Visual Attention Model Based on Particle Filter

Long Liu[1], Wei Wei[2,*], Xianli Li[1], Yafeng Pan[1] and Houbing Song[3]
[1] Key Laboratory of Shaanxi Province for Complex System Control and Intelligent Information Processing,
School of Automation and Information Engineering, Xi'an University of Technology
Xi'an, Shaanxi 710048 - China
[e-mail: liulong2010@139.com, lizizimu@126.com, daniel0914@126.com]
[2]School of Computer Science and Engineering, Xi'an University of Technology
Xi'an, Shaanxi 710048 - China
[e-mail: weiwei@xaut.edu.cn]
[3]Department of Electrical and Computer Engineering, West Virginia University
Montgomery, WV 25136 - USA
[e-mail: Houbing.Song@mail.wvu.edu]
*Corresponding author: Wei Wei

Received December 20, 2015; revised April 25, 2016; accepted July 7, 2016; published August 31, 2016

Abstract

The visual attention mechanism includes 2 attention models, the bottom-up (B-U) and the top-down (T-D), the physiology of which have not yet been accurately described. In this paper, the visual attention mechanism is regarded as a Bayesian fusion process, and a visual attention model based on particle filter is proposed. Under certain particular assumed conditions, a calculation formula of Bayesian posterior probability is deduced. The visual attention fusion process based on the particle filter is realized through importance sampling, particle weight updating, and resampling, and visual attention is finally determined by the particle distribution state. The test results of multigroup images show that the calculation result of this model has better subjective and objective effects than that of other models.

Keywords: fusion, particle filter, visual attention model, bayesian filter posterior, resampling

This work was supported by Science and technology research and development program of Shaanxi Province (2016GY-128) and the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20136118120010) and NSFC Grant (No. 11301414 & No.11226173).

http://dx.doi.org/10.3837/tiis.2016.08.020 **ISSN : 1976-7277**

1. Introduction

Visual attention research centers on computer stimulations of the primary response of the human visual system to the outside world. The main research task is to construct a rational mathematical model to calculate sensory effects similar to those of human vision. The research result is of crucial importance to target detection, tracking, and related technological development.

Itti and Koch [1–3] established for the first time a bottom-up (B-U) visual attention model focused on light intensity, color, and orientation features. By extracting those primary visual features, 3 visual channels were calculated both in parallel and independently. Ultimately, the result of visual attention was obtained through fusion calculation. Studies of the human visual mechanism show that attention is formed by the interaction of B-U and top-down (T-D) attention. This mechanism is called bidirectional attention. Itti and Koch's model did not include the influence of the human subjective thinking process, i.e., the influence of T-D visual information on attention, and thus failed to reflect the overall actual effects of human visual attention. Research concerning biological vision still does not accurately describe the formation mechanism of attention. For this reason, the construction of a bidirectional attention model [4, 5] remain a challenging problem for studies in this field. Bidirectional visual models are mainly divided into weighting fusion models (WFMs) [6, 7], bias control models (BCMs) [8–11], and pattern classification models (PCMs) [12–14].

The WFM obtains a final attention saliency map mainly through the fusion of B-U and T-D attention by linear weighting. Zhang Qin et al. [6] obtained a facial attention map by weighting fusion of the attention to human facial color and attention to target features such as color, intensity, and orientation. Fang et al. [7] took the directional features of the target as T-D attention, which was fused with Itti's attention model through proportional weighting. Note that that the fusion weighting coefficient of the WFM is set by experience and will lead to instability of attention results because it cannot adapt to visual image changes.

The BCM adjusts the 3 feature-fusion weighting coefficients of Itti's model by the bias value of T-D attention. Zhang Jing et al. [8] took similar distances as a T-D factor to regulate the corresponding B-U-featured weight coefficient to obtain an attention salience map. Alcides et al. [9] carried out self-organized neural network learning for particular static target features to generate the weight, thus regulating the calculation process of B-U attention to produce its attention salience map. Yu et al. [10, 11] established long-term-memory units of target features and calculated the distribution bias of the position probability by comparison with low-level features. For the BCM, false results are frequent owing to the difficulty in determining the control parameter.

The PCM realizes the learning process by taking T-D and B-U attention produced by the sample and human visual

attention as the input and output of the classifier respectively, and uses the decided classifier as the model to produce attention. The gaze-prediction model by Robert et al. [12] established the relations between areas with high attention salience and the related positions of the task. Judd et al. [13] held that such a relation is a nonlinear mapping one that can be resolved by training the classifier. Ali Borji [14] adopted AdaBoost to train several weak classifiers in his research on the nonlinear mapping relation between features and the visual attention salience map. The PCM model needs many samples to train the classifier, so it is largely influenced by sample features.

The above studies show that the fusion weighting coefficient of the WFM is set by experience, cannot adapt to visual image change, and will lead to instability of attention results. For the BCM model, false results are frequent owing to the difficulty in determining the control parameter. The PCM model needs many samples to train the classifier, and thus is largely influenced by sample features.

To reduce these inadequacies, this paper proposes a bidirectional visual fusion attention model based on particle filter. This model does not need to set a parameter or train a sample, is determined by the sparse distribution of particles, and will generate an attention saliency map in the end.

2. Bidirectional attention model

A bidirectional attention model based on particle filter was constructed. With B-U and T-D attention taken as its input value, this model controlled the importance sampling of particles through B-U attention, calculated and updated the particle weight, and then changed the distribution state of particles by resampling. The framework is shown in **Fig. 1**.

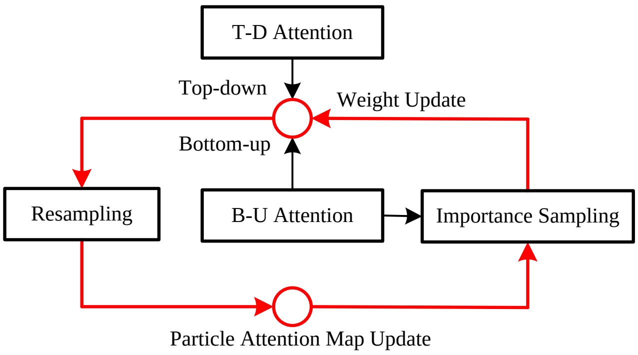


Fig. 1. Bidirectional (T-D and B-U) attention model based on particle filter

In **Fig. 1**, the most important step is the calculation of posterior probability for a weight update. There were 2 assumptions: (i) the time dynamic process conformed to the Markov one and (ii) observed values at different times were independent of each other and related only to the current state.

According to the particle filter theory, let the particle state be $\{X^{(i)}\}_{i=1}^{N}$ and the observed value Z^k. The density function of posterior probability at the k moment is approximately

$$p(X^T \mid Z^T) \approx \sum_{i=1}^{N} \lambda^{(i)} \cdot \delta(X^T - X_{(t)}^T) \tag{1}$$

where $\lambda^{(i)} \infty \dfrac{\pi(\hat{x}^{(i)})}{q(\hat{x}^{(i)})}$, $\pi(\cdot)$ is a probability density function, $q(\cdot)$ is an importance density function, and $i = 1, 2, \ldots, N$. The state of the bidirectional fusion attention saliency map is

assumed to be $x_{0:t}$. The observed value of B-U attention is $z_{1:t}^{T-D}$ and T-D attention $z_{1:t}^{B-U}$. In this paper, B-U and T-D attention are regarded as the input of Bayesian fusion estimation, so the posterior probability $p(X^T \mid Z^T)$ can be represented as $p(x_{0:t} \mid z_{1:t}^{B-U}, z_{1:t}^{T-D})$.
For particle weight updating calculation, the recurrence relation between $p(x_{0:t} \mid z_{1:t}^{B-U}, z_{1:t}^{T-D})$ and $p\left(x_{0:t} \middle| z_{1:t-1}^{B-U}, Z_{1:t-1}^{T-D}\right)$. Then the posterior probability value of the bidirectional fusion attention can be solved and deduced as follows:

$$p(x_{0:t} \mid z_{1:t}^{B-U}, z_{1:t}^{T-D})$$

$$= \frac{p(z_t^{B-U} \mid x_{0:t}, z_{1:t-1}^{B-U}, z_{1:t}^{T-D})\, p(x_{0:t} \mid z_{1:t-1}^{B-U}, z_{1:t}^{T-D})}{p(z_t^{B-U} \mid z_{1:t-1}^{B-U}, z_{1:t}^{T-D})} \tag{2}$$

$$= \frac{p(z_t^{B-U} \mid x_t, z_{1:t}^{T-D})\, p(x_{0:t} \mid z_{1:t-1}^{B-U}, z_{1:t}^{T-D})}{p(z_t^{B-U} \mid z_{1:t-1}^{B-U}, z_{1:t}^{T-D})} \tag{3}$$

$$= \frac{p(z_t^{T-D} \mid x_t, z_t^{B-U})\, p(z_t^{B-U} \mid x_t)}{p(z_t^{B-U} \mid z_{1:t-1}^{B-U}, z_{1:t}^{T-D})\, p(z_t^{T-D} \mid x_t)} \frac{p(z_t^{T-D} \mid x_{0:t}, z_{1:t-1}^{B-U}, z_{1:t-1}^{T-D})\, p(x_{0:t} \mid z_{1:t-1}^{B-U}, z_{1:t-1}^{T-D})}{p(z_t^{T-D} \mid z_{1:t-1}^{B-U}, z_{1:t}^{T-D})} \tag{4}$$

$$= \frac{p(z_t^{T-D} \mid x_t, z_t^{B-U})\, p(z_t^{B-U} \mid x_t)}{p(z_t^{B-U} \mid z_{1:t-1}^{B-U}, z_{1:t}^{T-D})\, p(z_t^{T-D} \mid x_t)} \frac{p(z_t^{T-D} \mid x_{0:t}, z_{1:t-1}^{B-U}, z_{1:t-1}^{T-D})\, p(x_t \mid x_{0:t-1}, z_{1:t-1}^{B-U}, z_{1:t-1}^{T-D})}{p(z_t^{T-D} \mid z_{1:t-1}^{B-U}, z_{1:t}^{T-D})} p(x_{0:t-1} \mid z_{1:t-1}^{B-U}, z_{1:t-1}^{T-D}) \tag{5}$$

$$= \frac{p(z_t^{T-D} \mid x_t, z_t^{B-U})\, p(z_t^{B-U} \mid x_t)}{p(z_t^{B-U} \mid z_{1:t-1}^{B-U}, z_{1:t}^{T-D})\, p(z_t^{T-D} \mid x_t)} \frac{p(x_t, z_t^{T-D} \mid x_{0:t-1}, z_{1:t-1}^{B-U}, z_{1:t-1}^{T-D})}{p(z_t^{T-D} \mid z_{1:t-1}^{B-U}, z_{1:t}^{T-D})} p(x_{0:t-1} \mid z_{1:t-1}^{B-U}, z_{1:t-1}^{T-D}) \tag{6}$$

$$= \frac{1}{k_t} p(z_t^{B-U} \mid x_t)\, p(z_t^{T-D} \mid x_t, z_t^{B-U})\, p(x_t, z_t^{T-D} \mid x_{0:t-1}) \frac{1}{p(z_t^{T-D} \mid x_t)} p(x_{0:t-1} \mid z_{1:t-1}^{B-U}, z_{1:t-1}^{T-D}) \tag{7}$$

$$= \frac{1}{k_t} p(z_t^{B-U} \mid x_t)\, p(z_t^{T-D} \mid x_t, z_t^{B-U})\, p(x_t \mid x_{0:t-1})\, p(z_t^{T-D} \mid x_t, x_{0:t-1}) \frac{1}{p(z_t^{T-D} \mid x_t)} p(x_{0:t-1} \mid z_{1:t-1}^{B-U}, z_{1:t-1}^{T-D}) \tag{8}$$

$$= \frac{1}{k_t} p(z_t^{B-U} \mid x_t)\, p(z_t^{T-D} \mid x_t, z_t^{B-U})\, p(x_t \mid x_{0:t-1})\, p(x_{0:t-1} \mid z_{1:t-1}^{B-U}, z_{1:t-1}^{T-D}) \tag{9}$$

The above deduction steps are simplified based on assumptions (i) and (ii), as follows:

i. From deducing steps in Equations (2) and (3), $p(z_t^{B-U} \mid x_{0:t}, z_{1:t-1}^{B-U}, z_{1:t}^{T-D}) = p(z_t^{B-U} \mid x_t, z_{1:t}^{T-D})$;

ii. similarly, from Equations (6) and (7), $p(x_t, z_t^{T-D} \mid x_{0:t-1}, z_{1:t-1}^{B-U}, z_{1:t-1}^{T-D}) = p(x_t, z_t^{T-D} \mid x_{0:t-1})$;

iii. the denominator of Equation (6), density $p(z_t^{B-U} \mid z_{1:t-1}^{B-U}, z_{1:t}^{T-D})$ and $p(z_t^{T-D} \mid z_{1:t-1}^{B-U}, z_{1:t}^{T-D})$ are independent of x_t and can be regarded as the constant k_t;

iv. and in accordance with assumption (ii), Equation (8) can be simplified as $p(z_t^{T-D} \mid x_t, x_{0:t-1}) = p(z_t^{T-D} \mid x_t)$.

According to the importance sampling theorem, the direct ratio of particle weight $\lambda^{(i)}$ is $\frac{p(\hat{x}_{0:t}^{(i)} \mid z_{0:t})}{q(\hat{x}_{0:t}^{(i)} \mid z_{0:t})}$, and can be represented as follows:

$$\lambda_t^{(i)} \propto \frac{p(\hat{x}_{0:t}^{(i)} \mid z_{0:t})}{q(\hat{x}_{0:t}^{(i)} \mid z_{0:t})} = \frac{p(z_t^{B-U} \mid \hat{x}_t^{(i)}) p(z_t^{T-D} \mid \hat{x}_t^{(i)}, z_t^{B-U}) p(\hat{x}_t^{(i)} \mid \hat{x}_{t-1}^{(i)}) p(\hat{x}_{0:t-1}^{(i)} \mid z_{1:t-1}^{B-U}, z_{1:t-1}^{T-D})}{q(\hat{x}_t^{(i)} \mid \hat{x}_{t-1}^{(i)}, z_t) q(\hat{x}_{t-1}^{(i)} \mid z_{t-1})} \tag{10}$$

$$\lambda_t^{(i)} \propto \lambda_{t-1}^{(i)} \cdot \frac{p(z_t^{B-U} \mid \hat{x}_t^{(i)}) p(z_t^{T-D} \mid \hat{x}_t^{(i)}, z_t^{B-U}) p(\hat{x}_t^{(i)} \mid \hat{x}_{t-1}^{(i)})}{q(\hat{x}_t^{(i)} \mid \hat{x}_{t-1}^{(i)}, z_t)} \approx \lambda_{t-1}^{(i)} \cdot p(z_t^{B-U} \mid \hat{x}_t^{(i)}) p(z_t^{T-D} \mid \hat{x}_t^{(i)}, z_t^{B-U}) \tag{11}$$

According to Equation (11), $p(z_t^{B-U} \mid \hat{x}_t^{(i)})$ represents the conditional probability of B-U attention observation in the current particle attention state, but $p(z_t^{T-D} \mid \hat{x}_t^{(i)}, z_t^{B-U})$ is that of T-D attention observation in B-U observation and the current particle attention state. Two values, $p(z_t^{B-U} \mid \hat{x}_t^{(i)})$ and $p(z_t^{T-D} \mid \hat{x}_t^{(i)}, z_t^{B-U})$, directly determine the weight value of the updated particles.

3. Calculation of attention saliency map

In this model, the salience of B-U visual attention is written as SM^{B-U} and that of T-D attention as SM^{T-D}. According to the bidirectional fusion attention model based on particle filter in Section 2, SM^{B-U} and SM^{T-D} are regarded as input values of the model to estimate the salience; the process of estimation is as follows: first, the importance sampling based on the value of SM^{B-U} is completed; then the weight of the particle filter is calculated by Equation (11); finally, the ultimate attention salience map is determined by the particle distribution density. The overall algorithm framework is shown in **Fig. 2**.

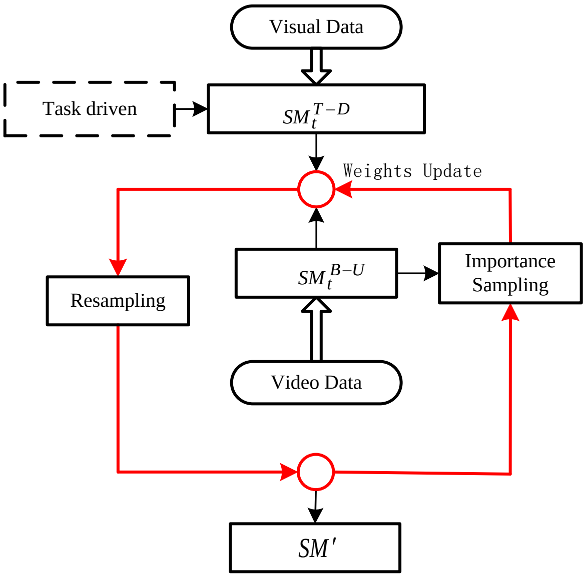


Fig. 2. Algorithm framework of saliency detection

3.1 Particle weight updating

According to Equation (11), the weight of the current state $\lambda_t^{(i)}$ is in direct proportion to the product of the weight of the previous time $\lambda_{t-1}^{(i)}$, $p(z_t^{B-U} \mid \hat{x}_t^{(i)}) p(z_t^{T-D} \mid \hat{x}_t^{(i)}, z_t^{B-U})$. $p(z_t^{B-U} \mid \hat{x}_t^{(i)})$ is the probability value of B-U attention in the current state, and $p(z_t^{T-D} \mid \hat{x}_t^{(i)}, z_t^{B-U})$ is the posterior probability value of T-D attention under B-U attention observational conditions. Those values are defined respectively as follows:

$$P(z_t^{B-U} \mid \hat{x}_t^{(i)}) = \frac{1}{\sqrt{2\pi} \cdot \delta_{B-U}} \cdot \exp\left(-\frac{(1 - SM^{B-U})^2}{2 \cdot \delta_{B-U}^2} \right) \tag{12}$$

$$P(z_t^{T-D} \mid \hat{x}_t^{(i)}, z_t^{B-U}) = \frac{1}{\sqrt{2\pi} \cdot \delta_{T-D}} \cdot \exp\left(-\frac{\left(1 - SM^{T-D} \cdot e^{SM^{B-U}}\right)^2}{2 \cdot \delta_{T-D}^2} \right) \tag{13}$$

For the B-U attention SM^{B-U}, the attention calculation method in Itti's model was adopted. T-D attention SM^{T-D} was measured by the similarity degree of task-related features. In this paper, the local binary pattern feature improved by Ojala et al. [18,19] was adopted.

3.2 Importance sampling and resampling

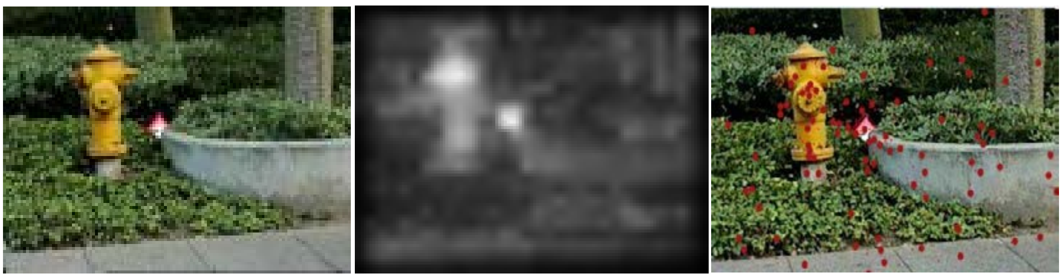
The B-U attention salience was used to regulate the density of Gaussian particle random sampling. Let $\hat{z}_1^{(i)} \sim U[0,1]$, $\hat{z}_2^{(i)} \sim U[0,1]$, and $i = 1,2,\ldots,N$ in an independent identical distribution, and make

$$\begin{cases} \hat{x}^{(i)} = \left(\sum_{i=1}^{N} z_1^{(i)} - N \cdot \mu_x \right) \Big/ \sqrt{N \cdot \delta_x^2} & i = 1,2 \ldots N \\ \hat{y}^{(i)} = \left(\sum_{i=1}^{N} z_2^{(i)} - N \cdot \mu_y \right) \Big/ \sqrt{N \cdot \delta_y^2} \end{cases} \tag{14}$$

where μ_x , μ_y , δ_x^2 , and δ_y^2 are the mean and variance of pseudorandom sequence $\hat{x}^{(i)}$ and $\hat{y}^{(i)}$ respectively. Through Equation (14), results of random Gaussian sampling can be produced in the area formed by the coordinate center (μ_x, μ_y). $SM_t^{\text{B-U}}(x, y)$ was assumed to be the saliency value of the saliency map on coordinate (x, y) at time t . The sampling density function was defined as follows:

$$\underset{(x,y)\in\Lambda}{\Gamma}\left(SM^{B-U}\right) = \frac{1}{\sqrt{2\pi}\delta_A} \cdot \exp\left(-\frac{1}{2}\left(\frac{SM^{B-U}(x_0, y_0) - u_A}{\delta_A} \right)^2 \right) \tag{15}$$

where i and j represent the abscissa and ordinate in the saliency map respectively; the mean was $u_A = \frac{\sum_{\{i,j|(i,j)\in SM^t\}} SM(i,j)}{N}$, and the variance was $\delta_A^2 = \frac{\sum_{\{i,j|(i,j)\in SM^t\}} [SM(i,j) - u]^2}{N}$. The result of particle sampling is shown in **Fig. 3**.



(a) (b) (c)

Fig. 3. Graphical overview of particle sampling.
(a) Original map. (b) B-U attention saliency map. (c) sampling results map.

After the recalculation of the particle weight, low-weight particles were weeded out by resampling to make them gather around the high-weight ones.

3.3 Attention saliency map

After resampling, the attention size was determined by the density of the particle distribution. The attention salience of the dense-particle-distribution area was remarkably higher and that of sparse distribution areas was remarkably lower, as shown in **Fig. 4**.

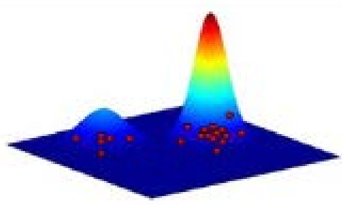


Fig. 4. Graphical overview of particle distribution and attention size

According to the distribution state of particles, in 2-dimensional space the size of the attention saliency can be defined as follows:

$$SM(x,y)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sigma^2}\cdot\varphi\left(\frac{x-x_i}{\sigma}\right)\cdot\varphi\left(\frac{y-y_i}{\sigma}\right) \tag{16}$$

where (x,y) is the spatial position of particle distribution, n is the number of particles, and $\sigma=h_n/\sqrt{n}$ is the window width. The 2-dimensional Gaussian windows were adopted as the window function. Then Equation (16) was turned into (17):

$$SM(x,y)=\frac{1}{2\pi n\sigma^2}\sum_{i=1}^{n}\exp\left(-\frac{(x-x_i)^2+(y-y_i)^2}{2\sigma^2}\right) \tag{17}$$

Fig. 5 shows the process of attention saliency estimation based on the particle filter.

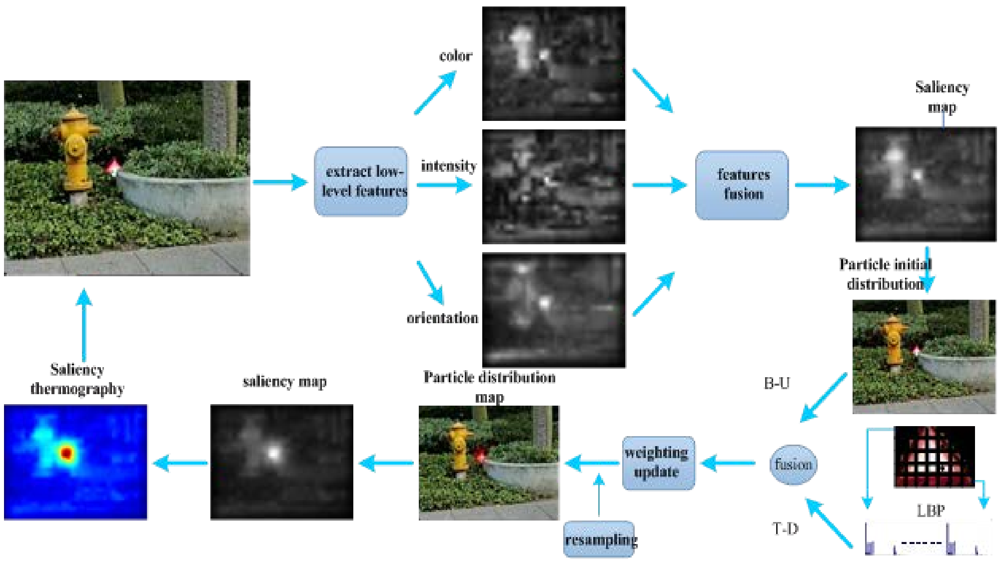


Fig. 5. Process of attention saliency estimation

4. Experimental results and discussions

In this section, the effect of the bidirectional visual attention model based on particle filter is tested and analyzed. In addition, visual attention saliency maps of our model and those of others (Itti's model[1], Fang's weighting fusion model[7], Zhang's bias control model[8], and Judd's pattern classification model[13–19]) are compared and analyzed. The saliency estimation accuracy is also evaluated. The experiment was completed on a Dell computer of 2.0 GHz and 1 GB RAM, and the programming environment was MATLAB 2012.

4.1. Environmental result analysis

Fig. 6 shows the experimental results. The testing image data were named "cola," "horse," "cup," "paper bag," "fire hydrant," "balloon," "oil drum," and "desk accessories." The first row in **Fig. 6** shows the original map; the second, the marked salience result (ground truth) of the map for comparing and calculating purposes; the third, the salience map of Itti's model; the forth, the result of Fang's WFM; the fifth, that of Zhang's BCM; the sixth, that of Judd's PCM; and the last, that of our model.

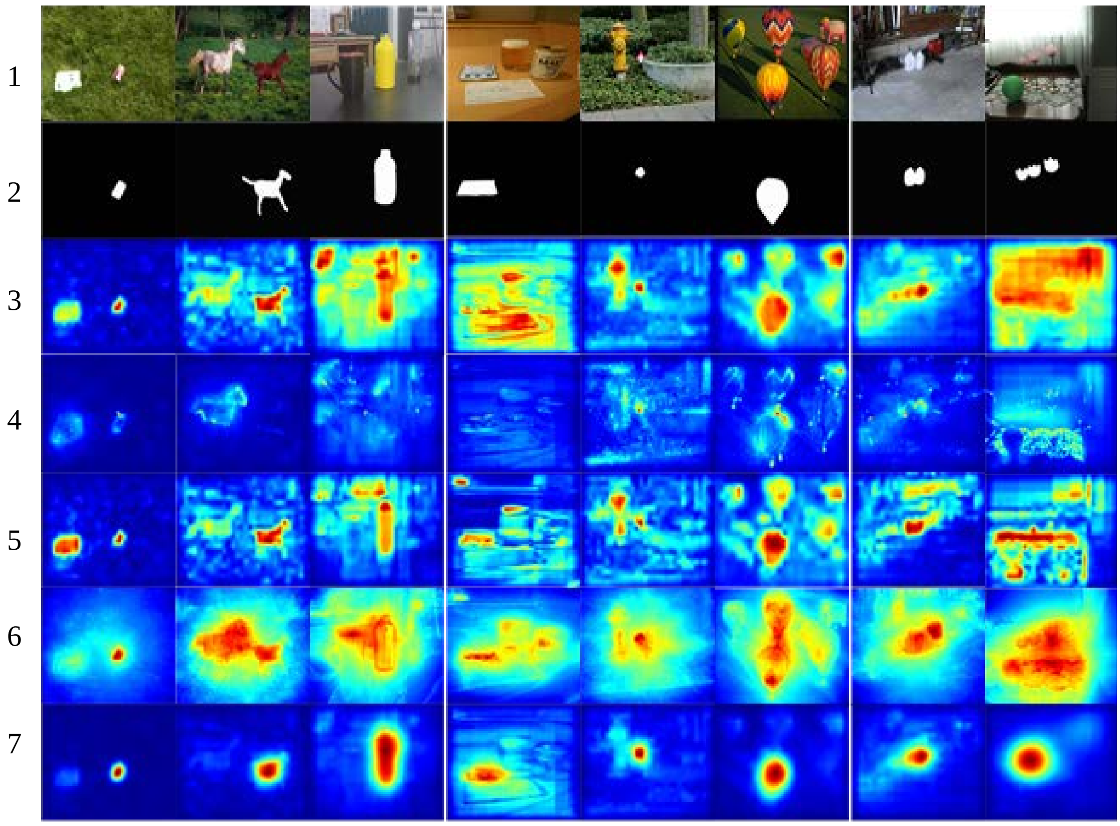


Fig. 6. Comparative experimental results

It can be seen from the experimental results that the attention saliency map of Itti's model (**Fig. 6**, row 3) was a simple presentation of areas with strong contrasts of visual stimulation. For the "cup," "paper bag," and "office supplies," the target area of attention was not clear due to the interference of light and background; In Fang's model (row 4), the fusion of the bidirectional attention was realized by linear weighting; the selection of the weighting coefficient had a large effect on the result, but obtaining an optimum result was difficult; for different images, the fixed weight coefficient had difficulty getting a consistent, substantial effect. For "cup," "paper bag," and "desk accessories," note that the clearing effect was not ideal. For "fire prevention," "balloon," and "oil drum" attention resulted in an only preliminary effect. "Cola" and "horse" results had the target attention region relatively close. Based on the feature similarity, Zhang's model (row 5) bias influenced the process of B-U feature fusion in a T-D manner, and thus it had a better effect than other methods. The method of Judd's model (row 6) is greatly influenced by the training samples, so it was far from accurate. In **Fig. 6**, it can be seen that compared to other methods the subjective effect of the estimation of the bidirectional fusion attention model proposed in this paper (row 7) was better; the saliency of task-related areas was significantly improved, the saliency of background areas was inhibited correspondingly, and thus a sharper saliency contrast was formed.

4.2 ROC performance index

A receiver operating characteristic (ROC) curve was applied in this paper to evaluate the estimation performance of attention saliency. Each pixel was set as a sample of the binary classifier. If the pixel value was greater than a certain threshold value, it would be the attention focus (positive sample); otherwise it would be a nonattention focus (negative sample). The true binary map of the image was the standard (ground truth); a series of corresponding values of the true positive rate (TPR) and false positive rate (FPR) were obtained by changing the threshold value. Then the ROC curve was drawn with the FPR and TPR as the lateral and longitudinal axes respectively. The TPR and FPR were calculated by Equations (18) and (19):

$$TPR = TP / nP \tag{18}$$

$$FPR = FP / nN \tag{19}$$

where nP and nN represent the number of positive and negative samples respectively and TP and FP are the number of positive samples judged true and false. **Fig.7** shows the ROC curves of the testing data. The closer the curve is to the top left corner, i.e., the TPR value is large and the FPR is small, the better the performance of the algorithm.

The green curve in **Fig. 7** is the result of Itti's model, which belongs to the B-U pattern. Its algorithm performance is better than those of others when the visual feature has a high contrast. Otherwise, its ROC effect is poor. The blue curve is the result of Fang's WFM, in which attention results of various images show large differences owing to the fixed fusion coefficient. The cyan curve is that of Zhang's bias control pattern, which produces good attention results and relatively stable quality by adopting the T-D bias pattern. The black curve is that of Judd’s PCM, in which statistical samples have an obvious influence on the attention result. The red curve shows the result from the proposed model. Clearly, with the most stable effect of the ROC curve among the test results, the proposed model is better.

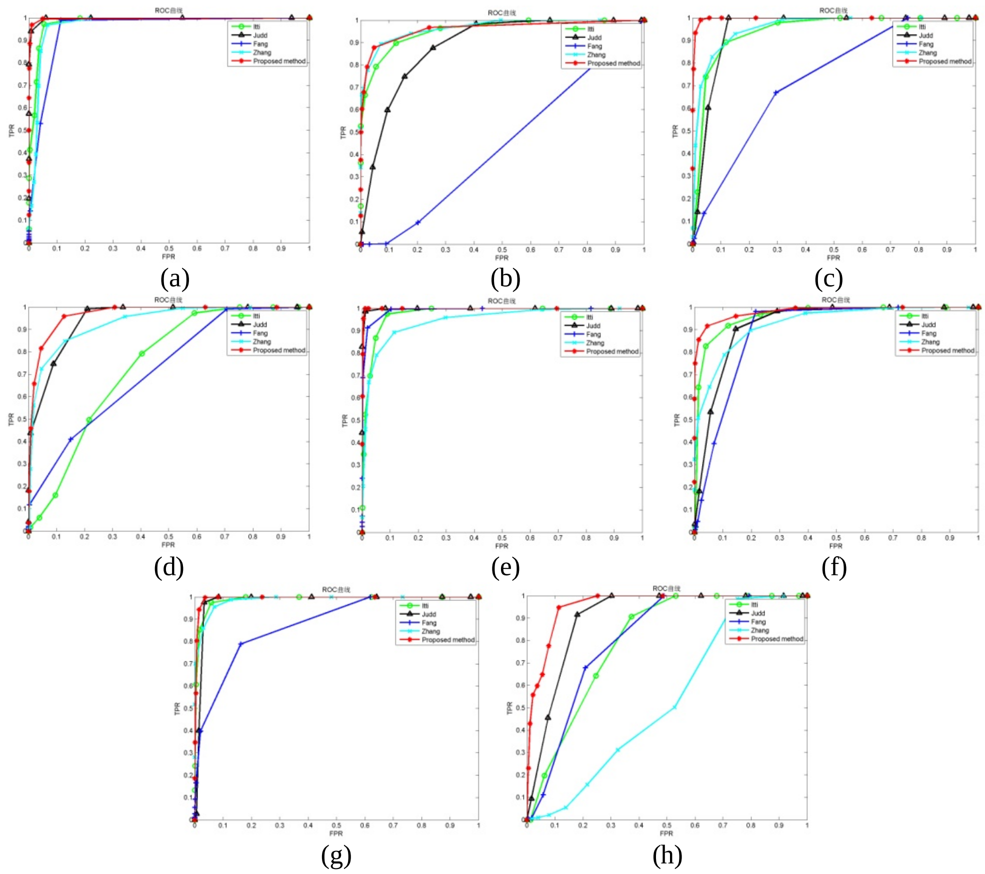


Fig. 7. ROC curve comparison. The closer the curve is to the top left corner, the better the performance of the algorithm. (a) “cola.” (b) “horse.” (c) “cup.” (d) “paper bag.” (f) “balloon.” (g) “oil drum.” (h) “desk accessories.”

5. Conclusion

On the basis of the Bayesian fusion theory, a bidirectional fusion attention model based on a particle filter is put forward, with Itti’s visual attention taken as B-U attention and task orientation attention taken as T-D attention. The attention salience is computed through the effective fusion of bidirectional attention within the particle filter’s framework and the particle distribution after filtering. The experimental results show that the model proposed in this paper, by which a relatively accurate attention salience map may be obtained, is of significant academic and practical value in related fields.

References

[1] Laurent Itti, Christof Koch and Ernst Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," *IEEE transactions on pattern analysis and machine intelligence*, Vol. 20, No. 11, November 1998. Article (CrossRef Link)

[2] Itti L, Koch C, "Computational Modeling of Visual Attention," *Nature Reviews Neuroscience*, 2 (3):193-203, 2001. Article (CrossRef Link).

[3] Navalpakkam V, Itti L, "Modeling the Influence of Task on Attention," *Vision Research*, 45(2):205-231, 2005. Article (CrossRef Link).

[4] Zechao Li,Jing Liu,Jinhui Tang,Hanqing Lu, "Robust Structured Subspace Learning for Data Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(10):2085-2098, 2015. Article (CrossRef Link).

[5] Zechao Li,Jing Liu,Yi Yang,Xiaofang Zhou,Hanqing Lu, "Clustering-Guided Sparse Structural Learning for Unsupervised Feature Selection," *IEEE Transactions on Knowledge and Data Engineering*,26(9):2138-2150, 2014. Article (CrossRef Link).

[6] Zhang-Qin Seak, Ang Li-Minn, Kah-Phooi Sen, "Face Segmentation Using Combined Bottom-up and Top-down Saliency Maps," in *Proc. of 2010 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT)*, Chengdu, 5:477-480, 2010. Article (CrossRef Link).

[7] Fang Y M, Lin W S, Lau C T, et al, "A Visual Attention Model Combining Top-Down and Bottom-Up Mechanisms for Salient Object Detection," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, Singapore, 1293-1296, 2011. Article (CrossRef Link).

[8] Zhang Jing, Zhuo Li, Gao Jingjing, et al, "A Study of Top-Down Visual Attention Model Based on Similarity Distance," *2nd International Congress on Image and Signal Processing*, CISP '09. Tianjin:1-5, 2009. Article (CrossRef Link).

[9] Benicasa A X, Marcos G Q, Zhao L, et al, "An Object-Based Visual Selection Model with Bottom-Up and Top-Down Modulations," in *Proc. of IEEE International Conference on Neural Networks* , Curitiba: Brazilian Symposium, 238-243, 2012. Article (CrossRef Link).

[10] Wei W, Yang X L, Zhou B, et al, "Combined energy minimization for image reconstruction from few views," *Mathematical Problems in Engineering*, 2012. Article (CrossRef Link).

[11] Yu Y L, George K I. Mann, et al, "An Object-Based Visual Attention Model for Robotic Applications," *IEEE Transaction on System*, 40(5):1398-1412, 2010. Article (CrossRef Link).

[12] Yu Y L, George K I. Mann, et al, "A Goal-Directed Visual Perception System Using Object-Based Top–Down Attention," *IEEE Transactions on Autonomous Mental Development*, 4(1):87-103, 2012. Article (CrossRef Link).

[13] Wei Wei,Qi Yong, "Information potential fields navigation in wireless Ad-Hoc sensor networks," *Sensors*, 11(5): 4794-4807, 2011. Article (CrossRef Link).

[14] Robert J. Peters and Laurent Itti, "Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention," in *Proc. of IEEE Conference on Computer and Pattern Recognition (CVPR'07)*, Minneapolis, MN1-7, 2007. Article (CrossRef Link).

[15] Tilke Judd,Krista Ehinger,Fredo Durand,et al, "Learning to Predict Where Humans Look," in *Proc. of IEEE International Conference on Computer Vision. Kyoto*, 106-2113, 2009. Article (CrossRef Link).

[16] Borji, "Boosting bottom-up and top-down visual features for saliency estimation," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 438-445, 2012. Article (CrossRef Link).

[17] Wei W, Xu Q, Wang L, et al, "GI/Geom/1 queue based on communication model for mesh networks," *International Journal of Communication Systems*, 27(11): 3013-3029, 2014. Article (CrossRef Link).

[18] Ojala T, Pietikäinen M & Mäenpää T, "A generalized Local Binary Pattern operator for multiresolution gray scale and rotation invariant texture classification," in *Proc. of Second International Conference on Advances in Pattern Recognition*, Rio de Janeiro, Brazil, 397-406, 2001. Article (CrossRef Link).

[19] Timo Ojala, Matti Pietikäinen and Topi Mäenpää, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, July 2002. Article (CrossRef Link).

Liu Long is an Associate Professor in the Department of Automation and Information Engineering at the Xi'an University of Technology in Shannxi, China. He obtained his PhD degree in Information and Communication Engineering from Xi'an Jiaotong University in 2005. His areas of research include target detection, target tracking, video understanding, and data fusion.

Li Xianli received her B.E.degree in Electrical Engineering and Information from Sichuan University,Chengdu,Sichuan,P.R.China,in June 2012. She is currently pursing a master's degree at the School of Automation and Information Engineering in Xi'an University of Technology. Her current research interests include target tracking , video understanding and data fusion.

Pan Yafeng received his B.E.degree in Information and Communication Engineering from Dalian Ocean University,Dalian,Liaoning,P.R.China,in June 2012.He is currently pursing a master's degree at the Faculty of Automation and Information Engineering in Xi'an University of Technology. His current research interests include target detection, target tracking and video understanding.

Wei Wei received his Ph.D. and M.S. Degrees from Xi'an Jiaotong University in 2011 and 2005, respectively. Currently he is an assistant Professor at Xi'an University of Technology. His research interests include wireless networks and wireless sensor networks application, mobile computing, distributed computing, pervasive computing and image processing.

Houbing Song (M'12–SM'14) received the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, in August 2012. In August 2012, he joined the Department of Electrical and Computer Engineering, West Virginia University Institute of Technology (WVU Tech), Montgomery, WV, where he is currently an Assistant Professor and the Founding Director of both the Security and Optimization for Networked Globe Laboratory (SONG Lab, www.SONGLab.us), and West Virginia Center of Excellence for Cyber-Physical Systems sponsored by West Virginia Higher Education Policy Commission. In 2007 he was an Engineering Research Associate with the Texas A&M Transportation Institute. He is the editor of six books, including Smart Cities: Foundations and Principles, Hoboken, NJ: Wiley, 2016, Cyber-Physical Systems: Foundations, Principles and Applications, Waltham, MA: Elsevier, 2016, and Industrial Internet of Things: Cybermanufacturing Systems, Cham, Switzerland: Springer, 2016. He has published more than 100 academic papers in peer-reviewed international journals and conferences. His research interests lie in the areas of cyber-physical systems, internet of things, cloud computing, big data, connected vehicle, wireless communications and networking, and optical communications and networking. Dr. Song is a member of ACM. Dr. Song was the very first recipient of the Golden Bear Scholar Award, the highest faculty research award at WVU Tech. Dr. Song is an Associate Editor for several international journals, including IEEE

Access and KSII Transactions on Internet and Information Systems, and a Guest Editor of more than 10 special issues. Dr. Song was the general chair of several international workshops, including the first IEEE International Workshop on Security and Privacy for Internet of Things and Cyber-Physical Systems (IOT/CPS-Security) within IEEE ICC, held on June 12, 2015, in London, UK, and the first IEEE International Workshop on Big Data Analytics for Smart and Connected Health within IEEE CHASE, held on June 27-29, 2016, in Washington D.C. Dr. Song also served as the technical program committee chair of the fourth IEEE International Workshop on Cloud Computing Systems, Networks, and Applications (CCSNA), held on December 6, 2015, in San Diego, USA. He is serving as the publicity chair of 2017 IEEE Wireless Communications and Networking Conference (WCNC) to be held on March 19-22, 2017 in San Francisco, CA, and the technical program committee chair of the fifth IEEE International Workshop on CCSNA to be held on December 4-8, 2016, in Washington D.C. Dr. Song has served on the technical program committee for numerous international conferences, including ICC, GLOBECOM, INFOCOM, WCNC, and so on.