IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON RECENT ADVANTAGES OF COMPUTER VISION BASED ON CHINESE CONFERENCE ON COMPUTER VISION (CCCV) 2017

Received December 6, 2017, accepted January 9, 2018, date of publication January 25, 2018, date of current version June 20, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2797880

Saliency Detection Method Using Hypergraphs on Adaptive Multiscales

FEILIN HAN[1], AILI HAN[2], AND JING HAO[2]
[1]College of Computer Science and Technology, Zhejiang University, Hangzhou 310007, China
[2]Department of Computer Science and Technology, Shandong University, Weihai 264209, China

Corresponding author: Aili Han (hanal@sdu.edu.cn)

This work was supported by the Shandong Provincial Natural Science Foundation of China under Grant ZR2016FM20.

ABSTRACT Saliency detection plays an important role in the fields of image processing and computer vision. We present an improved saliency detection method by means of hypergraphs on adaptive multi-scales (HAM). It first adjusts adaptively the ranges of pixel-values in *R*, *G*, *B* channels in an input image and uses the three ranges to determine adaptive scales for the construction of hypergraphs. And then, the image is modeled as multiple hypergraphs on adaptive scales in which hyper-edges are clustered by means of the agglomerative mean-shift method. The HAM method can get more single-scale hypergraphs and thus has higher accuracy than the previous ones because hypergraphs are constructed on adaptive multi-scales instead of fixed scales. Extensive experiments on the published benchmark datasets demonstrate that the HAM method has improved the performance of detecting salient objects in input images, especially for the images with narrow ranges of pixel-values.

INDEX TERMS Image processing, image analysis, object detection, saliency detection, hypergraph on adaptive multi-scales.

I. INTRODUCTION

Saliency detection aims to identify the most attractive region in an image. The applications include image segmentation [1], target recognition [2], image retrieval [3], [4], and so on. In order to extract salient objects in input images, many visual attention models have been proposed in the last twenty years.

Itti *et al.* [5] proposed a visual attention system inspired by animal visual systems in 1998. Since then, the researches on saliency detection have been developed rapidly. Ma *et al.* [6] designed a saliency detection method based on local contrast. Harel *et al.* [7] gave a graph-based visual saliency model. Cheng *et al.* [8] proposed a saliency detection method based on global contrast and space coherence. Li *et al.* [9] employed hypergraphs on a set of fixed scales to capture contextual attributes for saliency detection and proposed a contextual hypergraph modeling method for salient object detection (CHMS). Yuan *et al.* [10] presented a saliency detection method based on non-uniform quantification for *R*, *G*, *B* channels and weights for *L*, *a*, *b* channels. For more works related to saliency detection, please refer to [11]–[17].

The CHMS method [9] has good performance of capturing salient objects in most images, which has higher accuracy than many previous methods. Through the experiments, we find that, for the images with wide ranges of pixel-values (e.g., covering almost the whole range of [0, 255]), the CHMS method always has good performance; but for the images with narrow ranges of pixel-values (e.g., covering only the first or middle or last part of the range [0, 255]), the CHMS method usually cannot get very good performance. Further experiments show that hypergraphs on different scales have an important influence on experimental results.

Inspired by the experimental results of the CHMS method, we propose an improved saliency detection method using hypergraphs on adaptive multi-scales (HAM). The HAM method adaptively adjusts the ranges of pixel-values in *R*, *G*, *B* channels and detects salient objects on adaptive multi-scales, which provides more single-scale hypergraphs and thus gives better saliency maps, as shown in Fig. 1. The experimental results show that the HAM method can improve the performance of saliency detection, especially for the images with narrow ranges of pixel-values. The basic idea of adjusting adaptively the ranges of pixel-values in *R*, *G*, *B* channels in an image can be widely used in other applications in the fields of image processing, computer vision, or artificial intelligence.

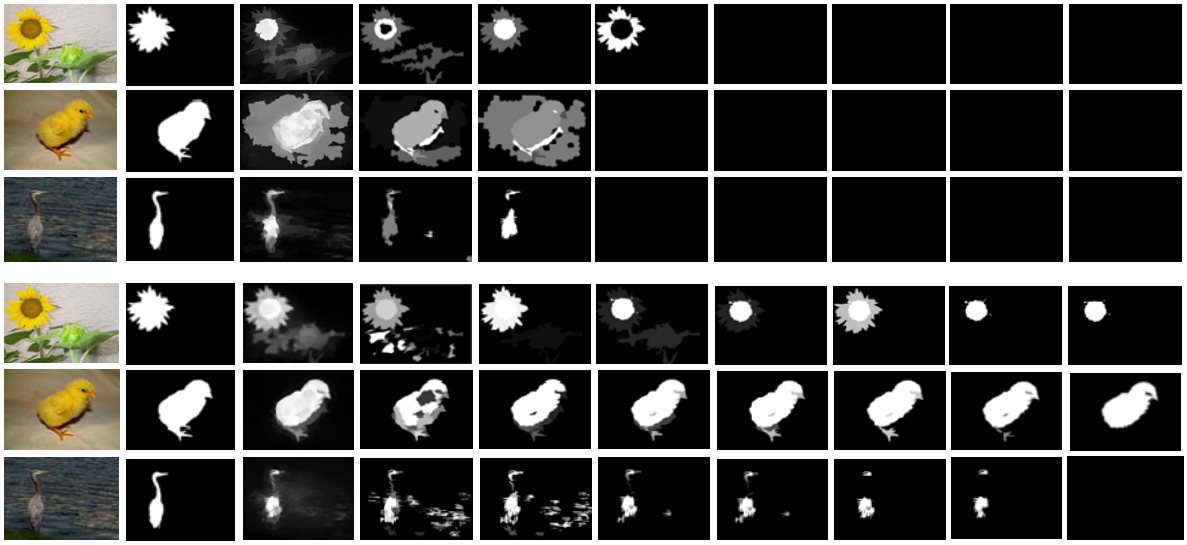

FIGURE 1. Comparison between the hypergraphs on fixed scales by the CHMS method and those on adaptive scales by the HAM methods for the images with narrow ranges of pixel-values. (*a*) The original image; (*b*) The ground truth; (*c*) The top three ones are the saliency maps by the CHMS method and the bottom three ones are the saliency maps by the HAM method; (*d*)-(*j*) The top three lines are the hypergraphs on the seven fixed scales 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75 by the CHMS method and the bottom three lines are the hypergraphs on adaptive scales by the HAM method. The experimental results show that the HAM method can get more single-scale hypergraphs and thus has better performance for the images with narrow ranges of pixel-values.

II. RELATED WORK

The CHMS method [9] uses hyper-edges on a set of fixed scales to capture the contextual properties of super-pixels, which improves significantly the performance of saliency detection. It first segments an input image into some super-pixels by means of SLIC [18] and then clusters super-pixels by means of agglomerative mean-shift [19].

In the CHMS method, an input image I is converted to a set of hypergraphs on fixed scales. Each hypergraph is denoted by $G = (V, E)$, where $V = \{v_i\}$ is a set of vertices corresponding to the super-pixels, and $E = \{e_j\}$ is a set of hyper-edges (a hyper-edge is a clique of super-pixels) that satisfies $\cup_{e_j \in E} e_j = V$.

The saliency of a hyper-edge is determined by the gradient magnitudes of the super-pixels within a narrow band along the boundary of the hyper-edge. For any hyper-edge e_j, the saliency score of e_j is defined as follows.

$$\Gamma(e_j) = \omega_{e_j} \left(\left\| I_g^* \bullet M_g(e_j) \right\|_1 - \rho(e_j) \right), \tag{1}$$

where ω_{ej} is the weight of e_j, I_g^* is the binary gradient map, $M_g(e_j)$ is a binary mask indicating the super-pixels within a narrow band along the boundary of the hyper-edge e_j, $\bullet$ is the elementwise dot product operator, $||.||_1$ is the 1-norm, and $\rho\ (e_j)$ is a penalty factor that is equal to the number of the super-pixels in the intersection of the hyper-edge e_j and the boundary super-pixels of the input image.

The saliency of a vertex (i.e., a super-pixel) is associated to the super-pixel and its contexts. For any hyper-pixel v_i in a hypergraph, the saliency score of v_i is defined as follows.

$$HSa(v_i) = \sum\nolimits_{e_j \in E} \Gamma(e_j) h(v_i, e_j), \tag{2}$$

where $\Gamma\ (e_j)$ is the saliency score of e_j, and $h(v_i, e_j)$ is the element value in the incidence matrix H. If $v_i \in e_j$, then $h(v_i, e_j) = 1$; otherwise $h(v_i, e_j) = 0$.

For more details, refer to [9]. We also refer to other works [20]–[23] to enrich the idea and improve the performance.

III. SALIENCY DETECTION USING HYPERGRAPHS ON ADAPTIVE MULTISCALES

The CHMS method [9] uses the seven empirical values 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75 as fixed scales to detect salient objects in an image, which has good performance for most images. It is widely accepted that there are great differences between different images. If all the pixel-values in an image lie in the first or middle or last part of the range [0, 255], there may be no hyper-edges for some scales which result in bad performance of saliency detection.

We propose an improved saliency detection method using hypergraphs on adaptive multi-scales (HAM). Different from the CHMS method, the HAM method uses adaptive scales instead of fixed scales. Thus, it can get more single-scale hypergraphs for the input images with narrow ranges of pixel-values, which results in higher accuracy.

A. ADAPTIVE SCALING OF PIXEL-VALUES

In order to adaptively adjust the ranges of pixel-values in R, G, B channels, we count for the pixel-values in each channel by means of histogram. The statistical results are used to determine the ranges of pixel-values covering more than 95% pixels in each channel. Selecting the range of covering more than 95% pixels is to avoid the influence of outliers [8]. The pixel-values outside the range are replaced by the nearest

pixel-values. And then, all the pixel-values are normalized and remapped to the range [0, 255]. For each channel $i, i \in \{R, G, B\}$, the remapped pixel-values are computed as follows.

$$\left(\left(\frac{I_i - low_i}{high_i - low_i}\right)^{\gamma} \cdot \left(high_{i,out} - low_{i,out}\right)\right) + low_{i,out}, \quad (3)$$

where I_i is a pixel-value in channel i in the input image I, low_i is the lower bound of the pixel-values in channel i, $high_i$ is the upper bounder of the pixel-values in channel i, $low_{i,out}$ is the lower bound of the remapped values in channel i, $high_{i,out}$ is the upper bound of the remapped values in channel i, γ indicates the curve shape. Here, $low_{i,out} = 0, high_{i,out} = 255$, $\gamma = 1$. Note that $\gamma = 1$ indicates a linear mapping.

Algorithm 1 Adaptive Scaling of Pixel-Values

Input: an *RGB* image I.
Output: the image I' after adaptive scaling.

(1) Count for the pixel-values in R, G, B channels to generate the histograms of R, G, B channels.
(2) Judge whether a pixel-value belongs to the range of covering more than 95% pixels in R, G, B channels. For R channel, the minimal pixel-value belonging to the range of covering more than 95% pixels is used as the lower bound low_r, and the maximal is used as the upper bound $high_r$. Same for G and B channels. Thus, the three ranges of pixel-values are $[low_r..high_r]$, $[low_g..high_g]$, and $[low_b..high_b]$.
(3) Normalize the three ranges of pixel-values, and then remap the normalized ranges of pixel-values to [0, 255]. For each channel $i, i \in \{R, G, B\}$, the normalization and remapping are done by the equation (3).

B. CONSTRUCTION OF HYPERGRAPHS ON ADAPTIVE SCALES

The CHMS method constructs hypergraphs on a set of fixed scales by means of agglomerative mean-shift [19]. Different from the CHMS method, the HAM method uses a set of adaptive scales instead of fixed scales. It first adjusts adaptively the ranges of pixel-values in R, G, B channels in an image to get three remapped ranges of pixel-values and then combines the remapped pixel-values with a set of fixed scales. The result of the two operations is equivalent to that of dealing with the original image on a set of adaptive scales. Thus, the HAM method uses hypergraphs on adaptive multi-scales to detect salient objects. The distance between any two pixels x_i and x_j on an adaptive scale μ is computed as follows.

$$dist = \left(\frac{x_i - x_j}{\mu}\right)^2. \quad (4)$$

The adaptive scaling of pixel-values in R, G, B channels in an image can result in more single-scale hypergraphs for the images with narrow ranges of pixel-values (e.g., covering only the first or middle or last part of the range [0, 255]). Thus, the performance of saliency detection has been improved by means of the HAM method. Take the three images shown in the first column in Fig. 1 as examples. When using the HAM method to detect salient objects, the number of single-scale hypergraphs is 7, 7, 6, respectively; whereas using the CHMS method to do that, the number of single-scale hypergraphs is 3, 2, 2, respectively, as shown in the (d)-(j) columns in Fig. 1. Extensive experiments show that, for the images with narrow ranges of pixel-values, the number of single-scale hypergraphs by the HAM method is always greater than that by the CHMS method; for the images with wide ranges of pixel-values, the two methods usually get similar number of single-scale hypergraphs. From the experimental results, it can be concluded that the HAM method can get more single-scale hypergraphs and thus has better performance than the CHMS method for the images with narrow ranges of pixel-values.

C. SALIENCY DETECTION USING HYPERGRAPHS ON ADAPTIVE MULTISCALES

The CHMS method uses hypergraphs on a set of fixed scales, no considering color difference of input images. For any input image, it takes the seven empirical values 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75 as fixed scales. In order to detect salient objects considering the factor of color differences in an image [10], we make an adaptive scaling of pixel-values in R, G, B channels and use hypergraphs on adaptive multi-scales to detect salient objects in an image. Extensive experiments show that the HAM method can improve the performance of detecting salient objects in input images, especially for the images with narrow ranges of pixel-values.

Algorithm 2 Saliency Detection Using Hypergraphs on Adaptive Multi-Scales

Input: an *RGB* image I.
Output: the saliency map S of the image I.

(1) Apply the Algorithm 1 to the image I to get the three ranges of pixel-values in R, G, B channels, which are normalized and remapped to [0, 255].
(2) Employ the SLIC algorithm [18] to over-segment the remapped image into some super-pixels to get a super-pixel image I'.
(3) Combine the super-pixel image with a set of fixed scales, which is equivalent to that of dealing with the input image on a set of adaptive scales. For each scale, use the agglomerative mean-shift method to cluster super-pixels into hyper-edges to get a hypergraph.
(4) Apply the Gaussian operator to the remapped image to get a gradient map I_g, and then use I_g to compute the saliency scores of hyper-edges and super-pixels to get a saliency map S_{SG}.
(5) Minimize a cost-sensitive classification function to get a LS-SVM based saliency map S_{SVM}.
(6) Linearly combine the saliency maps S_{SG} and S_{SVM} to get the final saliency map S.

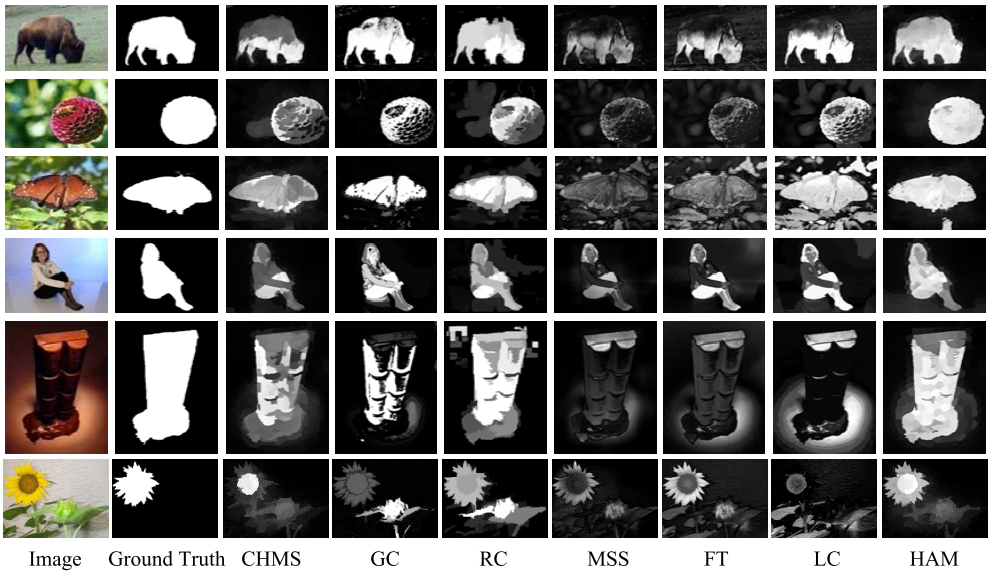

FIGURE 2. The detection results by the HAM method and six state-of-the-art methods on the MSRA-1000 dataset.

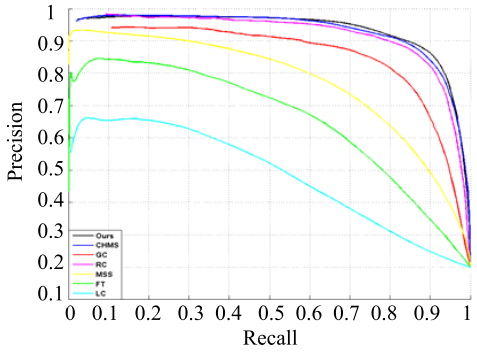

FIGURE 3. The *P-R* curves by the HAM method and six state-of-the-art methods on the MSRA-1000 dataset.

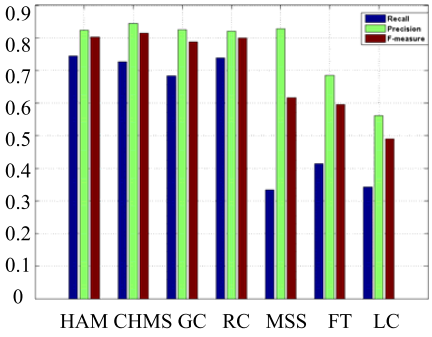

FIGURE 4. The *F*-measurements by the HAM method and six state-of-the-art methods on the MSRA-1000 dataset.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

We tested the HAM method on four published datasets, that is, the MSRA-1000 dataset [11], the SED dataset [12], the SOD dataset [13], and the ImgSal-50 dataset [13]. The MSRA-1000 dataset includes 1000 natural images with ground truth, which is one of the benchmark datasets for saliency detection. The SOD dataset has 300 very challenging images, in which the backgrounds and objects are complex. The ImgSal-50 dataset consists of 50 images in which the objects are clear, bigger and salient. The SED dataset consists of two subsets: SED1 and SED2, each includes 100 natural images with ground truth. The images in the SED1 subset have only one salient object, and the images in the SED2 subset have two salient objects. We also set up a new dataset marked as IMNR (the images with narrow ranges of pixel-values) which includes 50 images with smaller differences between the backgrounds and objects, i.e., with narrow ranges of pixel-values.

We use the *precision-recall* curves and F-measurements to evaluate the quantitative performances of the HAM method and six state-of-the-art methods. The F-measurements is computed as follows.

$$F_\beta = \frac{(1+\beta^2)P \times R}{\beta^2 P + R}, \tag{5}$$

where P is the precision rate, and R is the recall rate. We also select $\beta^2 = 0.3$.

The HAM method is compared with six state-of-the-art methods including CHMS [9], GC [14], RC [8], MSS [15], FT [11], LC [16]. Note that the CHMS method is based on LS-SVM [1] to obtain saliency maps. In order to do a fair comparison, the HAM method is also based on LS-SVM to do saliency detection. Extensive experiments show that the HAM method has improved the performance of saliency detection, which is the most competitive one in all the seven methods.

A. EXPERIMENTAL RESULTS ON THE MSRA-1000 DATASET

We use the HAM method and six state-of-the-art saliency detection methods to detect salient objects on the

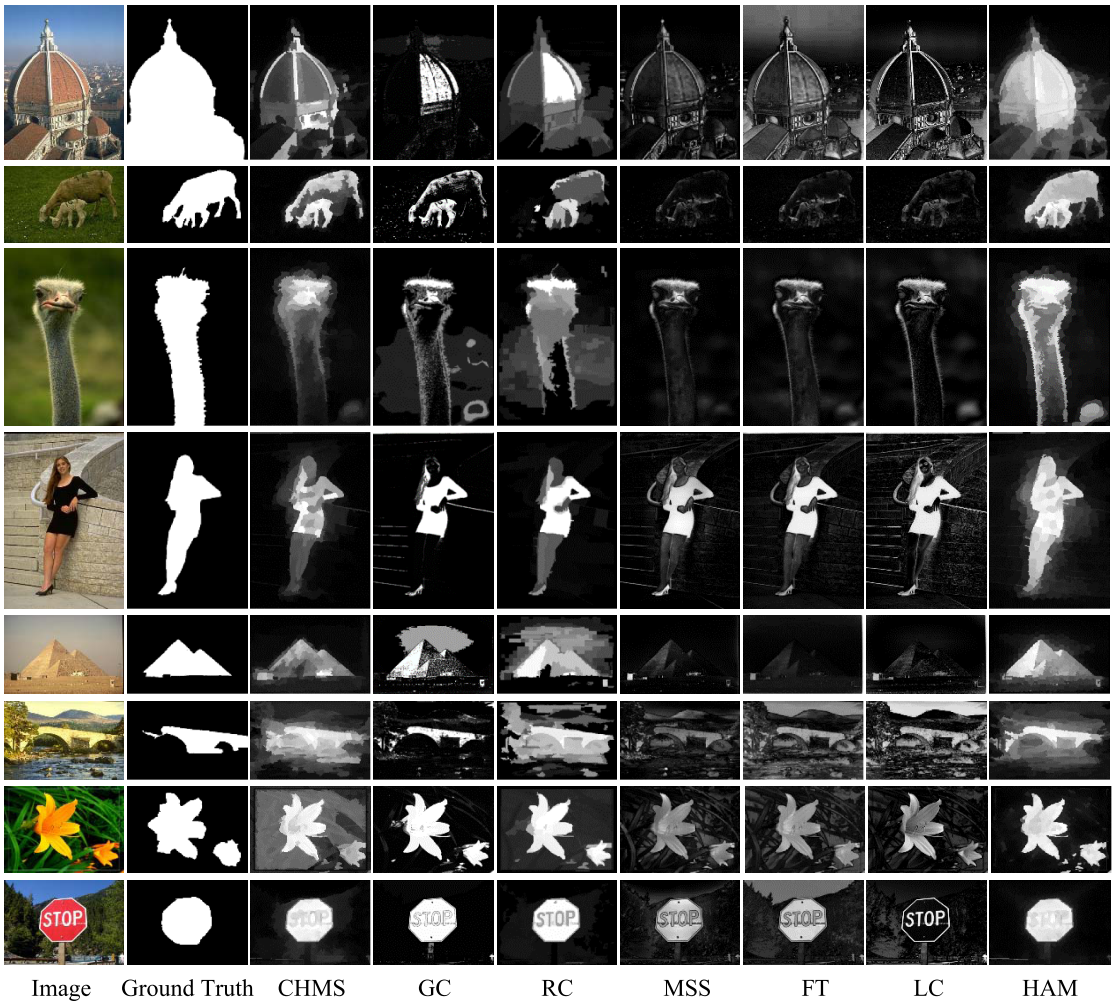

FIGURE 5. The detection results by the HAM method and six state-of-the-art methods on the SOD and ImgSal-50 datasets.

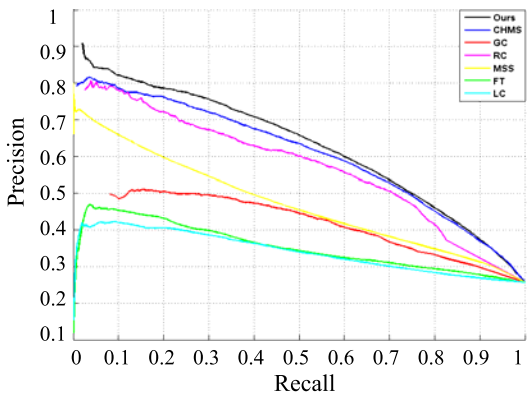

FIGURE 6. The *P-R* curves by the HAM method and six state-of-the-art methods on the SOD dataset.

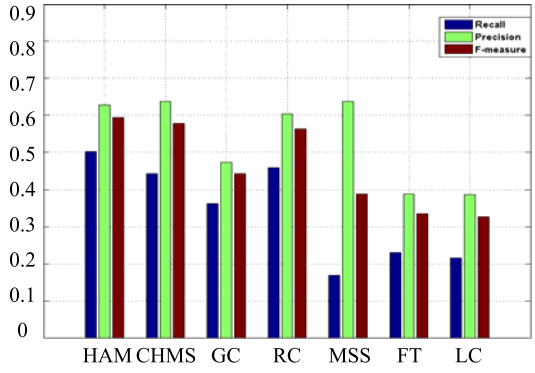

FIGURE 7. The *F*-measurements by the HAM method and six state-of-the-art methods on the SOD dataset.

MSRA-1000 dataset. Some of the experimental results are shown in Fig. 2. Extensive experiments show that the HAM method can more uniformly highlight objects and darken backgrounds than other methods. Furthermore, when the pixel-values of objects and backgrounds in an image are in a narrow range, the HAM method can better capture edge information of objects, uniformly highlight salient objects, and darken backgrounds.

The *Precision-Rate* (*P-R*) curves and *F*-measurements by the seven saliency detection methods on the MSRA-1000 dataset are shown in Fig. 3 and Fig. 4. From the experimental results, we can find that the HAM method has the best performance in all the seven methods, but its advantages are

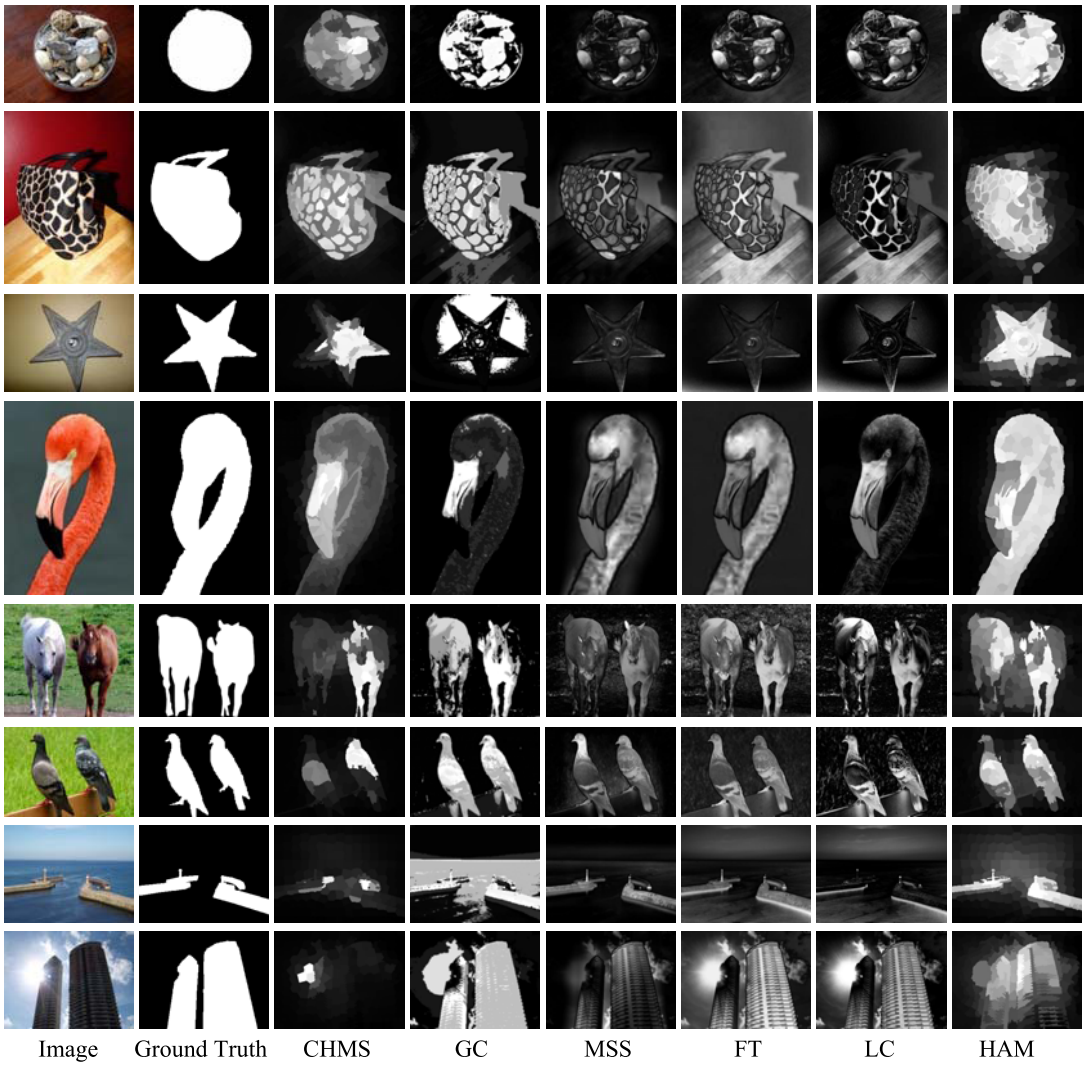

FIGURE 8. The detection results by the HAM method and five state-of-the-art methods on the SED dataset.

not obvious since most images in the MSRA-1000 dataset are with wide ranges of pixel-values and seldom of them are with narrow ranges of pixel-values.

B. EXPERIMENTAL RESULTS ON THE SOD AND IMGSAL-50 DATASETS

The HAM method and six state-of-the-art saliency detection methods are tested on the SOD and ImgSal-50 datasets. Some of the detection results are shown in Fig. 5, in which the results on top six lines are from the SOD dataset and the results on bottom two lines are from the ImgSal-50 dataset. The images in the SOD dataset have complex backgrounds or objects, which are more challenging than those in other datasets. Extensive experiments show that the HAM method can better detect objects in complex images and uniformly highlight objects. When dealing with complex images, the HAM method can find more complete contours of objects than other methods.

The *Precision-Rate* curves and F-measurements by the HAM method and six state-of-the-art saliency detection methods on the SOD dataset are shown in Fig. 6 and Fig. 7. The experimental results show that the *Precision-Rate* curves and F-measurements by the HAM method are always better than other six methods.

C. EXPERIMENTAL RESULTS ON THE SED DATASET

The HAM method and five state-of-the-art saliency detection methods are tested on the SED dataset. Some of the detection results are shown in Fig. 8. The SED dataset consists of two subsets: SED1 and SED2. Each image in SED1 has only one object, whereas each image in SED2 has two objects. The detection results on the top four lines in Fig. 8 are from the SED1 subset, and that on the bottom four lines are from the SED2 subset. Extensive experiments on the SED dataset show that the HAM method is superior to other five methods.

The *Precision-Rate* curves and F-measurements by the six detection methods on the SED1 and SED2 datasets are shown in Fig. 9, Fig. 10, Fig. 11, and Fig. 12, respectively. From the performances of the six methods on the SED1 and

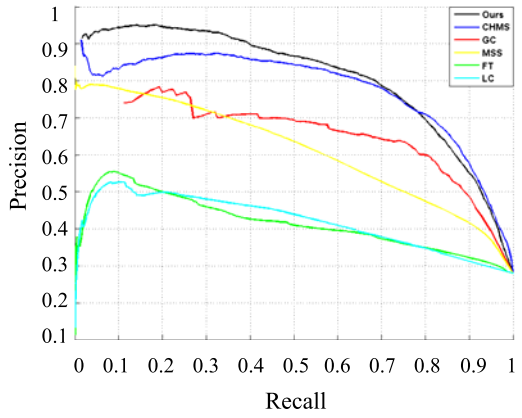

FIGURE 9. The *P-R* curves by the HAM method and five state-of-the-art methods on the SED1 dataset.

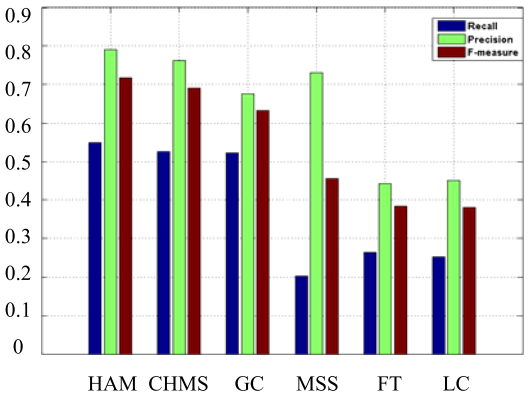

FIGURE 10. The *F*-measurements by the HAM method and five state-of-the-art methods on the SED1 dataset.

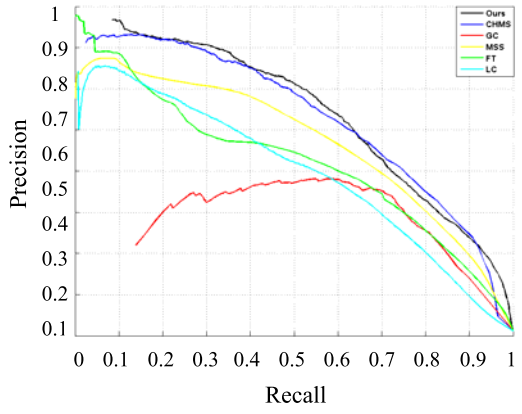

FIGURE 11. The *P-R* curves by the HAM method and five state-of-the-art methods on the SED2 dataset.

SED2 datasets, it can be found that the HAM method consistently performs better than other methods.

D. EXPERIMENTAL RESULTS ON THE IMNR DATASET

The HAM method and six state-of-the-art saliency detection methods are also tested on the IMNR dataset which includes 50 images with narrow ranges of pixel-values

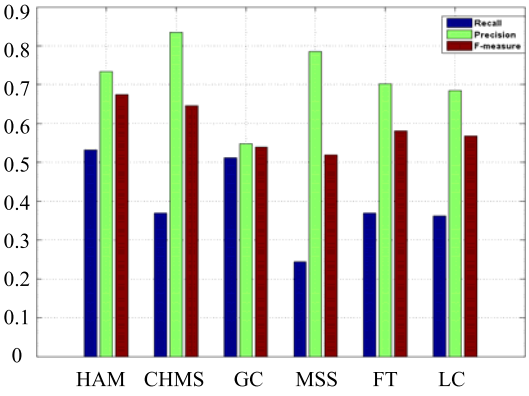

FIGURE 12. The *F*-measurements by the HAM method and five state-of-the-art methods on the SED2 dataset.

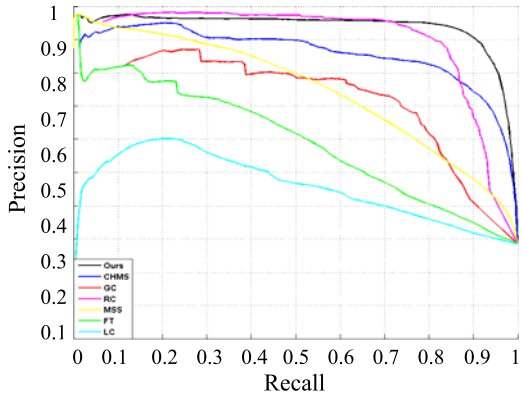

FIGURE 13. The *P-R* curves by the HAM method and six state-of-the-art methods on the IMNR dataset.

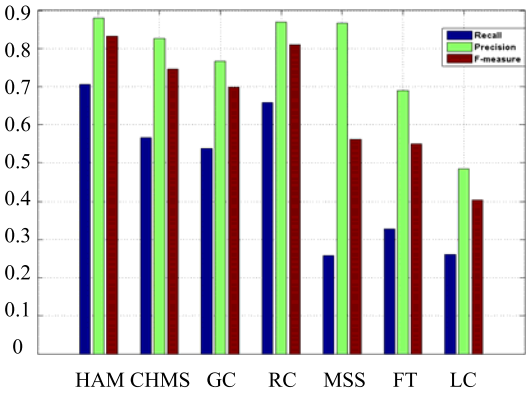

FIGURE 14. The *F*-measurements by the HAM method and six state-of-the-art methods on the IMNR dataset.

in *R*, *G*, *B* channels. The images in the IMNR dataset, which is with smaller differences between backgrounds and objects, are collected from the published benchmark datasets.

The *Precision-Rate* curves and *F*-measurements by the HAM method and six state-of-the-art methods on the IMNR

dataset are shown in Fig. 13 and Fig. 14, from which we can find that the HAM method has the best performance on the IMNR dataset in all the seven methods. Extensive experiments show that, for the images with narrow ranges of pixel-values, the performance by the HAM method are obviously better than other six methods. Thus, the HAM method is especially suitable for the images with small differences of pixel-values in objects and backgrounds.

V. CONCLUSION

The previous saliency detection methods based on hypergraphs uses fixed scales, no considering color difference of input images. In order to detect objects considering color differences of images, we propose an improved saliency detection method based on hypergraphs on adaptive multi-scales. The proposed method adaptively adjusts the ranges of pixel-values in R, G, B channels and uses hypergraphs on adaptive scales to detect salient objects in images. Extensive experiments on the published datasets show that the proposed method can improve the performance of saliency detection for the images with narrow ranges of pixel-values. For the images with wide ranges of pixel-values, the performance is also improved to some extent. Furthermore, the saliency maps by the HAM method are more smooth and robust than other methods. The basic idea of adaptively scaling pixel-values in R, G, B channels can be widely used in the fields of image processing, computer vision, or artificial intelligence.

REFERENCES

[1] L. Wang, J. Xue, N. Zheng, and G. Hua, "Automatic salient object extraction with contextual cue," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 105–112.

[2] V. Navalpakkam and L. Itti, "An integrated model of top-down and bottom-up attention for optimizing detection speed," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 2049–2056.

[3] X.-J. Wang, W.-Y. Ma, and X. Li, "Data-driven approach for bridging the cognitive gap in image retrieval," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jun. 2004, pp. 2231–2234.

[4] P. Xu *et al.*, "Cross-modal subspace learning for fine-grained sketch-based image retrieval," *Neurocomputing*, vol. 278, pp. 75–86, Feb. 2018.

[5] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[6] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proc. 11th ACM Int. Conf. Multimedia*, 2003, pp. 374–381.

[7] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 19. pp. 545–552, 2007.

[8] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, Mar. 2015.

[9] X. Li, Y. Li, C. Shen, A. Dick, and A. van den Hengel, "Contextual hypergraph modeling for salient object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 3328–3335.

[10] Y. Yuan, A. Han, and F. Han, "Saliency detection based on non-uniform quantification for RGB channels and weights for lab channels," in *Computer Vision*. Berlin, Germany: Springer, 2015, pp. 258–266.

[11] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 1597–1604.

[12] S. Alpert, M. Galun, R. Basri, and A. Brandt, "Image segmentation by probabilistic bottom-up aggregation and cue integration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8.

[13] J. Li, M. D. Levine, X. An, X. Xu, and H. He, "Visual saliency based on scale-space analysis in the frequency domain," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 996–1010, Apr. 2013.

[14] M. M. Cheng, J. Warrell, W. Y. Lin, and S. Zheng, "Efficient salient region detection with soft image abstraction," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1529–1536.

[15] R. Achanta and S. Süsstrunk, "Saliency detection using maximum symmetric surround," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 2653–2656.

[16] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proc. ACM Int. Conf. Multimedia*, 2006, pp. 815–824.

[17] W. H. Cheng, "A Visual attention based region-of-interest determination framework for video sequences," *IEICE Trans. Inf. Syst.*, vol. E88-D, no. 7, pp. 1578–1586, 2005.

[18] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[19] X. T. Yuan, B. G. Hu, and R. He, "Agglomerative mean-shift clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 2, pp. 209–219, Feb. 2012.

[20] Z. Ma *et al.*, "The role of data analysis in the development of intelligent energy networks," *IEEE Netw.*, vol. 31, no. 5, pp. 88–95, Sep./Oct. 2017.

[21] Z. Ma, J. H. Xue, A. Leijon, Z. H. Tan, Z. Yang, and J. Guo, "Decorrelation of neutral vector variables: Theory and applications," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 129–143, Jan. 2018.

[22] W.-H. Cheng, C.-W. Wang, and J.-L. Wu, "Video adaptation for small display based on content recomposition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 1, pp. 43–58, Jan. 2007.

[23] C.-C. Ho, J.-L. Wu, and W.-H. Cheng, "A practical foveation-based rate-shaping mechanism for MPEG videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 11, pp. 1365–1372, Nov. 2005.

[24] Q. Wang, Z. Meng, and X. Li, "Locality adaptive discriminant analysis for spectral–spatial classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2077–2081, Nov. 2017.

[25] J. Gao, Q. Wang, and Y. Yuan, "Embedding structured contour and location prior in siamesed fully convolutional networks for road detection," in *Proc. IEEE Int. Conf. Robot. Autom.*, Jun. 2017, pp. 219–224.

FEILIN HAN received the B.S. degree in multimedia technology from Shandong University, China, in 2014. She is currently pursuing the Ph.D. degree with the College of Computer Science and Technology, Zhejiang University, China. Her research interests include pattern recognition and machine learning, image and video processing, and video surveillance.

AILI HAN received the M.S. degree in computer science in 2001 and the Ph.D. degree in computer software from Shandong University, China, in 2008. She is currently a Professor of computer science with Shandong University. Her research interests include image processing, computer vision, algorithm design and analysis, and intelligent computing.

JING HAO received the B.S. degree in computer science from Shandong University, China, in 2014. She is currently pursuing the master's degree with the Department of Computer Science and Technology, Shandong University. Her research interests include image processing, saliency detection, and face recognition.