

## Research Article

# Multichannel Saliency Detection Based on Visual Bionics

Lidan Cheng , Tianyi Li , Shijia Zha , Wei Wei , and Jihua Gu

*College of Optoelectronics Science and Engineering, Soochow University, Suzhou Jiangsu Province 215000, China*

Correspondence should be addressed to Wei Wei; [weiwei0728@suda.edu.cn](mailto:weiwei0728@suda.edu.cn)

Received 28 May 2020; Revised 11 October 2020; Accepted 9 November 2020; Published 23 November 2020

Academic Editor: Liwei Shi

Copyright © 2020 Lidan Cheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Inspired by the visual properties of the human eyes, the depth information of visual attention is integrated into the saliency detection to effectively solve problems such as low accuracy and poor stability under similar or complex background interference. Firstly, the improved SLIC algorithm was used to segment and cluster the RGBD image. Secondly, the depth saliency of the image region was obtained according to the anisotropic center-surround difference method. Then, the global feature saliency of RGB image was calculated according to the colour perception rule of human vision. The obtained multichannel saliency maps were weighted and fused based on information entropy to highlighting the target area and get the final detection results. The proposed method works within a complexity of  $O(N)$ , and the experimental results show that our algorithm based on visual bionics effectively suppress the interference of similar or complex background and has high accuracy and stability.

## 1. Introduction

Saliency detection is an important research content in computer vision, which refers to the process of simulating human visual attention mechanism to accurately and quickly detect the most interesting regions in images. Borji et al. defined that saliency visually described the prominent target or area in the scene relative to its neighbouring area [1]. The human visual attention mechanism prioritizes a few significant areas or objectives, while ignoring or discarding others that are not, which can allocate computing resources selectively and greatly improve the efficiency of visual information processing. Therefore, the saliency computing model based on visual attention mechanism has been widely studied. When processing the input image or video, the computer can judge the importance of its visual information by detecting the saliency area. It has been widely applied in object detection and identification [2], image retrieval [3], video quality assessment [4], video compression [5], image cropping [6], and other fields.

The RGB image saliency detection model based on visual attention mechanism uses low-level feature contrast to calculate saliency [7, 8]. Typical of them are global feature comparison calculation model [9], local feature comparison

calculation model [10], and combination of global and local feature comparison model [11]. In order to improve the accuracy of detection, the saliency detection model was proposed based on prior knowledge [12]. Typical of them are position prior [13], background prior [14, 15], colour prior [16], shape prior [17], and boundary prior [18, 19].

However, most 2D image saliency detection models based on human visual attention mechanism ignore the fact that human visual attention mechanism is based on 3D scene. It shows that depth provides extra important information of saliency detection for RGB image. Desingh et al. discussed 3D saliency detection methods based on depth appearance, depth-induced blur, and centre-bias [20]. Niu et al. conducted depth saliency detection based on parallax contrast and professional knowledge in vertical photography [21]. Further, Ju et al. proposed a depth saliency detection model based on depth image anisotropic center-surround difference [22]. Ren et al. [23], respectively, proposed the saliency detection of RGB-D images against a complex background by combining the prior knowledge of depth, indicating the validity of depth information in 3D saliency detection. However, there are two challenges in the process of saliency detection of RGB-D images. The first is how to calculate the saliency of depth images under similar or

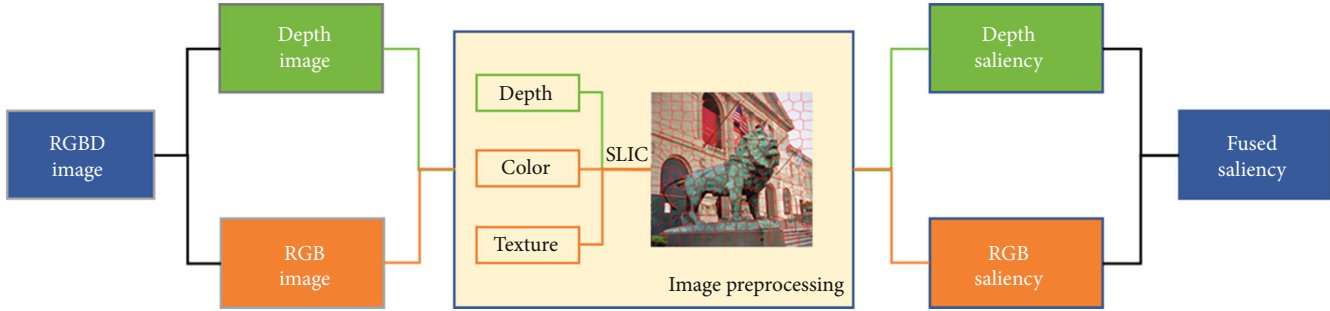


FIGURE 1: Our framework of saliency detection.

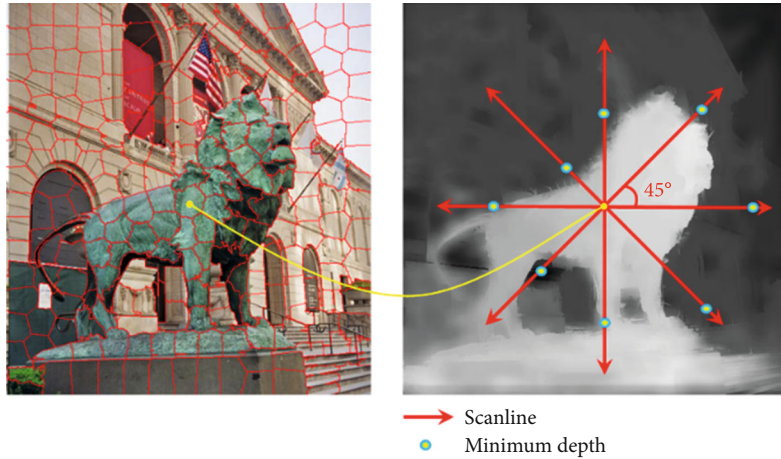


FIGURE 2: Example of the ACSD operator in a region.

complex background interference, and the second is how to combine the saliency map of depth image and RGB image to obtain the final result with a good performance. In this paper, we proposed a multichannel saliency detection method based on RGBD images, which has the following contributions:

- (1) On the basis of SLIC algorithm, colour, texture, and depth information are used to measure the distance of superpixel segmentation
- (2) Based on the perception rule of human vision, we introduced the depth information and global information of RGB image as two feature channels for saliency computing
- (3) The weighted features of depth saliency and colour saliency were fused by information entropy, and experiment shows that the algorithm has a good performance in case of background interference

## 2. Saliency Detection

The algorithm framework of this paper is shown in Figure 1. Combining the depth map with the RGB map to carry out image preprocessing and colour, texture, and depth information are introduced as the basis of superpixel segmentation. Then, the colour and depth information were calculated as

two feature channels of saliency map. As is shown in Figure 2, the depth saliency was obtained by the anisotropic center-surround difference (ACSD) method, and the global saliency of RGB image was calculated by global contrast method based on HSV space. Finally, information entropy is used to calculate the weights of two channels, respectively, and get the final fused saliency map.

## 3. Image Preprocessing

The human visual observation system takes the image region as the basic unit, and the saliency detection based on the region conforms to the visual characteristics of the human eyes. As a construction method of pixel region, superpixel technology has been widely used in computer vision field. Superpixel can quickly segment the image into subregions with certain semantics, which is conducive to the extraction of local features and the expression of structural information [24]. SLIC algorithm has obtained a good balance in the two aspects of edge fitting degree and compactness, which has an excellent comprehensive performance. When the SLIC is used to segment the left image, the obtained boundary is not accurate because of ignoring the mutual constraint relationship between the 2D and depth information. Therefore, colour, texture, and depth information are used to measure the distance of superpixel segmentation in this paper.

Converting the left image to the CIE Lab colour space and dividing the image into  $k$  superpixels. Here, each pixel has a unique identifier  $i$ . Extract the follow 7 d characteristics of each superpixel region as measurement property. It can be expressed:

$$S_{p^i} = \{l_i, a_i, b_i, C_{coni}, C_{cori}, E_i, d_i\} \quad i = 1, 2, 3, \dots, k, \quad (1)$$

where  $l_i, a_i,$  and  $b_i$  are the mean value of  $L, a,$  and  $b$  colour components of each superpixel region;  $C_{coni}, C_{cori},$  and  $E_i$  are the mean value of contrast, cross-correlation, and energy mean of gray level cooccurrence matrix of each superpixel region; and  $d_i$  is the depth value of each superpixel region. Then, we can describe the adjacent superpixel pair as  $S_{p^i j}$ :

$$S_{p^i j} = \{(S_{p^i}, S_{p^j})\} \quad i \in [1, k], j \in [1, k], i \neq j, \quad (2)$$

where  $S_{p^i j}$  superpixel pair with  $i$  and  $j$  as identifier,  $k$  is the number of superpixels of the image, and  $S_{p^i}$  and  $S_{p^j}$  are the 7 d characteristics of the adjacent superpixel pair. The number of adjacent superpixel pairs in each image is determined by SLIC superpixel segmentation.

Using colour, texture, and depth features to calculate the difference between all adjacent superpixel pairs  $S_{p^i j}$ .  $d_{lab}$ ,  $d_{glcm}$ , and  $d_{depth}$  are defined to describe the measurement of colour, texture and depth characters:

$$\begin{aligned} d_{lab} &= \sqrt{(l_i - l_j)^2 + (a_i - a_j)^2 + (b_i - b_j)^2}, \\ d_{glcm} &= \sqrt{(C_{coni} - C_{conj})^2 + (C_{cori} - C_{corj})^2 + (E_i - E_j)^2}, \\ d_{depth} &= \sqrt{(d_i - d_j)^2}. \end{aligned} \quad (3)$$

Then, the distance measurement of superpixel segmentation  $D_{ij}$  is

$$D_{ij} = \omega_1 \cdot \sqrt{\frac{\varepsilon + d_{lab}^2}{3}} + \omega_2 \cdot \sqrt{\frac{\varepsilon + d_{glcm}^2}{3}} + \omega_3 \cdot \sqrt{\varepsilon + d_{depth}^2}, \quad (4)$$

where  $\varepsilon = 10^{-4}$ . It is used to ensure the validity of the value.  $\omega_1, \omega_2,$  and  $\omega_3$  are the weight of colour, texture, and depth.

In the image, the greater the discreteness of a feature data set is, it means that the more influence this feature has on the image. Mean variance can effectively represent the degree of difference between data. Therefore, the global mean variance of colour, texture, and depth is used as the weight values of the three features  $\omega_1, \omega_2,$  and  $\omega_3$ .

If the difference between adjacent superpixels is less than a certain threshold  $th_1$ , the adjacent superpixel pair will be merged.

$$th_1 = \sqrt{\frac{\omega_1 \cdot (\bar{l} + \bar{a} + \bar{b})}{3}} + \sqrt{\frac{\omega_2 \cdot (\overline{C_{con}} + \overline{C_{cor}} + \overline{E})}{3}} + \sqrt{\omega_3 \cdot \bar{d}}, \quad (5)$$

where  $\bar{l}, \bar{a},$  and  $\bar{b}$  are the mean value of  $L, a,$  and  $b$  colour components of the image;  $\overline{C_{con}}, \overline{C_{cor}},$  and  $\overline{E}$  are the mean value of contrast, cross-correlation, and energy mean of gray level cooccurrence matrix of the image; and  $\bar{d}$  is the depth value of the image.

Finding all similar adjacent superpixel pairs and taking the upper left superpixel of the image as the starting point of clustering. The output after clustering contains  $n$  regions  $R_i, 1 \leq i \leq n$ .

#### 4. Depth Saliency Map

For each superpixel, the anisotropic center-surround difference (ACSD) value is calculated, and the value of center superpixel is assigned to each pixel within the region  $R_i$ . Performing an anisotropic scan along multiple directions, in each scanline, assuming the pixel with the minimum depth value as background and calculate the difference between the center pixel and background.  $L$  is the maximum scan length for each scanline. The typical value of  $L$  is a third of the diagonal length.

The anisotropic center-surround difference (ACSD) is summed over eight scanning directions  $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ,$  and  $315^\circ$ . The mathematical description of anisotropic center-surround difference (ACSD) measure is

$$\begin{aligned} S_d^m(x, y) &= d(x, y) - \min(d^m), \quad n \in [1, L], \\ S_d(x, y) &= \sum_{m=1}^8 S_d^m, \end{aligned} \quad (6)$$

where  $S_d^m$  indicates the ACSD value of pixel  $(x, y)$  along the scanline  $m$ .  $d(x, y)$  is the depth value of pixel  $(x, y)$ .  $n$  is the index of the pixels along the scan path  $m$ .  $\min(d^m)$  is the minimum depth value along the scanline  $m$ .  $S_d(x, y)$  is the ACSD value of pixel  $(x, y)$  which sums the center-surround difference values in eight directions.

#### 5. Global Saliency Map of RGB Image

Colour histogram is used to regularize the colour of the image to level 128 in order to reduce the computational complexity and save the storage space. On the other hand, the descending dimension algorithm for HSV colour space is proposed. With the decrease of saturation, any colour in HSV space can be described by the change of gray level. The intensity value determines the specific gray level of the conversion [25]. When the colour saturation is close to zero, all pixels look similar regardless of hue. With the increase of saturation, the pixels are distinguished by hue value.

Compared with colour saturation, human vision is more sensitive to hue and intensity. The pixels with lower



FIGURE 3: Saliency comparisons of different methods in terms of NJU400 dataset: (a) RGB image; (b) depth image; (c) ground truth; (d) GS; (e) MC; (f) MR; (g) WCTR; (h) ACS; (i) MSD.

colour saturation can be approximately represented by intensity level, while the pixels with higher colour saturation can be approximately represented by hue. Saturation value is used to determine whether each pixel can be represented by hue or intensity value, which is more consistent with the law of human visual perception. Saturation threshold  $th_2$  is

$$th_2 = 1 - \frac{0.8 I_v(x, y)}{255}, \quad (7)$$

where  $I_v(x, y)$  represents the  $V$  component value of a pixel. When the saturation value  $I_s(x, y)$  is greater than  $th_2$ , the pixel point is represented by the hue value  $I_h(x, y)$ ; when the saturation value is less than  $th_2$ , the pixel point is represented by the intensity value  $I_i(x, y)$ . The saliency of each pixel is

$$S_c(x, y) = \begin{cases} |\bar{I} - I_v(x, y)| I_s(x, y) \leq th_2 \\ |\bar{I} - I_h(x, y)| I_s(x, y) > th_2 \end{cases} \forall (x, y) \in R_i, \quad (8)$$

where  $I_s(x, y)$  is the saturation value of the pixel,  $I_h(x, y)$  is the hue value of the pixel,  $I_v(x, y)$  is the intensity value of the pixel, and  $\bar{I}$  is the mean value of all pixels.

## 6. Fusion of Saliency Map

When synthesizing the colour saliency map and the depth saliency map, the information entropy is used to calculate the weights of the channels.

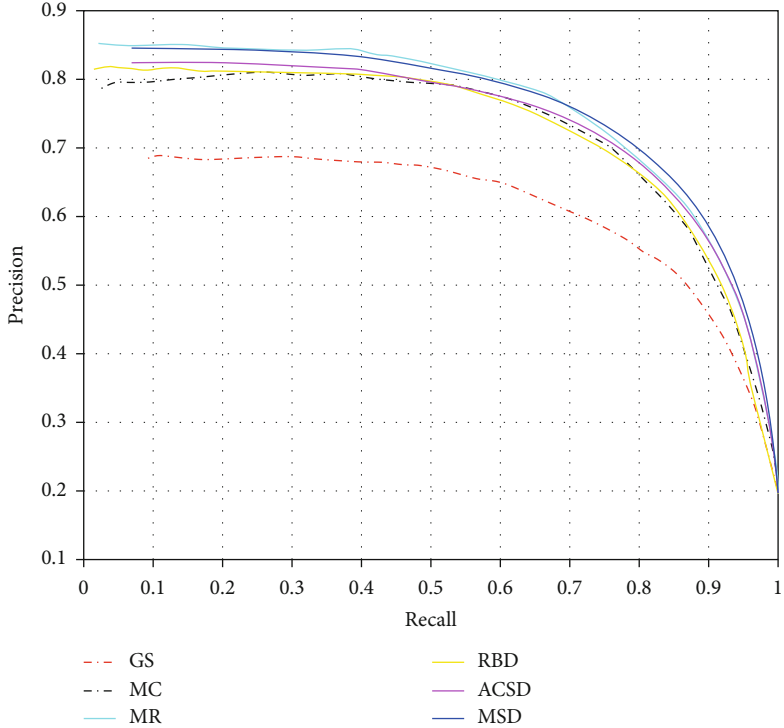
The information entropy of colour saliency is

$$H_c(R) = - \sum_{i=1}^n p_c(R_i) \log(p_c(R_i)), \quad (9)$$

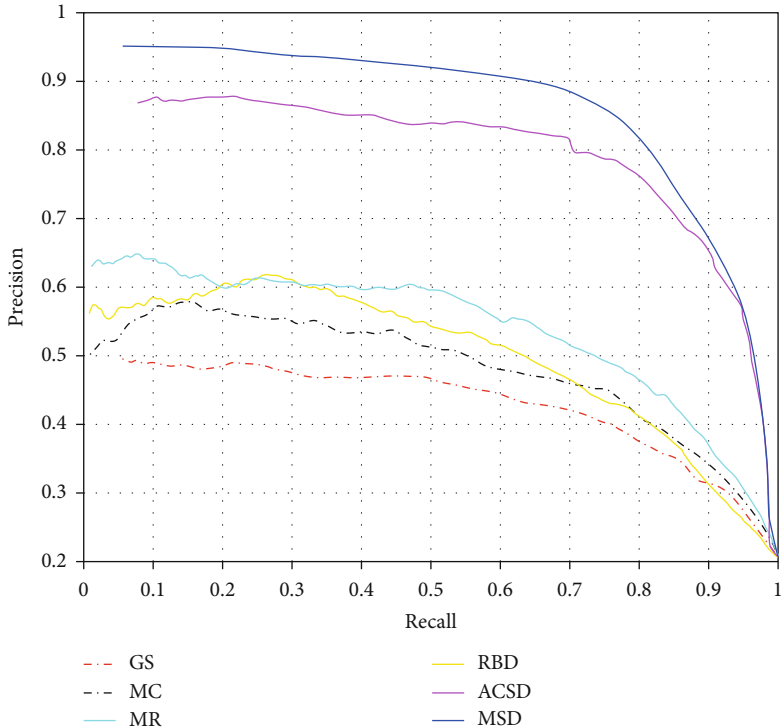
where  $p_c(R_i)$  is the ratio of the sum of  $R_i$  colour saliency values to the whole image.

The information entropy of depth saliency is

$$H_d(R) = - \sum_{i=1}^n p_d(R_i) \log(p_d(R_i)), \quad (10)$$



(a)



(b)

FIGURE 4: Continued.

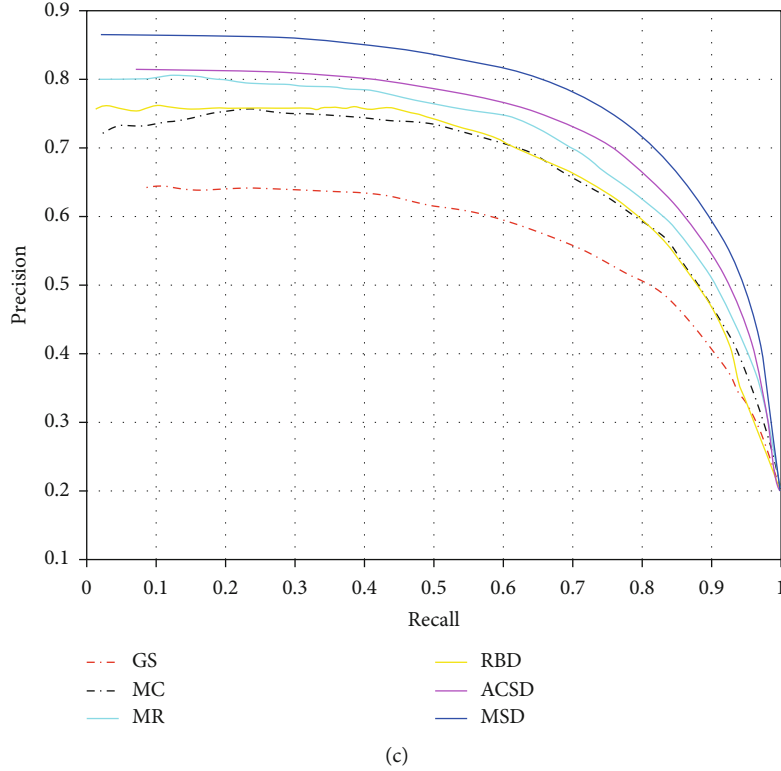


FIGURE 4: The precision-recall curves of different algorithms: (a) the precision-recall curves on the N group; (b) the precision-recall curves on the S/C group; (c) the precision-recall curves on full datasets.

where  $p_d(R_i)$  is the ratio of the sum of  $R_i$  depth saliency values of to the whole image.

The saliency map  $S_{\text{fuse}}(x, y)$  was obtained by fusing the two channels:

$$S_{\text{fuse}}(x, y) = \frac{H_c(R)}{H_c(R) + H_d(R)} S_c(x, y) + \frac{H_d(R)}{H_c(R) + H_d(R)} S_d(x, y) \forall (x, y) \in R_i. \quad (11)$$

## 7. Experimental Comparison

We show a few saliency maps generated by different algorithms in Figure 3.

The precision-recall curve is evaluated from two aspects: precision and recall. Precision refers to the ratio between the number of correct saliency pixels and the whole number of saliency pixels, which is used as the  $y$ -axis. Recall refers to the ratio of the number of correct saliency pixels to the number of true pixels, which is used as the  $x$ -axis.

The algorithms are tested on NJU400 datasets. Two test sets are divided from NJU400 according to the complexity and the similarity of the background. Four volunteers are invited to divide the raw datasets into the normal group (N group) and the similar/complex background group (S/C group). At last, 92 high quality and consistently labelled images are selected into the S/C group, and the rest are divided into the N group. The precision-recall curves of

TABLE 1: The performance of different algorithms tested on three groups.

Group	Algorithms	Precision
N group	GS	0.68
	MC	0.79
	MR	0.85
	RBD	0.81
	ACSD	0.84
	MSD	0.82
S/C group	GS	0.49
	MC	0.50
	MR	0.63
	RBD	0.56
	ACSD	0.87
	MSD	0.95
Full datasets	GS	0.64
	MC	0.72
	MR	0.80
	RBD	0.76
	ACSD	0.81
	MSD	0.86

different algorithms tested on the N group, S/C group, and full datasets are given in Figure 4. The performance of different algorithms tested on three groups is given in Table 1.

The proposed method works within a complexity of  $O(N)$ , and the evaluation on the results of these saliency detection algorithms in the S/C group shows that our algorithm has a better performance than other algorithms. In full datasets, it also performs well. By selecting the salient subset for further processing, the complexity of higher visual analysis can be reduced significantly. Many applications benefit from saliency analysis such as object segmentation, image classification, and image/video retargeting.

## 8. Conclusions

A new framework based on visual bionics for saliency detection under similar or complex background interference is proposed in this paper: First, we combine the depth map with the RGB map, and colour, texture, and depth information are introduced as the basis of superpixel segmentation. Second, the colour and depth information were calculated as two feature channels of saliency map. Finally, information entropy is used to calculate the weights of two channels, respectively, and get the final fused saliency map. The proposed method works within a complexity of  $O(N)$ , and the experimental results show that our saliency detection framework greatly reduces the error detection under similar and complex background and improves the overall saliency detection performance.

## Data Availability

The NJU400 datasets used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

## References

- [1] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 185–207, 2013.
- [2] W. Wang, J. Shen, L. Shao, and F. Porikli, "Correspondence driven saliency transfer," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5025–5034, 2016.
- [3] M. W. Jian, J. Y. Dong, and J. Ma, "Image retrieval using wavelet-based salient regions," *The Imaging Science Journal*, vol. 59, no. 4, pp. 219–231, 2013.
- [4] F. Xin, D. Yang, and Z. Ling, "Saliency variation-based quality assessment for packet-loss-impaired videos," *Acta Automatica Sinica*, vol. 37, no. 11, pp. 1322–1331, 2011.
- [5] R. Gupta and S. Chaudhury, "A scheme for attentional video compression," in *Proceedings of the 4th International Conference on Pattern Recognition and Machine Intelligence*, pp. 458–465, IEEE, Moscow, Russia, 2011.
- [6] Chenlei Guo and Liming Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 185–198, 2010.
- [7] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [8] Z.-P. Hu and M. Peng-Quan, "Graph presentation random walk salient object detection algorithm based on global isolation and local homogeneity," *Acta Automatica Sinica*, vol. 37, no. 10, pp. 1279–1284, 2011.
- [9] M. M. Cheng, N. J. Mitra, X. L. Huang, P. H. S. Torr, and S. M. Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015.
- [10] T. Yong, L. Yang, and D. Liang-Liang, "Image cell-based saliency detection via colour contrast and distribution," *Acta Automatica Sinica*, vol. 39, no. 10, pp. 1632–1641, 2013.
- [11] Y.-C. Guo, H.-J. Yuan, and P. Wu, "Image saliency detection based on local and regional features," *Acta Automatica Sinica*, vol. 39, no. 8, pp. 1214–1224, 2013.
- [12] W. Xu and T. Zhen-Min, "Exploiting hierarchical prior estimation for salient object detection," *Acta Automatica Sinica*, vol. 41, no. 4, pp. 799–812, 2015.
- [13] T. Liu, Z. Yuan, J. Sun et al., "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353–367, 2011.
- [14] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Computer Vision – ECCV 2012*, pp. 29–42, Springer, Berlin, Heidelberg, 2012.
- [15] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Columbus, Ohio, USA, 2014.
- [16] X. H. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 853–860, Providence, RI, USA, 2012.
- [17] H. Z. Jiang, J. D. Wang, Z. J. Yuan, T. Liu, N. N. Zheng, and S. P. Li, "Automatic salient object segmentation based on context and shape prior," in *Proceedings of 2011 British Machine Vision Conference*, pp. 110.1–111.10, Dundee, UK, 2011, BMVA Press.
- [18] C. Yang, L. H. Zhang, H. C. Lu, X. Ruan, and M. H. Yang, "Saliency detection via graph-based manifold ranking," in *Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3166–3173, Portland, OR, USA, 2013.
- [19] B. Jiang, L. Zhang, and H. Lu, "Saliency detection via absorbing markov chain," in *Proceedings of the IEEE international conference on computer vision*, Sydney, Australia, 2013.
- [20] K. Desingh, K. M. Krishna, D. Rajan, and C. V. Jawahar, "Depth really matters: improving visual salient region detection with depth," in *Proceedings of 2013 British Machine Vision Conference*, pp. 98.1–98.98, Bristol, England, 2013, BMVA Press.
- [21] Y. Z. Niu, Y. J. Geng, X. Q. Li, and F. Liu, "Leveraging stereopsis for saliency analysis," in *Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 454–461, Providence, RI, USA: IEEE, 2012.
- [22] R. Ju, L. Ge, W. J. Geng, T. W. Ren, and G. S. Wu, "Depth saliency based on anisotropic center-surround difference," in *Proceedings of 2014 IEEE International Conference on Image Processing*, pp. 1115–1119, Pairs, France: IEEE, 2014.

- [23] J. Q. Ren, X. J. Gong, L. Yu, W. H. Zhou, and M. Y. Yang, "Exploiting global priors for RGB-D saliency detection," in *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 25–32, Boston, MA, USA: IEEE, 2015.
- [24] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, *SLIC superpixels*, EPFL, Switzerland, 2010.
- [25] S. Sural, G. Qian, and S. Pramanik, "Segmentation and histogram generation using the HSV colour space for image retrieval," in *Proceedings. International Conference on Image Processing*, pp. 589–592, Rochester, NY, USA, September 2002.