ICIC Express Letters
Part B: Applications
Volume **7**, Number **11**, November **2016**

 ISSN 2185-2766
pp. **2491–2496**

AUTOMATIC SALIENT OBJECT SEGMENTATION VIA SHAPE PRIOR BASED ACTIVE CONTOUR MODEL

Shangbing Gao[1,2,3], Youdong Zhang[1], Jun Zhou[1] and Hao Zheng[3]

[1]The Key Laboratory for Traffic and Transportation Security of Jiangsu Province
Faculty of Computer and Software Engineering
[2]Jiangsu Provincial Key Laboratory for Advanced Manufacturing Technology
Huaiyin Institute of Technology
No. 1, Meicheng Rd., Huai'an 223003, P. R. China
luxiaofen_2002@126.com

[3]Key Laboratory of Trusted Cloud Computing and Big Data Analysis
Nanjing Xiaozhuang University
No. 3601, Hongjing Ave., Jiangning Dist., Nanjing 211171, P. R. China

Received June 2016; accepted September 2016

Abstract. *In this paper, we propose a novel model for unsupervised segmentation of viewer's attention object from natural images based on localizing region-based active contour (LRAC). Firstly, we proposed the saliency detection model via the multi-scale superpixel. Then, object-level shape prior is extracted combining saliency with object boundary information. Finally, this contour is improved by the edge-preserving to generate the initial contour for our automatic object segmentation system. In contrast with localizing region-based active contours that require considerable user interaction, the proposed method does not require it, i.e., the segmentation task is fulfilled in a fully automatic manner. Extensive experiments results on a large variety of natural images demonstrate that our algorithm consistently outperforms the popular existing salient object segmentation methods, yielding higher precision and better recall rates.*
Keywords: Active contour, Level set, Saliency, Image segmentation

1. **Introduction.** Object segmentation is one of the most important and challenging issues in image analysis and computer vision research. Most existing object segmentation systems adopt interaction-based paradigms [1,2], i.e., users are asked to provide segmentation cues manually and carefully.

Although the interaction-based methods are promising, they all pose a critical problem in which they need the users' semantic intention. Such manual labeling is time consuming and often infeasible. Thus, additional interactions are necessary when the seeds are not accurately provided. Specially, localizing region-based active contour (called LRAC) [3] is exactly one of the classic interaction-based methods. Segmentation results heavily depend on the initial contour selection. For this reason, developing a sophisticated fully automatic object segmentation method has been strongly demanded. By observing the fact that, under most circumstances, the salient parts of an image are usually consistent with interesting objects to be segmented; therefore, salient regions have been attempted to estimate. In contrast with existing interaction-based approaches that specify the object and background seeds by manual labeling, some methods (e.g., Liu et al.'s method [4], Achanta et al.'s method [5]) determine the seed locations based on the visual attention model. Since the accuracy of the visual attention model plays a crucial role in object segmentation, these algorithms also depend on the quality of the chosen saliency map.

To remedy such shortcoming, we pay close attention to salient object edge points rather than the saliency map itself. After the salient object edge points were detected, the region which is constrained by these corner points will be obtained. The boundary of this region

is close to the object edge. Thereby, the boundary of this region is used as the initial contour of LRAC model.

2. Related Work.

2.1. Localizing region-based active contour model (LRAC).

Shawn and Tannenbaum [3] proposed a natural framework that allows any region-based segmentation energy to be re-formulated in a local way. $B(x, y)$ is used to mask local regions. This function $B(x, y)$ will be 1 when the point y is within a ball of radius centered at x, and 0 otherwise.

The energy of this model is given as follows

$$E(\phi) = \int_{\Omega_x} \delta\phi(x) \int_{\Omega_y} B(x,y)(u_x - v_x)^2 dy dx \tag{1}$$

where u_x and v_x represent the intensity means in the interior and exterior of the contour localized by $B(x, y)$ at the point x. $\delta\phi(x)$ is given by

$$\delta\phi(x) = \begin{cases} 1 & \phi(x) = 0 \\ 0 & |\phi(x)| < \varepsilon \\ \frac{1}{2\varepsilon}\left\{1 + \cos\left(\frac{\pi\phi(x)}{\varepsilon}\right)\right\}, & \text{otherwise} \end{cases} \tag{2}$$

Minimizing the energy functional $E(\phi)$ with respect to ϕ, we derive the gradient descent flow

$$\begin{aligned} \frac{\partial\phi}{\partial t} =& \delta\phi(x) \int_{\Omega_y} B(x,y)\delta\phi(y) \cdot \left(\frac{(I(y) - u_x)^2}{A_u} - \frac{(I(y) - v_x)^2}{A_v}\right) dy \\ &+ \lambda\delta\phi(x) div\left(\frac{\nabla\phi(x)}{|\phi(x)|}\right) \end{aligned} \tag{3}$$

where $\lambda \geq 0$ is a fixed parameter, and A_u and A_v are the areas of the local interior and local exterior regions respectively given by

$$A_u = \int_{\Omega_y} B(x,y) \cdot H\phi(y) dy \tag{4}$$

$$A_v = \int_{\Omega_y} B(x,y) \cdot H(1 - \phi(y)) dy \tag{5}$$

$$H(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \tag{6}$$

Since the energy functional $E(\phi)$ is non-convex, it may introduce many local minimums. Moreover, (3) is not a global minimiser. Therefore, such an iterative process is prone to be local minima.

2.2. The saliency detection models.

Visual attention analysis has generally progressed on two fronts: bottom-up and top-down approaches. Thus, a number of very inspiring and mature saliency models have been recently introduced in the literature. Itti et al. [6] introduced a saliency model which was biologically inspired. Based on Itti's algorithm, many saliency models have appeared, such as, AC [7], graph-based visual saliency (GB) [8], frequency-tuned (FT) [5], saliency residual (SR) [9], CA [10], MZ [11], histogram contrast (HC) [12], and region contrast (RC) [12].

3. The Active Contour Model Based on Shape Prior.

In this section, we present in detail the proposed initial contour of LRAC. Firstly, the saliency detection method is proposed, and the object-level shape prior is then extracted combining saliency with object boundary information.

3.1. **Superpixel dissimilarity based saliency detection.** To get the superpixels, we use an adaptation of SLIC superpixels. SLIC superpixels segment an image using K-means clustering in RGBXY space. The RGBXY space yields local, compact and edge aware superpixels. Denote the superpixels as p_i, $i = 1, 2, \ldots, L$.

A superpixel is salient if the color of its pixels is unique. We should not, however, look at an isolated superpixel, but rather at its surrounding superpixels, which lead to a center-surrounding contrast. Thus, a superpixel p_i is considered salient if the appearance of the superpixel p_i is distinctive with respect to all other image superpixels. The color distance between two superpixels p_1 and p_2 is defined as:

$$dist_{color}(p_1, p_2) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} f(c_{1,i}) f(c_{2,j}) D(c_{1,i}, c_{2,j}) \tag{7}$$

where $f(c_{k,i})$ is the probability of the i-th color $c_{k,i}$ among all n_k colors in the k-th region p_k, $k = \{1, 2\}$, and $D(c_{1,i}, c_{2,j})$ is the color distance metric between pixels $c_{1,i}$ and $c_{2,j}$ in the $L * a * b$ space.

The positional distance between superpixels is also an important factor. Generally speaking, background superpixels are likely to have many similar patches both near and far-away in the image. However, the salient foreground superpixels tend to be grouped together. This implies that a superpixel p_i is salient when the superpixels similar to it are nearby, and it is less salient when the resembling superpixels are far away. Let $dist(p_i, p_j)$ be the Euclidean distance between the positions of superpixels p_i and p_j, which are represented by Euclidean distance between their centroids in the original image, and normalized by the larger image dimension. Based on the observations above, we define a dissimilarity measure between a pair of superpixels p_i and p_j as:

$$dissimilarity(p_i, p_j) = \frac{dist_{color}(p_i, p_j)}{1 + dist(p_i, p_j)} \tag{8}$$

A superpixel p_i is salient when $dissimilarity(p_i, q_k)$ is high for $\forall k \in [1, K]$. The saliency of superpixel p_i is defined as (we choose $K = 100$ in our experiments):

$$S_i = 1 - \exp\left\{-\frac{1}{K} \sum_{k=1}^{K} dissimilarity(p_i, q_k)\right\} \tag{9}$$

3.2. **The initial contour via shape prior.** In this section, we show how to extract shape prior, i.e., a salient closed contour, which combines saliency with boundary information. Our goal is to extract a closed contour C^*, which covers the salient object. Specifically, as shown in Figure 1, we first construct an edge map. The edge map consists of a set of line segments as illustrated in Figure 1(c), which are obtained from an edge detector, followed by a line fitting step. We refer to straight line segments as detected

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

FIGURE 1. An illustration of shape prior extraction: (a) input image, (b) saliency map, (c) detected line segments using Pb edge detector [14], (d) the initial contour, where white line means detected segments and gap-filling segments, (e) binary segmentation result

segments. Note that a detected segment may come from the boundary of the salient object, or the noise and texture of the object and background.

Our shape prior extraction can then be formalized to find an optimal closed contour C^* by identifying a subset of detected segments in E and connecting them together. Since the detected segments are disjoint, we construct additional line segments that fill the gaps between detected segments to form closed contours. We refer to these as gap-filling segments. Without knowing which gaps are along the resulting optimal contour, we construct a gap-filling segment between each possible pair of the endpoints of the different detected segments. In this way, a closed contour is defined as a cycle that traverses a set of detected and gap-filling segments alternately, as shown in Figure 1(d). The optimal closed contour C^* can be defined as:

$$C^* = \arg\min_C \frac{|C_G|}{\sum_{p\in C} S_m(p)} \tag{10}$$

where $|C_G|$ is the total length of gaps along the contour C, and $\sum_{p\in C} S_m(p)$ is the total saliency value of pixels located inside C. Ratio contour algorithm [13] can be employed to find such an optimal cycle in polynomial time which is used as the initial contour of LRAC.

4. **Experiments.** The experiments are implemented with Matlab code run on a Dell Dimension 4600 PC, with Pentium 4 processor, 2.80 GHz, 1 GB RAM, with Matlab 7.0 on Windows XP.

We perform experiments on the dataset which is provided by Achanta et al. in [5] which contains 1000 images, along with ground truth for each image in the form of accurate human-labeled masks for salient object.

To smooth the computed superpixels, we first merge those neighbouring regions whose d_{color} is less than 0.2, and d_{color} is the distance between the CIE L*a*b* histograms of two regions. To construct the edge map, we use the Pb edge detector [14].

Firstly, to measure the segmentation performance of our proposed algorithm, active contour model based on shape prior (AS) comprehensively, we compare AS algorithm with the Grabcut [15] algorithm using more saliency maps, i.e., the abovementioned 9 state-of-the-art saliency maps. To automatically initialize GrabCut, we use a segmentation obtained by binarizing the saliency map using a fixed threshold. We set the threshold to 0.3 empirically. Once initialized, we iteratively run GrabCut 4 times to improve the segmentation result. Figure 2 shows the comparison results. Here, we use the precision,

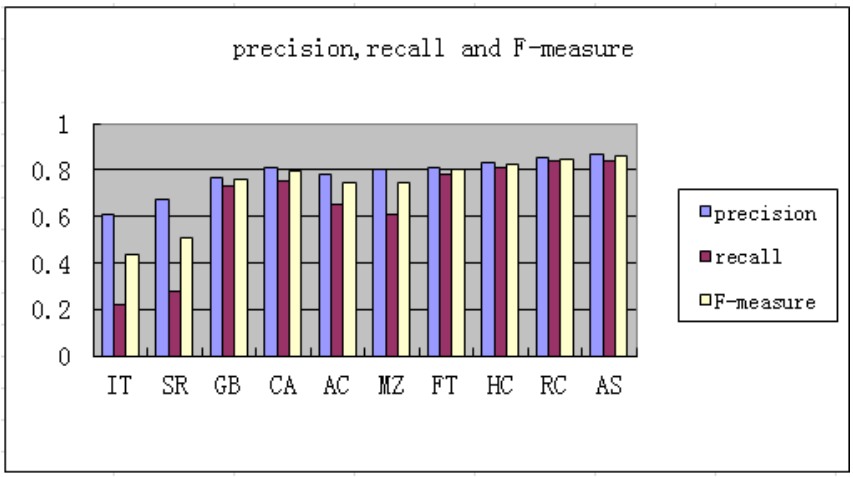

FIGURE 2. Precision-recall bars for the proposed algorithm and the Grabcut using different saliency maps. Our method AS shows high precision, recall, and F-measure values over the 1000-image database.

recall, and F-measure to evaluate the performance of our proposed model. It is shown in the figure that the proposed method significantly outperforms the abovementioned 9 models with respect to precision, recall, and F-measure.

We secondly measure the segmentation performance of the proposed algorithm, as compared with existing competitive automatic salient object segmentation methods, such as Liu et al.'s method [4] and Achanta et al.'s segmentation method [5]. Figure 3 shows the segmentation performance of the three methods. It is shown in the figure that the proposed method is more effective than the other two state-of-the-art algorithms.

FIGURE 3. Results of object segmentation. The leftmost is the original image. The segmentation results from the second left to right are obtained from [4,5] and the proposed algorithm.

5. **Conclusions.** In this paper, we propose superpixel-based saliency and object-level shape prior computation. Saliency map is computed based on multi-scale superpixels. And object-level shape prior is extracted combining saliency with object boundary information. We then integrate both of them into an active contour model framework, leading to binary segmentation of the salient object, where shape prior encourages segmentation boundary to be aligned with salient contour. In the future, we will proceed to extend the proposed model and its corresponding algorithms to the video object segmentation.

Acknowledgment. This work is sponsored by the National Natural Science Foundation of China (NSFC) #61402192, JiangSu Qing Lan Project, the open fund for the Key Laboratory for Traffic and Transportation Security of Jiangsu Province (TTS2015-05), the open fund of Jiangsu Provincial Key Laboratory for Advanced Manufacturing Technology (HGAMTL-1401), the Natural Science Foundation of Jiangsu Higher Education Institutions of China under Grant 14KJB580002, and six talent peaks project in Jiangsu Province (XYDXXJS-011). Prospective joint research projects of Jiangsu Province (BY2016061-01), the Science & Technology Fund of Huai'an under the Grants No. HAC2015023, HAG2015048.

REFERENCES

[1] S. Gao, J. Yang and Y. Yan, Saliency based localizing active contour for automatic natural object segmentation, *IET Image Processing*, vol.7, no.9, pp.787-794, 2013.

[2] Z. Liu, R. Shi, L. Shen, Y. Xue and K. N. Ngan, Unsupervised salient object segmentation based on Kernel density estimation and two-phase graph cut, *IEEE Trans. Multimedia*, vol.14, no.4, pp.1275-1289, 2012.

[3] L. Shawn and A. Tannenbaum, Localizing region-based active contours, *IEEE Trans. Image Processing*, vol.17, pp.2029-2039, 2008.

[4] X. Liu, S. Peng, Y. Cheung, Y. Tang and J. Du, Active contours with a joint and region-scalable distribution metric for interactive natural image segmentation, *IET Image Processing*, vol.8, no.12, pp.824-832, 2014.
[5] R. Achanta, S. Hemami, F. Estrada and S. Susstrunk, Frequency tuned salient region detection, *Proc. of IEEE Computer Vision and Pattern Recognition*, pp.1597-1604, 2009.
[6] L. Itti, C. Koch and E. Niebur, A model of saliency-based visual attention for rapid scene analysis, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.20, no.11, pp.1254-1259, 1998.
[7] R. Achanta, F. Estrada, P. Wils and S. Süsstrunk, Salient region detection and segmentation, *Proc. of International Conference of Computer Vision Systems*, pp.66-75, 2008.
[8] M. Zhao, L. Jiao, W. Ma, H. Liu and S. Yang, Classification and saliency detection by semi-supervised low-rank representation, *Pattern Recognition*, vol.51, pp.281-294, 2016.
[9] J. Wang, A. Borji, C. C. Kuo and L. Itti, Learning a combined model of visual saliency for fixation prediction, *IEEE Trans. Image Processing*, vol.25, no.4, p.1, 2016.
[10] S. Goferman, L. Zelnik-Manor and A. Tal, Context-aware saliency detection, *Proc. of IEEE Computer Vision and Pattern Recognition*, pp.12-20, 2010.
[11] E. Vig, M. Dorr and D. Cox, Large-scale optimization of hierarchical features for saliency prediction in natural images, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp.2798-2805, 2014.
[12] M. Cheng, N. Mitra and X. Huang, Salient object detection and segmentation, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.34, no.10, pp.1915-1926, 2012.
[13] S. Wang, T. Kubota, J. Siskind and J. Wang, Salient closed boundary extraction with ratio contour, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.27, no.4, pp.546-561, 2005.
[14] Q. Wei, D. Feng and H. Xie, Edge detector of SAR images using crater-shaped window with edge compensation strategy, *IEEE Geoscience & Remote Sensing Letters*, vol.13, no.1, pp.38-42, 2016.
[15] M. Cheng, V. Prisacariu, S. Zheng, P. Torr and C. Rother, DenseCut: Densely connected CRFs for realtime GrabCut, *Computer Graphics Forum*, vol.34, no.7, pp.193-201, 2015.