Research Journal of Applied Sciences, Engineering and Technology 7(1): 42-48, 2014
DOI:10.19026/rjaset.7.218
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: January 25, 2013 Accepted: March 02, 2013 Published: January 01, 2014

Research Article

A Computational Model of Visual Attention Based on Space and Object

Shuhong Li and Qiaorong Zhang
College of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou, China

Abstract: Object-based visual attention has got more and more attention in image processing. A computational model of visual attention based on space and object is proposed in this study. Firstly spatial visual saliency of each pixel is calculated and edges of the input image are extracted. Salient edges are obtained according to the visual saliency of each edge. Secondly, a graph-based clustering process is done to get the homogeneity regions of the image. Then the most salient homogeneity regions are extracted based on their spatial visual saliency. Perceptual objects can be extracted by combining salient edges and salient regions. Attention value of each perceptual object is computed according to the area and saliency. Focus of attention is shifted among these perceptual objects in terms of the attention value. The proposed computational model was tested on lots of natural images. Experiment results indicate that our model is valid and effective.

Keywords: Perceptual object, saliency map, salient region, visual attention, visual saliency

INTRODUCTION

Human and other primates can find the objects of interest in complex visual scenes quickly. This mechanism is called selective visual attention. With the help of visual attention mechanism, human and the primates can concentrate on the important things for further processing and ignore other unimportant things. Based on attention mechanism, the processing priority and resources are given to the important or interested information firstly to improve the computation and reaction speed. Visual attention mechanism has attracted lots of researchers in computer vision and image processing field in recent years.

Some researchers have proposed several computational models for visual attention. These computational models can be divided into two categories: space-based visual attention and object-based visual attention. The most influential space-based computational model was proposed by Itti *et al*. (1998). They computed visual saliency of each pixel in the feature maps such as intensity feature map, color feature map and orientation feature map at first. Then they fused these feature saliency maps to get an integrated saliency map. Finally the focus of attention is got and shifted based on saliency value of each part in the integrated saliency map. Saliency map is important in most of space-based visual attention models. Visual saliency of each pixel in the image is calculated. The focus of attention is set to the pixel with the maximal saliency value. The attention area is a rectangle or a circle whose center is the pixel with the maximal saliency. And the size of the attention area may be fixed (Itti *et al*., 1998; Itti and Kouch, 2001) or variable (Itti and Kouch, 2003; Dirk, 2006). Based on space-based visual attention models, the focus of attention is shifted from one spatial area to another area. So the integrity and validity of the object cannot be guaranteed.

In recent years, many researchers have been researching object-based visual attention. It is indicated that "object" plays an important role in visual attention (Lamy and Tsal, 2000; Yaoru, 2003a). According to the object-based visual attention model, the unit of attention is "object" not spatial location or area. People can adjust the shape and size of the attention region based on the interested object in object-based visual attention model. This is the advantage of object-based model by contrast to space-based model. It is very suitable to use object-based visual attention model in object detection and recognition applications.

In the object-based computational model, the most important problem is to define and extract the objects. Some researchers have proposed several approaches to define and extract objects. In the object-based visual attention model proposed by Yaoru and Robert (2003b) and Yaoru *et al*. (2008), but they did not give the definition of the object. The object was extracted using image segmentation method. Zhao *et al*. (2006) proposed a perceptual object definition and extraction method. They defined and extracted the objects using the inner continuity and similarity of the region and difference between the region and its surroundings.

Corresponding Author: Shuhong Li, College of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou, China

Their method was not suitable and not valid in some clustered images. Another object extraction method based on multi-scale analysis and grouping was proposed by Zou *et al*. (2006). But they only took edge information into account and neglected other features. Shao *et al*. (2008) presented a perceptual object detection approach using intrinsic dimensionality. But this method required manual adjustment in some complex images.

Another important problem is the focus of attention shift method in object-based computational model. Most of the current methods compute the attention value only based on the average saliency of the pixels in the perceptual objects. The location and the size of perceptual objects are neglected.

In this study, a new computational model of visual attention based on space and object is proposed. In the proposed model, we define the perceptual object using salient region, edges and homogeneity region. We extract salient edges using a saliency map and an edges detection algorithm. Homogeneity regions are clustered using a graph-based clustering algorithm. Then perceptual objects are extracted by combining salient regions, edges and homogeneity regions. Attention value of each perceptual object is computed according to its average visual saliency and its location and size. Focus of attention of shifts among these perceptual objects based on their attention value.

COMPUTATIONAL MODEL

Figure 1 shows the diagram of our proposed computational model of visual attention based on space and object.

Visual saliency calculation: Visual saliency of each pixel in the image is determined by the feature value of it and its neighbors. Firstly, we should extract visual features of the image. In this study, we use some visual features which are intensity, color and orientation. HSI (Hue, Saturation and Intensity) color space is more consistent with human color perception system than RGB color space. Therefore, we transform the input image from RGB space to HSI space using (1).

$$\begin{cases} H = \dfrac{1}{360}[90 - Arc\tan(\dfrac{F}{\sqrt{3}}) + \{0, G > B; 180, G < B\}] \\ S = 1 - [\dfrac{\min(R,G,B)}{I}] \\ I = \dfrac{(R+G+B)}{3} \\ F = \dfrac{2R-G-B}{G-B} \end{cases} \quad (1)$$

Channel I in (1) represents the intensity feature of the input image. Color feature of the image can be represented by H (Hue) channel and S (Saturation) channel. We can get four orientation feature maps by filtering the intensity feature map using four Gabor filters with orientation 0°, 45°, 90°, 135°, respectively.

In our computational model, we calculate three kinds of visual saliency to measure the visual saliency of each part in the image more correctly. They are local saliency, global saliency and rarity saliency.

Local saliency of each pixel in the image is measured by the contrast between the pixel and its neighbor pixels. In frequency domain, we can get the magnitude spectrum and phase spectrum of an image (Xuelei and Xiaoming, 2002; Peter and Walter, 2002). It is proved that phase spectrum is more important in image reconstruction than magnitude spectrum. Phase spectrum describes the value changing information at each pixel. Magnitude spectrum includes the feature value at each pixel. Therefore, we can get the local saliency using (2) by reconstructing the image with phase spectrum only:

$$\begin{cases} F(u,v) = \sum_{x=1}^{M}\sum_{y=1}^{m} f(x,y) e^{\frac{-j2\pi ux}{M}} e^{\frac{-j2\pi vy}{N}} \\ \qquad = R(u,v) + jI(u,v) \\ P(u,v) = \arctan(\dfrac{I(u,v)}{R(u,v)}) \\ S_{local}(x,y) = \dfrac{1}{M*N}\sum_{u=1}^{M}\sum_{v=1}^{N} I(u,v) e^{\frac{-j2\pi ux}{M}} e^{\frac{-j2\pi vy}{N}} \end{cases} \quad (2)$$

where,
f (x, y) : The value of pixel (x, y) in the feature map

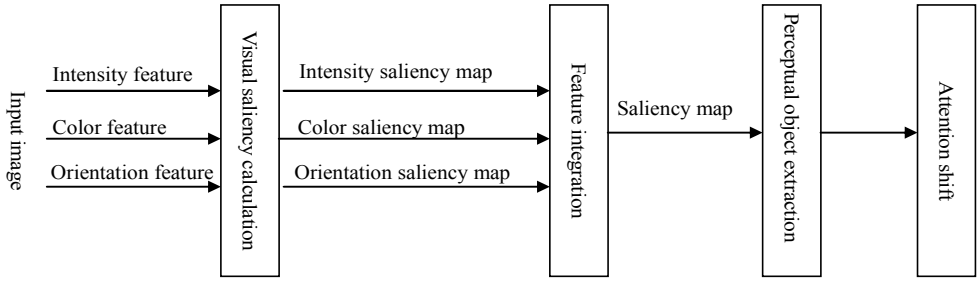

Fig. 1: Block diagram of the proposed model

Fig. 2: Example of saliency calculation

F (u, v) : The Discrete Fourier transform coefficients of f (x, y)
S_{local} (x, y) : The local saliency value of pixel (x, y)

Local saliency measures the contrast between the pixel and its neighbors so that the boundaries often have high local saliency values. Therefore we use global saliency besides local saliency. Global saliency of the pixels in a feature map can be calculated using (3):

$$\begin{cases} S_{Global}(x,y) = e^{\frac{|f(x,y)-f_{avg}(x,y)|}{f_{avg}(x,y)}} \\ f_{avg}(x,y) = \frac{1}{M*N}\sum_{x=1}^{M}\sum_{y=1}^{N} f(x,y) \end{cases} \quad (3)$$

In the proposed model, rarity saliency is also being computed. If the feature value of a pixel occurs less time, the rarity saliency value of the pixel is larger. The rarity saliency can be computed using (4):

$$S_{Rarity}(x,y) = \frac{1}{hist(f(x,y))} \quad (4)$$

where,
f (x, y) : The feature value of pixel (x, y) in the feature map
hist (·) : The histogram of the feature map

After local saliency map, global saliency map and rarity saliency map are generated, we should combine them into a feature conspicuity map. According to the effect of each saliency map, they are assigned different weights. We can get the feature conspicuity map using (5):

$$\begin{cases} V = \frac{1}{M*N}\sum_{i=1}^{M}\sum_{j=1}^{N}\left| f(x,y - \frac{1}{M*N}\sum_{i=1}^{M}\sum_{j=1}^{N} f(x,y)\right| \\ w_i = \frac{V_i}{\sum_{i=1}^{3} V_i} \\ C_F = w_1 * S_{Local} + w_2 * S_{Global} + w_3 * S_{Rarity} \end{cases} \quad (5)$$

The examples of local saliency, global saliency, rarity saliency and feature conspicuous map are shown in Fig. 2. Original images are shown in the top row and the images in the second row are their intensity feature maps. The local saliency maps, global saliency maps, rarity saliency maps and the intensity feature conspicuous maps are shown in the third, fourth, fifth and sixth row respectively.

Feature integration: When the feature conspicuity maps are integrated into an integration saliency map, they should be given different weights according to their effects on the overall saliency map (Yiqun *et al*., 2004). Four combination strategies are proposed in (Itti and Kouch, 2003). In the naive linear combination method, an equal weight was given to all the different features. So the method could not get a satisfying result. The second method was linear combination with learned weights. This method was better but it required a prior knowledge of the salient regions. The third method was global non-linear normalization method. The fourth method was the iterative non-linear method. Those two methods both used a local competition strategy in one feature map.

We propose a new strategy and process of feature integration. Salient area, salient point location and salient point distribution are used to measure the importance of the feature conspicuity maps and to calculate the weights.

Firstly salient points are needed to be extracted. We simply segment these feature saliency maps using a threshold T. Then we get a binary version of the feature conspicuity map using (6):

$$B(x,y) = \begin{cases} 1 & if \ \ C_F(x,y) \geq T \\ 0 & otherwise \end{cases} \quad (6)$$

where,
T : The threshold. We can compute the threshold using the MATLAB gray level function
C_F : The conspicuity map of different feature

In the obtained binary maps, the pixels with value 1 are defined as the salient points. Based on the rarity principle, if a feature saliency map has more salient points, the importance of the feature saliency map is less. We consider the number of salient points as the measure of salient point area using (7):

$$W_{area} = N \quad (7)$$

where, W_{area} represents the weight of salient point area and N means the number of salient points. When the number of salient pixels is larger than 70% of the area of the whole image, the weight of the feature saliency map is set to zero. This means it is not included when in feature integration.

Secondly, the regions near image center are often paid more attention to so that the regions near the image

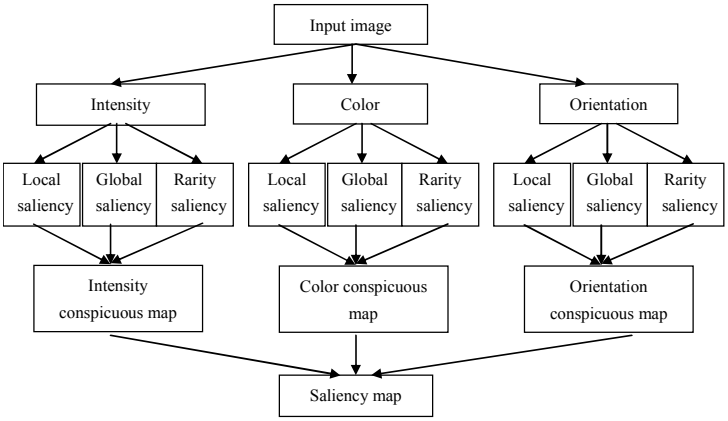

Fig. 3: Diagram of feature integration

center are more likely to be salient regions. So the salient point location can be considered as a criterion. Calculate the average distance between the salient points and the image center as the location criterion using (8):

$$W_{location} = \frac{1}{N}\sum_{i=1}^{N} Dist(sp_i, center) \quad (8)$$

where,
$W_{location}$: The weight of salient point location
N : The number of salient points
sp_i : Each salient point and center means the center of the image
Dist : The distance between two points

In addition, if the salient regions distribute separately in the feature saliency map, the feature saliency map is not very useful. So we compute the spatial distribution of salient points using (9) as another criterion:

$$W_{distribution} = \frac{1}{N}\sum_{i=1}^{N} Dist(sp_i, centroid) \quad (9)$$

where,
$W_{distribution}$: The weight of salient point spatial distribution
centroid : The center of the salient points

Finally, these feature conspicuity maps are fused together using (10):

$$\begin{cases} SM = \sum_{i=1}^{m} W_i * C_F^i \\ W_i = \dfrac{\frac{1}{W_{fi}}}{\sum_{i=1}^{m}\frac{1}{W_{fi}}} \\ W_{fi} = W_{area}^i + W_{location}^i + W_{distribution}^i \end{cases} \quad (10)$$

where,
SM : The integration saliency map
C_F^i : Each feature saliency map

The diagram of feature competition is shown in Fig. 3.

Perceptual object extraction: The process of perceptual object extraction has been described in our study (Qiaorong and Yafeng, 2010) in detail. Here we just give the main idea in short.

The perceptual object in our model is defined using (11) based on Gestalt theory and feature contrast theory. We use three attributes to describe a perceptual object which are homogeneity measure, edge (s) and saliency measure. Each attribute is described in detail below:

$$PO = \{HM, ES, SM\} \quad (11)$$

HM is the homogeneity measure of the inner part of an object. Based on the Gestalt perceptual organization theory, the pixels in an object are similar with each other in intensity, orientation, color, shape, texture or other features. We can cluster the pixels in the image into several homogeneity regions using graph-based clustering algorithm. An object may contain one or some homogeneity regions.

ES represents edges of the object. Based the Gestalt perceptual organization theory, an object is often closed and continuous. Edges mean the contour of an object and satisfy the Gestalt law of closure and continuity. Edges of the object can be extracted using edge detection algorithm. Then according to some rules, some tiny or unimportant edges are neglected and only important edges are preserved.

SM means the saliency measure of a perceptual object. Saliency of an object means the feature contrast between the object and its surroundings. In next section, we will introduce the algorithm to calculate saliency based on feature contrast. Based on the saliency value, saliency map of the input image will be obtained. According to the saliency map, we can extract salient regions in the image. Salient regions include pixels with

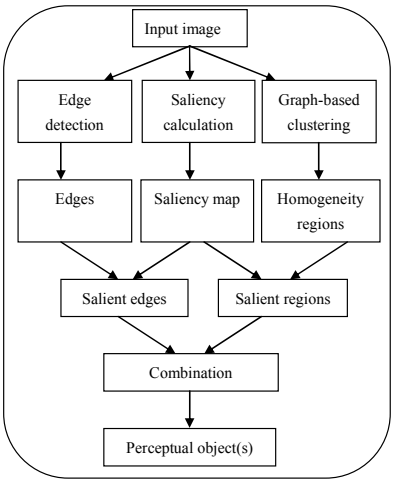

Fig. 4: Framework of perceptual object extraction

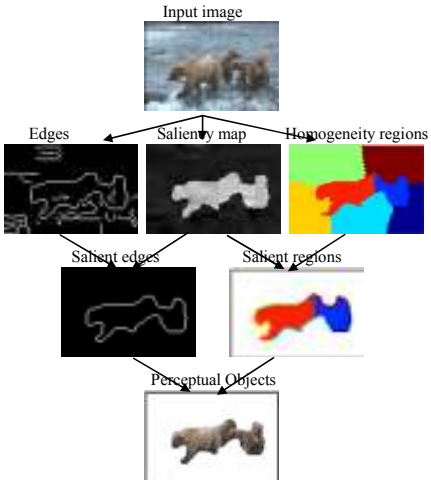

Fig. 5: Example of perceptual objects extraction

large saliency value. These regions are likely to contain objects and get the attention.

Perceptual objects can be extracted in an image based on the definition of perceptual object. The framework of extracting perceptual object is shown in Fig. 4. The detail of perceptual object extraction algorithm is described below:

- Calculate saliency value of each pixel based on feature value contrast and generate saliency map of the input image.
- Detect edges of the image. Extract salient edges and neglect edges which are not salient based on the saliency map.
- Cluster pixels which are near and similar in intensity, color and orientation using graph-based clustering algorithm. The input image is grouped into several homogeneity regions.
- Extract salient region (s) based on the saliency map and homogeneity regions. Salient regions are likely to contain objects and get the attention.
- Extract perceptual objects by combining salient regions and salient edges.

The example of perceptual object extraction is shown in Fig. 5.

Attention shift: Focus of attention shifts among the extracted perceptual objects according to their attention value. The attention value of each perceptual object is computed using (12) based on its average saliency value and its size:

$$\begin{cases} AV(O_i) = w_a \times Size(O_i) + w_s \times S_{avg}(O_i) \\ S_{avg}(O_i) = \dfrac{\sum_{j=1}^{Size(O_i)} S(x_j, y_j)}{Size(O_i)} \end{cases} \quad (12)$$

where,

$AV(O_i)$: The attention value of perceptual object O_i

$Size(O_i)$: The size of perceptual object O_i and it equals the number of its pixels

$S_{avg}(O_i)$: The average saliency of perceptual object O_i

$S(x, y)$: The average saliency value of the 3*3 neighbor field of pixel (x, y)

w_a & w_s : Weights of size and average saliency

Inhibition of return mechanism is used in attention shift. That means when one salient region has been focused on, the saliency of every pixel in the salient region is set to 0 using (13). Thus when we select another salient region, the salient regions which have been focused on won't be selected again:

$$AV(O_i) = \begin{cases} 0 & if\ O_i\ has\ been\ focused \\ AV(O_i) & otherwise \end{cases} \quad (13)$$

The example of perceptual object extraction and attention shift will be shown in below section.

EXPERIMENTAL RESULTS

To evaluate the performance of the proposed computation model of selective visual attention, we have tested it in many natural images. These images are downloaded from the internet or taken with a digital camera. The experiment results and analysis are described in detail.

Results: The proposed model has been tested on the computer with Intel Pentium 1.8 GHz and 512 M memory using more than 100 natural scene images. Figure 6 shows an input image.

The saliency map of the input image is shown in Fig. 7.

The extracted perceptual objects and the process of attention shift are shown in Fig. 8 (the size of the perceptual object is zoomed). The process of attention shift is illustrated by the order of perceptual objects.

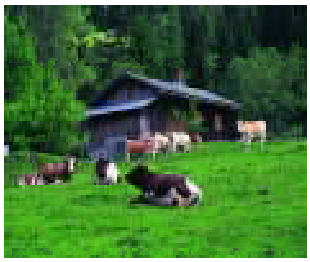

Fig. 6: The input image

Fig. 7: Saliency map

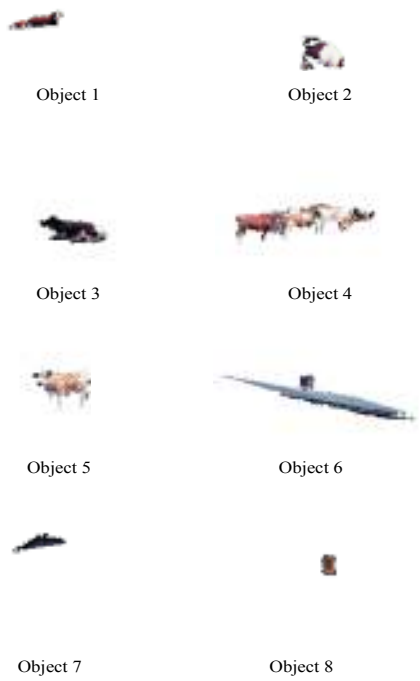

Fig. 8: Perceptual objects and attention shift

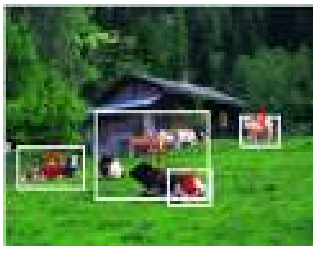

Fig. 9: Example of attention shift in space-based model

Comparison: To evaluate the validation of our proposed model, we also implemented the only space-based computational model. The attention shift process is shown in Fig. 9.

Comparing these two kinds of attention shift method (Fig. 8 and 9), we can find that the object-based attention model is reasonable and more appropriate to image processing and object recognition. In space-based model, the attention region is a rectangle or circle region whose center is the pixel with the maximal saliency value. The focus of attention is shifted only from one spatial region to another region. These problems are solved in object-based attention model.

DISCUSSION

The space-based model has some problems. From Fig. 9 we can see some selected regions don't contain a whole object and some selected regions include some background area. Sometimes attention will shift to a region which has no meanings. This is the defect of space-based attention model.

The object-based model can avoid the above problem of space-based model to a certain extent. But the object-based attention model needs perceptual clustering to segment the image into some regions before attention stage. And the same problem may exist in segmentation results. From Fig. 8 we can see the extracted objects sometimes contain several true objects and sometimes are only parts of an object. So the object in the model means perceptual object and is not the semantic object of the real world. We need more semantic knowledge to extract actual semantic objects.

CONCLUSION

We proposed a computational model for selective visual attention based on space and object in this study. Firstly, the spatial visual saliency is calculated. Then the salient edges and regions are extracted based on the spatial saliency. These salient edges and regions are combined to extract perceptual objects. According to the attention value of each perceptual object, the focus of attention is selected and shifted. This computational model can be used in image processing and actual machine vision.

Early vision features are important to construct the saliency map. A simple feature can not entirely represent the character of the salient region. Therefore, multiple features analysis is used in the proposed method. In this study, we consider colors, intensity as the features of the image. However, it is very likely that there are some other features such as edge and symmetry feature which also should be considered. What feature and how many features should be extracted according to the target will also be included in future study.

Some experiment results and discussion have been presented in this study. The proposed model is based on space and object. The problems of only space-based model are solved in the proposed model. But the proposed model also has some problems which need to be solved in the future study. Only bottom-up visual attention is researched in our model. The research on top-down visual attention to improve the saliency map will be included in future study.

ACKNOWLEDGMENT

This study was supported in part by a grant from Science Technology Project of Henan Province of China under Grant (No. 122300410379).

REFERENCES

Dirk, W., 2006. Interactions of visual attention and object recognition: Computational modeling, algorithms and psychophysics. Ph.D. Thesis, California Institute of Technology.

Itti, L, C. Kouch and E. Niebur, 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE T. Pattern Anal., 20: 1254-1259.

Itti, L. and C. Kouch, 2001. Computational modeling of visual attention. Nat. Rev. Neurosci., 2: 194-230.

Itti, L. and C. Kouch, 2003. Feature combination strategies for saliency-based visual attention systems. J. Elec. Imag., 10: 161-169.

Lamy, D. and Y. Tsal, 2000. Object features, object locations and object files: Which does selective attention activate and when. J. Exp. Psychol. Hum. Percept. Perform., 26: 1387-1400.

Peter, J.B. and M. Walter, 2002. Spatial frequency, phase and the contrast of natural images. J. Opt. Soc. Am., 19: 1096-1106.

Qiaorong, Z. and Z. Yafeng, 2010. Perceptual objects extraction based on saliency and clustering. J. Multimed., 5: 393-400.

Shao, J., J. Gao, Y. Zhao and X. Zhang, 2008. Perceptual object detection algorithm based on image intrinsic dimensionality. Chinese J. Sci. Instrum., 29: 810-815.

Xuelei, N. and H. Xiaoming, 2002. Statistical interpretation of the importance of phase information in signal and image reconstruction. Elsevier Science, June, 2002.

Yaoru, S., 2003a. Hierarchical object-based visual attention for machine vision. Ph.D. Thesis, University of Edinburgh.

Yaoru, S. and F. Robert, 2003b. Object-based visual attention for computer vision. Artif. Intell., 146: 77-123.

Yaoru, S., F. Robert, W. Fang and M.G. Herman, 2008. A computer vision model for visual-object-based attention and eye movements. Comput. Vis. Image Und., 112: 126-142.

Yiqun, H., X. Xing and M. Wei-Ying, 2004. Salient region detection using weighted feature maps based on the human visual attention model. LNCS, 3332: 993-1000.

Zhao, X.P., L. Wang and H. Zhan-Yi, 2006. A perceptual object based attention mechanism for scene analysis. J. Image Graph., 11: 281-288.

Zou, Q., S. Luo and Y. Zheng, 2006. A computational model of object-based attention using multi-scale analysis and grouping. Acta Electron. Sinica, 34: 559-562.