

Saliency Detection via Robust Seed Selection of Foreground and Background Priors

Muwei Jian^{1,2}, Ruihong Wang³, Hui Yu², Junyu Dong³, Yujuan Wang⁴, Yilong Yin⁵ and Kin-Man Lam⁶

¹ School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan, 250014, China

E-mail: jianmuwei@163.com Tel: +86-531-88596228

² School of Creative Technologies, University of Portsmouth, Portsmouth, UK

E-mail: hui.yu@port.ac.uk

³ Department of Computer Science and Technology, Ocean University of China, Qingdao, China

E-mail: {jingwang; dongjunyu}@ouc.edu.cn

⁴ Shandong Museum, Jinan 250014, China

E-mail: 59197120@qq.com

⁵ School of Software Engineering, Shandong University, Jinan 250101, China.

E-mail: ylyin@sdu.edu.cn

⁶ Centre for Signal Processing, Department of Electronic and Information Engineering,

The Hong Kong Polytechnic University, Kowloon, Hong Kong

E-mail: enkmlam@polyu.edu.hk Tel: +852-27662607

Abstract— Recently, saliency detection has become a research hotspot in both the computer-vision and image-processing fields. Among the diverse saliency-detection approaches, those based on the foreground and background-based model can achieve promising performance. Reliable seed selection for the foreground and background priors is a critical step for successful saliency detection. In this paper, we firstly exploit the spatial distribution of the extracted directional patches to predict the centroid of each salient object in an image. Then, we adopt the located centroids as the visual-attention center of the whole image to compute the superpixel-based center prior, which can facilitate the subsequent seed selection for the foreground and background-prior generation. Finally, the two individual foreground-based and background-based saliency maps are combined together into a plausible and authentic saliency map. Extensive experimental assessments on publicly available datasets demonstrate the effectiveness of our proposed model.

I. INTRODUCTION

With the exploitation of saliency detection, various models have been designed to imitate the visual attention mechanism of human beings [4-7]. Basically, saliency-detection frameworks can be divided into two categories [8, 9]: one is the bottom-up saliency detection model, while the other is the top-down saliency detection method. Roughly, the bottom-up saliency detection model is based on low-level visual features in images, including intensity, colour, contrast, texture, and orientation. From the biological point of view, the understanding of the visual attention mechanism for the bottom-up saliency detection model is unconscious and can be regarded as visual cues driven by centroid responding to the human visual system (HVS) [5-7]. On the contrary, the top-down computational model mainly relies on advanced understanding of image contents and requires prior knowledge of image scenes, which is special task-oriented towards actual applications. This commonly leads to the top-down saliency model lacking in usability and expansibility

Currently, most of the research work on saliency detection is based on the bottom-up modelling mechanism [8, 9]. In the early 1990s, Itti et al. [2] proposed a bottom-up saliency detection algorithm based on visual attention. In their model, an input image is firstly decomposed into three different channels, and then the center-surround contrast method is implemented to generate multi-scale image features, which are fused into a composite saliency map. In [3], Liu et al. devised a salient object detection method, which separates salient objects from the image background by taking multi-scale contrast, centered surrounded histogram, and colour space distribution into consideration. Based on the colour contrast scheme, Achanta et al. [10] proposed a frequency tuning (FT) method for salient region detection by considering the average colour difference of each pixel. According to the strategy of selective comparison, Wang et al. [11] devised a supervised Markov random field framework for saliency detection based on selective contrast. Later, in [12], a saliency detection model based on multi-instance learning was effectively designed, which combines the low-level, middle-level and high-level feature information in the process of computing saliency values. Yang et al. [13] proposed a graph regularization saliency-detection model, which combines contrast, smoothness priori and center priori based on convex hull for salient object detection. In [14], Tong et al. proposed a salient object detection method based on super-pixel and multi-scale analysis. The steering filter is optimized by the visual cues of local contrast, integrity and center deviation, to produce a final saliency map. Ma et al. [16] proposed a contrast-based image attention analysis model using a fuzzy growth algorithm for saliency-map generation. Based on the brightness information and spatial cues, Zhai and Shah [17] exploited global comparison to achieve saliency detection in video sequences. Harel et al. [18] proposed a graph-based visual saliency method for saliency

detection. Hou and Zhang [19] explored a spectral residual method based on logarithmic spectral analysis of images for saliency detection. In view of the comparison of local features in illumination, colour and motion cues, Rahtu [20] et al. designed a segmentation model based on conditional random field to separate salient objects from the background in images and videos. Murray et al. [21] proposed a nonparametric low-level saliency-detection method, which employs a spatial aggregation mechanism and a proportional weighting function to achieve saliency estimation. Goferman et al. [22] devised a situational awareness saliency-detection model based on a local-contrast method, which utilizes local low-level features, visual organization rules and high-level features to highlight salient objects in an image. Cheng et al. [7, 23] proposed a histogram-based contrast (HC) framework, and then combined spatial information to obtain a saliency map based on regional contrast (RC) for saliency detection. Qin et al. [24] designed a novel saliency detection method based on a cellular automated conduction mechanism, which combines colour and spatial contrast information in a Bayesian framework to produce a robust saliency map. In [25], Chen et al. defined two saliency principles: spatial compactness and colour contrast. Then, a super-pixel model based on homology similarity for saliency detection was proposed. Kong et al. [26] proposed a saliency detection framework by extending the Random Walk selection mechanism to separate the foreground objects from the background of the image. Jian et al. [27] introduced a saliency detection method by mixing a four-element distance-based Weber local descriptor and low-level priors. In [28], Li et al. designed a learning ranking method for salient object detection, which considers saliency detection as a cost-sensitive label sorting problem by sequencing saliency values in a descending order. To tackle saliency detection of underwater images, Jian et al. [29] proposed a saliency-detection model by combining the Weber descriptor based on quaternion distance and principal component analysis for underwater vision. Furthermore, a large underwater image database [30] was developed for saliency detection of underwater images.

On the other hand, since the top-down saliency detection model typically is designed with the focus on a single task and is also application-oriented, this requires large-scale training data with human labelled ground truths for supervised learning. Consequently, relatively few works have been designed based on the top-down saliency detection model. Oliva et al. [31] proposed a top-down saliency detection framework based on global scene configuration for object detection. In [32], Tao et al. designed a top-down-based saliency detection model for a traffic driving environment by imitating a driver's gaze behavior. Cholakkal et al. [33] proposed a top-down saliency detection method based on the locally constrained context sparse coding. Based on joint conditional random fields and dictionary learning, Yang et al. [34] devised a top-down saliency computational framework by applying a maximum margin algorithm to train the detection model based on fast inference algorithms.

The rest of the paper is organized as follows. In Section II, we will introduce the proposed saliency-detection model in detail. Experimental results on widely used databases are described in Section III. The paper closes with a conclusion and discussion in Section IV.

II. THE PROPOSED FRAMEWORK FOR SALIENCY DETECTION

A. Selection of the Spatial Center of Salient Objects

In this subsection, we exploit discrete wavelet frame transform (DWFT) for extracting directional patches in an image, and then utilize the spatial distribution of these directional patches to determine the potential center of salient objects. According to the principles of multi-resolution analysis (MRA), DWFT [1] is derived from Discrete Wavelet Transform (DWT), and both are sensitive to the variations in an image, which reflect the multiple directional properties. We represent the three decomposed subimages as LH^{DWFT} (the horizontal high-frequency details), HL^{DWFT} (the vertical high-frequency details), and HH^{DWFT} (the diagonal high-frequency details), respectively, and a low-pass filtered approximate sub-image LL^{DWFT} of the input image is also produced. These sub-images can be exploited to detect the directional patches/blocks in an image [35]. Then, the horizontal direction map based on each row of the sub-image $HL^{DWFT}(x, y)$ is defined as follows:

$$Dir_H(x, y) = \frac{1}{2W+1} |HL^{DWFT}(x, y)| \left\{ \sum_{r=-W}^W |HL^{DWFT}(x+r, y)| \right\}. \quad (1)$$

The number of adjacent coefficients, W , to be extracted can be set to the wavelet regularity value of orthogonal wavelets with compact support.

Similarly, by considering the columns in the $LH^{DWFT}(x, y)$ and $HH^{DWFT}(x, y)$ sub-images, the vertical direction map and the diagonal direction map can be defined. Since all these multi-directional maps are of the same size as the input image, they can be normalized, and then fused to form an integrated direction feature map. After that, directional patches can be extracted from the integrated direction feature map easily. Followed by [35], the $K = 15$ directional patches (Fig. 1(b)) with the largest values are selected from the global direction feature map, and the average centroid based on the directional patches, is calculated as the center prior estimation, as shown in Fig. 1(c).

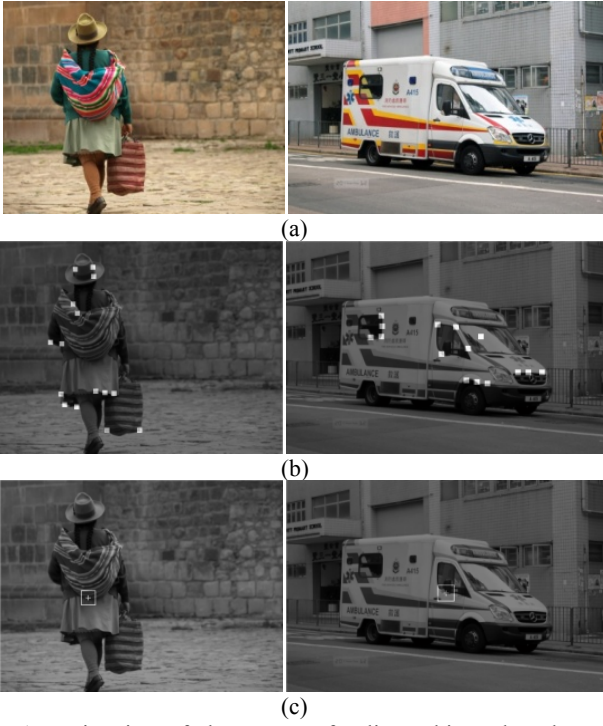


Fig. 1. Estimation of the center of salient objects based on the detected directional patches.

B. Center Prior Based on Centroid Superpixel

Superpixels have been widely used in image segmentation, saliency detection, and other related fields, because of their computational advantages. In our model, the simple linear iterative clustering algorithm (SLIC) [37] is applied for superpixel segmentation. According to human visual perception, people often look attentively at the central region of an object in an image, which means that the spatial center of a salient object plays an important role in saliency detection. The distance between a pixel position and the spatial center of a salient object can be defined as follows:

$$Cen(i) = \exp\left(\frac{-\sqrt{(x_i - x_0)^2 + (y_i - y_0)^2}}{2\sigma^2}\right), \quad (2)$$

where σ^2 controls the strength of spatial weighting, (x_0, y_0) are the coordinates of the spatial center of the salient object, and (x_i, y_i) are the coordinates of the pixel i .

Therefore, we can compute the center priori based on the centroid superpixel, in which the spatial center of the salient object is located. This can more accurately estimate the center of salient objects than directly using the superpixel located at the center of the image.

$$CenPri(j) = \frac{CenP(j)}{Area(j)}, \quad (3)$$

where $CenP(j)$ is the sum of the distances from all the pixels in the j^{th} superpixel to the spatial center (x_0, y_0) of the salient object, $Area(j)$ is the number of pixels in superpixel j , and the center priori $CenPri(j)$ is the distance of each superpixel to the centroid superpixel.

C. Colour Prior Based on Centroid Superpixel

Considering the fact that salient objects in an image generally have similar colour features, colour potentials are explored and utilized in saliency detection. The colour priori based on the centroid superpixel is proposed, as follows.

$$ColPri(j) = \frac{1}{\sqrt{\|Col(j) - Col(centroid)\|_2}}, \quad (4)$$

where $Col(j)$ is the mean colour feature vector of the j^{th} superpixel in the RGB colour space, and $Col(centroid)$ is the mean colour feature vector of the centroid superpixel, and $\|\cdot\|_2$ represents the Euclidean distance between the colour features of two superpixels.

D. Background and Foreground Cues

Most conventional background-based saliency-detection methods choose superpixels on the border of an image as background seeds [36, 43]. Firstly, based on the detected directional patches using DWFT with $K=15$ (see Fig. 1(b)), we utilize a convex hull by enclosing all the directional patches to label the salient object coarsely, and then prevent the background seed from appearing inside the convex hull to guarantee the effectiveness. Then, the colour and spatial distance between each superpixel and the foreground seed are calculated based on the selected background seed to obtain a background-cue saliency map [36].

Generally, a foreground salient object in an image has a visible contrast with the background, so the saliency map based on the foreground can highlight salient objects. In [39], Harris corner points are computed to form a coarse location of salient objects, and the rough region is used to generate the foreground seed. Different from the method [39], we use the centroid superpixel as an initial foreground seed. Those superpixels around this initial foreground seed are chosen as the foreground seeds to constitute the foreground seed set FG , using an adaptive threshold [5].

E. Saliency Fusion

Various types of visual cues are complementary to each other and have their own strengths and advantages. Firstly, the central priori saliency map and the colour priori saliency map are fused by pixel-by-pixel point multiplication:

$$Salmap_{cen-cor} = CenPri * ColPri, \quad (5)$$

where $*$ is the pixel-by-pixel multiplication operator.

Besides, with an attempt to further enhance the foreground cue of salient objects and adequately suppress the image background noises, a compounded saliency map is formed by

blending the foreground-cue and the background-cue saliency maps simultaneously, as follows [36]:

$$Salmap_{fore-back} = Sal_{back} \cdot (1 - \exp(-\gamma \cdot Sal_{fore})), \quad (6)$$

where γ is the equilibrium factor, which is set to 6 according to [36]. Then, the hybrid saliency map based on different types of visual cues is generated as follows:

$$Salmap_{final} = Sal_{cen-cor} \cdot (1 - \exp(-\gamma \cdot Sal_{fore-back})). \quad (7)$$

To make the saliency maps more constant and smooth, the final saliency map is polished by saliency diffusion and Gaussian attenuation between superpixels [36, 39].

III. EXPERIMENTS

In the experiments, a widely used database, the MSRA [6, 23] data set, was used to evaluate the performance of our proposed model, and to compare with other representative state-of-the-art saliency-detection methods. For comparison, six state-of-the-art saliency-detection models are considered, including HC (Histogram-based Contrast) [7], FT (Frequency Tuned) [10], MSS (Multi-Scale Superpixels) [14], LC (Local spatial saliency Calculation) [17], RC (Region Contrast) [23], and FD (Foreground and Background seed selection) [36].

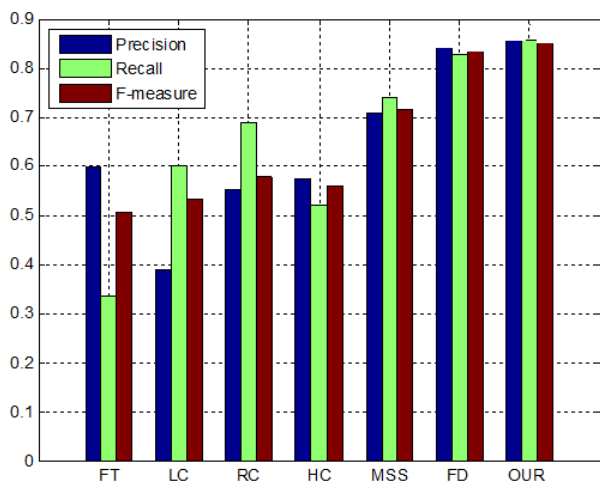


Fig. 2. Performance of the proposed model compared with state-of-the-art methods on the MSRA data set.

To quantitatively evaluate the effectiveness of the seven state-of-the-art saliency-detection methods, the average Precision, Recall, and F -measure are computed for objective analysis. Fig. 2 compares the performance of the proposed model with the other state-of-the-art models. We can see that the proposed algorithm achieves higher performance than the other methods in terms of precision, recall and F -measure.

In addition to the performance indices, Fig. 3 shows the visual results of the different state-of-the-art models. We can observe that our proposed method can generate more accurate saliency maps than the other methods.

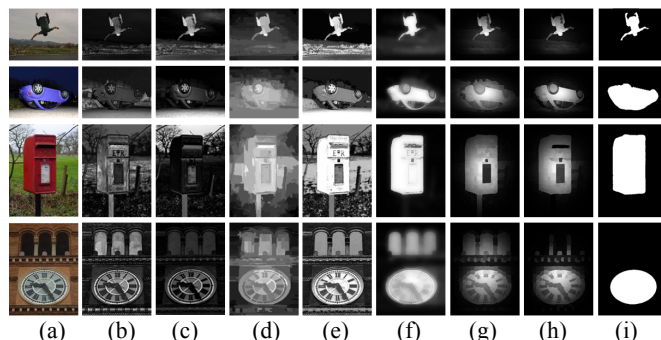


Fig. 3. Visual results of the different state-of-the-art models on the MSRA data set. (a). Input images; (b). FT[10]; (c). LC[17]; (d). RC[23]; (e). HC[7]; (f). MSS[14]; (g). FD[36]; (h).Our method; (i). GTs.

IV. CONCLUSIONS

In this paper, we propose a saliency-detection model based on the centroid of salient objects, two distinct categories of visual cues are synthetically integrated into a robust saliency map, to achieve the particular goal of accomplishing foreground reliability and background efficiency. Experimental results on an open-access database have shown that the proposed algorithm is efficient.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (NSFC) (61601427); Royal Society - K. C. Wong International Fellowship (NIF\R1\180909).

REFERENCES

- [1] M. Unser, Texture classification and segmentation using wavelet frames, *IEEE Trans. Image Process.* 4 (11) (1995) 1549-1560.
- [2] L. Itti, C. Koch, E. Niebur, A model of saliency based visual attention for rapid scene analysis, *IEEE Trans. PAMI.* 20 (11) (1998) 1254-1259.
- [3] T. Liu, J. Sun, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," in *IEEE CVPR*, 2007, pp. 1-8.
- [4] Y. Fang, W. Lin, Z. Fang, Z. Chen, C. W. Lin, C. Deng, Visual acuity inspired saliency detection by using sparse features, *Information Sciences*, Volume 309, 2015, pp. 1-10.
- [5] N. OTSU. A threshold selection method from gray level histograms. *IEEE Trans. Syst. Man Cybern.*, 9 (1) (1979), pp. 62-66.
- [6] http://research.microsoft.com/en-us/um/people/jiansun/SalientObject/salient_object.htm
- [7] M. Cheng, G. Zhang, N. Mitra, X. Huang, S. Hu. Global Contrast based Salient Region Detection, *IEEE CVPR*, pp. 409-416, 2011.
- [8] A. Toet, Computational versus psychophysical bottom-up image saliency: a comparative evaluation study, *IEEE Trans. PAMI.* 33 (11), 2011, 2131-2146.
- [9] A. Borji, D.N. Sihite, L. Itti, Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study, *IEEE Trans. Image Process.* 22 (1) (2013) 55-69.
- [10] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk. Frequency-tuned salient region detection, in *IEEE CVPR*, 2009, pp. 1597-1604.
- [11] Q. Wang, Y. Yuan, P. Yan, X. Li, Visual saliency by selective contrast, *IEEE Trans. CSVT.* 23 (7), (2013) 1150-1155.

- [12] Q. Wang, Y. Yuan, P. Yan, X. Li, Saliency detection by multiple-instance learning, *IEEE Trans. Cybern.* 43 (2) (2013) 660-672.
- [13] C. Yang, L. Zhang, H. Lu, Graph-regularized saliency detection with convex-hull-based center prior, *IEEE SPL*, 20 (7) (2013) 637-640.
- [14] N. Tong, H. Lu, L. Zhang, X. Ran, Saliency detection with multi-scale superpixels, *IEEE Signal Process. Lett.* 21 (9) (2014) 1035-1039.
- [15] W. Zhu, S. Liang, Y. Wei, J. Sun, Saliency optimization from robust background detection, in: *IEEE CVPR*, 2014, pp. 2814-2821.
- [16] Y.F. Ma, H.J. Zhang, Contrast-based image attention analysis by using fuzzy growing, in: *ACM MM*, 2003, pp. 374-381.
- [17] Y. Zhai, M. Shah. Visual attention detection in video sequences using spatiotemporal cues, in: *ACM MM*, 2006, pp. 815-824.
- [18] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, *Adv. Neural Inf. Process. Syst.* (2006) 545-552.
- [19] X. Hou, L. Zhang, Saliency detection: a spectral residual approach, in: *IEEE CVPR*, 2007, pp. 1-8.
- [20] E. Rahtu, J. Kannala, M. Salo, J. Heikkilä, Segmenting salient objects from images and videos, in: *European Conference on Computer Vision*, 2010, pp. 366-379.
- [21] N. Murray, M. Vanrell, X. Otazu, C. A. Parraga, Saliency estimation using a non-parametric low-level vision model, in: *IEEE CVPR*, 2011, pp. 433-440.
- [22] S. Goferman, L. Zelnik-Manor, A. Tal, Context-aware saliency detection, *IEEE Trans. IEEE CVPR.* 34 (10) (2012) 1915-1926.
- [23] M. Cheng, N.J. Mitra, X. Huang, S. Hu, Global contrast based salient region detection, *IEEE Trans. PAMI.* 37 (3) (2015) 569-582.
- [24] Y. Qin, H. Lu, Y. Xu, H. Wang, Saliency detection via cellular automata, in: *IEEE CVPR*, 2015, pp. 110-119.
- [25] Z. Chen, H. Wang, L. Zhang, Y. Yan, H. M. Liao, Visual saliency detection based on homology similarity and an experimental evaluation, *J. Visual Commun. Image Represent.* 40 (2016) 251-264.
- [26] Y. Kong, L. Wang, X. Liu, H. Lu, X. Ruan, Pattern mining saliency, in: *ECCV*, vol. 9910, 2016, pp. 583-598.
- [27] M. Jian, Q. Qi, J. Dong, X. Sun, Y. Sun, K. M. Lam, Saliency detection using quaternionic distance based weber local descriptor and level priors, *Multimedia Tools Appl.* (2017) 1-18.
- [28] Z. Li, C. Lang, S. Feng, T. Wang, Saliency ranker: a new salient object detection method, *J. Visual Commun. Image Represent.* 50 (2018) 16-26.
- [29] M. Jian, Q. Qi, J. Dong, X. Sun, K. M. Lam, Integrating QDWD with pattern distinctness and local contrast for underwater saliency detection, *J. Visual Commun. Image Represent.* 53 (2018) 31-41.
- [30] M. Jian, Q. Qi, H. Yu, J. Dong, C. Cui, X. Nie, H. Zhang, Y. Yin, K. M. Lam, The extended marine underwater environment database and baseline evaluations, *Applied Soft Computing*, Vol. 80, pp. 425-437, 2019..
- [31] A. Oliva, A. Torralba, M. S. Castelhana, J. M. Henderson, Top-down control of visual attention in object detection, in: *IEEE ICIP*, vol. 1, 2003.
- [32] D. Tao, K. Yang, Y. Li, H. Yan, Where does the driver look? top-down-based saliency detection in a traffic driving environment. *IEEE Trans. on Intell. Transport. Systems*, 17(7), 2051-2062, 2016.
- [33] H. Cholakkal, D. Rajan, J. Johnson, Top-down saliency with locality-constrained contextual sparse coding, *BMVC* (2015).
- [34] J. Yang, M.H. Yang, Top-down visual saliency via joint crf and dictionary learning, *IEEE Trans. PAMI.* 39 (3) (2017) 576-588.
- [35] M. Jian, W. Zhang, H. Yu, C. Cui, X. Nie, H. Zhang, Y. Yin, Saliency detection based on directional patches extraction and principal local colour contrast, *Journal of Visual Communication and Image Representation*, 57 (2018): 1-11.
- [36] J. Wang, H. Lu, X. Li, N. Tong, W. Liu, Saliency detection via background and foreground seed selection, *Neurocomputing*, 152 (2015): 359-368.
- [37] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Sabine, SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. PAMI*, 2012, 34(11): 2274-2282.
- [38] D. R. Martin, C. C. Fowlkes, J. Malik, Learning to detect natural image boundaries using local brightness, colour, and texture cues. *IEEE Trans. PAMI.* 26 (5) (2004), pp. 530-549.
- [39] Y. Xie, H. Lu, Visual saliency detection based on Bayesian model, in: *IEEE ICIP*, 2011, pp. 645-648.