

*

VISUAL ATTENTION BASED SMALL OBJECT SEGMENTATION IN NATURAL IMAGES

Wen Guo^{1,2,3}, Changshen Xu^{1,3}, Songde Ma¹, Min Xu⁴

¹ Institute of Automation, Chinese Academy of Sciences, Beijing, China

² Department of Electric Engineering, Shandong Institute of Business and Technology, Yantai, China

³ China-Singapore Institute of Digital Media, 119615, Singapore

⁴ Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia

{wguo, csxu, sdma}@nlpr.ia.ac.cn, min.xu@uts.edu.au

ABSTRACT

Small object segmentation is a challenging task in image processing and computer vision. In this paper we propose a visual attention based segmentation approach to segment interesting objects with small size in natural images. Different from traditional methods which use the single feature vectors, visual attention analysis is used on local and global features to extract the region of interesting objects. Within the region selected by visual attention analysis, Gaussian Mixture Model (GMM) is applied to further locate the object region. By incorporation of visual attention analysis into object segmentation, the proposed approach is able to narrow the searching region for object segmentation so as to increase the segmentation accuracy and reduce the computational complex. Experimental results demonstrate that the proposed approach is efficient for object segmentation in natural images, especially for small objects. The proposed method outperforms traditional GMM based segmentation significantly.

Index Terms – Gaussian Mixture Model (GMM), Visual Saliency, Segmentation.

1. INTRODUCTION

Object segmentation is widely used in object tracking and recognition. The approaches of object segmentation can be classified into 3 categories [1]: global knowledge based, region based and edge based. Although these methods have obtained certain meaningful result on object segmentation, it is still difficult to extract objects of attention from an image. For small objects of attention in natural images, the state-of-the-art approaches have not achieved remarkably satisfactory results because the small objects are sensitive to noise and cluttered background. In this paper, we focus on small object segmentation in natural images.

Recently, visual attention has been borrowed into computer vision and multimedia field [4-6]. Many researchers find that visual attention is helpful for object detection, recognition, segmentation [14] and tracking. In [8], a rough region was segmented based on pre-attentive features using visual attention. The bottom-up attention [9] was utilized to extract location, size and shape of objects from images. Itti [10] presented a visual attention system inspired by the

behaviour and the neuronal architecture of the early primate visual system. Itti's visual attention model has been widely used in object attention detection.

Attention object detection attracts ever-increasing researchers' efforts, yet it is still a challenging task, especially for natural images. In natural images, it is very difficult to find the object of attention due to the small objects and complicated background. Moreover, the gap between machine learning and human perception makes the task even harder. The existing methods of natural image segmentation include minimization active contour, edge flow, MRF, kernel density estimation, spline regression, and GMM. Among these methods, GMM is more robust than other methods due to its region based essence. In order to implement GMM, EM algorithms [3] and some other co-training algorithms [2] were applied to find the optimal parameters. These algorithms improved the segmentation accuracy significantly. Our previous research of using Co-EM strategy [2] for natural image segmentation demonstrated promising segmentation results. However, it is still difficult for small object segmentation due to two reasons. Firstly, the objects of attention are very small, which might be identified as background. Secondly, the objects of attention might be easily distorted by the color, texture of the surroundings. Therefore, the objects become too insignificant to be extracted.

Small object is defined as the object which occupies less than 20 percent of the input image. In this paper, we introduce visual attention to segment the small objects of attention. If the object is big enough, it will be possible to segment the object directly using traditional segmentation approaches. On the other hand, if the object is very big, visual attention may focus on the details of the object. In this paper, we firstly apply visual attention analysis based on local features to locate the rough region containing the object of attention which can help to reduce the region of search, avoid affect of the background and accurately find the candidates of attention objects. Secondly, the Gaussian Mixture Model (GMM) [3] is utilized based on global features to segment objects of attention from the rough region. By this way, we segment small objects of attention with two contributions. 1) Both local and global features are used efficiently for small object segmentation. 2) The segmented results of small objects are relatively interested to the users.

* This work is partially supported by NSFC Grant #60970092, 60970105.

2. VISUAL ATTENTION BASED OBJECT SEGMENTATION

As shown in Fig. 1, the proposed method consists of two parts, visual attention detection and GMM based segmentation.

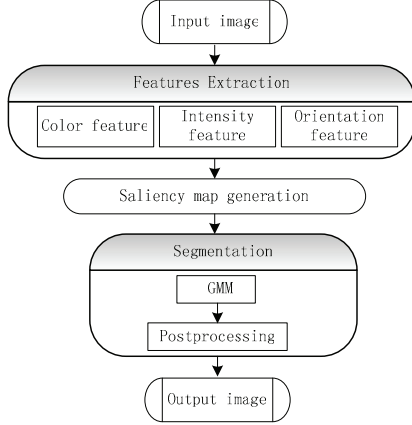


Fig.1. Framework of object segmentation

The proposed visual attention detection method is based on the work on Itti [10] and Harel [13]. The visual attention model is implemented following three steps: feature extraction, saliency computation and inherit of return. Color, intensity and orientation are firstly extracted to generate feature vectors for the image. Each feature is computed by a set of linear “centre-surround” operation. The Gaussian pyramids are used to created spatial multi-scales. Features are normalized into $[0, 1]$. In order to obtain the rough region of attention, we compute the saliency map to represent the conspicuity at all locations in the visual field.

2.1. Features Extraction

Color, intensity and orientation are used as features to generate saliency map. For the RGB color model (r, g and b being the red, green and blue channels), an intensity image I is represented as $I = (r + g + b) / 3$. Gaussian pyramid $I(\sigma)$ is created from I , where σ is the scale. The r, g and b channels are normalized by I in order to decouple hue from intensity. Four broadly-tuned colour channels are generated as follows:

$$\begin{cases} R = r - (g + b) / 2; G = g - (r + b) / 2 \\ B = b - (g + b) / 2; Y = (r + g) - |r - g| / 2 - b \end{cases} \quad (1)$$

Four Gaussian pyramids $R(\sigma), G(\sigma), B(\sigma)$ and $Y(\sigma)$ are created from those color channels.

Centre-surround differences between “centre” fine scale c and a “surround” coarse scale s yield the feature maps. The intensity feature map is constructed from the intensity contrast, which is detected by neurons sensitive. Here the sensitivities are computed by:

$$I(c, s) = |I(c) \odot I(s)| \quad (2)$$

where \odot marks the cross-scale difference between two maps.

The color feature maps are constructed from the color channels, which are represented by a “color double-opponent” system [10]. In the centre of the receptive fields, neurons are stimulated by one color and inhibited by another one, while the converse may be true in the surround. Accordingly, maps $RG(c, s)$ are created in the model to simultaneously account for red/green and green/red double opponency (3), $BY(c, s)$ for blue/yellow and yellow/blue double opponency (4).

$$RG(c, s) = |(R(c) - G(c)) \odot (G(s) - R(s))| \quad (3)$$

$$BY(c, s) = |(B(c) - Y(c)) \odot (Y(s) - B(s))| \quad (4)$$

To obtain orientation, we need an appropriate filter set and a method to extract the measure of orientation out of the output. Garbor pyramids decomposition [10] is widely used for local orientation calculation. Therefore, the local orientation information is obtained from I by using oriented Garbor pyramids $O(\sigma, \theta)$, where σ is the scale, $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ is the preferred orientation, and the orientation feature maps $O(c, s, \theta)$ encode, as a group, local orientation contrast between the centre and surround scales

$$O(c, s, \theta) = |I(c, \theta) \odot I(s, \theta)| \quad (5)$$

2.2. Saliency Map Generation

After visual features are calculated, these features create a feature map for the image. We should consider how to utilize these feature maps to generate visual attention map. It is not a desirable way to combining all feature maps into a saliency map because too much noise or less-salient objects present in majority maps. The most important issue for saliency map generation is how to set different weights to feature maps. A neural network is applied to find out these weights.

Itti introduced a map normalization operator $\mathbb{N}(\bullet)$ [10], which globally promotes maps in which conspicuous location is present. The operator includes three principals. By $\mathbb{N}(\bullet)$, the feature maps are combined into three “conspicuity maps”, \tilde{I} for intensity, \tilde{C} for color, and \tilde{O} for orientation:

$$\tilde{I} = \bigoplus_{c, s} \mathbb{N}(I(c, s)) \quad (6)$$

$$\tilde{C} = \bigoplus_{c, s} [\mathbb{N}(RG(c, s)) + \mathbb{N}(BY(c, s))] \quad (7)$$

$$\tilde{O} = \sum_{\theta} \mathbb{N}(\bigoplus_{c, s} \mathbb{N}(O(c, s, \theta))) \quad (8)$$

where \oplus marks point-by-point addition. The saliency map is obtained by averaging three normalized conspicuity maps.

A concept of “activation” is proposed in [13], which is unusual corresponding to high values in its neighborhood. Two weight computational formulations are employed to calculate some outbound edge of nodes. In his paper, we use graph-based activation to generate features. The graph-based

scheme is also implemented for feature map normalization. The experimental results gave a better performance than Itts' saliency map.

2.3. Segmentation for Objects of Attention

GMM has been widely used for image segmentation. The most important issue for GMM implementation is to get suitable parameters. EM algorithm [3] has been proved to be efficient for GMM parameter estimation. Extended GMM-EM methods have been proposed, such as Co-EM [2], which also show good performance. However none is able to successfully segment small objects in natural images.

In this paper, we focus on segmentation of small objects in natural images. For an input image, we firstly calculate the visual attention saliency map to locate the rough region of the objects of attention. Then GMM is used to segment the objects from the rough attention region.

The detailed implementation is described as follows.

(1) Color feature maps, intensity feature maps and orientation feature maps are firstly calculated according to Eq.(1)-Eq.(3). Feature maps are combined into a saliency map by averaging three normalized conspicuity maps. Graph based visual saliency algorithm is used to generate the saliency map.

(2) According to statistic conspicuity of the saliency map, select a plausible threshold T to generate a binary image.

$$f(i, j) = \begin{cases} 0, & f(i, j) > T \\ 1, & f(i, j) < T \end{cases} \quad (8)$$

(3) GMM is used to model the rough object region. The parameter estimation of GMM uses EM algorithm [3] including two steps.

I. E-step: The posterior probability of sample x_j at the t th step is calculated as:

$$p(i | x_j; \Theta^{(t)}) = \frac{\alpha_i^{(t)} p(x_j; \theta_i^{(t)})}{\sum_{i=1}^n \alpha_i^{(t)} p(x_j; \theta_i^{(t)})} \quad (9)$$

II. M-step: The $(t+1)$ th step $\Theta^{(t+1)}$ is updated as:

$$\alpha_i^{(t+1)} = \frac{1}{n} \sum_{j=1}^n p(i | x_j; \Theta^{(t)}) \quad (10)$$

$$\mu_i^{(t+1)} = \frac{\sum_{j=1}^n x_j p(i | x_j; \Theta^{(t)})}{\sum_{j=1}^n p(i | x_j; \Theta^{(t)})} \quad (11)$$

$$\Sigma_i^{(t+1)} = \frac{\sum_{j=1}^n x_j p(i | x_j; \Theta^{(t)}) [(x_j - \mu_j^{(t+1)})(x_j - \mu_j^{(t+1)})^T]}{\sum_{j=1}^n p(i | x_j; \Theta^{(t)})} \quad (12)$$

where $\theta_i = \{\mu_i, \Sigma_i\}$ is the i th component mean vector and the covariance matrix of the random variable of the X , Θ is the set of θ_i , p is the probability density function, α_i are the weights which satisfy $\alpha_i > 0$ and $\sum_{i=1}^k \alpha_i = 1$, $i = 1, 2, \dots, n$. We label the connected region, then calculate the area of every

region and select a threshold S . If the area satisfies $\sum_i \sum_j f(i, j) < S$, delete the region. A rectangle is generated according to the width and height of the region to label the rough object region. The rectangle is 10 percent bigger than the rough region in order to remain enough candidates for object detection.

3. EXPERIMENTS

50 images containing small objects are used to test the performance of our method. 30 images are selected from Berkeley dataset. 10 natural images and 10 road sign images are selected from the Internet. In Fig. 2, we give 16 representative images to show the segmentation results. Since our method is an improvement of traditional GMM for small object segmentation, we also implement the traditional GMM based segmentation with RGB features and Gabor features respectively for comparison. The original images and related experimental results are shown in Fig. 2.

As shown in column (e), we can find that the segmentation by using traditional GMM is not satisfactory. For image 3(e), the segmentation result is distorted by the noise from background. For image 1(e), the object has not been segmented. Although the small object in image 3(f) has been segmented, some regions in background are also recognized as objects. Images in column (d) are the segmentation result by using our method. It is obvious that our segmentation method provides the best result. More results can be seen in Fig. 4.

The proposed method segments object based on visual attention. Objects not attracting human attention are difficult to be segmented by our method. An example is shown in Fig. 3. The small yellow flower in the bottom of the image is segmented as object of attention instead of the brown bird which is considered as an object in the image. However, objects of attention (e.g. yellow flower in Fig.3) are more significant for image understanding than other objects.

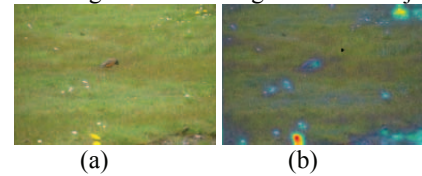


Fig.3. An example of segmentation. (a) original image (b) saliency map

4. CONCLUSIONS

We have proposed a novel segmentation method for small objects of attention in natural images. The proposed method combines visual attention analysis and Gaussian Mixture Model to segment the objects of interests. The proposed method outperforms the traditional GMM based segmentation, especially for small objects detection.

Currently, the proposed method is implemented for small object segmentation, which is the initial step for small object

tracking. In the future, our study will be extended to small object tracking.

REFERENCES

- [1] Sonka,M., Hilavac,V., Boyle,R. Image Processing, Analysis, and Machine Vision, 1998, 2nd edn.Brooks/Cole.
- [2] Z.Li, J.Chen, Q.Liu, etc, "Image Segmentation Using Co-EM Strategy" Lecture Notes in Computer Science, ACCV2007, pp. 827-836.
- [3] K.K.Yiu,M.W.Mak, C.K.Li, "Gaussian Mixture Model and Probabilistic Decision-based Neural Networks For Pattern Classification: A Comparative Study," Neural Computing and Applications, 1999, vol. 8, pp. 235-245
- [4] Yun Zhai, Mubarak Shan, "Visual Attention Detection in Video Sequences Using Spatiotemporal Cues," Proceeding of the 14th ACM International Conference on Multimedia, 2006, pp. 815-824.
- [5] Ken Fukuchi, Miyazato,K., Kimura,A., "Saliency-based video segmentation with graph cuts and sequentially updated priors," IEEE ICME 2009, pp. 638-641.
- [6] H.Liu, S.Jiang, Q.Huang, etc, "A Generic Virtual Content Insertion System Based on Visual Attention Analysis," Proceeding of the 16th ACM International Conference on Multimedia, 2008, pp. 379-388.
- [7] Itti,L., Koch,C., Niebur,E., "A model for saliency-based visual attention for rapid scene analysis," IEEE Trans. PAMI, 1998, vol. 20, pp. 1245-1259.
- [8] Walther,D., Itti,L., Riesenhuber,M., ect. "Attentional selection for object recognition – a gentle way," Proceedings of the Second International Workshop on Biologically Motivated Computer Vision, 2002, pp. 472-479.
- [9] Rutishauser,U., Walther,D. Koch,C., etc, "Is bottom-up attention useful for object recognition," Proceeding of ICCV 2004, pp. 37-44.
- [10] G.Hua, Z.Liu, Z.Wu, "Iterative Local-Global Energy Minimization for Automatic Extraction Objects of Interest," IEEE Trans. PAMI, 2006, vol. 28, pp. 1701-1706.
- [11] D.Comaniciu, P.Meer, "Mean-Shift: A Robust Approach toward Feature Space Analysis," IEEE Trans. PAMI, 2002, vol. 24, pp. 1-18.
- [12] S. Xiang, F.Nie, C.Zhang, etc, "Interactive Natural Image Segmentation via Spline Regression," IEEE Trans. Image Processing, 2009, vol. 18, pp. 1623-1632.
- [13] Harel,J., Koch,C., Perona,P., "Graph-Based Visual Saliency," Advances in Neural Information Processing Systems, MIT Press, 2007, pp. 545-552
- [14] D.Michael, U.Martin, H.Martin, ect, Saliency driven total variation segmentation, Proceeding of ICCV2009, pp. 817-824,2009.

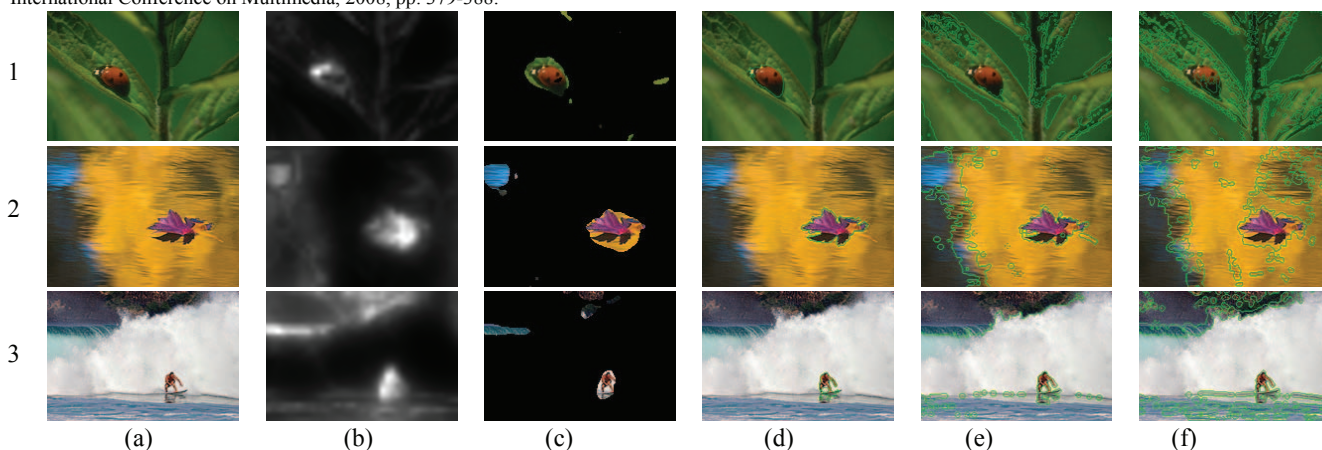


Fig.2. (a) Original image, (b) Visual saliency map of original image, (c) Plausible partial saliency display, (d) Segmentation result using our approach, (e) Segmentation result using GMM, (f) Segmentation result using Garbor feature.

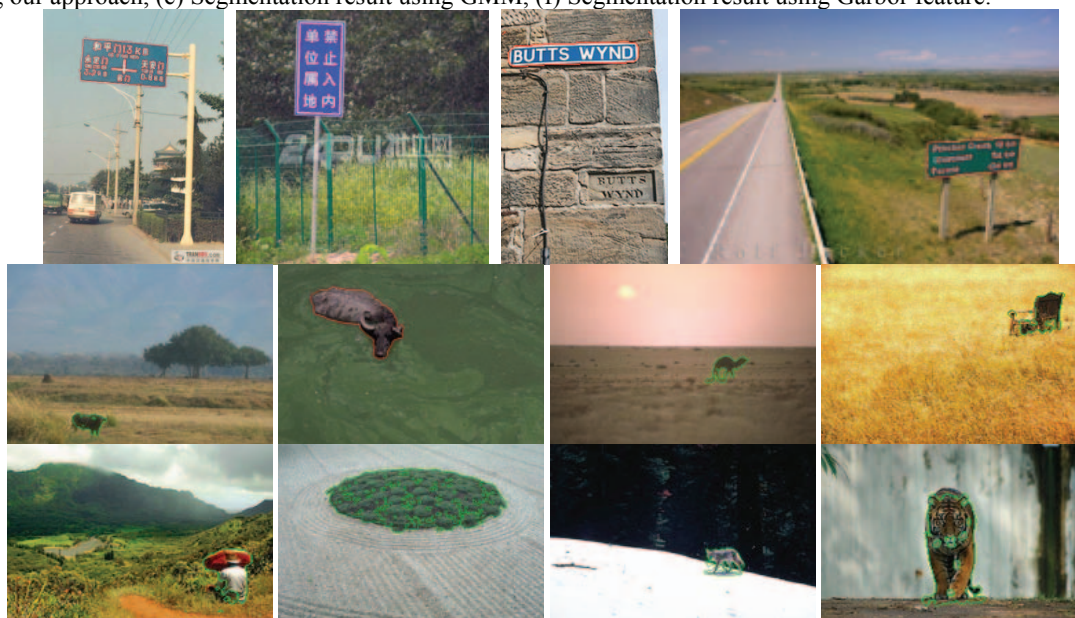


Fig.4. More segmentation results of objects of attention