An Iterative Co-Saliency Framework for RGBD Images

Runmin Cong, Jianjun Lei, *Senior Member, IEEE*, Huazhu Fu, Weisi Lin, *Fellow, IEEE*, Qingming Huang, *Senior Member, IEEE*, Xiaochun Cao, *Senior Member, IEEE*, and Chunping Hou

***Abstract*—As a newly emerging and significant topic in computer vision community, co-saliency detection aims at discovering the common salient objects in multiple related images. The existing methods often generate the co-saliency map through a direct forward pipeline which is based on the designed cues or initialization, but lack the refinement-cycle scheme. Moreover, they mainly focus on RGB image and ignore the depth information for RGBD images. In this paper, we propose an iterative RGBD co-saliency framework, which utilizes the existing single saliency maps as the initialization, and generates the final RGBD co-saliency map by using a refinement-cycle model. Three schemes are employed in the proposed RGBD co-saliency framework, which include the addition scheme, deletion scheme, and iteration scheme. The addition scheme is used to highlight the salient regions based on intra-image depth propagation and saliency propagation, while the deletion scheme filters the saliency regions and removes the non-common salient regions based on interimage constraint. The iteration scheme is proposed to obtain more homogeneous and consistent co-saliency map. Furthermore, a novel descriptor, named depth shape prior, is proposed in the addition scheme to introduce the depth information to enhance identification of co-salient objects. The proposed method can effectively exploit any existing 2-D saliency model to work well in RGBD co-saliency scenarios. The experiments on two RGBD co-saliency datasets demonstrate the effectiveness of our proposed framework.**

***Index Terms*—Common probability, depth shape prior (DSP), iterative optimization, RGBD co-saliency framework, three schemes.**

I. Introduction

HUMAN visual system serves as a filter for selecting the salient and interesting regions for further processing. In computer vision community, saliency detection methods are proposed to simulate this characteristic of early primate visual system, which aim at capturing the most salient and informative regions in an image, and have been applied in a wide range of vision applications, such as image retrieval [1], sensation enhancement [2], [3], foreground annotation [4], image segmentation [5], [6], image quality assessment [7], image retargeting [8], [9], and coding [10]. Numerous saliency detection models [11]–[17] focus on detecting the salient objects from a single image, which achieve encouraging performance on the public benchmarks.

As an emerging and challenging issue, co-saliency detection has been attracting more attentions in recent years. Different from the traditional single saliency detection model, co-saliency detection methods focus on discovering the common salient objects in multiple images. Based on the definition of co-saliency [18]–[22], two main properties of the co-salient object should be owned simultaneously: 1) the target objects should be salient in individual image and 2) all co-salient objects should be common among multiple images. In fact, the categories, intrinsic characteristics and locations of these objects are entirely unknown, and the background of the scenes is different. Fig. 1 provides an example of co-saliency detection. The two dogs in the last three images are salient in the single image saliency model. However, from the perspective of the image group, only the black dog is the common salient object in this image group. Therefore, co-saliency detection is a more challenging issue compared to saliency detection, while the extracted co-salient regions are more useful in many computer vision tasks, such as object co-localization [23], [24], image matching [25], foreground co-segmentation [26], [27], and co-detection [28]. Most existing co-saliency detection models [29]–[37] focus on designing the complete and independent algorithms to discover the common salient regions and obtaining the satisfactory performances.

Manuscript received January 31, 2017; revised August 1, 2017; accepted October 22, 2017. Date of publication November 21, 2017; date of current version December 14, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61722112, Grant 61520106002, Grant 61731003, Grant 61332016, Grant 61620106009, Grant U1636214, and Grant 61602344, and in part by the National Key Research and Development Program of China under Grant 2017YFB1002900. This paper was recommended by Associate Editor H. Lu. *(Corresponding author: Jianjun Lei.)*

R. Cong is with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China, and also with the School of Computer Engineering, Nanyang Technological University, Singapore 639798 (e-mail: rmcong@tju.edu.cn).

J. Lei and C. Hou are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: jjlei@tju.edu.cn; hcp@tju.edu.cn).

H. Fu is with the Ocular Imaging Department, Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore 138632 (e-mail: huazhufu@gmail.com).

W. Lin is with the School of Computer Engineering, Nanyang Technological University, Singapore 639798 (e-mail: wslin@ntu.edu.sg).

Q. Huang is with the School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 100190, China (e-mail: qmhuang@ucas.ac.cn).

X. Cao is with the State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China, and also with the School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 100190, China (e-mail: caoxiaochun@iie.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2017.2771488

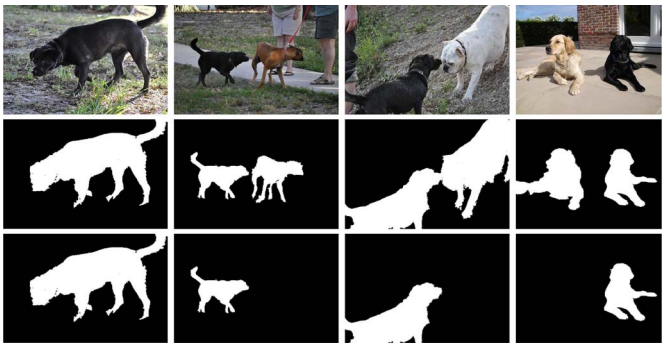


Fig. 1. Example of the co-saliency detection. The first row is the input RGB images in an image group, the second row shows some saliency detection results, and the third row represents some co-saliency detection results. From the perspective of the image group, only the black dog is the common salient object in the co-saliency detection model.

In fact, the single saliency map produced by the existing saliency model can be considered as an initialization in co-saliency detection. Moreover, the existing co-saliency detection methods mainly rely on the designed cues or initialization, and lack the refinement-cycle. In this paper, we propose an effective co-saliency framework based on the refinement-cycle model, which integrates the addition scheme, deletion scheme, and iteration scheme. The addition scheme is used to enrich the saliency regions through the depth propagation and saliency propagation. The inter saliency model is formalized as common probability calculation to capture the inter-image correspondence in the deletion scheme. Moreover, the iterative optimization scheme is designed to achieve more superior co-saliency result in our framework.

In addition, the depth information from the RGBD image has been demonstrated the usefulness for many computer vision tasks [38], [39], and has been introduced into many saliency models to enhance the detection performance [40]–[45]. Niu *et al.* [40] used the global disparity contrast and domain knowledge to capture the depth information. Peng *et al.* [41] calculated the depth saliency from multicontextual contrast including the local context, global context and background context. Feng *et al.* [43] proposed a local background enclosure (LBE) feature to evaluate the depth saliency. However, in the above methods, the information of the relevant and similar objects in a sequence of images is ignored and not exploited. In this paper, a novel depth descriptor, named depth shape prior (DSP), is proposed to capture the shape attributes from the depth map to improve the co-saliency detection performance.

In summary, most of existing co-saliency methods aim to design a single forward pipeline which generates the co-saliency map based on the designed cues directly, but lack the refinement-cycle scheme and ignore the depth information for RGBD images. Thus, in this paper, we propose an iterative RGBD co-saliency framework, which utilizes the additional depth information and employs the existing RGB saliency map as the initialization in a refinement-cycle model to produce the final RGBD co-saliency map. The major contributions of the proposed method are as follows.

1) An iterative co-saliency framework for RGBD images is proposed, which integrates addition scheme, deletion scheme, and iteration scheme. The intra-image propagations and the inter-image constraints are incorporated into a cyclic model to achieve the co-saliency detection.
2) A novel depth descriptor, named DSP, is proposed to capture the shape attributes from the depth map and enhance the identification of co-salient objects from RGBD images.
3) A superpixel-level common probability function among multiple images is calculated to exploit the inter-image corresponding relationship in the deletion scheme.
4) An iterative updating strategy is designed to obtain more homogeneous and consistent co-saliency result in the iteration scheme.

The rest of this paper is organized as follows. Section II reviews the related works of saliency and co-saliency detection briefly. Section III details the proposed RGBD co-saliency detection framework. The experimental results and analysis with quantitative evaluation are presented in Section IV. Finally, the conclusion is provided in Section V.

II. Related Work

In the last decade, a number of methods [11]–[17], [46]–[50] have been presented to identify the salient regions from an RGB image, which can be applied in many computer vision tasks. Cheng *et al.* [11] proposed a regional contrast-based saliency detection algorithm, which simultaneously evaluates global contrast differences and spatial weighted coherence scores. Li *et al.* [13] proposed a novel graph-based saliency detection algorithm, which formulates the pixel-wised saliency maps using the regularized random walks ranking. Shi *et al.* [15] designed a hierarchical saliency model to compute the saliency cues using weighted color contrast on three image layers. More recently, the theory of deep learning has been applied in saliency detection, and has achieved remarkable performance. Chen *et al.* [47] built a saliency model with two stacked convolutional neural networks (CNNs), and the coarse-level and fine-level representation learning are utilized to learn the saliency representation in a progressive manner. Lee *et al.* [49] integrated the hand-crafted features and high-level features into the saliency model to enhance performance of saliency detection. To address the blurry boundary of the salient object, Li and Yu [50] proposed an end-to-end deep contrast network, which includes pixel-level fully convolutional stream and segment-wise spatial pooling stream.

In addition, human can perceive the surroundings by an additional depth cue that is provided by stereopsis, which plays an important role in the human visual system. The depth information has demonstrated its usefulness for many computer vision tasks [27], [41], [51]. Some saliency detection methods integrated with depth information, named RGBD saliency, are proposed in recent years. In [41], considering low-level feature

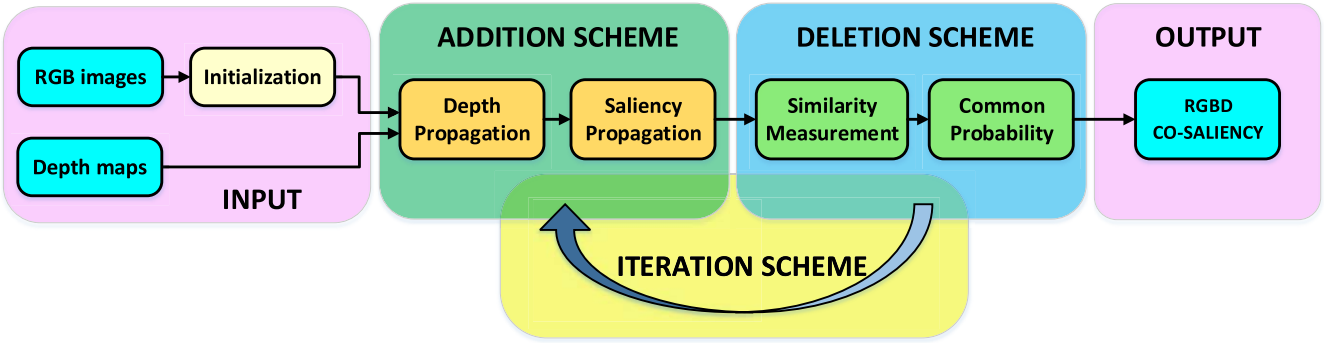


Fig. 2. Flowchart of the proposed RGBD co-saliency framework.

contrast, mid-level region grouping and high-level object-aware priors, Peng *et al.* proposed a multistage RGBD model to discover the salient regions. Ju *et al.* [42] proposed a depth-aware anisotropic center-surround difference measurement to evaluate the depth saliency. Feng *et al.* [43] used the LBE feature to compute the RGBD saliency. Considering the quality of depth map, Cong *et al.* [44] proposed an RGBD saliency model by combining depth confidence analysis and multiple cues fusion.

Different from modeling the human visual mechanism from a single image, co-saliency detection makes effort to discover the common salient objects when people are reviewing a group of related images. There are many methods have been proposed to achieve the co-saliency detection. In [20], the single-image and multi-image saliency maps are linearly combined to detect the co-salient regions. However, this method is only available for image pairs. In [22], a cluster-level method that integrates five saliency cues is proposed to compute the inter saliency. Liu *et al.* [29] proposed a co-saliency model based on hierarchical segmentation, which includes the fine-level and coarse-level segmentation. In [30], co-saliency detection is formulated as a two-stage saliency propagation problem that uses the single-image saliency map to propagate pairwise saliency values. Cao *et al.* [33] proposed a general fusion framework for saliency and co-saliency detection, which introduces the self-adaptively weighted scheme via rank constraint. Li *et al.* [34] proposed a two-stage guided scheme to obtain the guided saliency map for each single image through ranking framework, and the final co-saliency map is produced by fusing these guided saliency maps. To detect the salient objects from complex natural scenes with small-scale high-contrast backgrounds, Huang *et al.* [35] presented a saliency detection method via multiscale low-rank analysis, and introduced a GMM-based co-saliency detection method. Recently, the learning-based co-saliency models are drawing more attention from researchers due to its superior performance. Zhang *et al.* [36] exploited CNN with additional transfer layers to generate the higher-level features, and these features are used to discover the co-salient objects via Bayesian framework. In [37], the co-saliency detection is formulated under a multiple-instance learning (MIL) framework, and self-paced learning regime is introduced into the MIL framework for selecting training samples in a theoretically sound manner.

However, all the above-mentioned co-saliency models focus on RGB images, and ignore the effectiveness of depth information to enhance the identification of co-salient objects from the cluttered and changing natural scene. In [56], an RGBD co-saliency model using bagging-based clustering is proposed. This paper is different with method in [56].

1) The motivation is different. The method in [56] utilized the bagging-based clustering in a single forward pipeline, which proceeds from the RGBD saliency map. By contrast, this paper aims at designing an iterative co-saliency framework, which includes addition, deletion, and iteration schemes to achieve the co-saliency detection. Moreover, we employ the 2-D saliency map as the initial map, and incorporate the additional depth information in the model. Thus, our method can convert any 2-D saliency map into the RGBD co-saliency map.
2) The use of depth information is different. The method in [56] focuses on the basic depth distribution features, e.g., depth value, depth range, and HOG histogram on the depth map. By contrast, a novel depth descriptor named DSP is proposed in our method, which captures the shape attributes from the depth map and improves the performance of the co-saliency detection by using the depth consistency and shape attributes.

III. Proposed Method

The proposed RGBD co-saliency framework is introduced in this section. Fig. 2 shows the framework of the proposed method. Our method is initialized by the existing 2-D saliency maps, and then three schemes are employed to generate the final RGBD co-saliency map. The addition scheme is used to grow the initialized saliency map from the perspective of intra-image, the deletion scheme is designed to suppress the non-common regions from the perspective of inter-image, and the iteration scheme is exploited to obtain more homogeneous and consistent co-saliency map. Some visual examples of the proposed method are shown in Fig. 3.

Notations: Given N input RGB images $\{I^i\}_{i=1}^N$, and the corresponding depth maps are denoted as $\{D^i\}_{i=1}^N$. The M_i single saliency maps for image I^i produced by existing single image saliency models are represented as $S^i = \{S_j^i\}_{j=1}^{M_i}$. In our method,

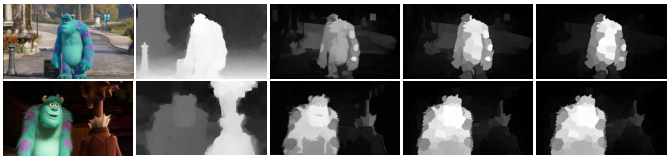


Fig. 3. Some examples of the proposed method. (a) RGB image. (b) Depth map. (c) Initialized saliency map. (d) Co-saliency map without iteration. (e) Final co-saliency map with iteration.

the superpixel-level region is regarded as the basic unit for processing. Thus, each RGB image I^i is abstracted into superpixels $R^i = \{r_m^i\}_{m=1}^{N_i}$ using SLIC algorithm [52] first, where N_i is the number of superpixels for image I^i.

A. Initialization

The proposed co-saliency framework aims at discovering the co-salient objects from multiple images in a group with the assistance of existing 2-D saliency maps. Therefore, some existing saliency maps produced by 2-D saliency models are used to initialize the framework. It is well known that different saliency methods own different superiority in detecting salient regions. In a way, these saliency maps are complementary in some regions, thus, the fused result can inherit the merits of the multiple saliency maps, and produce more robust and superior detection baseline. In our method, the simple average function is used to achieve a more generalized initialization result. The initialized saliency map for image I^i is denoted as

$$S_f^i(r_m^i) = \frac{1}{M_i}\sum_{j=1}^{M_i} S_j^i(r_m^i) \tag{1}$$

where $S_j^i(r_m^i)$ denotes the saliency value of superpixel r_m^i produced by jth saliency method for image I^i, and M_i is the number of saliency maps for image I^i. In our experiments, five saliency methods including RC [11], DCLC [12], RRWR [13], HS [15], and BSCA [16], are used to produce the 2-D initialized saliency map. Some examples of the initialized saliency map are shown in Fig. 3(c). From the figures, we can see that the initialized result produces an impressive baseline for later co-saliency detection.

B. Addition Scheme

In this section, the addition scheme is designed to extend the saliency region based on the intra-image constraint by using two propagation algorithms. First, a novel depth descriptor, named DSP, is proposed to capture the depth cue and produce an RGBD saliency result in depth propagation. Then, saliency propagation is utilized to optimize and improve the saliency result furtherly.

1) Depth Propagation: After initialization, the merits of the different saliency maps are inherited into the initialized saliency map. The depth information is introduced into the framework to enhance the identification of salient objects due to its usefulness in saliency detection. In general, the depth map owns the following properties.

i) The salient object appears higher depth value compared to the backgrounds.
ii) The high quality depth map can provide sharp and explicit boundary of the object.
iii) The interior depth value of the object should be smoothness and consistency.

Inspired by these observations, a depth descriptor, namely DSP, is proposed to capture the shape attributes from the depth map and improve the performance of the co-saliency detection by using the depth consistency and shape attributes. The proposed DSP descriptor is based on depth propagation and region grow. Several identified superpixels are selected as the seeds first, and then the DSP map can be calculated via depth constraints.

For each image I^i, the top K superpixels with higher initialized saliency value are selected as the root seeds, which is represented as $\{r_{rk}^i\}_{k=1}^K$, and the corresponding DSP map DSP_k^i is initialized as zero.

For each root seed, we determine a set of child nodes $\{r_{cp}^i\}$ to depict the depth shape based on the depth smoothness and consistency constraints. In the l-loop diffusion, the superpixels direct neighboring the $(l-1)$-loop child nodes are selected as the l-loop child nodes only if they satisfy the following two constraints.

i) *Depth Smoothness:* The depth difference between the neighbor superpixel and $(l-1)$-loop child seeds is less than a certain threshold T_1, as $|d_{nq}^i - d_{c,l-1}^i| \le T_1$, where d_{nq}^i is the depth value of the neighbor superpixel r_{nq}^i, and $d_{c,l-1}^i$ is the average depth value of $(l-1)$-loop child seeds.
ii) *Depth Consistency:* The depth difference between the neighbor superpixel and root seed should be smaller than a specific threshold T_2, as $|d_{nq}^i - d_{rk}^i| \le T_2$, where d_{rk}^i is the depth value of the root seed r_{rk}^i.

Be noted that the child node in the first loop diffusion is initialized by the root seed in our method, and the two thresholds are set to 0.1 and 0.2, respectively. The DSP value of the child node r_{cp}^i in the l-loop is defined as

$$\mathrm{DSP}_k^i\left(r_{cp}^i\right) = 1 - \min\left(\left|d_{cp,l}^i - d_{c,l-1}^i\right|, \left|d_{cp,l}^i - d_{rk}^i\right|\right) \tag{2}$$

where $d_{cp,l}^i$ is the depth value of the child node r_{cp}^i in the l-loop, $d_{c,l-1}^i$ is the average depth value of $(l-1)$-loop child node set, and $|\cdot|$ is the absolute value function. Then, the next loop diffusion will be continued until there is no neighboring superpixel satisfies the depth constraints.

In our method, the top K root seeds are selected for each image I^i to improve the robustness, and K DSP value maps are obtained for each image. Therefore, the final DSP map is defined as

$$\mathrm{DSP}^i(r_m^i) = \frac{1}{K}\sum_{k=1}^{K} \mathrm{DSP}_k^i(r_m^i) \tag{3}$$

where K is the number of the root seeds, which is fixed to 10 in all the experiments.

To achieve more superior and stable saliency result, the initialized RGB saliency and the DSP map are combined in

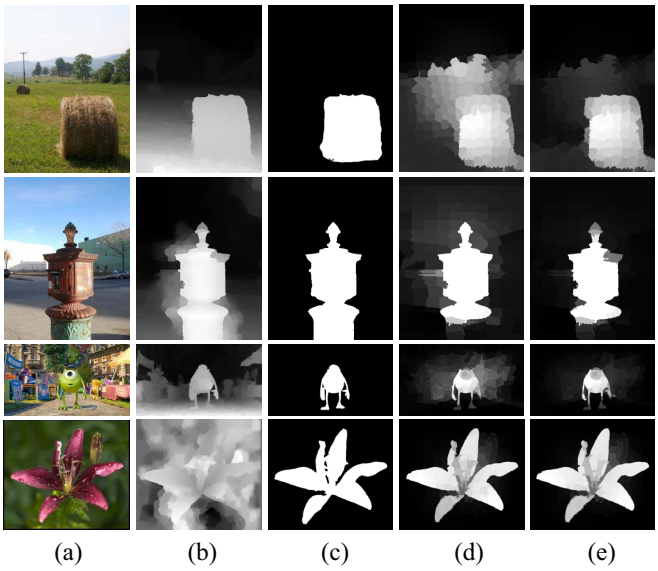


Fig. 4. Examples of the depth propagation. (a) RGB image. (b) Depth map. (c) Ground truth. (d) RGB saliency result. (e) RGBD saliency result with DSP descriptor.

our method. Because the bad depth map may degenerate the accuracy of DSP map generation, we introduce the depth confidence measure λ_d to evaluate the quality of the depth information, which is defined as [44]

$$\lambda_d = \exp((1 - m_d) \cdot \mathrm{CV} \cdot H) - 1 \tag{4}$$

where m_d is the mean value of the whole depth image, CV denotes coefficient of variation, and H is the depth frequency entropy, which denotes the randomness of depth distribution. A larger λ_d value represents greater reliability of the input depth map. More details of depth confidence measure can be found in [44]. Thus, the RGBD saliency that integrates initialized saliency map and DSP map weighted by depth confidence measure according to the depth quality is defined as

$$S^i_{dp}(r^i_m) = (1 - \lambda^i_d) \cdot S^i_f(r^i_m) + \lambda^i_d \cdot S^i_f(r^i_m) \cdot \mathrm{DSP}^i(r^i_m) \tag{5}$$

where λ^i_d is the depth confidence measure for image I^i, $S^i_f(r^i_m)$ represents the initialized saliency value of superpixel r^i_m, and $\mathrm{DSP}^i(r^i_m)$ is the DSP value of superpixel r^i_m. The obtained saliency map is normalized into [0, 1]. With this depth confidence measure, the poor-quality depth map will be limited in the combination with RGB feature to avoid the degradation of the RGBD co-saliency result. Fig. 4 shows some examples of the depth propagation. Comparing with the RGB saliency maps, some background regions around the salient object are suppressed effectively through depth propagation, such as the lawns in the first image, the roads in the second image, and the buildings in the third image. Moreover, the RGBD saliency model is more robust. Even if the quality of depth map is bad, such as the last row in Fig. 4, our model still achieves better result by highlighting the RGB saliency component while DSP descriptor cannot exploit accurate shape attributes from the poor-quality depth map.

2) Saliency Propagation: With the obtained RGBD saliency map, the saliency propagation is conducted to further optimize the result. In our method, the superpixels are classified into three groups based on the saliency value first, which is denoted as the saliency seed superpixels, background seed superpixels, and the unknown superpixels. Then, saliency propagation is used to propagate the saliency of unknown superpixels on the graph from the saliency and background seeds.

For image I^i, a graph $G^i = (\vartheta^i, \varepsilon^i)$ among superpixels is constructed first, where ϑ^i denotes the node set which corresponds to the superpixels, and ε^i is the link set among adjacent nodes. The affinity matrix $\boldsymbol{W}^i = [w^i_{uv}]_{N_i \times N_i}$ is defined as the similarity between two adjacent superpixels

$$w^i_{uv} = \begin{cases} \exp\left(-\frac{\|\boldsymbol{c}^i_u - \boldsymbol{c}^i_v\|_2 + \lambda^i_d \cdot |d^i_u - d^i_v|}{\sigma^2}\right), & \text{if } r^i_v \in \Omega^i_u \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $\boldsymbol{c}^i_u$ and d^i_u denote the mean Lab color and depth value of superpixel r^i_u, Ω^i_u represents the neighbor set of superpixel r^i_u, $\|\cdot\|_2$ is the two-norm of vector, λ^i_d is the depth confidence measure, and σ^2 is a constant control parameter.

In our method, the seed superpixels are selected based on RGBD saliency value produced by depth propagation. Then, the top κ superpixels with higher saliency values are considered as the saliency seeds, and the bottom κ superpixels with lower saliency values are treated as the background seeds. In our experiments, κ is set to 10. The initialized propagation score of the superpixel is defined as follows:

$$S^i_0(r^i_n) = \begin{cases} 1, & \text{if } r^i_n \in \Psi_F \\ 0, & \text{if } r^i_n \in \Psi_B \\ S^i_{dp}(r^i_n), & \text{otherwise} \end{cases} \tag{7}$$

where Ψ_F represents the saliency seed set, and Ψ_B denotes the background seed set.

Using the labeled seeds, the saliency is propagated on the graph, and the score with saliency propagation is achieved by

$$S^i_{\mathrm{sp}}(r^i_m) = \sum_{n=1}^{N_i} w^i_{mn} \cdot S^i_0(r^i_n) \tag{8}$$

where w^i_{mn} is the element of the affinity matrix.

C. Deletion Scheme

The addition scheme is used to improve and optimize the saliency map from the perspective of intra-image. On the other hand, the inter-image information plays an important role in co-saliency detection. Therefore, a deletion scheme is designed to capture the corresponding relationship among multiple images, which aims to suppress the common and non-common backgrounds, and enhance the common salient regions from the perspective of multiple images. In our deletion scheme, a superpixel-level similarity measurement is constructed to represent the similarity relationship between two superpixels. Then, a common probability function using the similarity measurement is used to calculate the likelihood of each superpixel belonging to the common regions.

1) Multiple Cues-Based Similarity Measurement: In deletion scheme, the color cue, depth cue, and saliency cue are combined into a measurement to evaluate the similarity between two superpixels.

a) RGB similarity: The color histogram and texture histogram [53], [54] are used to represent the RGB feature on the superpixel level, which are denoted as HC_m^i and HT_m^i, respectively. Then, the Chi-square measure is employed to compute the feature difference. Thus, the RGB similarity is defined as

$$S_c\left(r_m^i, r_n^j\right) = 1 - \frac{1}{2}\left[\chi^2\left(HC_m^i, HC_n^j\right) + \chi^2\left(HT_m^i, HT_n^j\right)\right] \quad (9)$$

where r_m^i and r_n^j are the superpixels in image I^i and I^j, respectively, and $\chi^2(\cdot)$ denotes the Chi-square distance function.

b) Depth similarity: Two depth consistency measurements, namely depth value consistency and depth contrast consistency, are composed of the final depth similarity measurement, which is defined as

$$S_d\left(r_m^i, r_n^j\right) = \exp\left(-\frac{W_d\left(r_m^i, r_n^j\right) + W_c\left(r_m^i, r_n^j\right)}{\sigma^2}\right) \quad (10)$$

where $W_d(r_m^i, r_n^j)$ is the depth value consistency measurement to evaluate the inter image depth consistency, due to the fact that the common regions should appear similar depth values

$$W_d\left(r_m^i, r_n^j\right) = \left|d_m^i - d_n^j\right|. \quad (11)$$

$W_c(r_m^i, r_n^j)$ describe the depth contrast consistency, because the common regions should represent more similar characteristic in depth contrast measurement

$$W_c\left(r_m^i, r_n^j\right) = \left|D_c\left(r_m^i\right) - D_c\left(r_n^j\right)\right| \quad (12)$$

with

$$D_c\left(r_m^i\right) = \sum_{k \neq m}\left|d_m^i - d_k^i\right| \exp\left(-\frac{\left\|\boldsymbol{p}_m^i - \boldsymbol{p}_k^i\right\|_2}{\sigma^2}\right) \quad (13)$$

where $D_c(r_m^i)$ denotes the depth contrast of superpixel r_m^i, $\boldsymbol{p}_m^i$ denotes the position of superpixel r_m^i, and σ^2 is a constant.

c) Saliency similarity: Inspired by the prior that the common regions should appear more similar in single saliency map compared to other regions, the output saliency map from the addition scheme is used to define the saliency similarity measurement in this paper

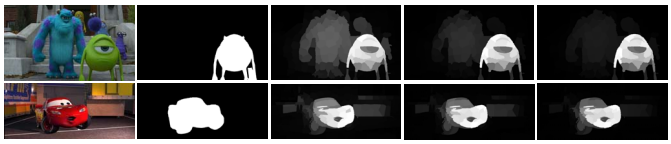
$$S_s\left(r_m^i, r_n^j\right) = \exp\left(-\left|S_{\text{sp}}^i\left(r_m^i\right) - S_{\text{sp}}^j\left(r_n^j\right)\right|\right) \quad (14)$$

where $S_{\text{sp}}^i(r_m^i)$ is the saliency score of superpixel r_m^i based on (8).

d) Combination similarity: Based on these cues, the combination similarity measurement is defined as the average of the three similarity measurements

$$S_M\left(r_m^i, r_n^j\right) = \frac{S_c\left(r_m^i, r_n^j\right) + S_d\left(r_m^i, r_n^j\right) + S_s\left(r_m^i, r_n^j\right)}{3} \quad (15)$$

where $S_c(r_m^i, r_n^j)$, $S_d(r_m^i, r_n^j)$, and $S_s(r_m^i, r_n^j)$ are the normalized RGB, depth, and saliency similarities between superpixel r_m^i and r_n^j, respectively. A larger $S_M(r_m^i, r_n^j)$ value corresponds to greater similarity between two superpixels.



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

Fig. 5. Some examples of the iteration scheme. (a) RGB image. (b) Ground truth. (c) Initial saliency map through the addition and deletion scheme. (d) Saliency map after the first iteration. (e) Final saliency result.

2) Common Probability: For co-saliency detection, it is necessary to discriminate whether the selected salient objects are common or not. Thus, how to determine the common objects is a key point for co-saliency detection. In general, the common object is defined as the object with repeated occurrence in multiple images. Based on this definition, the common probability function is used to evaluate the likelihood that a superpixel belongs to the common regions, and it is defined as the sum of maximum matching probability among different images. For each superpixel r_m^i, only the most matching superpixel r_k^j in image I^j is selected for calculation, which is denoted as

$$r_k^j = \arg\max_{n \in [1, N_j]} S_M\left(r_m^i, r_n^j\right) \quad (16)$$

where r_k^j is the most matching/similar superpixel in image I^j for superpixel r_m^i based on the maximum combination similarity score, and N_j represents the number of superpixles in image I^j.

Then, these selected superpixels from different images are used to calculate the common probability

$$P_c^i\left(r_m^i\right) = \frac{1}{N-1}\sum_{j=1, j \neq i}^{N} S_M\left(r_m^i, r_k^j\right) \quad (17)$$

where r_k^j is the most matching superpixel in image I^j for superpixel r_m^i, and N denotes the number of images in an image group. Finally, the updated co-saliency map of deletion scheme is denoted as

$$S_{\text{del}}^i\left(r_m^i\right) = S_{\text{sp}}^i\left(r_m^i\right) \cdot P_c^i\left(r_m^i\right) \quad (18)$$

where $S_{\text{sp}}^i(r_m^i)$ is the saliency score of superpixel r_m^i produced by the addition scheme. Fig. 3(d) shows the co-saliency map after addition and deletion schemes. Compared with the initialized saliency map shown in Fig. 3(c), the co-salient object appears to be more consistency and the backgrounds are effectively suppressed.

D. Iteration Scheme

In order to obtain more superior co-saliency map, an iterative scheme is designed in our framework, as shown in Fig. 2. The iterative scheme works as a refinement model to combine the addition and deletion steps and refine the co-saliency map in loop. In the iteration scheme, a heuristic termination strategy is set by checking the maximum iteration number $I_{\max}$ and the difference between two iterations. Specifically, the second termination condition is introduced to check whether the saliency result becomes stable or

Algorithm 1 Overall Framework

Input: The RGB images and depth maps in an image group.
Output: The co-saliency map for each image.

```
for each image in the group do
    Obtain the initialized saliency map using Eq. (1);
    repeat
        Conduct the addition scheme using Eqs. (2-8);
        Conduct the deletion scheme using Eqs. (9-18);
    until D_t^i ≤ ζ or t ≥ I_max
end for
```

not, which is formulated as the average difference between two iteration results

$$D_t^i = \left(\frac{1}{\Pi} \sum \left| S_{\text{del}}^i(t) - S_{\text{del}}^i(t-1) \right| \right) \leq \zeta \tag{19}$$

where $S_{\text{del}}^i(t)$ is the co-saliency map produced after the tth iteration optimization, Π represents the number of pixels in the co-saliency map, and ζ is a given threshold to determine whether the iteration should be terminated or not, which is set to 0.1 in all experiments. Until $D_t^i \leq \zeta$, the iteration will be terminated and output the final co-saliency map, otherwise, the iteration will continue. Some visual examples of the iteration scheme are shown in Fig. 5. The third column shows the original co-saliency result, and the first iteration and the final co-saliency maps are shown in the last two columns of Fig. 5. From the figure, we can see that the initial co-saliency map is improved obviously with the iteration processing. For example, the cartoon with blue hair (named Sulley) is suppressed effectively since it is not a common object in the image group. Similarly, the background regions around the red car are also suppressed through the iteration scheme. The overall framework of the proposed method is summarized in Algorithm 1.

IV. Experiments

In this section, the proposed RGBD co-saliency framework is evaluated on two RGBD co-saliency datasets. The qualitative and quantitative comparison with other state-of-the-art methods are presented. In addition, the analysis and discussion are conducted, which include the analysis of each module in the framework and the discussion of one-for-one option co-saliency framework.

A. Experimental Settings

The proposed co-saliency framework is evaluated on two RGBD benchmarks: 1) the RGBD Coseg183 dataset[1] [27] and 2) the RGBD Cosal150 dataset.[2] The RGBD Coseg183 dataset contains 183 images with pixel-level ground-truth that are distributed in 16 indoor scenes. For more comprehensive comparison and analysis, we collect 21 image sets containing totally 150 images from RGBD NJU-1985 dataset [42] for RGBD co-saliency detection, which is called RGBD Cosal150 dataset. Pixel-level ground-truth for each image is manually labeled in the dataset.

Three quantitative criteria are adopted to evaluate the co-saliency map, which include the precision-recall (PR) curve, F-measure, and area under curve (AUC) score. The precision and recall score are computed by thresholding the saliency map into a binary map, and comparing the binary map against the ground truth. The PR curve demonstrates the relationship between precision and recall of saliency map at different thresholds. F-measure [55] is defined as the weighted mean of precision P and recall R, which is denoted as

$$F_\beta = \frac{(1+\beta^2)P \times R}{\beta^2 \times P + R} \tag{20}$$

where β^2 is set to 0.3 that emphasizes the precision more than recall. In addition, AUC evaluates the object detection performance, which is computed as the area under the standard ROC curve. In the proposed method, the number of superpixels for each image is set to 200, the maximum iteration number is set to 5 for balancing the computational complexity and performance, and the method is implemented in MATLAB 2014a on a Quad Core 3.5-GHz workstation with 16-GB RAM. The proposed method costs average 42.67 s to process one image. The project is available on our website.[3]

B. Comparison With State-of-the-Art Methods

In this section, we compare the proposed method with eight state-of-the-art methods, which include RC [11], DCLC [12], RRWR [13], HS [15], BSCA [16], CCS [22], SCS [34], and LRMF [35]. The first five single image saliency methods are regarded as the input of the proposed framework, and the last three methods are the state-of-the-art co-saliency methods.

For subjective evaluation, the visual examples on two datasets are shown in Figs. 6 and 7, which consist of three image groups on RGBD Cosal150 dataset, i.e., the group of cartoon named Mike, red car, and statue, as well as three groups on RGBD Coseg183 dataset, i.e., the group of white cap, computer, and red flashlight. From Fig. 6, we can see that the single image saliency methods (e.g., RC, HS, and RRWR) fail to discover the co-salient objects effectively and accurately. Taking the group Mike as an example, the common salient object is the green cartoon with big eye. However, many non-common objects, such as the cartoon with blue hair and the purple snake, are detected as the salient objects in the single saliency models. In addition, some background regions are not effectively suppressed in groups red car and statue, such as the trees and nonsalient cars. In a word, the single saliency detection methods fail to detect the common salient objects in co-saliency scenarios. Therefore, it is essential that a co-saliency framework should be designed to convert the single saliency map into co-saliency result. The co-saliency map produced by our framework is shown in the last row of Fig. 6. Compared with the single saliency maps, the common salient regions are highlighted more consistent and accurate, and the backgrounds are suppressed effectively.

[1] http://hzfu.github.io/proj_rgbdseg.html
[2] https://rmcong.github.io/proj_RGBD_cosal.html
[3] https://rmcong.github.io/proj_RGBD_cosal_tcyb.html

Fig. 6. Visual comparison of different saliency and co-saliency detection methods on RGBD Cosal150 dataset.

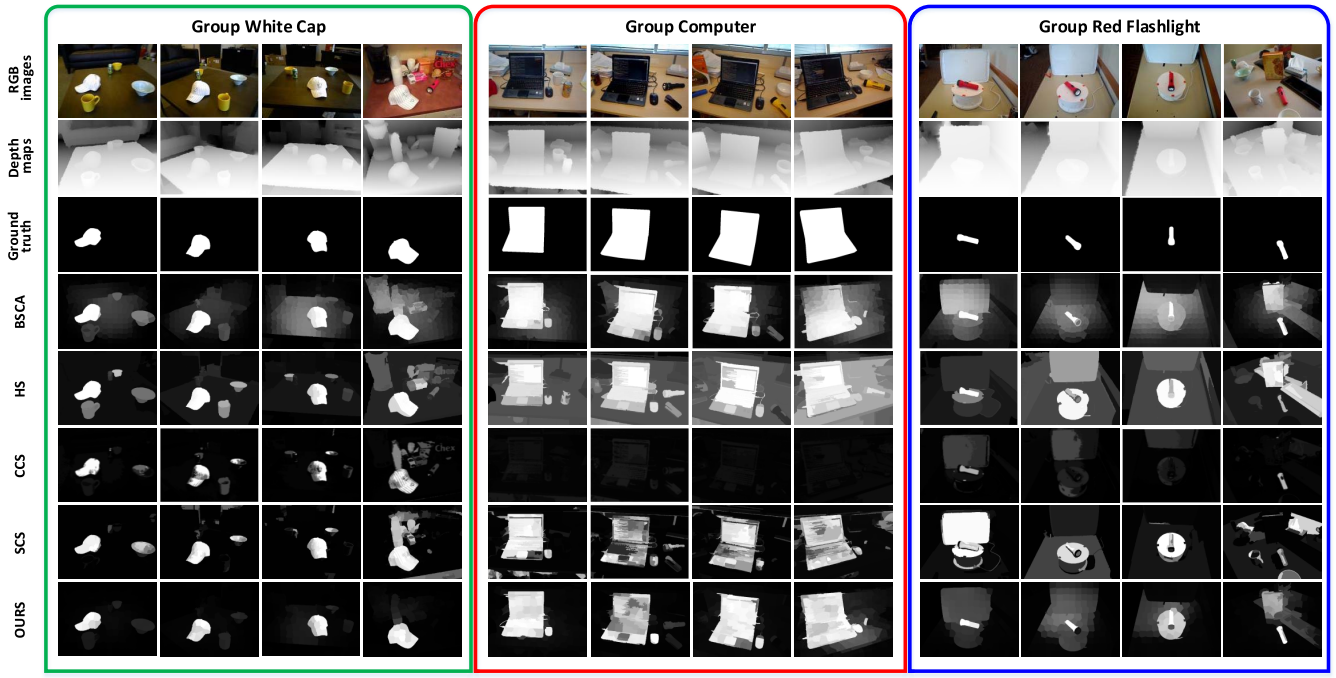


Fig. 7. Visual comparison of different saliency and co-saliency detection methods on RGBD Coseg183 dataset.

To further evaluate the proposed method, three state-of-the-art co-saliency methods are introduced for comparison. From the figures, it indicates that the proposed approach can effectively highlight the common salient regions from the image group, and robustly suppress the background regions even when the salient regions exhibit large variations in shape and direction or the background is very complex and interferential. In contrast, the RGBD Coseg183 dataset is more difficult and challenging for co-saliency detection, and some visual examples are shown in Fig. 7. The proposed method achieves better performance compared with the other saliency and co-saliency detection methods. For example, the non-common objects, e.g., the white bowl and yellow cup, are effectively suppressed in the white cap group compared to other methods. Moreover, in computer group, the consistency and homogeneity of the salient object is improved obviously compared with others. In red flashlight group, the red flashlight using our method is highlighted more effective than others. However,

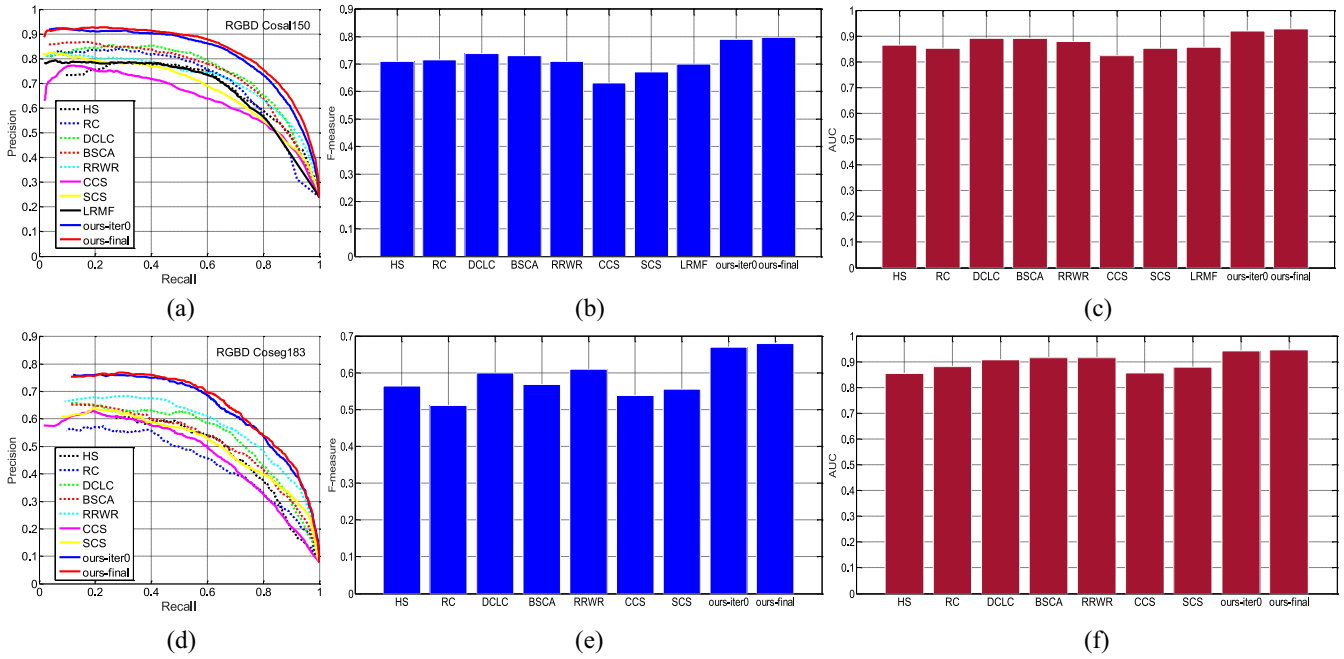


Fig. 8. Quantitative comparisons between the proposed method and the state-of-the-art methods on two datasets. Notice that "ours-iter0" means the co-saliency without iteration scheme, and "ours-final" denotes the co-saliency result with iteration scheme. (a)–(c) PR curves, F-measure, and AUC scores on RGBD Cosal150 dataset. (d)–(f) PR curves, F-measure, and AUC scores on RGBD Coseg183 dataset.

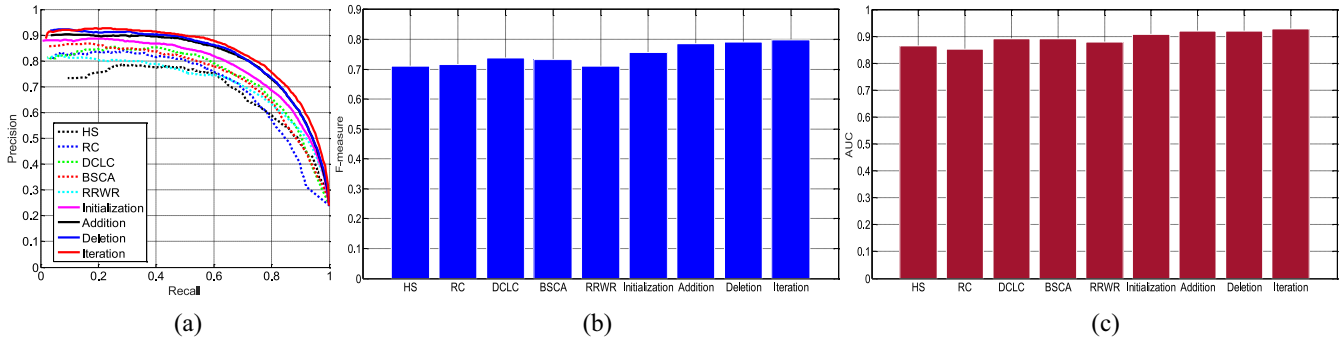


Fig. 9. Quantitative comparisons of the each part of the proposed framework on RGBD Cosal150 dataset. (a) PR curves. (b) F-measure. (c) AUC scores.

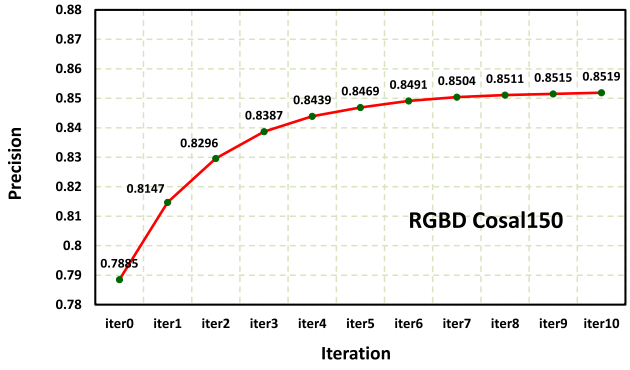


Fig. 10. Average precision of each iteration on RGBD Cosal150 dataset.

some backgrounds are still retained in the final result due to the small size and complex scene.

The quantitative comparison results including the PR curves, F-measure, and AUC scores are reported in Fig. 8. As can be seen, on the RGBD Cosal150 dataset, the proposed method achieves the highest precisions of the whole PR curves, the largest F-measure and AUC score compared with other methods. The same conclusion can be drawn from the results on RGBD Coseg183 dataset. From the PR curves on both two datasets, it can be seen that the final co-saliency result (the red line) reaches the highest level in all curves, and the performance of co-saliency framework is obviously superior to the five original single image saliency models. It also demonstrates that the proposed co-saliency framework achieves the goal of converting the single saliency results into co-saliency scenarios. The F-measure and AUC scores also support the conclusion. In the proposed RGBD co-saliency detection framework, we aim to design a many-for-one structure, i.e., multiple single saliency maps input and one co-saliency map output, to synthesize the superiority of different single saliency maps. In order to prove the effectiveness and versatility of the proposed algorithm, another one-for-one option is also implemented and evaluated, and the relevant results will be discussed in Section IV-E.

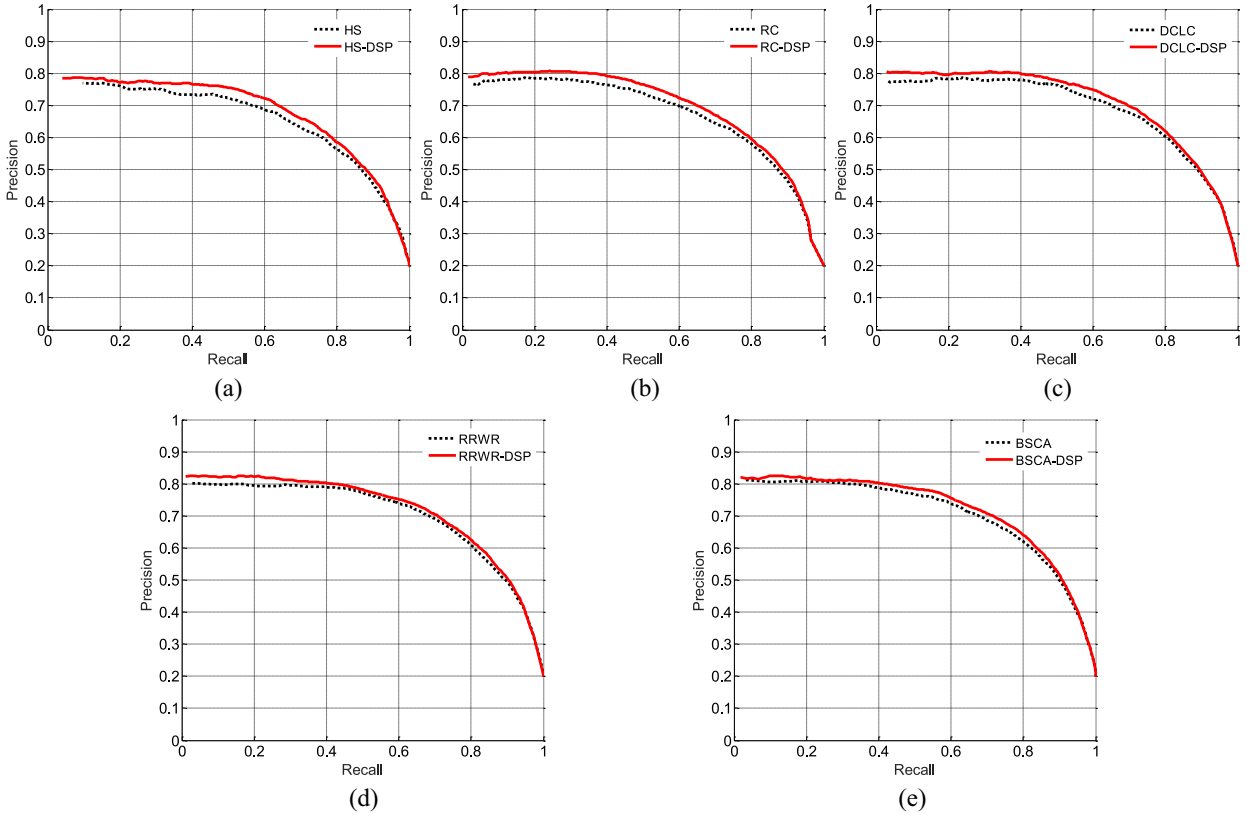


Fig. 11. Quantitative evaluation of the DSP on RGBD400 dataset. The black line in each PR curve denotes the original RGB saliency result, and the red line represents the saliency result with DSP. The PR curves for (a) HS, (b) RC, (c) DCLC, (d) RRWR, and (e) BSCA methods.

C. Module Analysis

In this section, we comprehensively evaluate each module of the proposed framework including the initialization, addition scheme, deletion scheme, and iteration scheme. The quantitative comparisons on RGBD Cosal150 dataset are shown in Fig. 9, and the evaluation result of the iteration scheme is represented in Fig. 10. In the initialization process, the original five saliency maps are integrated to produce a baseline for co-saliency detection, and its PR curve is marked as carmine in Fig. 9(a). Compared with the PR curves and F-measure of the original single saliency results, it indicates that the initialization result achieves better performance and produces a preferable baseline for later co-saliency detection. In addition scheme, DSP is proposed to introduce the depth information into the framework, and label propagation is used to further optimize the saliency result. As shown in Fig. 9, the saliency map through the addition scheme (the black line in PR curves) is improved significantly compared to the initialization result (the carmine line), and the F-measure and AUC score also achieve higher scores. Then, the deletion scheme is conducted to introduce the inter-image corresponding information into the framework and produce the initial co-saliency map. All the quantitative measurements in Fig. 9 show that the initial co-saliency result without iteration scheme obtains the best performance compared to other modules.

In addition, an iteration scheme is designed to further update the co-saliency map and achieve more consistent result. The PR curves in Fig. 9(a) demonstrate that the performance is obviously improved using the iteration scheme, in which the blue line denotes the initial co-saliency result and the red line represents the final co-saliency result with iteration scheme. With the conduct of the iteration scheme, the performance of co-saliency detection is continually optimized according to the F-measure and AUC score. In other words, all of these measurements illustrate the effectiveness of the iterative mechanism proposed in our framework. In order to verify the rationality of the iteration termination condition, an experiment of ten iterations without termination conditions are conducted on RGBD Cosal150 dataset, and the detailed quantitative comparison results are shown in Fig. 10. From the average precision curve, we can see that, with the iterative progress, the performance of the algorithm tends to be stable gradually. In general, the termination conditions will not be satisfied after the first iteration, and its improved level is most noticeable. Moreover, most of the images will satisfy the termination condition after 3–4 iterations, that is, the co-saliency map no longer appears obvious changes. Thus, it also demonstrates that the maximum iteration number of 5 is reasonable in the experiments.

D. Evaluation of the Depth Shape Prior

In the framework, a novel depth descriptor, namely DSP, is proposed to introduce the depth information to assist the identification of the co-salient objects. Introducing the DSP into RGB saliency model, the 2-D saliency model will turn into an RGBD saliency model and achieve a better performance.

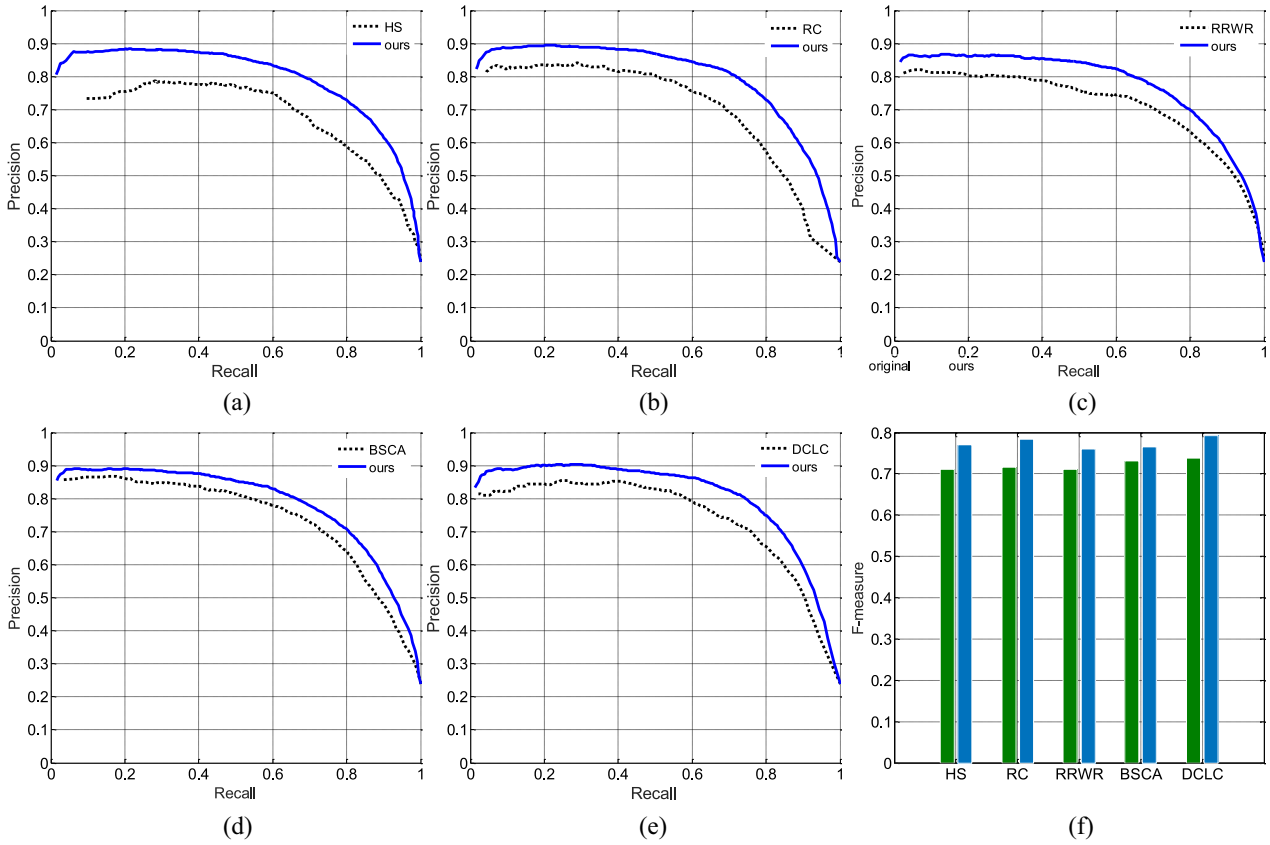


Fig. 12. Quantitative evaluation of one-for-one option for our framework on RGBD Cosal150 dataset. The black line in each PR curve denotes the original RGB saliency result, and the blue line represents the final co-saliency result using the proposed framework. The PR curves with different input saliency maps, i.e., (a) HS, (b) RC, (c) RRWR, (d) BSCA, and (e) DCLC. (f) F-measure of the one-for-one framework.

TABLE I
F-MEASURE OF THE DSP ON THE RGBD400 DATASET

	HS	RC	DCLC	RRWR	BSCA
Without DSP	0.6661	0.6732	0.6914	0.7040	0.7022
With DSP	0.6904	0.6914	0.7094	0.7132	0.7136
Percentage Gain	3.65%	2.70%	2.60%	1.31%	1.62%

TABLE II
F-MEASURE OF THE DSP ON THE RGBD COSAL150 DATASET

	HS	RC	DCLC	RRWR	BSCA
Without DSP	0.7101	0.7163	0.7385	0.7106	0.7318
With DSP	0.7294	0.7457	0.7642	0.7294	0.7502
Percentage Gain	2.72%	4.10%	3.48%	2.65%	2.51%

In this section, we evaluate the performance of DSP on RGBD400 saliency dataset [42], and the relevant results are shown in Fig. 11 and Table I. Five different 2-D saliency maps produced by BSCA, RC, HS, RRWR, and DCLC methods are used as the original saliency maps. In PR curves, the black line denotes the original RGB saliency result, and the red line represents the saliency result with DSP. From the PR curves shown in Fig. 11, it can be seen that the saliency result with DSP achieves the higher precisions of the whole PR curves compared to the 2-D saliency results, and the F-measure also arrives at the consistent conclusion from Table I. For the F-measure, the maximum percentage gain achieves 3.65% for HS method, and the average percentage gain achieves 2.38%. In order to further illustrate the effectiveness of DSP descriptor in our model, we have conducted an experiment on the RGBD Cosal150 dataset, and the results are shown in Table II. From the table, it can be seen that the F-measure achieves the maximum percentage gain of 4.10% for RC method, and the average percentage gain achieves 3.09%. These experiments demonstrated that the depth information could improve the performances of the co-saliency. In other words, the DSP can be used as an independent descriptor that converts the 2-D saliency map into RGBD saliency map.

E. Discussion

The proposed RGBD co-saliency detection framework is designed as a many-for-one model, that is, multiple single 2-D saliency maps input and one RGBD co-saliency map output. In fact, our framework can also achieve one-for-one model. In other words, if there is only one saliency map is embedded into the framework, it can also output one RGBD co-saliency map. The experimental comparison is reported

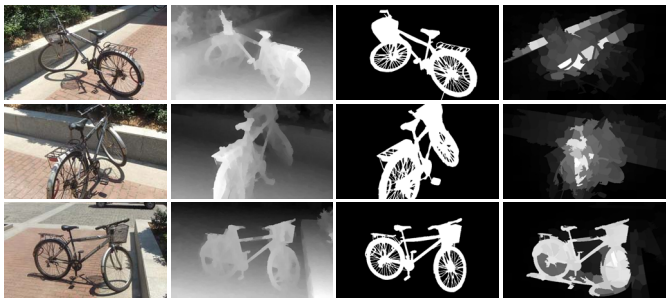


Fig. 13. Some challenging examples for our RGBD co-saliency detection model. (a) RGB image. (b) Depth map. (c) Ground truth. (d) Co-saliency result produced by our framework.

in Fig. 12. The PR curves and F-measure demonstrate that the one-for-one option also achieves the transformation from single image saliency map to RGBD co-saliency map, and obtains better performance of co-saliency detection. In general, the better the saliency map is, the better the co-saliency map achieves. This is, of course, the reason why the multiple saliency maps are fused at first in the proposed framework. It can provide a better baseline for later detection in order to achieve more accurate and stable co-saliency result. However, as the results shown in Fig. 12, the proposed framework can also acquire satisfying result when only one saliency map is embedded.

In addition, we discuss some degenerated cases of the proposed RGBD co-saliency framework in this section. The details are shown in Fig. 13. In this group, there are lots of textures and detail regions, such as the spokes and back seat of the bike in each image. These regions are difficult to detect completely and accurately for our framework. The main reason is that our co-saliency model focuses on low-level features extraction without learning, which could not capture the details very well. In the future, the low-level traditional features and the high-level learning features can be integrated to further describe the object characteristics of the image.

V. Conclusion

In this paper, an iterative RGBD co-saliency framework is proposed to convert the existing 2-D saliency model into RGBD co-saliency scenario. Three schemes are integrated into the framework, which include addition scheme, deletion scheme and iteration scheme. The addition scheme is used to optimize the single saliency map and introduce the depth information into the framework using a novel descriptor named DSP. The deletion scheme aims at capturing the inter-image constraints and suppressing the non-common regions using a common probability function, which is formulated as the likelihood of each superpixel belonging to the common regions. Finally, an iterative scheme is designed to obtain more homogeneous and consistent co-saliency map. The comprehensive comparison and discussion on two RGBD co-saliency datasets have demonstrated that the proposed method outperforms other state-of-the-art saliency and co-saliency models.

References

[1] Y. Gao, M. Shi, D. Tao, and C. Xu, "Database saliency for fast image retrieval," *IEEE Trans. Multimedia*, vol. 17, no. 3, pp. 359–369, Mar. 2015.

[2] J. Lei *et al.*, "Depth sensation enhancement for multiple virtual view rendering," *IEEE Trans. Multimedia*, vol. 17, no. 4, pp. 457–469, Apr. 2015.

[3] J. Lei *et al.*, "Depth map super-resolution considering view synthesis quality," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1732–1745, Apr. 2017.

[4] X. Cao, C. Zhang, H. Fu, X. Guo, and Q. Tian, "Saliency-aware nonparametric foreground annotation based on weakly labeled data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1253–1265, Jun. 2016.

[5] J. Peng, J. Shen, and X. Li, "High-order energies for stereo segmentation," *IEEE Trans. Cybern.*, vol. 46, no. 7, pp. 1616–1627, Jul. 2016.

[6] W. Wang, J. Shen, and F. Porikli, "Saliency-aware geodesic video object segmentation," in *Proc. CVPR*, Boston, MA, USA, 2015, pp. 3395–3402.

[7] K. Gu *et al.*, "Saliency-guided quality assessment of screen content images," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1098–1110, Jun. 2016.

[8] Y. Fang *et al.*, "Objective quality assessment for image retargeting based on structural similarity," *IEEE J. Emerg. Sel. Topic Circuits Syst.*, vol. 4, no. 1, pp. 95–105, Mar. 2014.

[9] J. Shen, D. Wang, and X. Li, "Depth-aware image seam carving," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1453–1461, Oct. 2013.

[10] Z. Pan *et al.*, "Fast reference frame selection based on content similarity for low complexity HEVC encoder," *J. Vis. Commun. Image Represent.*, vol. 40, pp. 516–524, Oct. 2016.

[11] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *Proc. CVPR*, Colorado Springs, CO, USA, Jun. 2011, pp. 409–416.

[12] L. Zhou, Z. Yang, Q. Yuan, Z. Zhou, and D. Hu, "Salient region detection via integrating diffusion-based compactness and local contrast," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3308–3320, Nov. 2015.

[13] C. Li, Y. Yuan, W. Cai, Y. Xia, and D. D. Feng, "Robust saliency detection via regularized random walks ranking," in *Proc. CVPR*, Boston, MA, USA, Jun. 2015, pp. 2710–2717.

[14] J. Lei *et al.*, "A universal framework for salient object detection," *IEEE Trans. Multimedia*, vol. 18, no. 9, pp. 1783–1795, Sep. 2016.

[15] J. Shi, Q. Yan, L. Xu, and J. Jia, "Hierarchical image saliency detection on extended CSSD," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 4, pp. 717–729, Apr. 2016.

[16] Y. Qin, H. Lu, Y. Xu, and H. Wang, "Saliency detection via cellular automata," in *Proc. CVPR*, Boston, MA, USA, Jun. 2015, pp. 110–119.

[17] Q. Wang, Y. Yuan, P. Yan, and X. Li, "Saliency detection by multiple-instance learning," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 660–672, Apr. 2013.

[18] D. Zhang, H. Fu, J. Han, A. Borji, and X. Li. (2017). *A Review of Co-Saliency Detection Technique: Fundamentals, Applications, and Challenges*. [Online]. Available: https://arxiv.org/abs/1604.07090

[19] H.-T. Chen, "Preattentive co-saliency detection," in *Proc. ICIP*, Hong Kong, Sep. 2010, pp. 1117–1120.

[20] H. Li and K. N. Ngan, "A co-saliency model of image pairs," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3365–3375, Dec. 2011.

[21] K.-Y. Chang, T.-L. Liu, and S.-H. Lai, "From co-saliency to co-segmentation: An efficient and fully unsupervised energy minimization model," in *Proc. CVPR*, Colorado Springs, CO, USA, Jun. 2011, pp. 2129–2136.

[22] H. Fu, X. Cao, and Z. Tu, "Cluster-based co-saliency detection," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3766–3778, Oct. 2013.

[23] K. Tang, A. Joulin, L.-J. Li, and L. Fei-Fei, "Co-localization in real-world images," in *Proc. CVPR*, Columbus, OH, USA, Jun. 2014, pp. 1464–1471.

[24] A. Joulin, K. Tang, and L. Fei-Fei, "Efficient image and video co-localization with frank-wolfe algorithm," in *Proc. ECCV*, Zürich, Switzerland, Sep. 2014, pp. 253–268.

[25] A. Toshev, J. Shi, and K. Daniilidis, "Image matching via saliency region correspondences," in *Proc. CVPR*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8.

[26] H. Fu, D. Xu, B. Zhang, S. Lin, and R. K. Ward, "Object-based multiple foreground video co-segmentation via multi-state selection graph," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3415–3424, Nov. 2015.

[27] H. Fu, D. Xu, S. Lin, and J. Liu, "Object-based RGBD image co-segmentation with mutex constraint," in *Proc. CVPR*, Boston, MA, USA, Jun. 2015, pp. 4428–4436.
[28] X. Guo *et al.*, "Robust object co-detection," in *Proc. CVPR*, Portland, OR, USA, Jun. 2013, pp. 3206–3213.
[29] Z. Liu, W. Zou, L. Li, L. Shen, and O. L. Meur, "Co-saliency detection based on hierarchical segmentation," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 88–92, Jan. 2014.
[30] C. Ge, K. Fu, F. Liu, L. Bai, and J. Yang, "Co-saliency detection via inter and intra saliency propagation," *Signal Process. Image Commun.*, vol. 44, pp. 69–83, May 2016.
[31] M.-M. Cheng, N. J. Mitra, X. Huang, and S.-M. Hu, "SalientShape: Group saliency in image collections," *Vis. Comput.*, vol. 30, no. 4, pp. 443–453, Aug. 2014.
[32] X. Cao, Z. Tao, B. Zhang, H. Fu, and X. Li, "Saliency map fusion based on rank-one constraint," in *Proc. ICME*, San Jose, CA, USA, Jul. 2013, pp. 1–8.
[33] X. Cao, Z. Tao, B. Zhang, H. Fu, and W. Feng, "Self-adaptively weighted co-saliency detection via rank constraint," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4175–4186, Sep. 2014.
[34] Y. Li, K. Fu, Z. Liu, and J. Yang, "Efficient saliency-model-guided visual co-saliency detection," *IEEE Signal Process. Lett.*, vol. 22, no. 5, pp. 588–592, May 2015.
[35] R. Huang, W. Feng, and J. Sun, "Saliency and co-saliency detection by low-rank multiscale fusion," in *Proc. ICME*, Turin, Italy, Jun./Jul. 2015, pp. 1–6.
[36] D. Zhang, J. Han, C. Li, and J. Wang, "Co-saliency detection via looking deep and wide," in *Proc. CVPR*, Boston, MA, USA, Jun. 2015, pp. 2994–3002.
[37] D. Zhang *et al.*, "A self-paced multiple-instance learning framework for co-saliency detection," in *Proc. ICCV*, Santiago, Chile, Dec. 2015, pp. 594–602.
[38] B. Chen, J. Yang, B. Jeon, and X. Zhang, "Kernel quaternion principal component analysis and its application in RGB-D object recognition," *Neurocomputing*, vol. 266, pp. 293–303, Nov. 2017.
[39] R. Cong *et al.*, "Co-saliency detection for RGBD images based on multi-constraint feature matching and cross label propagation," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 568–579, Feb. 2018.
[40] Y. Niu, Y. Geng, X. Li, and F. Liu, "Leveraging stereopsis for saliency analysis," in *Proc. CVPR*, Providence, RI, USA, Jun. 2012, pp. 454–461.
[41] H. Peng, B. Li, W. Xiong, W. Hu, and R. Ji, "RGBD salient object detection: A benchmark and algorithms," in *Proc. ECCV*, Zürich, Switzerland, Sep. 2014, pp. 92–109.
[42] R. Ju, Y. Liu, T. Ren, L. Ge, and G. Wu, "Depth-aware salient object detection using anisotropic center-surround difference," *Signal Process. Image Commun.*, vol. 38, pp. 115–126, Oct. 2015.
[43] D. Feng, N. Barnes, S. You, and C. McCarthy, "Local background enclosure for RGB-D salient object detection," in *Proc. CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 2343–2350.
[44] R. Cong *et al.*, "Saliency detection for stereoscopic images based on depth confidence analysis and multiple cues fusion," *IEEE Signal Process. Lett.*, vol. 23, no. 6, pp. 819–823, Jun. 2016.
[45] W. Wang, J. Shen, Y. Yu, and K.-L. Ma, "Stereoscopic thumbnail creation via efficient stereo saliency detection," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 8, pp. 2014–2027, Aug. 2017.
[46] H. Li, H. Lu, Z. Lin, X. Shen, and B. Price, "Inner and inter label propagation: Salient object detection in the wild," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3176–3186, Oct. 2015.
[47] T. Chen, L. Lin, L. Liu, X. Luo, and X. Li, "DISC: Deep image saliency computing via progressive representation learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1135–1149, Jun. 2016.
[48] S. He, R. W. H. Lau, W. Liu, Z. Huang, and Q. Yang, "SuperCNN: A superpixelwise convolutional neural network for salient object detection," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 330–344, Dec. 2015.
[49] G. Lee, Y.-W. Tai, and J. Kim, "Deep saliency with encoded low level distance map and high level features," in *Proc. CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 660–668.
[50] G. Li and Y. Yu, "Deep contrast learning for salient object detection," in *Proc. CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 478–487.
[51] A. Wang, J. Lu, J. Cai, T.-J. Cham, and G. Wang, "Large-Margin multi-modal deep learning for RGB-D object recognition," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 1887–1898, Nov. 2015.
[52] R. Achanta *et al.*, "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.
[53] T. Leung and J. Malik, "Recognizing surfaces using three-dimensional textons," in *Proc. ICCV*, Sep. 1999, pp. 1010–1017.
[54] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.
[55] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, "Frequency-tuned salient region detection," in *Proc. CVPR*, Miami, FL, USA, Jun. 2009, pp. 1597–1604.
[56] H. Song, Z. Liu, Y. Xie, L. Wu, and M. Huang, "RGBD co-saliency detection via bagging-based clustering," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1722–1726, Dec. 2016.

Runmin Cong received the M.S. degree from the Civil Aviation University of China, Tianjin, China, in 2014. He is currently pursuing the Ph.D. degree in information and communication engineering with Tianjin University, Tianjin.

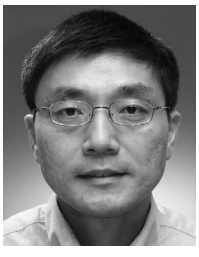
He was a visiting student with Nanyang Technological University, Singapore, from December 2016 to February 2017. His current research interests include saliency detection, 3-D imaging, and computer vision.

Jianjun Lei (M'11–SM'17) received the Ph.D. degree in signal and information processing from the Beijing University of Posts and Telecommunications, Beijing, China, in 2007.

He was a Visiting Researcher with the Department of Electrical Engineering, University of Washington, Seattle, WA, USA, from August 2012 to August 2013. He is currently a Professor with Tianjin University, Tianjin, China. His current research interests include 3-D video processing, virtual reality, and artificial intelligence.

Huazhu Fu received the B.S. degree in mathematical sciences from Nankai University, Tianjin, China, in 2006, the M.E. degree in mechatronics engineering from the Tianjin University of Technology, Tianjin, in 2010, and the Ph.D. degree in computer science from Tianjin University, Tianjin, in 2013.

He was a Research Fellow with Nanyang Technological University, Singapore, for two years. He is currently a Research Scientist with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore. His current research interests include computer vision, image processing, and medical image analysis.



Weisi Lin (M'92–SM'98–F'16) received the Ph.D. degree from King's College, London University, London, U.K.

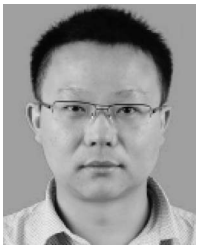
He is a Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. His current research interests include image processing, perceptual signal modeling, video compression, and multimedia communication, in which he has published over 180 journal papers and 230 conference papers, authored two books, and holds seven patents.

Dr. Lin has been an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, and the IEEE SIGNAL PROCESSING LETTERS. He has been the Technical Program Chair of the IEEE ICME 2013, PCM 2012, QoMEX 2014, and the IEEE VCIP 2017. He has been an Invited/Panelist/Keynote/Tutorial Speaker in over 20 international conferences, as well as a Distinguished Lecturer of the IEEE Circuits and Systems Society from 2016 to 2017, and the Asia–Pacific Signal and Information Processing Association from 2012 to 2013. He is a fellow of IET, and an Honorary Fellow of the Singapore Institute of Engineering Technologists.

Qingming Huang (SM'08) received the bachelor's degree in computer science and the Ph.D. degree in computer engineering from the Harbin Institute of Technology, Harbin, China, in 1988 and 1994, respectively.

He is a Professor with the University of Chinese Academy of Sciences, Beijing, China, and an Adjunct Research Professor with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing. He has published over 300 academic papers in prestigious international journals, including the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and top-level conferences, such as ACM Multimedia, ICCV, CVPR, IJCAI, and VLDB. His current research interests include multimedia video analysis, image processing, computer vision, and pattern recognition.

Dr. Huang is an Associate Editor of *Acta Automatica Sinica*, and a Reviewer of various international journals, including the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and the IEEE TRANSACTIONS ON IMAGE PROCESSING. He has served as the General Chair, the Program Chair, the Track Chair, and a TPC Member for various conferences, including ACM Multimedia, CVPR, ICCV, ICME, PCM, and PSIVT.



Xiaochun Cao (SM'14) received the B.E. and M.E. degrees in computer science from Beihang University, Beijing, China, and the Ph.D. degree in computer science from the University of Central Florida, Orlando, FL, USA.

He has been a Professor with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China, since 2012. He was a Research Scientist with ObjectVideo Inc., Reston, VA, USA, for about three years. From 2008 to 2012, he was a Professor with Tianjin University, Tianjin, China. He has authored and co-authored over 120 journal and conference papers.

Dr. Cao was a recipient of the Piero Zamperoni Best Student Paper Award at the International Conference on Pattern Recognition in 2004 and 2010 and nominated for the University of Central Floridas University Level Outstanding Dissertation Award for his dissertation. He is on the Editorial Board of the IEEE TRANSACTIONS OF IMAGE PROCESSING. He is a fellow of IET.

Chunping Hou received the M.Eng. and Ph.D. degrees in electronic engineering from Tianjin University, Tianjin, China, in 1986 and 1998, respectively.

Since 1986, she has been the faculty of the School of Electronic and Information Engineering, Tianjin University, where she is currently a Full Professor and the Director of the Broadband Wireless Communications and 3D Imaging Institute. Her current research interests include 3-D image processing, 3-D display, wireless communication, and the design and applications of communication systems.