize TV-norm, which achieves good performance on texture images and fast converge property with a GPU implementation.

(III) Convex inference for fast global convergence. In order to find the global solution of our energy function which consists of conditional models of region and boundary, we convert the energy function into a convex form by relaxing the constraint of discrete labeling of each pixel. Further, we introduce a Bregman iteration method [15] to enforce the coupling of region and boundary term. With this algorithm, these two energy terms can be quickly optimized separately and stepwise, and the cumulative errors from minimizing computation for the two terms are mutually cancelled in the Bregman iteration process. A multigrid scheme is further developed to deal with the dilemma of high-order features (subtle localization vs. discriminative power), and it also greatly accelerates convergence.

In the literature, some traditional methods [16, 17, 18] minimize energy functions by a gradient descent procedure which often gets "stuck" at local minimum. Bresson et al. [24] recently extended the active contour model into a convex formulation, and obtained more reliable results by some variational techniques, like Primal-Dual algorithm [4, 14, 24], Bregman iteration schemes [15, 19]. Also, benefiting from the global convergence property, graph based optimization algorithms draw wide attention, such as Graph Cuts [1], Shortest Path [2], Random Walk [3]. However, this type of algorithms is limited by using very local image features with pairwise interactions, and it often leads to rough segment boundaries (Fig.9).

The remainder of the paper is organized as follows. Sec.2 introduces the problem formulation with the coupled representation for segmentation. Sec.3 presents the discriminative learning methods for region and boundary. Sec.4 proposes the inference algorithm for solving the segmentation process. Sec.5 shows a series of experiments and comparisons. Finally, Sec.6 concludes this paper.

2. PROBLEM FORMULATION

We consider the continuous image domain Λ as shown in Fig.1. The interactive segmentation aims to decompose the image domain Λ into "foreground" Λ_R^+ and "background" Λ_R^-, meanwhile "boundary" Λ_B^+ and the rest Λ_B^-. The two partitions are strictly coupled by $\Lambda_B^+ = \partial\Lambda_R^+$. Thus we firstly present a coupled representation of both region assignment and boundary presence for this problem, and a solution is denoted as two coupled partitions:

$$\Pi = \begin{cases} \Lambda = \Lambda_R^+ \cup \Lambda_R^- \\ \Lambda = \Lambda_B^+ \cup \Lambda_B^- \end{cases} \quad s.t. \quad \Lambda_B^+ = \partial\Lambda_R^+. \tag{1}$$

Let u be the characteristic function $\mathbf{1}(x \in \Lambda_R^+)$ for region label, which takes value 1 as foreground, and 0 as background, and v be the boundary characteristic function $\mathbf{1}(x \in \Lambda_B^+)$, which takes 1 as boundary, and 0 as non-boundary. The discriminative probability test introduced is based on the dual region/boundary representation, with the backbone

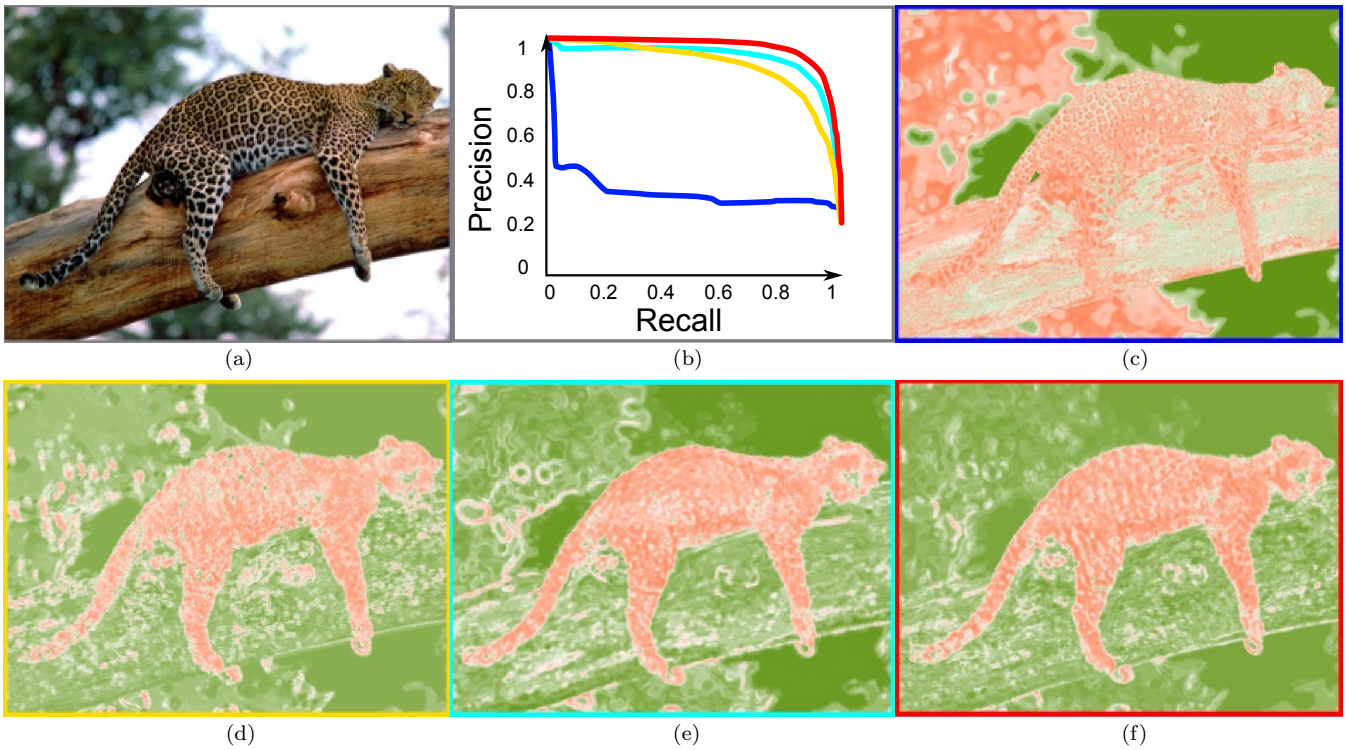


Figure 2: Discriminative power of different image features for appearance model learned from Leopard image (a). Probability ratio maps estimated by generative learning with Gaussian mixture model (c), discriminative learning with gradient histogram (d), color histogram (e) and all features (f) (Higher color intensity means higher probability), and their corresponding PR curves (b).

equation being:

$$E(\Pi) = -\int u \log \frac{P_R(+|I,S)}{P_R(-|I,S)} dx - \alpha \int v \log \frac{P_B(+|I)}{P_B(-|I)} dx, \quad s.t. \quad v = |\nabla u|. \tag{2}$$

The region test $T_R = -\log\{P_R(+|I,S)/P_R(-|I,S)\}$ is generalized from a region competition model [6] to a conditional probability test, which directly induces inter-class competition between foreground and background regions. Given foreground scribbles S_+ and background scribbles S_- as positive and negative training samples, the region appearance model $P_R(+|I)$ is learned by an online discriminative classifier considering multi-scale image features. Further, a prior model $P_R(+|S)$ representing user's intention behind scribbles is also incorporated by a model calibration scheme explained in Sect. 3.3.

The boundary test $T_B = -\log\{P_B(+|I)/P_B(-|I)\}$ offers a statistic explanation for the edge stop function (a decreasing function of image gradient in Geodesic Active Contour model [17]). It naturally embraces the state-of-the-art edge detectors [8, 9], which are trained offline. The parameter α serves as a weight between two tests.

In this way, the optimization problem is presented as searching for the best answer of two *hypothesis testing* combined: whether a specific pixel belongs to foreground, and whether a specific pixel belongs to boundary. The models for computing the region and the boundary representations are independent, yet the processes for computing the dual representations are tightly coupled with both processes interacting with and constraining each other by $v = |\nabla u|$.

3. DISCRIMINATIVE LEARNING

The conditional model $p(Y|I)$ specifies the probabilities of possible label sequences Y given an observed image I. Therefore, it does not expend much effort to model the observed image I. Meanwhile, the conditional probability of the label sequence can depend on arbitrary, dependent features (rich image features), which means we can model the conditional probability of a pixel $P(Y(x)|I_{\Lambda(x)})$ given a large neighborhood $I_{\Lambda(x)}$, or even an entire image I. Our conditional probability model combines discriminative image features that essentially improve the predicting accuracy as shown in Fig.2, and user's intention that is helpful for the extremely ambiguous data as illustrated in Fig.3.

3.1 Learning appearance models

For a two class problem, we express the probability in an exponential family form $P(+|I) = \frac{1}{Z}\exp\{-\sum_i \lambda_i h_i\}$. Thus, the log-posterior ratio is expressed as a simple linear form,

$$\log \frac{P(+|I)}{P(-|I)} = \sum_i \lambda_i h_i, \tag{3}$$

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. λ_i is its corresponding coefficient. This simple formula is connected to a logistic regression by fitting a binomial log-likelihood or equivalent to a Boosting algorithm by optimizing an exponential criterion [20].

Region appearance model is learned from user's scribbles. We randomly draw 200-400 positive and negative training

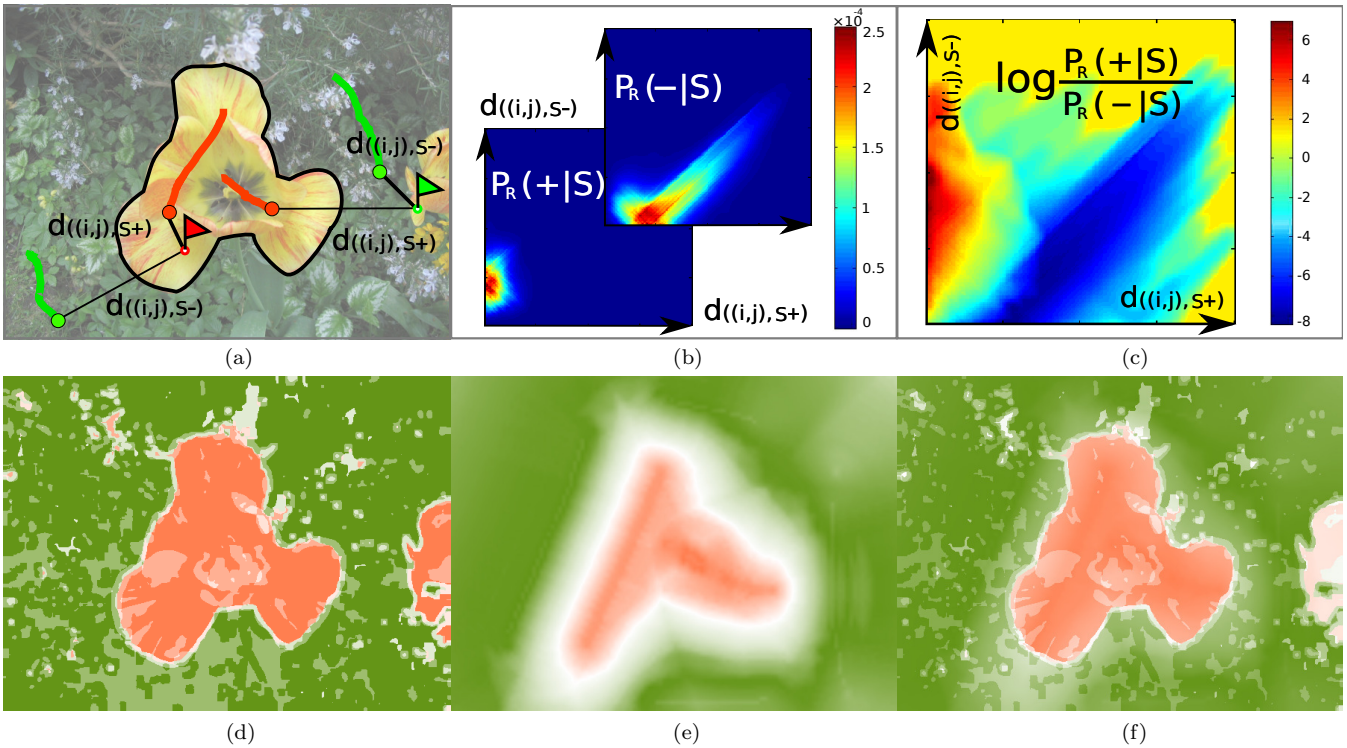


Figure 3: (a) Spatial prior indicates foreground pixels (such as red flag) located nearer to foreground scribbles, while the background pixels are closer to background scribbles. (b, c) The prior model discriminates foreground and background label by a large statistical divergence. With the spatial prior (e) incorporated, some ambiguous appearances (d) (such as an undesirable flower on the right) were suppressed in the combined model (f).

samples from pixels on foreground scribbles and background scribbles respectively. The GentleBoost [20] algorithm is applied to additively pursue most discriminative features robustly. *Boundary appearance model* is trained offline on the Berkeley dataset [21]. We apply a sigmoid transformation to each feature response $h(I) = sigmoid(r(I))$ where the sigmoid function maps the filter response $r(I)$ to a real value in $[0, 1]$. And we estimate the feature parameters λ by the logistic regression.

3.2 Feature design

Informative features capture the essence of an image pattern and greatly facilitate the learning stage of a discriminative algorithm. In the region model, we use two types of local histograms with different window sizes to characterize local image statistics around a specific pixel. They are histogram of color (HoC) and histogram of oriented gradients (HoG) [22]. For HoC, we choose HSV color space with 12-bin Hue, 8-bin Saturation and 8-bin intensity Value. For HoG, we first compute a local gradient for each intensity, and then project all gradients around a specific pixel into 12-bin histogram according to different orientations. In the boundary model, we prepare 12-bin boundary features by applying some gradient filters (4 different scales) to each color component. In our implementation, by using an integral image technique of histogram calculation [23], the computational time for the feature preparation of a 500*500 image is generally less than 0.2 seconds on a desktop PC before the users are involved.

3.3 Incorporating user's intention

In this study, we notice that interactive scribbles imply strong spatial prior to the user's intention. Pixels on the scribbles are deterministically labeled into corresponding regions. Meanwhile the other pixels that are close to the scribbles, are still preserving a high probability of belonging to the corresponding regions. In order to embed this prior information, a model calibration scheme is applied by,

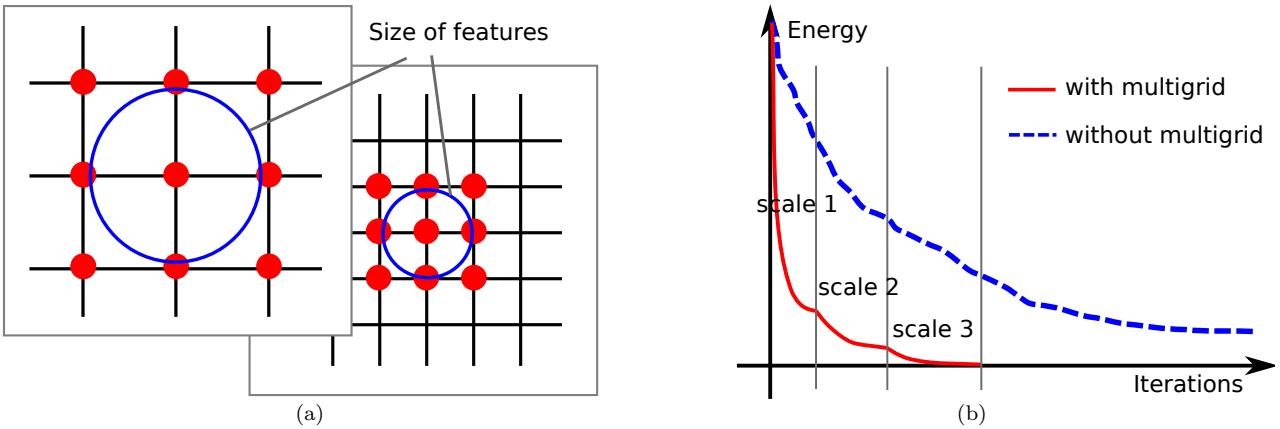
$$\log\frac{P_R(+|I,S)}{P_R(-|I,S)} = \log\frac{P_R(+|I)}{P_R(-|I)} - \log\frac{P_R(+)}{P_R(-)} + \log\frac{P_R(+|S)}{P_R(-|S)}. \quad (4)$$

The above equation essentially updates the old label prior $P_R(+)$ to $P_R(+|S)$. Since $P_R(+|I)$ is the posterior probability learned from the discriminative classifier, the old prior $P_R(+)$ only counts the frequency of labels without considering the prior behind scribbles. The new prior $P_R(+|S)$ models user's intention based on spatial closeness to scribbles. As a result, the updated predictor is a combination of both the image likelihood from the discriminative learner and the prior on scribbles.

We define $P_R(+_{(i,j)}|S)$ for each pixel (i, j) as its label frequency f with respect to Euclidean distances $d((i, j), S_+)$ and $d((i, j), S_-)$. The distance characterizes the closeness of the pixel (i, j) to its nearest pixel on scribbles $(i', j') \in S$.

$$\begin{aligned} P_R(+_{(i,j)}|S) &= f(+_{(i,j)}|d((i,j),S_+), d((i,j),S_-)), \\ d((i,j),S) &= \frac{1}{D_{max}} \min_{(i',j')\in S} \parallel (i,j),(i',j') \parallel, \end{aligned} \quad (5)$$

where D_{max} is the maximum Euclidean distance of the im-



(a) (b)

Figure 4: (a) The dilemma of high-order features (subtle localization vs. discriminative power); (b) convergence of energy with/without multigrid.

age (diagonal length). The prior is estimated in a non-parametric form.

As illustrated in Fig.3, the prior model defined above discriminates between foreground and background labels by a large statistical divergence. Taking advantage of the strong prior to the user's intention, the ambiguous appearances of the undesired flower on the right were thus suppressed in the combined model.

4. CONVEX INFERENCE

For optimizing the proposed energy function in Eq. 2, the Level Sets method [18] is a standard approach. However, it is a slow technique, and it only determines a local minimum. Fortunately, it is proved [24] that the energy in the form of Eq. 2 has an equivalent convex formulation by relaxing the discrete characteristic function u into a continuous interval $u \in [0,1]$, and $v = |\nabla u| \in [0,1]$. The optimal solution is obtained by thresholding the u into $\{0,1\}$ given a simple threshold μ.

Theorem 1: Call u^* any minimizer of $E(u)$ in Eq. 2 (which is a global minimizer by convexity), then the characteristic functions,

$$\mathbf{U}_R(x) = \begin{cases} 1 & if \quad u(x) > \mu, \\ 0 & otherwise \end{cases} \qquad \mu \in [0,1], \tag{6}$$

are also global minimizers of Eq. 2. Minimizing the energy with respect to u is equivalent to minimize a relaxed formula which takes $\mu \in [0,1]$. [25]

4.1 Bregman iteration

In order to compute a minimizer of Eq. 2 with relaxed target variable, T. Goldstein et al. [15] introduced a fast and accurate minimization algorithm, Bregman iteration. As reported in [19], this algorithm is much faster and more accurate than the Primal-Dual algorithm used in [4, 14], because of the quadratic convergence property rather than the linear convergence of Projection's algorithm, and it is faster than Graph Cuts [1] algorithm as well.

The key of this algorithm is using Bregman iteration to enforce "*Coupling*" of the region term T_R and boundary term T_B of Eq. 2 by solving the following unconstrained problem,

$$\begin{aligned} (u^{k+1}, v^{k+1}) &= \arg\min_{u\in[0,1],v} |v| T_B + u T_R + \frac{\beta}{2} \| v - \nabla u - \epsilon^k \|_2^2 \\ \epsilon^{k+1} &= \epsilon^k + \nabla u^{k+1} - v^{k+1} \end{aligned} \tag{7}$$

where ϵ^k is the cumulative error in the iteration k, and it can be mutually cancelled on the next iteration $k+1$ by the Bregman iteration process.

Without any hard constrain, the solving of the two energy terms of region and boundary is separated, and it is extremely fast. The minimizing solution for the region term u^{k+1} is characterized by the optimality condition with a fast approximation of "Gauss-Seidel" iteration:

$$\mu \nabla u = T_R + \mu div(\epsilon^k - v^k), u \in [0,1]. \tag{8}$$

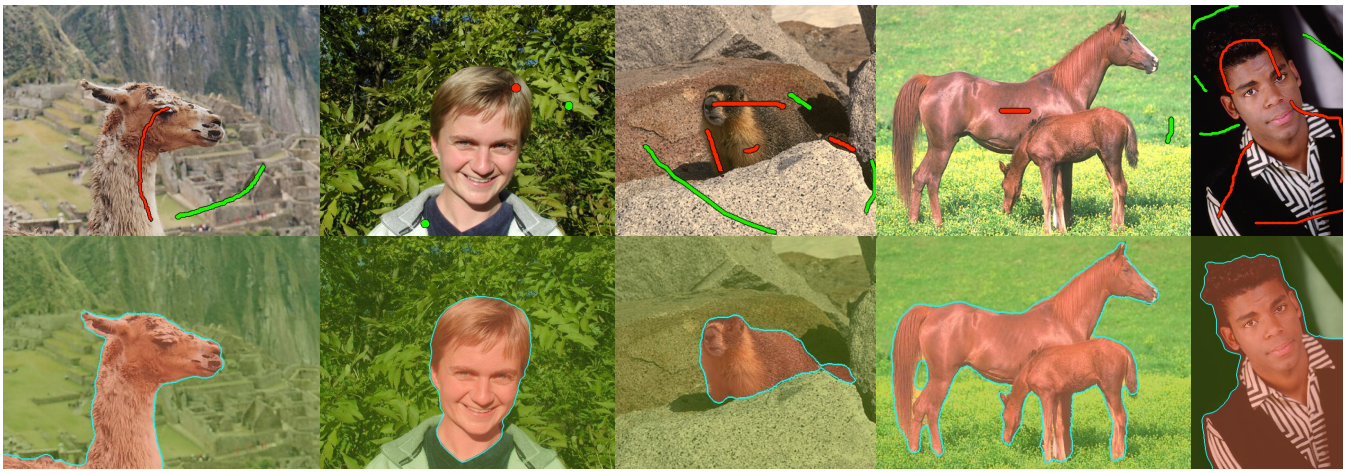
And the solution of minimizing boundary term v^{k+1} is given by soft-thresholding the "shrinkage" operator:

$$v_{i,j}^{k+1} = \frac{\nabla u^{k+1} + \epsilon^k}{|\nabla u^{k+1} + \epsilon^k|} max(|\nabla u^{k+1} + \epsilon^k| - \mu^{-1} T_B, 0). \tag{9}$$

4.2 Solving with multigrid

The high-order features greatly improve the discriminative power of classification, as we showed in Fig.2. However, it sometimes hurts the accuracy of localization. As shown in Fig.4(a), the larger the range of the feature window, the more information it carries to discriminate between different image patterns. At the same time, it results in a weaker ability to localize a point. That is what we called the dilemma of high-order features.

In our method, we deal with this problem by a simple coarse-to-fine scheme. Firstly, we use the largest feature window with strong discriminative power, and infer a coarse partition with down-sampling. After that, we use the bilinear interpolation to get an initial solution of the next scale, and update the model with the finer features (smaller window) to predict a subtle probability for the uncertain pixels ($u \in [0.1, 0.9]$). We keep changing the scale of domain and the scale of the features until reaching the pixel level. This scheme not only preserves high discriminative power but also brings high precision of boundary localization. Moreover, it accelerates convergence of the Bregman iteration dramatically. The convergence rate is presented in Fig.4(b).



(a) (b) (c) (d) (e)

Figure 5: Some segmentation results on MSRC [12] and Berkeley dataset [21]. (best shown in color)

Method	Bai et al. [2] IJCV2009	Grady [3] PAMI2006	Couprie et al.[5] ICCV2009	Boycov et al.[1] ICCV2001	Unger et al.[4] BMVC2008	Our method
Region precision	0.50	0.56	0.58	0.69	**0.73**	**0.79**
Boundary precision	0.05	0.10	0.11	0.14	**0.16**	**0.21**
Average running time	**0.52 s**	1.58 s	0.73 s	0.65 s	0.84 s	**0.12 s**

Table 1: Quantitative evaluation on LHI dataset [26].

5. EXPERIMENTAL ANALYSIS

The proposed system is tested on several challenging images from Berkeley segmentation dataset [21], MSRC dataset [12], and LHI interactive segmentation benchmark [26]. In the experiment, an Intel Core 2 Duo E7300(2.67 GHz) CPU and 2GB RAM PC is used as the experimental platform.

We firstly start with an evaluation of system performance on large datasets and show its superiority over state-of-the-art methods on both accuracy and efficiency. Then we explicitly analyze each component in our proposed system and discuss their contributions.

5.1 Qualitative evaluation

Fig.5 presents some results of our algorithm on several popular test images of Berkeley segmentation dataset [21] and MSRC dataset [12].

In Fig.5 (a,b,d) with very few user scribbles (small training set), the most discriminative patterns between foreground and background are easily captured by discriminative learning. Even for some very ambiguous image patterns like Fig.5 (c), the algorithm still performs well after more scribbles being placed.

Fig.5 (e) is an extreme case in which there are overlapping image patterns (eg. black and flat pattern) between the foreground and background. In this situation, the algorithm can't tell any difference according to local features only. Fortunately, spatial prior helps to give a reasonable explanation for each image region in our system. Actually, local ambiguous patterns are nontrivial and widely exist in natural images, and spatial prior eliminates some obvious false alarm around the scribbles.

5.2 Quantitative evaluation

Interactive systems are generally hard to evaluate quantitatively due to the subjectivity and variation of human interference. In this paper, we evaluate our method on the [3]LHI interactive segmentation benchmark [26]. The benchmark provides several natural images, corresponding ground-truths and three users' scribbles for each image.

We compare our system to 5 state-of-the-art approaches: [4]Boycov et al.[1] (ICCV2001), [5]Bai et al.[2] (IJCV209), [6]Grady [3] (PAMI2006), [7]Unger et al.[4] (BMVC2008), [8]Couprie et al.[5] (ICCV2009). The quantitative results are presented in Tab.1, some sample images are shown in Fig.10.

Accuracy Two evaluation criteria of accuracy are applied: (1) Region precision, $\Lambda_R^+(a) \cap \Lambda_R^+(b)/\Lambda_R^+(a) \cup \Lambda_R^+(b)$, measures an overlap rate between a result foreground and the corresponding ground truth foreground; (2) Boundary precision, $1/D(\Lambda_B^+(a), \Lambda_B^+(b))$, 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.

Efficiency We also compute the average running time for these methods over all images on LHI database (Tab.1). Our algorithm is many times faster than others. Then we used

[3]http://www.imageparsing.com/interactivesegmentation.html
[4]http://vision.csd.uwo.ca/code/
[5]http://www.tc.umn.edu/~baixx015/example.htm
[6]http://www.cns.bu.edu/~lgrady/software.html
[7]http://gpu4vision.icg.tugraz.at/
[8]http://sourceforge.net/projects/powerwatershed/

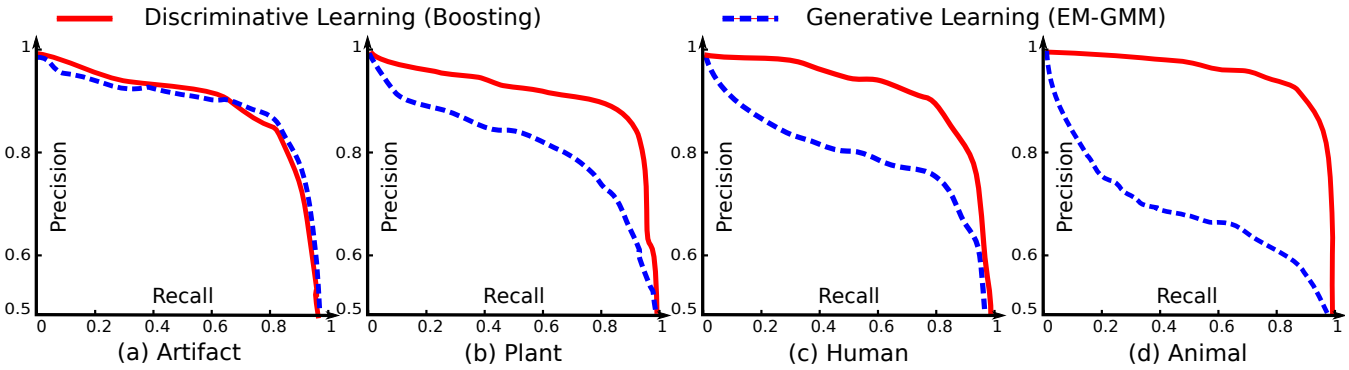


Figure 6: The contribution of discriminative learning on different image categories of LHI dataset [26].

the flower image (Fig.3) as a standard test image, and resized it to different resolutions. The computational costs for these algorithms with respect to different image resolutions are illustrated in Fig.8. Besides that, we also compared our system with a GPU implementation of Random Forests and Total Variation algorithm, which is a recent improvement of TVSeg [4]. Our algorithm achieves 50 times faster than the GPU implementation on a 500×350 image.

5.3 Algorithm analysis

From the results mentioned above, our algorithm generally outperforms other state-of-the-art algorithms both on accuracy and efficiency. This is owing to some major contributions in this paper, including discriminative learning with rich image features, spatial prior, and convex inference algorithm we applied. Next, we will discuss the contributions of each of these components.

Discriminative learning Based on 4 different categories of LHI dataset [26], we test Gaussian Mixture Model (being popularly used in other algorithms) on the HSV color space and discriminative learning of Boosting with rich image features. We plot the PR curve for each category on average as displayed in Fig.6. In artifact category, images are almost flat color patterns without many complex textures. GMM model shows a good ability to estimate the true color distribution, and Boosting achieves similar performance. In the categories of plant, human, and animal, it is hard to effectively separate the complex object from the natural background by simple color intensities in GMM, but boosting can still capture the distinct texture pattern or shading effect by rich image features.

Spatial prior In Fig.7, we plot the segmentation precision curve with respect to the percentage of user's scribbles. With our proposed spatial prior, the interaction process is accelerated, and the system achieves high precision quickly. Even if a lot of user interactions are imposed, the prior still helps to get a higher precision due to its ability to handle some local ambiguities.

Inference algorithm Actually, our proposed energy form can be solved by traditional Level sets method [18], TV optimization of Primal-Dual [4] and Bregman Iteration [15], as well as some discrete graph-based optimization algorithm (known as solving different norm of boundary regularization [5]), such as Shortest Path (Geodesics) [2], Random Walker [3] and Graph Cuts [1]. In Fig.9, we set the same input energy terms for these algorithms, and then optimize the energy by different algorithms. We can see that continuous

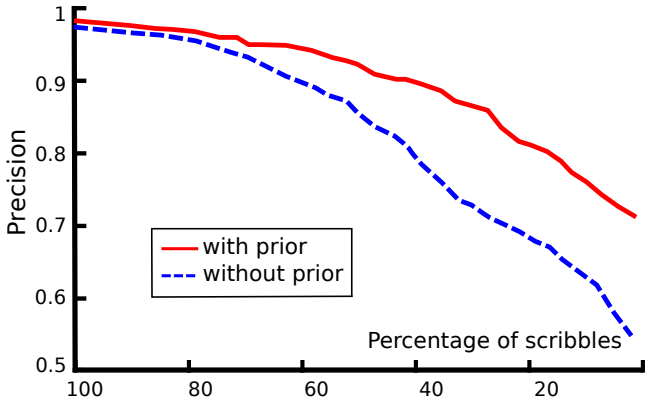


Figure 7: The improvement on segmentation precision due to the combination of spatial prior.

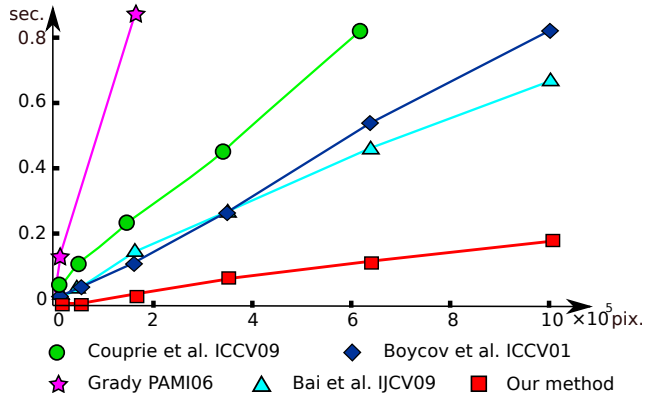


Figure 8: The computational cost of inference algorithms with respect to image sizes.

Method	Learning	Inference	Total
Santner et al.[14] (2009)	1.14 s	0.5 s	1.64 s
Our system	**0.015** s	**0.016** s	**0.031** s

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.

methods (d-f) achieve better sub-pixel precision than the discrete algorithms, and Level sets method will get "stuck" in a local minimum, e.g. the front legs of fox image. Primal-Dual algorithm used in [4, 14] suffers from some local noise, which is also reported in [19].

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 latest demos/executables of CO3 are available at `http://www.stat.ucla.edu/~ybzhao/research/co3/`.

7. ACKNOWLEDGMENTS

The authors would like to thank Tianfu Wu, Liang Lin, Zhangzhang Si, Mingtian Zhao, Yanbiao Duan and anonymous reviewers for their constructive suggestions. This work at the Lotus Hill Research Institute is supported by China 863 Program 2007AA01Z340, 2009AA01Z331 and NSF China grants 60970156, 60728203, and the work at Beijing Jiaotong University is supported by China 863 Program 2007AA01Z168, NSF China grants 60975078, 60902058, 60805041, 60872082, 60773016, Beijing Natural Science Foundation 4092033, and Doctoral Foundations of Ministry of Education of China 200800041049.

8. REFERENCES

[1] Boykov, Y.Y., Jolly, M.P.: Interactive graph cuts for optimal boundary region segmentation of objects in n-d images. In: ICCV (2001) vol.1 105–112

[2] Bai, X., Sapiro, G.: Geodesic matting: A framework for fast interactive image and video segmentation and matting. IJCV 82 (2009) 113–132

[3] Grady, L.: Random walks for image segmentation. PAMI (2006) 1768–1783

[4] Unger, M., Pock, T., Trobin, W., Cremers, D., Bischof, H.: TVSeg - interactive total variation based image segmentation. In: BMVC, Leeds, UK (2008)

[5] Couprie, C., Grady, L., Najman, L., Talbot, H.: Power Watersheds: A New Image Segmentation Framework Extending Graph Cuts, Random Walker and Optimal Spanning Forest. In: ICCV(2009) Page 731–738

[6] Zhu, S.C., Yuille, A.: Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation. PAMI 18 (1996) 884–900

[7] Tu, Z., Zhu, S.C.: Image Segmentation by Data Driven Markov Chain Monte Carlo. PAMI vol. 24 no. 5 (2002) pp. 657–673

[8] Konishi, S., Yuille, A.L., Coughlan, J.M., Zhu, S.C.; Statistical Edge Detection: Learning and Evaluating Edge Cues. PAMI Volume 25 , Issue 1 (2003) Pages: 57–74

[9] Martin, D.R., Fowlkes, C.C., Malik, J.: Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues. PAMI Volume 26, Issue 5 (2004) Pages: 530–549

[10] Paragios, N., Deriche, R.: Geodesic Active Regions and Level sets Methods for Supervised Texture Segmentation. IJCV (2002) Page 223–247

[11] Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: From contours to regions: An empirical evaluation. In: CVPR (2009) 2294–2301

[12] Rother, C., Kolmogorov, V., Blake, A.: Grabcut: Interactive foreground extraction using iterated graph cuts. ACM ToG 23 (2004) 309–314

[13] Avidan, S.: Spatialboost: Adding spatial reasoning to adaboost. In: ECCV (2006) 386–396

[14] Santner, J., Unger, M., Pock, T., Leistner, C., Saffari, A., Bischof, H.: Interactive texture segmentation using random forests and total variation. In: BMVC, London, UK (2009)

[15] Goldstein, T., Osher, S.: The split bregman method for l1-regularized problems (2008) UCLA CAM Report 08–29.

[16] Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. IJCV Vol.1 (1988) 321–331

[17] Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. IJCV 22 (1997) 61–79

[18] Osher, S., Sethian, J.A.: Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations. J. of Computational Physics 79 (1988) 12–49

[19] Goldstein, T., Bresson, X., Osher, S.: Geometric applications of the split bregman method: Segmentation and surface reconstruction. UCLA CAM Report (2009) 09–06.

[20] Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting. Annals of Stat. 28 (2000)

[21] Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: ICCV (2001) vol.2 416–423

[22] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. IJCV. Vol.2. (2005) 886–893

[23] Viola, P., Jones, M.: Robust real-time face detection. IJCV (2004) Page 137–154

[24] Bresson, X., Esedoglu, S., Vandergheynst, P., Thiran, J.P., Osher, S.: Fast global minimization of the active contour/snake model. J. of Mathematical Imaging and Vision 28 (2007) 151–167

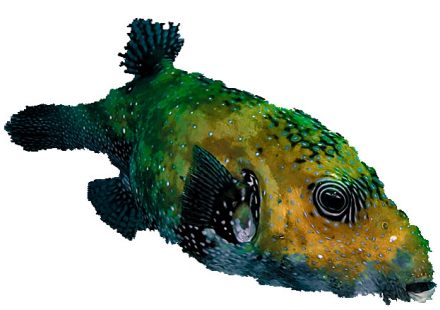
[25] Chan, T., Esedoglu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. UCLA CAM Report (2004) 04–54.

[26] Yao, B., Yang, X., Zhu, S.C.: Introduction to a large-scale general purpose ground truth database: Methodology, annotation tool and benchmarks. In: EMMCVPR (2007) 169–183

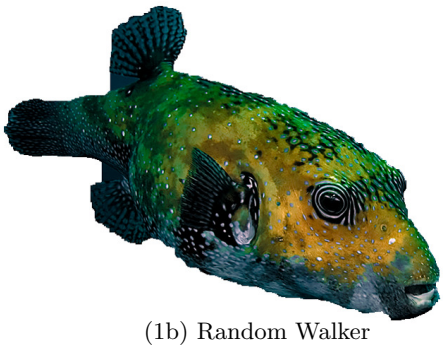
[27] Luo, J., Savakis, A.E.: Self-supervised texture segmentation using complementary types of features. Pattern Recognition 34 (2001), pp. 2071–2082

(1)

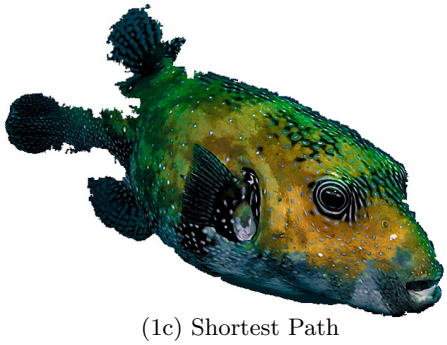
(2)



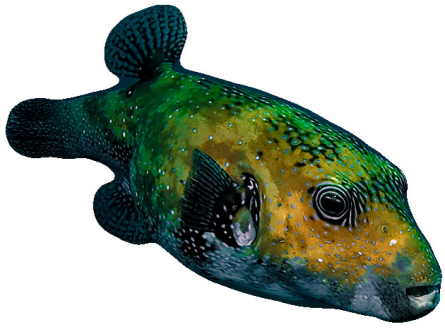
(1a) Graph Cuts



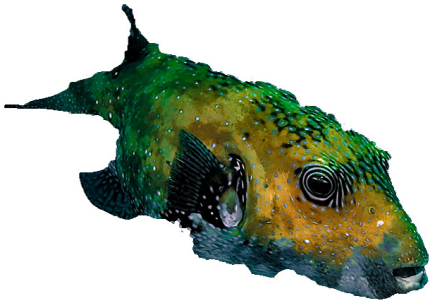
(1b) Random Walker



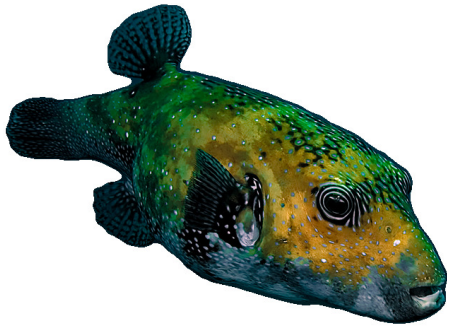
(1c) Shortest Path



(1d) Level Sets

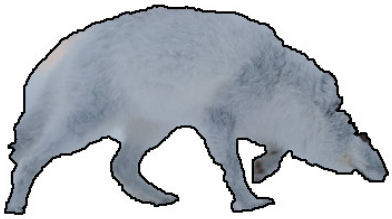


(1e) Primal-Dual



(1f) Bregman Iteration (ours)

(2a) Graph Cuts



(2b) Random Walker

(2c) Shortest Path

(2d) Level Sets

(2e) Primal-Dual

(2f) Bregman Iteration (ours)

Figure 9: Results of different inference algorithms under same energy setting.

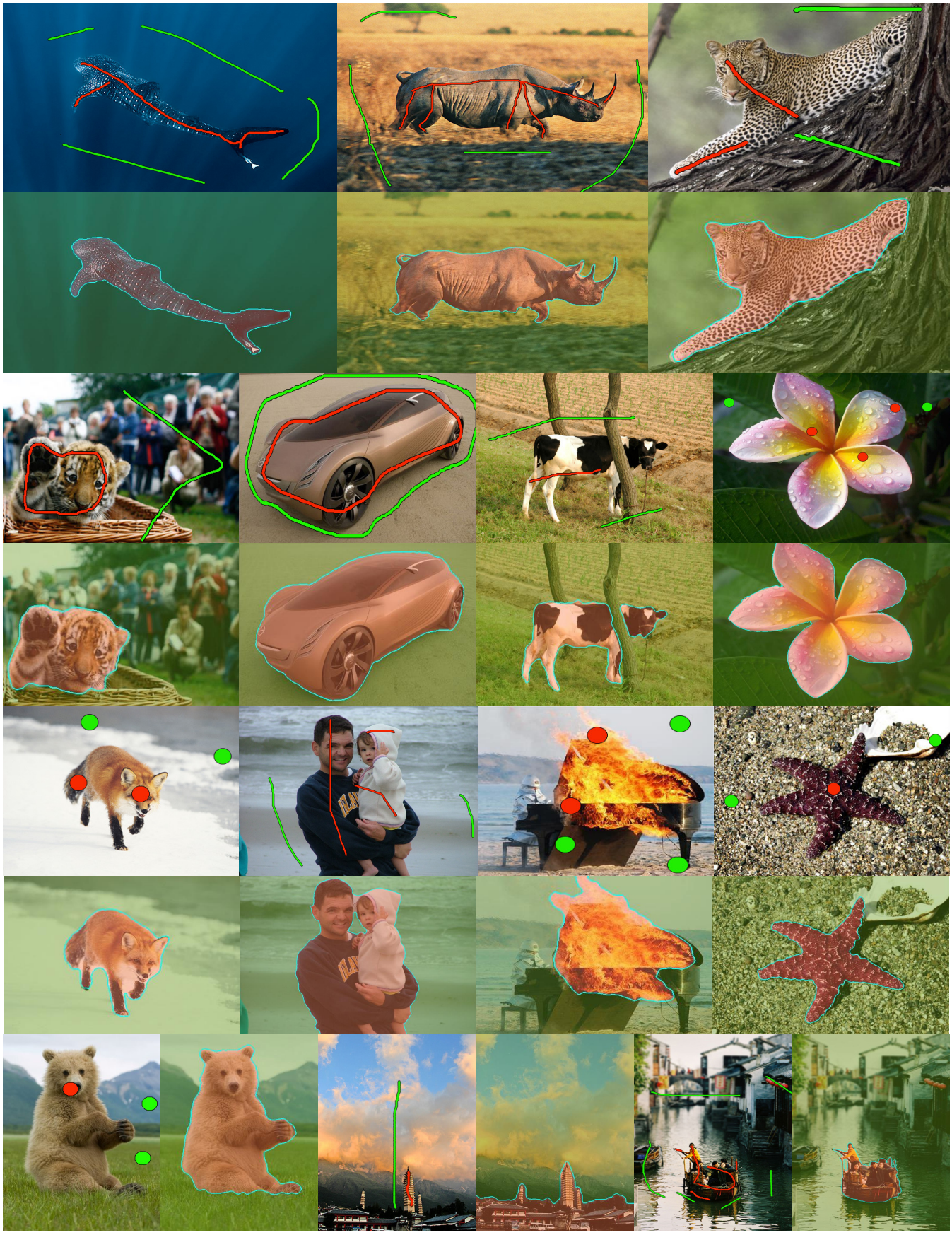


Figure 10: Sample results on LHI dataset [26]. (This figure is best presented with scale 500% in color.)