

Research on Visual Saliency Model Based on CovSal Algorithm and Histogram Contrast

Guoan Yang

Xi'an Jiaotong University, No.28
Xianning West Road, Xi'an, Shaanxi,
710049 China

gayang@mail.xjtu.edu.cn

Xinyu Zhang

Xi'an Jiaotong University, No.28
Xianning West Road, Xi'an, Shaanxi,
710049 China

2505891873@qq.com

Zhengzhi Lu

Xi'an Jiaotong University, No.28
Xianning West Road, Xi'an, Shaanxi,
710049 China

lu947867114@stu.xjtu.edu.cn

Yuhao Wang

Xi'an Jiaotong University, No.28
Xianning West Road, Xi'an, Shaanxi,
710049 China

313888139@qq.com

Junjie Yang

Xi'an Jiaotong University, No.28
Xianning West Road, Xi'an, Shaanxi,
710049 China

2523775890@qq.com

Wenyi Ren

Northwest A&F University, No.22
Xi'nong Road, Yangling, Shaanxi,
712100 China

renovelhuman@gmail.com

ABSTRACT

This paper proposes a novel visual saliency model based on the CovSal algorithm of the region covariance matrices and histogram contrast (HC) method. First, we give a new CovSal algorithm of the local saliency contrast by improving the center-surround segmentation method. Second, we add the HC algorithm of the global saliency contrast, and then we sparsify the global saliency map using the low-rank matrix. Finally, we integrate the local and global saliency maps through combining both the CovSal and HC contrast. Present paper proposes a new visual saliency model that combines the local and global saliency contrast algorithms, simultaneously retaining their advantages and eliminating their drawbacks. For example, the proposed model reduces the influence of background and texture details. The experimental results show that the performance of the visual saliency model proposed in this paper has been improved compared to the CovSal algorithm presented by Erdem and Erdem.

CCS Concepts

• Information systems→Information retrieval→Query representation→Document topic models→Retrieval models and ranking.

Keywords

visual saliency model; region covariances; histogram contrast algorithm; CovSal algorithm; low rank matrix.

1. INTRODUCTION

The human visual system (HVS) plays a crucial role in terms of humans' visual perception, understanding, and exploration of external environments. The HVS receives massive amounts of visual information every moment, but even in the face of complex natural scenes, the HVS can also quickly search for areas of

interest. This ability is called *visual attention*. With the continuous development of related fields such as neurobiology, cognitive psychology, and computer science in recent years, the visual attention mechanism, which usually aims to process complex visual data by simulating the human visual perception mechanism through mathematical models, has gradually become a research hot topic. The saliency model of the visual attention mechanism not only can filter out unimportant interference information in the natural scene but also can select the most salient object in the scene and consequently guide the movement of human eyeballs.

The research shows that the human visual attention mechanisms can be divided into two types: (1) fast unconscious data-driven bottom-up mechanism and (2) conscious task-driven top-down mechanism. These two types of mechanisms exist in the process of human visual attention simultaneously. Among them, the bottom-up saliency model is mainly used to predict foreground (human eye fixation) that focus on the saliency models of the local or/and global contrast [1]. The former bottom-up model is mainly used to define saliency through center-surrounded segmentation contrast. This method takes less consideration of the global impact and can only obtain some highly salient regions, so it is not ideal for processing complex background images. However, it can be improved through the multiscale analysis method. The top-down model takes into account the overall image relationship, and the significance of each part is defined by their difference from the entire image. Thus, it creates a highlight of the complete area, but the introduced image features are relatively few. In order to solve the aforementioned problems, this paper proposes a model that combines local contrast and global contrast for a visual saliency model. Harel et al. (2006) proposed a graph theory based visual saliency (GBVS) model [2]. Bruce (2007) proposed a saliency model (AIM model) based on information maximization theory [3]. After that, Zhang et al. (2008) proposed a Bayesian detective framework of the saliency model based on the natural statistics [4]. Hou and Zhang (2007) proposed a frequency-domain spectral residual (SR) algorithm [5]. Cheng (2015) proposed a saliency model based on histogram contrast (HC) that generated each pixel's saliency degree according to the contrast in the color histogram; they also added the spatial structure characteristics of an image based on the HC algorithm and proposed an region contrast (RC) algorithm based on block contrast [6].

The main contents of this paper are as follows: Section 2 describes relevant research and its results. Section 3 presents the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Request permissions from Permissions@acm.org.
ICCA 2018, October 12–14, 2018, Tokyo, Japan

© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-6563-5/18/10...\$15.00
<https://doi.org/10.1145/3284516.3284528>

algorithm for a novel visual saliency model. Section 4 presents experimental results and their analysis. Section 5 gives the conclusion and suggestions for further research.

2. RELATED VISUAL SALIENCY MODEL

2.1 Visual Saliency Model Based on the Difference of Center-Surrounded Statistics

The saliency model based on the difference between the center-surrounded statistics mainly uses the region covariance matrix to integrate the features nonlinearly [7]. Firstly, the method divides an image into non-overlapping image blocks with equal size on every scale and extracts the feature images. The region covariance of any an image block and the mean value of the feature image can quickly be calculated by using the integral image. Secondly, for one image block, we can calculate the differences between it and its surrounding image blocks with the same size. Note that the difference is proportional to the geodesic distance between the region covariances of the image blocks, whereas it is inversely proportional to the position distance between the image blocks. Finally, the visual saliency model further projects the region covariance matrix into the Euclidean space through the Cholesky factorization and assembles them with the mean value as feature vectors. Note that, when measuring the differences mentioned previously, the Euclidean distance between the feature vectors can be utilized to replace the geodesic distance between region covariance. Consequently, a first-order statistic is introduced into the visual saliency model. The saliency of the central image block is defined as the mean value of the difference between it and its closest m image blocks, where m is a parameter that needs to be set manually. Because the visual saliency model achieves multiscale saliency detection, the scale transformation can be implemented by changing the size of the image block matrix and the final saliency map is obtained through the multiplication and re-smoothing of the saliency maps in each scale.

2.2 Visual Saliency Model Based on HC Algorithm

In recent years, the HC algorithm based visual saliency model has been one of the most typical research results that is a bottom-up model based on global contrast analysis, which contrasts an image block with all the other image blocks in the entire image to achieve the calculation of the visual saliency. The visual saliency model can strip a large number of targets out of their surroundings, exhibit excellent performance compared to the visual saliency model of a local contrast, and produce higher saliency values near the edges. In the visual saliency model based on HC algorithm, the saliency value of each pixel is defined as the color difference of all the other pixels compared with this one, as follows [6]:

$$S(I_k) = S(c_i) = \sum_{j=1}^n f_j D(c_i, c_j)$$

where c_i denotes a color value in a pixel position, n denotes the number of different colors, f_j stands for a frequency produced color value c_i in an image, and $D(c_i, c_j)$ represents the Euclidean distance between two kinds of color values.

2.3 Performance Evaluation Method of Visual Saliency Model

In order to make an accurate assessment to the performance of the visual salience model in term of predicting the performance of

human eye fixation, this paper uses a bottom-up model to create evaluation indicators, which are described as follows.

AUC is the area under receiver operating characteristics curve [8], NSS is normalized scanpath saliency metric, Similarity is the similarity score proposed by Judd and colleagues [9], and EMD is the earth mover's distance. The bigger the former three indicators (AUC, NSS, and Similarity) are, the better the model performance is, whereas EMD is the dissimilarity score between two-probability distributions, and the smaller EMD is, the better the model performance is.

3. PROPOSED VISUAL SALIENCY MODEL

3.1 Title Region Covariance Matrix-CovSal algorithm

Region covariance matrices (RCMs) of the feature points in a visual image were first proposed as a compact region descriptor by Tuzel et al. [10]. Since then, it has effectively been utilized in high-level of computer vision applications for solving various problems such as texture recognition, object detection, object tracking, and so on. The region R in a RCMs model can be represented as the covariance matrix C_R of the feature points F :

$$C_R = \frac{1}{n-1} \sum_{i=1}^n (F_i - \mu)(F_i - \mu)^T \quad (1)$$

$$F_i = [L_i \ a_i \ b_i \ \left| \frac{\partial I_i}{\partial x} \right| \ \left| \frac{\partial I_i}{\partial y} \right| \ x \ y]$$

where $\{F_i\}$, $i = 1, \dots, n$ denotes the d -dimensional feature points in the region R , and μ is the mean value of these points. Note that L , a , and b denote color space; I denotes the pixels in an image; and x and y denote image coordinates. Here we have a total of seven parameters. The covariance matrix is a symmetric matrix with its diagonal elements representing the feature variances and its nondiagonal elements representing the correlations of the feature points. In [10], the researchers also proposed a fast algorithm for computing covariance matrices of rectangular regions by using the first- and the second-order integral image representations with $O(d^2)$ computational complexity [11]. Here, the covariance matrix is a symmetric semi-positive definite matrix that can be defined as a strictly symmetric positive definite matrix by adding a small value to its diagonal. Note that covariance matrices lie on Riemannian manifold space but not on Euclidean space. The distance between two covariances C_1 and C_2 can be expressed as the geodesic distance between two points in the Riemannian manifold space [12], as follows:

$$d_{geo}(C_1, C_2) = \sqrt{\sum_{i=1}^d \ln^2 \lambda_i(C_1, C_2)} \quad (2)$$

where $\{\lambda_i(C_1, C_2)\}$ $i = 1, 2, \dots, d$ are the generalized eigenvalues of C_1 and C_2 , and calculated as follows:

$$\lambda_i C_1 x_i - C_2 x_i = 0, \quad i = 1, 2, \dots, d. \quad (3)$$

where x_i is the generalized eigenvector of C_1 and C_2 , respectively.

The region covariance is an effective regional descriptor that can naturally fuse the features to not need to use the normalized feature or weight parameters, suppress noises and make it easy to calculate by subtracting the mean value μ , and possess partially rotation invariant properties as the statistical value of feature in the region. Otherwise, the dimensionality of their region covariance matrix for any size areas is the same; hence, we can compare the features in different areas directly. Therefore, we can take advantage of the region covariance as the saliency feature of the local contrast in difference between center-surrounded statistics for the bottom-up models.

3.2 Low Rank Matrix and Sparse Representation

Research on visual characteristics and sparse representation of images was first proposed by Olshausen and Field [13], whose work revealed that visual reception possesses multiresolution characteristics, directional selectivity, band-pass characteristics, and anisotropic properties because an image can be described as a matrix such that a sparse matrix can enable it to represent an image well using fewer data. The robust principal component analysis (RPCA) [14] is the sparse representation method; it can enable that matrix X is decomposed as both the low-rank matrix A and the sparse matrix E in a constraint condition, as follows:

$$\begin{aligned} \min \|A\|_* + \lambda \|E\|_1 \quad (4) \\ s. t. \quad X = A + E \end{aligned}$$

Where $\|\cdot\|_*$ denotes the nuclear norm of a matrix, $\|\cdot\|_1$ denotes the norm of ℓ_1 space, and λ ($\lambda \geq 0$) is a balance coefficient. Here, Eq. (4) is a convex relaxation problem of rank function, matrix E contains the full saliency region, and matrix A contains the non-saliency region. Generally, the balance coefficient can be described as follows:

$$\lambda = \frac{1}{\sqrt{\max(m, n)}}$$

where m and n stand for the number of pixels in an image for horizontal and vertical directions.

When people observe a natural scene, they do not need to capture the full information of the scene image. Rather, just some salient areas are called to attention by human eyes. Therefore, the saliency region can be expressed by a sparse matrix, and the redundant information can be eliminated as noise in the visual image. From the cognitive sciences we know that a visual image can divide into two parts, the sparse part and the redundant part, as follows:

$$Info(Image) = Info(Redundancy) + Info(Saliency) \quad (5)$$

In Eq. (5), $Info(Redundancy)$ characterizes the highly related part in the visual information processing that represents the nonsalient regions such as background and texture detail of an image. $Info(Saliency)$ characterizes the uncorrelated part that attracts the most interest of people; it represents the salient region of a visual image. Therefore, in this paper we can decompose the parts of background and salient regions of a visual image into the low-rank matrix and the sparse matrix. Moreover, the convex relaxation problem is solved for Eq. (4) by using the inexact augmented Lagrangian multiplier (IALM), and the augmented Lagrangian function is constructed as follows:

$$L(A, E, \Gamma, \mu) = \|A\|_* + \lambda \|E\|_1 + \frac{\mu}{2} \|X - A - E\|_F^2 + \langle \Gamma, X - A - E \rangle \quad (6)$$

where the IALM updates A , E , and Γ iteratively to minimize $L(A, E, \Gamma, \mu)$, and $\|\cdot\|_F$ denotes the Frobenius norm.

3.3 Visual Saliency Model Based on the Local Contrast

Generally, the initial saliency maps consist of the local and global contrast saliency maps using the bottom-up model. In the local contrast algorithm, we mainly aim at improving the difference between the center-surrounded statistics. Among them, the local contrast saliency map is obtained by dividing the original image into non-overlapping small blocks with the same size on the five scales, and then computing the image saliency successively by calculating the difference between the center-surrounded statistics. The feature vector of each small block is composed of color, direction, and position, in which each block's covariance is composed of a 7×7 matrix and their mean value consists of a 7×1 vector. Here, the test images of 512×512 pixels are adopted from the MIT 300 image database (http://saliency.mit.edu/results_mit300.html). The size of small blocks at the five scales are $K = \{8, 16, 32, 64, 128\}$, respectively.

This paper improved the segmentation method in Erdem and Erdem [7] and applied simultaneously the two center-surrounded segmentation methods for creating the local contrast saliency map. The first segmentation, the method using the CovSal algorithm [7], is to divide the central block into the ‘‘center,’’ and the ‘‘surround,’’ composed of the surrounding $(2 \times m + 1)^2 - 1$ ($m = 2$ in Fig. 1) squares with the same size, as shown in Fig. 1.

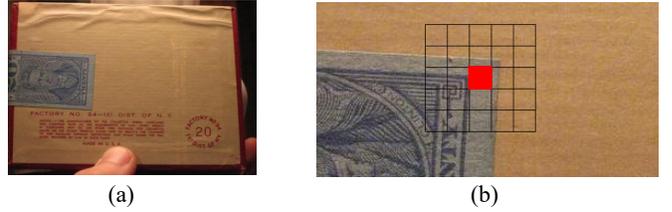


Fig. 1: The segmentation method adopted by the CovSal algorithm: (a) the original image; and (b) the segmentation method

The saliency of each block is defined as the smallest value of $S(R_i)$ obtained from the surrounding $(2 \times m + 1)^2 - 1$ square images, and $S(R_i)$ can be expressed as follows:

$$S(R_i) = \frac{1}{m} \sum_{j=1}^m d(R_i, R_j) \quad (7)$$

The three components R_i , R_j , and $d(\cdot)$ denote the central block, the surrounded block, and the distance between them, respectively. The distance $d(\cdot)$ can be calculated by

$$d(R_i, R_j) = \frac{\text{covdist}(C_i, C_j)}{1 + \|x_i + x_j\|} \quad (8)$$

where C_i and C_j represent, respectively, the covariance matrices of region R_i and R_j , $\text{covdist}(C_i, C_j)$ represents the geodesic distance between two points in Riemann space, and x_i and x_j stand for the image coordinates of region R_i and R_j , respectively. Thus, the

covariance distance is weighted by the opposite spatial distance while determining the region features, thereby one can reduce the influence of neighboring regions in terms of visually similarity.

A second segmentation method in the local contrast [15], as shown in Fig. 2, is based on an assumption accepted widely, namely that if an image block is salient in all the scales then it can be regarded as saliency area. The saliency of each block is defined as the smallest value of $S(R_i)$ obtained from all the surrounding square images, and the calculation of $S(R_i)$ is expressed in Eq. (9):

$$S(R_i) = \text{covdist}(C_i, C_j) + \text{Dist}(M_i, M_j) \quad (9)$$

where $\text{Dist}(M_i, M_j)$ denotes the Euclidean distance between mean values, and M_i and M_j denote the mean values of features corresponding to both the central block and surrounding blocks, respectively.

In this paper, we combine the preceding two kinds of segmentation method for local contrast and simultaneously consider that if the human’s observation is on image center, then a saliency region is generally concentrated in the center area of the image. Therefore, it can’t be expanded to the surrounded blocks of the margins, so we set its saliency as zero to be the smallest value. In this paper, we presented a new saliency map using two different segmentation methods at each scale that the maps are normalized and weighted according to $0.5 \times S_1 + 0.55 \times S_2$, and the final local contrast saliency map is obtained by integrating the five initial saliency maps at different scales to be the same size, then smoothing them, and finally achieving the map-chain multiplication, as given by Eq. (10):

$$\text{Sal}(k) = G_\sigma(k) * \prod_{c \in MS} G_\sigma(k) * S_{ci}(k) \quad (10)$$

where $G_\sigma(k)$ denotes a Gaussian filter that plays a role of smoothing and blurring images, and S_{ci} represents a maximum value of normalized saliency map in different scales. The Gaussian standard deviation σ here is $0.02 \times W$, where W is the width of an image. MS refers to the initial local contrast saliency map at each scale. The local contrast saliency map obtained by our approach possesses the best performance compared with the CovSal algorithm, as shown in Fig. 3.

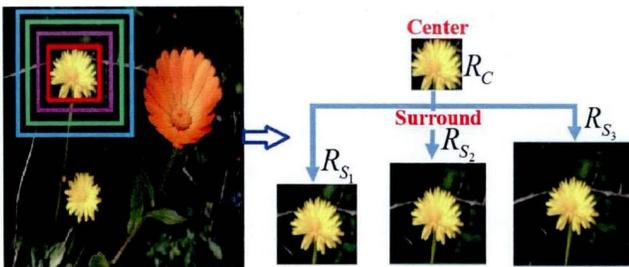


Fig.2: Segmentation method proposed in [15]

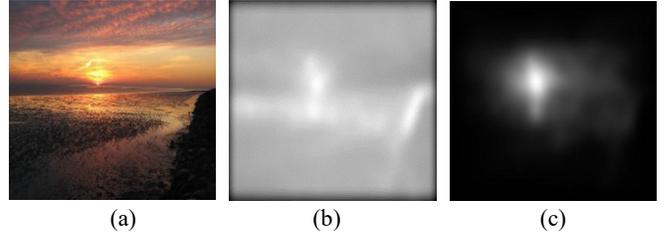


Fig.3: Comparison of the saliency maps between CovSal and proposed algorithm: (a) original image; (b) CovSal algorithm; (c) local contrast saliency map proposed in this paper

3.4 Visual Saliency Model Based on Global Contrast

Firstly, for the visual saliency model based on the global contrast, this paper presented a new HC algorithm with the low-rank matrix. It uses the low-rank matrix decomposition to sparsify the saliency maps in the HC algorithm. Because the local contrast algorithm divides an image into small blocks that results in many disadvantages, such as the saliency map being based on the local contrast, the algorithm can’t ensure the integrality of the saliency map, so we hope to use a global contrast saliency map to improve this drawback, for example using the RGB color space. From Fig. 4 we can see that the local contrast saliency maps do not provide a highly salient representation for the large salient regions with identical colors or texture details, for instance the body parts of the dog, whereas the HC algorithm can effectively handle that. However, the HC algorithm has retained completely the texture details of the original image, but these texture details do not contribute to the final saliency map. Therefore, we adopted low-rank matrix decomposition to remove the background or texture details, treating it as noise as much as possible. The effect of the algorithm proposed in this paper is shown in Fig. 4.

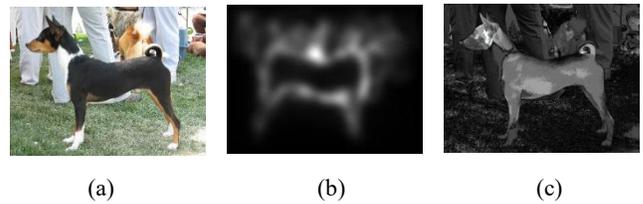


Fig. 4; Comparison of saliency maps of different algorithms: (a) original image; (b) saliency map proposed local algorithm; (c) saliency map of HC algorithm

Secondly, this paper uses the IALM to obtain an initial saliency map based on global color contrast of the HC algorithm, and then performs the low-rank decomposition of the saliency map. The resulting sparse matrix is normalized such that it is eventually easy to fuse the saliency map obtained from the local contrast. From Fig. 5 we can see that the global contrast saliency map Fig. 5(d) based on our algorithm is excellent compared with Fig. 5(b), which adopted the normal HC algorithm. It not only effectively restrained the background and texture details of the saliency map but also the salient parts are fully preserved.

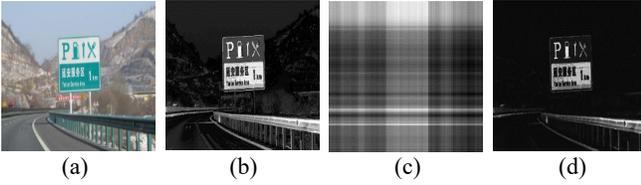


Fig.5: Saliency map based on HC and low-rank matrix decomposition: (a) original image; (b) initial saliency map obtained by HC algorithm; (c) low-rank part of initial saliency map after low-rank decomposition; (d) sparse part of initial saliency map after low-rank decomposition

Finally, the present paper fused both the local and global contrast saliency maps through a weighted method in the ratio of 5:2 and created a final saliency map, as shown in Fig. 6. It can be seen from the effect of the saliency map that the visual saliency model proposed in this paper better implemented a saliency representation for the overall or large saliency region.

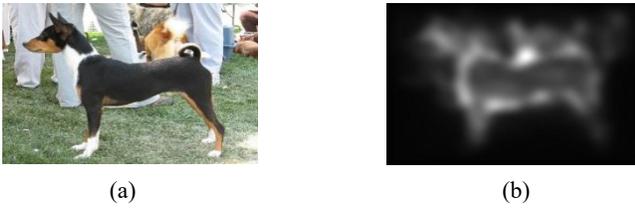


Fig. 6 Saliency map of the proposed algorithm after integration: (a)original image; (b) visual saliency map after integration

3.5 Proposed Algorithm for the Visual Saliency Model

The overall algorithm flowchart of this paper is shown in Fig. 7.

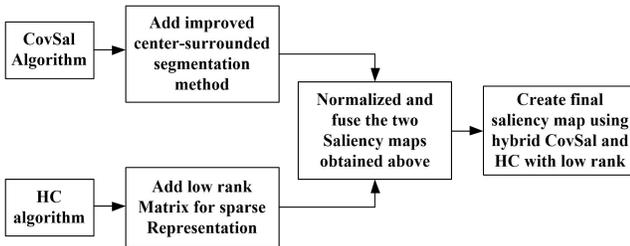


Fig. 7: The overall flow chart of the algorithm in this paper

4. Experimental Results and Analysis

The experimental test set used in this paper is the MIT 300 launched by the MIT Vision Group, which consists of 39 subjects' eye tracking datasets for 300 different natural images and is an authoritative test set for visual saliency models. The objective comparison method for the algorithm performance uses NSS, AUC, EMD, and Similarity score.

The performance of the visual saliency model is tested by the MIT Saliency Benchmark Team [16-18], as shown in Table 1. Experimental results show that the hybrid CovSal and HC algorithm with low-rank matrix decomposition in the visual saliency model proposed in this paper is superior to the CovSal algorithm proposed by Erdem and Erdem [7], as follows:

1) This paper improved the CovSal algorithm. Compared to the CovSal algorithm, the following changes were observed: AUC

decrease 0.006, NSS decrease 0.01, Similarity increase 0.003, and EMD increase 0.049;

2) This paper proposed the hybrid CovSal and HC algorithm with low-rank decomposition. Compared to the CovSal algorithm, the following changes were observed: AUC is increased by 0.002, NSS is increased by 0.01, EMD is decreased by 0.292, and Similarity is increased by 0.005.

Table 1. Objective comparison of saliency model performance

	AUC	NSS	EMD	Similarity
Itti et al. (1998)	0.750	0.97	4.560	0.405
Harel et al. (2007)	0.801	1.24	3.574	0.472
Torralba et al. (2006)	0.684	0.69	4.715	0.343
Hou & Zhang(2007)	0.682	1.01	5.368	0.319
Zhang et al.(2008)	0.672	0.68	5.088	0.340
Bruce & Tsotsos (2009)	0.751	0.79	4.236	0.390
Goferman et al. (2010)	0.742	0.95	4.900	0.390
Judd et al. (2009)	0.810	1.18	4.450	0.420
RC algorithm(2014)	0.790	1.18	3.480	0.408
CNN-VLM algorithm (2015)	0.790	1.18	4.550	0.430
Original Covsal (2013)	0.806	1.22	3.390	0.470
Improved CovSal algorithm	0.800	1.21	3.341	0.473
Our proposed hybrid Covsal/HC algorithm with low Rank	0.806	1.191	3.625	0.475

5. CONCLUSIONS

This paper combines the CovSal local saliency algorithm and the HC global saliency algorithm to extract their respective advantages and eliminate their drawbacks; therefore, the performance of the bottom-up visual saliency model is improved. A new center-surrounded segmentation approach is presented in the local contrast visual saliency model, and a low-rank matrix sparse representation method is added to the global contrast visual saliency model. The experimental results show that the visual saliency model proposed in this paper exhibits some improvement over other representative saliency models. For future research, the method of weighted integration of local and global saliency maps can further be improved, for example by adding an adaptive integration method to the weighting. The traditional visual feature representation algorithm has reached a bottleneck period, and deep learning and deep convolutional neural networks are required to further improve the performance of the saliency model.

6. ACKNOWLEDGMENT

This work was partly supported by the National Natural Science Foundation of China under grant No. 61673314, No.61573272, No.11504297.

7. REFERENCES

- [1] Itti, L., Koch, C., and Niebur, E. 1998. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, Vol. 20, No. 11, pp. 1254–1259.
- [2] Harel, J., Koch, C., and Perona, P. 2006. Graph-Based Visual Saliency. *International Conference on Neural Information Processing Systems*, MIT Press, 2006:545–552.
- [3] Bruce, N., and Tsotsos, J. 2007. Attention Based on Information Maximization. *Journal of Vision*, 2007, 7(9):950–950.
- [4] Zhang, L., Tong, M., H., Marks, T., K., et al. 2008. SUN: A Bayesian Framework for Saliency Using Natural Statistics. *Journal of Vision*, 2008, 8(7):32, 1–20.
- [5] Hou, X., and Zhang, L. 2007. Saliency Detection: A Spectral Residual Approach. *IEEE Conference on Computer Vision and Pattern Recognition*, 2007:1–8.
- [6] Cheng, M., M., Mitra, N., J., Huang, X., et al. 2015. Global Contrast Based Salient Region Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(3):569–582.
- [7] Erdem, E., and Erdem, A. 2013. Visual Saliency Estimation by Nonlinearly Integrating Features Using Region Covariances. *Journal of Vision*, 2013, 13(4):11, 1–20.
- [8] Garcia-Diaz, A., Leborán, V., Fdez-Vidal, X., R., et al. 2012. On the Relationship between Optical Variability, Visual Saliency, and Eye Fixations: A Computational Approach. *Journal of Vision*, 2012, 12(6):17, 1–22.
- [9] Judd, T., Durand, F., and Torralba, A. 2012. A Benchmark of Computational Models of Saliency to Predict Human Fixations. *MIT Technical Reports*, 2012, <http://hdl.handle.net/1721.1/68590>.
- [10] Tuzel, O., Porikli, F., and Meer, P. 2006. Region Covariance: A Fast Descriptor for Detection and Classification. *Computer Vision-ECCV 2006*, Springer Berlin Heidelberg, 2006:589–600.
- [11] Viola, P., and Jones, M. 2001. Rapid Object Detection Using a Boosted Cascade of Simple Features. *IEEE International Conference on Computer Vision and Pattern Recognition, CVPR 2001*, 2001, vol. 1, pp. 511–518.
- [12] Tuzel, O., Porikli, F., and Meer, P. 2008. Pedestrian Detection via Classification on Riemannian Manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(10):1713–1727.
- [13] Olshausen, B., A., and Field, D., J. 1996. Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images.” *Nature*, 1996, vol. 6583, no. 381, pp. 607–609.
- [14] Hyvarinen, A. 1999. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 1999, 10(3):626–634.
- [15] Li, Y., and Xu, J. 2014. A Multi-Factor Integrated Model of Visual Attention.” *International Conference on Computational Intelligence for Modelling, Control and Automation, CIMCA 2014*, pp. 82–86.
- [16] Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F. A. Oliva, and A. Torralba, *MIT Saliency Benchmark*. Available at: <http://saliency.mit.edu>
- [17] Judd, T., Durand, F., and Torralba, A. 2012. A Benchmark of Computational Models of Saliency to Predict Human Fixations. *MIT technical report*, 2012 (<http://hdl.handle.net/1721.1/68590>)
- [18] Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., Durand, F. 2016. What do different evaluation metrics tell us about saliency models? *arXiv:1604.03605*, 2016 (<https://arxiv.org/abs/1604.03605>)