Salient Object Segmentation Based on Automatic Labeling

Lei Zhou[1], Chen Gong[1], YiJun Li[1], Yu Qiao[1], Jie Yang[1], and Nikola Kasabov[2]

[1] Institute of Image Processing and Pattern Recognition, Shanghai Jiaotong University, China
[2] Auckland University of Technology, New Zealand

Abstract. This paper proposes an automatic salient object extraction framework. Firstly, the saliency model are developed by applying the low level color features and the boundary prior. The initial salient regions are extracted by adaptive thresholding. Multiple classifiers are trained with extracted initial region, which reflect color information of images or adopt label propagation. Then, the labels for segmentation are generated automatically via classifier composition. Finally, the conditional random field (CRF) model based on multi-feature fusion is applied for salient object segmentation. Empirical study reveals that the proposed algorithm achieves satisfying performance.

Keywords: saliency detection, automatic object segmentation, automatic labeling, conditional random field.

1 Introduction

Object segmentation is a challenging problem in computer vision and it has wide applications in areas such as object recognition, image classification and image retrieval, etc. Therefore, many methods have been proposed to extract interesting objects automatically. Salient object extraction can be formulated as a binary labeling problem which assigns a unique label to each pixel (belonging to salient object or background), and the labeling problem is often formulated as a minimization of the energy [1]. In the past few years, many energy formulations have been developed which adopt either markov random field (MRF) or conditional random field (CRF). The efficiency of methods mainly lies in how the appearance cues, such as color, texture or valuable high level information, are defined and incorporated into the segmentation model. In the context of salient object segmentation based on saliency map, the key issue is how to utilize the saliency model efficiently. Many works focus on incorporating the saliency information into segmentation model directly. In [2], the saliency map obtained via maximal symmetric surrounding region is directly exploited to construct the data term for graph cut. In [3], saliency map and color similarity are used to define the two complementary data terms and the weights for the two data terms are set adaptively. There also exist works which extract the initial regions of salient objects based on the saliency models, so as to define more discriminative features for

M. Lee et al. (Eds.): ICONIP 2013, Part III, LNCS 8228, pp. 584–591, 2013.

segmentation. In [4], the seeds of salient object/background are selected manually by thresholding on the saliency map and they are updated iteratively. In [5], an iterative unsupervised salient object segmentation approach based on kernel density estimation (KDE) and two-phase graph cut is proposed. In [6], a CRF model is constructed to integrate cues such as color and context information. There are also many strategies which extract initial salient regions based on the proposed saliency map of high quality or design schemes that are more robust than thresholding based on existing saliency models. In [7], convex hull analysis is performed on several binary object masks which are generated by diverse saliency maps, to select the most compact shape to represent the object. In [5], an initial segmentation is generated by thresholding on kernel density estimation based saliency model. In [6], an adaptive selection mechanism is designed to select the minimal connected region of the saliency map as the initial segmentation of the salient object according to three measures, connectivity, convexity and saliency.

In order to enhance the segmentation reliability, especially for complicated images, and improve the overall segmentation quality, we propose an efficient saliency model which exploits the low level features, such as color contrast and color distribution, and the boundary prior. Then, a initial segmentation of the salient object is selected according to the saliency map. To extract the initial region more precisely, color Gaussian mixture model (GMM) and a semi-supervised label propagation method are applied to generate seeds automatically via classifier composition. Then, the seeds for salient objects and background are generated automatically, and they are used to train the appearance cues for segmentation. Finally, a CRF model is constructed for obtaining the final segmentation results. The main contributions of the paper are summarized below:

1 An automatic segmentation algorithm that can extract object from background without any interaction is proposed.

2 A saliency model that integrates low level features of images and prior information related to boundaries of images is developed.

3 An automatic labeling scheme based on classifiers composition is presented.

2 Saliency Model Combination

Color Contrast and Color Distribution: Color contrast is inspired by the observation that color components of a salient object may have a strong contrast to their surroundings. Assume that an image is divided into regions (or superpixels) $R_i, i \in \{1, 2, 3, ..., N\}$. Then, region i's color contrast saliency S_i^{con} is computed according to the definition in [8]:

$$S_i^{con} = \sum_{j \neq i} D_c(R_i, R_j) D_s(R_i, R_j), \tag{1}$$

where $D_c(R_i, R_j)$ is the color distance between the two regions, and $D_s(R_i, R_j)$ is the spatial distance between the regions R_i and R_j.

The distribution of color information in R_i, D_i^{dist} is defined in eq. (2) according to the definition in [8]:

$$D_i^{dist} = \sum_{j \neq i} w_{ij}^C (p_j - \overline{p_i})^2. \tag{2}$$

In eq. (2), p_i describes the average position of superpixel i and $\overline{p_i}$ is the weighted average position. w_{ij}^C is the weight corresponding to color similarity between the region i and region j. The regions with higher distribution variances may have lower saliency, so we define the color distribution saliency as:

$$S_i^{dist} = 1 - D_i^{dist}. \tag{3}$$

Boundary Prior: In an image, the object near to the boundary is less-likely to be the salient object. Geodesic distance is computed based on the nearest background nodes Ω_B which are selected by an method similar to [9]. For the pixel m, the distance is defined as $g(m) = \min_{s \in \Omega_B} d_g(s, m)$. The geodesic distance is computed based on the length of a discrete path:

$$L(\Gamma) = \sum_{i=1}^{n-1} \sqrt{(1 - \gamma_g) d(\Gamma^i, \Gamma^{i+1})^2 + \gamma_g \parallel \nabla(\Delta^i) \parallel^2}. \tag{4}$$

where Γ is an arbitrary discrete path with pixels defined as $\{\Gamma^1, ..., \Gamma^n\}$. $d(\Gamma^i, \Gamma^{i+1})$ is the Euclidean distance between two points (Γ^i and Γ^{i+1}). Then the distance is defined as $d_g(a, b) = \min_{\Gamma \in P_{a,b}} L(\Gamma)$. We use the parameter γ_g to weight two kinds of distances: the Euclidean distance and the distances computed based on image gradient. For quick computation, the fast marching algorithm is used [10] to compute the geodesic distances. Then, the saliency model related to boundary prior is $S_i^{Bd} = g(i)$.

Similar to [8], the nonlinear combination of color contrast, color distribution and boundary prior S^{cmb} is defined by eq. (5),

$$S_i^{cmb} = S_i^{con} \times S_i^{dist} \times S_i^{Bd}. \tag{5}$$

The initial salient object region extracted based on saliency map is defined as $INTR = \{i | S_i^{cmb} \geq \eta\}$, the background region is $INTB = \{i | S_i^{cmb} < \eta\}$. The adaptive threshold $\eta = 1.5 \times S_{mean}$ where S_{mean} is the mean saliency over the entire saliency map.

3 Classifier Composition for Automatic Labeling

3.1 Classifier Based on Color Information

The basic features for pixel p is RGB color and the feature vector is represented as $I_p = RGB_p$. We define FG to represent the salient object and BG to represent the background. The color information contained in the sets $INTR$ and $INTB$ are modeled as Gaussian mixture model (GMM), respectively. Let the color

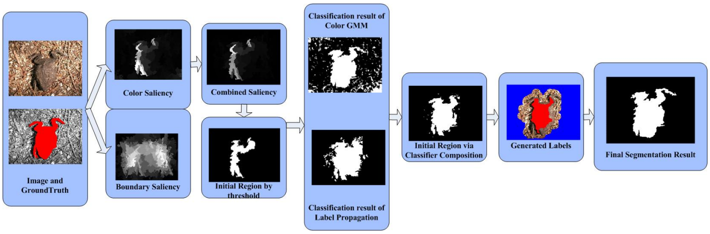

Fig. 1. The flow-chart of automatic seed generation. The pixels marked as red are the labels for objects and pixels marked as blue are background labels.

models be represented by GMM $\{\alpha_c, \mu_c, \Sigma_c\}_{c=1}^{C}$ in the RGB color space, where $\alpha_c, \mu_c, \Sigma_c$ represent the set of weight, mean color and covariance matrix of the $c-$th component, respectively. The mixture distribution of I_x can be formulated as a linear superposition of Gaussians in the form:

$$V(I_x|l) = \sum\nolimits_c \alpha_{cl} N(I_x|\mu_{cl}, \Sigma_{cl}), l \in \{FG, BG\}, \tag{6}$$

where $\{\alpha_{cl}, \mu_{cl}, \Sigma_{cl}\}$ represent the weight, the mean color and the covariance matrix of the $c-$th component learned from color information of class l, $l \in \{FG, BG\}$. In our experiments, GMM with 5 components are used to represent the color models in each class. Then, the posterior probability at each pixel p of the image is:

$$\mathrm{P}_{gmm}(F_p = l|I_p) = \frac{V(I_p|l)}{V(I_p|FG) + V(I_p|BG)}, l \in \{FG, BG\}. \tag{7}$$

The basic classifier is a function mapping the image space to figure-ground classification space:

$$H_{col}(p(I_i; F_i)) = \begin{cases} 1, p(F_i = FG|I_i) > p(F_i = BG|I_i) \\ 0, p(F_i = FG|I_i) \leq p(F_i = BG|I_i), \end{cases} \tag{8}$$

where $p(F_i|I_i)$ is the posterior probability associated with label F_i at pixel i. F_i is the label at pixel i and $F_i \in \{FG, BG\}$.

3.2 Classifier Based on Label Propagation

Given a point set $X = \{x_1, ..., x_l, x_{l+1}, ..., x_n\}$ and a label set $L = \{1, ..., c\}$. The indication vector is $y = \{y_1, ..., y_n\}^T$, in which $y_i = 1$ if x_i is labeled as salient object, and $y_i = 0$ otherwise. We set $y_i = 1$ for pixel $i \in INTR$. Let $f : X \rightarrow R^n$ represent a propagation function which assigns a value f_i to each point x_i. A graph $G = (V, E)$ is built on the points set. The edges E are weighed by an affinity matrix $W = [w_{ij}]_{n\times n}$. Given the graph, the degree matrix is

$D = diag\{D_{11}, ..., D_{nn}\}$, where $D_{ii} = \sum_j w_{ij}$. Similar to [11], the optimization label propagation problem is:

$$\min_{\mathbf{f}: f(x) \in \mathbb{R}} \quad Q(\mathbf{f}) = \frac{1}{2} \sum_{k=1}^{n} \sum_{j=1}^{n} \omega_{kj} \left(\frac{1}{\sqrt{D_{kk}}} f_k - \frac{1}{\sqrt{D_{jj}}} f_j \right)^2 + \theta \sum_{k=1}^{n} (f_k - y_k)^2, \quad (9)$$

where θ controls the balance between the smoothness constraint and fitting constraint. The result function with unnormalized Laplacian matrix is:

$$f^* = (D - \alpha W)^{-1} y, \quad (10)$$

where $\alpha = \frac{1}{1+\theta}$. f^* can be also interpreted as a probability and we define $P_{lp} = f^*$. Then, the classifier related to label propagation is described as:

$$H_{lab}(f(I_i)) = \begin{cases} 1, f^*(I_i) > \tau \\ 0, f^*(I_i) \leq \tau, \end{cases} \quad (11)$$

where τ is the adaptive threshold and we set $\tau = 1.5 \times \frac{\sum f^*(I_i)}{n}$.

3.3 Automatic Labeling

To divide the image region into several regions via Classifier Composition. Two basic pixel sets $A = \{i | H_{col}(i) > 0\}$ and $B = \{i | H_{lab}(i) > 0\}$ are defined. $\bar{A}$ and $\bar{B}$ are the related complements.

$$\begin{array}{l} C1 = \{i | i \in A \cap B\}, C2 = \{i | i \in A \cap \bar{B}\}, \\ C3 = \{i | i \in B \cap \bar{A}\}, C4 = \{i | i \in \bar{B} \cap \bar{A}\}. \end{array} \quad (12)$$

We use pixels in $C1$ to generate the foreground seeds LF. The pixels with low saliency value is contained in set $SAL = \{i | S^{cmb}(i) < 0.1\}$. The pixel in set $D = C4 \cup SAL$ is utilized to generate the background seeds. We shrink the initial region $C1$ to avoid inexact boundaries and form an accurate object labels. By shrinking pixels in set D, a ring region is obtained which are taken as the background labels LB. The process of automatical labeling is illustrated in Fig. 1.

4 Formulation of Salient Segmentation

Image segmentation can be modeled with a conditional random field (CRF). Consider a random field F defined over a set of variables $\{F_1, F_2, ..., F_N\}$. The domain of each variable is a set of labels $L = \{\ell_1, \ell_2, ..., \ell_k\}$. Let $I = \{I_1, .., I_N\}$ be the observed data corresponding to image information and N is the image dimension. I_i is the feature vector at pixel i and $I_i = \{RGB_i, LAB_i, S_i^{cmb}\}$. F_i represents the label assigned to pixel i. In our model, two features, color model P_{gmm} and label propagation probability P_{lp} are used. Let w be an $N \times 2$ matrix and $w = \{w_1, w_2\}$, where $w_i = [w_{i1}, w_{i2}, ..., w_{iN}]^T$ is an N-dimensional vector

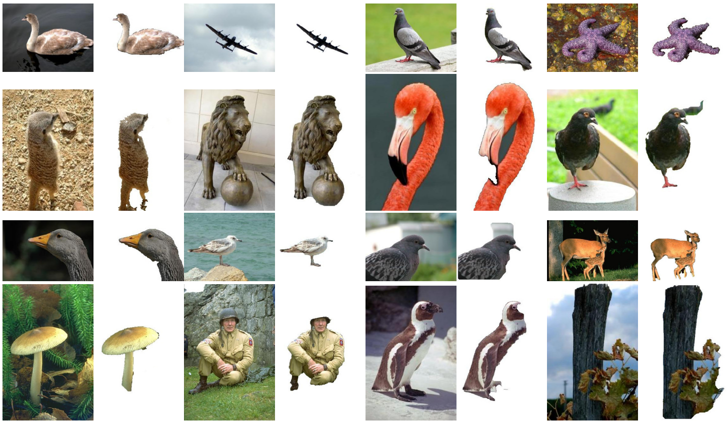

Fig. 2. Segmentation results of our method. The corresponding results are listed next to the original images.

related to a feature. We formulate the segmentation problem as a binary labeling task and the energy function $E(F|I,w)$ takes the form:

$$E(F|I,w) = w_1^T E_1^{Col}(F|I) + w_2^T E_1^{Lp}(F|I) + \tau E^{Pair}(F|I). \tag{13}$$

The energy function $E_1^{Col}(F|I)$ is the energy related to color information and $E_1^{Lp}(F|I)$ represents the single-site clique potentials related to label propagation probability. $E^{Pair}(F|I)$ is the pair-site clique potentials and the parameter τ is a control weight for the pairwise constraint. We set $\tau = 1$ in the experiments. In the process of CRF construction, labels LF, LB (described in section 3.3) are used to compute P_{gmm} and P_{lp}. Then, the common unary potential for two features is:

$$\begin{aligned} &E_1^{Col/Lp}(F|I) = \{V_1^{Col/Lp}(F_1), ..., V_N^{Col/Lp}(F_N)\}^T, \\ &V_1{}^{Col/lp}(F_q = l) = -\log(\mathrm{P}_{gmm/lp}(I_q; F_q = l)). \end{aligned} \tag{14}$$

In the experiments, we set $w_i = \{0.5, 0.5\}^T$. The pairwise term between neighbor nodes is computed based on the low-level features (such as RGB color, LAB color and saliency value). The related pixel pairwise term is defined as:

$$E^{pair}(i,j) = exp(-|I_i - I_j|/2\sigma^2), i,j \in NEB, \tag{15}$$

where NEB is set of pixels in neighborhood and $\sigma = 0.5$ for the experiments.

5 Performance Evaluation

In this section, we evaluate the performance of our method on Berkeley [15] and Weizmann [13] databases. Some segmentation results are illustrated in Fig. 2.

Table 1. Performance comparison of our method with other segmentation methods: F-measures of our method and 4 state-of-the-art segmentation algorithms by evaluating them on the Weizmann single object database.

Methods	F-measure(%)	Remarks
[6] With auto-context cues	0.91±0.013	Automatic
Proposed Framework	0.89±0.002	Automatic
[6] Without auto-context cues	0.88±0.011	Automatic
[12] Unified approach	0.87±0.011	interactive
[13] Cues integration	0.86±0.012	Automatic
GMM+Initial	0.86±0.011	Automatic
Label Propagation+Initial	0.85±0.012	Automatic
[14] Texture segmentation	0.83±0.016	Automatic

Empirical results show that our method can extract salient object efficiently, and is able to deal with the images with weak boundary or complex background. The F-measure score ($F = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision}$) is computed as well for objective comparison on Weizmann single object segmentation database, and the result is listed in Table 1. We compare our method with four state-of-the-art methods and the F-measure scores of these methods are quoted from [6]. Based on the initial seeds (described in Section 3.3), the scores of segmentation results using CRF (described in Section 4), by GMM classifier in eq. (8) and by Label Propagation classifier in eq. (11) are presented in Table 1 as well. It is noticed that the F-measure score of our method (using CRF) outperforms all the baselines except for the result of [6], which applies the context cues. The performance of our method can be further improved by integrating more discriminative features or refined iteratively.

6 Conclusion

In this paper, we have proposed a framework to extract salient objects from images automatically. Firstly, we propose a saliency model to estimate the initial object region exploiting the low level color features and prior information. Secondly, the seeds are generated automatically by classifiers composition to obtain more precise initial region. Finally, a CRF model is constructed to extract the salient objects. Experimental results show that the proposed method can achieve better performance than baselines on some popular segmentation benchmarks. In future work, we will explore how to incorporate high-level classifiers into the proposed segmentation model.

Acknowledgments. This research is partly supported by NSFC, China (No: 61273258, 31100672)Ph.D. Programs Foundation of Ministry of Education of China (No.20120073110018).

References

1. Boykov, Y.Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in nd images. In: 2001 Eighth IEEE International Conference on Computer Vision (ICCV), vol. 1, pp. 105–112. IEEE (2001)
2. Achanta, R., Susstrunk, S.: Saliency detection using maximum symmetric surround. In: 2010 17th IEEE International Conference on Image Processing (ICIP), pp. 2653–2656. IEEE (2010)
3. Rahtu, E., Kannala, J., Salo, M., Heikkilä, J.: Segmenting salient objects from images and videos. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part V. LNCS, vol. 6315, pp. 366–379. Springer, Heidelberg (2010)
4. Jung, C., Kim, B., Kim, C.: Automatic segmentation of salient objects using iterative reversible graph cut. In: 2010 IEEE International Conference on Multimedia and Expo (ICME), pp. 590–595. IEEE (2010)
5. Liu, Z., Shi, R., Shen, L., Xue, Y., Ngan, K.N., Zhang, Z.: Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut. IEEE Transactions on Multimedia 14(4), 1275–1289 (2012)
6. Xue, J., Wang, L., Zheng, N., Hua, G.: Automatic salient object extraction with contextual cue and its applications to recognition and alpha matting (2013), `http://www.sciencedirect.com/science/article/pii/S0031320313001581`
7. Park, K.T., Moon, Y.S.: Automatic extraction of salient objects using feature maps. In: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 1, p. I–617. IEEE (2007)
8. Fu, K., Gong, C., Yang, J., Zhou, Y.: Salient object detection via color contrast and color distribution. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) ACCV 2012, Part I. LNCS, vol. 7724, pp. 111–122. Springer, Heidelberg (2013)
9. Wei, Y., Wen, F., Zhu, W., Sun, J.: Geodesic saliency using background priors. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 29–42. Springer, Heidelberg (2012)
10. Sethian, J.A.: Fast marching methods. SIAM Review 41(2), 199–235 (1999)
11. Zhou, D., Bousquet, O., Lal, T.N., Weston, J., Schölkopf, B.: Learning with local and global consistency. Advances in Neural Information Processing Systems 16(753760), 284 (2004)
12. Bagon, S., Boiman, O., Irani, M.: What is a good image segment? A unified approach to segment extraction. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 30–44. Springer, Heidelberg (2008)
13. Alpert, S., Galun, M., Basri, R., Brandt, A.: Image segmentation by probabilistic bottom-up aggregation and cue integration. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8. IEEE (2007)
14. Galun, M., Sharon, E., Basri, R., Brandt, A.: Texture segmentation by multiscale aggregation of filter responses and shape elements. In: Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV), pp. 716–723. IEEE (2003)
15. Martin, D.R., Fowlkes, C.C., Malik, J.: Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(5), 530–549 (2004)