Multimed Tools Appl
https://doi.org/10.1007/s11042-017-5576-y

RGBD co-saliency detection via multiple kernel boosting and fusion

Lishan Wu[1,2] · **Zhi Liu**[1,2] · **Hangke Song**[1,2] · **Olivier Le Meur**[3]

Received: 11 May 2017 / Revised: 9 November 2017 / Accepted: 21 December 2017

Abstract RGBD co-saliency detection, which aims at extracting common salient objects from a group of RGBD images with the additional depth information, has become an emerging branch of saliency detection. In this regard, this paper proposes a novel framework via multiple kernel boosting (MKB) and co-saliency quality based fusion. First, on the basis of pre-segmented regions at multiple scales, the regional clustering by feature bagging is exploited to generate the base co-saliency maps. Then the clustering-based samples selection is performed to select the most similar regions with high saliency from different images in the image set. The selected samples are utilized to learn a MKB-based regressor, which is applied to all regions at multiple scales to generate the MKB-based co-saliency maps. Finally, to make full use of both MKB and clustering-based co-saliency maps, a co-saliency quality criterion is proposed for adaptive fusion to generate the final co-saliency maps. Experimental results on a public RGBD co-saliency detection dataset demonstrate that the proposed co-saliency model outperforms the state-of-the-art co-saliency models.

Keywords Co-saliency detection · RGBD images · Multiple kernel boosting · Fusion

✉ Zhi Liu
liuzhisjtu@163.com

Lishan Wu
wlsxxrs@163.com

Hangke Song
hksong0209@163.com

Olivier Le Meur
olemeur@irisa.fr

[1] Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai 200444, China

[2] School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China

[3] IRISA, University of Rennes 1, 35042 Rennes, France

Published online: 05 January 2018

1 Introduction

Saliency detection, with the goal of highlighting salient objects, has become a research hotspot and plays an important role in computer vision tasks. A number of early works [6, 12, 14, 20] have been performed to explore effective and efficient methods for salient object detection, especially when the deep neural networks were adopted for saliency detection [11, 22]. As an emerging branch of visual saliency, co-saliency detection, which discovers the common salient objects in a group of images, quickly gains research interests in the recent years. Thanks to its superior scalability, co-saliency detection has been widely used in image/video co-segmentation [29, 30], object co-localization [7], and weakly supervised object detection [24].

Since the early work [13] which calculates pixel-level saliency using the information of related images, more and more co-saliency models [4, 8, 18, 19, 21, 31, 32] have been proposed. These models assume that the co-salient objects existing in a set of related images should share consistency on low-level features. Li et al. [18] considered the co-saliency as a linear combination of the single-image saliency maps, which are computed by three available saliency models, and the multi-image saliency map, which is calculated by constructing a co-multilayer graph, to generate the co-saliency map for image pairs. Li et al. [19] then extended co-saliency detection from image pairs to a group of images. Fu et al. [8] used contrast, spatial, and global correspondence to measure the pixel-level cluster saliency and generate the final co-saliency maps by fusing the single-image saliency and multi-image saliency. Liu et al. [21] combined contrast, object prior and global regional similarity, on the basis of hierarchical segmentation to generate co-saliency maps, which effectively utilize the information of both coarse scale and fine scale. Ye et al. [31] used color and SIFT features to generate co-salient exemplars, of which saliency values are propagated to recover co-salient objects. Cao et al. [4] proposed a saliency fusion framework to obtain co-saliency maps by making full use of the relationship of multiple saliency cues and the obtained self-adaptive weights.

With the emergence of depth cameras, stereo cameras and Kinect sensors, the depth information can be captured simultaneously with the original RGB images, to generate the RGBD images, which facilitate a number of image/video processing tasks. With the additional depth information as well as the RGB information, more and more saliency models [10, 15, 17, 25, 27] have been proposed for RGBD images. Song et al. [25] proposed a multi-stage depth-aware saliency model for salient region detection, which learns a discriminative saliency fusion (DSF) regressor based on random forest to estimate saliency measures of regions, and achieves a better saliency detection performance for RGBD images. Guo et al. [10] proposed a two-step saliency evolution strategy, which fully explores the potential of color cue and depth cue in the whole procedure of salient object detection, to ensure the high precision and completeness of the detected salient objects. Recently, Fu et al. [9] fused some high-performing RGB saliency models, co-saliency models and RGBD saliency models to generate RGBD co-saliency maps, but the relevance of depth cues among different RGBD images is ignored. Song et al. [26] proposed a RGBD co-saliency detection model using bagging-based clustering, which effectively utilizes different features for different RGBD image sets. Nonetheless, RGBD co-saliency detection has not been fully explored yet.

Therefore, this paper proposes a novel RGBD co-saliency detection framework using multiple kernel boosting (MKB) and co-saliency quality-based fusion as shown in Fig. 1. On the basis of region pre-segmentation results at multiple scales, the regional clustering by feature bagging is exploited to generate the base co-saliency (BCS) maps. Then the clustering-based samples selection is performed to gather training samples, and the MKB algorithm is brought in co-saliency detection for RGBD images so as to generate the MKB-based co-saliency (MCS)

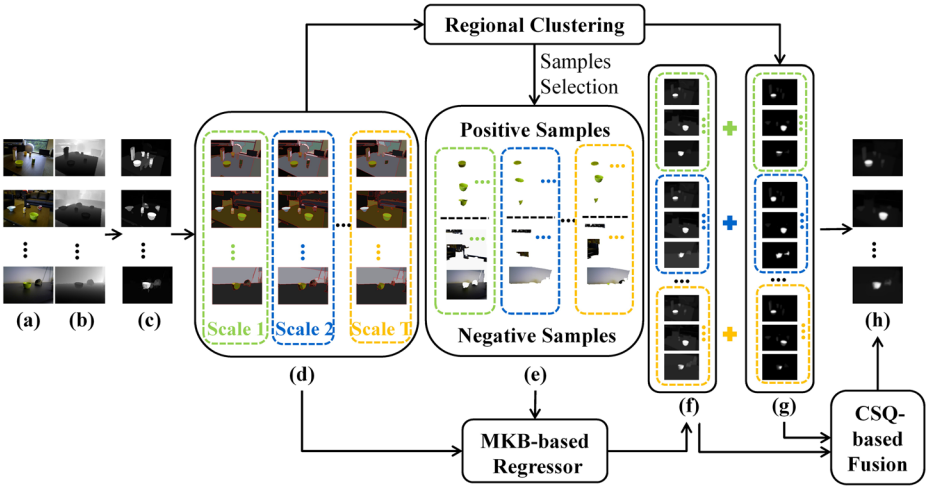

Fig. 1 (Better viewed in color) Framework of RGBD co-saliency model via multiple kernel boosting and fusion. (**a**) Color images, (**b**) depth maps, (**c**) RGBD single saliency maps, (**d**) region pre-segmentation results at multiple scales, (**e**) positive samples and negative samples at multiple scales, (**f**) MCS maps, (**g**) BCS maps, (**h**) final co-saliency maps

maps. Finally, a co-saliency quality (CSQ) criterion is proposed for adaptive fusion and generating final co-saliency maps.

The main contributions of our RGBD co-saliency model lie in the following three aspects:

1) In order to select reliable samples for training, we exploit the fuzzy c-means clustering (FCM) method [2] to adaptively determine the number of clusters, instead of using a fixed number of clusters as in [26]. In addition, a multi-scale segmentation is performed before clustering so as to obtain more versatile samples for training. Compared to [26], our method improves the quality of clustering, and thus effectively enhances the reliability of samples selection for co-saliency detection.
2) We propose to use the MKB algorithm to generate MCS maps, which can highlight co-salient objects more completely. The features used in our MKB algorithm are dynamically selected according to different clustering processes operating at multiple scales, which are different from the fixed features used in [27, 28].
3) In order to make full use of complementary advantages of various MCS maps and BCS maps at different scales, the proposed CSQ criterion is exploited for adaptive fusion, which enables to achieve the better co-saliency detection performance.

The rest of this paper is organized as follows. The proposed RGBD co-saliency model is presented in Section 2. Experimental results are shown in Section 3, and the conclusion is given in Section 4.

2 Proposed RGBD co-saliency model

2.1 Generate clustering-based co-saliency maps

Given a group of RGBD images $\{I_n\}_{n=1}^{N}$ (color images with their depth maps) as shown in Fig. 1, the normalized RGBD single saliency map for each image is generated using the

RGBD saliency model in [25]. At each scale, the pre-segmentation result of each color image with Q regions is generated using the *gPb-owt-ucm* method [1]. We use $\left\{R_q^n\right\}_{q=1}^{Q}$ to represent each region in the image I_n, and use the mean of RGBD single saliency values of all pixels in R_q^n as the regional saliency of R_q^n. Then R_q^n is selected as a candidate object region for clustering, if its regional saliency is higher than the threshold θ, which is set to 0.25 for a relatively higher recall rate. Following the work in [26], we perform the regional clustering by feature bagging, which is able to choose the appropriate features for different image sets with different scenes. The features are the same as those used in [26], including color, depth and geometric properties of regions, with a total of 24 dimensions. For the p^{th} clustering process, a random number $FN^p (1 \le FN^p \le 24)$ is generated as the number of regional features used in this clustering process, and a FN^p-dimensional feature vector is generated for each region. Let C^p denote the p^{th} clustering result and C_x^p denote the x^{th} cluster in the p^{th} clustering result, the cluster-level co-saliency (CCS) of C_x^p is defined as follows:

$$CCS_x^p = MS_x^p \cdot \exp\left(1 - NCD_x^p\right) \cdot \exp\left(CCO_x^p / N\right), \tag{1}$$

where MS_x^p is the mean of regional saliency values of all regions contained in C_x^p. We compute the Euclidean distance between the feature vector of each region in C_x^p and the cluster center of C_x^p, and calculate the cluster distance CD_x^p as the mean of all such Euclidean distances in C_x^p. All the cluster distances with the p^{th} clustering result are normalized into the range of [0, 1] to obtain NCD_x^p for each cluster C_x^p. It is obvious that the lower NCD_x^p is, the more similar the regions in C_x^p are. The cluster co-occurrence rate CCO_x^p is defined as the number of images involved in C_x^p. Therefore, using Eq. (1), a cluster is assigned with a higher co-saliency value, if the regions in this cluster have the higher regional saliency values, show the higher similarity and occur in more images.

In order to obtain the better clustering results and select more reliable samples, we use the fuzzy c-means clustering (FCM) method [2] to adaptively determine the number of clusters. FCM exploits the membership function U (fuzzy partition matrix) to determine the possibility of a region belonging to each cluster. The number of clusters has a significant impact on the accuracy of clustering, so we utilize the membership function to find an appropriate number of clusters. The membership function $U = \{u_{xv}\}_{1 \le x \le X,\ 1 \le v \le V}$ should meet the following three conditions: $\sum_{x=1}^{X} u_{xv} = 1$, $u_{xv} \in [0, 1]$ and $0 \le \sum_{v=1}^{V} u_{xv} \le V$. X denotes the number of clusters, and V denotes the number of candidate regions that need to be classified. In order to find an appropriate number of clusters, we define the following function to maximize,

$$G(u, X) = \frac{1}{V} \sum_{x=1}^{X} \sum_{v=1}^{V} {u_{xv}}^2. \tag{2}$$

It has been proved in [2] that the function $G(u, X)$ has an exclusive maximum on the well-defined intervals, and the number of clusters, X, which corresponds to this maximum value, is the appropriate number of clusters. For a balance between the runtime and clustering performance, the search range of the number of clusters for each clustering process is set to $[\min(round(V/n), 20), 25]$. The analysis of how to determine the search range of the number of clusters is given in Section 3.4.

Based on the p^{th} clustering result, each region in the cluster C_x^p is assigned with the same cluster-level co-saliency value, CCS_x^p, and the co-saliency values of all the other regions, which are not selected as the candidate object regions, are uniformly set to 0. With a total of P

clustering results, for each image I_n, a total of P weak co-saliency (WCS) maps, $\left\{WCS_n^p\right\}_{p=1}^{P}$, are generated by the above process accordingly. In our experiments, the number of clustering processes, P, is set to 150.

In order to find the most reliable positive samples, we adopt the criterion in [26] to evaluate the quality of each clustering result. Specifically, for the p^{th} clustering result, its clustering quality (CQ) is defined as follows:

$$CQ^p = SR_{xx}^p \cdot \exp\left(\frac{CCO_{xx}^p/N}{RN^p} - DA_{xx}^p\right), \tag{3}$$

where $SR_{xx}^p = \sum_{x \neq xx}\left[CCS_{xx}^p / CCS_x^p\right]$ is the cluster separation rate, and CCS_{xx}^p is the largest cluster-level co-saliency value among all the clusters in the p^{th} clustering result. A larger value of SR_{xx}^p indicates that the object regions are highlighted better from background regions with the p^{th} clustering result. DA_{xx}^p is the variance on the number of regions belonging to the cluster C_{xx}^p in each image. A small value of DA_{xx}^p indicates that all images tend to have the same number of the highlighted object regions, and this indicates the higher reliability of the p^{th} clustering result. RN^p is the average number of non-adjacent regions. RN^p tends to be small when the highlighted object regions are more concentrated, and this indicates the better quality of the p^{th} clustering result.

Based on the CQ values of all clustering results, the WCS maps with the best clustering result, which has the largest CQ value, are selected as the base co-saliency (BCS) maps, $\{BCS_n\}_{n=1}^{N}$.

2.2 Samples selection

The regions in the cluster C_{xx}^b, which has the largest CCS value in the best clustering result, are selected as the positive samples. To guarantee that each image provides at least one positive sample, when one image is not involved in the cluster C_{xx}^b, the region with the maximal regional saliency value in the image will be selected as the positive sample. The negative samples are generated as follows: 1) The regions with a regional saliency value lower than the threshold, 0.1 in our work, are selected as the negative samples; 2) Since background regions usually locate around image borders in most cases, the regions connecting image borders are also selected as the negative samples.

As stated above, we can select samples at one segmentation scale. It is well known that the accuracy of saliency detection is sensitive to the number of regions as well as the size of regions, for the reason that salient objects probably appear at different scales. Therefore, we choose five different scales for the *gPb-owt-ucm* method [1] based on the maximum region number (MAX-N) and the minimum segmentation size (MIN-S), i.e. the ratio of minimum region area to the average area of all regions. The five different scales are set to be MAX-N = 50 with MIN-S = 0.1, MAX-N = 50 with MIN-S = 0.2, MAX-N = 110 with MIN-S = 0.1, MAX-N = 110 with MIN-S = 0.2, and MAX-N = 110 with MIN-S = 0.5. For clarity, the base co-saliency (BCS) maps at the t^{th} scale is denoted as $\left\{BCS_n^t\right\}_{n=1}^{N}$. As shown in Fig. 1, the clustering-based samples selection is performed individually at each scale. After that, all the selected samples at all scales are gathered together to constitute the training set, which will be used to learn a MKB-based regressor in Section 2.3.

2.3 Generate MKB-based co-saliency maps

Inspired by [28, 33], in which a strong classifier based on MKB is learned to measure saliency with a set of feature descriptors and kernels, we introduce MKB for co-saliency detection in this paper. Different from [28], which exploits a set of fixed features like color and local binary pattern (LBP), we exploit the regional features used in the best clustering processes at each segmentation scale to describe each sample.

Algorithm 1 presents the pseudo-code for learning the MKB-based regressor. At the beginning, for all the five scales, positive samples are labeled with +1, and negative samples are labeled with −1, respectively. Therefore, we have a total of H training samples, $\{r_i, l_i\}_{i=1}^{H}$, where r_i is the i^{th} sample and l_i is its binary label. Then a total of M ($M = \Omega_K \times \Omega_F$) support vector machines (SVMs) with different kernels and different features are generated. Here, $\Omega_K = 4$ indicates four kernels, i.e. linear kernel, polynomial kernel, RBF kernel and sigmoid kernel, and $\Omega_F = 5$ indicates the five feature sets used in the best clustering processes at the five scales. Let r denote a sample, and a total of M SVMs with $\{z_m(r)\}_{m=1}^{M}$ are generated. The AdaBoost method is used to combine the SVMs into the final regressor according to their weights. After a total of J iterations (J is adaptively equal to M), the MKB-based regressor $Y(r)$ is learned from the training set.

Algorithm 1 Pseudo code of MKB process

Input: The training set $\{r_i, l_i\}_{i=1}^{H}$, and kernel functions $\{k_m\}_{m=1}^{M}$.

Output: The regressor $Y(r)$.

1: The decision function of a single-kernel SVM as $z_m(r)$ is trained with the kernel k_m by using the training set $\{r_i, l_i\}_{i=1}^{H}$, and a total of M ($M = \Omega_K \times \Omega_F$) SVMs are generated.

2: Initialize the weights for samples as $w_1(i) = 1/H$ ($i = 1,...,H$).

3: **for** j =1: J

For the SVM classifiers, obtain a set of decision functions, $\{z_m(r)\}, m = 1,2,...,M$.

For each SVM, calculate the regression error:

$$\varepsilon_m = \frac{\sum_{i=1}^{H} w(i)\left|z_m(r_i)\right|\left(\mathrm{sgn}(-l_i z_m(r_i)) + 1\right)/2}{\sum_{i=1}^{H} w(i)\left|z_m(r_i)\right|}.$$

Select the optimal SVM $z_j(r)$ with the minimum regression error $\varepsilon_j = \min_{1 \le m \le M} \{\varepsilon_m\}$.

Then compute the combination coefficient:

$$\beta_j = \frac{1}{2}\log\frac{1-\varepsilon_j}{\varepsilon_j} \cdot \frac{1}{2}\left[\mathrm{sgn}\left(\log\frac{1-\varepsilon_j}{\varepsilon_j}\right) + 1\right].$$

Update each weight for samples as $w_{j+1}(i) := \frac{w_j(i) e^{-\beta_j l_i z_j(r_i)}}{2\sqrt{\varepsilon_j(1-\varepsilon_j)}}$.

Then perform normalization on all weights.

end for

4: The regressor is obtained as $Y(r) = \sum_{j=1}^{J} \beta_j z_j(r)$.

The learned regressor $Y(r)$ is then used to predict the co-saliency values of all regions at five scales, and to generate the MKB-based co-saliency (MCS) maps, $\{MCS_n^t\}_{t=1}^{T}, (T = 5)$, for each image I_n. Then following [28], the graph cut-based refinement [16] is exploited to enhance the smoothness of MCS maps. Figure 1 shows at different scales some examples of MCS maps, which effectively highlight the co-salient objects and simultaneously suppress irrelevant regions.

2.4 Co-saliency quality based fusion

For the better co-saliency detection performance, we first linearly combine each MCS map with its corresponding BCS map to obtain the integrated co-saliency (ICS) map as follows:

$$ICS_n^t = 0.5 \cdot MCS_n^t + 0.5 \cdot BCS_n^t. \tag{4}$$

Then an adaptive fusion method based on the quality of different ICS maps is proposed to generate the final co-saliency maps. For ICS_n^t, its co-saliency quality (CSQ) is defined as follows:

$$CSQ_n^t = SSDI_n^t \cdot DSSI_n^t. \tag{5}$$

In Eq. (5), at each scale, the color histogram in the *Lab* color space for positive samples in each image is calculated, and the chi-square distance between each pair of such color histograms is calculated by the following means. $SSDI_n^t$ (Same Scale with Different Images) denotes at the t^{th} scale the similarity between the positive samples with each image I_n and the positive samples with all the other images. $DSSI_n^t$ (Different Scales with the Same Image) denotes the similarity between the positive samples with each image I_n at the t^{th} scale and the positive samples with the same image at other scales. A higher value of either term indicates a higher similarity among the positive samples, and thus a higher reliability of the positive samples. Specifically, $SSDI_n^t$ and $DSSI_n^t$ are defined as follows:

$$SSDI_n^t = 1 - \frac{1}{2} \sum_{n_j=1, n_j \neq n}^{N} \chi^2\left(H_n^t, H_{n_j}^t\right), \tag{6}$$

$$DSSI_n^t = 1 - \frac{1}{2} \sum_{t_j=1, t_j \neq t}^{T} \chi^2\left(H_n^t, H_n^{t_j}\right), \tag{7}$$

where H_n^t is the color histogram for the positive samples with each image I_n at the t^{th} scale, and $\chi^2(\cdot)$ denotes the chi-square distance.

The CSQ value is calculated for each ICS map using Eq. (5). A higher CSQ value indicates a better quality of the corresponding ICS map. With CSQ values, for each image I_n, all the ICS maps are integrated to generate the final co-saliency (FCS) map at region level via the adaptive fusion as follows:

$$FCS_n = \sum_{t=1}^{T} CSQ_n^t \cdot ICS_n^t. \tag{8}$$

Furthermore, to generate the final co-saliency maps at pixel level for the better visualization, a small Gaussian kernel as suggested in [3] is convolved on the region-level FCS maps. As shown in the rightmost part of Fig. 1, by integrating all the ICS maps at all scales based on their CSQ values as the weights, the co-salient objects are uniformly highlighted and background regions are effectively suppressed in the final co-saliency maps.

3 Experimental results

3.1 Experimental setting

We performed experiments on a public dataset [9] for RGBD co-saliency detection, with 183 RGBD images from 16 object classes. Our RGBD co-saliency detection results are compared with four existing co-saliency models, i.e. CB [8], HS [21], ODR [31] and SACS [4], for color images, as well as two state-of-the-art RGBD co-saliency models, i.e. FFS [9] and BC [26]. The source codes of CB, HS, ODR, SACS and BC are provided by their authors, while the co-saliency maps of FFS are provided by its authors.

3.2 Qualitative comparison

Figure 2 shows a subjective comparison of co-saliency maps for several image sets. All co-saliency maps are normalized into the same range of [0, 255]. As shown in the leftmost example of Fig. 2, the other co-saliency models falsely highlight some irrelevant background regions, while our model can better suppress irrelevant regions more completely. As for some complicated image sets, in which the common objects have a small size, such as the green bowls in the 2nd example and the soda cans in the 3rd example, the other models cannot accurately highlight the co-salient objects as a whole and also cannot uniformly suppress background regions, while our model achieves the better performance than other models. In addition, with the help of depth features, our model maintains the best performance on highlighting co-salient object regions as well as suppressing background regions, when the color of co-salient objects is quite similar to the color of background (see the rightmost example in Fig. 2). We can see from Fig. 2 that our model is able to handle some complicated images with heterogeneous objects, low-contrast objects and cluttered background better than the other models.

3.3 Quantitative comparison

The performances of different co-saliency models are evaluated objectively by using the precision-recall (PR) curve and the F-measure with its coefficient β^2 set to 1, which weights precision and

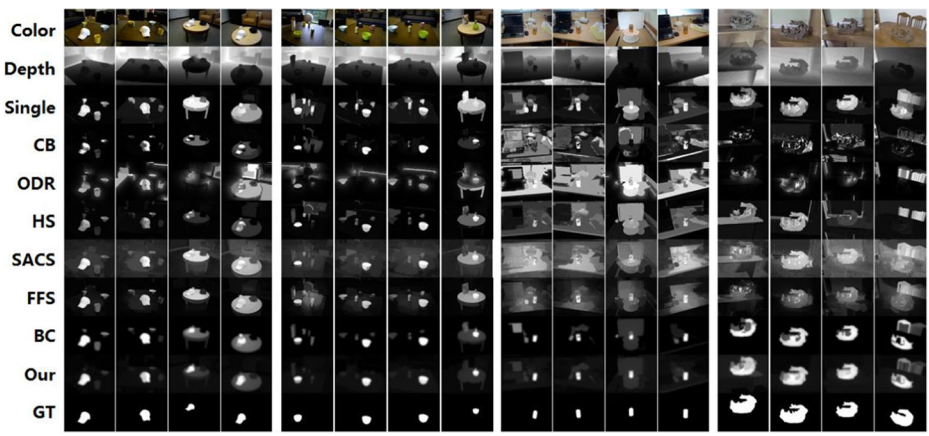

Fig. 2 Examples of co-saliency detection on four image sets

recall equally. It can be seen from Fig. 3 that our model outperforms all the other models. As shown in Table 1, our model also achieves the highest F-measure. Both Fig. 3 and Table 1 indicate that our model achieves the better performance than all the other models for RGBD co-saliency detection.

The performance contributions of different parts in our model are further evaluated. Recall that in our model, we utilize five different scales, i.e. MAX-N = 50 with MIN-S = 0.1, MAX-N = 50 with MIN-S = 0.2, MAX-N = 110 with MIN-S = 0.1, MAX-N = 110 with MIN-S = 0.2, and MAX-N = 110 with MIN-S = 0.5, which are numbered consecutively as T1 to T5. The co-saliency maps generated by linear combination of MCS maps and BCS maps at each scale from T1 to T5, respectively, are denoted as from ICS-T1 to ICS-T5 accordingly. The F-measures achieved by the results from ICS-T1 to ICS-T5 are shown in Table 2. It can be seen from Table 2 that the performances with different scales are close, and the best performance among all scales is achieved with the scale T1 (MAX-N = 50 with MIN-S = 0.1). Therefore, we just select the scale T1 for the related analysis on one scale in the following.

The different versions of our model are defined as follows. 1) Single: the RGBD single saliency maps obtained by [25]; 2) BCS-T1: the cluster-level co-saliency maps with the largest CQ values at the scale T1; 3) MCS-T1: the MKB-based co-saliency maps at the scale T1; 4) ICS-T1: the co-saliency maps by linear combination of MCS maps and BCS maps at the scale T1; 5) Linear: the linear fusion of the ICS maps at all scales; 6) Ours-K-means: the final co-saliency maps output by our model with the K-means clustering method [23] instead of the FCM clustering method; 7) Ours: the final co-saliency maps output by our model.

The PR curves of the above versions are shown in Fig. 4 and the analysis is given as follows: 1) The PR curve of BCS-T1 is obviously higher than that of RGBD single saliency model [25]. This indicates that the clustering process plays an important role in RGBD co-saliency detection. 2) The PR curve of MCS-T1 is higher than that of BCS-T1. This indicates that the use of MKB with different features and different kernels benefits co-saliency detection. Such a performance elevation demonstrates the effectiveness of our MKB-based regressor. 3) The PR curve of ICS-T1 is higher than that of MCS-T1 and BCS-T1. This indicates that the simple linear combination of MCS maps and BCS maps can boost the co-saliency detection

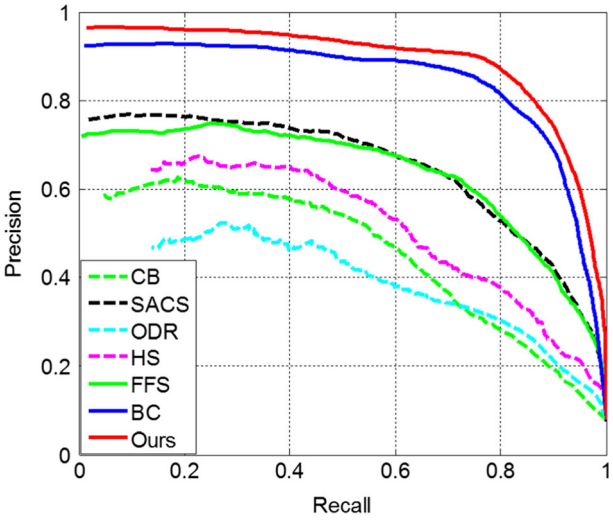

Fig. 3 (Better viewed in color) PR curves of different co-saliency models

Table 1 F-measures (larger is better) and average running time (ART) per image (seconds) of different co-saliency models. The best F-measure and the least ART are marked with boldface

Model	CB	SACS	ODR	HS	FFS	BC	Ours
F-measure	0.451	0.473	0.386	0.480	0.531	0.709	**0.789**
ART	9.42	**1.38**	98.33(43.89)	71.58(17.14)	–	60.02(5.58)	142.96(88.52)

Note for FFS, only the co-saliency maps are provided by the authors, while the code is not available. So, the ART of FFS is not obtained

Note that the number in the bracket excludes the time of generating segmentation results using the *gPb-owt-ucm* method

performance. 4) The PR curves show that our complete model outperforms the single-scale model (ICS-T1) as well as the simple linear fusion of the ICS maps at all scales (Linear), due that our complete model can adaptively weight more the better ones of the ICS maps. 5) The PR curves show that our complete model obviously outperforms its variant (Ours-K-means), and this demonstrates the effectiveness of the fuzzy c-means clustering method used in our model.

Besides, for RGBD co-saliency detection, we should never ignore the contribution of depth information, and thus the versions of our model without depth features are tested. As show in Fig. 4, the performances of those versions without depth features are worse than those with depth features. This indicates that depth features are important for the better co-saliency detection performance. The average F-measure achieved by our model without depth features (Ours-noDepth) is 0.738, which is quite lower than the average F-measure achieved by our model, 0.789. Therefore, both PR curves and F-measures demonstrate the effectiveness of introducing depth features into RGBD co-saliency detection.

3.4 Further analysis

Since our model exploits the FCM clustering method, the number of clusters is a factor that has the effect on co-saliency detection performance. Recall that X denotes the number of clusters, and V denotes the number of candidate regions that need to be classified. Here we vary the value of X to search for an appropriate range with the better co-saliency detection performance. We first evaluated on all categories of the dataset with different values of X using two settings: (1) X is fixed with 2, 5, 10, 15, 20, 25, 30 and 35, respectively; (2) X is adaptively set to $round(V/n)$ with the round operation, where n is the expected number of regions in each cluster on average, and we tested different values of n from 2 to 9. As shown in the top two rows of Table 3, the F-measures with $X = 20$ and $X = 25$ are the largest under the first setting. As shown in the 3rd row and the 4th row of Table 3, the F-measure with $X = round(V/6)$ is the largest under the second setting.

Based on the above observations, the search range of the number of clusters for each clustering process is set to $[\min(round(V/n), 20), 25]$. Such a setting for the search range ensures that the range of [2, 15] is always included in the search range considering the good performance with $X = 20$ and $X = 25$. We also tested other search ranges by varying the value of n, and the results shown in the bottom two rows of Table 3 verify that such a setting for the search range achieves the best

Table 2 F-measures (larger is better) with different scales. The best F-measure is marked with boldface

The Pre-segmentation Scales	ICS-T1	ICS-T2	ICS-T3	ICS-T4	ICS-T5
F-measure	**0.758**	0.753	0.738	0.745	0.747

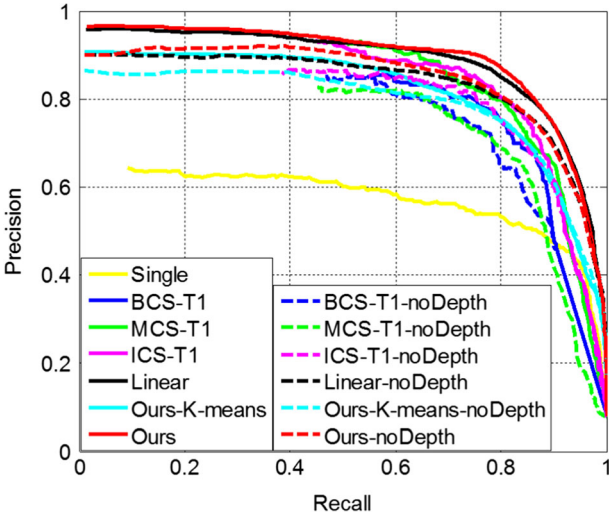

Fig. 4 (Better viewed in color) Contribution analysis on each component in our model and comparison with variants of our model

performance. Besides, the PR curves of our model with these search ranges are shown in Fig. 5, which also demonstrates the best performance of our model with the above setting of search range.

3.5 Computational complexity

Our model is implemented using MATLAB on a PC with an Intel Core i7 3.4GHz CPU and 8GB RAM. We performed the statistic of computation time on RGBD co-saliency dataset [9], in which each image has a resolution of 640 × 480. The average running time (ART) per image of our model and other models is calculated and shown in the bottom row of Table 1. The average processing time per image using our model is 142.96 s, in which the pre-segmentation step using the *gPb-owt-ucm* method [1] takes 54.44 s, the generation of BCS maps takes 54.67 s, samples selection takes 31.62 s, the generation of MCS maps and ICS maps takes 0.09 s and 1.45 s, respectively, and the fusion takes 0.69 s. The computation cost of our model is higher than other models due to the two main time-consuming parts in our model. One is the pre-segmentation step using the *gPb-owt-ucm* method, which occupies 38.1% processing time but can be effectively parallelized and accelerated using a GPU implementation [5] to a few seconds. The other is the generation of BCS maps by the clustering process, which occupies 38.2% processing time, due that we search for an appropriate

Table 3 F-measures (larger is better) with different values of X and different settings of search range. The best F-measures are marked with boldface

X	2	5	10	15	20	25	30	35
F-measure	0.637	0.760	0.747	0.753	**0.762**	**0.762**	0.757	0.744
V/n in $X = round(V/n)$	$V/9$	$V/8$	$V/7$	$V/6$	$V/5$	$V/4$	$V/3$	$V/2$
F-measure	0.754	0.755	0.759	**0.764**	0.736	0.756	0.748	0.727
V/n in the search range $[\min(round(V/n), 20), 25]$	$V/9$	$V/8$	$V/7$	$V/6$	$V/5$	$V/4$	$V/3$	$V/2$
F-measure	0.773	0.772	0.766	**0.789**	0.771	0.765	0.773	0.767

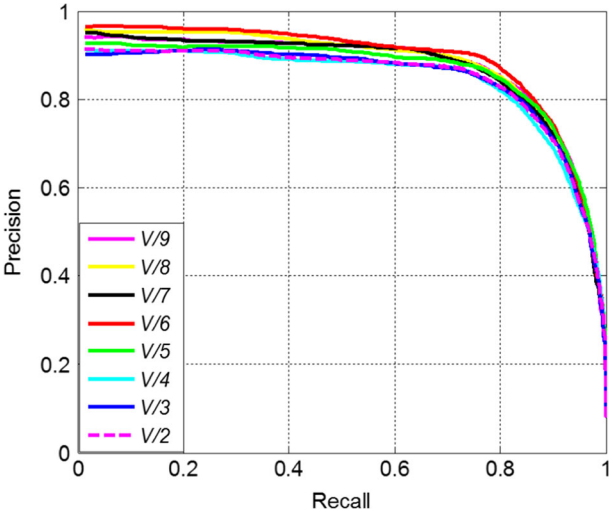

Fig. 5 (Better viewed in color) Performance comparison on different settings of search range of the number of clusters (different settings of V/n in the search range $[\min(round(V/n), 20), 25]$ are shown on the legend)

range of the number of clusters. We also calculated the ART per image using the variant of our model with the K-means clustering method (Ours-K-means), which takes 205.47 s, higher than 142.96 s using our model. Compared to Ours-K-means with the F-measure of 0.647, our model reduces the computation cost with the better co-saliency detection performance.

4 Conclusion

This paper proposes a novel framework for RGBD co-saliency detection via MKB and co-saliency fusion. On the basis of pre-segmented regions at multiple scales, the BCS maps are generated by the regional clustering with feature bagging. Then the clustering-based samples selection is performed to obtain the training samples, which are used to learn a MKB-based regressor. The MCS maps are generated by predicting the co-saliency values of all regions at multiple scales with the learned regressor. Finally, after a linear combination of BCS maps and MCS maps, the final co-saliency maps are generated by using the co-saliency quality weighted fusion. Experimental results on a public RGBD co-saliency detection dataset show that our model consistently outperforms the state-of-the-art co-saliency models.

Acknowledgements This work was supported by the National Natural Science Foundation of China under Grant No. 61771301, and by the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning.

References

1. Arbelaez P, Maire M, Fowlkes C, Malik J (2011) Contour detection and hierarchical image segmentation. IEEE Trans Pattern Anal Mach Intell 33(5):898–916
2. Bezdek JC (2013) Pattern recognition with fuzzy objective function algorithms. Springer Science & Business Media, New York, pp 99–108

3. Borji A, Itti L (2012) Exploiting local and global patch rarities for saliency detection. In: Proc. of IEEE conference on computer vision pattern recognition, pp 478–485
4. Cao X, Tao Z, Zhang B, Fu H, Feng W (2014) Self-adaptively weighted co-saliency detection via rank constraint. IEEE Trans Image Process 23(9):4175–4186
5. Catanzaro B, BY S, Sundaram N, Lee Y, Murphy M, Keutzer K (2009) Efficient, high-quality image contour detection. In: Proc. of IEEE conference on computer vision, pp 2381–2388
6. Cheng MM, Mitra NJ, Huang X, Torr P, SM H (2015) Global contrast based salient region detection. IEEE Trans Pattern Anal Mach Intell 37(3):569–582
7. Cho M, Kwak S, Schmid C, Ponce J (2015) Unsupervised object discovery and localization in the wild: part-based matching with bottom-up region proposals. In: Proc. of IEEE conference on computer vision and pattern recognition, pp 1201–1210
8. Fu H, Cao X, Tu Z (2013) Cluster-based co-saliency detection. IEEE Trans Image Process 22(10): 3766–3778
9. Fu H, Xu D, Lin S, Liu J (2015) Object-based RGBD image co-segmentation with mutex constraint. In: Proc. of IEEE conference on computer vision and pattern recognition, pp 4428–4436
10. Guo J, Ren T, Bei J (2016) Salient object detection for RGB-D image via saliency evolution. In: Proc. of IEEE international conference multimedia and expo, pp 1–6
11. Huang X, Shen C, Boix X, Zhao Q (2015) SALICON: reducing the semantic gap in saliency prediction by adapting deep neural networks. In: Proc. of IEEE international conference on computer vision, pp 262–270
12. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20(11):1254–1259
13. Jacobs DE, Goldman DB, Shechtman E (2010) Cosaliency: where people look when comparing images. In: Proc. of ACM symposium on user interface software and technology, pp 219–228
14. Jiang H, Wang J, Yuan Z, Wu Y, Zheng N, Li S (2013) Salient object detection: a discriminative regional feature integration approach. In: Proc. of IEEE conference on computer vision pattern recognition, pp 2083–2090
15. Ju R, Liu Y, Ren T, Ge L, Wu G (2015) Depth-aware salient object detection using anisotropic center-surround difference. Signal Process Image Commun 38:115–126
16. Kolmogorov V, Zabin R (2004) What energy functions can be minimized via graph cuts? IEEE Trans Pattern Anal Mach Intell 26(2):147–159
17. Lang C, Nguyen T, Katti H, Yadati K, Kankanhalli M, Yan S (2012) Depth matters: influence of depth cues on visual saliency. In: Proc. of European conference on computer vision, pp 101–115
18. Li H, Ngan KN (2011) A co-saliency model of image pairs. IEEE Trans Image Process 20(12):3365–3375
19. Li H, Meng F, Ngan K (2013) Co-salient object detection from multiple images. IEEE Trans Multimedia 15(8):1896–1909
20. Liu Z, Zou W, Le Meur O (2014) Saliency tree: a novel saliency detection framework. IEEE Trans Image Process 23(5):1937–1952
21. Liu Z, Zou W, Li L, Shen L, Le Meur O (2014) Co-saliency detection based on hierarchical segmentation. IEEE Signal Process. Lett. 21(1):88–92
22. Liu N, Han J, Zhang D, Wen S, Liu T (2015) Predicting eye fixations using convolutional neural networks. In: Proc. of IEEE conference on computer vision and pattern recognition, pp 362–370
23. MacKay D (2003) Information theory, inference and learning algorithms. Cambridge University Press, Cambridge, pp 284–292
24. Siva P, Xiang T (2011) Weakly supervised object detector learning with model drift detection. In: Proc. of IEEE international conference on computer vision, pp 343–350
25. Song H, Liu Z, Du H, Sun G (2016) Depth-aware saliency detection using discriminative saliency fusion. In: Proc. of IEEE international conference on acoustics, speech and signal processing, pp 1626–1630
26. Song H, Liu Z, Xie Y, Wu L, Huang M (2016) RGBD co-saliency detection via bagging-based clustering. IEEE Signal Process. Lett. 23(12):1722–1726
27. Song H, Liu Z, Du H, Sun G, Le Meur O, Ren T (2017) Depth-aware salient object detection and segmentation via multiscale discriminative saliency fusion and bootstrap learning. IEEE Trans Image Process 26(9):4204–4216
28. Tong N, Lu H, Ruan X, Yang M (2015) Salient object detection via bootstrap learning. In: Proc. of IEEE conference on computer vision pattern recognition, pp 1884–1892
29. Wang Z, Liu R (2013) Semi-supervised learning for large scale image cosegmentation. In: Proc. of IEEE international conference on computer vision, pp 393–400
30. Wang W, Shen J, Li X, Porikli F (2015) Robust video object cosegmentation. IEEE Trans Image Process 24(10):3137–3148
31. Ye L, Liu Z, Li J, Zhao WL, Shen L (2015) Co-saliency detection via co-salient object discovery and recovery. IEEE Signal Process. Lett. 22(11):2073–2077
32. Zhang D, Fu H, Han J, Wu F (2016) A review of co-saliency detection technique: fundamentals, applications, and challenges. arXiv preprint, arXiv:1604.07090

33. Zhou X, Liu Z, Sun G, Ye L, Wang X (2016) Improving saliency detection via multiple kernel boosting and adaptive fusion. IEEE Signal Process Lett 23(4):517–521

Lishan Wu received the B.E. degree from Zhejiang University of Technology, Hangzhou, China, in 2015. She is currently pursuing the M.E. degree at the School of Communication and Information Engineering, Shanghai University, Shanghai, China. Her research interests include saliency detection and salient object segmentation.

Zhi Liu received the B.E. and M.E. degrees from Tianjin University, Tianjin, China, and the Ph.D. degree from Institute of Image Processing and Pattern Recognition, Shanghai Jiaotong University, Shanghai, China, in 1999, 2002, and 2005, respectively. He is currently a Professor with the School of Communication and Information Engineering, Shanghai University, Shanghai, China. From Aug. 2012 to Aug. 2014, he was a Visiting Researcher with the SIROCCO Team, IRISA/INRIA-Rennes, France, with the support by EU FP7 Marie Curie Actions. He has published more than 150 refereed technical papers in international journals and conferences. His research interests include image/video processing, machine learning, computer vision and multimedia communication. He was a TPC member/session chair in ICIP 2017, PCM 2016, VCIP 2016, ICME 2014, WIAMIS 2013, etc. He co-organized special sessions on visual attention, saliency models, and applications at WIAMIS 2013 and ICME 2014. He is an area editor of *Signal Processing: Image Communication* and served as a guest editor for the special issue on *Recent Advances in Saliency Models, Applications and Evaluations* in *Signal Processing: Image Communication*. He is a senior member of IEEE.

Hangke Song received the B.E. degree from Hangzhou Dianzi University, Hangzhou, China, in 2014, and the M.E. degree from Shanghai University, Shanghai, China, in 2017. His research interests include saliency detection and salient object segmentation.

Olivier Le Meur received the Ph.D. degree from University of Nantes, Nantes, France, in 2005. He was with the Media and Broadcasting Industry from 1999 to 2009. In 2003, he joined the Research Center of Thomson-Technicolor, Rennes, France, where he supervised a Research Project concerning the modeling of human visual attention. He has been an Associate Professor of image processing with the University of Rennes 1 since 2009. In the SIROCCO Team of IRISA/INRIA-Rennes, his current research interests include human visual attention, computational modeling of visual attention, and saliency-based applications, such as video compression, objective assessment of video quality, and retargeting.