2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

Salient region detection based on Local and Global Saliency

Peng Wang, Zhi Zhou, Wei Liu and Hong Qiao

Abstract—**a new and effective salient region detection method based on local and global saliency information is proposed. To keep the completeness of salient regions, the input image is segmented into several regions firstly. Then for each region, local saliency and global saliency are generated respectively. The local saliency is computed by multi-scale neighborhood contrast, and the global saliency is measured according to global spatial distribution and inter-region isolation of features. Based on the local saliency and global saliency, the final saliency can be obtained by the weighted combination of them. The comparison experiment results demonstrate the effective performance of the proposed algorithm on salient region detection.**

I. Introduction

Visual attention has been studied by researchers in different areas, such as physiology, psychology and computer vision, and how to detect and segment the salient regions in an input image is an important task in many applications. The extraction of salient regions is able to provide useful information for different tasks, such as image compression, object detection and segmentation, as well as object recognition and tracking.

There are two manners in salient region detection: the bottom-up manner [1-4] and the top-down manner [14-17], and also some methods combine the two manners together for context-aware saliency detection [21]. The bottom-up manner can find the salient regions without any priori-knowledge. The top-down manner is knowledge/task-driven, and the salient regions are extracted through a perception processing in which a training process is usually required. The aim of salient region extraction is to detect the most distinctive regions or pixels in the input image, which is usually by data-driven, and these methods are based on the bottom-up manner.

For an input image, the saliency is usually measured based on the difference or contrast from each pixel's local neighborhood. One of the early successful models was proposed by Itti [1], which was developed from the first explicit computational model for bottom-up visual attention by Koch et al[18]. In Itti's model, a center-surround difference operator on different image features was used to compute the multi-scale feature maps, and then the obtained feature maps over different scales were combined and normalized to form a final salience map. Similar methods of saliency computation were used in [3-6,10]. The model presented in [3] also constructed the feature map of color by local contrast, and they introduced new feature maps of edges and symmetry. S. Feng et al. [4] used a linear combination of contrasts in the Gaussian image pyramid to simply define the multi-scale contrast features. Ma [5] generated the saliency map by local neighborhood contrast of the LUV image based on contrast and fuzzy growing. Liu et al. [6] extracted salient regions on color by learning local, regional and global features, and the local feature was represented by multi-scale contrast of color. The method proposed by Achanta [10] also caculated saliency through multi-scale local contrast of features between a region and its neighborhood. There were also some other methods to calculate saliency locally. In Kadir and Brady's work [2], saliency was measured by scale localized feature with high entropy based on local complexity. The approach in Aziz and Mertsching [7] extracted different regions first, and then computed the properties based on different features for each region. The saliency was finally obtained by combining different regions according to their saliency values. Weijer [8] detected salient points with color and shape distinctiveness, which were determined by the local differential structure of image. H.-Y. Chen et al. [9] proposed a block-based visual attention model using the standard deviation to reduce the computational complexity. These methods performed effectively in detecting salient regions which contrast drastically with local environment.

Methods mentioned above usually only take the local difference or contrast information into account, and the global information is ignored. The global saliency in the image has been proposed in some recent works [6][11][12][19]. In [6], color spatial variance was defined as the global feature, and it was assumed that a particular color with less spatial variance was likely to be salient. In [11], a saliency detection method based on color and orientation was presented, and the global saliency information, such as the spatial distribution of different color, was used to detect salient regions. A region merging method was proposed in [12] to segment salient regions using global image information. In these methods, the object with distinctive feature compared to the whole image is more likely to be detected as salient object. In [19], a novel global method is proposed, and the spectral residual of an image is used to detect the salient regions.

Local saliency information and global saliency information are usually mutually complementary in saliency detection. In this paper, we propose a new method for salient region detection with the combination of local and global image information. To keep the completeness of the salient regions, we segmented the input image into several regions first. Then the local saliency of each region is computed by multi-scale

Peng Wang, Wei Liu and Hong Qiao are with Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: peng_wang@ia.ac.cn).

Zhi Zhou is with School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore.

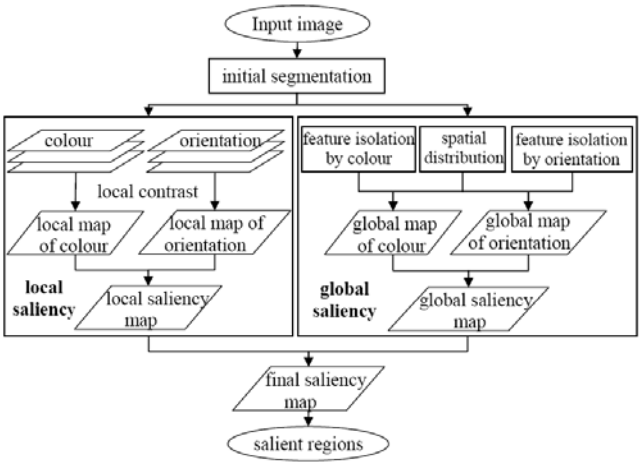


Fig. 1. Framework of the proposed method

neighborhood contrast, and the global saliency is measured according to global spatial distribution and inter-region isolation of features. Color and orientation features are used in both local saliency and global saliency calculation, due to their mutual complementarity. The final saliency is obtained by the weighted combination of local saliency and global saliency, and the weights are determined based on the entropy of the obtained saliency maps.

The paper is organized as follows. Section II describes the framework and details of the proposed saliency detection algorithm. Some experimental results are given in Section III, and conclusion is presented in Section IV.

II. Saliency detection using local and global image information

The algorithm proposed in this section aims to detect and segment the most salient regions from the input image by considering local saliency and global saliency simultaneously. Fig. 1 provides an overview of the proposed local and global saliency based automatic detection and segmentation algorithm, which can be divided into three steps. First, the input image is segmented into several regions, with the purpose of keeping the completeness of the salient regions. Then the local saliency of each region is computed by multi-scale neighborhood contrast, and the global saliency is measured according to global spatial distribution and inter-region isolation of features. Color and orientation features are used in both local saliency and global saliency calculation, due to their mutual complementarities. Third, the final saliency is obtained by the weighted combination of local saliency and global saliency, and the weights are determined based on the entropy of the obtained saliency maps.

A. Segmentation of the input image

Most existing methods of salient region detection [1, 2, 6, 10, 11] usually directly calculate the saliency in pixel level, and this will result in that the extracted salient objects may not be connected regions but scattered pixels. In order to keep the completeness of the salient objects, here HSEG algorithm [13] is introduced to roughly segment the input images into multiple regions before saliency detection. HSEG algorithm is a local homogeneity based segmentation method, which needs few control parameters to be tuned. Compared to other well-known image segmentation algorithms like mean-shift, HSEG needs less calculation and is easy to implement. The segment result of HSEG is satisfying for further processing.

Using HSEG algorithm, we can get an H-map of the input image. Then based on the H-map, region growing method is used to segment the input image into multiple regions. Furthermore, region merging is processed to prevent over-segmentation based on similarity of regions on color histogram, and each region is described by the histogram, such as 4×4×4-bins histogram in our experiments, in which the number of bins is determined empirically.

B. Local Saliency Generation

Based on the segmentation results above, the saliency of each region is calculated by considering the local and global image information simultaneously (Fig. 1), due to the mutual complementarities between the local and global saliency in salient region detection and segmentation. In the calculation of both local and global saliency, color and orientation features are used. RGB color space is used as the color feature, and orientation feature is represented by the output of Gabor filters in four directions: 0, 45, 90, and 135 degrees.

Local saliency is important in salient region detection, showing the distinctiveness between a region and its neighborhood. A region with high difference from its neighborhood should be a salient region. Lots of local saliency detection methods have been proposed, such as [1], [2], [5] and [10]. In this subsection, we will propose a new local saliency detection method which is partly illuminated by [10].

In an input image, the local saliency value of a pixel in the image is determined as a sum of local contrast values at different scales.

$$LocContr_{(x,y)}^{Fea} = \sum_s \| \boldsymbol{F}^{Fea}(x,y) - \boldsymbol{M}^{Fea}(x,y)_s \| \qquad (1)$$

where *Fea* represents the feature used, i.e., *Fea*= {*color, orientation*}. NR_s denotes the surrounding region of a pixel, and *s* denotes the scale of the region NR_s. In (3), the scales of the region NR_s are set to be 1/4, 1/8 and 1/16 of the scale of the input image, respectively. $\boldsymbol{F}^{Fea}(x,y)$ represents the feature vector at location (x,y), and $M(x,y)_s$ is the mean feature vector of pixels in region NR_s at scale *s*, which can be computed as

$$M^{Fea}(x,y)_s = \frac{\sum_{(p,q)\in NR_s} F^{Fea}(p,q)}{N} \qquad (2)$$

where N is the number of pixels in region NR_s with the scale of *s*.

Using (1), the local saliency value of each pixel in the input image can be obtained. Then the local saliency value of each segmented region can be computed as

$$LocSal_i^{Fea} = \frac{\sum_{(x,y)\in I} LocContr_{(x,y)}^{Fea}}{N_i} \qquad (3)$$

where N_i denotes the number of pixels in region *i*, and *I* denotes the set of pixels belonging to region *i*.

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

Fig. 2 Local saliency maps generation. From left to right: (a) input image, (b) local saliency map generated by color feature, (c) local saliency map generated by orientation feature, (d) the final local saliency map, (e) detection and segmentation results based on the local saliency map.

Through (1) and (3), we can obtain two local saliency maps using color and orientation features respectively. Then the final local saliency can be generated by the combination of $LocSal_i^{color}$ and $LocSal_i^{orientation}$

$$LocSal_i = LocSal_i^{color} + \beta LocSal_i^{orientation} \quad (4)$$

where β denotes the weight decided by the ratio of entropy in color local saliency map to that of orientation local saliency map.

Fig. 2 shows the local saliency maps generated by the proposed method. Fig. 2 (b) and (c) are the local saliency maps generated by color and orientation features respectively, and Fig. 2 (d) shows the final local saliency map. Fig. 2 (e) shows the detection and segmentation results based on the local saliency map, only the region of hair is detected and segmented as the salient region.

C. Global Saliency Generation

As shown in Fig. 2, when we only use the local image information to detect the salient region, only some highly local contrasted parts of the salient region are detected and segmented from the input image. To obtain a better detection and segmentation result, we will introduce a global saliency detection process to work together with the local saliency detection process in this subsection. Saliency detection based on the global information has been proposed [6][11][12], and the proposed method detects the global saliency by considering the spatial distribution and feature isolation of each region simultaneously.

Since a wide dispersed feature (such as color and orientation used in this paper) is less possible to be contained in a salient region, global spatial distribution of a specific kind of feature is a good way to describe the salient region[6]. Therefore, the spatial distributions of different region can be used to evaluate the compactness and saliency of the region.

Let $Cen_i^{sp} = \begin{bmatrix} MeanX_i \\ MeanY_i \end{bmatrix}$ denotes the spatial center coordinate of the ith region. $MeanX_i$ and $MeanY_i$ are the mean values of coordinate x and y in the ith region respectively

$$MeanX_i = \frac{\sum_{(x,y)\in I} x}{N_i} \quad (5)$$

$$MeanY_i = \frac{\sum_{(x,y)\in I} y}{N_i} \quad (6)$$

The intra-spatial distribution of a region can be measured by the spatial variance of pixels in this region. There may be some regions with the same intra-spatial distribution, but the relative compactness may be different. Therefore, the relative compactness of a region with respect to the other region and the intra-spatial distribution should be used together to represent the compactness of a region. Then the spatial distribution of the ith region can be computed as

$$DISTRI_i = \sum_j \frac{\sum_{(x,y)\in I} \| X - Cen_j^{sp} \|^2}{N_i} \quad (7)$$

where $DISTRI_i$ denotes the spatial distribution of the ith region, and $X = [x, y]^T$ denotes the pixel coordinate. N_i denotes the number of pixels in region i, and I denotes the set of pixels belonging to region i.

The feature isolation shows the distinctiveness of a region from others in feature domain, which also contributes greatly to the saliency of a region.

Color and orientation features are used in the calculation of feature isolation respectively. The feature isolation of a region is measured by distance of feature vectors between pixels in this region and pixels in other regions. Then the feature isolation of the ith region can be computed as

$$ISO_i^{Fea} = \sum_j \frac{\sum_{(x,y)\in I} \| Fea(x,y) - Mean_j^{Fea} \|^2}{N_i} \quad (8)$$

where ISO_i^{Fea} denotes the feature isolation of the ith region, Fea={*color, orientation*}. $Fea(x,y)$ represents the fearure vector at location (x,y), and $Mean_j^{Fea}$ denotes the mean feature vector value of pixels in jth region.

For each region, the global saliency can be measured based on its spatial distribution and feature isolation. A region with lower spatial distribution in spatial domain and higher feature isolation in feature domain is more likely to be the salient region. Therefore, the global saliency of the ith region can be computed as

$$GloSal_i^{Fea} = \frac{ISO_i^{Fea}}{DISTRI_i} \quad (9)$$

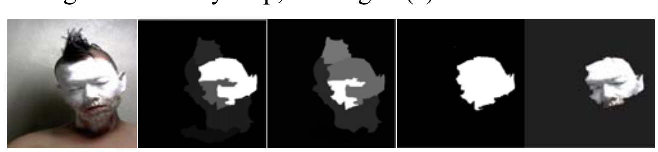
where $GloSal_i^{Fea}$ denotes the global saliency of the ith region for feature Fea.

Then, the final global saliency can be measured by the combination of the color global saliency and orientation global saliency (9)

$$GloSal_i = GloSal_i^{color} + \gamma GloSal_i^{orientation} \quad (10)$$

where γ denotes the weight decided by the ratio of entropy in color global saliency map to that of orientation global saliency map.

Fig. 3 shows the generation of global saliency map. Fig. 3(b) and (c) are the global saliency maps generated by color and orientation features, respectively. Fig. 3 (d) shows the final global saliency map, and Fig. 3 (e) shows the detection



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

Fig.3. Global saliency maps generation. From left to right: (a) input image, (b) global saliency map generated by color feature, (c) global saliency map generated by orientation feature, (d) the final global saliency map, (e) detection and segmentation results based on the global saliency map.

result based on global saliency. In next step, we will combine the local saliency with the global saliency together to obtain the final detection result.

D. Generation of Final Saliency Map

Based on the mutual complementarities between the local and global saliency, in this section, we will combine these two kinds of saliency together with different weights. The weight for each saliency map is determined automatically according to the information contained in the corresponding map.

The local map or global map is represented by n-bin histograms, and the corresponding entropy E is measured by

$$E = -\sum_{i=1}^{n} P_i \log_2(P_i) \tag{11}$$

where P_i is the probability that the pixel gray value falls into the ith bin. Higher value of entropy means fewer information in the map. The saliency map with more information should have higher influence in the generation of the final saliency map, and it should be with a bigger value of combination weight. The weight can be assigned as

$$W = E_{Max} - E \tag{12}$$

where E_{Max} means the maximum entropy value of a map which changes with the number of histogram bins.

Let W_{Loc} and W_{Glo} denote the combination weights for local and global saliency respectively, which can be calculated using the method above, then the final saliency of the ith region can be measured as

$$Sal_i = W_{Loc} \times LocSal_i + W_{Glo} \times GloSal_i \tag{13}$$

Fig. 4 shows the process of salient region detection with the proposed method. The final saliency map is generated by combination based on the mutual complementary of local saliency and global saliency. It obtains a more perfect result with regions of both hair and face being detected to be salient.

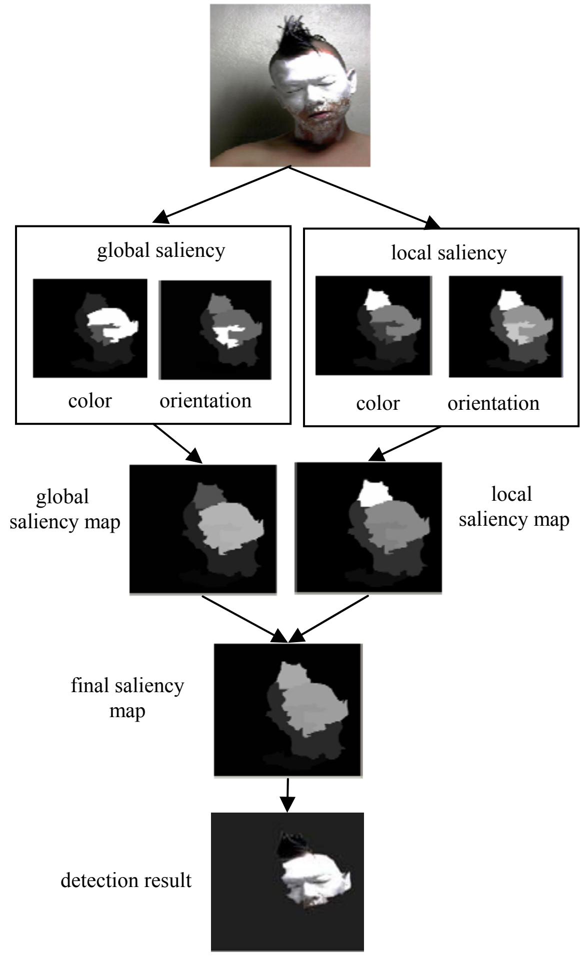
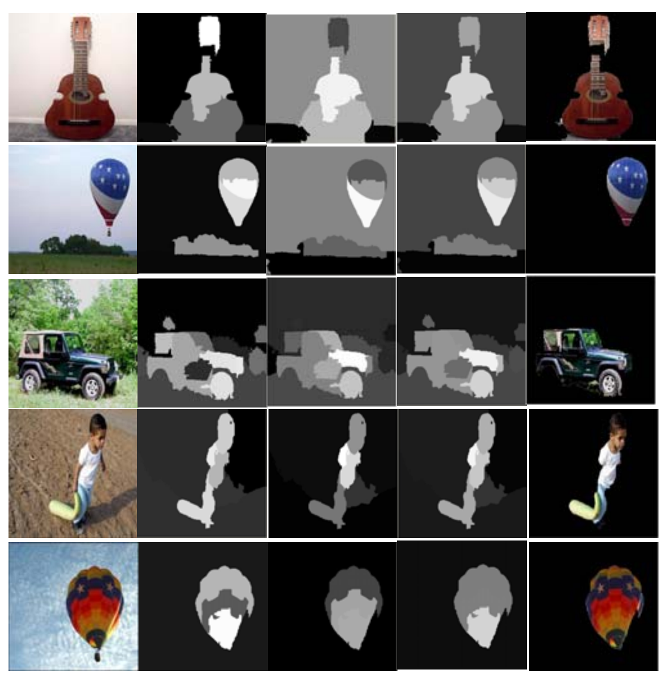


Fig. 4. Overview of the process of salient region detection with the proposed method.

III. Experimental Results

We experimentally evaluate the performance of the proposed method on images from the database MSRA (provided by Microsoft Research Asia) [6] using a computer with the configuration of Core duo CPU and 2G RAM. The proposed method is also compared with other saliency detection methods, such as the Itti's method [1] which is based on the local image information, and the Gopalakrishnan's method [11] using the global color information.

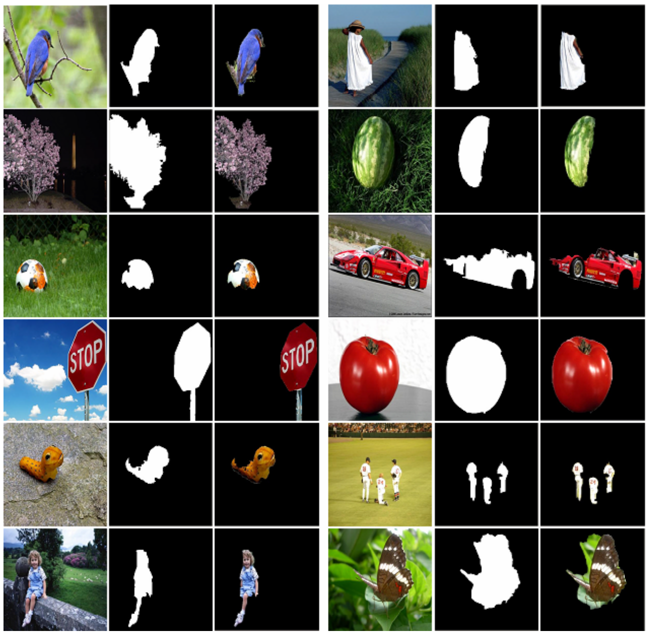
Fig. 5 shows the detection and segmentation performance of the proposed method. Fig. 5 (a), (b) and (c) are the input image, the local saliency map, and the global saliency map respectively. Fig. 5 (d) shows the binary saliency map of the proposed method with the combination of local and global saliency, and Fig. 5 (e) shows the detection results of the proposed method. The proposed method can effectively extract the most salient regions from the input images, and the mutual Complementary of the local and global saliency guarantees the completeness of the extracted salient region.



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

Fig.5. Salient region detection base on the proposed method. From left to right: (a) input image, (b) the local saliency map, (c) the global saliency map, (d) the binary map of the final saliency map, (e) final detection result.

Fig. 6 shows some experimental results of the proposed method. Fig. 6 (a), (b) and (c) are the input images, the binary



(a) (b) (c) (a) (b) (c)

Fig.6. Results of the proposed method. From left to right: (a) input image, (b) binary saliency map, (c) final detection results.

saliency map, and the detection and segmentation results, respectively. In most cases, the proposed method can effectively detect the salient regions from the input images.

In order to well evaluate the proposed method, the performance of the proposed method is compared with other saliency detection methods. Fig. 7 (a) shows the input images, and Fig. 7 (b), (c) and (d) show the saliency maps generated by the Itti's method [1], the Gopalakrishnan's method [11]

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

Fig.7. Saliency maps generated by different methods. From left to right: (a) input image, (b) Itti's method [1], (c) Gopalakrishnan's method [11], (d) the proposed method.

and the proposed method, respectively. The proposed method outperforms the other two methods. Fig. 8 shows the detection results using the proposed method (Fig. 8 (b)), and the frequency-tuned method [20] (Fig. 8 (c)).

We also evaluate the quantitative performance of different methods. The MSRA database we used contains images with salient regions marked as labeled rectangles, which is called "ground truth". Similar to [6], in this paper, an objective evaluation of the algorithm is carried out based on precision, recall and F-Measure.

Precision is the ratio of correctly detected salient region to the detected region, and recall is the ratio of correctly detected region to the "ground truth". F-Measure is an overall performance measurement defined as

$$F-Measure = \frac{(1+\alpha)\times precision \times recall}{(\alpha \times precison + recall)} \tag{14}$$

Fig. 9 shows the comparison of average precision, recall and f-measure values between the proposed method, the Itti's method [1], Gopalakrishnan's method [11], and frequency-tuned method [20] (Achanta's method) . The proposed method obtains a best detection and segmentation results over the other two methods which only use the local or global image information.

IV. Conclusion

In this paper, a new method of salient region detection based on local and global image information is proposed. The input image is firstly segmented into several regions, then for each region, the local saliency is computed by multi-scale neighborhood contrast, and the global saliency is measured according to global spatial distribution and inter-region isolation of features. The final saliency is obtained by the weighted combination of local saliency and global saliency, and the weights are determined based on the entropy of the obtained saliency maps. Experimental results demonstrate that effectiveness of the proposed method.

Acknowledgment

This work was partly supported by the NNSF (National Natural Science Foundation) of China under the grants 61100098, 61379097, 61033011 and 61210009.

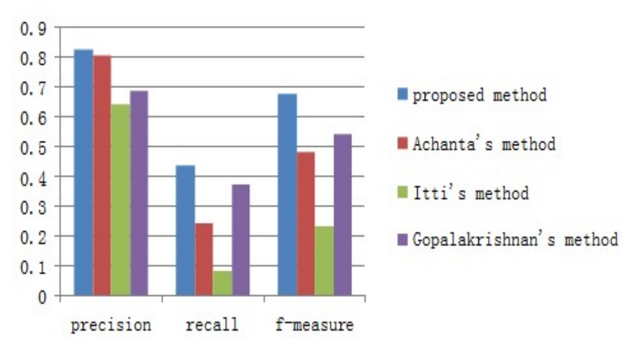
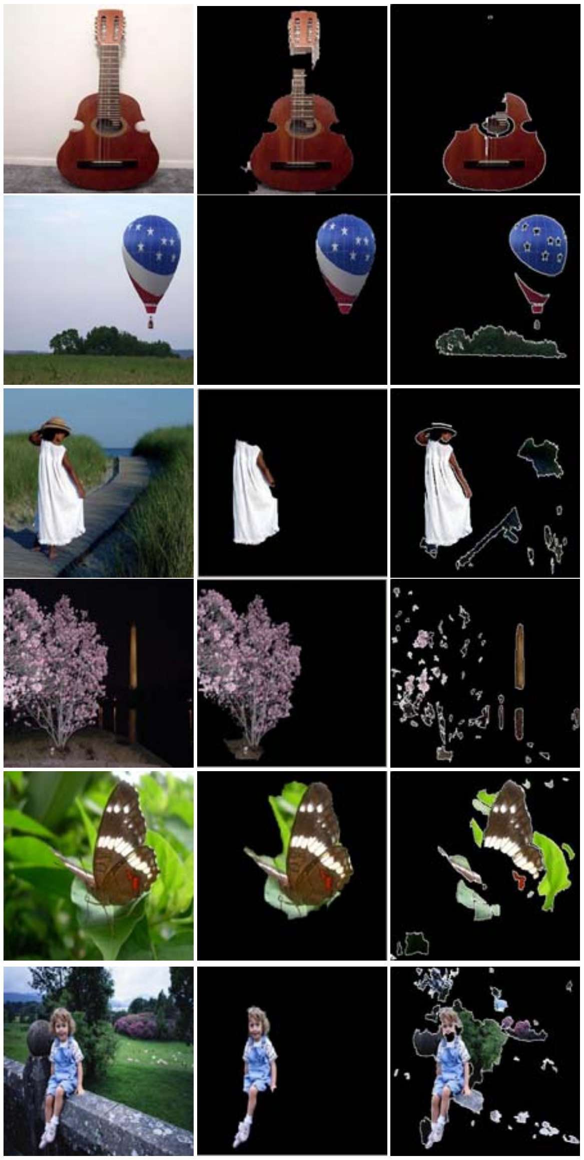


Fig. 9. Comparison of precision, recall, f-measure with other methods based on images from the MSRA database.



(a) (b) (c)

Fig. 8. Salient region detection results with different methods. From left to right: (a) input image, (b) proposed method, (c) frequency-tuned method [20] (Achanta's method).

REFERENCES

[1] 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.

[2] T. Kadir and M. Brady, "Saliency, scale and image description, " *International Journal of Computer Vision*, vol. 45, no. 2, pp. 83–105, 2001.

[3] S. J. Park, J. K. Shin and M. Lee, "Biologically inspired saliency map model for bottom-up visual attention," *Proc. of the Second International Workshop on Biologically Motivated Computer Vision*, pp. 418–426, 2002.

[4] S. Feng, D. Xu, X. Yang, "Attention-driven salientedge(s) and region(s) extraction with application to CBIR," *Signal Processing*, vol. 90, pp. 1-15, 2010.

[5] Y. F. Ma and H. J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," *Proc. of the eleventh ACM international conf. on Multimedia*, pp. 374-381, 2003.

[6] T. Liu, J. Sun, N. Zheng, X. Tang and H. Y. Shum, "Learning to detect a salient object," *Proc. of IEEE Computer Society Conf. on Computer and Vision Pattern Recognition*, pp. 1–8, 2007.

[7] M. Z. Aziz and B. Mertsching, "Fast and robust generation of feature maps for region-based visual attention," *IEEE Trans. on Image Processing,* vol. 17, no. 5, pp. 633–644, 2008.

[8] J. Weijer, T. Gevers and A. D. Bagdanov, "Boosting color saliency in image feature detection, " *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 28, no. 1, pp. 150–156, 2006.

[9] H. -Y. Chen, J. -J. Leou, "Saliency-directed image interpolation using particle swarm optimization, " *Signal Processing*, vol. 90, pp. 1676–1692, 2010.

[10] R. Achanta, F. Estrada, P. Wils and S. Susstrunk, "Salient region detection and segmentation, " *Proc. 6th international conf. on Computer vision systems*, pp. 66-75, 2008.

[11] V. Gopalakrishnan, Y.Q. Hu and D. Rajan, "Salient region detection by modeling distributions of color and orientation, " *IEEE Trans. on Multimedia*, vol. 11, pp. 5, pp. 892–905, 2009.

[12] C. M. Kuo, Y. H. Kuan and N. C. Yang, "Color-based image salient region segmentation using novel region merging strategy, " *IEEE Trans. on Multimedia*, vol. 10 no. 5, pp. 832–845, 2008.

[13] F. Jing, M. Li, H. Zhang and B. Zhang, "Unsupervised image segmentation using local homogeneity analysis, " Proc. of IEEE Int. Symposium on Circuits and Systems, 2003.

[14] V. Navalpakkam and L. *I*tti, "An integrated model of top-down and bottom-up attention for optimizing detection speed, " Proc. of IEEE Computer Society Conf. on Computer and Vision Pattern Recognition, vol. 2, pp. 2049–2056, 2006.

[15] S. Frintrop, G. Backer and E. Rome, "Goal-directed search with a top-down modulated computational attention system," *Proc. of the Annual Meeting of the German Association for Pattern Recognition*, pp. 117–124, 2005.

[16] F. Orabona, G. Metta and G. Sandini, "Object-based Visual Attention: a Model for a Behaving Robot," *3rd International Workshop on Attention and Performance in Computational Vision within CVPR,* 2005.

[17] Z. D. Li and J. Chen, "On Semantic Object Detection with Salient Feature," *Proc. of the 4th Int. Symposium on Advances in Visual Computing,* vol. 5359, no. 2, pp. 782–791, 2008.

[18] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, pp. 219–227, 1985.

[19] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," *Proc. of IEEE Computer Society Conf. on Computer and Vision Pattern Recognition*, pp. 1-8, 2007.

[20] R. Achanta, et al., "Frequency-tuned Salient Region Detection," *Proc. of IEEE Computer Society Conf. on Computer and Vision Pattern Recognition,* pp. 1597-1604, 2009.

[21] S. Goferman et al. "Context-Aware Saliency Detection, " *IEEE Trans.Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1915–1926, 2012.