Saliency Detection with Sparse Prototypes: An Approach Based on Multi-Dictionary Sparse Encoding

Jun Wang*, Zemin Wu, Chang Tian and Lei Hu

College of Communications Engineering, Army Engineering University of PLA, Nanjing, China.

Abstract. This paper proposes a bottom-up saliency detection algorithm based on multi-dictionary sparse recovery. Firstly, the SLIC algorithm is used to segment the image into superpixels in multilevel and atoms with a high background possibility are selected from the boundary superpixels to construct the multi-dictionary. Secondly, sparse recovery of the entire image is achieved using multi-dictionary to get sub-saliency maps from the perspective of sparse recovery errors. The final saliency map is generated in a weighted fusion manner. Experimental results on three public datasets demonstrate the effectiveness of our model.

1 Introduction

Visual saliency detection aims to identify the most distinctive regions in an image that draw human attention. This topic has received increasing interest in recent years because of its wide range of applications such as image segmentation, image retrieval, target recognition, image classification, and video compression. Although much significant progress has been made, salient object detection remains a challenging problem.

Sparse representation has been widely used in computer vision task in recent years. The typical sparse representation based saliency detection models have the following two types. The first type is the centre-surround model. Li et al. [1] and Han et al. [2] reconstruct a central image patches using its nearby local regions (i.e., using the nearby local regions as dictionary to recovery the centre patch) and then defined its saliency degree using sparse reconstruction error. However, a given patch is usually very similar to those nearby. As a result, the internal region of a large object cannot be detected well. The second method is based on the hypothesis that image boundaries belong to the background in most cases. Li et al. [3] and Jia et al. [4] propose to construct a background dictionary using superpixels at image boundary and perform sparse recovery to get the saliency map, as shown in Figure 2(a). Recovery error is used to determine saliency degree. In general, their performance is better than that of centre-surround model. However, they are imprecise for using the entire boundary superpixels as dictionary atoms if the object appears on the boundary.

Different from pervious works, in this paper, we propose a bottom-up saliency detection algorithm with sparse prototypes. First, the SLIC algorithm [5] is used to segment the image into superpixels in multi-level and reliable atoms is selected from boundary super-pixels to construct multi-dictionary. Second, sparse encoding of the entire image is achieved based on these dictionaries. The initial sub-saliency maps are generated according to the sparse recovery errors. The final saliency map is obtained through sub-saliency maps’ fusion. Our algorithm flow is shown in Fig. 1 and we name it ‘MDSR’ in the following.

2 Dictionary construction

To better capture intrinsic structural information and improve computational efficiency, an input image $I$ is segmented into $N$ super-pixels $\{p_i\}_{i=1}^N$ by the simple linear iterative clustering (SLIC) algorithm. The mean colour features $c_i$ in CIE-Lab colour space and mean...
normalized coordinates \( x_i \) of pixels are used to describe each superpixel \( p_i \).

\[
p_i = [c_i; x_i]
\]

Therefore, the entire image can be represented as:

\[
I = [p_1, p_2, \ldots, p_N]
\]

It has been shown that image boundaries are good visual cues for background models which can be exploited for saliency detection [3, 4]. On the other hand, salient object are likely to appear at the centre of a scene. But these assumptions may not always hold. So instead of directly extracted all the image boundary superpixels to construct dictionary, we firstly compute the background possibility of each boundary superpixel:

\[
P_{bg}(p_i) = 1 - \exp \left(-\frac{BC^2(p_i)}{2\delta^2}\right)
\]

\( BC(p_i) \) stands for the boundary connectivity of superpixel \( p_i \), which was proposed by Zhu et al. [6] to estimate the probability that an image patch belongs to background. It is defined by:

\[
BC(p_i) = \frac{L(p_i)}{\sqrt{A(p_i)}}
\]

\[
L(p_i) = \sum_{j=1}^{N} \exp \left(-\frac{d^2_{geo}(p_i, p_j)}{2\delta^2}\right) \cdot \delta(p_j \in I^{hl})
\]

\[
A(p_i) = \sum_{j=1}^{N} \exp \left(-\frac{d^2_{geo}(p_i, p_j)}{2\delta^2}\right)
\]

Where \( L(p_i) \) is the length along the image boundary \( I^{hl} \). \( \delta() \) is 1 for superpixels on the image boundary and 0 otherwise. \( A(p_i) \) represents the spanning area of \( p_i \), and \( d^2_{geo}(p_i, p_j) \) indicates the geodesic distance between the superpixel \( p_i \) and \( p_j \) in the CIE-Lab colour space.

Then we select the boundary superpixels with its background possibility value bigger than \( \xi \) as the atoms. \( \xi \) is defined as bellow:

\[
\xi = \frac{1}{2K} \sum_{j=1}^{K} P_{bg}(p_j)
\]

\( K \) denotes the number of all boundary superpixels. The dictionary \( D \) is constructed based on the selected atoms, as shown in Fig. 2(b).

\[ \text{Fig. 2. Comparison of dictionary construction: (a) traditional methods which directly use the entire boundary superpixels as the atoms of dictionary. (b) Our method that select the superpixels with high background possibility as the atoms (the superpixels marked with small white square).} \]
two objects. Most objects in SED2 are located at the boundary.

For each image, we obtained 10 saliency maps using the proposed model and other 9 state-of-the-art methods: GB [10], SR [11], PCA [12], DSR [3], LR [13], RPCA [14], BFS [15], RBD [6], and BSCA [16]. We use AUC value and MAE value to evaluate all algorithms. For each method, a binary map is obtained by segmenting each saliency map with a given threshold \( T \in [0, 255] \) and then compared with the ground truth mask \( G \) to compute the true positive and false positive rates to plot the ROC curve and compute the area under ROC curve to get the AUC value. The MAE metric calculates the average absolute error between the saliency map and ground truth mask of test image \((W \times H \text{ resolution})\) in pixel:

\[
MAE = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} |S(i, j) - G(i, j)|
\]

When in experiment, we segment every test image at three scale (i.e., we set \( M = 3 \)) for accuracy and efficiency. The segment parameters are \([100 20]\) and \([300 90]\). The parameter \( \delta \) in equation (3) is set to 1. The sparse reconstruction parameter \( \lambda \) in equation (8) is set to 0.01.

### 4.2 Experiment results

As shown in Table 1, our proposed algorithm outperforms other 9 methods lower MAE and higher AUC scores, which indicates the predicted saliency map is more similar to ground truth. Note that, the advantage of our algorithm is more obvious than others in SED2 datasets, demonstrating the effectiveness of MDSR in addressing the boundary saliency problems. A visual comparison of several saliency maps is give in Fig. 3. For the first example, the image background is complicated and there is an interference object at the boundary. For the other test images, the salient object all contacts the image boundary. As can be seen clearly, compared with the saliency maps generated by other algorithms, our results can highlight the salient object more clearly and completely under different scenes, which demonstrates the robustness of our model.

### Table 1: Comparison of AUC and MAE value on three public datasets. (The best performance are marked with red color)

<table>
<thead>
<tr>
<th></th>
<th>ASD</th>
<th>MSRA5000</th>
<th>SED2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC</td>
<td>MAE</td>
<td>AUC</td>
</tr>
<tr>
<td>GB</td>
<td>.8858</td>
<td>.2289</td>
<td>.8711</td>
</tr>
<tr>
<td>SR</td>
<td>.7651</td>
<td>.2299</td>
<td>.7553</td>
</tr>
<tr>
<td>RPCA</td>
<td>.8994</td>
<td>.2090</td>
<td>.8724</td>
</tr>
<tr>
<td>PCA</td>
<td>.9702</td>
<td>.1552</td>
<td>.9481</td>
</tr>
<tr>
<td>LR</td>
<td>.9584</td>
<td>.1885</td>
<td>.9238</td>
</tr>
<tr>
<td>BFS</td>
<td>.9702</td>
<td>.1285</td>
<td>.9481</td>
</tr>
<tr>
<td>DSR</td>
<td>.9792</td>
<td>.0809</td>
<td>.9530</td>
</tr>
<tr>
<td>BSCA</td>
<td>.9783</td>
<td>.0855</td>
<td>.9533</td>
</tr>
<tr>
<td>RBD</td>
<td>.9702</td>
<td>.0659</td>
<td>.9481</td>
</tr>
<tr>
<td><strong>Ours</strong></td>
<td>.9810</td>
<td>.0558</td>
<td>.9527</td>
</tr>
</tbody>
</table>

Fig. 3. Visual comparison between our results and others. (a) Input. (b) SR. (c) GB. (d) RPCA. (e) PCA. (f) LR. (g) DSR. (h) BFS. (i) RBD. (j) BSCA. (k) Ours. (l) Ground truth mask.

### 5. Conclusion

In this paper, we propose a bottom-up saliency detection algorithm via multi-dictionary sparse encoding. To handle the misjudgment problem when salient objects appear at the image boundary, we segment the image in multilevel. In each segment level, we compute the background possibility of superpixels located at image boundaries and then select the reliable ones as the atoms of dictionary. Based on the multi-dictionary, sparse encoding of the whole image is achieved and the saliency map \( s \) are generated according to the sparse recovery errors. Experimental results demonstrate the effectiveness of our algorithm.

It is necessary to explain that our method also has good expansibility. If we have a strict requirement for the accuracy of the result in practice, we can set a larger \( M \). However, if we emphasis on the speediness, we can set a smaller \( M \). In the further research, we will focus on the relationship between the speediness and accuracy of our algorithm.

### References


