LETTER

Visual Attention Guided Multi-Scale Boundary Detection in Natural Images for Contour Grouping

Jingjing ZHONG†(a), Student Member, Siwei LUO(b), Member, and Qi ZOU†(c), Nonmember

SUMMARY Boundary detection is one of the most studied problems in computer vision. It is the foundation of contour grouping, and initially affects the performance of grouping algorithms. In this paper we propose a novel boundary detection algorithm for contour grouping, which is a selective attention guided coarse-to-fine scale pyramid model. Our algorithm evaluates each edge instead of each pixel location, which is different from others and suitable for contour grouping. Selective attention focuses on the whole saliency objects instead of local details, and gives global spatial prior for boundary existence of objects. The evolving process of edges through the coarsest scale to the finest scale reflects the importance and energy of edges. The combination of these two cues produces the most saliency boundaries. We show applications for boundary detection on natural images. We also test our approach on the Berkeley dataset and use it for contour grouping. The results obtained are pretty good.

key words: boundary detection, selective attention, global spatial prior, contour grouping

1. Introduction

Boundary is an important and fundamental cue for perceptual grouping, shape recognition and other middle and high visual tasks. It is a contour in the image plane that represents changes in pixel ownership from one object or surface to another [1]. The result of boundary detection is the input of contour grouping [10], [11], and initially affects the performance of a grouping algorithm. It is a pity that grouping algorithms pay little attention to the inputs, although human intervention is sometimes used in order to improve grouping performance. Here the results of our model are used for contour grouping especially.

The most famous algorithm of boundary detection is canny edge detector [3]. It looks for local discontinuities in image brightness, and locates the edges by the maximum gradients. It has good performance on edge location, but prefers neither boundaries nor noises. In order to get rid of noise, other cues are used in combination with brightness. Joshi [4] uses gradient computation and surround suppression to detect contours, while Lee [7] detects salient contours by orientation energy distribution. Martin etc. have done outstanding works. They provide an empirical basis for research on image segmentation and boundary detection, which called the Berkeley segmentation dataset, and have developed new boundary detection algorithms [1] on the dataset. They combine local brightness, color, and texture to compute the posterior probability of a boundary existing at each image location and orientation. Global or middle and high level information has also been used to improve performance. [8] brings in global information on the basis of [1], while [9] combines the results of image segmentation and gestalt rules. All of these algorithms perform quite well on the benchmark, but they all treat boundaries on the pixel scale and do not take the edge as a whole.

In this paper we propose a novel multi-scale boundary detection algorithm for contour grouping using selective attention guided global spatial prior information. Different from other algorithms, our model is selective attention guided and treats boundaries on the edge scale, which leads to several advantages. Firstly, selective attention guides our attention to the boundary of saliency objects instead of trivial noises. Secondly, using edges instead of pixels is much more appropriate for contour grouping. And thirdly, the existent period of an edge crossing scales pops out important edges naturally. We show applications for boundary detection on natural images. We also test our approach on Berkeley segmentation dataset and do contour grouping based on the results of the model. The performance of our model is pretty good.

The rest of the paper is organized as follows. Section 2 introduces selective attention briefly and describes how we get selective attention guided spatial prior. Single-scale and multi-scale boundary detection are explained in Sect. 3 and 4 respectively in detail. Section 5 presents the experimental results and the performance evaluations. Finally, conclusion appears in Sect. 6.

2. Selective Attention Guided Spatial Prior

Selective attention mechanisms are very effective in human visual system. The most important function of selective visual attention is to direct our gaze rapidly towards objects of interest in our visual environment. Itti and Koch [2] introduce a computational model of visual attention, which generates saliency maps to guide shifts of attention. Interesting objects, especially which are with abundant texture, are highlighted in a saliency map (see Fig. 1 (b)). The region information of salient objects provides spatial prior for boundary, and eliminates details inside of the objects.
2.1 Saliency Region Detection

Saliency map is a combination of different feature maps, including colors, intensity, orientations, symmetric information and so on. In this paper, following [5] we decompose natural images into 4 double-opponent color (red, blue, green, and yellow) pyramids, one intensity pyramid and 4 orientation \( \theta = \left[ 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4} \right] \) pyramids to create feature maps. Saliency map can be calculated according to the following equations:

\[
S(x) = \gamma_C S_C(x) + \gamma_O S_O(x)
\]

\[
S_C(x) = \frac{1}{\sum_{i}^{n} S_C(x, y_i) \cdot g_{\text{gau}}(x, y_i)}
\]

\[
S_O(x) = \frac{1}{\sum_{i}^{n} \tilde{C}_O(x, NHO(k)) \cdot g_{\text{gau}}(k)}
\]

Where \( \gamma_C, \gamma_O \) are the weighting coefficients for the color-intensity salience and orientation salience. \( S_C(x, y) \) is the color-intensity contrast and \( g_{\text{gau}} \) is the Gaussian weighted distance between pixel \( x \) and \( y \). \( \tilde{C}_O(x, NHO(k)) \) is the orientation contrast between \( x \) and its \( k \)-th neighborhood (more details can be found in [5]).

Saliency map is a gray image, and its histogram (see Fig. 1 (c)) generally has two peaks. With this character, we can use Otsu’s method [12] to search for the threshold and separate salient regions (see Fig. 1 (d)) from the saliency map. The threshold is calculated by Eq. (2):

\[
T = \arg \max_t \sigma_t^2
\]

\[
= \arg \max_t \omega_1(t) \omega_2(t) [\mu_1(t) - \mu_2(t)]^2
\]

Where \( \omega_1 \) and \( \mu_1 \) are class probabilities and class means respectively.

2.2 Spatial Prior Computation

The boundaries of salient regions contribute spatial prior to boundary detection. Since the saliency map is a blurred image and the segmentation of salient regions is not rigorous, we cannot get a simple and exactly located boundary of a salient region from itself directly. However, spectral graph theory, particularly, the Normalized Cuts criterion [6] can help us. The boundary of a salient region is located in three steps.

- Do Ncut on the original image and salient region image separately, and segment them into \( N \) regions (we set \( N \) to be 16/8 in our experiments, see Fig. 1 (e)).
- Emerge regions according to the salient regions and inapparent ones on both the two Ncut results.
- Combine these two segmentation results, and the boundary (see Fig. 1 (f)) is exactly the boundary of salient regions.

Spatial prior is decided by the boundary of salient regions. Considering location offset in different scales, the spatial prior is computed by a sub-Gaussian function as in Eq. (3).

\[
sp(x, y) = C(\theta, q) \cdot \exp(-d(x, y)/\theta \cdot \sqrt{q})
\]

\[
d(x, y) = \min_{a, b \in B} \sqrt{(x - a)^2 + (y - b)^2}
\]

Where \( C(\theta, q) \) is the function of \( \theta \) and \( q \):

\[
C(\theta, q) = \frac{q}{2 \theta \Gamma(1/q)}
\]

Where \( \Gamma(\cdot) \) is a Gamma function. In order to give a proper guide, we choose \( \theta = q = 5 \) in our experiments.

3. Single-Scale Boundary Detection

According to comparison in [1], we only use brightness cue to do edge detection in our algorithm, just like what we have done in our previous work.

3.1 Edge Detection

The edges of an image on a scale get from the convolution of the original image and the Gaussian filter of the scale.

\[
R(x, y, \sigma) = G(x, y, \sigma) \otimes I(x, y)
\]

Where \( G(x, y, \sigma) \) is the Gaussian filter of scale \( \sigma \), and \( I(x, y) \) is the brightness image, \( R(x, y, \sigma) \) is the result of convolution.

Edges are the extrema of brightness changes, that is to say the extrema of brightness gradients. So here we use the extremum of the convolution gradient to locate edges.

\[
\begin{align*}
\frac{\partial^2 R(x, y, \sigma)}{\partial \sigma^2} &= 0 \\
\frac{\partial^2 R(x, y, \sigma)}{\partial v_3^2} &< 0
\end{align*}
\]
\[ v = \arctan \frac{R_y}{R_x} \]  \hspace{1cm} (6)

Where \( v \) is the orientation of gradient, \( R_x \) and \( R_y \) stand for first partial derivative.

### 3.2 Edge Evaluation

The probability of an edge belonging to the boundary is decided by two aspects. The first one is the importance of the edge itself, and the second one is the spatial prior information.

\[ p(e) = \alpha l(e) \cdot g(e) + \beta \sum_{(x,y) \in e} sp(x,y) \]  \hspace{1cm} (7)

Where \( l(e) \) is the length of edge \( e \), \( g(e) \) is the mean value of brightness gradients of \( e \), \( sp(x,y) \) is the spatial prior which we calculate in Eq. (3).

### 4. Multi-Scale Boundary Detection

We have an observation that on low-resolution images it becomes possible for a boundary detection algorithm to explore a much larger portion of the boundaries of objects instead of details inside. We expect that at a coarse scale we can pay more attention on the edges with high probability to be a boundary. We generate a pyramid with 5 levels for the images, and do single-scale boundary detection on each image in the pyramid.

#### 4.1 Spatial Prior Updating

We start from the coarsest scale, and updating the spatial prior following Eq. (8) if there is an edge located on the position with high spatial prior.

\[ sp(x,y,\sigma_{i+1}) = sp(x,y,\sigma_i) + w \cdot sp_e(x,y,\sigma_i) \]

\[ \text{s.t. } sp(x,y,\sigma_i) > 0 \]  \hspace{1cm} (8)

Where \( sp(\cdot) \) is calculated according to Eq. (3), and \( \sigma_{i+1} \) is the next finer scale to \( \sigma_i \).

#### 4.2 Edge Energy Computation

We use edge energy to evaluate the probability of an edge belonging to the boundary. Edge energy exists in the whole life of an edge, starting from its appearance at a finer scale to its disappearance at a coarser scale.

\[ E(e) = \int_{\sigma_a}^{\sigma_d} \frac{p(e,\sigma) - \overline{p(\sigma)}}{\text{var}(p(\sigma))} \ d\sigma \]  \hspace{1cm} (9)

Where \( \sigma_a \) and \( \sigma_d \) are the scale starting to appear and disappear respectively. \( \overline{p(\sigma)} \) and \( \text{var}(p(\sigma)) \) are the mean and variation of all \( p(e,\sigma) \) at scale \( \sigma \).

Edges with high energy are most probably on the boundary of an image. As a matter of experience, the first 20%~30% edges can remain the primary boundaries. And it is appropriate for contour grouping.

### 5. Experimental Results and Evaluation

All of our test images come from Berkeley segmentation dataset [1] and iLab [2], including natural objects and man-made objects. The size of image differs from 160*240 to 321*481. Our algorithm shows good performance on the images with complex texture objects. We give a few boundary detection results in Fig. 2 comparing with other algorithm.

We apply a precision-recall (PR) framework to evaluate the performance of our model. Human-marked boundaries from the Berkeley segmentation dataset are used as ground truth. Although the result of our algorithm is not the best one, our result is comparable with others, and much better than other algorithms using only one brightness cue. The comparison is shown in Table 1.

![Fig. 2](image.jpg)

**Table 1** Comparison of proposed method with others [13].

<table>
<thead>
<tr>
<th>Method</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Probability of Boundary</td>
<td>0.70</td>
</tr>
<tr>
<td>Ultrametric Contour Maps</td>
<td>0.67</td>
</tr>
<tr>
<td><strong>Our Proposed Method</strong></td>
<td><strong>0.66</strong></td>
</tr>
<tr>
<td>Brightness/Color/Texture Gradients</td>
<td>0.65</td>
</tr>
<tr>
<td>Brightness Gradients</td>
<td>0.62</td>
</tr>
</tbody>
</table>
Since our boundary detection algorithm is especially for contour grouping, we take the results of our algorithm as the inputs of contour grouping (see Fig. 3). It shows that our algorithm can help to improve the performance of the perceptual grouping model introduced in [10] remarkably.

6. Conclusion

We propose a novel algorithm of multi-scale boundary detection for contour grouping. We use the saliency map to provide spatial prior guidance, and update spatial prior scale by scale. We detect boundaries by evaluate edges instead of pixels, which makes our algorithm suitable for contour grouping tasks. Experiments on man-made images and natural images show the good performance of our algorithm. However, limited by the edge detection method, there are small gaps on the boundaries detected by our algorithm. We will do further research to avoid this problem.

Acknowledgments

The research is supported by National High Technology Research and Development Program of China (2007AA01Z168), National Nature Science Foundation of China (60805041, 60872082, 60773016) and Doctoral Foundations of Ministry of Education of China (20050004001).

References