

## A novel method of Visual Attention for Targets Detection

Zhuo Zhang, Xinnan Fan\*, Xuewu Zhang, Haiyan Xu, Min Li  
 Jiangsu Key Laboratory of Power Transmission and Distribution Equipment Technology  
 Hohai University  
 Nanjing, China  
 e-mail: zhangz@hhuc.edu.cn

**Abstract**—Inspired by the research of human visual system in neuroanatomy and psychology, the paper proposes a two-way collaborative visual attention model for target detection. In this new method, bottom-up attention information cooperates with top-down attention information to detect a target rapidly and accurately. Firstly, the statistical prior knowledge of target and background is applied to optimize bottom-up attention information in different feature space and scale space. Secondly, after the SNR of salience difference between target and interference is computed, the bottom-up gain factor is obtained. Thirdly, the gain factor is applied to adjust bottom-up attention information extraction and then to maximize the salience contrast of target and background. Finally, target is detected by adjusted saliency. Experimental results show that the proposed model in this paper can improve the real-time capability and reliability of target detection.

**Keywords**—target detection; visual attention mechanism; two-way collaborative mechanism

### I. INTRODUCTION

Human visual system can quickly and accurately detect target in a large amount of input information by limit resources. Many studies show that visual attention mechanism plays an important role in that process. The working principle and computing strategy of visual attention mechanism are that secondary information for visual processing task is discarded and limit resources are concentrated to deal with more important information [1].

This selective attention mechanism makes human visual system has strong robustness and high efficiency in complex environment target detection. Therefore, research on the target detection based on visual attention mechanism has important significance.

At present, the visual attention mechanism is mainly made of data driven and task driven, known as bottom-up mechanism and top-down mechanism [2]. Bottom-up mechanism is that human would be always interested in salient things without an attentional goal. Top-down is that we would be always blind to things we do not wish to see if we have top-down attention alone. Top-down attentional mechanism is more according with task and Bottom-up attentional mechanism is often faster and more potent. Many theories or research framework have put emphasis on one of them [3]-[5]. And main research is concentrated on bottom-up model [6]-[7], meanwhile, there are some studies devoted to developing top-down model. In 1998, the University of California, Itti [2], [8] proposed a visual attention model that

based on the Primary Visual Feature, has become one of the most classic visual attention model and Stentiford in 2003 put forward using Visual Attention Map to represent the image salience [9]. In 1990, Rimey [10] proposed that using enhanced Hidden Markov Model to realize data-driven selective attention mechanism. Subsequently, Salah [11] applied the Observed Markov Model to simulate the task-driven attention mechanism, and have achieved good results in digital recognition and face recognition. Based on the research in Itti, in 2002, Navalpakkam constructed a task-oriented attention model [12]. Both Itti model and attention map are a popular way to imitate bottom-up attentional mechanism.

This paper combining with copper surface defects detection, explore and research a kind of accurate, fast interesting target detection methods that based on visual attention mechanism. Firstly, the model extracts attention information of bottom-up, taking both features of spatial domain and frequency domain and saliency detection of feature space and scale space into consideration. Then, considering the differences of detecting technique and environment complexity, it applies the statistical prior knowledge of target and background to obtain a bottom-up gain factor. Finally, it uses the gain factor to construct a kind of bottom-up and top-down attention information cooperative visual attention mechanism and attempt to solve complex and severe working environment defect detection problem.

### II. TWO-WAY COLLABORATIVE ATTENTION MODEL

The researchers in the target detection field has applied task-oriented attention mechanism to achieve many applications as face recognition, image retrieval and etc. However, those applications are mostly based on the bottom-up model and perform task-oriented attention by adjusting the salient areas scale, orientation, size or shape. Shortcoming of these task-oriented models is that they are lack of an effective definition of high-level information to implement real task driven target detection. Some researchers tried to combine bottom-up and top-down attention. For example, Oliver in 2006 proposed a real-time computer vision and machine learning system to simulation and identification human behavior. This system used Bayesian model and closed feedback loop to combine top-down and bottom-up attention information. For known target, this method can achieve a high accuracy of

detection ,but it still has difficulty in identifying the unknown target in real-time detection.

One of the key challenges in integrating data-driven and task-driven attention mechanism is to find the optimal top-down influence on bottom-up. This paper intends to construct a two-way collaborative visual attention model, as the figure 1 shows. In order to meet the requirements of real-time capability of online target detection, the model based on human visual perceptual system theory, is on existing foundation of the salient detection of targets in complex environment, such as natural environment, and it expands knowledge base structure to add function of focus shift module. Also, on the base of extraction information from the bottom up attention ,the model searches interested region from the original image. Then, under the guidance of target detection task ,it considers the statistical probability of the target and background knowledge to optimize the different the bottom up attention information at feature space and scale space from .The process is that through analysis of target and background, calculating the target and interference salience difference SNR to get the bottom-up gain factor and applying the gain factor to adjust the bottom up attention information extraction and to maximize the salience contrast of the target and background. Consequently, the target detection real-time capability and robustness is improved.

### III. TARGET DETECTION METHODS BASED ON COLLABORATIVE ATTENTION MECHANISM

The input information processing of the proposed model is as follows:

#### A. Bottom up attention information extraction

Bottom up attention information can be obtained by calculating various salient features in an image. As the feature description of surface defect is varied, such as grey statistics characteristics, texture features, two-dimensional geometry features, color characteristics, etc. Therefore, Choosing suitable features to describe saliency is the key of targets detection. According to the characteristics of the surface defect target on copper strip and meeting the actual needs of inspection process ,the paper selects the primary visual features, spectral texture features and wavelet statistical features to describe the saliency of targets defect.

The computation processes of these three kinds of features saliency are as follows:

##### (1) Calculation of primary visual feature saliency

Firstly, through the decomposition of Gaussian pyramid and Gabor pyramid at different scales, 9 intensities features, 36 colors features and 36 orientations features are obtained .

According to the 81 features, center-surround differences  $\ominus$  between a “center” fine scale  $c$  and a “surround” coarser scale  $s$  is computed then to get 42 feature maps including 6 for intensity, 12 for color, and 24 for orientation. Afterwards, local iteration strategy is applied to feature maps respectively to form feature attention maps.

##### (2) Calculation of texture feature saliency of frequency domain

After the value of image texture feature  $S(\theta)$  is computed, intensity of the image histogram is used to estimate the local probability density function  $p(a, sc, l)$  of  $S(\theta)$  ,where  $p(a, sc, l)$  means the probability of  $S(\theta)$  .Then, texture features saliency  $ST(sc, l)$  is calculated by the following formula :

$$ST(sc, l) = - \sum p(a, sc, l) \log_2 p(a, sc, l)$$

##### (3) Calculation of statistical saliency

After multi-scale statistical method is used to compute the statistical characteristics of the image, dissimilarity measure  $SM(sc, l) \frac{n!}{r!(n-r)!} \{ \}$  is obtained by the following formula:

$$SM(sc, l) = \frac{sc^2}{sc-1} (T^2(sc, l) - T^2(sc-1, l))$$

Where the region  $l$  meets the condition  $T^2 > UCL$  .

#### B. Top-down attention information extraction

Different feature attention can stimulate saliency detection to accelerate the process of target detection. Therefore, according to the target or interference prior information  $\Theta$  , the model computes target and background statistics knowledge and then using different weights for top-down component to optimized bottom-up map. The target and background saliency ratio SNR is expressed as

$$SNR = \frac{E[S_T(A)]}{E[S_D(A)]} = \frac{\sum_{j=1}^N g_j \sum_{i=1}^n g_{ij} E_{T|V} [SS_{iT}(sc, l)]}{\sum_{j=1}^N g_j \sum_{i=1}^n g_{ij} E_{D|V} [SS_{iD}(sc, l)]}$$

Then the optimal choice of weights of bottom up attention information is as

$$g_{ij} = \frac{SNR_{ij}}{\frac{1}{n} \sum_{k=1}^n SNR_{kj}} \quad (4)$$

$$g_j = \frac{SNR_j}{\frac{1}{N} \sum_{k=1}^N SNR_k} \quad (5)$$

#### C. Calculation of Two-way collaborative Saliency

After the target and background probability distribution functions are applied to estimate the top-down gain factor  $g_{ij}$  and  $g_j$ , the bottom-up saliency value obtained in the step A is combined with the gain factor in the step B to compute the final saliency map just as

$$S_j(sc, l) = \sum_{i=1}^n g_{ij} S_{ij}(sc, l) \quad (6)$$

$$S(sc, l) = \sum_{j=1}^N g_j S_j(sc, l) \quad (7)$$

D. attention Focus shift

In order to obtain the entire information of the image, the focus of attention is circularly shifted in the process of salient regions detection. Selection and transfer of attention focus determines the position and its important degree of an interested region. Saliency information of the regions is obtained through calculation of bottom-up and top-down attention information. Furthermore, it is necessary for the model to sort the regions on saliency value and to obtain salient region of the original image. As a result, the current salient region is that saliency value is maximum among all alternate regions and the remaining areas are known as potential salient regions. Meanwhile, the competitive mechanism of winner-takes-all is used for the salient region mutual inhibition in target interest map and the strongest attractive region will be the next salient area.

The model is also inspired by two characteristics of the human vision system known as

1) Adjacent priority principle. Human vision tends to choose the region which is adjacent to current salient region as the next saliency region.

2) Inhibition return principle. Human vision inhibits the salient regions that has been selected.

Therefore, it is necessary for attention focus selection and transfer to consider the influences of saliency value, adjacent distance and inhibition of return. Consequently, the model defines inhibition of return factor  $IR_k$  and adjacent priority factor  $PE_k$  to control attention focus shift.

Inhibition of return factor  $IR_k$  is defined as follows

$$IR_k = \begin{cases} 1 & k \text{ has not been selected} \\ 0 & k \text{ has been selected} \end{cases} \quad (8)$$

Adjacent priority control factor  $PE_k$  is defined as follows

$$PE_k = DIS_k / DIS_{sum} \quad (9)$$

$DIS_k$  is the distance of k-th potential saliency region to the current region,  $DIS_{sum}$  represents the distance of all the potential saliency regions to the current saliency region.

After that, the attraction of salient region is expressed as

$$AR_k = IR_k \cdot PE_k \cdot S(sc, l) \quad (10)$$

The winner-take-all neural network is used to find attention focus from the saliency map, to select the candidate region and to obtain salient region. Then, the model determines the potential transfer targets according to the current salient regions. And, searching for the next salient point and shifting attention focus on adjacent priority and inhibition return mechanism.

Repeating all of above, it can obtain a series of salient regions.

IV. EXPERIMENT AND RESULT ANALYSIS

To verify the effectiveness of the proposed model in this research, the paper uses images of copper strip with the size of 256×256 from the company of XingRong, in Changzhou,

China. The sample set has 1200 images including 1050 defect images and 150 nondestructive images and the image library contains lower resolution defect target images and weak defect images.

The number of attention focus transfer directly affects the efficiency of interested regions detection. Therefore, 1200 sample copper images is used to compute statistic number of the target attention saccades in the Markov model and the enhanced attention model.

Performance comparison of the observable Markov model and the improved attention model is shown as Table 1. From the table, it is concluded that the classification accuracy of the improved attention model and observable Markov model is approximately but the Markov model requires 5 saccades to complete target region detection and the improved attention model only need 1.7 times average saccades, which greatly improve the real-time performance. Also it is proved the proposed method effectiveness in the high speed on-line detection.

TABLE I. PERFORMANCE COMPARISON

method	The average number of saccades	Positioning accuracy (%)	
		train	test
Observable Markov model	5	97.45	94.40
The proposed model	1.7	97.56	94.48

V. CONCLUSION

A new kind of bottom-up and top-down cooperative visual attention model is proposed for targets detection in this paper. The model is established based on the obtained prior knowledge of target and background. Firstly, the statistical prior knowledge of target and background is applied to optimize bottom-up attention information in different feature space and scale space. Secondly, after the SNR of salience difference between target and interference is computed, the bottom-up gain factor is obtained. Thirdly, the gain factor is applied to adjust bottom up attention information extraction and then to maximize the salience contrast of target and background. Finally, target is detected by adjusted saliency. The experimental results show that the method proposed in this paper have effective guidance for targets detection, and relative to the Markov model, the results of detection accuracy and real-time capability have been greatly improved. For future research, it is necessary to optimize the algorithm for improving accuracy of targets detection.

ACKNOWLEDGMENT

The paper is supported both by “National Natural Science Foundation of China”(Grant No. 61273170) and “Research Fund for the Doctoral Program of Higher Education of China”(Grant No. 20120094120023), it also is

sponsored both by “the Fundamental Research Funds for the Central Universities”(Grant No.BZX/10B101-09) and “Chang Zhou Science and technology supporting program” (Grant No. CE20120101)

REFERENCES

[1] Simons D.J. Chabris C.F. “Gorillas in our midst: sustained inattention blindness for dynamic events” Pion Ltd.Perception.London vol.28, pp.1059-1074, April 1999  
 [2] Itti L., Koch C.” A saliency-based search mechanism for overt and covert shifts of visual attention” Vision research vol.40, pp.1489-506, 2000  
 [3] S.W.Luo, Information Processing Theory of Visual Perception, Electronics Industry Publishers, China,2006  
 [4] DeAngelis, GC, Ohzawa I and Freeman RD”Receptive-field dynamics in the central visual pathways” Trends in Neurosciences vol.18, pp.451-458, 1995  
 [5] Duncan J., Humphreys G.W” Visual search and stimulus similarity”Psychological VOL.96 pp.433-458, 1989  
 [6] O.Le Meur, P.Le Callet, D.Barba” coherent computational approach to model bottom-up visual attention” J .IEEE T PATTERN ANA.A VOL.28 pp.802-817,2006

[7] H.H Bülthoff, S.W Lee, T.Poggio, C.Wallraven, Editors. “Biologically inspired saliency map model for bottom-up visual attention”.Proceedings of the Second International Workshop on Biologically Motivated Computer Vision pp.22-24,2002  
 [8]Itti, L.,C. Koch, and E. Niebur.” A model of saliency-based visual attention for rapid scene analysis” IEEE Transactions on Pattern Analysis and Machine Intelligence,V OL.20,pp.1254-1259,1998  
 [9] F.W.M.Stentiford” An attention based similarity measure with application to content based information retrieval”. Proceedings of the Storage and Retrieval for Media Databases conference, SPIE Electronic Imaging,VOL.5021,pp.221-232.2003  
 [10] R.D.Rimey, C.M.Brown,”Selective attention as sequential behavior: Modeling eye movements with an augmented hidden markov mode”,Department of Computer Science, University of Rochester, USA,1990  
 [11] A.Salah, E.Alpaydin, L. Akarun.” A selective attention-based method for visual pattern recognition with application to handwritten digit recognition and face recognition”J. IEEE T PATTERN VOL.24,pp.420-425,2002  
 [12] Navalpakkam, V. and Itti, L.” A Goal Oriented Attention Guidance Model”, Biologically Motivated Computer Vision, Springer,pp. 81—118,2002

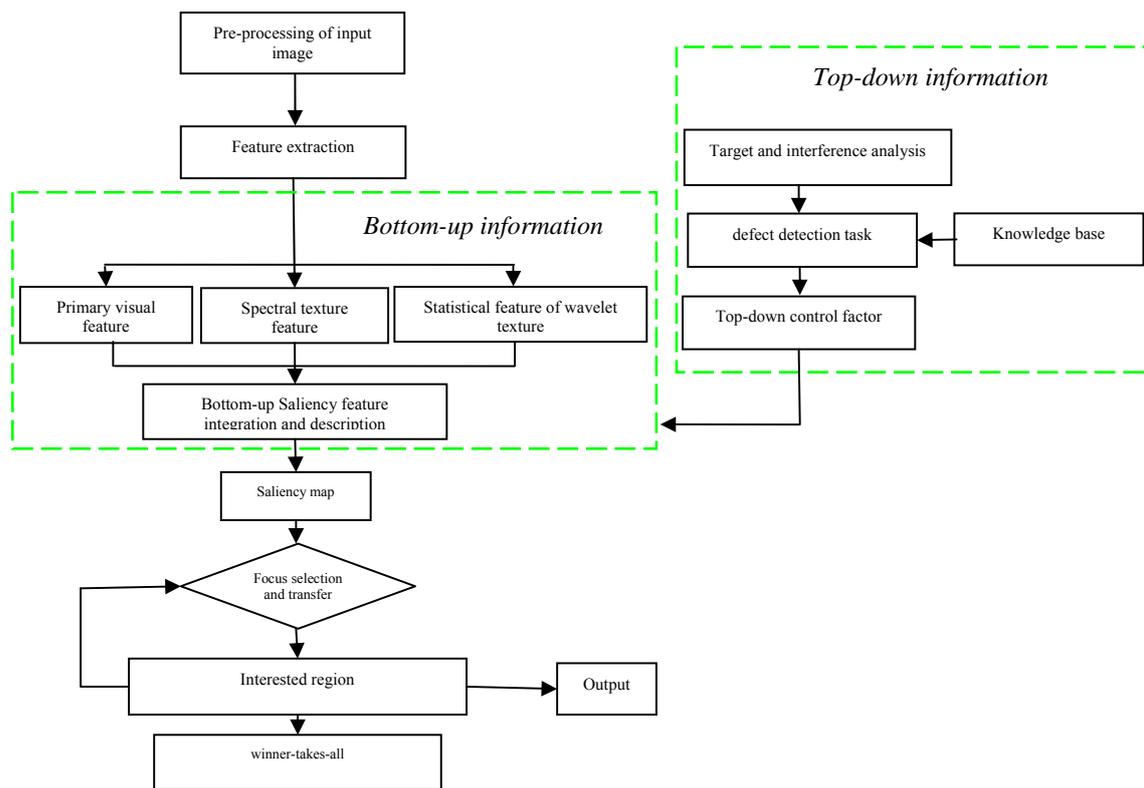


Figure 1 Two-way collaborative visual attention model for target detection