

# An efficient segmentation method using saliency object detection

Chongbo Zhou · Chuancai Liu

Published online: 2 February 2014  
© Springer Science+Business Media New York 2014

**Abstract** Automatic co-segmentation is a challenging task because it lacks of prior cues. In this paper, an efficient region contrast based method is proposed for salient object detection and segmentation. The coarse location information of the salient object and the background is first estimated based on the distribution of the detected key-points. Histograms of the estimated foreground and background are calculated as their features. An image is then over-segmented into super-pixels and their histograms are computed. The saliency of a super-pixel is obtained according to the similarity coefficients between the super-pixel and the estimated foreground/background. With the saliency map, the salient object in the image is extracted using a graph cut based optimized framework. The proposed method is compared with state-of-the-art methods on the widely used dataset, and the experiments show that it overall obtains more accurate results.

**Keywords** Object segmentation · Saliency detection · Similarity measurement

## 1 Introduction

Automatic object segmentation is a fundamental research topic which plays an important role in many computer vision applications, such as object detection, object recognition, and scene understanding. Many segmentation methods [6, 12, 20–22, 24] have been proposed to deal with this issue under supervision model. Saliency of images provides prior cues of the foreground objects and unsupervised automatic segmentation based on saliency detection has attracted extensive attention. Recently, visual saliency detection studies [3, 9, 18, 31] have started to focus on the saliency detection in full-resolution, the easy implementation, and the efficiency of an algorithm. In order to meet these requirements, the generic knowledge of salient regions is utilized in the methods, where the knowledge includes frequency information and contrast cues. Meanwhile the position details of salient objects are often assumed or

---

C. Zhou (✉) · C. Liu  
School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing,  
Jiangsu 210094, China  
e-mail: zhouchongbo@163.com

C. Zhou  
School of Physics and Engineering, Qufu Normal University, Qufu, Shandong 273165, China

estimated as strong cues [14, 28, 37]. In this paper, a simple but effective salient object detection and segmentation method is proposed. Firstly, key-points are provided using a corner detection method. A convex hull encompassing the refined key-points is obtained as the estimated salient foreground region, and the other parts of the image are considered as the background. Secondly, the image is over-segmented into super-pixel, and the saliency of each super-pixel is computed according to the Bhattacharyya coefficients between the super-pixel and the estimated foreground/background regions. At last, the salient object is extracted from the background using graph cut based optimization framework.

Many current segmentation methods work in a supervised manner since completely unsupervised segmentation lack the necessary contextual information to accurately separate an image into coherent regions [14, 15]. On the other hand, supervised methods can produce good results, but usually require large workload spent on manually labeled training datasets. Interactive or semi-supervised segmentation methods attempt to address the disparity between fully automatic and fully supervised segmentation, but typically still require human intervention on each individual image. Considering the fact that the same or similar object appear in each image in segmentation, we seek a fully unsupervised method. In this paper, a simple but effective salient object detection and segmentation is proposed. Firstly, key-points are provided using a corner detection method. A convex hull encompassing the refined key-points is obtained as the estimated salient foreground region, and the other parts of the image are considered as the background. Secondly, the image is over-segmented into super-pixel, and the saliency of each super-pixel is computed according to the Bhattacharyya coefficients between the super-pixel and the estimated foreground/background regions. At last, the salient object is extracted from the background using graph cut based optimization framework.

The remainder of this paper is organized as follows. Super-pixel representation (including how to over-segment an image into super pixels) and their similarity measurements are introduced in Section 2. The selection and refinement of seed super-pixels are detailed in Section 3. With the saliency map, the salient object in the image is extracted using a graph cut based optimized framework in Section 4. Experiments are performed and results are provided in Section 5. Finally, concluding remarks are given in Section 6.

## 2 Representations of super-pixels and their similarity measurements

In our method, an initial segmentation is required to partition the image into super-pixels, i.e. homogeneous regions, for merging. Merging super-pixels rather than merging pixels is often used for object segmentation [20–22]. In this paper, we use mean shift method proposed in [10] for initial segmentation because it can well preserve the object boundaries which contribute to excellent results.

A super-pixel can be described in many aspects, such as the color, edge, texture, shape and size of the region [5, 30]. Among them the color histogram is an effective descriptor to represent the object color feature statistics and it is widely used in pattern recognition and object tracking [11, 34]. In the context of region merging based segmentation, the color histogram is more robust than other feature descriptors. This is because the initially segmented super-pixels of the desired object often vary a lot in size and shape, while the colors of different regions from the same object will have high similarity [29]. Besides the color cues, texture features have also been proven effective in texture classification, image segmentation, image retrieval, and generic object recognition [25, 27, 36, 38]. In order to incorporate texture cues into color histograms, we exploit color histograms used in [32] to represent the super-pixels. The histogram representation of [32] considers color and texture cues simultaneously since

each pixel is represented by a region centered on it. In our current implementation, each pixel is represented by its  $3 \times 3$  neighborhood. To reduce the computational cost, the image is first quantized in each channel in RGB color space. We uniformly quantize each color channel into 16 levels and produce a single-channel index image where the intensity of pixels may have  $U = 16 \times 16 \times 16 = 4096$  different values. Therefore, the histogram of each super-pixel is calculated in the feature space of 4096 bins. This case is equivalent to the case in [32] where the number of cluster centers in the K-means clustering operation is set to 4096. The histogram of a super-pixel is computed as follows. The pixels in the super-pixels are firstly replaced by the neighbor pixels (including themselves) in the  $3 \times 3$  neighborhood in feature space. Then all these pixels are counted to produce histograms. The histograms are normalized to eliminate the effect of the area of the super-pixels.

After obtaining histograms of super-pixels, we seek a way to measure the similarity between two super-pixels. There are some well-known goodness-of-fit statistical measurements such as Euclidean distance, Bhattacharyya coefficient, and the log-likelihood ratio statistic. Here we choose the Bhattacharyya coefficient used in [11, 29] to measure the similarity between super-pixels. Let  $R$  and  $Q$  denote two super-pixels and  $H_R$  and  $H_Q$  denote their color histograms. The similarity is defined as follows:

$$\text{sim}(R, Q) = \sum_{u=1}^U \sqrt{H_R^u \cdot H_Q^u} \quad (1)$$

where  $H_R^u$  and  $H_Q^u$  are the corresponding  $u$ th elements in the normalized histograms  $H_R$  and  $H_Q$  respectively, and  $U$  denotes the number of bins in the histogram. Unlike the Euclidean distance, the higher the Bhattacharyya coefficient of  $R$  and  $Q$  is, the higher their similarity is.

We also tried the Euclidean distance here, but the result is much worse, which is partly due to the fact that, the Bhattacharyya distance can measure not only the means of sample sets, but the standard deviations as well.

### 3 Selection of seed super-pixels using saliency maps

In our method, no priors are provided unlike interactive or supervised segmentation where interactive strokes or training images are provided. Seed super-pixels are detected automatically and the combination of the results of different saliency detection methods on different types of color space is exploited to produce initial seed regions. The initial seed regions are further refined by eliminating the dissimilar ones based on the assumption that the detected foreground super-pixels in all the images are highly similar unless they are mixed with background super-pixels.

#### 3.1 Finding saliency map using different methods on different types of color space

It is widely recognized that salient object detection is very helpful in computer vision and image processing [1, 35]. However, it is still a challenging task because there is no method yet in the current literature that can detect saliency accurately for all images. In order to achieve robust saliency detection, we adopt a strategy of linear combination of different detection result as [23]. Unlike [23], we perform saliency detection on different types of color space using different detection methods. The combined saliency map denoted by  $S$  is represented as follows:

$$S = \sum_{k=1}^K \sum_{l=1}^L S_l^k \quad (2)$$

where  $S_l^k$  denotes the normalized saliency map obtained using method  $k$  on the color space type  $l$ . Note that, here we implicitly set the weight of all saliency detection algorithms and all color spaces to be exact the same. From Eq. (2), we can see that if a pixel is identified as a salient pixel by most of algorithms on most of color space, it will have a high saliency value. Otherwise, it will be regarded as a background pixel.

Note that, we never implicitly assume that, either the proposition “salient pixels are all similar” or “salient pixels belong to the foreground” is true. We only assume that, salient super-pixels can serve as good seeds in most cases for the initialization of the process of our proposed method.

In our work, we exploited three methods for generating saliency maps on three types of color space. The three saliency maps are Itti’s model saliency [17], frequency-tuned saliency [2], and spectral residual saliency [16] since the combination of these three methods have been proven effective [23]. The first is the well-known saliency model which mimics the visual search process of human. The saliency map is computed using multi-scale image features in a bottom-up manner. The second estimates the center-surround contrast using color and luminance features based on a frequency-tuned approach. The third saliency model employs the log-spectrum of an input image, and extracts the spectral residual of an image in the spectral domain. The three saliency models perform the saliency detection in different ways, and advantages of each method are expected to be exploited in such a combination. The three types of color space are RGB, Lab and HSI. Pixels in each detection result have normalized salient value in the range [0, 1]. We use adaptive threshold to determine which pixel is a salient point. The adaptive threshold  $T_a$  is computed as [2] as follows:

$$T_a = \frac{2}{W \times H} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} S_l^k(x, y) \tag{3}$$

where  $S_l^k(x, y)$  is corresponding saliency map,  $(x, y)$  is the position of a pixel,  $H$  is the height of the image, and  $W$  is the width.

### 3.2 Refinement of seed super-pixels

Refining the initially selected elements according to certain rules is usually a necessary process in seed selection in order to obtain more accurate results [6, 12]. The implicit assumption in co-segmentation is that the images contain an identical or similar object. This can help to refine the seed regions since the super-pixels belonging to foreground class have high similarity. The background super-pixels which are misclassified as seed super-pixels are regarded as outliers. We simply use statistical methods for multivariate outlier detection to detect these outliers and eliminate them. Suppose that  $O_{seed}^q = \{R_1^q, R_2^q, \dots, R_{N^q}^q\}$  is the seed set including  $N^q$  seed super-pixels corresponding to the image  $I^q$  where  $q=1, 2, \dots, N$ , and  $N$  is the number of images to be co-segmented simultaneously. All the seed sets have  $N^{all} = \sum_{q=1}^N N^q$  seed super-pixels.

Statistical methods for multivariate outlier detection often indicate those observations that are located relatively far from the center of the data distribution [4]. Several distance measures can be implemented for such a task. The Mahalanobis distance is a well-known criterion which depends on estimated parameters of the multivariate distribution. Given  $N^{all}$  observations whose number of dimensions are  $U$ , denote the sample mean vector by  $\bar{H}$  and the sample covariance matrix by  $V$ , thus

$$V = \frac{1}{N^{all}-1} \sum_{q=1}^N \sum_{i=1}^{N^q} (H_i^q - \bar{H})(H_i^q - \bar{H})^T \tag{4}$$

The Mahalanobis distance for each multivariate data point  $n, n=1, 2, \dots, N^{all}$ , is denoted by  $M_n$  and given by

$$M_n = \left( \sum_{i=1}^n (H_1^q - \bar{H}) V^{-1} (H_1^q - \bar{H})^T \right)^{1/2} \tag{5}$$

Accordingly, those observations with a large Mahalanobis distance are indicated as outliers. The outliers sorting in descending order are eliminated one by one, until one of the set has only one element per image.

#### 4 Object extraction using graph cut based optimization

To extract a salient object from an image using the saliency map, the simplest method is to segment the saliency map using a fixed or adaptive threshold [6]. Though the threshold method is simple, it often cannot provide satisfactory segmentation results. Therefore more sophisticated schemes are needed and many saliency map models have been applied to object segmentation. Ma and Zhang [37] use region growing and fuzzy theory on their saliency maps to locate objects of attention with rectangles. Han et al. [8] employ a Markov random field to group pixels of the prominent object based on their saliency maps and low level features such as color, texture and edge. Achanta et al. [6] average saliency values within segments and realize object extraction using an adaptive threshold method. Cheng et al. [24] apply the saliency map results to initialize the iterative Grab cut algorithm [7] for object segmentation. Xie et al. [9] use the graph cut algorithms [19] in their work. In our method, given the images and corresponding saliency maps, the object extraction is also formulated as a labeling problem as [7, 19].

In the labeling based representation, for an image  $I$  containing  $N$  super-pixels, each super-pixel  $S_i$  is assigned a label  $l_i$ , where  $i=1, 2, \dots, N$ . For figure-ground segmentation, i.e. the 2-class segmentation case, the label  $l_i$  takes 0 or 1, where 0 denotes the background label and 1 denote the foreground label. Each super-pixel is labeled as foreground or background, and thus there are  $2^N$  labeling schemes for the image. In the viewpoint of energy minimization, each labeling corresponds to a potential energy value, also called a penalty cost. Not only individual super-pixels but also neighboring ones are considered when computing the energy values [26]. Denote a labeling by  $L=(l_1, l_2, \dots, l_N)$ , and an energy function  $\mathbb{E} : 2^N \rightarrow R$ , which maps any labeling  $L$  to a real number  $\mathbb{E}(L)$ , can be written as:

$$\mathbb{E}(L) = \lambda \cdot \sum_{i \in V} u(l_i | S_i) + \sum_{(i,j) \in E} p(l_i, l_j | S_i, S_j) \cdot \delta(l_i, l_j) \tag{6}$$

Where  $V$  denotes the set of indexes of pixels,  $E$  denotes the set of all adjacent super-pixel pairs, the unary term  $u(l_i | S_i)$  denotes the cost of assigning  $l_i$  to  $S_i$ , the pair-wise term  $p(l_i, l_j | S_i, S_j)$  denotes the cost of assigning  $l_i$  and  $l_j$  to adjacent  $S_i$  and  $S_j$ ,  $\delta(l_i, l_j)$  is an indicator function, and  $\lambda$  is a positive weighing coefficient specifying a relative importance of the unary term versus the pair-wise term. In our work, the values of a saliency map is normalized to the scope of [0, 1], and the saliency value of each super-pixel is regarded as the probability that the super-pixel belongs to the foreground. Then the unary term  $u(l_i | S_i)$  is defined as:

$$u(l_i = 1 | S_i) = -\ln Sal(S_i) \tag{7}$$

$$u(l_i = 0 | S_i) = -\ln(1 - Sal(S_i)) \tag{8}$$

The pair-wise term is defined as

$$p(l_i, l_j | S_i, S_j) = \exp\left(-\frac{\|f_{S_i} - f_{S_j}\|^2}{2\sigma^2}\right) \quad (9)$$

where  $\sigma$  is standard deviation of histograms, and  $\|\cdot\|$  denotes the Euclidean distance between the two histograms  $f_{S_i}$  and  $f_{S_j}$ .

## 5 Experiments and results

In this section, experiments were performed to evaluate the performance of the proposed method. The proposed method was then compared with some state-of-the-art ones on two datasets: the MSRC-v2 dataset and a dataset used in [1]. Finally, the deficiency of our current method was discussed.

### 5.1 Evaluation of the saliency detection methods

For the dataset used in [6], 1000 images with labeled ground truths provided selected from 5000 consistent images of a public dataset given by [6]. The proposed method is compared with 3 state-of-the-art methods, including FT [6], RC [24], and BS [9], which output full-resolution saliency maps and almost outperform the other methods in current published work [9, 24].

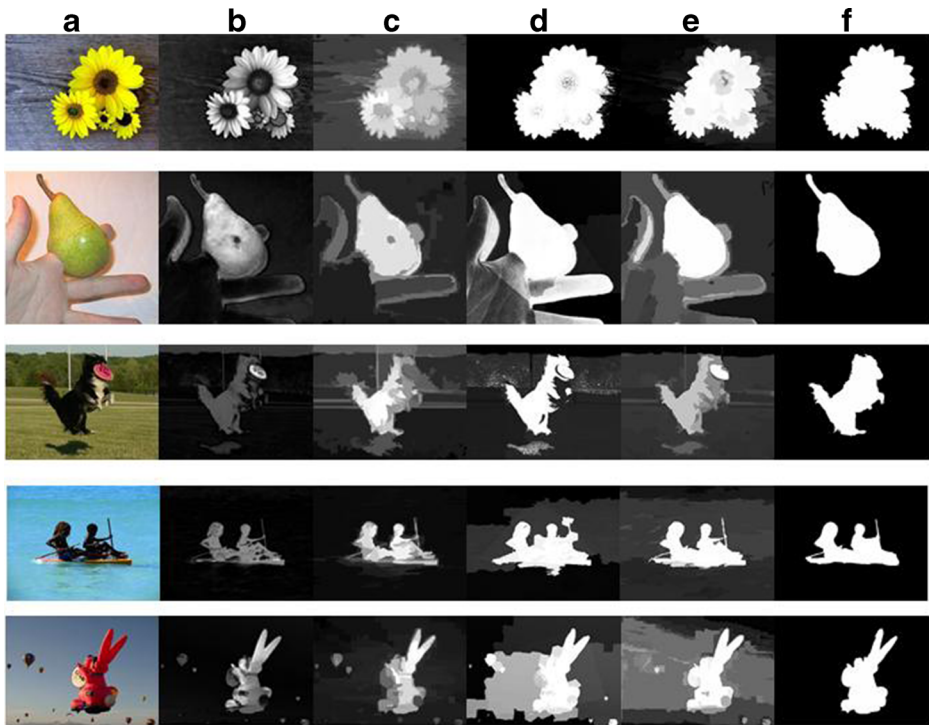
Some sample saliency detection results are shown in Fig. 1. The FT method computes saliency at each pixel by its color contrast to the average of the whole image, but it does not work well when the salient region and the background have similar color. The RC method estimates regional saliency based on color as well as spatial positions, and it performs well in highlighting small salient objects. However, it also inevitably identifies small background patches incorrectly. The BS method can provide a strong saliency map, which highlights the salient object and weakens the background regions simultaneously; nevertheless it probably ignores small distinct parts in the estimated hull. The proposed method generates uniformly highlighted salient regions, and overall, it is able to better estimate saliency maps.

The true usefulness of a saliency map is determined by the application, and the saliency maps are evaluated in the context of salient object segmentation. To quantitatively evaluate the performance of the proposed method, the precision and accuracy are computed to quantitatively evaluate the performance. The precision and recall are defined as

$$\text{Precision} = \frac{TP}{TP + FP} \quad (10)$$

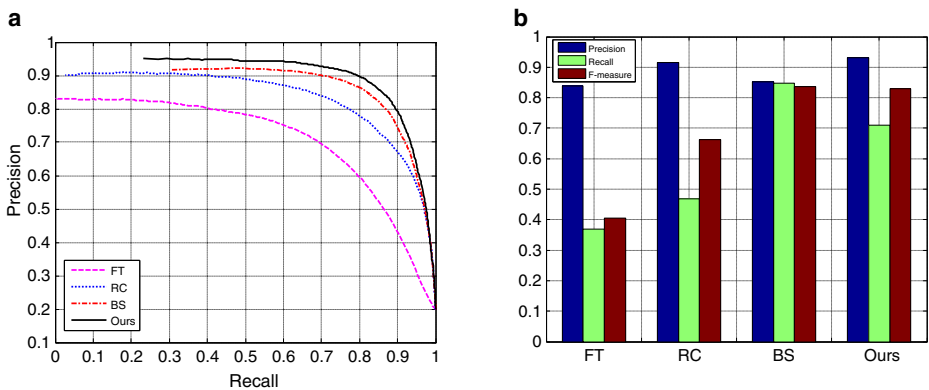
$$\text{Recall} = \frac{TP}{TP + FN} \quad (11)$$

where  $TP$  (True Positives),  $FP$  (False Positives), and  $FN$  (False Negatives) denote the number of correctly classified object pixels, the number of background pixels but classified as object, and the number of object pixels but classified as background respectively. Obviously, the higher the precision and recall are, the better the method is. The precision and recall values on the 1000 images are computed by varying the threshold from 0 to 255 as [6]. The resulting precision and recall curves are shown in Fig. 2a. This curve provides a reliable comparison of how well various saliency maps highlight salient regions in images. As shown in the curve



**Fig. 1** Visual comparison of saliency maps: **a** original images, **b** FT maps, **c** RC maps, **d** BS maps, **e** our maps, and **f** ground-truths

objectively, the proposed method achieves the best results. At the maximum recall, all methods have the same low precision value. This happens at the threshold zero, where all pixels from the saliency maps of each method are classified as positives, leading to an equal value for true and false positives for all methods.



**Fig. 2** Quantitative results: **a** precision-recall curves for naive thresholding of saliency maps on 1000 images, **b** mean precision, recall, and F-measure values for our segmentation method using different saliency maps as initialization

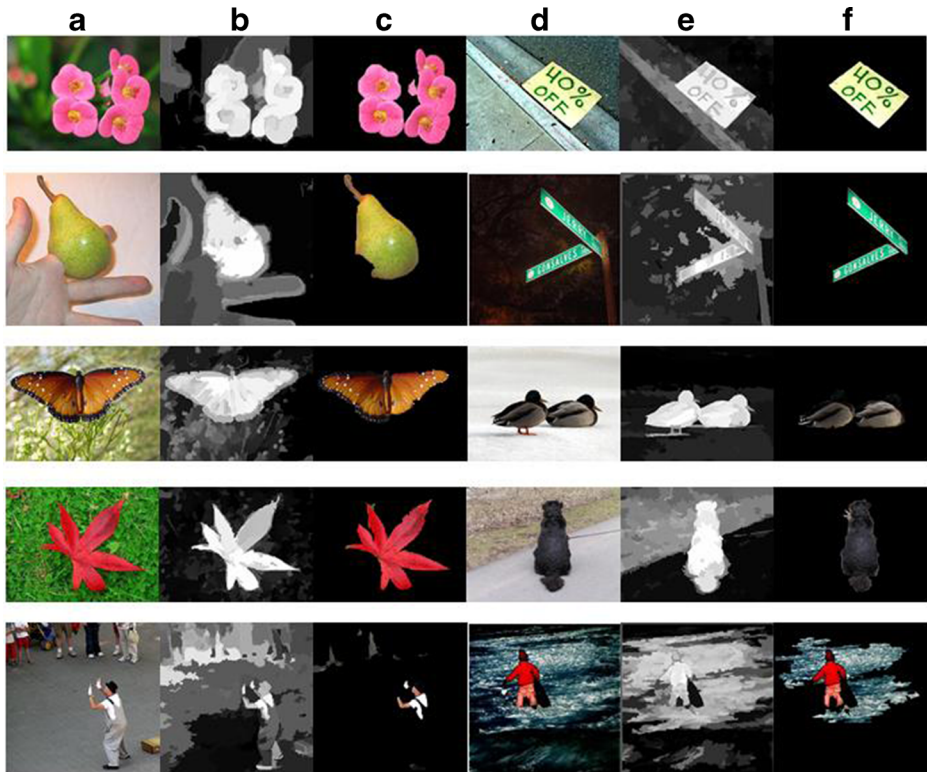


## 5.2 Effectiveness of the proposed method

Since the proposed method is designed for automatically unsupervised segmentation, especially for extraction of homogenous object from the background, it will perform well on images with homogenous foregrounds. Figure 3 shows some high quality results on image groups where the number of images in each group is more than 2. As mentioned before, it can be seen from the results that the proposed method can obtain satisfactory segmentations if the objects occurring in the image pair or group have consistent color or texture features.

## 5.3 Experiments on MSRC-v2 dataset

In this subsection, performance was evaluated on a subset of MSRC-v2 database [33]. The proposed method was compared with a variety of state-of-the-art methods on the dataset. The performance measurement used here is the IU defined above. Table 1 gives quantitative results. In comparison to results on other datasets, the performance on MSRC dataset is relatively low. This is mainly because that many image groups in this dataset depict different instances of the same class. Note that the images in MSRC-v2 dataset have multiple labels. We used the main object category for each MSRC image as foreground, and the rest of the pixels as background as [19]. The numbers in brackets in the first column denote the numbers of images in the corresponding image groups. The proposed method achieves the best performance for 9 out of 14 object classes.



**Fig. 3** Some visual results using saliency maps: columns **a** and **d** are original images, **b** and **e** are saliency maps, **c** and **f** are segmentation results



**Table 1** Performance comparison on MSRC-v2 dataset with various settings of BiCos [7]

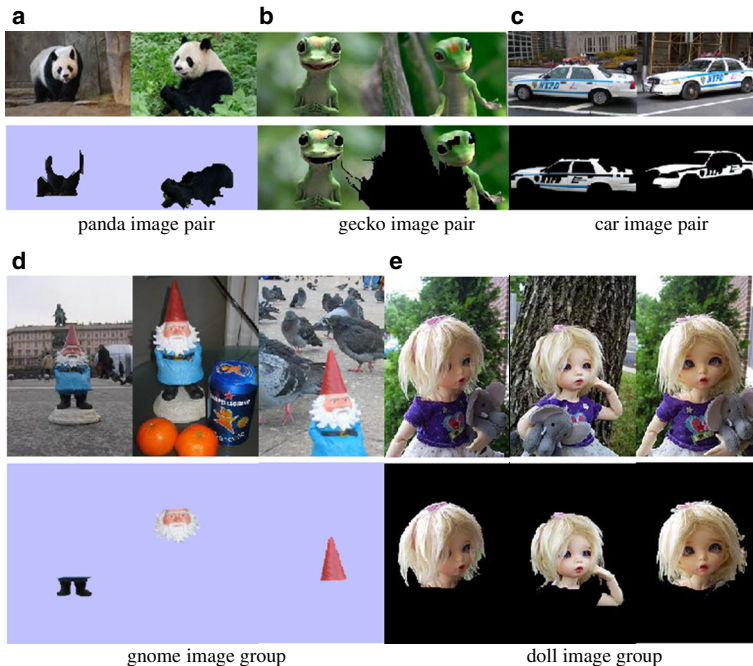
Class	BiCos [7] (1 iter.)	BiCos [7] (5 iter.)	BiCos-MT [7] (1 iter.)	BiCos-MT [7] (5 iter.)	Our proposed
Bike (30)	28.4	29.3	29.9	30.3	31
Bird (30)	70.2	71.2	71.2	71.2	71
Car (30)	59.5	59.9	60.3	60.5	62
Cat (24)	71.4	71.5	71.2	71.5	73
Chair (30)	41.8	42.0	42.6	43.4	46
Cow (30)	73.8	74.2	74.6	74.6	77
Dog (26)	66.4	66.9	67.1	67.2	69
Face (30)	28.6	28.9	28.6	28.8	29
Flower (30)	59.3	59.9	60.7	61.4	62
House (30)	40.0	40.7	41.0	41.7	42
Plane (30)	50.0	50.2	50.1	50.8	52
Sheep (30)	71.4	71.5	72.3	72.8	76
Sign (30)	34.8	35.1	35.4	36.0	37
Tree (30)	61.9	62.2	62.1	62.4	63
Average	54.1	54.6	54.8	55.2	56.4

The reason for the good performance is that the images have homogenous objects or the objects have distinct features. The proposed method works worst in bike, face and sign object categories. The reason for the bad performance is that object instances in these categories vary greatly in color or texture with cluttered backgrounds, which leads to poor results.

Finally, two versions of the well-known Bi-level Co-Segmentation (BiCos) method [7] with various settings are compared and tabulated accordingly in Table 1. It is seen that, even for BiCos-MT with 5 iterations, our proposed method is still with some slight advantages.

#### 5.4 Discussions

The proposed method is designed for unsupervised foreground/background segmentation, i.e. automatic extraction of objects from the backgrounds, based on the fact that the same or similar object appears in each image. The accuracy of seed detection and the performance of merging process highly depend on the assumption that each image has homogenous foreground or background. To ensure accurate seeds as possible, the initial detection result is further refined by eliminating possible outliers. On the other hand, the refinement weakens the diversity of foreground super-pixels and may result in an incomplete seed results. The deliberately designed merging process is fast and easy to implement, and it performs well on the images which satisfy the assumption. However, the proposed method performs poorly on image groups which do not satisfy the assumption. Figure 4 demonstrates some low quality results encountered in our experiments. In Fig. 4a, the panda has two classes of distinct white and black parts and only one class is left after the refinement; therefore, the incomplete seeds lead to a partial result. In Fig. 4b, the gecko has the similar color with the background and it result in misclassification of background parts as foreground ones. The car in Fig. 4c has several kinds of distinct parts and only saliency parts are segmented out as the case of Fig. 4a. There are three images in the gnome image group of Fig. 4d, but saliency regions in each image correspond to different parts of the gnome. The saliency region in the first image is the boots, the saliency region in the second image is the face, and the saliency region in the third image is the hat. The different saliency regions



**Fig. 4** Some low quality results. **a** panda image pair. **b** gecko image pair. **c** car image pair. **d** gnome image group. **e** doll image group

result in poor segmentations. In Fig. 4e, the true object is the doll, but the detected saliency regions of the images are the faces, parts of the doll. In our future work, we will introduce interactive inputs instead of refinement to ensure accurate seeds and meanwhile preserve diversity of foreground; and we will exploit more discriminative representation for super-pixels to make the merging process more robust.

## 6 Conclusions

In this paper, a contrast based saliency detection method is proposed and foreground object segmentation using saliency maps is performed. The method is simple, but efficient and effective. Experiments are performed on the widely used dataset, and the results show that the proposed method overall produces more accurate detection results than state-of-the-art methods.

**Acknowledgments** This work is partially supported by the National Natural Science Fund of China (Grant Nos. 60632050, 9082004, 61202318), and the Technology Project of provincial University of Fujian Province (Grant No. JK2011040).

## References

1. Achanta R, Estrada F, Wils P, Susstrunk S (2008) Salient region detection and segmentation. In: Proceedings of International Conference on Computer Vision Systems, Santorini, Greece, pp 66–75
2. Achanta R, Hemami S, Estrada F, Susstrunk S (2009) Frequency tuned salient region detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami, USA, pp 1597–1604

3. Batra D, Kowdle A, Parikh D, Luo J, Chen T (2010) iCoseg: interactive co-segmentation with intelligent scribble guidance. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, USA, pp 3169–3176
4. Ben-Gal I (2005) Outlier detection. In: Maimon O, Rockach L (Eds) Data mining and knowledge discovery handbook: a complete guide for practitioners and researchers, Kluwer Academic Publishers, pp 131–148
5. Birchfield S (1998) Elliptical head tracking using intensity gradients and color histograms. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Santa Barbara, USA, pp 232–237
6. Carreira J, Sminchisescu C (2010) Constrained parametric min-cuts for automatic object segmentation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, USA, pp 3241–3248
7. Chai Y, Lempitsky V, Zisserman A (2011) BiCoS: a bi-level co-segmentation method for image classification. In: Proceedings of International Conference on Computer Vision, Barcelona, Spain, pp 2579–2586
8. Chang K, Liu T, Lai S (2011) From co-saliency to co-segmentation: an efficient and fully unsupervised energy minimization model. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, USA, pp 2129–2136
9. Cheng D, Figueiredo M (2007) Cosegmentation for image sequences. In: Proceedings of International Conference on Image Analysis and Processing, Modena, Italy, pp 635–640
10. Comaniciu D, Meer P (2002) Mean shift: a robust approach toward feature space analysis. *IEEE Trans Pattern Anal Mach Intell* 24(5):603–619
11. Comaniciu D, Ramesh V, Meer P (2003) Kernel-based object tracking. *IEEE Trans Pattern Anal Mach Intell* 25(5):564–577
12. Endres I, Hoiem D (2010) Category independent object proposals. In: Proceedings of European Conference on Computer Vision, Heraklion, Greece, pp 575–588
13. Hochbaum D, Singh V (2009) An efficient algorithm for co-segmentation. In: Proceedings of International Conference on Computer Vision, Kyoto, Japan, pp 269–276
14. Holzinger A (2012) On knowledge discovery and interactive intelligent visualization of biomedical data—challenges in human-computer interaction & biomedical informatics. Conference on e-Business and Telecommunications (ICETE 2012), pp IS9–IS20
15. Holzinger A (2012) Biomedical informatics: computational sciences meets life sciences. BoD, Norderstedt
16. Hou X, Zhang L (2007) Saliency detection: a spectral residual approach. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, USA, pp 1–8
17. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. *IEEE Pattern Anal Mach Intell* 20(11):1254–1259
18. Joulin A, Bach F, Ponce J (2010) Discriminative clustering for image co-segmentation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, USA, pp 1943–1950
19. Joulin A, Bach F, Ponce J (2012) Multi-class cosegmentation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Providence, USA, pp 542–549
20. Kumar M, Turki H, Preston D, Koller D (2011) Learning specific-class segmentation from diverse data. In: Proceedings of International Conference on Computer Vision, Barcelona, Spain, pp 1800–1807
21. Lee Y, Grauman K (2010) Collect-cut: segmentation with top-down cues discovered in multi-object images. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, USA, pp 3185–3192
22. Lee Y, Grauman K (2012) Object-graphs for context-aware visual category discovery. *IEEE Trans Pattern Anal Mach Intell* 34(2):346–358
23. Li H, Ngan KN (2011) A co-saliency model of image pairs. *IEEE Trans Image Process* 20(12):3365–3375
24. Liu G, Lin Z, Yu Y, Tang X (2010) Unsupervised object segmentation with a hybrid graph model. *IEEE Trans Pattern Anal Mach Intell* 32(5):910–924
25. Liu G, Zhang L, Hou Y, Li Z, Yang J (2010) Image retrieval based on multi-texton histogram. *Pattern Recogn* 43(7):2380–2389
26. Long Y, Huang Y (2006) Image based source camera identification using demosaicking. In: Proceedings of IEEE 8th Workshop on Multimedia Signal Processing, Victoria, Canada, pp 419–424
27. Malik J, Belongie S, Leung T, Shi J (2001) Contour and texture analysis for image segmentation. *Int J Comput Vis* 43(1):7–27
28. Mukherjee L, Singh V, Dyer C (2009) Half-integrality based algorithms for cosegmentation of images. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami, USA, pp 2028–2035
29. Ning J, Zhang L, Zhang D, Wu C (2010) Interactive image segmentation by maximal similarity based region merging. *Pattern Recogn* 43(2):445–456
30. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell* 24(7):971–987
31. Rother C, Kolmogorov V, Minka T, Blake A (2006) Cosegmentation of image pairs by histogram matching—incorporating a global constraint into mrf. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, New York, USA, pp 993–1000

32. Schroff F, Criminisi A, Zisserman A (2006) Single-histogram class models for image segmentation. In: Proceedings of the Indian Conference Computer Vision, Graphics and Image Processing, Madurai, India, pp 82–93
33. Shotton J, Winn J, Rother C, Criminisi A (2006) Textonboost: joint appearance, shape and context modeling for multi-class object recognition and segmentation. In: Proceedings of European Conference on Computer Vision, Graz, Austria, pp 1–15
34. Swain M, Ballard D (2002) Color indexing. *Int J Comput Vis* 7(1):11–32
35. Tang L, Li H (2010) Extract salient objects from natural images. In: Proceedings of International Symposium on Intelligent Signal Processing and Communications Systems, Chengdu, China, pp 1–4
36. Varma M, Zisserman A (2005) A statistical approach to texture classification from single images. *Int J Comput Vis* 62(1):61–81
37. Vicente S, Kolmogorov V, Rother C (2010) Cosegmentation revisited: models and optimization. In: Proceedings of European Conference on Computer Vision, Heraklion, Greece, pp 465–479
38. Winn J, Criminisi A, Minka T (2005) Object categorization by learned universal visual dictionary. In: Proceedings of International Conference on Computer Vision, Beijing, China, pp 1800–1807



**Chongbo Zhou** received the B.S. degree in Electronics and Information Engineering from Qufu Normal University in 2002, and the M.S. degree in Circuits and Systems from Shanghai University in 2005. In 2013, he received the Ph.D. degree in Computer Science from Nanjing University of Science and Technology, China. His research interests include image segmentation and pattern recognition.



**Chuancai Liu** is a full professor in the school of computer science and engineering of Nanjing University of Science and Technology, China. He obtained his Ph.D. degree from the China Ship Scientific Research Center in 1997. His research interests include pattern recognition and computer vision.