Saliency Level Set Evolution

Jincheng Mei[1] and Bao-Liang Lu[1,2,⋆]

[1] Center for Brain-Like Computing and Machine Intelligence
Department of Computer Science and Engineering
Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering
[2] Shanghai Jiao Tong University
800 Dong Chuan Road., Shanghai 200240, China
bllu@sjtu.edu.cn

Abstract. In this paper, we consider saliency detection problems from a unique perspective. We provide an implicit representation for the saliency map using level set evolution (LSE), and then combine LSE approach with energy functional minimization (EFM). Instead of introducing sophisticated segmentation procedures, we propose a flexible and lightweight LSE-EFM framework for saliency detection. The experimental results demonstrate our method outperforms several existing popular approaches. We then evaluate several computation strategies independently. The comparisons results indicate their effectiveness and strong abilities in combatting saliency confusions.

Keywords: Saliency Detection, Level Set Evolutionm, Computer Vision.

1 Introduction

Saliency detection has drawn much attentions in computer vision society. It aims to extract anomalous objects and informative structures from images. Recently, it has been widely employed in computer vision and multimedia applications [1].

In traditional saliency detection, objects are usually extracted through post-processing of saliency map. For example, Hou and Zhang [2] proposes a simple thresholding strategy to extract proto-objects from images. Achanta et al. [3] suggest a widely adopted adaptive thresholding methods one step further. Unlike the methods mentioned above, Gopalakrishnan et al. [4] formulate saliency detection as a foreground and background labeling problem.

In this paper, we notice the interpretability of level set methods, which enable us to readily compute the saliency posterior probability. From this perspective, we propose a novel method, named as Saliency Level Set Evolution (SLSE), for separating salient objects and contexts of an image. As a classical method for shape representation, level set uses the *zero level set* of a 3D level set function (LSF) to represent a closed 2D curve [5]. This implicit representation enables numerical computations without curve and surface parametrization.

⋆ Corresponding author.

C.K. Loo et al. (Eds.): ICONIP 2014, Part II, LNCS 8835, pp. 170–177, 2014.

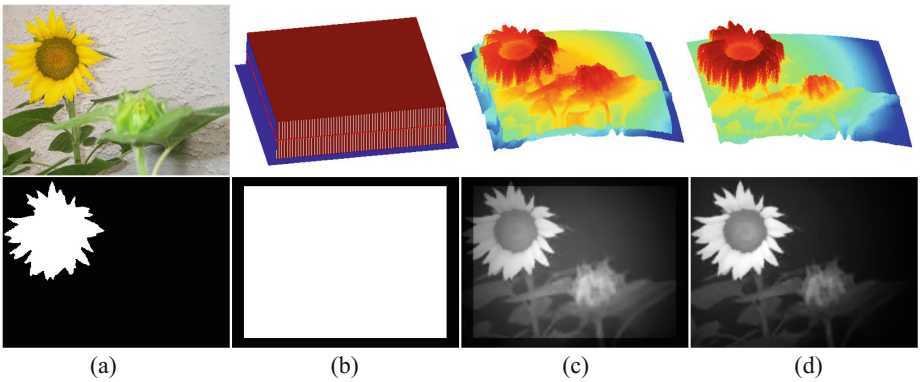

Fig. 1. Saliency level set evolution. (a) An input image and its ground truth; (b) The initial LSF and its binary map. The red contour marks the zero level set; (c) The LSF and binary map after 5 iteration rounds; (d) After 30 rounds.

As shown in Fig.1(b), we firstly initialize the level set with *real number* values. This enables us to evolve the LSF without explicitly determining the binary labels. After several rounds of evolution, we obtain the final LSF as shown in Fig.1(c)-(d). Saliency map is novelly defined as the final normalized LSF.

The main contributions of this paper include:

1) We first provide level set representation for saliency and propose saliency detection method via LSE. This framework is interpretably plausible and provides interfaces for further adaptive improvements.

2) Experimental results show that SLSE outperforms many recently proposed methods in MSAR-1000 [3] and SED1 [6] datasets.

The rest is organized as follows: Section 2 introduces the LSE-EFM formulation. Section 3 demonstrates the experiment results and provides some discussions. Finally, Section 4 gives the conclusion about our work.

2 LSE-EFM Framework

The interpretability of combining LSE with saliency comes from two points: *on the one hand*, because of the ill-posed nature of saliency, it is questionable to directly appoint each pixel a binary label. A more appropriate representation is to utilize a binary random variable to provide a probability description. *On the other hand*, saliency detection can also be viewed as a contour selection *i.e.*, to exactly segment out salient objects is equivalent to select contours which accurately envelope the object regions. Based on these two points, we introduce the level set framework [5].

2.1 Level Set Representation

In our formulation, every pixel $\mathbf{z}$ of an image I is represented as a 5D vector $I(\mathbf{z}) = [G \ P]^\top = [L, a, b, x, y]^\top$ where $G = [L, a, b]$ is the value in the CIE Lab color space [7], and $P = [x, y]$ denotes the position of $\mathbf{z}$. Background and foreground regions are denoted as R_0, R_1, respectively. Using the level set terminology, we get:

$$\begin{cases} \mathbf{z} \in R_0 & \text{if } \phi(\mathbf{z}) < 0, \\ \mathbf{z} \in R_1 & \text{if } \phi(\mathbf{z}) \geq 0, \end{cases} \tag{1}$$

where ϕ is the *level set function* (LSF) [5]. The set $\mathcal{B} = \{\mathbf{z} | \phi(\mathbf{z}) = 0\}$ is called the *zero level set*, which indicates the boundary partitioning of R_0 and R_1. With the mapping rule:

$$\mathcal{C}(\mathbf{z}) = i \quad \text{if } \mathbf{z} \in R_i, \ i \in \{0, 1\}, \tag{2}$$

we get a level set representation based on ϕ instead of explicitly assigning each pixel a binary label.

2.2 Energy Functional Minimization

We firstly review the Bayesian framework [8], in which the saliency $\mathcal{S}$ of a given image I is defined as a posterior probability:

$$\mathcal{S} = p(\mathcal{C}|I), \tag{3}$$

where $\mathcal{C}$ is the salient region matrix with each element a binary random variable. We take the negative logarithm of the post probability and get the energy functional E [9] as:

$$E(\mathcal{C}; I) \sim -\log p(\mathcal{C}|I) \tag{4}$$

$$\sim \underbrace{-\log p(I|\mathcal{C})}_{\text{likelihood term}} \underbrace{-\log p(\mathcal{C})}_{\text{prior term}} \tag{5}$$

where $\log p(I)$ is independent of $\mathcal{C}$ thus omitted. Following [10], the energy functional is defined as a mutual information with a curve length penalty term,

$$E(\mathcal{C}; I) = \underbrace{-|\Omega| \hat{I}(\mathcal{C}; I)}_{\text{data term}} + \underbrace{\alpha \oint_{\mathcal{B}} ds}_{\text{penalty term}} \tag{6}$$

where $|\Omega|$ denotes the number of pixels in I, and $\mathcal{B}$ is a closed boundary. $\hat{I}$ is the mutual information which can be expanded as:

$$\hat{I}(\mathcal{C}; I) = H(I) - \sum_{i=0,1} \frac{|R_i|}{|\Omega|} H(I|\mathcal{C} = i) \tag{7}$$

Note the entropy $H(I)$ is also independent of $\mathcal{C}$ thus dropped. $H(I|\mathcal{C}=i)$ can be approximated by applying the weak law of large numbers and a nonparametric kernel density estimation (KDE) [11],

$$H(I|\mathcal{C}=i) = -\frac{1}{|R_i|}\int_{R_i} \log \hat{p}_i(I(\mathbf{z}))d\mathbf{z} \tag{8}$$

$$= -\frac{1}{|R_i|}\int_{R_i} \log\left(\int_{R_i} K(I(\hat{\mathbf{z}}) - I(\mathbf{z}))d\hat{\mathbf{z}}\right)d\mathbf{z} \tag{9}$$

where $\hat{p}_i(I(\mathbf{z})) \triangleq p(I(\mathbf{z})|\mathbf{z} \in R_i)$ indicates the likelihood of observing $I(\mathbf{z})$ when the pixel $\mathbf{z}$ belongs to region R_i, $\hat{p}_i(I(\mathbf{z}))$ is estimated employing a fast improved Gaussian transform [12], and $K(x) \propto \exp\{-\frac{x^2}{2\sigma^2}\}$ is a Gaussian kernel function.

2.3 Gradient Flow

Since both the energy explanation and the contour explanation are reasonable in the level set framework, the energy functional can be denoted as $E(\mathcal{B})$ as well as $E(\mathcal{C}; I)$. Thus the gradient flow of $E(\mathcal{B})$ in (4) is

$$\frac{\partial E(\mathcal{B})}{\partial t} = \left[\log\frac{\hat{p}_0(I(\mathbf{z}))}{\hat{p}_1(I(\mathbf{z}))} - \alpha\kappa\right]\boldsymbol{N} = v\boldsymbol{N}, \tag{10}$$

where $\boldsymbol{N} = \nabla\phi/|\nabla\phi|$ is the outward unit normal vector of $\mathcal{B}$, and $\kappa = \nabla\cdot(\nabla\phi/|\nabla\phi|)$ is the curvature of $\mathcal{B}$, which is defined as the divergence of $\boldsymbol{N}$.

2.4 Saliency Computation

As shown in Fig.1(b), we can initialize the LSF as a central box function. An alternative is to moderately narrowing the scope by thresholding. This leads to faster evolution. For thresholding, we adopt OTSU method [13] here.

During the evolution procedure, there exists different strategies. Recall in Section 2.1, every pixel is represented as a 5D vector. Actually, luminance contrast does not contribute as much information as color contrast for saliency [14]. This leads to weight tuning for luminance, *i.e.*, suppressing the luminance weight during the evolution. Another cue is that salient objects generally favor more compact spatial distributions than contexts [15]. This leads to weight tuning for position. Totally, the $\hat{p}_i(I(\mathbf{z}))$ in (8) is modified as:

$$\hat{p}_i(I(\mathbf{z})) \triangleq p(I(\mathbf{z})|\mathbf{z} \in R_i) \tag{11}$$

$$= p([\omega_L L(\mathbf{z}), a(\mathbf{z}), b(\mathbf{z}), \omega_{P^i} P(\mathbf{z})]^\top|\mathbf{z} \in R_i) \tag{12}$$

In the experiments, we set $\omega_L = 0.15$, $\omega_{P^0} = 0$, and $\omega_{P^1} = 0.25$. In each iteration, gradient flow (10) is employed without step searching:

$$\phi(k+1) = \phi(k) - \Delta\cdot v \tag{13}$$

where $\Delta = 10$ empirically is the step rate. We set the maximum iteration number as 30, since the convergence is generally very fast in the experiments. Finally, we normalize the LSF to obtain the saliency map.

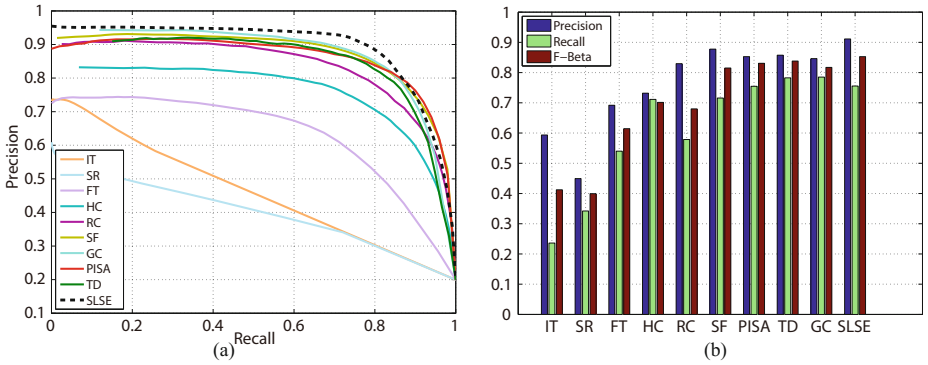

Fig. 2. (a) Average PR curves of different approaches on the MSRA-1000 dataset; (b) Average precision, recall, and F_β-measure using the adaptive thresholding. The proposed SLSE method achieves the best performance.

3 Experiments

3.1 Performance Evaluation

We evaluate our SLSE approach on two benchmarks. One is MSRA-1000, which contains 1000 natural images selected from the MSRA dataset [16] and accurate human-labeled ground truth [3]. The other is SED1 [6]. We compare SLSE with nine baselines following the selection criteria: citation number (IT [17], SR [2], FT [3]), recency (PISA [18], TD [19], GC[20]), and relation with our approach (HC, RC [1], SF [15], PISA [18]).

In particular, HC, RC [1] focus on color contrast, SF [15] proposed spatial distribution cue, and PISA [18] ensemble color and spatial feature. The proposed SLSE method combines color, spatial and luminance information.

Following [3], we evaluate the performance of each method using two metrics: the Precision-Recall (PR) curve and the F_β-measure where $\beta^2 = 0.3$ as suggested in [3]. In the first evaluation, each saliency map is segmented by thresholds varying within $[0, 255]$ and then compared to the ground truth to get PR values. An average PR curve is generated for each method. In the second evaluation, each saliency map is segmented by an adaptive threshold. The PR values is used to calculate their harmonic mean value F_β-score.

Fig.2 and Fig.3 demonstrate the comparisons on MSRA-1000 and SED1, respectively. As shown in Fig.2, the proposed approach outperforms most methods, including recently proposed HC, RC [1], TD [19] and GC [20]. Notably, SLSE performs better than SF [15] and PISA [18] within a wide (near 90%) recall range, and has comparable performance at high recall rates. While on the SED1 dataset, SLSE also achieves better performance than PISA in both metrics.

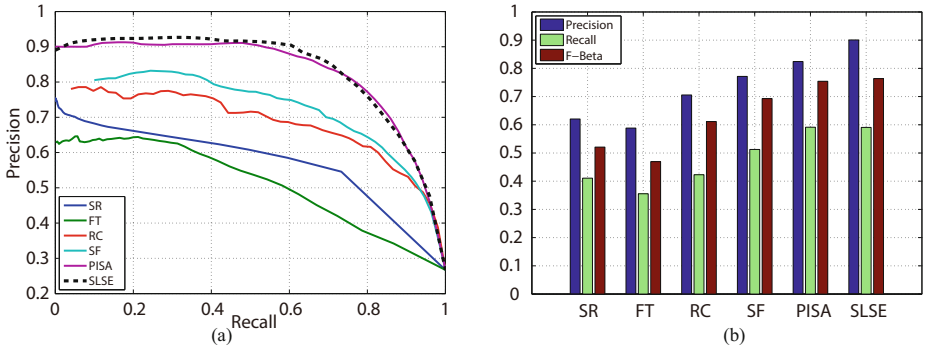

Fig. 3. Evaluation results on the SED1 dataset

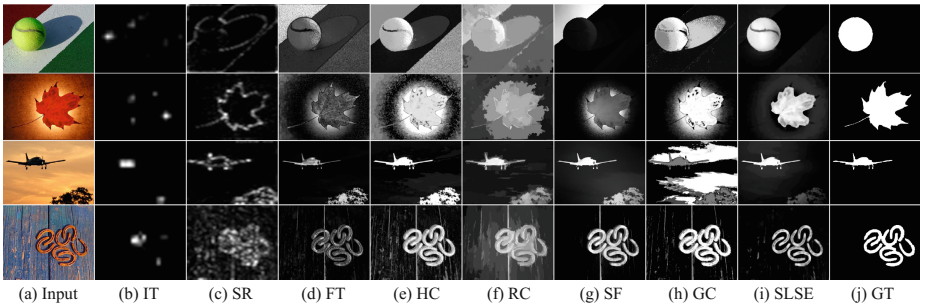

Fig. 4. Visual comparison of the existing approaches to our method (SLSE) and ground truth (GT). Here we compare with IT [17], SR [2], FT [3], HC, RC [1], SF [15] and GC [20]. PISA [18] and TD [19] are not included since there is no public implementation or result. SLSE consistently generates saliency maps closest to ground truth (GT). See the text for discussion.

3.2 Visual Comparison

Fig.4 presents a visual comparison. SLSE provides the best results overall. In particular, we notice that for images with compact bright or shadow regions, such as the first two examples, all of the existing methods used fail to detect the object or falsely highlight contexts while SLSE extracts the object accurately. This validates the consideration about weight tuning for luminance. The third example contains two seemingly similar objects. Only SLSE correctly highlight the ground truth object, attributing to the different constraints on object and background. The last image have relatively more salient context. Most methods falsely detect the vertical line as a salient object while SLSE not, owing to the benefits of level set method, which can hold the LSF contour close to the natural object boundaries.

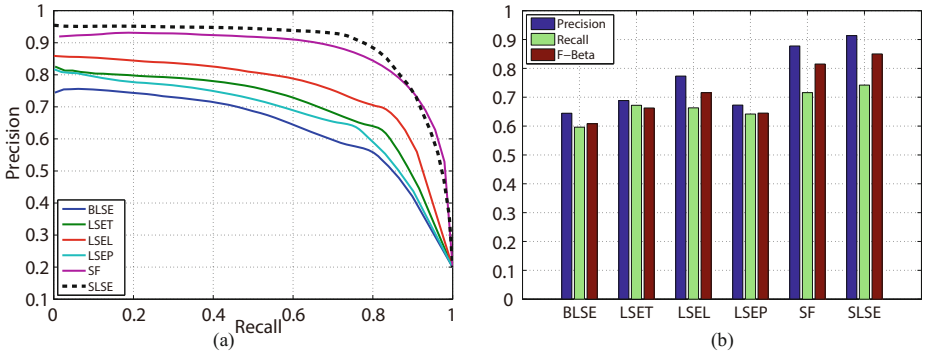

Fig. 5. Experimental comparison of different methods on the MSRA-1000 dataset. (a) PR curves; (b) Average precision, recall, and F_β-measure.

3.3 Strategy Comparison

We briefly introduce several computation strategies in Section 2.4. For completeness, we experimentally verify their effects. In Fig.5, we compare results of employing different strategies on the MSRA-1000 dataset. BLSE denotes basic LSE without using any strategy; LSET denotes using only thresholding; LSEL denotes using only luminance tuning; LSEP denotes using only position tuning; and SLSE represents using all the computation strategies.

From Fig.5, we see that the performance of BLSE is the worst. We notice that utilizing any single strategy only slightly improves the performance. And the best result is achieved by SLSE (outperforms the baseline SF [15] method). Following the comparison, we observe that the strategies assist LSE and are complementary to each other, providing effective detection cues to visual saliency.

3.4 Discussion

The LSE framework can easily ensemble different cues for saliency detection, and the computation is independent of the feature dimensionality [12]. We exploit color, luminance and spatial information here, all of which are low-level features. Some high-level features, such as shape, text and face should be prudently considered. How to ensemble both high-level and low-level information in our framework is an open question.

4 Conclusions

In this paper, we have proposed a flexible LSE-EFM framework for saliency detection. The proposed framework is different from the existing methods and provides a level set representation for saliency and transforms the evolution into an EFM problem. The proposed SLSE method is extensively evaluated on two public datasets and achieves good performance. We have further validated the computation strategies and the results demonstrate their effectiveness.

Acknowledgments. This work was partially supported by the National Natural Science Foundation of China (Grant No. 61272248), the National Basic Research Program of China (Grant No. 2013CB329401), and the Science and Technology Commission of Shanghai Municipality (Grant No. 13511500200).

References

1. Cheng, M.M., Zhang, G.X., Mitra, N.J., Huang, X., Hu, S.M.: Global contrast based salient region detection. In: CVPR (2011)
2. Hou, X., Zhang, L.: Saliency detection: A spectral residual approach. In: CVPR (2007)
3. Achanta, R., Hemami, S., Estrada, F., Süsstrunk, S.: Frequency-tuned salient region detection. In: CVPR (2009)
4. Gopalakrishnan, V., Hu, Y., Rajan, D.: Random walks on graphs to model saliency in images. In: CVPR (2009)
5. Osher, S., Fedkiw, R.: Level set methods and dynamic implicit surfaces, vol. 153. Springer (2003)
6. Alpert, S., Galun, M., Basri, R., Brandt, A.: Image segmentation by probabilistic bottom-up aggregation and cue integration. In: CVPR (2007)
7. Hunt, R.W.G., Pointer, M.R.: Measuring colour. John Wiley & Sons (2011)
8. Zhang, L., Tong, M.H., Marks, T.K., Shan, H., Cottrell, G.W.: Sun: A bayesian framework for saliency using natural statistics. Journal of Vision 8(7) (2008)
9. Chang, J., Fisher, J.: Efficient mcmc sampling with implicit shape representations. In: CVPR (2011)
10. Kim, J., Fisher III, J.W., Yezzi, A., Çetin, M., Willsky, A.S.: A nonparametric statistical method for image segmentation using information theory and curve evolution. IEEE Trans. Image Process. 14(10), 1486–1502 (2005)
11. Parzen, E.: On estimation of a probability density function and mode. The Annals of Mathematical Statistics 33(3), 1065–1076 (1962)
12. Morariu, V.I., Srinivasan, B.V., Raykar, V.C., Duraiswami, R., Davis, L.S.: Automatic online tuning for fast gaussian summation. In: NIPS (2008)
13. Otsu, N.: A threshold selection method from gray-level histograms. Automatica (1975)
14. Einhäuser, W., König, P.: Does luminance-contrast contribute to a saliency map for overt visual attention? European Journal of Neuroscience 17(5), 1089–1097 (2003)
15. Perazzi, F., Krahenbuhl, P., Pritch, Y., Hornung, A.: Saliency filters: Contrast based filtering for salient region detection. In: CVPR (2012)
16. Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.-Y.: Learning to detect a salient object. IEEE Trans. Patt. Anal. and Mach. Intell. 33(2), 353–367 (2011)
17. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Patt. Anal. and Mach. Intell. 20(11), 1254–1259 (1998)
18. Shi, K., Wang, K., Lu, J., Lin, L.: Pisa: Pixelwise image saliency by aggregating complementary appearance contrast measures with spatial priors. In: CVPR (2013)
19. Scharfenberger, C., Wong, A., Fergani, K., Zelek, J.S., Clausi, D.A.: Statistical textural distinctiveness for salient region detection in natural images. In: CVPR (2013)
20. Cheng, M.-M., Warrell, J., Lin, W.-Y., Zheng, S., Vineet, V., Crook, N.: Efficient salient region detection with soft image abstraction. In: ICCV (2013)