Research Article

Salient Object Detection Based on Background Feature Clustering

Kan Huang,[1,2] Yong Zhang,[1,2] Bo Lv,[1,2] and Yongbiao Shi[1,2]

[1]*University of Chinese Academy of Sciences, Beijing 100049, China*
[2]*Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China*

Correspondence should be addressed to Kan Huang; huangkan15@mails.ucas.ac.cn

Automatic estimation of salient object without any prior knowledge tends to greatly enhance many computer vision tasks. This paper proposes a novel bottom-up based framework for salient object detection by first modeling background and then separating salient objects from background. We model the background distribution based on feature clustering algorithm, which allows for fully exploiting statistical and structural information of the background. Then a coarse saliency map is generated according to the background distribution. To be more discriminative, the coarse saliency map is enhanced by a two-step refinement which is composed of edge-preserving element-level filtering and upsampling based on geodesic distance. We provide an extensive evaluation and show that our proposed method performs favorably against other outstanding methods on two most commonly used datasets. Most importantly, the proposed approach is demonstrated to be more effective in highlighting the salient object uniformly and robust to background noise.

1. Introduction

As an important preprocessing technique in computer vision to reduce computational complexity, saliency has attracted much attention in recent years. Saliency detection has boosted a wide variety of applications in computer vision tasks, such as image retrieval [1, 2], image compression [3], content-aware image editing [4, 5], and object detection and recognition [6].

According to previous works, saliency models can be categorized into bottom-up approaches [7–9] and top-down approaches [10, 11]. Most top-down approaches make use of prior knowledge and are based on supervised learning. While these supervised approaches can effectively detect salient regions and perform overall better than bottom-up approaches, it is still expensive to perform the training process, especially data collection. Compared to top-down approaches, bottom-up based approaches do not need data collection and training process, meanwhile requiring little prior knowledge. These advantages make bottom-up approaches more efficient and easy to implement in a wide range of real computer vision applications. In this paper, we focus on bottom-up approach.

In recent years, most works deploy color contrast [7, 8] as the measurement of saliency. And some other works utilize color rarity [12] to produce saliency map. Algorithms based on these cues usually work well in situations where foreground object appears to be distinctive in terms of contrast or rarity. But they may fail in dealing with images containing various background patterns. Work in [13] tends to focus on pattern distinctness in saliency detection. It defines a pixel to be salient if the patch corresponding to that pixel has less contribution to the whole image in PCA system. So it will perform badly in cases where salient object is overwhelming in size in an image. By assuming most of the narrow border of the image as background region, information of backgroundness priors can be exploited in calculating saliency maps, as done in [14, 15]. Though backgroundness priors seem to play a role in locating salient objects, treating patches on the border directly as background patches would incur serious problems, especially when salient objects touch the image border or present similar features to a certain part of the image border. Thus, looking for an effective way to construct background distribution, from which salient objects can be separated, seems to be necessary.
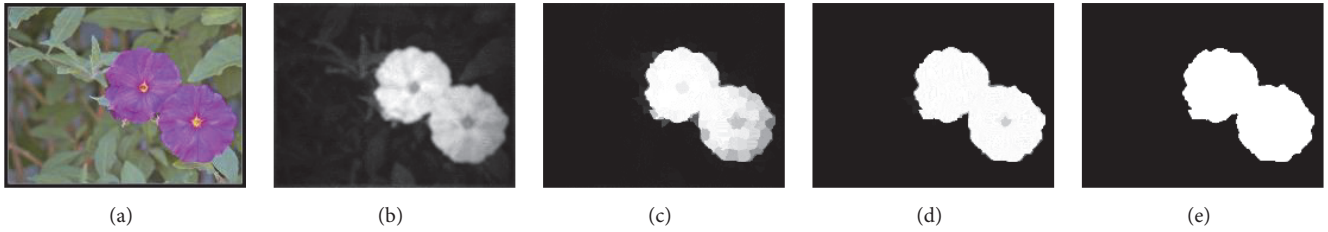
FIGURE 1: Main framework of the proposed approach. (a) Input image. (b) Coarse saliency map. (c) Saliency map after edge-preserving element-level filtering. (d) Final saliency map after geodesic distance refinement. (e) Ground truth.

In this work, we suggest that clustering features within the pseudobackground region (border region of the image) would give a suitable representation of the background distribution. Thus we first construct background distribution in this way and then calculate saliency value of every location as the distance from background distribution. In the following, we employ a series of refinement methods to enhance the salient region and make salient objects stand out uniformly.

There are two main contributions underlying our proposed approach: (1) we propose a novel way to construct background distribution of an image based on clustering method. In this way, statistical and structural information of background can be fully exploited to construct background distribution, which would allow for a subsequently more accurate saliency measurement, and (2) we propose a new refinement framework, which is composed of edge-preserving element-level filtering and upsampling based on geodesic distance. The new refinement framework has an excellent ability to uniformly highlight the salient object and removing background noise.

2. Proposed Approach

Our proposed approach adopts a coarse-to-fine strategy in which every step fully utilizes the information from previous step. There are three main steps composing the proposed approach. First, the patches from image border are used to construct background distribution from which a coarse saliency map is derived. Second, the coarse saliency map is enhanced by an edge-preserving element-level filtering. Third, final saliency map is obtained by refinement based on geodesic distance. The main framework is depicted in Figure 1. Each step will be illustrated in detail in each section below.

2.1. Background Distribution. Since backgroundness prior has shown its importance in salient object detection, we aim to find an effective way to establish the background distribution, from which saliency map can be calculated. Similar to work in [11], we define the pseudobackground region as the border region of the image. Some improved methods [15–17] separate the border region into several certain parts, such as top, bottom, left, and right, and deal with them, respectively. To our consideration, this may fail to exploit as much as possible the statistical information about backgroundness priors. In our approach, we construct the background distribution with the help of clustering algorithm. To implement

clustering on all patch features within pseudobackground region, provide a way to construct background distribution, which allows for automatically and substantially exploiting statistical information of the backgroundness prior. Work in [18] also uses clustering algorithm to establish background distribution and thus generate a coarse saliency map based on the distribution. Different from our method, it only considers color information in superpixel level. By contrast, patch feature used in our method represents not only color information but also structural information of a neighborhood of a location. The effectiveness of patch feature has already been demonstrated in [13, 19].

For the feature representation of each patch, we choose to use the feature in CIE LAB color space, which possesses the property of perceptual uniformity and has also proved its effectiveness in many saliency algorithms [20, 21]. Thus, for example, if we take a 7×7 patch from an image, the feature of this patch will be represented by a 147×1 column vector (pixelNumber $\times$ channelNumber). This kind of representation can represent both color and pattern information. Here, we denote this vectorized feature for image patch i to be $f(i)$.

After patch features have been drawn from the border region of an image, we treat them as a collection of background feature samples and use K-means clustering algorithm [22] to cluster the samples. Compared to separating border region into certain parts and dealing with them, clustering the features within border region into different groups would provide a novel way to exploit statistical and structural information of the backgroundness prior. Consequently, this would find clusters that best represent background feature distribution. Larger clusters have more contribution to background distribution and smaller clusters have less contribution. Through exercising with different numbers of clusters, we found that clustering the samples into four clusters could maintain relatively good performance; thus we choose to cluster the features into four groups. From the above process, we can get four feature clusters, sample number of each cluster is denoted $N(l)$, and centroid location value of each cluster is denoted, $C(l)$, $l \in \{1, 2, 3, 4\}$.

In most of the previous work, a variety of measurements have been used to measure image saliency. The most commonly used measurement is rarity and contrast. But, as stated in Section 1, such measurement may not be suitable for more general cases. A more intuitive saliency measurement can be defined as the difference from background distribution which describes the backgroundness prior. In our approach, we
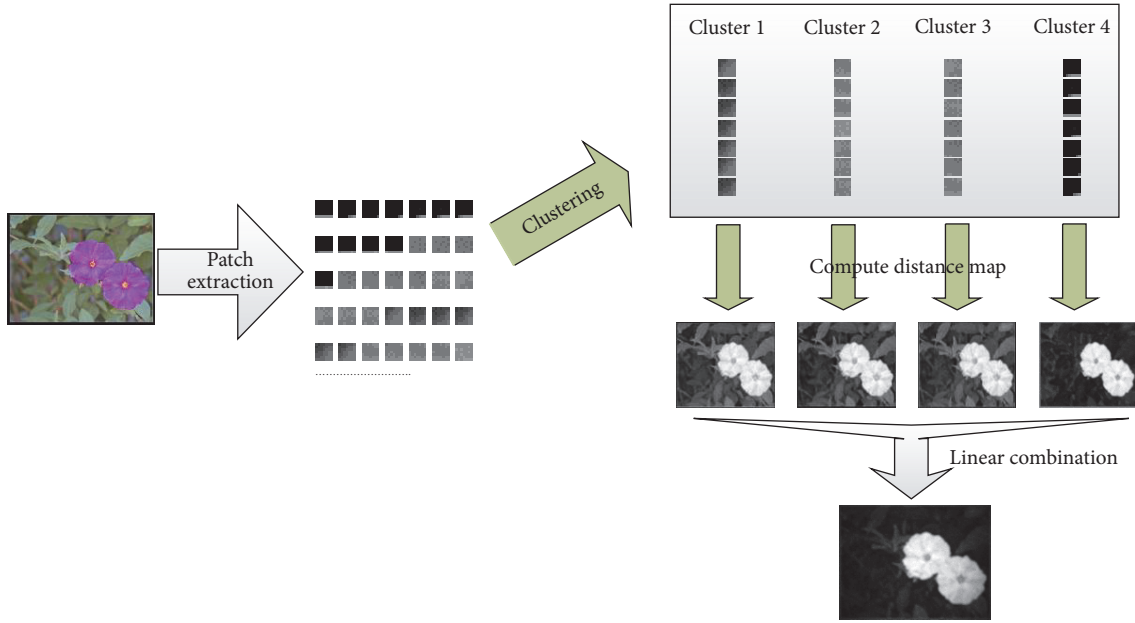
FIGURE 2: Coarse saliency map generation process.

consider a location to be salient if patch feature corresponding to that location is distant from background distribution. For the distance measurement, we adopt Euclidean distance as the distance metric.

As we get four background feature clusters, we then calculate a distance map for each cluster. In each distance map, value of every location is defined as the Euclidean distance between the patch feature of that location and the centroid value of this feature cluster. Distance map with respect to each cluster is computed by

$$D^{(l)}(i) = \left\| f(i) - C(l) \right\|_2, \quad (1)$$

where i denote patch location in image, $l \in \{1, 2, 3, 4\}$.

In this way, four distance maps are derived. And the four maps are normalized to $[0, 1]$. In the following, we create a coarse saliency map through combination of the four distance maps. The combination procedure is defined as follows:

$$S_{\text{coarse}} = \sum_{l=1}^{4} w_l \times D^{(l)}, \quad (2)$$

where "$\times$" denotes element-wise multiplication and w_l denotes weight for each distance map. w_l is defined as follows:

$$w_l = \frac{N(l)}{\sum_{i=1}^{4} N(i)}, \quad (3)$$

where $w_l \in [0, 1]$. Obviously, the weight can be seen as the ratio of sample number in each cluster to whole sample number. In this way, coarse saliency map is more similar to distance maps which are produced from larger clusters. Larger clusters have a stronger representative power for representing the real background. So distance maps produced from these larger clusters are more different from background

and meanwhile are more likely to highlight salient regions. The whole process of producing coarse saliency map is shown in Figure 2 (patches are zoomed in for observation purpose).

2.2. Edge-Preserving Element-Level Filtering. In the previous section, a coarse saliency map is generated. We aim to further discriminate salient object from background and highlight salient object uniformly in this section. Before starting the substantial operation, we decompose image into homogeneous basic elements and define saliency on element level hereafter. This is based on the observation that an image can be decomposed into basic, structurally representative elements that abstract away unnecessary detail and meanwhile allow for a clear and intuitive definition of saliency. Each element should locally abstract the image into perceptual homogenous regions. For detailed implementation, we use a linear spectral clustering method [23] to abstract the image into homogenous elements.

In this section, we adopt a method called edge-preserving element-level filtering to enhance salient region. This step is motivated by the observation that (1) salient object always tends to have sharp edges to background and (2) a region is likely to be salient if it is surrounded by salient regions. The operation of edge-preserving element-level filtering deals with basic elements described above. Its mathematical formula is described in (4) which we will explain in next paragraph. This operation possesses several properties, such as preserving sharp edges, refining saliency value of an element by combining saliency values of its surrounding elements.

After decomposing the image into homogenous elements, coarse saliency value of each element (superpixel) is calculated by averaging the coarse saliency value S_{coarse} of all its pixels inside. Then, for jth superpixel, if its coarse
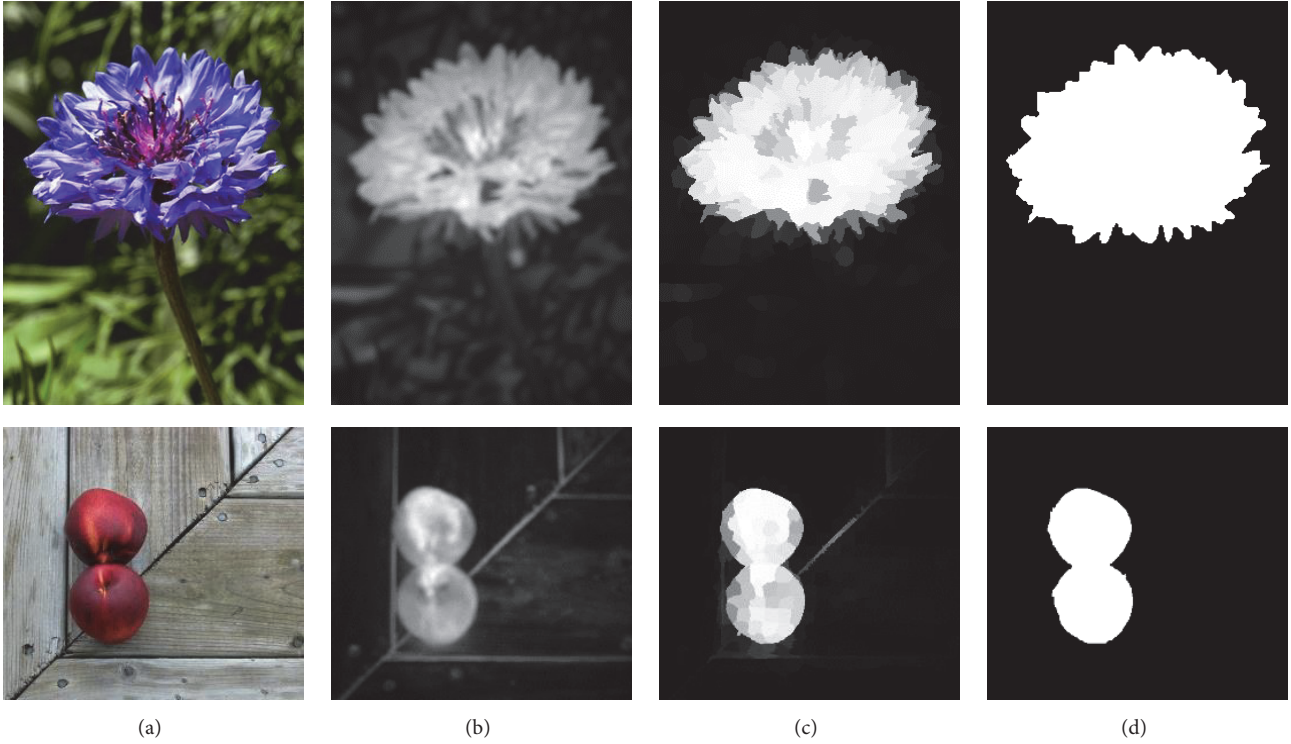
| (a) | (b) | (c) | (d) |

FIGURE 3: Edge-preserving element-level filtering. (a) Input image. (b) Coarse saliency map. (c) Saliency map after edge-preserving element-level filtering. (d) Ground truth.

saliency value is labeled as $S^c(j)$, the saliency value of the qth superpixel is measured by

$$S^f(q) = d(q) \cdot S^c(q) + [1 - d(q)] \cdot \sum_{j \in \Theta} w_{qj} S^c(j), \quad (4)$$

where $j \in \Theta$ are a set of superpixels for which qth superpixel is the neighbor within a certain distance, w_{qj} is a weight based on Euclidean distance, and $d(q) \in \{0, 1\}$ indicate whether qth superpixel is touching sharp edges in image. Formula (4) can be interpreted as follows: if qth superpixel is touching sharp edges in image, then saliency value of qth superpixel remains unchanged as its coarse saliency value, but if qth superpixel is not directly connected to sharp edges, then saliency value of qth superpixel is determined by saliency values of its surrounding superpixels.

Given superpixel positions $P(q)$ and $P(j)$, weight w_{qj} in (4) is defined as Euclidean distance between positions of qth and jth superpixel:

$$w_{qj} = \|P(q) - P(j)\|_2. \quad (5)$$

The indicator $d(q)$ in (4) is determined according to the following process: all superpixels contiguous to qth superpixel are searched first, then the largest variance among saliency values of these superpixels is computed, and finally if the variance is larger than a certain threshold, the qth superpixel is regarded as touching sharp edges, thus setting $d(q) = 1$, otherwise setting $d(q) = 0$.

Results of edge-preserving element-level filtering are illustrated in Figure 3. It is obvious that coarse saliency

maps are strongly enhanced after this procedure. Sharp edges of salient object have been well preserved and background noises have been largely removed. Work in [14] adopted an operation called "context-based error propagation," which is similar to edge-preserving element-level filtering in this section. It smooths region saliency by taking all other region elements in the same feature cluster into consideration, while our method refines region saliency using its spatially adjacent regions. Obviously, our method is more intuitive and simple and, above all, preserves sharp edges well.

2.3. Geodesic Distance Refinement. The final step of our proposed approach is refinement with geodesic distance [24]. The motivation underlying this step is based on the consideration that determining saliency of an element as weighted sum of saliency of its surrounding elements, where weights are correspond to Euclidean distance, has a limited performance in uniformly highlighting salient object. We tend to find some other solutions that could enhance regions of salient object more uniformly. From recent works [25, 26], we found that the weights may be sensitive to geodesic distance. Thus, saliency value based on geodesic distance is defined as follows:

$$S(p) = \sum_{i=1}^{N} \delta_{pi} S^f(i), \quad (6)$$

where N is the total number of superpixels and δ_{pi} is the weight based on geodesic distance between pth and ith superpixels.
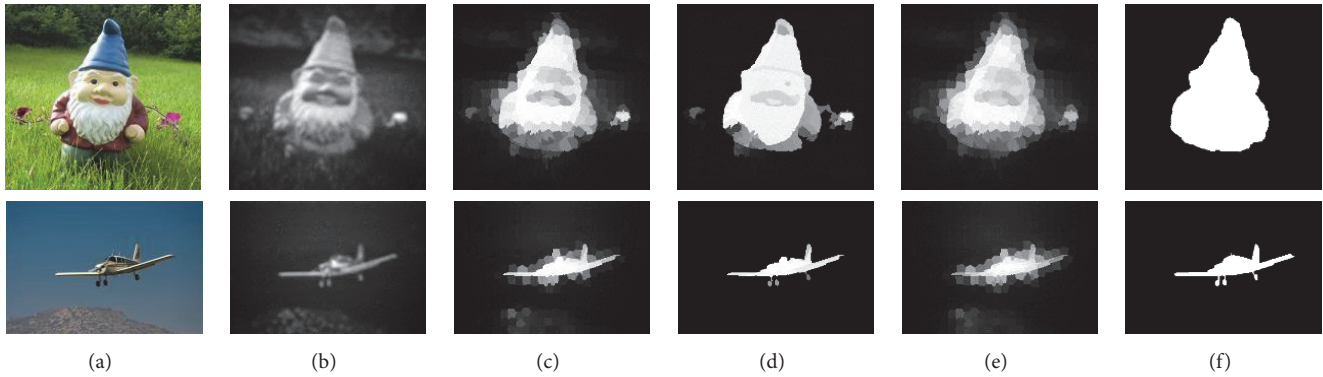
| (a) | (b) | (c) | (d) | (e) | (f) |

FIGURE 4: Geodesic distance refinement and comparison with Euclidean distance refinement. (a) Input image. (b) Coarse saliency map. (c) Saliency map after edge-preserving element-level filtering. (d) Saliency map using geodesic distance refinement. (e) Saliency map using Euclidean distance refinement. (f) Ground truth.

The weight values are produced similar to work in [25]. First, an undirected weight graph has been constructed by connecting all adjacent superpixels (a_k, a_{k+1}) and assigning their weight $d_c(a_k, a_{k+1})$ as the Euclidean distance between their saliency values which are derived in the previous Section 2.2. Then, the geodesic distance between two superpixels $d_g(p, i)$ can be defined as accumulated edge weights along their shortest path on the graph:

$$d_g(p, i) = \min_{a_1 = p, a_2, \ldots, a_n = i} \sum_{k=1}^{n-1} d_c(a_k, a_{k+1}). \quad (7)$$

In this way, we can get geodesic distance between any two superpixels in the image. Then, the weight δ_{pi} is defined as

$$\delta_{pi} = \exp\left\{ -\frac{d_g^2(p, i)}{2\sigma_c^2} \right\}, \quad (8)$$

where σ_c is the deviation for all d_c values. From formula (8), we can easily conclude that when p and i are in a flat region, saliency value of i would have a higher contribution to saliency value of p, and when p and i are in different regions between which a steep slope existed, saliency value of i tends to have a less contribution to saliency value of p.

In order to demonstrate the effectiveness of refinement using geodesic distance, we compare geodesic distance and Euclidean distance when they are used as combination weights. Figure 4 shows the experiment results. Refinement based on geodesic distance performs much better than refinement based on Euclidean distance, in that saliency objects are distinctively and uniformly highlighted.

3. Experiment

We provide an extensive comparison of our approach to several state-of-the-art methods on two most common-used datasets. The two benchmark datasets include ASD [7] and MSRA10K [8]. The ASD dataset contains 1000 images, while MSRA10K contains 10000 images, both with pixel-level ground truth.

3.1. Qualitative Evaluation. Figure 5 presents a qualitative comparison of our approach with other six outstanding methods. The other methods include Itti [27], GBVS [28], FT [7], CAS [19], PCA [13], wCtr [25], and DRFI [11]. It can be seen that results of Itti and GBVS only successfully detect some fuzzy locations of salient object, which makes it far from real application. CAS highlights more on object edges. While PCA performs overall better that these methods just mentioned before, there still exist failing occasions when salient objects are overwhelming in size. FT has a simple and efficient implementation, but problems occur when global cues play a role in saliency detection as FT focuses on local cues. Method of wCtr has a discriminative power to highlight salient objects but would go unstable when some nonsalient regions are also clustered as salient objects and are hardly connected to image boundary. Different from these methods which are all bottom-up based approaches, DRFI incorporates high level priors and supervised learning, which make it outstanding among other approaches. Our method is based on bottom-up processing; it is easy for us to conclude from the experiment results that our method provides more favorable results compared with others and even performs better than supervised approach like DRFI. An important tip is that our method is implemented on the original resolution of the input image without any rescale preprocessing.

3.2. Metric Evaluation. Similar to [7, 8], we deploy precision and recall as the evaluation metric. Precision corresponds to the percentage of salient pixels correctly assigned, while recall corresponds to the fraction of assigned salient pixels in relation to the ground truth number of salient pixels. Precision-recall (PR) curve can be produced by comparing the results with the ground truth through varying thresholds within the range $[0, 255]$. PR curve is an important metric in performance evaluation. PR curve with respect to each dataset is given in Figure 6.

In addition to precision and recall, we compute the *F*-measure, which is defined as follows:

$$F_\beta = \frac{(1 + \beta^2) \cdot \text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}}, \quad (9)$$

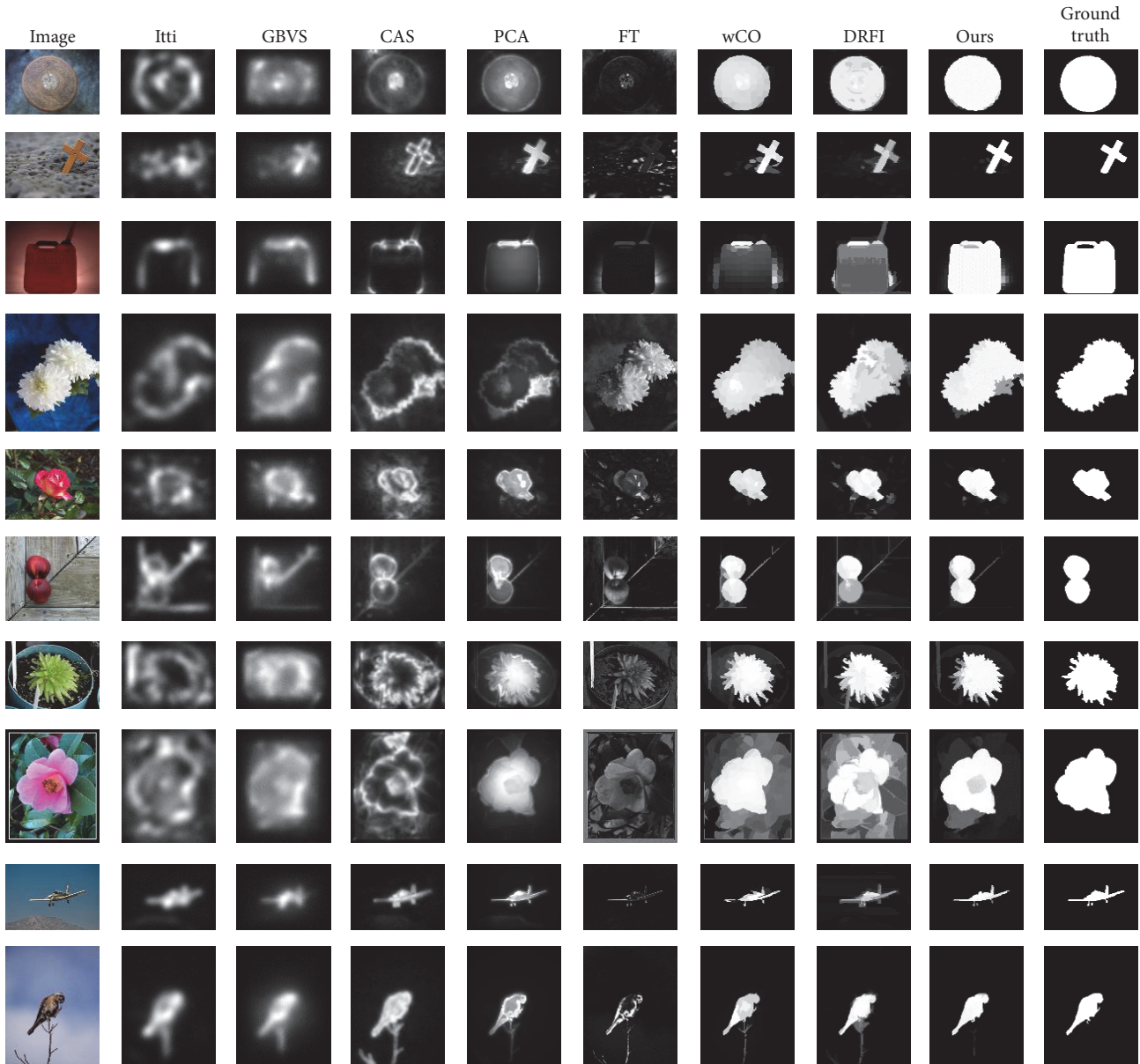| Image	Itti	GBVS	CAS	PCA	FT	wCO	DRFI	Ours	Ground truth

FIGURE 5: Visual comparison of state-of-the-art approaches to our method and ground truth.

where we set $\beta^2 = 0.3$. In Figure 7, we show the precision, recall, and F-measure values for adaptive threshold, which is defined as twice the mean saliency of the image. Figures 6 and 7 show that the proposed method achieves best among all these methods on the two datasets.

3.3. Performance. In order to evaluate algorithms thoroughly, we also compare the average running time of our approach to other methods on the benchmark images. Timing has been taken on a 2.6 GHz CPU with 8 GB RAM. The environment is Matlab 2016a installed on Windows. Running times of algorithms are listed in Table 1.

CAS method is slower because of its exhaustive nearest-neighbor search. Our running time is similar to that of PCA method. Both methods involve location-wise patch feature extraction and comparing. Our method spends most of the

running time on the first step of feature clustering (21%) and subsequent coarse saliency map computation (25%). Time of our method is moderate among these methods. Considering that our method is implemented on the original resolution of input image, computational cost of our method is still favorable against other methods.

4. Conclusion

In this paper, an effective bottom-up salient object detection method is presented. In order to utilize the statistical and structural information of backgroundness priors, background distribution based on feature clustering has been constructed. That distribution allows for computing a coarse saliency map. Then two refinement steps, including edge-preserving element-level filtering and upsampling based on

FIGURE 6: Precision-recall (PR) curves for all algorithms. (a) PR curves on MSRA10K dataset. (b) PR curves on ASD dataset.

FIGURE 7: Precision, recall, and *F*-measure for adaptive thresholds. (a) Measured on MSRA10K dataset. (b) Measured on ASD dataset.

TABLE 1: Comparison of running times.

Method	Ours	GBVS	CAS	Itti	FT	PCA	DRFI	wCtr
Time (s)	2.65	0.70	7.68	0.895	0.44	2.68	3.238	0.723
Code	Matlab	Matlab, C++	Matlab	Matlab	Matlab	Matlab	Matlab	Matlab

geodesic distance, have been applied to obtain the final saliency map. In the final map, salient objects are uniformly highlighted and background noises are removed thoroughly. The experimental evaluation exhibits the effectiveness of the proposed approach compared with the state-of-the-art methods.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

References

[1] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu, "Sketch2Photo: internet image montage," *ACM Transactions on Graphics*, vol. 28, no. 5, pp. 124–124, 2009.

[2] S.-M. Hu, T. Chen, K. Xu, M.-M. Cheng, and R. R. Martin, "Internet visual media processing: a survey with graphics and vision applications," *Visual Computer*, vol. 29, no. 5, pp. 393–405, 2013.

[3] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000 still image coding system: an overview," *IEEE Transactions on Consumer Electronics*, vol. 46, no. 4, pp. 1103–1127, 2002.

[4] G.-X. Zhang, M.-M. Cheng, S.-M. Hu, and R. R. Martin, "A shape-preserving approach to image resizing," *Computer Graphics Forum*, vol. 28, no. 7, pp. 1897–1906, 2009.

[5] M.-M. Cheng, F.-L. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Repfinder: finding approximately repeated scene elements for image editing," *ACM Transactions on Graphics*, vol. 29, no. 4, pp. 83:1–83:8, 2010.

[6] Z. Ren, S. Gao, L.-T. Chia, and I. W.-H. Tsang, "Region-based saliency detection and its application in object recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 5, pp. 769–779, 2014.

[7] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 1597–1604, Miami, Fla, USA, June 2009.

[8] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015.

[9] Z. Wang and B. Li, "A two-stage approach to saliency detection in images," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 965–968, Las Vegas, Nev, USA, April 2008.

[10] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Proceedings of the 12th International Conference on Computer Vision (ICCV '09)*, pp. 2106–2113, Kyoto, Japan, October 2009.

[11] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: a discriminative regional feature integration approach," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 2083–2090, Portland, Ore, USA, June 2013.

[12] F. Perazzi, P. Krahenhuhl, Y. Pritch, and A. Hornung, "Saliency filters: contrast based filtering for salient region detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 733–740, June 2012.

[13] R. Margolin, A. Tal, and L. Zelnik-Manor, "What makes a patch distinct?" in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 1139–1146, Portland, Ore, USA, June 2013.

[14] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV '13)*, pp. 2976–2983, Sydney, Australia, December 2013.

[15] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 3166–3173, Columbus, Ohio, USA, June 2013.

[16] T. Zhao, L. Li, X. H. Ding, Y. Huang, and D. L. Zeng, "Saliency detection with spaces of background-based distribution," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 683–687, 2016.

[17] J. Han, D. Zhang, X. Hu, L. Guo, J. Ren, and F. Wu, "Background prior based salient object detection via deep reconstruction residual," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 8, pp. 1309–1321, 2015.

[18] Y. Qin, H. C. Lu, Y. Q. Xu, and H. Wang, "Saliency detection via Cellular Automata," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR' 15)*, pp. 110–119, Boston, Mass, USA, June 2015.

[19] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1915–1926, 2012.

[20] Y. Xie and H. Lu, "Visual saliency detection based on Bayesian model," in *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP '11)*, pp. 645–648, Brussels, Belgium, September 2011.

[21] S. Li, H. C. Lu, Z. Lin, X. Shen, and B. Price, "Adaptive metric learning for saliency detection," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3321–3331, 2015.

[22] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithms: analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 881–892, 2002.

[23] Z. Li and J. Chen, "Superpixel segmentation using Linear Spectral Clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR' 15)*, pp. 1356–1363, Boston, Mass, USA, June 2015.

[24] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[25] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 2814–2821, IEEE, Columbus, Ohio, USA, June 2014.

[26] K. Fu, C. Gong, I. Y. H. Gu, and J. Yang, "Geodesic saliency propagation for image salient region detection," in *Proceedings of the 20th IEEE International Conference on Image Processing (ICIP '13)*, pp. 3278–3282, Melbourne, Australia, September 2013.

[27] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[28] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proceedings of the 20th Annual Conference on Neural Information Processing Systems (NIPS '06)*, pp. 545–552, Vancouver, Canada, December 2006.