applied optics

Selective visual attention based clutter metric with human visual system adaptability

Bo Zheng,[1,2] Xiao-Dong Wang,[1,*] Jing-Tao Huang,[1] Jian Wang,[1] and Yang Jiang[1]

[1]Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
*Corresponding author: 2629164931@qq.com

Received 26 May 2016; revised 11 August 2016; accepted 21 August 2016; posted 22 August 2016 (Doc. ID 267063); published 20 September 2016

Most existing clutter metrics are proposed based on fixed structural features and experienced weight measures. In this paper, we identify the clutter as selective visual attention effects and propose a type of clutter metric. First, adaptive structural features are extracted from the blocks with an edge-structure similarity to the target. Next, the confusing blocks are selected by the similarity threshold based on the attention guidance map. The clutter is estimated by quantifying the effects of confusing blocks on target acquisition performance. The comparative field experiments, with a Search_2 dataset, show that the proposed metric is consistent with the adaptability of the human visual system (HVS) and outperforms other metrics.

OCIS codes: (110.2960) Image analysis; (110.2970) Image detection systems; (110.3000) Image quality assessment; (110.3925) Metrics.

http://dx.doi.org/10.1364/AO.55.007700

1. INTRODUCTION

Currently, models of target acquisition performance in viewing images affected by background, such as ACQUIRE-LC or Detect05 [1–3], have the experienced constant of N50 statistics to determine the spectral natures of background and the effects on the human detection process. However, this methodology tends to have less than the desired accuracy. Background clutter, that is, the similarity between the target and its background, is the typical feature confusing observers and affecting target acquisition performance [4]. However, in the present models, the background clutter is not taken into consideration. Therefore, it is fundamentally important to quantify the clutter and explore the relationship between its effect and target acquisition performance.

Several metrics of background clutter have been proposed. The statistical variance (SV) metric [5] and its derivations [6,7] are a first class clutter metric. These metrics only rely on the average variance of the whole scene and are too simple to assess the clutter numerically. The most commonly used human visual system (HVS) based metric is the second class clutter metric that outperforms the first class clutter. The probability of edge (POE) metric [8], based on the high sensitivity of HVS to edges, is inferior to the distribution of edge (DOE) metric [9], which emphasizes the importance of structure and weakens the influence of scene illumination on target detection. The gradient features based edge strength similarity *metric* (ESSIM) [10] measures the similarity of the background on the regional gradient distribution and quantifies the clutter by the image structure similarity metric. The target structure similarity measure (TSSIM) metric [11] quantifies the similarity of a target to its background in terms of luminance, contrast, and structure, and its performance is seriously affected by the selection of constants. POE, DOE, ESSIM, and TSSIM quantify the clutter in terms of fixed structural features and experienced weight measures, while they have limited success in predicting target acquisition performance in a complex background. The contrast sensitivity function based (CSFCMS) metric [12] weights the visual differences between the target and its background in the spatial frequency domain with the Mannos–Sakrison contrast sensitivity function. The hidden Markov model (HMM)-based (HMMC) metric [13] greatly describes the target and quantifies the clutter by optimizing the HMM parameters of the target sequence. The brain cognitive model based (BSD) metric [14] obtains the similarity map by using the structural similarity index [15] and weights the similarity of the blocks in the map according to the brain cognitive characteristics. CSFCMS, BSD, and HMMC introduce the cognitive characteristics [16] of HVS and quantify the clutter by weighting the similarity of the target to its background in accordance to the fixed weight measure.

In detecting the desired target in a complex scene, it is shown that HVS is highly adaptive for extracting structural features and is automatic in selecting the blocks relevant to the target [17,18]. None of the conventional clutter metrics conform to the characteristics of HVS mentioned above. In this paper, we introduce the widely accepted selective visual attention (SVA) [19,20] mechanisms and quantify the clutter by

simulating the attention characteristics of HVS. Based on the SVA research, we suppose that the features are extracted from the higher salient background blocks rather than from the entire scene by the top-down selective attention of HVS. In addition, the similar background blocks confuse the observers and affect the target acquisition performance only if the similarity is beyond a certain threshold. Under these two assumptions, we further propose two thresholds—saliency threshold (SAT) and similarity threshold (SIT)—for the selection of salient blocks in feature extraction and similar blocks in target detection, respectively. Next, we propose an adaptive methodology to determine the thresholds by the root-mean-square-error (RMSE) surface, which takes the adaptability of HVS into account. Introducing the proposed clutter metric into the target detection probability prediction model and then applying the model to the Search_2 dataset [21], we achieve a prediction accuracy of 0.0507 in RMSE.

2. SELECTIVE VISUAL ATTENTION BASED CLUTTER METRIC

Similar to acquiring the information from a scene, observers will extract features to reduce the amount of incoming visual data for higher-lever cognitive processing. In the target acquisition performance experiment, the observers detect the target in a complex background, which is identified as a task-dependent process with top-down SVA. The important features related to the target will attract the observers' attention more and be extracted in the experiment [22–24]. Based on this fact, we propose a new feature extraction model. First, the intensity map and orientation map of the image edge points are combined into edge-structure information. Second, the edge-structure information of each background block is calculated, and the similarity to the target is quantified. Finally, the edge-structure similarity map is generated. This map indicates the saliency of each background block for top-down selective attention and guides the defined SAT to select the background blocks. The features are extracted from the target and the selected blocks are extracted by principal component analysis [25,26].

The sparse representation based clutter metric (SRC) proposed by Yang *et al.* quantifies the background clutter by finding the sparsest representation of the target against the background in the feature domain [27]. First of all, the background is divided into blocks twice the size of the target. Next, the similarity between the background blocks and the target is described by sparse representation [28], which goes well with the nature of capturing the sparse functions of the input signal of HVS. Finally, the clutter is quantified by summing up the absolute value of the nonzero elements of the sparse representation.

Our work originates from this research. Most of the coefficients of the sparsest target representation are zero except those corresponding to the similar background blocks, and the bigger the coefficient is, the higher the similarity of the block [29]. The sparsest vector is the attention guidance map that directs attention to shift among the background blocks while detecting the target. Previous metrics take all the similar blocks as the confusing blocks that will attract attraction and affect target acquisition performance, therefore quantifying the clutter by summing up and weighting all the similar blocks. However, they are not consistent with the selective characteristic of the HVS. In this paper, we suppose that only the blocks with similarity beyond a particular level will obscure the observers' view and affect the target acquisition performance. Thus, in our clutter metric, the confusing blocks are selected by the defined SIT, and the clutter is quantified by summing up their effects on the target acquisition performance.

The flow chart of the process is presented in Fig. 1, and the main steps are as follows:

1. All of the target and background images are transferred from gray space to illumination space by the following formula:

$$Y = k\left(\frac{g - g_t}{c + g - g_t}\right)^r, \tag{1}$$

where g and Y are the gray-level value and display-luminance value of the image, respectively; k is a proportion constant; g_t is the gray-level value corresponding to the dark current; c is the half saturation level; and r is the display gamma coefficient. The background image is divided into N overlapping blocks with the same size as the target.

2. By filtering the target image with vertical and horizontal Sobel operators, we get the edge intensity and orientation maps. We obtain the total number (TN) of edge points with intensity above half of the averaged intensity in the intensity map and the

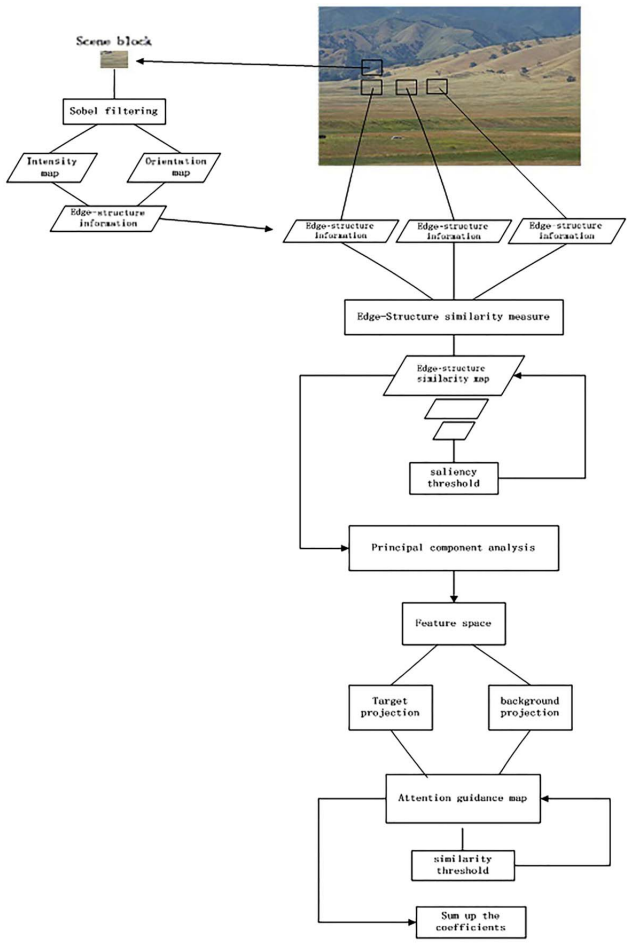

Fig. 1. Flow chart of the model.

orientation vector (TO) composed of the orientation values at the corresponding positions in the orientation map. Next, we mark the number of orientation values in TO lying in six equidistant orientation spaces from 0° to 180° as TON_k. The target edge-structure information of D_t is defined as

$$D_t = (\mathrm{TON}_k/\mathrm{TN}), \qquad k = 1, 2, 3\ldots, 6. \tag{2}$$

With the same procedure for the background blocks, the background edge-structure information of D_r is obtained where r denotes the rth background block.

3. The edge-structure similarity measure [9] of C_r of the rth background block is given by

$$C_r = \|D_r - D_t\|_2, \tag{3}$$

where $\|\cdot\|_2$ is the L_2 norm and C_r represents the similarity of the rth background block. Depending on the definition above, we can draw the inference that the smaller the C is, the higher the similarity of the corresponding block. C_rs of all blocks compose the edge-structure similarity map of the entire background.

4. The target image and the background blocks with C_r values below the defined CT in the edge-structure similarity map are arranged as columns to compose the data space of $X \subset R^{M\times N}$ for the principle component analysis.

5. The observation matrix of $X \subset R^{M\times N}$ is as the feature domain, and the principal component of $y \subset R^{D\times N}$ $(D \ll M)$ is extracted from $X \subset R^{M\times N}$ with the feature domain. The target projection of $t \subset R^{D\times 1}$ and the background projection of $b \subset R^{D\times N}$ to the feature domain are preserved in $y \subset R^{D\times N}$.

6. The sparsest representation of $t \subset R^{D\times 1}$s is calculated by

$$\arg\min \|s\|_1 \quad \text{subject to } t = bs. \tag{4}$$

Because of the low rank of the background and $D \ll N$, s can be solved by the minimum l_1 norm. The solved s is the attention guidance map for observers.

7. The SVA metric is defined as the sum of the absolute values of the elements in the sparse vector of s that are beyond the defined SIT:

$$\mathrm{SVA} = \sum_{j=1}^{I} |s_j| \qquad (s_j > \mathrm{SIT}), \tag{5}$$

where I is the number of the elements.

3. THRESHOLD DETERMINATION

The edge-structure similarity map indicates the attraction of each block, and the blocks for feature extraction are selected by the SAT. The attention guidance map directs the attention to shift among the blocks, and the blocks affecting target acquisition performance are selected by the SIT. We propose an adaptive method to determine the thresholds by pursuing the lowest point in the RMSE surface.

The target detection probability prediction model [30] is as follows:

$$\mathrm{PD}_{\mathrm{pred}} = \frac{(X/X_{50})^E}{1 + (X/X_{50})^E}, \tag{6}$$

where $\mathrm{PD}_{\mathrm{pred}}$ and X are the detection probability and clutter metric value, respectively. X_{50} and E are constants assigned by least square fitting of Eq. (6).

The similarity threshold is defined as dynamic values from 0 to maximum in the edge-structure similarity maps, and the saliency threshold is defined as the dynamic values from 0 to the maximum in the attention guidance maps. Introducing a pair of the thresholds into the SVA metric, we obtain the corresponding clutter metric value and the detection probability prediction. Then the RMSE between the prediction and the subjective detection probability is acquired. We create a diagram of a two-dimensional RMSE surface where x and y coordinates represent saliency and SIT, respectively, and the z coordinate represents the RMSE value.

Figures 2 and 3 shows the RMSE surface and its contour plot, respectively. The thresholds corresponding to the lowest RMSE value conform to the adaptability of HVS. The lower ridge is along the SAT coordinate and at the SIT coordinate of 0.40, which indicates that HVS cannot distinguish the target from the background. The lowest point on the ridge is at the SAT coordinate of 0.23, indicating that HVS is sensitive to the blocks. According to the above discussion, the SAT and the SIT are determined as 0.23 and 0.40, respectively.

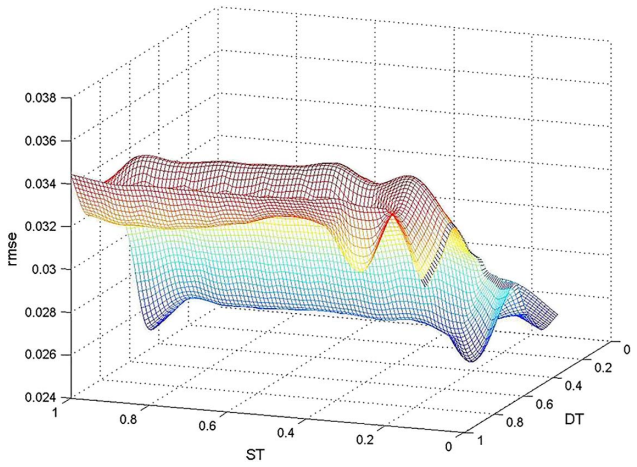

Fig. 2. Two-dimensional root-mean-square-error surface.

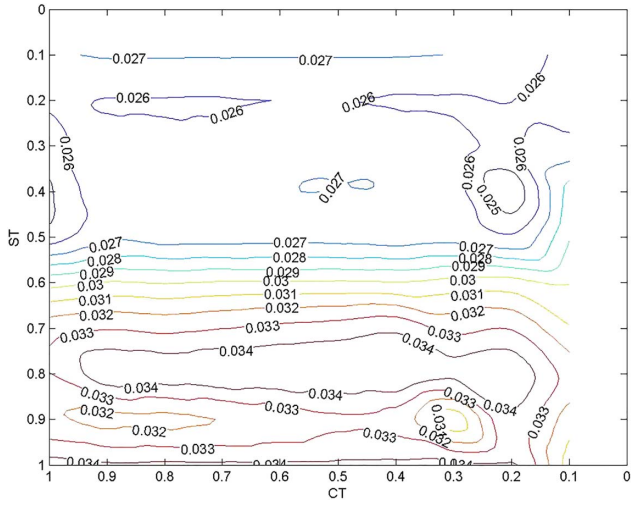

Fig. 3. Contour plot of the root-mean-square surface.

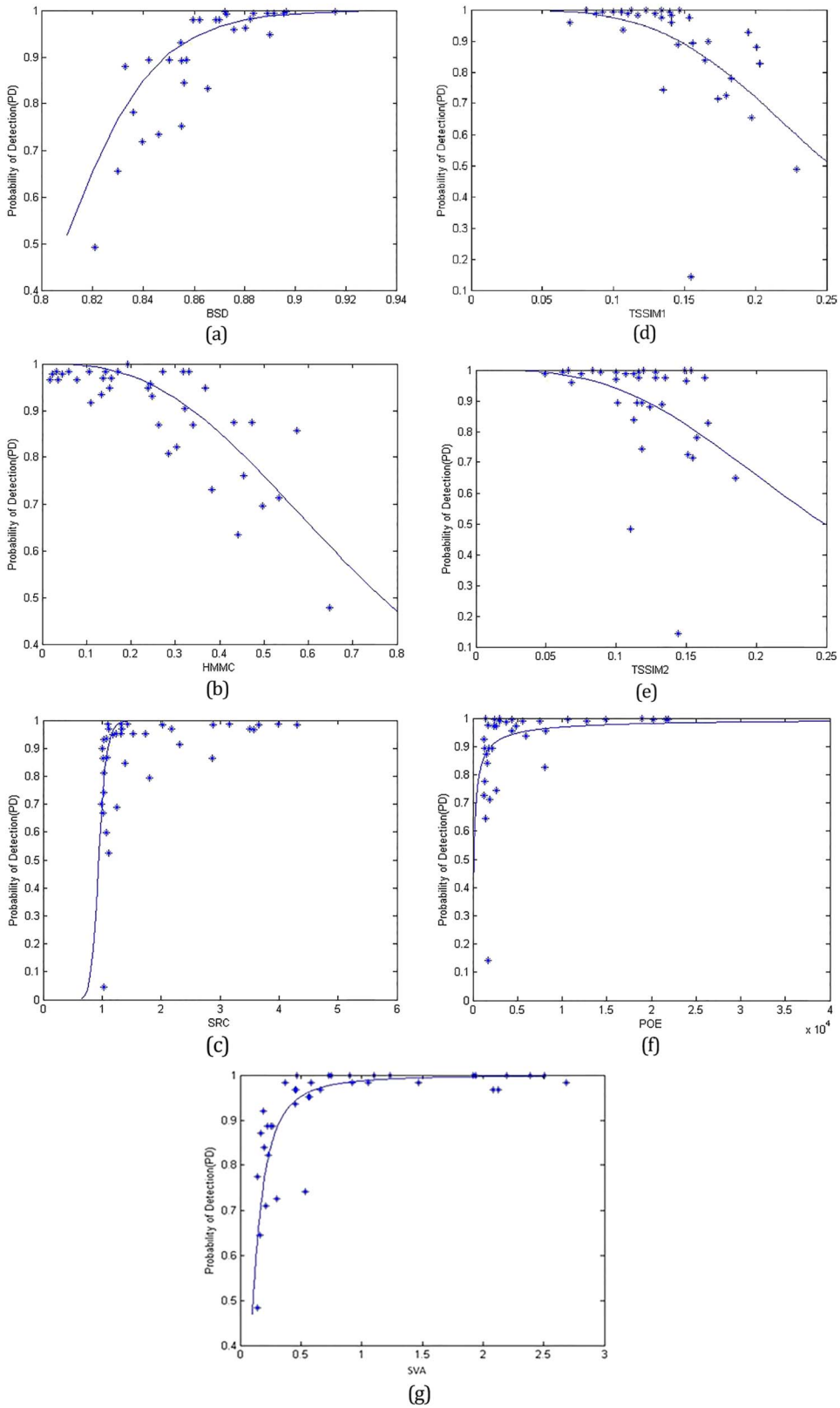

Fig. 4. Scatter plots of experimental target detection probability versus clutter metric values. Each sample point represents one target image in the Search_2 dataset: (a) BSD, (b) HMMC, (c) SRC, (d) TSSIM1, (e) TSSIM2, (f) POE, (g) SVA.

4. EXPERIMENTAL RESULT AND ANALYSIS

Similar to [13,14], the Search_2 database is used to validate the clutter metric. Thirty-nine original images with only one search target are used to assess the clutter metrics of BSD, HMMC, TSSIM1, TSSIM2, SRC, and POE; Eq. (6) is used to predict the detection probability by clutter metrics. For TSSIM1, $C_1 = (0.01L)^2$ and $C_2 = (0.03L)^2$, where L is the dynamic range of the pixel values (255 for 8-bit grayscale images) [31]. For TSSIM2, both C_1 and C_2 were selected to be a very small constant of 2×10^{-16} (eps in Matlab). We define the threshold in POE to be 0.7 for experience.

The scatter plots in Figs. 4(a)–4(g) display the relationship between the subjective experiment results and clutter metric values, with the solid curves being the regression curves of the detection probability prediction. To assess the extent of the agreement between the prediction and the subjective data, the adopted measures contain RMSE, Pearson linear correlation coefficient (PLCC), and Spearman's rank correlation coefficient (SRCC). The detailed results as well as the curve fitting parameters are given in Table 1.

From Figs. 4(a)–4(g), we can find that SVA, BSD, and HMMC congregate around the regression's solid lines more closely than do SRC, TSSIM1, TSSIM2, and POE. This can be also found in Table 1, where both correlation coefficients and RMSE of the three metrics correlate better with the subjective experiment data. These metrics illustrate their superior performance. The test results of SVA show that SVA outperforms BSD and HMMC in RMSE value, while it does not perform the best in other two correlation coefficients.

SVA introduces SVA and takes the HVS adaptability into consideration in the threshold determination. TSSIM1 and TSSIM2 quantify the clutter in terms of luminance, contrast, and structure, and POE is based on the structure of edge and experienced threshold. Their lower performance indicates that clutter metrics with fixed structure and thresholds have limited success in predicting target acquisition performance. Both SVA and SRC introduce sparse representation of the target against the background. SRC finds the sparsest representation in a no-target-defined feature domain and sums up all similarity vector elements as the clutter. Meanwhile, SVA acquires the sparsest representation in the target-defined feature domain and only sums up the vector elements beyond the SIT. BSD quantifies the clutter by weighting the structural similarity according to the information content weight measure, and HMMC automatically searches the most desired information of the target by optimizing the HMM parameters. Both of them introduce the adaptability of HVS and so possess advantages over other metrics except for SVA. The test results show that the two assumptions conform to the perception characteristics of HVS, and the adaptive threshold determination method is effective.

SVA with the lowest RMSE can be thought superior to others in precision, while it does not perform the best in the correlation coefficients of PLCC and SRCC. The reason is explained by analyzing the vertex positions of PLCC and SRCC surfaces. By the adaptive threshold determination method, we can also get the PLCC and SRCC surfaces. As is shown in Figs. 5 and 6, the vertices of the PLCC and SRCC surfaces correspond to the highest correlation coefficients, which are, respectively, 0.9280 and 0.9306, much higher than that of BSD and HMMC. Table 2 lists the vertex positions of RMSE, PLCC, and SRCC surfaces. We find that they are not coincident, which can be explained by the effect of target local contrast on the standard deviations of SAT and SIT of vertices. First of all, images of the Search_2 dataset are divided into four groups according to the target local contrast. Then the RMSE, PLCC, and SRCC vertices of each group are obtained by the proposed threshold determination method. Finally, diagrams of the standard deviation of SIT and SAT are given in Figs. 7 and 8, respectively, where the x coordinate represents the target local contrast and the y coordinate represents the standard deviation. It can be seen that the deviations gradually decrease as target local contrast increases, implying that the vertex positions get closer. Thus, we can conclude that the contrast affects the vertex positions and separates them. It is our future work to take the target local contrast into the model.

Table 1. Performance Comparison among Different Metrics

METRICS	X_{50}	E	RMSE	PLCC	SRCC
BSD	0.8087	45.8438	0.0557	0.879	0.805
HMMC	0.7649	−2.7066	0.0665	0.8223	0.7145
SRC	0.9429	14.4091	0.1283	0.5796	0.7003
TSSIM1	0.2529	−3.9987	0.1357	0.5799	0.7100
TSSIM2	0.24848	−3.0253	0.1541	0.3771	0.3675
POE	121.3247	0.7772	0.1387	0.4446	0.6398
SVA	0.09152	1.766	0.0507	0.7951	0.7125

Table 2. Vertices of Different Surfaces

TEST	SIT	SAT	VALUE
RMSE	0.40	0.23	0.0507
PLCC	0.44	0.24	0.9280
SRCC	0.82	0.33	0.9306

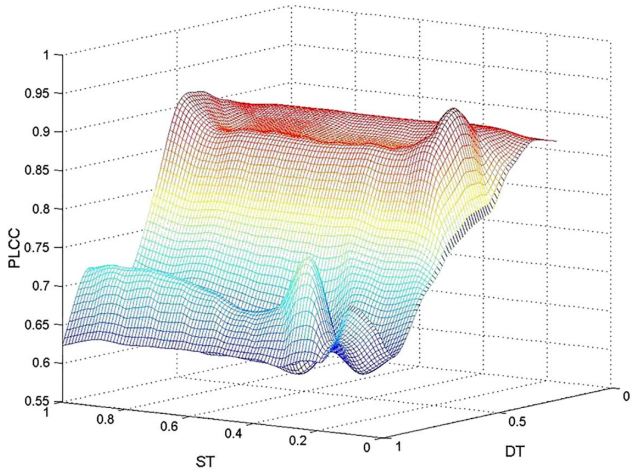

Fig. 5. Two-dimensional PLCC surface.

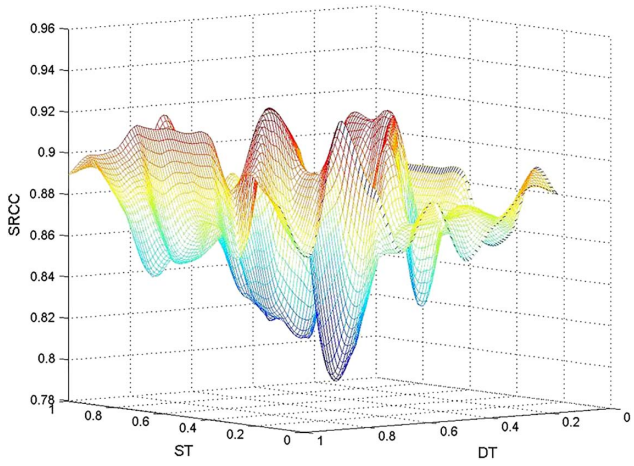

Fig. 6. Two-dimensional SRCC surface.

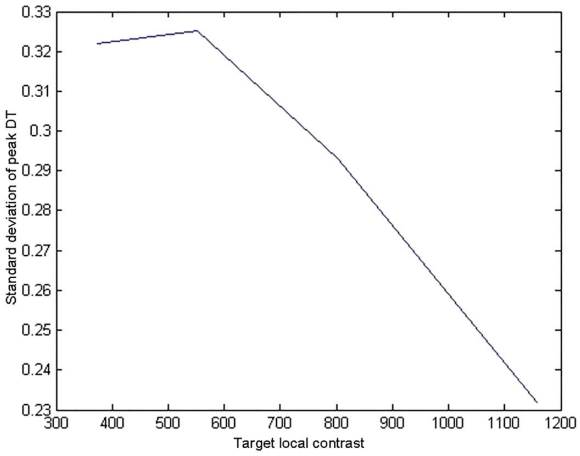

Fig. 7. Relationship between standard deviation of vertex SIT and target local contrast.

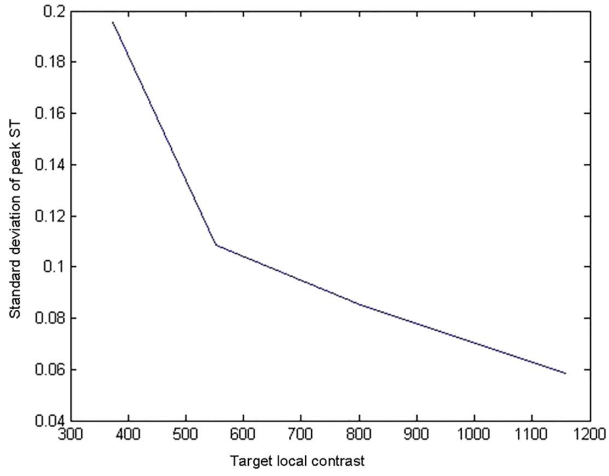

Fig. 8. Relationship between standard deviation of vertex SAT and target local contrast.

5. CONCLUSION

The selective visual attention mechanisms are introduced into clutter metrics, and two assumptions are presented in this paper. According to these assumptions, two thresholds for feature extraction and confusing block determination adaptively are further put forward. Next, we propose the adaptive threshold determination method, which simulates the adaptability of HVS. The high performance of the proposed clutter metric shows that the assumptions are consistent with the selective characteristics of HVS to some extent.

The SVA metric and the assumptions can be helpful in (1) establishing a HVS-based target acquisition model to find out the sensitive parameters and to guide the optimization design of an optoelectronic imaging system, (2) guiding the clutter suppression algorithm by quantifying the image clutter after suppression, and (3) assessing image quality by quantifying the similarity to the complete reference image based on the characteristic of HVS.

However, the analysis of misalignment of the vertices of the three surfaces suggests that local contrast is a very important clue toward improving the target acquisition model, and we are continuing efforts to integrate it with the clutter metric.

Funding. Key Scientific and Technological Projects of Jilin Province of China (20140204058GX, 20140204030GX).

Acknowledgment. We thank Dr. Alexander Toet for kindly providing us the Search_2 dataset. The Search_2 dataset has been made available online [32].

REFERENCES

1. R. G. Riggers, E. L. Jacobs, R. H. Vollmerhausen, B. O'Kane, M. Self, S. Moyer, J. G. Hixson, G. Page, K. Krapels, D. Dixon, R. Kistner, and J. Mazz, "Current infrared target acquisition approach for military sensor design and wargaming," Proc. SPIE **6207**, 1–17 (2006).
2. T. Meitzler, E. Sohn, H. Singh, and G. Gerhart, "Detection probability using relative clutter in infrared images," IEEE Trans. Aerosp. Electron. Syst. **34**, 955–962 (1998).
3. J. A. D'Agostino, W. Lawson, and D. L. Wilson, "Concepts for search and detection model improvements," Proc. SPIE **3063**, 14–22 (1997).
4. W. R. Reynolds, "Toward quantifying infrared clutter," Proc. SPIE **1311**, 232–240 (1990).
5. D. E. Schmieder and M. R. Weathersby, "Detection performance in clutter with variable resolution," IEEE Trans. Aerosp. Electron. Syst. **AES-19**, 622–630 (1983).
6. H. Singh, V. Gantam, M. Bhaskara, and S. Singh, "Two dimensional clutter: a new definition," in *Proceedings of the 36th Midwest Symposium on Circuits and Systems* (IEEE, 1993), pp. 88–91.
7. S. R. Rotman, G. Tidhar, and M. L. Kowalczyk, "Clutter metrics for target detection systems," IEEE Trans. Aerosp. Electron. Syst. **30**, 81–91 (1994).
8. G. Tidhar, G. Reiter, Z. Avital, Y. Hadar, S. R. Rotman, V. George, and M. L. Kowalczyk, "Modeling human search and target acquisition performance: IV. Detection probability in the cluttered environment," Opt. Eng. **33**, 801–808 (1994).
9. Q. Li, C. Yang, and J. Q. Zhang, "Target acquisition performance in a cluttered environment," Appl. Opt. **51**, 7668–7673 (2012).
10. X. C. Min, S. Z. Lin, and L. Y. Peng, "Metrics of image background clutter by introducing gradient features," Opt. Precis. Eng. **23**, 1838–1844 (2015).
11. H. Chang and J. Zhang, "New metrics for clutter affecting human target acquisition," IEEE Trans. Aero. Electron. Syst. **42**, 361–368 (2006).

12. X.-Q. Chu, C. Yang, and Q. Li, "Contrast-sensitivity-function-based clutter metric," Opt. Eng. **51**, 067003 (2012).
13. Q. Li, C. Yang, J.-Q. Zhang, and D.-Y. Zhang, "Hidden Markov models for background clutter," Opt. Eng. **52**, 0731108 (2013).
14. D. Xu, Z. Shi, and H. B. Luo, "A structural difference based image clutter metric with brain cognitive model constraint," Infrared Phys. Technol. **57**, 28–35 (2013).
15. Z. Wang and A. C. Bovik, "A universal image quality index," IEEE Signal Process. Lett. **9**, 81–84 (2002).
16. B. D. Vaughan, "Soldier-in-the-loop target acquisition performance prediction through 2001: integration of perceptual and cognitive models," in Army Research Lab Aberdeen Proving Ground Md Human Research and Engineering Directorate, Technical Report, ARL - TR - 3833, (July 2006).
17. L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. Pattern Anal. Mach. Intell. **20**, 1254–1259 (1998).
18. A. Yarbus, *Eye Movements During Perception of Complex Objects* (Springer, 1967).
19. L. Itti, *Models of Bottom-up and Top-down Visual Attention* (California Institute of Technology, 2000).
20. D. Walther, U. Rutishauser, C. Koch, and P. Perona, "Selective visual attention enables learning and recognition of multiple objects in cluttered scenes," Comput. Vis. Image Understanding **100**, 41–63 (2005).
21. A. Toet, P. Bijl, and J. M. Valeton, "Image dataset for testing search and detection models," Opt. Eng. **40**, 1760–1767 (2001).
22. V. Navalpakkam and L. Itti, "A goal oriented attention guidance model," in *Proceedings of Biologically Motivated Computer Vision* (Springer, 2002), pp. 453–461.
23. P. Van de Laar, T. Heskes, and S. Gielen, "Task-dependent learning of attention," Neural Netw. **10**, 981–992 (1997).
24. J. M. Henderson, J. R. Brockmole, M. S. Castelhano, and M. Mack, "Visual saliency does not account for eye movements during visual search in real-world scenes," in *Eye Movements: A Window on Mind and Brain* (Elsevier, 2007), pp. 537–562.
25. A. K. Mishra and B. Mulgrew, "Bistatic SAR ATR using PCA-based features," Proc. SPIE **6234**, 62340U (2006).
26. G. N. Ali, P.-J. Chiang, A. K. Mikkilineni, G. T.-C. Chiu, E. J. Delp, and J. P. Allebach, "Application of principal components analysis and Gaussian mixture models to printer identification," in *Proceedings of NIP and Digital Fabrication Conference* (Society for Imaging Science and Technology, 2004), pp. 301–305.
27. C. Yang, J. Wu, Q. Li, and J.-Q. Zhang, "Sparse-representation-based clutter metric," Appl. Opt. **50**, 1601–1605 (2011).
28. J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," IEEE Trans. Pattern Anal. Mach. Intell. **31**, 210–227 (2009).
29. B. J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," Proc. IEEE **98**, 1031–1044 (2010).
30. D. L. Wilson, "Image-based contrast-to-clutter modeling of detection," Opt. Eng. **40**, 1852–1857 (2001).
31. Z. Wang and A. C. Bovik, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process. **13**, 600–612 (2004).
32. A. Toet, "The Search_2 dataset," 2014, http://figshare.com/articles/The_Search_2_dataset/1041463, DOI: 10.6084/m9.figshare.1041463.