J Inf Process Syst, Vol.9, No.4, pp.592-601, December 2013
http://dx.doi.org/10.3745/JIPS.2013.9.4.592

pISSN 1976-913X
eISSN 2092-805X

Small Object Segmentation Based on Visual Saliency in Natural Images

Huynh Trung Manh* and Gueesang Lee*

Abstract—Object segmentation is a challenging task in image processing and computer vision. In this paper, we present a visual attention based segmentation method to segment small sized interesting objects in natural images. Different from the traditional methods, we first search the region of interest by using our novel saliency-based method, which is mainly based on band-pass filtering, to obtain the appropriate frequency. Secondly, we applied the Gaussian Mixture Model (GMM) to locate the object region. By incorporating the visual attention analysis into object segmentation, our proposed approach is able to narrow the search region for object segmentation, so that the accuracy is increased and the computational complexity is reduced. The experimental results indicate that our proposed approach is efficient for object segmentation in natural images, especially for small objects. Our proposed method significantly outperforms traditional GMM based segmentation.

Keywords—Gaussian Mixture Model (GMM), Visual Saliency, Segmentation, Object Detection.

1. INTRODUCTION

Object segmentation has been widely used in many applications, such as object tracking and object recognition. Object segmentation can be classified into the following three categories: global knowledge based segmentation, region based segmentation, and edge based segmentation. Although these traditional methods have achieved acceptable results, they still do not perform well for images that contain small objects. For small objects in a natural image, the previous state-of-art methods have not obtained satisfactory results because the small object is sensitive to noise and cluttered backgrounds. In this paper, we focus on the small object segmentation in natural images.

Recently, the applications of visual attention in computer vision have become popular, which include object detection, recognition, segmentation, and tracking. Visual saliency is the perceptual quality that makes an object, a person, and pixel stand out in relation to its neighbours and thus, capture our attention. Saliency estimation methods can be classified into the following three different types: biologically based, computational, or a combination of these two methods. Generally, most of the recent methods convey low features such as intensity, color, or orientation

※ This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), which is funded by the Ministry of Education, Science and Technologies(2013-056480 and 2013-006535) and the MSIP(Ministry of Science, ICT&Future Planning), Korea, under the ITRC (Information Technology Research Center) support program that is supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2013-H0301-13-3005).
Manuscript received February 25, 2013; accepted April 7, 2013.
Corresponding Author: Gueesang Lee (gslee@chonnam.ac.kr)
* Dept. of Electronics and Computer Engineering, Chonnam National University, Gwangju, 500-120, Korea (trungmanhhuynh@gmail.com, gslee@chonnam.ac.kr)

in order to determine the contrast of image regions. In the first category, Itti et al. [1] and Koch and Ullman [2] proposed their method based on biologically plausible architecture by creating centre-surround contrast by using a Different of Gaussian (DoG) approach. Some of the methods that are based only on pure computational models [3,4] estimate saliency by using centre-surround feature distances. The third category of approaches incorporate biological models into computational ones [5], create feature maps with Itti's method but perform the normalization using a graph based method. In this paper, we implement our computational model to extract the location, size, and shape of objects from images.

Attention object detection is attracting the ever-increasing 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 the natural image segmentation include minimizing the 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, the EM algorithm [6, 11] and some other co-training algorithms were applied to find the optimal parameters. These algorithms significantly improved the segmentation accuracy. Our previous research on the Co-EM strategy for natural image segmentation demonstrated promising segmentation results. However, it is still difficult to do small object segmentation due to two reasons. First, the objects of attention are very small and might be identified as part of the background. Second, the objects of attention might be easily distorted by the color and texture of the surroundings. Therefore, the objects become too insignificant to be extracted.

A small object is defined as the object, which occupies less than 20% 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 first applied a visual attention analysis that is based on local features to locate the rough region that contains the object of attention, which can help to reduce the search region, avoid the effect of the background, and accurately find the attention object candidates. Secondly, we utilized the Gaussian Mixture Model [6] to segment objects of attention from the rough region.

This paper is organized as follows: the proposed method is introduced in Section 2. The experimental results and comparison to existing methods are mentioned in Section 3. Finally, we give the conclusions about our method and future works in Section 4.

2. Proposed Approach

Our proposed method consists of the following two parts: visual attention detection and GMM based segmentation, as shown in Fig.1. The proposed visual attention detection method is based on our novel saliency detection method. In Section 2.1, we will show that our saliency model significantly outperforms several of the state-of-art methods. Our models are mainly based on the band-pass filtering that uses DoG to obtain the appropriate middle frequency of an image. Furthermore, we also make the assumption that a person will mainly concentrate on the center of the image and as such, that the saliency value of the pixels surround the center should be higher

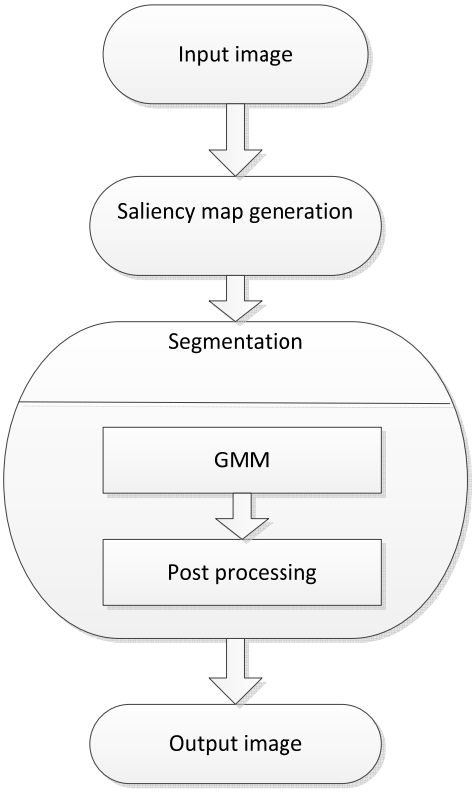

Fig.1. Framework of object segmentation

than others.

There are two contributions in our method: (1) we introduce an innovative visual attention method and (2) we combine GMM and our saliency method to successfully segment the small object.

2.1 Saliency Map Generation

In this section, we introduce a novel model for saliency object detection, which is used in the first step of our method. Recently, there are several approaches for visual saliency detection, yet problems still remain. The saliency maps that have been created from some previous methods [1,2, 3, 7, 8] have the problem of low resolution. Some methods generate maps with ill-defined boundaries [1, 2, 7], which render it useless for post application. Others fail to highlight the whole object [1, 8] or they highlight the smaller salient region instead of the larger one. These limitations have been successfully handled by the frequency-tuned model, which was proposed by Achanta et al. [4]. This method provides the following three criteria for highlighting a salient object: well-defined boundary, full resolution, and effective computation. This work exploits not only most of the low frequency content but also the high frequency content to obtain the saliency map. The method is based on band-pass filtering:

$$DoG(x,y)=\frac{1}{2\pi}\left[\frac{1}{\sigma_1^2}e^{-\frac{(x^2+y^2)}{2\sigma_1^2}}-\frac{1}{\sigma_2^2}e^{-\frac{(x^2+y^2)}{2\sigma_2^2}}\right] \tag{1}$$

Where, σ_2 defines the high frequency cut-off. In order to remove the noises that are contained in very high frequency we set σ_2 as small. σ_1 defines the low-frequency cut-off so we enlarged σ_1 to infinity to take all of the low frequency to highlight the whole objects. Therefore, we can re-write it as:

$$S=\left\|I_\mu(x,y)-I_f(x,y)\right\| \tag{2}$$

Where, S is the final saliency map, I_μ is the average of all the pixels in an image, and $I_f$is the blurred image, which is obtained by the small binomial Gaussian kernel. However, there are two limitations of this method: first, it fails to highlight the object when the number of pixels for the salient object are larger than half of the total number of pixels of an image; second, undesired salient regions are also highlighted when the background is complex or when illumination effects occur. Next, we propose using the maximum symmetric surround method to handle these problems by varying the bandwidth of the centre-surround filter. This means that the bandwidth of this filter is very large in the center of images and becomes smaller at the borders (Fig. 5). The proposed method efficiently remove the non-objects that surround the boundaries of images but that still highlight entire objects around their centers. It also efficiently handles salient objects, even better than [4] in the presence of the complex illumination.

The limitation still exists in this method, when it fails to evaluate the saliency value for each pixel by considering the distance between each one of the borders of the image precisely. For example, if the image borders cut the salient object, they are treated as the background and is less likely to be detected. In this paper, we increased the saliency value for each pixel with an integrated model, which conveys the information about the distance from each pixel to each border and to the center of the image. Therefore, our visual saliency detection algorithm not only attains all of the advantages of previous methods but also deals with their problems.

In this paper, we propose an integrated model that makes the following three contributions: the salient object is highlighted with a well-defined boundary; non-objects or complex textures are completely removed; and the problem of there being a large salient region in the image is handled. To achieve these things, we precisely evaluated the saliency value for each pixel by using a model that combines a maximum symmetric surround model with a exponential ellipse saliency model. Therefore, the saliency value at the given pixel is obtained as:

$$\begin{aligned} S &= \vartheta(x,y).\left\|I_\mu(x,y)-I_f(x,y)\right\| \\ &= \left(1-e^{\left(\frac{x-\rho(i)}{R}\right)^2+\left(\frac{y-\rho(j)}{C}\right)^2}\right).\left\|I_\mu(x,y)-I_f(x,y)\right\| \end{aligned} \tag{3}$$

Where $\rho(i)$ and $\rho(j)$ is the location of the pixel in the center of image that follows the x-axis and y-axis, respectively. (R, C) is the size of the input image. $I_\mu(x, y)$ is the average of the sub-

image at the center pixel(x,y) as shown by:

$$I_{\mu}(x,y) = \frac{1}{A}\sum_{i=x-x_o}^{x+x_o}\sum_{j=y-y_o}^{y+y_o} I(x,y) \tag{4}$$

Where

$$x_o = \min(x, w-x)$$
$$y_o = \min(y, h-y)$$
$$A = (2x_o+1)(2y_o+1)$$

Our exponential ellipse model considers the high saliency value for the pixel in the center of image and decreases the amount of pixels that surround the borders by an exponential level. By applying this model, we were able to completely remove the complex textures. As shown in Fig.3, the pixels that are surround the center of the image should have a high saliency. The dataset only contains two types of images that are [400x300] and [300x400] in size. Hence, we constructed two models that are appropriate and consistent for indicating the saliency for each pixel. The smaller ellipse shape indicates that there is a higher saliency level for the pixels that are located in that ellipse.

2.2 Segmentation for Objects of Attention Based on the GMM

The Gaussian Mixture Model (GMM) is popular for colour clustering and image segmentation. The GMM has been widely used for image segmentation. The most important issue for GMM implementation is to get suitable parameters. The EM algorithm [3] has been proven to be efficient for GMM parameter estimation. Extended GMM-EM methods have been proposed, such as Co-EM [2], which also shows good performance. However, none of them are able to successfully segment small objects in natural images. In this paper, we focus on the segmentation of small objects in natural images. For an input image, we first calculate the visual attention saliency map to locate the rough region of the objects of attention. Then we used the GMM to segment the objects from the rough attention region.

The parameters estimation of the GMM uses the EM algorithm and includes these two steps:

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

$$p(l, x_j; \Theta^{(t)}) = \frac{\alpha_l^{(t)} p(x_j; \theta_l^{(t)})}{\sum_{i=1}^{n} \alpha_l^{(t)} p(x_j; \theta_l^{(t)})} \tag{5}$$

$$\alpha_i^{(t+1)} = \frac{1}{n}\sum_{j=1}^{n} p(i \mid x_j; \Theta^{(t)}) \tag{6}$$

$$\mu_i^{(t+1)} = \frac{\sum_{j=i}^{n} x_j \, p(i \mid x_j; \Theta^{(t)})}{\sum_{j=1}^{n} p(i \mid x_j; \Theta^{(t)})} \tag{7}$$

$$\sum_i^{(t+1)} = \frac{\sum_{j=i}^{n} x_j \, p(i \mid x_j; \Theta^{(t)})[(x_j - \mu_j^{(t+1)})(x_j - \mu_j^{(t+1)})^T]}{\sum_{j=1}^{n} p(i \mid x_j; \Theta^{(t)})} \tag{8}$$

Where $\theta_i = \{\mu_i, \Sigma_i\}$ is the i th component mean vector 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 that satisfy $\alpha_i \rangle 0$ and $\sum_{i=1}^{k} \alpha_i = 1$.

3. Experimental Results

3.1 Saliency Map Comparison

First, we compared our saliency maps with 7 state-of-the-art methods. These methods were created by Itti et al. [3], Ma and Zhang [5], Harel et al. [6], Hou and Zhang [8], and Achanta et al. [1,9,10], which are referred to as IT98, MZ03, HA07, HO07, AC08, FT08, and MS10, respectively. We refer to our method as MA13. To compare the quality of our saliency map with previous methods, we used the precision-recall based method [1]. We did the segmentation for each level of a grayscale image from 0 to 255 and calculated their precision and recall with 1,000 public

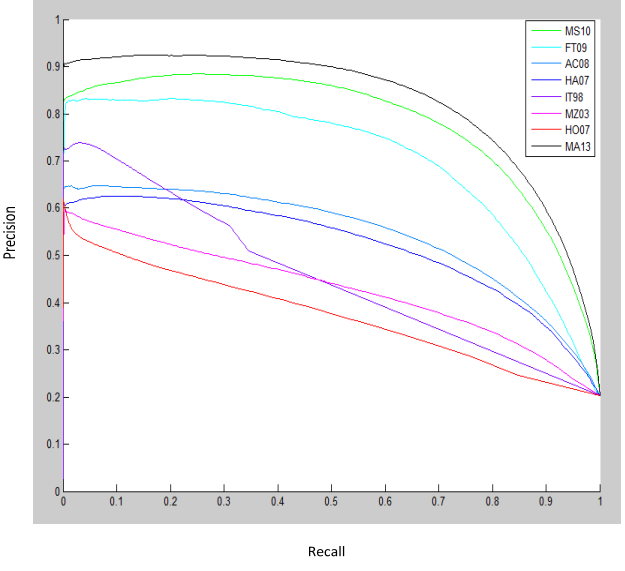

Fig. 2. Precision-recall curve using the ground truth. Our saliency method (black-line), which we used on over 1,000 public images, shows the best performance compared to 7 previous state-of-the-art methods

images and their ground truth data. As shown in Fig. 5, the results show that our method has better quality than other saliency maps. We concluded that our visual saliency method is well suited for small object detection in our system.

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

with TP, FP, FN denote true-positive, false-positive, false-negative, respectively.

3.2 Segmentation Based GMM Results

The images containing small objects were selected from the Berkeley dataset and from the Internet. In Fig. 4 we show the comparison between the standard GMM and our proposed method. In the second column, we show the saliency map, which highlights the small object before the second stage. In the third column, the standard GMM failed to segment the small objects and caused a lot of noise as well; and the final column shows our results. We can easily see that our method outperforms and achieves much better results as compared to the original GMM. Figure 5 shows more segmentation results from our method.

The proposed method segments an object based on visual attention. Objects that do not attract attention from people will be difficult to segment with our method. For example, in Figure 6, the mountainside captures a person's attention instead the boats. Hence, it is highlighted in the saliency map and the segmentation results become false.

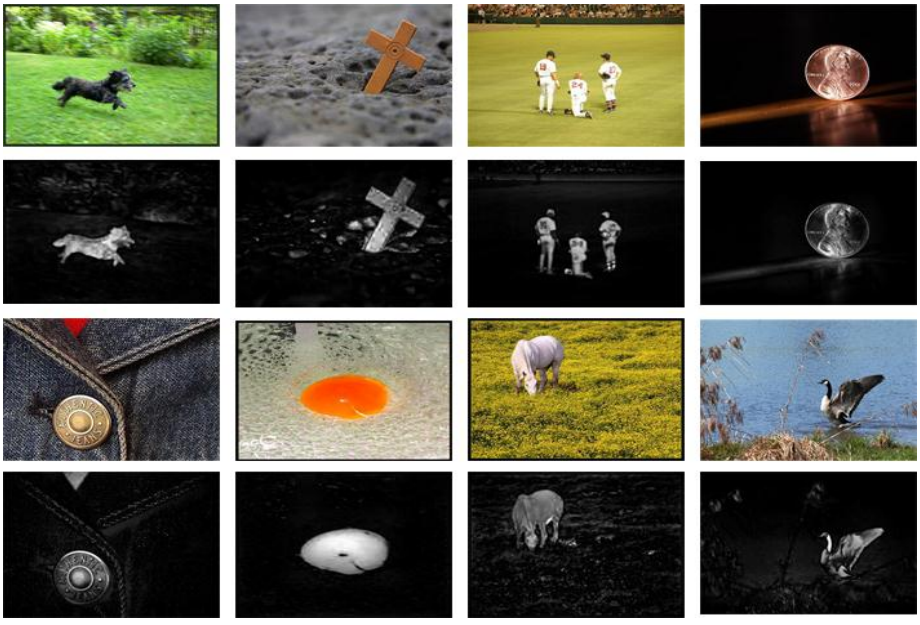

Fig. 3. Comparison of saliency maps. Our method produces the maps that have better quality and highlights the whole object with a well-defined boundary, as well as in complex textures.

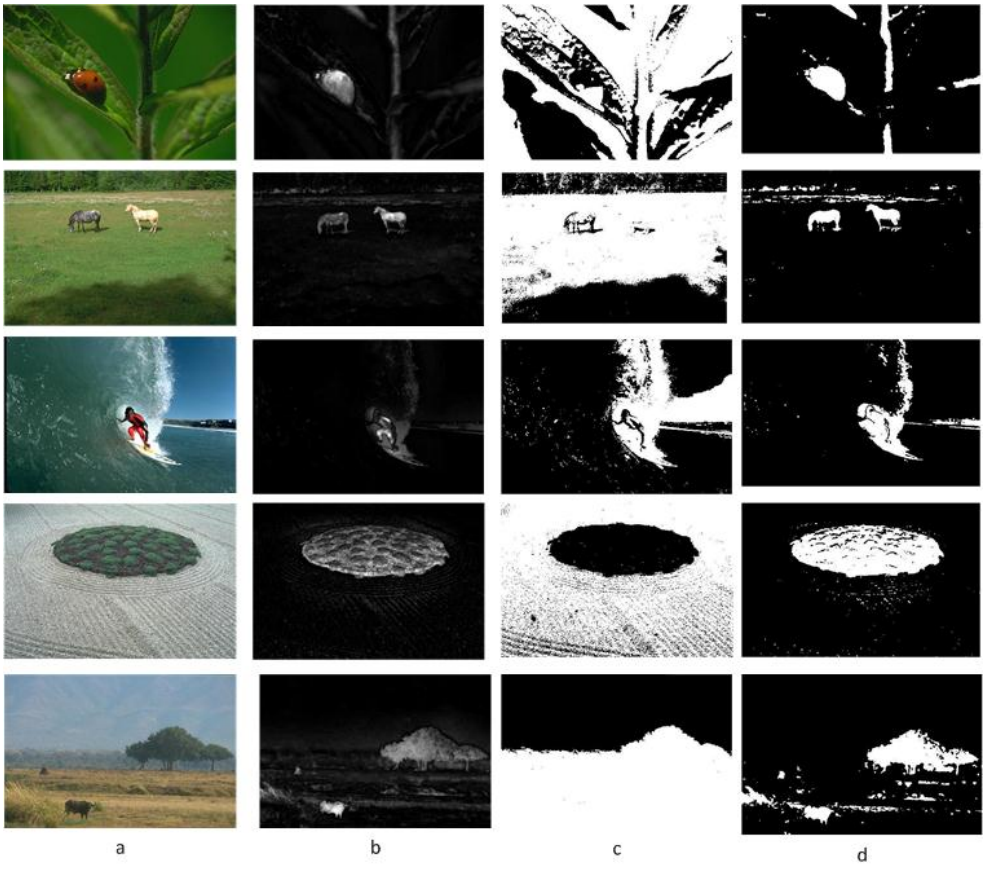

Fig. 4. (a) Original image, (b) Visual saliency map of the original image, (c) Segmentation results using the standard GMM, (d) Segmentation results from our method.

Fig. 5. Some more results from our proposed method.

Fig. 6. Failure case

4. CONCLUSION

In this paper, we have proposed a novel segmentation method for small objects of attention in natural images. We have successfully constructed a system that can be used for small object segmentation, which is based on visual attention. The main contribution of our method relies on our novel Saliency detection in combination with the standard GMM in order to segment the small object. Nevertheless, there were some failure cases because a small object cannot capture human attention, but our method generally outperforms the standard GMM. Future work should focus on improving the saliency detection method for further segmentation. Currently, the proposed method is implemented for only small object segmentation, which is the initial step for small object tracking and recognition.

REFERENCES

[1] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," Advances in Neural Information Processing Systems , 2007, pp.545–552.

[2] S. Frintrop, M. Klodt, and E. Rome, "A real-time visual attention system using integral images," in International Conference on Computer Vision Systems, 2007.

[3] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp.1–8.

[4] R.Achanta, S.Hemami, F.Estrada, and S. Susstrunk," Frequency-tuned salient region detection", in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp.1597-1604 .

[5] R. Achanta, F. Estrada, P. Wils, and S. Susstrunk, "Salient region detection and segmentation," International Conference on Computer Vision Systems, 2008, vol. 5008, pp.66–75.

[6] 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.

[7] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, November, 1998, vol. 20, no. 11, pp. 1254–1259.

[8] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," ACM International Conference on Multimedia, November, 2003, pp.374–381.

[9] R. Achanta, S.Susstrunk, "Saliency detection using maximum symmetric surround," IEEE 17th International Conference on Image Processing, 2010.

[10] 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.

[11] Z.Li, J.Chen, Q.Liu, etc, "Image Segmentation Using Co-EM Strategy," Lecture Notes in Computer Science, ACCV2007, pp.827-836.

Huynh Trung Manh

2012 He received the B.S degree in Computer Engineering from Ho Chi Minh University of Technology, Viet Nam. 2012- Present He is currently a master student in Department of Electronics and Computer Engineering, Chonnam National University, Korea. Research Interests: Image processing, Computer Vision and Object Tracking.

Gueesang Lee

1980 He received the B.S degree in Electrical Engineering from Seoul National University.1982 He received the M.S degree in Electrical Engineering from Seoul National University.1991 He received Ph.D. degree in Computer Science from Pennsylvania State University. He is currently a professor of the Department of Electronics and Computer Engineering in Chonnam National University, Korea. Research Interests: Image processing, computer vision and video coding.