Journal of Information Hiding and Multimedia Signal Processing ©2018 ISSN 2073-4212
Ubiquitous International Volume **9**, Number **4**, July **2018**

An Improved Two-steps Saliency Detection Algorithm based on Binarized Normed Gradients and Nuclear Norm Model in Video Sequences

Xiuyan Tian[1,2], Haifang Li[1], Hongxia Deng[1]

[1]College of Information and Computer,
Taiyuan University of Technology, 030600, China

[2]Department of Economy & Management,
Yuncheng University, 044000, China

Received June, 2017; revised January, 2018

Abstract. *Due to the Binarized Normed Gradients(BING) based saliency region detection suffers from low detection precision and high error detection rate, a novel improved image saliency regions detection method is proposed in this paper, which combines Binarized Normed Gradients for coarse location and low-rank decomposition model for fine selection. Firstly, Histogram Threshold(HT) method is adopted to adaptively compute the optimal threshold value so as to reduce boundary detection error; Secondly, on the basis of candidate saliency regions detected by gradients boxes for coarse location, robust principal component analysis method is used to obtain the low-rank component for building a background model, and eliminate the background regions by background differential method. Qualitative analysis shows that our proposed method in this paper has obvious advantages in the precision and recall rate compared with other existing methods , which demonstrates that our method is a highly efficient and reliable saliency regions detection.*
Keywords: Saliency region detection; Binarized Normed Gradients; Low-rank model; Precision and recall; Coarse location; Fine selection.

1. **Introduction.** Saliency detection is an important basic research in the fields of image analysis and computer vision, and is widely used in many application, such as object recognition and image retrieval. An image has often the larger background area, while the interested saliency region only account for a sub-fraction. Therefore, saliency region detection can narrow the processing area in the subsequent action, reduce the background interference, and improve the efficiency of image analysis and understanding [1-3].

Essentially there are two types of saliency detection method: threshold segmentation [4-7] and multi-proposal voting [8-9]. Threshold segmentation focus on finding out the area with the most visible difference, which is usually used in early algorithm. In order to improve the applicability for different types of image and integrity of the results, a saliency detection algorithm is proposed by Russell et al[3]. It combines the adaptive threshold merging with a new background selection strategy. Saliency region detection algorithm based on seed segmentation firstly selects multiple seed regions from an image, and then starts clustering from the seed region, and separates the saliency region on the basis of color, edge, texture and so on. Ref [7] proposed a novel seed-based graph cut segmentation, which generates a series of seed segmentation region and combines the structural learning method to obtain the saliency region. The advantage of these threshold segmentation is that the detection accuracy of saliency region is high, but there is a large

amount of calculation. In addition, the segmentation effect is greatly affected by the position of seed region. If the seed point is not properly selected, the segmentation effect will be significantly reduced. Multi-proposal voting method mainly adopts traversal with sliding window, where the salience of each window is voted on basis of its different image content and finally uses voting score to detect saliency region. The goal of generating object proposals is to create a relatively small set of candidate bounding boxes that cover the objects in the image. Nowadays, multi-proposal voting is more widely used because of its strong robustness to the illumination of the image, the contrast of the local contrast, and it has a good detection effect to saliency region with big difference for color. In [8], the gradient and closed boundary features in the image are extracted, and the Bayesian framework is used to describe these features, where the bounding box of the sliding window is voted to detect the saliency region of the image according to the voting score. In 2014, Cheng et al. proposed an BING boxes method that quickly locates object region with high precision and is a novel method for generating object bounding box proposals using edges[9]. The method does not use the machine learning algorithm to perform a saliency learning for object region, but rather a traditional image processing. Firstly, the edge response of each pixel in the image is calculated, and then the edges are grouped according to the similarity of the edge direction. Affinities are computed between edge groups based on their relative positions and orientations such that groups forming long continuous contours have high affinity. The affinity of the edge group is taken as the measure strategy, thus the saliency region in the image is detected by the edge voting. Since the computational efficiency of the edge is very high, the method can quickly detect saliency region, which is very valuable in the field of large-data processing (such as image retrieval). The experimental results show that BING method[9] can accelerate the detection efficiency, and can determine the location of the region with high precision. Therefore, it is an important algorithm in the field of saliency region detection.

Although the detection efficiency of the saliency region detection method based on the gradient box described in [9] is very high, there is a problem that the recall rate and the accuracy are low for saliency region. In this paper, we improve the method by using the Histogram Threshold method to adaptively calculate the optimal threshold of the edge magnitude, which replaces the fixed threshold method used in [9] so as to reduce the detection error of boundary point. Two-level detection method is adopted in paper, where saliency region based on the gradient box is used as the coarse localization region, then fine selection is performed on that basis. The low-rank background model is constructed by using the robust principal component analysis method, and the background region is eliminated by the background difference, which reduces the false detection of saliency region. According to these improvement, the objective is to improve the recall rate and accuracy of saliency region detection in the case of ensuring the high efficiency.

The rest of our paper is organized as follows: the flowchart description of our proposed saliency region detection algorithm is detailed in Section 2. Section 3 shows the results of the saliency region detection under different images, and the analysis is given qualitatively and quantitatively. The conclusion and the next work will be described in last Section.

2. **Our proposed Saliency Region Detection.**

A novel salient regions detection method is proposed in this paper, which combines Binarized Normed Gradient boxes for coarse localization with low-rank model for fine selection. Its basic framework is described in Fig 1.

Firstly, it uses edge boxes detection method proposed in [9], to realize coarse localization for the saliency regions; in order to reduce boundary detection error, the selection of the

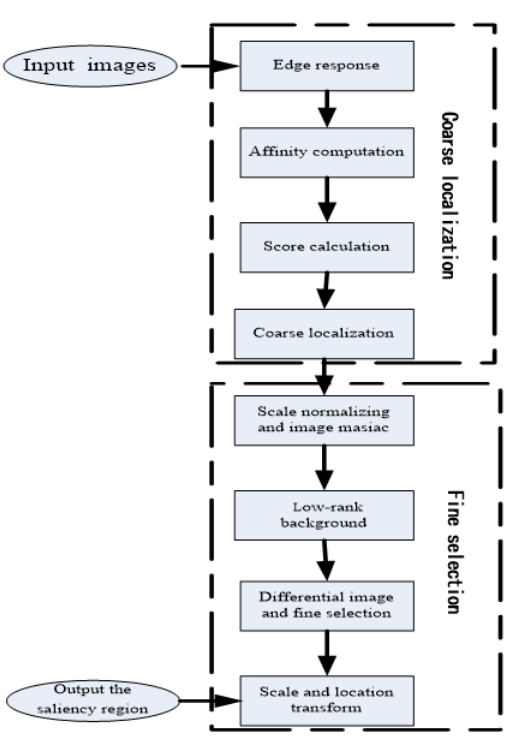

FIGURE 1. Basic framework

detection boundary threshold is improved by adopting Histogram Threshold method so as to adopt adaptively the optimal threshold; Secondly, the candidate saliency region from coarse localization is proposed by scale normalizing and image mosaic,then robust principal component analysis is adopted to get the low-rank component of the spliced image so as to build background model, thus the background difference method is used to achieve fine selection of salient region, which can solve the problems with high error detection and low detection precision in saliency region proposed by [9].

2.1. **Binarized Normed Gradient method for coarse localization.** In [9], an image saliency regions detection method based on bounding boxes is proposed. it is assumed that the saliency object is complete in the image, so that the bounding box of the saliency object is surely inside the image. Based on this assumption, this method traverses every different scale window in the image to find the internal boundary point, where it uses the results of voting from the edge groups to determine whether the content of the window is salient, so as to finalize the location of the saliency region. In addition, the method supposes that the edge information can best describe the object, and can be used to extract the candidate region with high quality. Compared with the traditional multi-scale sliding window scanning strategy, the edge boxes method can eliminate redundant windows to get a few candidate windows with high quality. Therefore, the outstanding advantage of Binarized Normed Gradient method is to detect very fast with high quality. The saliency regions detection based on BING method consists of four parts, which are briefly summarized as follows.

(a) Edge response computation For an image, the edge response of each pixel should be calculated firstly. The edge response is obtained by the structured edge detector that is described in [10]. The edge detector can effectively detect the object boundary and its computational efficiency is very high. After the edge response is obtained, non-maximum suppression is used on the obtained edge response to search for the boundary peak. In this

way, each pixel can obtain an edge magnitude value and the direction . In [9], the pixel with the edge magnitude value greater than 0.1 is defined as the boundary point, and the boundary curve formed by connecting the boundary points is defined as the contour curve. However, using fixed threshold to search boundary points is of high efficiency, but it has false boundary points detection and missing boundary points detection. Especially for the images under different illumination, the image boundary point detection accuracy is not stable. In order to improve the accuracy for image boundary point detection and the robustness for different illumination, Histogram Threshold method is used to calculate the optimal segmentation threshold of the edge magnitude adaptively. In particular, linear transform method is used to convert edge magnitude values of the pixel within the integer range of [0,99], the edge magnitude values of all pixels constitute a grayscale image in 100 levels of gray. Then, Histogram Threshold-based adaptive image segmentation algorithm is used to calculate the optimal segmentation threshold, the details is described in [11]. Finally, the pixels whose edge magnitude is greater than the adaptive threshold are defined as the boundary points. In this way, the segmentation threshold obtained by the adaptive algorithm has stronger robustness to images under different illumination, and the boundary points detection has higher precision.

(b) Affinity computation In order to improve operation efficiency, the boundary points are grouped, and the affinity of each edge group is calculated as the basis for voting. The construction method of edge groups is as follows: the greedy search algorithm is used to combine the boundary points in neighborhood 8, it wont stop until the direction difference sum of boundary points in the edge group exceeds a threshold T (T = $\pi/2$). After the edge group is obtained, the affinity of the edge group is calculated by the average position and average direction of the boundary points in the edge group, denoted as follows:

$$a(s_i, s_j) = |\cos(\theta_i - \theta_{ij})\cos(\theta_j - \theta_{ij})|^{\gamma} \tag{1}$$

where, S_j and S_J represent two edge groups respectively, θ_i and θ_i represent the average direction of boundary points for and respectively, θ_{ij} is the angle between the average position x_i of edge group s_i and the average position x_j of edge group x_j . γ is used to moderate the sensitivity of affinity along with the change of direction, where γ is equal to 2 in this paper. Intuitively, if the angle between these average positions of two edge groups is similar to the direction of these two edge groups, these two edge groups have greater affinity. If these two edge groups can be separated by more than two pixels, their affinity is set to zero. In order to improve computational efficiency, only the affinity that exceeds a fixed thresh T2 will be stored (T2 is valued as 0.05), the other affinities are set to zero.

(c) Voting score calculation For the voting to bounding box b from edge group s_i, the score of voting can be represented as:

$$h_b = \frac{\sum m_i \omega_b(s_i)}{2(b_w + b_h)^{\kappa}} \tag{2}$$

where b_w and $b_h$represent the width and the height of the bounding box respectively. It is noted that, the divisor in the above equation is the length of the bounding box rather than the area, it is because no matter how much the scale is, every edge has the width of a single pixel. Nevertheless, in order to avoid the deviation caused by excessive amount of edge points in big windows, κ is set to 1.5 for restriction. $m_i$is the accumulative sum of magnitude values of all the boundary points in the edge group s_i. $w_b(s_i)$ is the weight

of the edge group s_i to the bounding box b, which is denoted as :

$$\omega_b(s_i) = \begin{cases} 1, & s_i \in b \\ 1 - \max\limits_{T} \prod\limits_{j}^{|T|-1} a(t_j, t_{j-1}), x_i \in b \ \ \& \ \ s_i \notin S_b \\ 0, & else \end{cases} \tag{3}$$

where S_b represents the a collection of all edge groups that overlap the boundary of the box $b; T$ represents an orderly path for the edge group whose length is |T|with the starting point $t_1 \in S_b$ and the ending point $t_{|T|} = s_i$. If such path is not existed, $\omega_b(s_i)$=1. Equation 3 is to search for the path with the highest affinity among the edge group and the edge groups that overlap with the boundary of the bounding box. Since the affinity of all paired edge groups is set to zero, so the calculation in this process is highly efficient. Given that the closer the edge groups are to the bounding box boundary, the greater contribution the edge groups will have, the bounding box b^{in} on the center, need subtract a boundary amplitude value, which is denoted as:

$$h_b^{in} = h_b - \frac{\sum_{p \in b^{in}} m_p}{2(b_w + b_h)^{\kappa}} \tag{4}$$

where the width and the height of b^{in} are $b_w/2$ and $b_h/2$, respectively.

(d)Coarse localization To locate the image saliency regions, the sliding window search strategy is used in this paper, where it traverses all the candidate bounding boxes with different locations, different scales and different pixel aspect ratios, to calculate the voting scores. The translation scale is the same as the step length of the pixel aspect ratio α, whose value is 0.65. After the sliding window pass, the greedy iterative search strategy is adopted to search for the largest h_b^{in} on different locations, scales, and pixel aspect ratios. At the end of each iteration, the search step is halved. When the translation step is less than 2 pixels, the search terminates. Then, the sorting operation is based on the voting score, the location of the bounding box of which the maximum voting score exceeds 0.5, and the corresponding score is stored. Finally, non-maximum suppression is carried out on the ordered bounding boxes. In particular, if IoU of a bounding box is β times as large as IoU of the bounding box that has higher voting score, the bounding box with lower voting score should be deleted. The empiric value of β is set to 0.55. Thus, IoU refers to the ratio of the overlap region between the candidate box (Detection Result) and the bounding box (Ground Truth) with the object truth value, to the total area of both bounding boxes. When IoU is greater than the threshold T_{IoU}, the bounding box detection is considered to be correct. After non-maximum suppression, the locations of remaining bounding boxes are the saliency regions determined by coarse localization.

2.2. **Low-rank background model for fine selection.** Let's assume that the number of bounding boxes obtained from coarse localization is N, and scale normalized and image mosaic are carried out for image patches contained in these N bounding boxes; then, robust principal component analysis method is used to construct the image low-rank background model after image mosaic; and then differential method is used on the image mosaic and the background model to select saliency regions with more precision; finally, convert the bounding boxes in the selected saliency regions to the scale and location of the original image, and obtain the final detection result for the saliency regions. The details are described as follows:

(a) Scale normalized and and image mosaic The size of the image patch is $w_i \times h_i$. In this paper, the height of these N image patches is normalized to $H(H = 30)$. Similarly, the width of image patches is normalized with the same scale. In this way, after

scale normalized, the size of the image patch is $\left(\mathrm{Int}\left(\frac{H}{h_i}\times w_i\right)\right)\times H$, whereInt$(\cdot)$ denotes rounding operation. Then, N image patches with the same height are concatenated together along the horizontal direction so as to obtain the image mosaic , whose size is $\left(\sum_{\mathrm{i}=1}^{N}\mathrm{Int}\left(\frac{H}{h_i}\times w_i\right)\right)\times H$.

(b) Low-rank background construction As for mosaic image , it consists not only of saliency region, but also the background region, thus robust principal component analysis can be adopted to regard the image matrix as the interference matrix, which can be decomposed into the sum of a low rank matrix and a sparse matrix, as shown in Eq.(5).

$$F = L + S \tag{5}$$

where L is the low-rank matrix denoted as the background region, whileS is the sparse matrix representing the saliency region. The low-rank decomposition process can be mathematically described by optimization method, denoted as follows:

$$\begin{aligned} &\min_{L,S} \quad \mathrm{rank}(L) + \lambda\|S\|_0 \\ &\text{s.t} \quad F = L + S \end{aligned} \tag{6}$$

where rank $(\cdot)$ and $\|\cdot\|_0$ are denoted as the rank of matrix and L_0 norm, respectively;λ is a regularization parameterwhose value is written as

$$\lambda = \frac{1}{\sqrt{\max\left(\left(\sum_{i=1}^{N}\mathrm{Int}\left(\frac{H}{h_i}\times w_i\right)\right), H\right)}} \tag{7}$$

where $\max(\cdot,\cdot)$ is the maximum operation. In Eq.(6), it is difficult to solve the rank calculation and L_0norm minimization problem. The rank$(L)$is replaced with the nuclear norm $\|\mathrm{L}\|_* = \sum_i \sigma_i$ $\|\mathrm{A}\|_* = \sum_i \sigma_i$ where $\sigma_i$is the singular value of data matrix L, and L_0 norm is replaced with L_1 norm. Simple replacement yields a tractable optimization problem:

$$\begin{aligned} &\min_{L,S} \quad \|L\|_* + \lambda\|S\|_1 \\ &\text{s.t} \quad F = L + S \end{aligned} \tag{8}$$

where $\|\|_*$,$\|\|_1$ are denoted as the nuclear norm and L_1 norm, respectively. We use the non-exact Augmented Lagrangian multiplication described in [12] to solve the robust principal component analysis problem in Eq. (8) and restore the low rank matrix as the background image. See [12] for detailed information.

(c) Differential image and fine selection According to above description, differential image between mosaic imageFand low-rank background F_0can be written as

$$F_d(x,y) = \left| F(x,y) - \frac{1}{9}\sum_{i=-1}^{1}\sum_{i=-1}^{1} F_0(x+i, y+j) \right| \tag{9}$$

$F_d(x,y)$ is the value of the pixel (x,y) in the differential image. In the process of difference, each pixel in the background image is valued the mean of Neighbourhood 8, so as to reduce the noise interference caused by the reconstruction of the low-rank background image. Then, the Histogram Threshold method is used to obtain the optimal segmentation threshold T_{opt} for the differential image adaptively. When the value of the pixel in the differential image is greater than T_{opt} , the corresponding pixel point is determined to be the foreground with a value of 255; Otherwise, it is determined to be the background

with a value of 0. Next, the top-hat transform in mathematical morphology is used to filter binary images after segmentation so as to remove the noise and fill the hole. Finally, connectivity scanning of neighborhood 8 is adopted to obtain the size and the location of bounding boxes in the foreground region, thus the remaining bounding boxes are the saliency regions after fine selection. Here, the size of the bounding boxes in the saliency regions, the sequence number and the relative location in the original bounding box should be recorded.

(d) Scale and location transform For each remaining bounding box in the saliency regions after fine selection, according to the scale transform ratio used in scale normalized for corresponding original bounding box, and the location of this bounding box, as well as the relative location relationship between the bounding box and the original bounding box, recover the scale and the location of the bounding box in the saliency region . In this way, through the coarse localization by edge boxes and the fine selection by low-rank background model, false detection and missing detection problems in [9] can be reduced. The concrete details will be discussed in experiment.

3. **Experiments and Analysis.** In order to validate the performance of our proposed saliency region detection, it will be compared with the existing saliency methods described in references [7], [8] and [9], so as to qualitatively and quantitatively evaluate its performance for these methods. All parameters in comparative algorithms are fixed for all experiments to demonstrate the robustness and stability of our proposed method. We will firstly introduce the selected data-set in the experiment, then elaborate the quantitative evaluation index and the qualitative analysis for the performance, and finally show the results of the comparison experiment.

3.1. **The data sets.** The PASCAL VOC 2007 used by reference [9] is still adopted in this experiment. It is also the commonly used data-set for saliency test. The data set contains 9963 images, where the number of saliency regions is different in each image, but all saliency regions have been manually labeled; In order to compare the results with those of reference 8, the data-set of MSRA-1000 is also used in this experiment for saliency regions detection. In this experiment, the value of IoU between the bounding boxes obtained from the calculation and the bounding boxes from the manually labeled saliency regions, can be used to determine whether the detection result is correct. For each detected saliency region, if its IoU value is greater than the threshold T_{IoU}, the detection result for this saliency region is considered to be correct. In reference 9, T_{IoU} in the experiment is set to 0.7, Statistically, T_{IoU} is valued as 0.65, 0.7, 0.75, 0.8 and 0.85 in this experiment to test the performance of saliency region detection, respectively so as to analyze the performance of our proposed method.

3.2. **Performance evaluation index.** The true usefulness of a saliency map is determined by the application, and the saliency maps are evaluated in the context of salient object segmentation. To quantitatively evaluate the performance of the proposed method, the precision and accuracy are computed to quantitatively evaluate the performance. The precision and recall are defined as:

$$\mathrm{Pr}\,ecision = \frac{TP}{TP + FP} \times 100 \tag{10}$$

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \times 100 \tag{11}$$

where TP(True Positives), FP(False Positives), and FN (False Negatives) denote the number of correctly classified object pixels, the number of background pixels but classified

as object, and the number of object pixels but classified as background respectively. Obviously, the higher the precision and recall are, the better the method is. In addition, in order to evaluate the detection efficiency of various methods, Frame Rate is also adopted in this paper, which specifically refers to the average number of images being processed per second, in fps. Considering that the detection efficiency is closely related to the software and the hardware platform where the algorithm is in operation, the same software and hardware platform and the same data set for test are used in the comparative tests for the efficiency of various detection methods. Thus, All of the experiments are run under Visual Studio 2013 on PCs with an Intel quad-core i5 CPU at 3.20 GHz and 16GB memory.

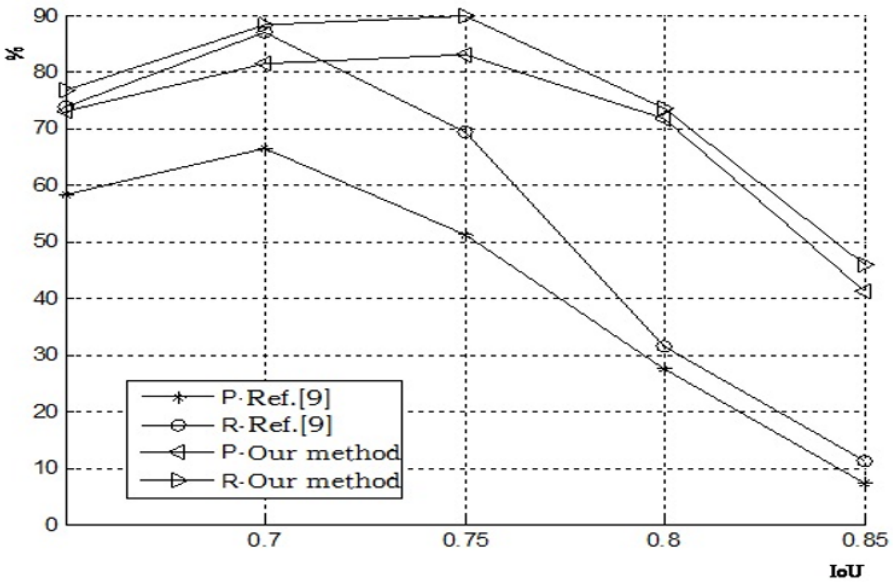

FIGURE 2. Detection index comparison for different T_{IoU}

3.3. **Visual comparisons.** In order to illustrate the detection advantages of our proposed algorithm, the visual effect is compared for some complex scenes detection, as shown in Fig.2 and Fig.3. Since the reference [9] adopts image processing based on edge information to detect salient region instead of machine learning algorithms, then the edges are grouped based on the criteria of similar direction, and the affinity of the edge group is taken as the measure for voting score in box, the salient regions for the test image are detected by the BING voting. Therefore, the proposed method has significant difference with the image segmentation based saliency detection. For the fairness and justice of performance comparison among different algorithms, the results of algorithms are divided into two categories. In this paper, the coarse localization is firstly used for saliency detection, then the low-rank background model is adopted to select the saliency regions with higher precision. Fig 2 shows the comparison results of our algorithm with references [7] and [8] for saliency regions segmentation. It shows that the the low-rank background image obtained by robust principal component analysis method and image masiac, has contained most of the background information, and the saliency regions that are segmented by image deferential have been already close to the ground-truth regions. And the segmentation effect of reference [7] is greatly related to the seeded region. If the seed regions were chosen incorrectly, the segmentation effect would be greatly reduced, and the excessive segmentation will occur, as shown in Fig 2 (b). In reference [8], the gradients and closed boundary features are used for region segmentation, but the edge is not

dealt with in this method, which makes so much background information as the object, as shown in Fig 2 (c). Therefore, our proposed algorithm for saliency regions detection has a great detection effect. Fig 3 shows that our proposed method can obtain the best saliency region. It is worthwhile to be noted that, although the colors of the foreground and the background are very similar, the reasonable saliency image can also be generated by the proposed method in this paper. However, other methods can only detect parts of the object, and many corresponding background regions are also detected at the same time. The experimental results show that the saliency regions obtained by our proposed method in this paper have more overlap with ground-truth regions, and less missing detection than other methods. The comparison methods have been found to have failed to detect saliency regions. By analyzing some comparison results, it can be found that when the background is complex and has uneven distribution, some background regions are detected as saliency objects. In addition, the algorithm is difficult to distinguish the object when the color of saliency region is in line with that of the background. Qualitative analysis shows that our proposed method for salient regions detection greatly reduces the false salient regions detection and improves the detection precision through the coarse location by edge boxes and the fine selection by the low-rank background model.

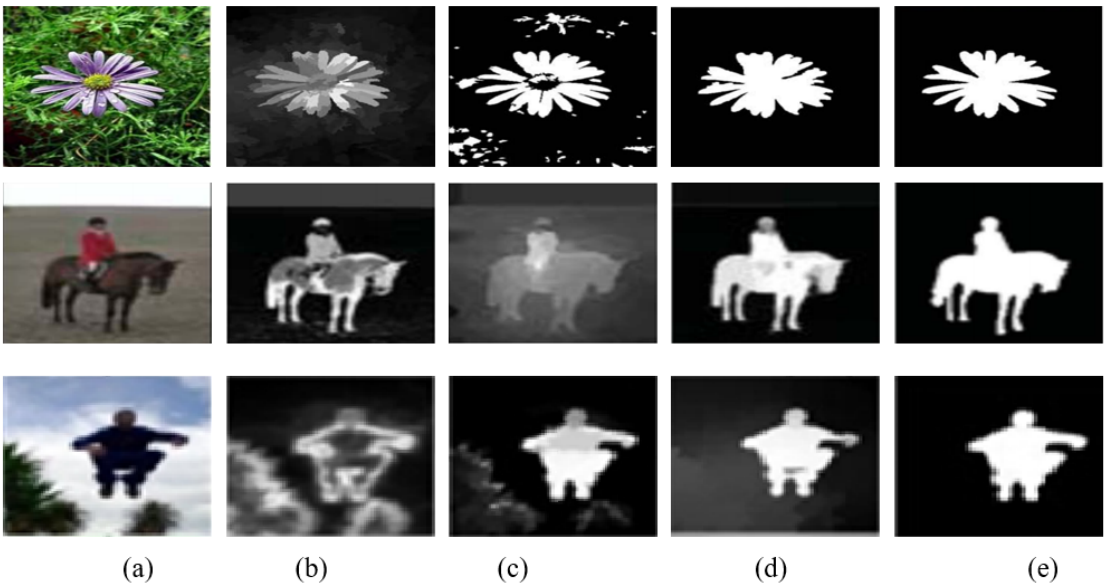

FIGURE 3. Visual comparisons. (a) Original image; (b) Result in[7](c) Result in[8] ;(d)Our method; (e) Ground-truth regions

3.4. **Quantitative comparison.** In order to exactly evaluate the performance results, compare firstly our proposed method with reference [9] as T_{IoU} is given different values, where T_{IoU} is valued as 0.65, 0.7, 0.75, 0.8 and 0.85, respectively. The corresponding precision and recall are as shown in Fig 4. With the increase of T_{IoU}, the recall rate of reference [9] declines significantly ,while the decline of recall rate is not significant in our proposed method, as shown in Fig 4.This illustrates that the bounding box of saliency region detected by our proposed algorithm has higher precision. The main reason is that the adaptive threshold is adopted for extracting the boundary points instead of the fixed threshold in our method, which can reduce the boundary point extraction error. In addition, we perform further fine selection with the low-rank background model on the bounding boxes detected by [9], so that some false detection regions from coarse location can be eliminated, especially the over-sized bounding box detected by coarse location when

the object contour and the background region exist intersection. In addition, whatever the value is from this precision perspective, the precision index for our proposed method are all higher than the reference [9]. The main reason is our method use two-level saliency detection, which obtain further fine selection with the low-rank background model on the bounding boxes detected by [9], so that the false detection error is reduced. According to Fig 4, the value of $T_{IoU}$is set to 0.75 in this paper. Under the condition of $T_{IoU} = 0.75$, the detection index for saliency regions is analyzed statistically in our proposed method and Refs [7], [8], [9] . Fig 5 shows the comparison results of the precision (P) and the recall rate (R), and the comparison results of Frame Rate (FR) are given in Table 1.

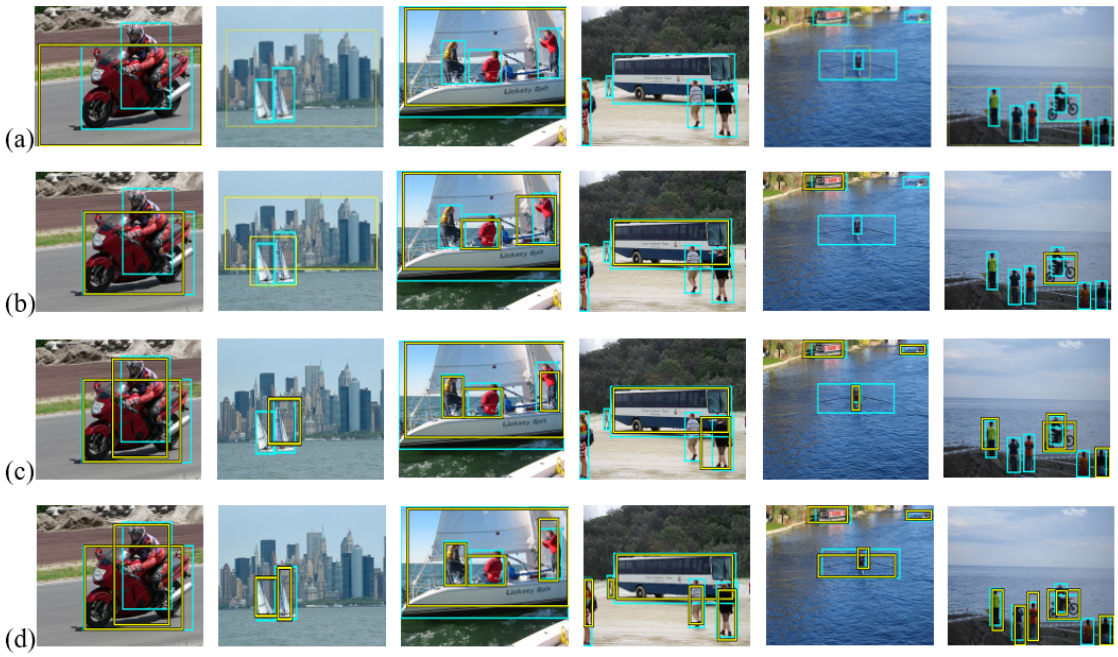

FIGURE 4. Saliency detection for different algorithms in video sequences (blue for Ground-truth region, yellow for the detected region) (a)Result in[7]; (b)Result in[8]; (c)Result in[9]; (d) Our detection result

As shown in Figure 5, the Precision and Recall indexes are higher than those of the other three methods, especially the precision index. This is because our proposed method combines two detection stages including edge boxes for coarse location and low-rank background model for fine selection, to reduce the phenomenon of false positives, so as to increase the detection precision. Among these three compared methods, although the Precision and Recall indexes in references [7] and [8] are higher than those of Literature [9], the recall rate (R) of these two methods are much lower than that of [9]. Although the frame rate of the proposed method is also a little lower than that of [9], the difference is not significant, and the Precision and Recall indexes of the proposed method are much higher than those of [9]. Therefore, our proposed method in this paper has better detection performance for saliency regions than the other three methods .

4. **Conclusions.** Aiming at the problem that the saliency detection method in [9] suffers from low detection precision and high error detection rate, a novel image saliency regions detection method is proposed, which combines two detection stages including BING step for coarse location and low-rank background model for fine selection. Compared with the Ref. [9], the improvement of our method is mainly in two aspects. Firstly, the proposed

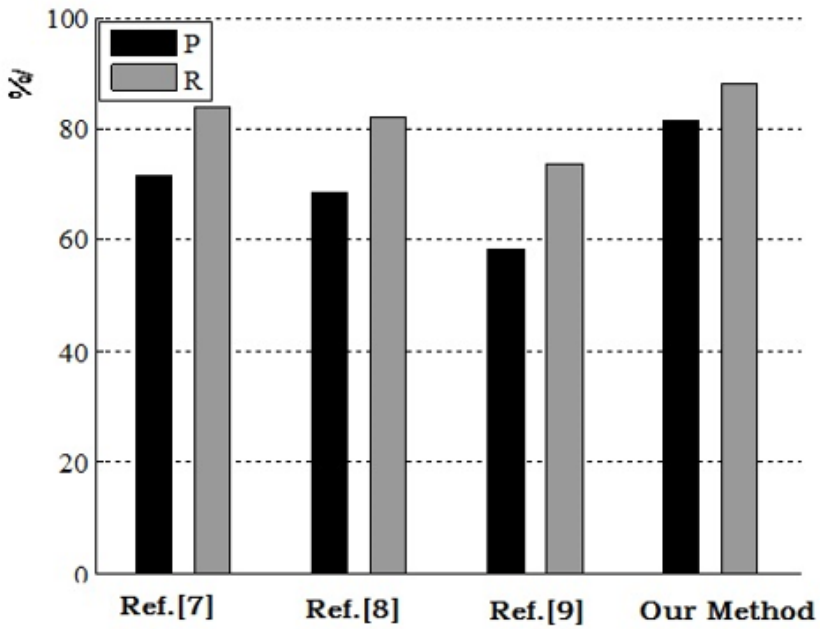

FIGURE 5. Index comparison for different methods

method makes improvement on threshold selection for boundary points detection, which uses Histogram Threshold method for adaptively computing the optimal threshold value to reduce boundary detection error. Secondly, on the basis of candidate saliency regions detected by BING for coarse location, it uses robust principal component analysis method to obtain the low-rank component for building a background model, so as to eliminate the background regions based on background differential method. Qualitative analysis shows that our proposed method in this paper has obvious advantages in the precision and recall rate compared with other contrast methods , which shows that the method is a highly efficient and reliable saliency regions detection.

Acknowledgments. This work was supported by the National Natural Science Foundation of Shanxi Province (No.2014021022-5), and the Technological Project of State Grid Corporation of China (No.5205301500).

REFERENCES

[1] M. M. Cheng, N. J. Mitra, X. Huang, et al. Global Contrast Based Salient Region Detection, *Journal of IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 37, no. 3, pp. 569-582, 2015.

[2] C. Rother, T. Minka, A. Blake, and V. Kolmogorov. Cosegmentation of image pairs by histogram matching - incorporating a global constraint into mrfs. In CVPR, 2016.

[3] B. Russell, A. Efros, J. Sivic, W. Freeman, and A. Zisserman. Using multiple segmentations to discover objects and their extent in image collections. In CVPR, 2013.

[4] P. Rantalankila, J. Kannala, E. Rahtu, Generating Object Segmentation Proposals Using Global and Local Search, *IEEE Conference on Computer Vision and Pattern Recognition.*, pp. 2417 - 2424. 2014

[5] F. Schaffalitzky and A. Zisserman. Viewpoint invariant texture matching and wide baseline stereo. In ICCV, 2011.

[6] J. Carreira, C. Sminchisescu. Cpmc: Automatic object segmentation using constrained parametric min-cuts *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.34, no. 7, pp. 1312-1328, 2012.

[7] I. Endres, D. Hoiem, Category-Independent Object Proposals with Diverse Ranking, *Journal of IEEE Transactions on Pattern Analysis & Machine Intelligence*, 36, no. 2, pp. 222-34, 2014.
[8] B Alexe, T Deselaers, V Ferrari. Measuring the Objectness of Image Windows, *Journal of IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 34, no. 11, pp.2189-2202, 2012.
[9] M. Cheng, Z. Zhang, W. Lin, et al. BING: Binarized Normed Gradients for Objectness Estimation at 300fps, *computer vision and pattern recognition:* 3286-3293, 2014..
[10] P Dollr, C L Zitnick. Structured forests for fast edge detection, *Proceedings of the IEEE International Conference on Computer Vision,*, pp. 1841-1848. 2013.
[11] D. N. Bhargava, A. Kumawat, D. R. Bhargava, Threshold and binarization for document image analysis using Histogram Thresholds Algorithm, *Journal of International Journal of Computer Trends & Technology*, vol.17, no. 5, pp. 272-275, 2014.
[12] N. B. Erichson, C. Donovan, Randomized low-rank Dynamic Mode Decomposition for motion detection, *Journal of Computer Vision and Image Understanding*, vol. 146, pp. 40-50, 2016.