CrossMark

ORIGINAL PAPER

Salient object detection via images frequency domain analyzing

Chao He[1] · Zhenxue Chen[1] · Chengyun Liu[1]

Abstract Salient object detection has become an important direction in image processing and computer vision. The traditional center-priori theory believes that salient target should be closer to the central area of the image. However, false detection will often occur when the salient object is closer to the image boundary. So, this paper obtains center coordinates of the salient object by using Harris corner detection algorithm and convex hull. Accordingly, an improved center-priori saliency detection model is obtained by applying the frequency-tuned method. And then, the local saliency is set up by wavelet transforming which has the local characteristic information representation ability in the time domain and frequency domain. In addition, we obtain the global saliency by spectral residual analyzing. Finally, an advanced center-priori saliency model is established. The experimental results show that the model in this paper has better detection effects and higher target detection rates.

1 Introduction

Saliency detection also plays important roles in many computer vision applications such as moving object detection [1–3], tracking [4–6], and action recognition [7,8]. In the late 1990s, Itti proposed biological visual attention mechanism and image saliency theory based on the biology for the first time [9]. Itti's model studied man's visual attention

✉ Zhenxue Chen
chenzhenxue@sdu.edu.cn

[1] School of Control Science and Engineering, Shandong University, Jinan 250061, China

transfer mechanism when people observe things in the nature. And it obtains saliency maps by using intensity, color, and direction channels of the input image [10]. Then, it fused all saliency maps and obtained ideal detection effect. In recent years, machine vision got great development and was applied to many fields of image processing, artificial intelligence, and others. At the same time, image saliency theory, as an important branch of machine vision, has made great progress. Many new detection methods explain saliency from the perspectives of biological vision, statistics, and information theory.

The paper is organized as follows. In Sect. 2, we briefly review the previous researches that have made a contribution to image saliency theories and target detection methods. Then, we carefully describe our method for image saliency target detection in Sects. 3, 4, 5, and 6. The results of our experiments could be found in Sects. 7 and 8. Finally, we summarize this paper in Sect. 9.

2 Previous related works

The existed saliency researches could be divided into two categories [11,12]: (1) the bottom–up model based on stimulating and data-driving. Itti's biological visual saliency model [13] detects salient object by extracting image's primary characteristics of colors, intensity, and directions. Then, feature maps were multi-scale analyzed by establishing a nine-layer Gaussian pyramid. Finally, feature maps were obtained by using center-around operator. They were fused to obtain finally saliency map. Harel's saliency detection was based on graph theory [14], and it used the characteristics of Markov random field to construct two-dimensional images of Markov chain on the basis of Itti model. Yang et al. [15] used background and foreground as prior knowledge and quadrilateral node as benchmark for the first round of sorting.

The signification maps were binary-transformed and was the second sort benchmark. Finally, saliency map was obtained, and the significant detection model based on graph sort was set up. Hou's spectral residual method [16] did not need to deal with the color, intensity, and other characteristics, but analyzed images logarithmic spectrum and calculated residual. Then, the final saliency maps were obtained by Fourier inverse transform. Achanta proposed saliency detection method [17] which was based on the images frequency analysis. (2) The top–down model was based on task and knowledge such as Maybank proposed image significant area matching method [18], and Yang and others obtained specific target by dictionary learning conditions and random testing method [19].

Currently, most of saliency models use early visual features to analysis images. This includes colors, texture, intensity, and directions. In this paper, we use the HSV color space to analyze visual color and intensity features. The texture map is obtained by using texture filter which can be found by a MATLAB function called *rangefilt*. The directions feature is obtained by wavelet transforming. Therefore, wavelet transform has the ability to extract the images' horizontal, vertical, and diagonal direction information.

In general, the salient object usually locates in the central field of the vision when people observe the surrounding scenery. Based on this, many papers put forward center-priori method to obtain saliency objects. But, in the reality captured images, the most significant things in the off-center position even appear in the image boundary. This leads to the traditional center-priori theory which is not consistent with actual image characteristics. So, this paper gets image's saliency points by using Harris [20] corner detection algorithm and obtains convex hull of the regional [21]. Then, the area within the convex hull is regarded as false foreground information and the outside is false background information. Finally, we could get an improved center-priori coefficient by using frequency-tuned method.

Shannon information coding theory [22] thinks that any information should contain important part and redundant part, and it is a key problem how to extract the effective information and remove redundant part in the field of modern information code. As a kind of complex information, images can be identified by people directly and divided into sensitive and insensitive parts which correspond to the significant and not-significant parts of images. Wavelet transform has good local information representation ability [23]. Especially in recent years, because of its ability of analysis on spatial and frequency domain in multi-scale transform at the same time, it is concerned by scholars [24–27] in the field of visual attention. Wavelet transform can effectively suppress the image's high-frequency information by analyzing images' local visual features (colors, texture, directions, intensity), and reserve the significant part. We can obtain local fea-

tures saliency maps by wavelet transforming. Hou's spectral residual theory would not analyze specific features. But to get global saliency map, it is necessary for image logarithmic transformation. The two methods have their advantages and disadvantages for different types of images. Therefore, this paper combines local features and global features to get saliency map. To enhance salient detection results, we use improved center-priori knowledge based on frequency-tuned method. The experimental results show that our method is better.

3 The improved center-priori coefficient based on frequency-tuned method

Traditional center-priori theories [28] think the salient object usually located in the center of vision, as shown in Fig. 1a. Therefore, the paper regards it as priori knowledge which is simulated by Gaussian function. Gaussian model can highlight center part and suppress surrounding, as shown in Fig. 2a.

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}} \tag{2}$$

Gaussian model can simulate the visual center-priori theory, but the image is two dimensional. Therefore, the two-dimensional Gaussian function is used to implement it, such as Formula (2). Gaussian model can match to the priori knowledge. It is shown in Fig. 2b.

In fact, the salient region usually deviates from center in many images, as shown in Fig. 1b. Traditional center-priori knowledge and actual situation have a huge deviation, which causes detection results to deteriorate. In response to this problem, the paper proposes Harris corner detection and convex hull theory to get image salient object center coordinate. Harris corner detection algorithm calculates grayscale image curvature distribution and the maximum curvature key angular point. Then, it calculates convex hull geometric center which is regard as the center of the salient object.

This paper gets key angular point and convex hull in original image, such as Fig. 3a. Then, we could obtain Fig. 3b

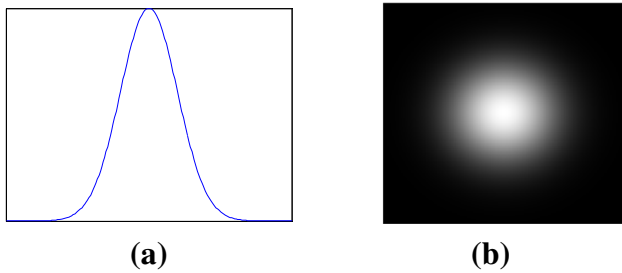
Fig. 1 Saliency target images

(a)   **(b)**

Fig. 2 Center-priori model

(a)   **(b)**   **(c)**

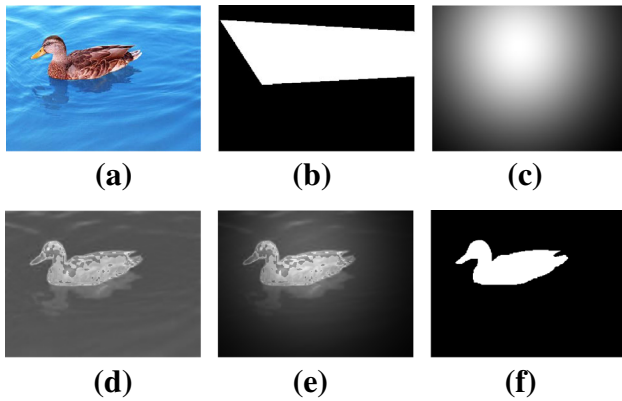**(d)**   **(e)**   **(f)**

Fig. 3 Improved center-priori coefficient

by binary image segmenting. In Fig. 3b, the white region is convex hull area which could be regarded as false foreground; the black region is false background. We establish a two-dimensional Gaussian model by computing the center coordinates of the convex hull. The model $G(x, y)$ is preliminary center-priori coefficient. It is mapped to the two-dimensional space in Fig. 3c. Figure 3f is artificial significant target segmentation map. Compared with Fig. 3b, f, we can find that the false foreground region includes a part of the background information, but in most images, the false background region contained the real background. Based on this, compared to the enter image, the false background region having more similar visual features to the real background region. Therefore, we compute the mean false background feature vector I_μ not the enter image and get frequency-tuned coefficient $S(x, y)$ by formula (3). $I_{\omega hc}$ is the corresponding image pixel vector value in the Gaussian blurred version (using a 5*5 separable binomial kernel) of the original image, and $\| \|$ is the L2 norm. The frequency-tuned coefficient is mapped to the two-dimensional space in Fig. 3d. Figure 3e is improved by center-priori coefficient $\Theta(x, y)$ based on frequency-tuned method. It can be obtained by formula(4).

$$S(x, y) = \left\| I_\mu - I_{\omega hc}(x, y) \right\| \tag{3}$$

$$\Theta(x, y) = e^{[g(x,y) \times S(x,y)]} \tag{4}$$

4 Local features saliency model based on wavelet transform

Wavelet transform can transform features image from spatial domain to frequency domain. And it can break images down into high- and low-frequency parts. The paper detects features of colors, intensity, textures, and direction in HSV color space. To get multi-scale feature image, low-frequency part of each layer needs a secondary decomposition [29]. Each low-frequency feature map is dealt by inverse wavelet transform to get local feature saliency maps. (In this paper, four-layer wavelet decomposition structure is established).

$$[A_N^c, H_S^c, V_S^c, D_S^c] = WT_N[I_c(x)] \tag{5}$$

The respective characteristics of image are obtained by Formula (5). Then, two-dimensional wavelet reconstruction algorithm is used to reconstruct feature maps in order to get feature saliency maps.

$$S_i(x) = g(x) * IWT[c(x) * A_S^c(x)] \tag{6}$$

where $g(x)$ is Gaussian filter; $c(x)$ is mean filter.

Finally, four kinds of feature saliency maps are fused by Formula (7) for final local feature saliency maps.

$$\text{LsMap} = g(x) * \left(\sum_{i=1}^{4} k_i \cdot S_i / \max\{S_i, 0.5\} \right) \tag{7}$$

5 Global features saliency model based on spectral residual

Hou's image spectral residual analysis method is based on the properties of natural image similarity. It is a statistical method of image saliency detection. It need not analyze the features. Its core idea is to remove redundant information and enhance regions of interest (ROI) in the image. This property is also known as $1/f$ law. It states that the amplitude $A(f)$ of the averaged Fourier spectrum of the ensemble of natural images obeys a distribution: $E\{A(f)\} \propto 1/f$. Hou found the average amplitude spectrum logarithmic was local linear by studying a lot of images. The spectral residual is obtained by calculating the difference in images' amplitude and average logarithm amplitude spectrum. Finally, the global saliency map is obtained by logarithmic inverse transformation of spectral residual [30]. Spectral residual calculation Formula:

$$R(f) = \log[A(f)] - h_n * \log[A(f)] \tag{8}$$

where h_n is mean filter.

The steps of Hou's saliency detection are shown as follows:

Step 1 The input images are Fourier-transformed, and their amplitudes $A(f)$ are calculated.

$$A(f) = \Re(F[I(x)]) \tag{9}$$

Step 2 Images' phase spectrums $P(f)$ are calculated.

$$P(f) = \Gamma(F[I(x)]) \tag{10}$$

Step 3 Spectral residuals $R(f)$ are calculated by Formula (8).

Step 4 Global feature maps GsMap(x) are obtained.

$$\text{GsMap}(x) = g(x) * F^{-1}[\exp(R(f) + P(f))]^2 \tag{11}$$

6 The model of images saliency detection

Existed researches show [31] that in the process of biology observing around things, the images are equal bandwidth processed by retina. The researchers also found that eyeball is more sensitive to low-frequency elements. Therefore, the low-frequency elements are usually regarded as visual significant goals. The low- and high-frequency information of multi-scale image features is obtained by wavelet decomposition. Local features saliency maps are obtained by removing high-frequency information and reconstructing low-frequency images. It simulates the process of biological vision. However, wavelet transform has good local information representation capability. It is easy to cause redundant information is detected. So, we enhance the saliency target by fusing global feature saliency information and suppressing redundant information. Hou's method is based on information theory. It uses logarithmic spectrum and some mathematical operation to analysis original images. It effectively removes the image background and obtains saliency maps. The final saliency map is obtained by fusion algorithm (13). Images' significant detection is shown in Fig. 4. The experimental results show that the precision is higher by fusing local and global features saliency maps.

$$\text{FsMap} = M[\text{LsMap}(x) \times e^{\text{GsMap}(x)}] * g(x) \tag{12}$$
$$\text{SaliencyMap} = \Theta_{(x)} \times FsMap \tag{13}$$

FsMap is non-center-priori saliency map; SaliencyMap is the center-priori saliency map. $M(\cdot) = (\cdot)^{\ln\sqrt{2}}/\sqrt{2}$ is a fusion algorithm [23].

In general, visual focus is more close to the center of people's view. But, due to the shooting angle or actual demand, saliency target maybe deviated from center of image and even appear in the image border area. So, the improved center-priori saliency model is built by Harris corner detection and convex hull based on the frequency-tuned method.
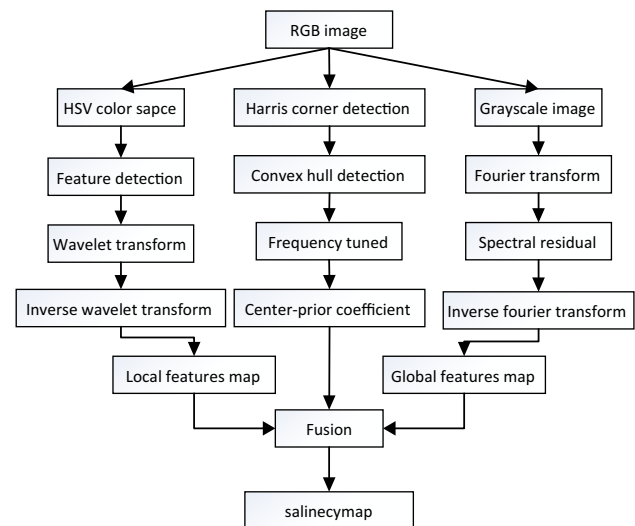
Fig. 4 Model of saliency detection

7 Experiment results and analysis of the MSRA database

This paper used MSRA database [32] for different measurement model. It contains 1000 RGB images which are used to test the detection results. In addition, the database contains binary images which are marked manually. The significant target is marked as "1." Background is marked as "0". In addition, this paper adopts evaluation mechanism that is widely recognized: *P–R* curve, *F*-Measure chart, ROC curve [33] and AUC value. This paper adopts *F*-Measure to evaluate the detection performance of local features saliency map, global feature saliency map, non-center-priori fusion model and center-priori model.

Figure 5 contains local saliency map (LsMap), global saliency map (GsMap), saliency map (FsMap), and center-priori saliency map (SaleincyMap). Wavelet transform saliency model can effectively detect significant target and get target outline. But, at the same time, background texture was detected because of its strong local information detection ability. It reduces detection accuracy as shown in Fig. 5b. Hou's method is almost not influenced by complex texture, but can cause significant target detection which is
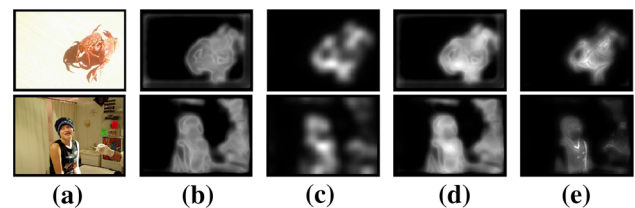
Fig. 5 Feature saliency maps. **a** is original map. **b** is local saliency map. **c** is global saliency map. **d** is non-center-priori saliency map. **e** is final saliency map
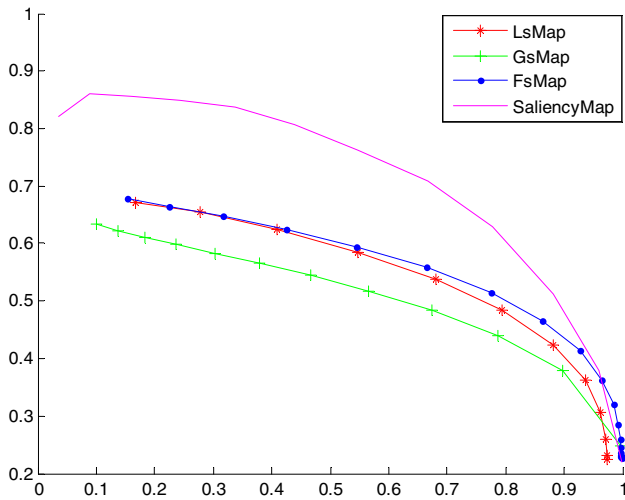
Fig. 6 Features saliency maps *P–R* curve

Fig. 7 *F*-Measure histogram

tion, as shown in Fig 5.

$$\text{Precision} = \frac{|M \cap G|}{|M|}$$

$$\text{Recall} = \frac{|M \cap G|}{|G|},$$

$$F_{\text{measure}} = \frac{(1+\alpha) \times P \times R}{\alpha \times P + R} \quad (14)$$

where M means saliency map to a binary mask, G means ground-truth map. The F_α is proposed as a weighted harmonic mean of Precision and Recall with a nonnegative weight α. It is often set to 0.3 to raise more importance to the Precision value. The reason for weighting precision more than recall is that recall rate is not as important as precision.

In order to measure the performance of the proposed model, we use ROC curve to measure Itti's model, wavelet transform method, Hou's spectral residual, and Achanta' FT model. The calculations of TPR and FPR are described as Formula (15). AUC is refers to the area under ROC curve. If the AUC value is bigger, the detection result is better. So, Fig. 8 shows that our method is better, but images' edge is blurry sometimes.

$$\text{TPR} = \frac{|M \cap G|}{|M|} \quad \text{FPR} = \frac{|M \cap G|}{|M \cap G| + |\overline{M} \cap \overline{G}|} \quad (15)$$

where TPR is true-positive rate, FPR is false-positive rate, and AUC is area under ROC curve.

Figure 9 shows ROC curve of several models which shows that our method is better. The area under ROC curve of our method is maximum, which is shown in Fig. 10.

8 Experimental results and analysis for the CSSD

Although images from MSRA database have a large variety in their content, background structures are primarily simple and smooth. And, in this database, most images contain only a single salient object. To represent the situations that natural

incomplete as shown in Fig. 5c. Therefore, we fused two kinds of feature maps by Formula (12), and the effect was shown in Fig. 5d. Fusion algorithm makes saliency maps more layered and makes it easier to distinguish between salient object and background. In order to further increase significantly the detection precision, we would enhance saliency target by adopting center-priori method based on the frequency-tuned method. The results show that the background is suppressed and the saliency target is further highlighted by adopting improved center-priori coefficient, such as Fig. 4e.

Figure 5 shows features maps for a few images, and our method is better intuitively. In order to further enhance the persuasiveness, we use 1000 images in the MSRA10K database for the testing experiments and *P–R* curve was obtained. (*P–R* algorithm as Formula (14), the Precision *P* refers to accuracy; the Recall *R* is recall rate, it is overlap ratio for significant regional and artificial mark target). In the process of calculation *P–R* parameter, the paper used fixed threshold segmentation algorithm. We need to calculate a set of *P*, *R* for each threshold. Finally, the *P–R* curve is obtained as shown in Fig. 6. It shows that our method can improve the effect of detection. The precision is improved effectively by using center-priori coefficient and histogram of the image enhancement process. In addition, the *F*-Measure chart is a comprehensive evaluation for *P–R* value. We adopted the Otsu threshold segmentation algorithm [34] for getting *F*-Measure value. The saliency map is needed to be transformed to binary image *M* from grayscale image and is dealt by Formula (14) with manual marked image *G*. Figure 7 shows that shooting rate of our approach is higher. And, the comprehensive evaluation *P–R* is better. The method in this paper can highlight saliency target and suppress background informa-
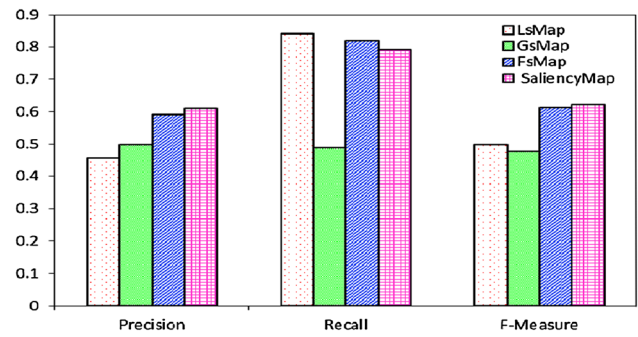
Fig. 8 Detection effects of various models, **a** Input **b** Itti [13] **c** WT [23] **d** SR [16] **e** FT [17] **f** Ours
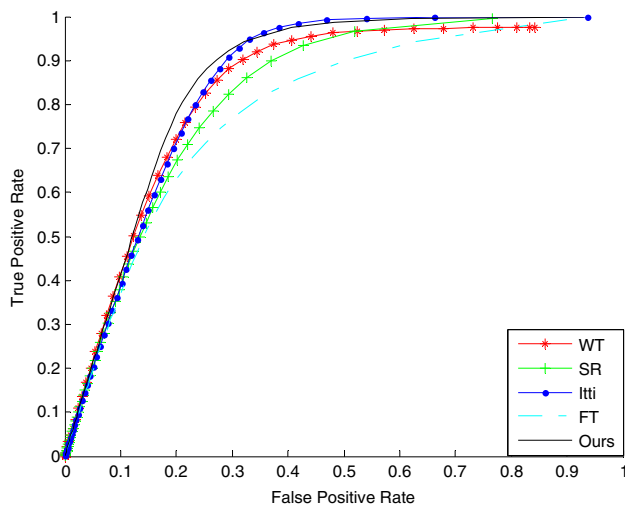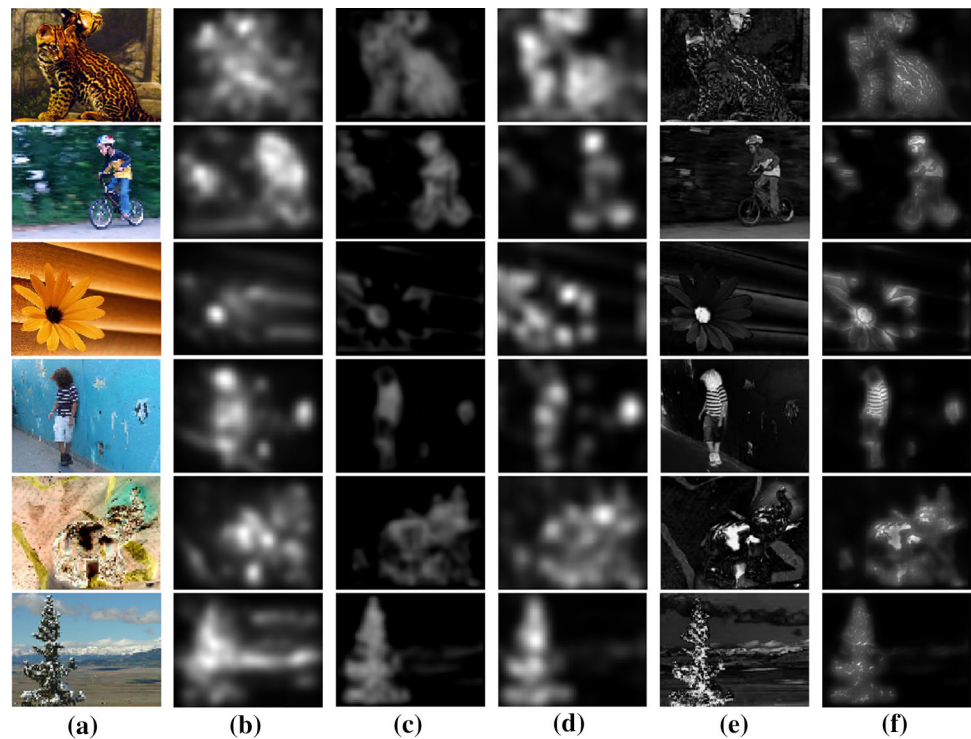

(a)　　(b)　　(c)　　(d)　　(e)　　(f)

Fig. 9 ROC curve of several models

Fig. 10 AUC value of several models

brightness area. In addition, Figs. 12 and 13 show that our approach is better.

9 Conclusions

The paper extracts image feature and obtains low-frequency information of image features by wavelet decomposing. Then, the local features saliency maps are obtained by wavelet reconstruction. Wavelet transform has obvious advantage in detecting images texture. But, it is not effective to highlight saliency target. Hou's method can effectively makes target prominent. But target edge texture information is lack. Therefore, we fuse the two methods to enhance the detection effects. In the end, it is conducive to further highlight the salient object and suppress background by using center-

images generally fall into, we use Complex Scene Saliency Dataset (CSSD) [35]. It includes many semantically meaningful but also structurally complex images for evaluation. Figures 11, 12, and 13 are the results of the qualitative and quantitative analyses.

Figure 11 shows that our model is effective for the complicated scenarios and multi-object saliency detection. Compared with other models, our method is more effective to highlight salient object and suppress background area, but also retain original images' information instead of a vague
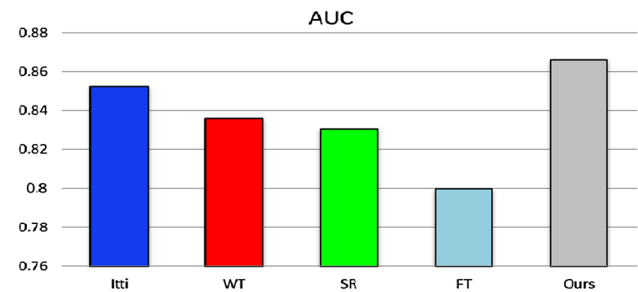
Fig. 11 Detection effects of various models, **a** Input **b** Itti [13] **c** WT [23] **d** SR [16] **e** FT [17] **f** Ours
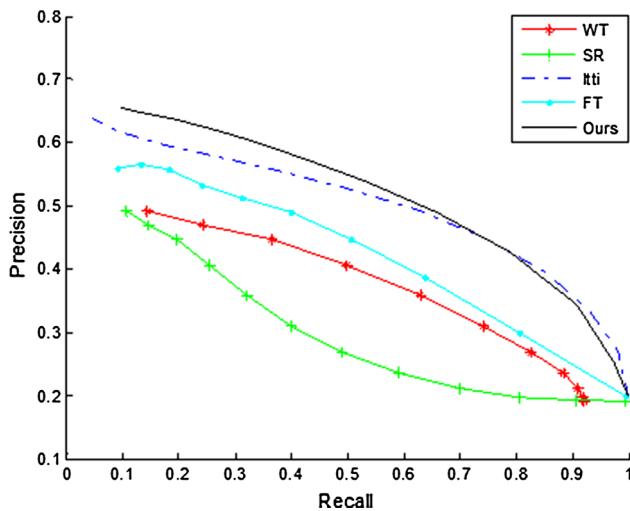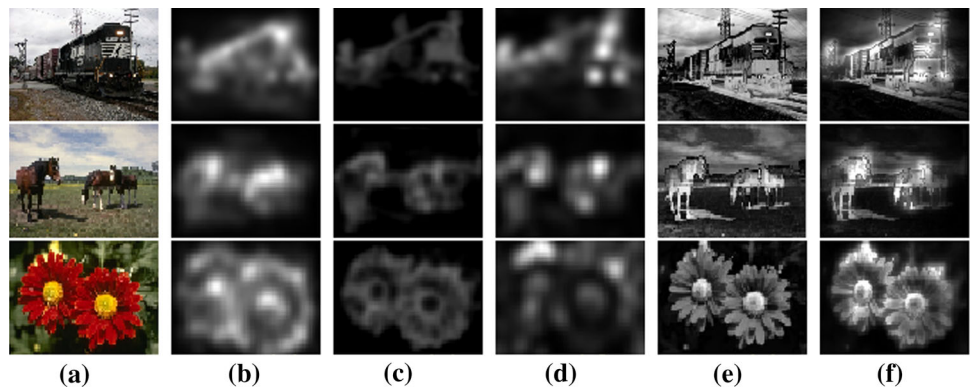

(a) (b) (c) (d) (e) (f)
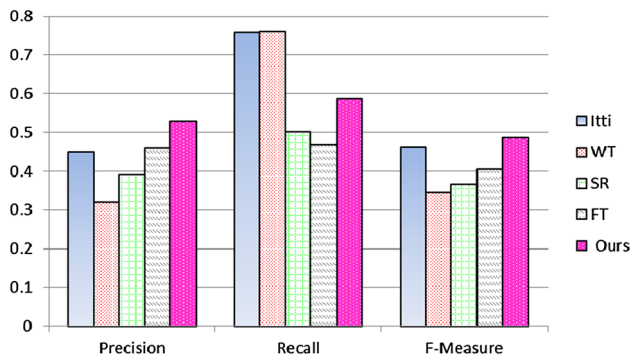

Fig. 12 *P–R* curve of the several models

Fig. 13 *F*-Measure histogram

priori based on the frequency-tuned method. The final results show that our model detection effect is better.

References

1. Shanableh, T.: Saliency detection in MPEG and HEVC video using intra-frame and inter-frame distances. Signal Image Video Process. **10**(4), 703–709 (2016)
2. Zhang, S.P., Yao, H.X., Liu, S.H.: Dynamic background subtraction based on local dependency histogram. IJPRAI **23**(7), 1397–1419 (2009)
3. Zhang, S.P., Kasiviswanathan, S., Yuen, P.C., Harandi, M.: Online dictionary learning on symmetric positive definite manifolds with vision applications. AAAI. 3165–3173 (2015)
4. Zhang, S.P., Zhou, H.Y., Yao, H.X., Zhang, Y.H., Wang, K.Q., Zhang, J.: Adaptive normalhedge for robust visual tracking. Signal Process. **110**, 132–142 (2015)
5. Zhang, S.P., Zhou, H.Y., Jiang, F., Li, X.L.: Robust visual tracking using structurally random projection and weighted least squares. IEEE Trans. Circuits Syst. Video Technol. **25**(11), 1749–1760 (2015)
6. Zhang, S.P., Yao, H.X., Sun, X., Lu, X.S.: Sparse coding based visual tracking: review and experimental comparison. Pattern Recogn. **46**(7), 1772–1788 (2013)
7. Jiang, F., Zhang, S.P., Wu, S., Gao, Y., Zhao, D.B.: Multi-layered gesture recognition with kinect. J. Mach. Learn. Res. **16**, 227–254 (2015)
8. Zhang, S.P., Yao, H.X., Sun, X., Wang, K.Q., Zhang, J., Lu, X.S., Zhang, Y.H.: Action recognition based on overcomplete independent components analysis. Inf. Sci. **281**, 635–647 (2014)
9. Itti, L.: Models of Bottom-Up and Top-Down Visual Attention. California Institute of Technology, Pasadena (2000)
10. Liu, Z.G., Yang, Y., Ji, X.H.: Flame detection algorithm based on a saliency detection technique and the uniform local binary pattern in the YCbCr color space. Signal Image Video Process. **10**(2), 277–284 (2016)
11. Hamel, S., Guyader, N., Pellerin, D., Houzet, D.: Contribution of color in saliency model for videos. Signal Image Video Process. **10**(3), 423–429 (2016)
12. Singh, N., Agrawal, R.K.: Combination of Kullback–Leibler divergence and Manhattan distance measures to detect salient objects. Signal Image Video Process. **9**(2), 427–435 (2015)
13. Itti, L., Koch, C., Niebur, E.: Model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. **20**(11), 1254–1259 (1998)

14. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. Adv Neural Inf. Process. Syst. **19**, 545–552 (2007)

15. Yang, C., Zhang, L.: Saliency detection via graph-based manifold ranking. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3166–3173. (2013)

16. Hou, X.D., Zhang, L.: Saliency detection: A spectral residual approach. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. (2007)

17. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1597–1604. (2009)

18. Maybank, S.: A probabilistic definition of salient regions for image matching. Nuerocomputing **120**, 4–14 (2013)

19. Yang, J., Yang, M.: Top-down visual saliency via joint CRF and dictionary learning. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2296–2303. (2012)

20. Cui, J., Xie, J.: Corner detection on finger vein images using the improved Harris algorithm. Opt. Int. J. Light Electron Opt. **125**(17), 4668–4671 (2014)

21. Yang, C., Zhang, L.: Graph-regularized saliency detection with convex-hull-based center prior. Signal Process. Lett. **20**(7), 637–640 (2013)

22. Tu, X.B.: Image Significant Algorithm and Evaluation Research. Nanjing Aeronautics and Astronautics University, Nanjing (2011)

23. Neverz, I., Lin, W., Fang, Y.: A saliency detection model using low-level features based on wavelet transform. IEEE Trans. Multimed. **15**(1), 96–105 (2012)

24. Biswas, A.: VidTIMIT audio visual phoneme recognition using AAM visual features and human auditory motivated acoustic wavelet features. In: IEEE Conference on Recent Trends in Information Systems, pp. 428–433. (2015)

25. Su, X.: Study on optimal wavelet decomposition level in infrared and visual light image fusion. In: IEEE Conference on Measuring Technology and Mechatronics Automation, pp. 616–619. (2010)

26. Ge, Y.: Application research on wavelet analysis in denoising of visual measurement edge detection. In: International Symposium on Computational Intelligence and Design, pp. 343–346. (2011)

27. Velasco, R., Luis, E.: Visual servoing using discrete wavelet transform. In: World Automation Congress. pp. 1–6. (2012)

28. Yang, C.: Salient Object Detection Based on Prior Integration and Manifold Ranking. Dalian university of technology, Dalian (2013)

29. Siraj, S., Sudheer, S.: Performance and analysis of high capacity steganography of color images involving wavelet transform. Opt. Int. J. Light Electron Opt. **126**(23), 3755–3760 (2015)

30. Zhang, J., Wang, M., Zhang, S.P., Li, X.L., Wu, X.D.: Spatiochromatic context modeling for color saliency analysis. IEEE Trans. Neural Netw. Learn. Syst. **27**(6), 1177–1189 (2016)

31. Simpson, W., Mcfadden, S.: Spatial frequency channels derived from individual differences. Vis. Res. **45**(21), 2723–2727 (2005)

32. Cheng, M., Zhang, G.: Global contrast based salient region detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 409–416. (2011)

33. Goferman, S., L. Manor, Z., Tal, A.: Context-aware saliency detection. In: IEEE conference on Computer Vision and Pattern Recognition, pp. 2376–2383. (2010)

34. Borj, A., Cheng, M.M., Jiang, H.: Salient object detection: a survey. arXiv preprint arXiv: 1441.5857 (2014)

35. Yan, Q., Xu, L., Shi, J.P., Jia, J.: Hierarchical saliency detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1155–1162 (2013)