LETTER

Contour Grouping and Object-Based Attention with Saliency Maps

Jingjing ZHONG[†a)], ***Student Member***, **Siwei LUO**[†b)], ***Member***, ***and*** **Jiao WANG**[†c)], ***Student Member***

SUMMARY The key problem of object-based attention is the definition of objects, while contour grouping methods aim at detecting the complete boundaries of objects in images. In this paper, we develop a new contour grouping method which shows several characteristics. First, it is guided by the global saliency information. By detecting multiple boundaries in a hierarchical way, we actually construct an object-based attention model. Second, it is optimized by the grouping cost, which is decided both by Gestalt cues of directed tangents and by region saliency. Third, it gives a new definition of Gestalt cues for tangents which includes image information as well as tangent information. In this way, we can improve the robustness of our model against noise. Experiment results are shown in this paper, with a comparison against other grouping model and space-based attention model.
key words: *contour grouping, object-based attention, saliency map*

1. Introduction

Contour grouping, a branch of perceptual grouping, is a middle level vision task which can identify object boundaries in noisy images, and its results are very useful for high-level vision problems. A whole process of contour grouping mainly includes three steps (Fig. 1): First, a set of edges (tangents) are detected from input images. Second, affinity matrix of the edges or relationship between edges is computed according to Gestalt rules. And third, graph partition, spectral clustering, or some other optimal methods are used to get final results which may be open, closed, symmetric, or convex boundaries.

At first contour grouping methods are mainly used in synthetic images [1]. In these methods, they only consider the edge information, such as proximity, continuity, and sometimes closure. In order to improve the grouping performance for natural images, convexity and symmetry, which are important characteristics of natural objects, are introduced into the contour grouping methods [2], [5]. The images include abundant information, which edges can't tell us. Making use of the information, Stahl further adds region areas enclosed by edges and intensity feature maps extracted from natural images to the methods, and he gets much better grouping results in some natural images [4], [8]. But for these methods, there still remain several questions that may degrade the grouping performance.

Manuscript received June 22, 2009.
Manuscript revised August 5, 2009.
[†]The authors are with School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China, 100044.
a) E-mail: jingjing_zhong@163.com
b) E-mail: swluo@bjtu.edu.cn
c) E-mail: wangjiao0828@163.com
DOI: 10.1587/transinf.E92.D.2531

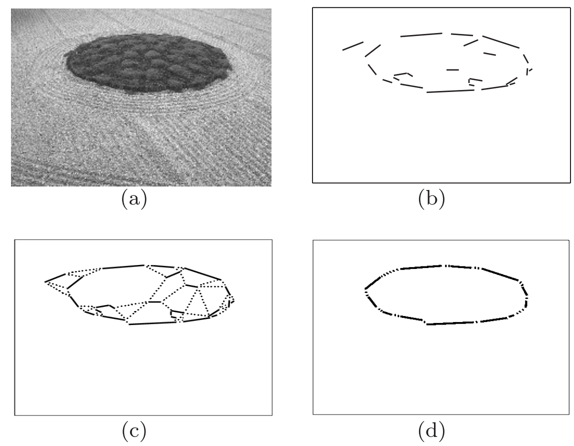

Fig. 1 The process of contour grouping. (a) The original image, (b) the first stage: edge detection, (c) the second stage: affinity computation, (d) the third stage: grouping.

- Most of these contour grouping methods only consider edge information to compute Gestalt cues, and do not consider any image information. For example, it is difficult to define similarity of edges only from edges themselves, while we can use the image color on either side of each edge to define it.
- Most of these contour grouping methods detect multiple boundaries only according to the grouping cost without considering any grouping strategies. The grouping results will be more accordant with human vision when considering visual attention mechanism.

Object-based attention, which corresponds to space-based attention, suggests that visual attention can directly select discrete objects rather than continuous spatial locations within the visual field [6]. The most famous object-based attention model is proposed by Sun [3]. It introduces a hierarchical attention movement model but remains the problem of object definition.

In this paper, we present a new contour grouping model that address these problems. First, we introduce a new definition of Gestalt cues for tangents. In our definition, the relationship between tangents is decided not only by tangents themselves but by image color. Second, we add the saliency map to the grouping cost computation. The grouping results satisfy gestalt rules and visual saliency simultaneously. Third, we expand the grouping model to construct a hierarchical object-based attention model, which can locate perceptual objects and give a shift of object-based attention.

The rest of this paper is organized as follows: Section 2 describes the contour grouping model. Section 3 gives the definition of Gestalt cues in detail, which describe the relationship between tangent pairs. Section 4 describes the construction of Saliency maps. Section 5 introduces an object-based attention model which is formed when the grouping model is repeated in a hierarchical way. Section 6 reports experiment results on natural color images. Section 7 concludes this paper.

2. Contour Grouping for Salient Boundaries

In our paper, we follow the three steps for contour grouping, First, edges are detected from natural images using the method we proposed in [9]. These edges are separated into tangents with a curvature tolerance. And then each tangent is represented by two sibling directed tangents with opposite directions. Second, Gestalt cues are computed between directed tangent pairs. We mainly consider proximity, continuity, and similarity in this paper. These three cues are weighted summed according to the inferential power of these cues. Third, we construct a grouping cost which is decided by Gestalt cues and saliency map. The boundary with the minimum grouping cost is the grouping result of our model.

The grouping cost for a closed boundary is computed following Eq. (1):

$$\mathfrak{C}(B) = \frac{G(B)}{|S(R(B))|} = \frac{\sum_{e_i e_j \in B} (\lambda_1 prox_{ij} + \lambda_2 cont_{ij} + \lambda_3 simi_{ij})}{\left| \int \int_{R(B)} s(x,y) dx dy \right|} \tag{1}$$

where $G(B)$ is the sum value of Gestalt cues of directed tangent pairs which are successive tangents in the boundary. A smaller value means better proximity, continuity and similarity. $R(B)$ is the region enclosed by the closed boundary B, and $S(R(B))$ the salient value of region $R(B)$. $s(x,y)$ represents the salient value of pixel (x,y) in saliency map. A bigger value means more salient. Our goal is to find the boundary that encloses a region with as many salient pixels as possible. The optimization of the grouping cost $\mathfrak{C}(B)$ can follow the ratio contour algorithm proposed in [10].

3. Gestalt Cues Definition

We consider three observable cues that we expect to be most powerful for our grouping method: proximity, continuity and similarity. Observable data relating two directed tangents $\vec{e_i}$ and $\vec{e_j}$ is shown in Fig. 2. More details are discussed in another paper.

- Proximity: A function of the length d_{ij}, the distance between the end point of $\vec{e_i}$ and the start point of $\vec{e_j}$.
- Continuity: A function of the orientation change $(\alpha+\beta)$ between $\vec{e_i}$ and $\vec{e_j}$.

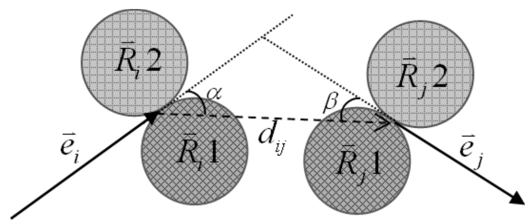

Fig. 2 Observable data relating two directed tangents.

- Similarity: A function of the sum difference in image colors between region $\vec{R_i}1$ and $\vec{R_j}1$, $\vec{R_i}2$ and $\vec{R_j}2$. The difference is measure by χ^2 distance of HSV color histogram.

These three cues are weighted summed to represent the affinity of tangent pairs according to their inferential power [11]. In this paper, the three parameters λ_1, λ_2 and λ_3 in Eq. (1) are set to be 1, 0.37 and 0.62, respectively.

4. Saliency Map Construction

Saliency measures how different a pixel, a region, or an object contrasts with its surroundings, and depends on various factors, such as color, intensity, texture, orientation between the target and its neighbors. A saliency map presents every pixel's saliency in an image, and guides the shift of visual attention. The computation of saliency map can follow [3]. The saliency map is calculated by combining the color, intensity, and orientation salience of the image. We decompose images into 4 double-opponent color (red, blue, green, and yellow) pyramids, one intensity pyramid and 4 orientation ($\theta = [0, \pi/4, \pi/2, 3\pi/4]$) pyramids.

$$\begin{aligned} s(x,y) &= \gamma_{CI} s_{CI}(x,y) + \gamma_O s_O(x,y) \\ s_{CI}(x,y) &= \frac{\sum_i S_{CI}(p,p'_i) \cdot d_G(p,p'_i)}{\sum_i d_G(p,p'_i)} \\ s_O(x,y) &= \frac{\sum_i \widehat{C_O}(p,p'_i) \cdot d_G(p,p'_i)}{(\xi+\omega) \cdot m_r \cdot \sum_i d_G(p,p'_i)} \end{aligned} \tag{2}$$

Where γ_{CI}, γ_O are the weighting coefficients for the color-intensity salience and orientation salience. $S_{CI}(p,p'_i)$ is the color-intensity contrast and d_G is the Gaussian weighted distance between pixel (x,y) and its neighborhood pixels (x_i,y_i), respectively. $\widehat{C_O}(p,p'_i)$ is the orientation contrast between pixel (x,y) and its neighborhood pixels (x_i,y_i) (for details see [3]).

We construct a new saliency map based on Eq. (2) following three steps (see Fig. 3). First, we apply an over segmentation on the image to divide the image into N regions, which are labeled as $R_1, R_2, \ldots, R_N$. We denote the average salient value of these regions to be $\bar{S}_i, i = 1, 2, \ldots N$. Second, we group these regions into two clusters according to

Fig. 3 Saliency map construction. (a) Original image, (b) segmentation result, (c) the new saliency map used for grouping.

the average salient value $\bar{S}_i$. That is all regions which satisfy $\bar{S}_i > \bar{S}$ belong to the cluster R^+, and the remaining belong to the cluster R^-, where

$$\bar{S} = \frac{\sum_{i=1}^{N} \bar{S}_i}{N}.$$

And third, the salient values are rearranged which satisfy

$$s(x,y) \in \begin{cases} [0,1] & (x,y) \in R^+ \\ [-1,0] & (x,y) \in R^- \end{cases}.$$

In order to make a balance between R^+ and R^-, we further set

$$\sum_{(x,y)\in R^+} s'(x,y) = -\sum_{(x,y)\in R^-} s'(x,y)$$

Without loss of generality, we set $s'(x,y) \in [0,1], (x,y) \in R^+$, then $s'(x,y) \in [\tau, 0], (x,y) \in R^-$,where

$$\tau = \frac{\sum_{(x,y)\in R^+} s(x,y)}{\sum_{(x,y)\in R^-} s(x,y)}.$$

By doing these, we can make an unbiased choice between R^+ and R^- regions, even when the areas of these two regions are with great difference.

5. Object-Based Attention Model

We extend the contour grouping model in a hierarchical way, and construct a hierarchical object-based attention model. In psychologic experiments, psychologists have found that shifting attention within an object should be easier than shifting attention between objects, which is also called same-object effect [7]. That is to say, when the object is attended, all the details in the object will be attended thereafter. According to the effect, We can repeat the contour grouping model in this way:

- Run the contour grouping model to detect the most salient boundary B_1. The image is separated into two regions R_{B_1} and $R_{\bar{B}_1}$ by B_1.
- If the homogeneity of region R_{B_1} is good enough, go to next step directly. Otherwise, take region R_{B_1} as a new image, construct a new saliency map on it, and repeat these three steps.
- Inhibition of return is used for region R_{B_1} and tangents in R_{B_1}. If the homogeneity of $R_{\bar{B}_1}$ is good enough, the whole process is accomplished. Otherwise, construct a new saliency map on region $R_{\bar{B}_1}$, and run the three steps on it.

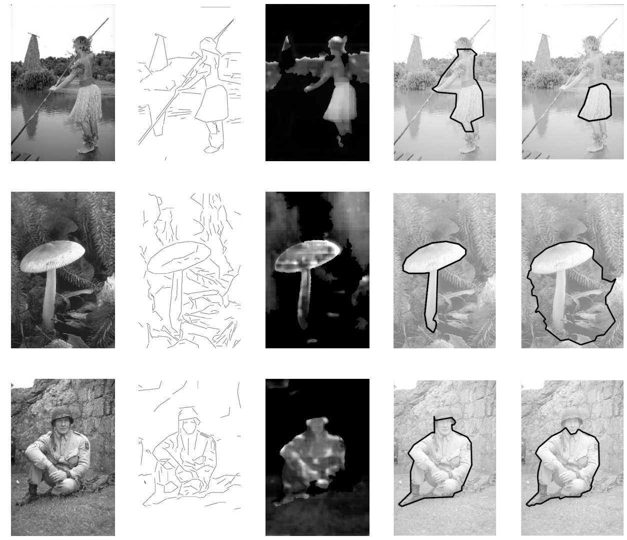

Fig. 4 Contour grouping results of our model. From left to right: the original image, detected tangents, saliency map, result of our model, corresponding result of [4].

We follow the method in [13] to measure the homogeneity of a region using image intensity. After this iteration, we construct a hierarchical object-based attention model. For each detected objects, We pay attention to the detailed sub-objects in it first, then shift the attention to the objects out of its region.

6. Experimental Results and Evaluation

We run our contour grouping method on a set of natural images. The experimental results are compared with a previous edge grouping method [4], which introduces region areas. In Fig. 4 we give our grouping results and the comparison one on three images. We can see that, by using a saliency map, the results of our model are close to the boundaries of actual objects.

We also illustrate the result of object-based attention model which is an extension of our contour grouping model. Figure 5 gives the detected boundaries of objects and the shift of visual attention compared with a proto-objects attention model in [12]. The detected perceptual objects in our model correspond to meaningful actual objects, and the shift of attention is hierarchical as we described in Sect. 5, while the salient regions are meaningless in Walther's model and the shift is out of order.

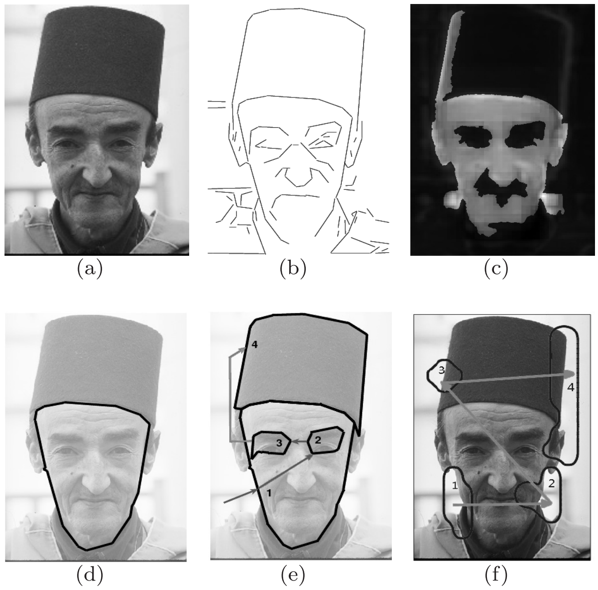

Fig. 5 Object-based attention. (a) Original image, (b) detected tangents, (c) saliency map, (d) contour grouping result, (e) shift of object-based attention, (f) shift space-based attention.

7. Conclusion

In this paper we propose a contour grouping model which combines Gestalt cues of directed tangent pairs and saliency maps. We make use of image information in two ways: the definition of the similarity of tangent pairs and the saliency of regions which are enclosed by boundaries, while most other models only consider edges. It ensures that our grouping model is more robust against noise. The extension of the grouping model constructs an object-based attention model, which can detect perceptual objects from the images and give the hierarchical shift of visual attention. Experiments on natural images compared with other models show that our model performs more favorable.

Acknowledgments

This work is supported by National High Technology Research and Development Program of China (2007AA01Z168), National Nature Science Foundation of China (60805041, 60872082, 60773016), Beijing Natural Science Foundation (4092033) and Doctoral Foundations of Ministry of Education of China (200800041049).

References

[1] L.R. Williams and K.K. Thornber, “A comparison of measures for detecting natural shapes in cluttered backgrounds,” Int. J. Comput. Vis., vol.34, no.2/3, pp.81–96, 2000.
[2] N. Adluru, L.J. Latecki, R. Lakaemper, T. Young, X. Bai, and G. Ari, “Contour grouping based on local symmetry,” IEEE Intereneational Conference on Computer Vision, pp.1–8, 2007.
[3] Y.R. Sun and R. Fisher, “Object-based visual attention for computer vision,” Artif. Intell., vol.146, pp.77–123, 2003.
[4] J.S. Stahl and S. Wang, “Edge grouping combining boundary and region information,” IEEE Trans. Image Process., vol.16, no.10, pp.2590–2606, 2007.
[5] S. Wang, J.S. Stahl, A. Bailey, and M. Dropps, “Global detection of salient convex boundaries,” Int. J. Comput. Vis., vol.71, no.3, pp.337–359, 2007.
[6] J. Duncan, “Converging levels of analysis in the cognitive neuroscience of visual attention,” Phil. Trans. R. Soc. Lond. B., vol.353, pp.1307–1317, 1998.
[7] J. Avrahami, “Objects of attention, objects of perception,” Perception & Psychophysics, vol.61, pp.1604–1612, 1999.
[8] J.S. Stahl, K. Oliver, and S. Wang, “Openg boundary capable edge grouping with feature maps,” IEEE Computer Society Workshop on Perceptual Organization in Computer Vision, pp.1–8, 2008.
[9] J.J. Zhong, S.W. Luo, and Q. Zou, “Visual attention guided multiscale boundary detection in natural images for contour grouping,” IEICE Trans. Inf. & Syst., vol.E92-D, no.3, pp.555–558, March 2009.
[10] S. Wang, T. Kubota, J. Sisikin, and J. Wang. “Salient closed boundary extraction with ratio contour,” IEEE Trans. Pattern Anal. Mach. Intell., vol.27, no.4, pp.546–561, 2005.
[11] J.H. Elder and R.M. Goldberg, “Ecological statistic of gestalt laws for the perceptual organization of contours,” J. Vision, vol.2, pp.324–353, 2002.
[12] D. Walther and C. Koch, “Modeling attention to salient proto-objects,” Neural Netw., vol.19, pp.1395–1407, 2006.
[13] K.B. An, J. Sun, and W.N. Du, “Homogeneity based image objective quality metric,” IEICE Trans. Inf. & Syst., vol.E89-D, no.10, pp.2682–2685, Oct. 2006.