Salient object detection employing robust sparse representation and local consistency

Yi Liu[a,b], Qiang Zhang[a,b,*], Jungong Han[c], Long Wang[d]

[a] *Key Laboratory of Electronic Equipment Structure Design, Ministry of Education, Xidian University, Xi'an Shaanxi 710071,China*
[b] *Center for Complex Systems, School of Mechano-Electronic Engineering, Xidian University, Xi'an Shaanxi 710071,China*
[c] *Department of Computer Sciences and Digital Technologies, Northumbria University, Newcastle upon Tyne NE1 8ST, U.K*
[d] *Center for Systems and Control, College of Engineering, Peking University, Beijing 100871, China*

Abstract

Many sparse representation (SR) based salient object detection methods have been presented in the past few years. Given a background dictionary, these methods usually detect the saliency by measuring the reconstruction errors, leading to the failure in the complex scene. In this paper, we propose to replace the traditional SR model with a *robust* sparse representation (RSR) model, for salient object detection, which replaces the least squared errors by the sparse errors. Such a change dramatically improves the robustness of the saliency detection in the existence of non-Gaussian noise, which is the case in most practical applications. By virtual of RSR, salient objects can equivalently be viewed as the sparse but strong "outliers" within an image so that the salient object detection problem is reformulated to a sparsity pursuit one. Moreover, we jointly utilize the representation coefficients and the reconstruction errors to construct the saliency measure in the proposed method. Finally, we integrate a local consistency prior among spatially adjacent regions into the RSR model in order to uniformly highlight the whole salient object. Experimental results demonstrate that the proposed method significantly outperforms the traditional SR

*Corresponding author. Address: P.O.Box 183, Department of Automatic Control, Xidian University, No.2 South TaiBai Road, Xi'an, Shaanxi Province, 710071, China. Tel: +86 029 88604397. Email address : qzhang@xidian.edu.cn (Q. Zhang)

based methods and is competitive with some current state-of-the-art methods, especially for those images with complex structures.

Keywords: Salient object detection, robust sparse representation, local consistency, complex structures

1. Introduction

Visual saliency refers to identifying certain regions of a scene, which stand out from their surroundings and catch immediate attention [1]. As an important branch of visual saliency, salient object detection has attracted a wide range of attention. Generally, it is essentially a binary segmentation problem [2] starting by detecting the attractive objects in a scene followed by a segmentation procedure that extracts the entire objects from the background. It has been widely applied to many fields, such as image segmentation [3], classification [4], cluster [5], recognition [6], content-based image retrieval [7] and image fusion [8].

Recently, sparse representation (SR) has been exploited to salient object detection [9, 10, 11, 12] as a result of its successful applications in many computer vision and image processing tasks, such as face recognition [13], image classification [14], and so on. In these SR based methods, the salient object detection is normally carried out in three steps. First, input images are divided into many patches or super-pixels. Secondly, an over-complete dictionary is constructed, which helps to encode the feature vectors collected from those patches or super-pixels. Thirdly, the saliency value for each patch or super-pixel is measured according to its representation coefficients or residual errors.

For the SR based salient object detection methods, there are two important issues: dictionary construction and saliency measure. Earlier methods are prone to adopt the surrounding patches of each test patch as the dictionary [9, 10]. Due to the fact that the edges of salient objects have high contrast against their surrounding patches, such SR based methods usually assign higher salient values to the edges rather than the whole objects, as illustrated in Fig. 1(b). Recently, some boundary priors [15] are integrated into these methods based on

the assumption that backgrounds are usually distributed on the boundary of an image. Under this assumption, the patches or super-pixels near the boundary of an image are often selected to construct a background dictionary [12, 16, 17]. As shown in the first row of Fig. 1(c), these methods could overcome the shortcomings of those methods with the surrounding patches as the dictionary.

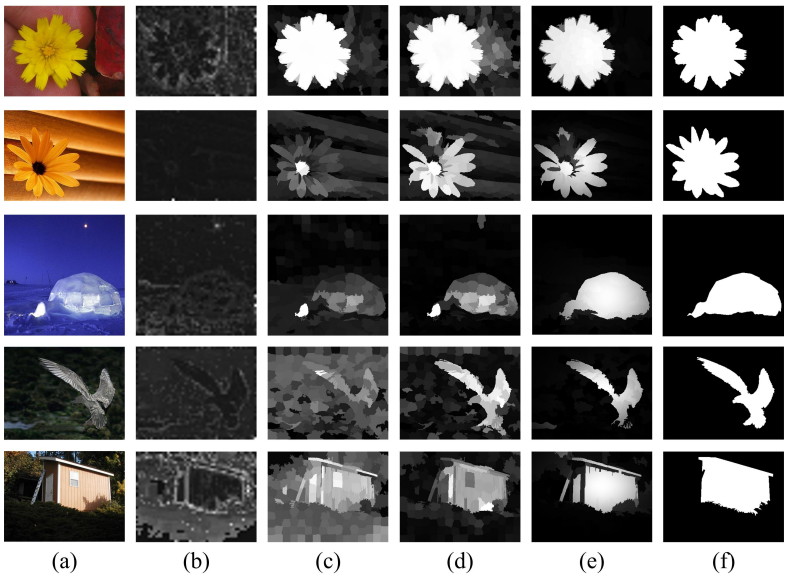


Figure 1: Typical challenging examples for SR based salient object detection methods. (a) Original images; (b) SR based method with surrounding patches as the dictionary [10]; (c) SR based method with background templates near the image boundary as the background dictionary; (d) Proposed RSR method with the background dictionary but without the local consistency prior; (e) Proposed RSR method with the background dictionary as well as the local consistency prior; (f) Ground truth.

With respect to the saliency measure, most SR based salient object detection methods employ either the sparseness (i.e., the coding length) of the representation coefficients or the reconstruction errors, especially the latter, to define the saliency measure [9, 10, 11], because there is an assumption that natural signals can be represented or approximately represented as a linear combination of a "few" atoms from a redundant dictionary [9, 10].

However, the traditional SR model employed in those salient object detec-

tion methods imposes a sparsity constraint on the representation coefficients to achieve the sparse coding of each test image patch or super-pixel. It basically minimizes the sum of squared reconstruction errors, therefore tending to be sensitive to the non-Gaussian noise as well as sparse "outliers" [13, 18]. Two undesirable results will be obtained when the residual errors are used as the saliency measure, especially for those methods based on the background dictionary. One is that many regions belonging to the foreground will not be highlighted when the foreground object and the background look similar, as shown in the second and third rows in Fig. 1(c). The other one is that the background will not be well suppressed, as shown in the last two rows in Fig. 1(c).

Moreover, there generally exist strong spatial correlations among the local regions in an image, i.e., the spatially adjacent patches or super-pixels with similar features should have similar saliency values. But in most of the existing salient object detection methods, this local consistency is often ignored, and the saliency of each image patch or super-pixel is computed independently. As a result, the whole salient object could not be uniformly highlighted, as shown in the third and last rows in Fig. 1(d). Besides, background can not be well suppressed, resulting parts of the background being falsely taken as the salient regions, as shown in the first and fourth rows in Fig. 1(d).

In this paper, we aim to detect the salient object in an image with complex structures by addressing the two problems mentioned above. More specifically, to enhance the algorithm robustness against the non-Gaussian noise, we replace the least squared reconstruction errors with the sparse reconstruction errors. In another word, we impose a $l_{2,1}$-norm minimization constraint on the reconstruction errors to ensure the column-sparsity of the error matrix. It can be interpreted as that the salient objects are sparsely distributed "outliers" within an image and seeking such "outliers" is equivalent to a sparsity pursuit problem, which can be solved by a robust sparse representation (RSR) model [18]. When applied to the detection of salient objects, RSR is expected to possess higher distinctiveness between the foreground objects and their backgrounds,

as shown in Fig. 1(d). Besides, based on the local consistency, the spatially adjacent pathes or super-pixels with similar features should have similar saliency values. Thus they should possess similar sparse representation coefficients as well as reconstruction errors when they are sparsely coded by using RSR with respect to the same background dictionary. We achieve that by introducing two Laplacian regularizations with respect to the representation coefficients and reconstruction errors, respectively, into the RSR model. As a result, the whole salient object can be uniformly highlighted and the background can also be well suppressed, a shown in Fig. 1(e). Eventually, an object function taking both the above mentioned aspects into account is minimized, thus helping to generate the saliency map.

In summary, our paper differs from the existing works in three aspects:

(1) We employ the RSR model, instead of the traditional SR model, in our proposed method. To our best knowledge, this is *the first attempt* to apply the RSR model to the detection of salient objects. By virtue of RSR, the salient object is modeled as sparse but strong "outliers" within an image so that the salient object detection can be accomplished by solving a sparsity pursuit problem.

(2) We involve a local consistency prior among spatially adjacent regions by imposing two Laplacian regularizations on the representation coefficients and reconstruction errors in our proposed method. This is different from the existing method in [17], in which only a Laplacian regularization term is imposed on the representation coefficients.

(3) Two saliency measures are defined based on the representation coefficients and the reconstruction errors, respectively, and the two saliency measures are fused to obtain the final saliency measure. Especially, in the representation coefficient based saliency measure, *the sparseness and magnitude information of the representation coefficients* are jointly employed.

The remainder of this paper is organized as follows. Section 2 briefly reviews the related work. Section 3 describes the proposed salient object detection method in detail. Experimental results and conclusions are given in Section 4

and Section 5, respectively.

2. Related work

2.1. Contrast-based salient object detection methods

During the last few years, numerous salient object detection methods have been proposed [9, 10, 11, 12, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28], among which the contrast-based methods are most popular [9, 10, 15, 16, 21, 22, 26, 27]. These contrast-based methods can be further divided into local-contrast based and global-contrast based ones, respectively.

Local-contrast based methods tend to highlight a certain region with high visual attention with respect to its small neighborhoods [17, 21, 22, 29]. The earlier local-contrast based methods are designed for only saliency detection [19], but in recent years, they are extended to salient object detection [21, 29]. The common observation is that these methods tend to produce higher saliency values near the edges instead of uniformly highlighting the whole salient objects.

As opposed to those local-contrast based ones, global-contrast based methods evaluate the saliency value of each pixel or region with respect to the entire image [22, 23]. In other words, these methods aim to capture the holistic rarity or uniqueness from an image. Compared with the local-contrast based methods, global-contrast based methods can obtain more uniform detection results and have attracted more attentions [20, 22, 23].

In view of the advantages of local-contrast based and global-contrast based methods, combining such two methods has been studied in recent years [24, 25, 30, 31]. For example, in [24], a salient object detection method was presented, which integrated a global saliency estimation via a high-dimensional color transform (HDCT) and a local saliency estimation via regression. The two resulting saliency maps complemented each other to obtain a final saliency map. Similarly, in [25], the authors presented a coding-based saliency measure by exploring both global and local cues for saliency computation.

2.2. SR based salient object detection methods

Recently, some salient object detection methods have been presented based on the sparse representation (SR) theory. In the earlier SR based methods, the predefined dictionary is simply selected from the surrounding patches of each given test patch. Such approaches generally produce higher saliency values at the boundaries of the object [9, 10]. In essence, these methods can be categorized as the local-contrast based ones.

Lately, several boundary priors [15] have been integrated into these SR based methods considering the fact that the background is usually distributed on the boundary of an image. Based on this assumption, the patches or super-pixels near the image boundary are often selected to construct a global background dictionary. Then, the saliency value for each patch or super-pixel could be measured according to reconstruction errors with respect to the predefined background dictionary. For example, in [16], saliency was measured via the reconstruction errors with respect to the background templates obtained from the image boundary. To accurately locate the salient object, the authors [17] constructed a heuristic background dictionary to increase the discriminating power and representation efficiency of the traditional SR model. The heuristic background dictionary was obtained from the image boundary where the foreground noises had been removed from the border regions. Due to the fact that all of the test image super-pixels are reconstructed from the same background dictionary, these methods can be seen as global-contrast based ones. In general, these methods could more uniformly highlight the whole salient object in most cases than those methods using the surrounding patches as the dictionary.

3. Proposed algorithm

In this section, we will first briefly introduce the robust sparse representation (RSR) model in [18] and then explain how we apply it to the detection of salient objects.

3.1. Robust sparse representation model

Let $X = [x_1, x_2, ..., x_N]$ be an observed data matrix of size $d \times N$, each column of which is a data vector $x_i \in R^d$. Given a dictionary $D \in R^{d \times M}$ with M prototype atoms, the RSR model is defined as follows [18]

$$\min_{Z,E} \|Z\|_1 + \lambda \|E\|_{2,1} \quad s.t.\ X = DZ + E \tag{1}$$

where $\|Z\|_1$ denotes the l_1-norm of the matrix Z and is defined as $\|Z\|_1 = \sum_{ij} |Z(i,j)|$. $\|E\|_{2,1}$ denotes the $l_{2,1}$-norm of the matrix E and is defined as $\|E\|_{2,1} = \sum_j \sqrt{\sum_i (E(i,j))^2}$. $Z(i,j)$ and $E(i,j)$ are the (i,j)-th entries of the matrices Z and E, respectively. The parameter $\lambda > 0$ is employed to balance the effects of the two components in Eq. (1).

Revealed in Eq. (1), the RSR model imposes the sparsity constraint on the representation coefficients matrix Z to sparsely code each image patch or super-pixel. And the $l_{2,1}$-norm is imposed on the reconstruction errors matrix E to ensure that it is sparse in column. In the traditional SR model, the conventional least-squared reconstruction error is employed, which tends to be sensitive to the non-Gaussian noise, such as sparse "outliers". Differently, in the RSR model, a so-called sparse reconstruction error is employed, which improves the robustness of the RSR model against the non-Gaussian noise or sparse but strong "outlier", and helps to select the discriminative patches or super-pixeles.

Actually, the $l_{2,1}$-norm has been widely utilized in different tasks, including feature selection [32] and subspaces clustering [5]. In our method, each column in the matrix E corresponds to the RSR reconstruction errors for each super-pixel. By using the $l_{2,1}$-norm minimization, E is ensured to be sparse in column, i.e., some columns in E will be forced to be zero ones. This indicates that the super-pixels corresponding to these columns can be well reconstructed by using the given background dictionary and are thus seen as background ones. In contrast, the super-pixels corresponding to those non-zero columns can not be well reconstructed by using the given background dictionary. In other words, these super-pixels are significantly different from those background super-pixels and are thus seen as foreground salient ones. Therefore, the $l_{2,1}$-norm minimization

on the error matrix E can be used to detect those foreground super-pixels that are distinct from the background ones.

3.2. RSR based salient object detection

Fig. 2 illustrates the diagram of the proposed method, mainly consisting of three parts: image over-segmentation and feature extraction, robust sparse coding with local consistency, saliency map generation and propagation. In the following contents, we will elaborate each part.

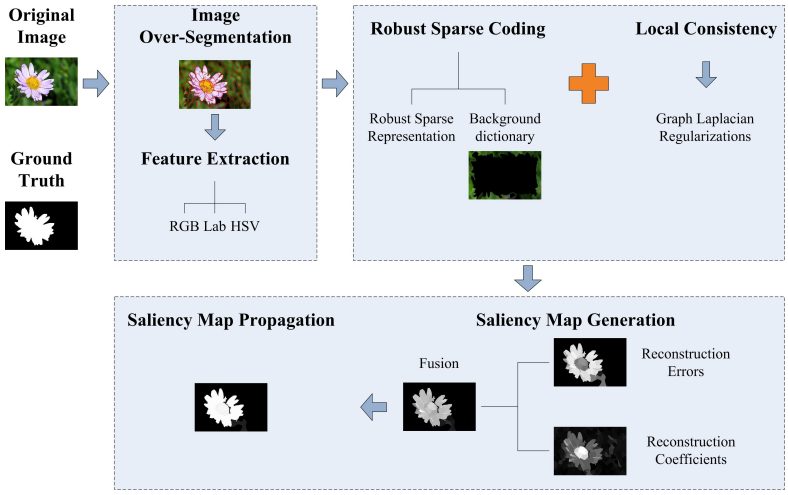


Figure 2: Diagram of the proposed salient object detection.

3.2.1. Image over-segmentation and feature extraction

In our proposed method, the input image I is first over-segmented into N super-pixels $S = [s_1, s_2, \ldots, s_N]$ by using the simple linear iterative clustering (SLIC) algorithm [33] due to its simplicity and efficiency. For each super-pixel s_i, a feature vector $x_i \in R^m$ of dimension $m = 9$ is constructed, which includes its red-, green-, and blue-components in RGB color space, its lightness- and two color-opponent-components in the CIELab color space, and its hue-, saturation-, and value-components in the HSV color spaces. On top of it, the feature vector for each super-pixel is generated via averaging all of the feature vectors of the

pixels contained in the current super-pixel. Finally, horizontally stacking the feature vectors of all superpixels produces a feature matrix $X \in R^{m\times N}$ for the input image, i.e., $X = [x_1, x_2, \ldots, x_N] \in R^{m\times N}$.

3.2.2. Background dictionary construction

Similar to other SR based salient object detection methods, the predefined dictionary plays an important role in the proposed RSR based method. Motivated by its successful applications in the salient object detection methods [12, 15, 16, 17, 27, 28], we also use the image boundary prior to construct a background dictionary $D = [d_1, d_2, \ldots, d_K] \in R^{m\times K}$ in the RSR model. In such a background dictionary D, $d_i \in R^m$ denotes the feature vector of a super-pixel near the image boundary, and K refers to the number of the dictionary atoms. Here, we simply extract all the super-pixels that directly connect the image border to construct the background dictionary.

3.2.3. Robust sparse coding with local consistency

Using the dictionary $D \in R^{m\times K}$ constructed in the previous sub-section, each super-pixel could be sparsely represented if directly applying the RSR model in Eq. (1). However, this ignores the strong correlations among the spatially local regions [17], and thus will be unavoidable for some isolated regions in the detected result. In principle, the spatially adjacent super-pixels with similar features should have similar saliency values [17]. Correspondingly, these super-pixels will have similar representation coefficients and reconstruction errors when sparsely coded by using RSR with respect to the same background dictionary. We achieve that by imposing two Laplacian regularizations on the representation coefficients and reconstruction errors. The resulting salient object detection model is formulated as follows.

$$\begin{aligned} \min_{Z,E} \|Z\|_1 + \lambda_1\|E\|_{2,1} + \lambda_2 tr(ZLZ^T) + \lambda_3 tr(ELE^T) \\ s.t. \quad X = DZ + E \end{aligned} \tag{2}$$

where $Z \in R^{K\times N}$ and $E \in R^{m\times N}$ are the representation coefficients matrix and reconstruction errors matrix corresponding to the input image X, respective-

ly. D is the predefined background dictionary. The Laplacian regularizations $tr(ZLZ^T)$ and $tr(ELE^T)$ are defined as

$$tr(ZLZ^T) = \frac{1}{2}\sum_{i,j}^{N} \|z_i - z_j\|_2^2 \, \omega_{ij} \tag{3}$$

$$tr(ELE^T) = \frac{1}{2}\sum_{i,j}^{N} \|e_i - e_j\|_2^2 \, \omega_{ij} \tag{4}$$

In Eq. (3) and Eq. (4), z_i and e_i denote the i-th columns of the matrices Z and E, respectively. The weight ω_{ij} implies the similarity between the i-th and j-th super-pixels and will be discussed later in detail. Based on these weights, an affinity matrix $W \in R^{N\times N}$ with its (i,j)-th entry $W_{i,j} = \omega_{ij}$ and a diagonal degree matrix $C \in R^{N\times N}$ with its i-th diagonal element $C_{i,i} = \sum_j W_{i,j}$ are constructed. The Laplacian matrix L is thus defined as $L = C - W$. λ_1, λ_2, and λ_3 are three positive trade-off parameters.

The weight ω_{ij} is computed by Eq. (5) in this paper.

$$\omega_{ij} = \begin{cases} \exp(-\frac{\|p_i - p_j\|_2^2}{2\sigma_p^2}) \cdot \exp(-\frac{\|x_i - x_j\|_2^2}{2\sigma_f^2}), & \text{if } s_i \text{ and } s_j \text{ are spatially adjacent,} \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

where $p_i, p_j \in R^2$ denote the center positions of the super-pixels s_i and s_j. $x_i, x_j \in R^m$ are their feature vectors, respectively. σ_p and σ_f are two scalars, and are experimentally set to $\sqrt{0.5}$ and 1, respectively.

In Eq. (5), the component, $\exp(-\frac{\|p_i - p_j\|_2^2}{2\sigma_p^2})$, denotes the spatial distance between the two super-pixels s_i and s_j, and the component, $\exp(-\frac{\|x_i - x_j\|_2^2}{2\sigma_f^2})$, indicates their feature similarity. Eq. (5) ensures that two super-pixels with smaller spatial distance and more similar features are assigned to higher weight values, thus preserving the local consistency among the spatially adjacent super-pixels with similar features.

Fig. 3 illustrates the validity of the two Laplacian regularization terms. Compared with those detection results obtained by the RSR model without considering the local consistency (i.e., Fig. 3(b)), the performance is improved to some extent by imposing one of the two Laplacian regularization terms (i.e.,

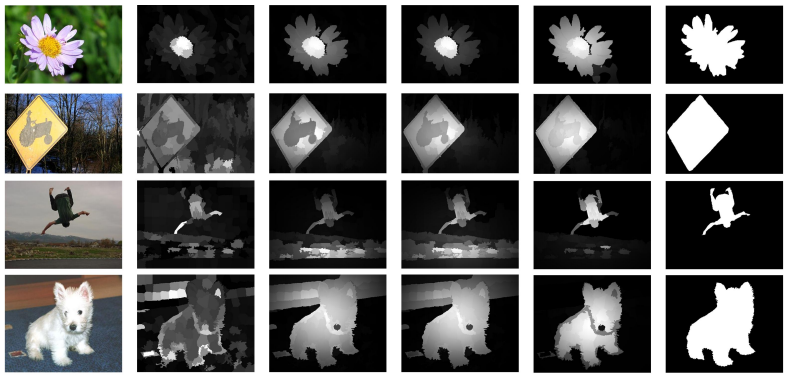


Figure 3: Illustrations of the validity of the two Laplacian regularization terms. (a) Original image; Saliency maps obtained by the RSR model (b) without considering the local consistency; (c) imposing the Laplacian regularization only on the representation coefficients; (d) imposing the Laplacian regularization only on the reconstruction errors; (e) imposing two Laplacian regularizations on the representation coefficients and reconstruction errors; (f) Ground truth.

Fig. 3(c) and (d)). As well, compared with the detection results obtained by imposing one Laplacian regularization term (i.e., Fig. 3(c) and (d)), the performance is further improved by using two Laplacian regularization terms. More specifically, the foreground salient object is more uniformly highlighted (as shown in the first two rows in Fig. 3), and the background noise is also better suppressed (as shown in the last two rows in Fig 3) by using two Laplacian regularization terms than using one of the two Laplacian regularization terms.

The optimization algorithm is convex and can be solved by various methods. In this paper, we jointly adopt the Alternating Direction Method of Multipliers (ADMM) [34] and a modified Sparse Reconstruction by Separable Approximation (SpaRSA)-based method [35] to solve the optimization problem in Eq. (2). This requires the minimization of the following augmented Lagrangian function:

$$\begin{aligned} L(Z,E,Y) = \|Z\|_1 + \lambda_1\|E\|_{2,1} + \lambda_2 tr(ZLZ^T) + \lambda_3 tr(ELE^T) \\ + \langle Y, X - DZ - E\rangle + \tfrac{\mu}{2}\|X - DZ - E\|_F^2 \end{aligned} \tag{6}$$

where Y is a Lagrangian multiplier, introduced to remove the equality con-

straint in Eq. (2), and μ is a penalty parameter. $< \cdot >$ denotes the Euclidean inner product of two matrices. Clearly, this problem becomes unconstrained, and can be minimized with respect to Z and E, respectively. Algorithm 1 summarizes the calculations of the optimization model. More details can be seen in Appendix A.

Algorithm 1 Solving the optimization problem in Eq. (6).

Input: Feature matrix X, parameters λ_1, λ_2, λ_3, and Laplacian matrix L.

Output: Z and E

1: **intialize:** $Z^{(0)} = \mathbf{0}$, $E^{(0)} = \mathbf{0}$, $Y^{(0)} = \mathbf{0}$, $\mu^{(0)} = 1$, $\mu_{\max} = 10^{10}$, $\rho = 1.1$, $t = 0$, $\varepsilon_1 = 10^{-3}$, and $\varepsilon_2 = 10^{-6}$.

2: **repeat**

3: Fix E and update Z using Eq. (A.4).

4: Fix Z and update E using Eq. (A.9).

5: Update the multiplier Y: $Y^{(t+1)} = Y^{(t)} + \mu^{(t)}(X - DZ^{(t)} - E^{(t)})$.

6: Update μ:

$$\mu^{(t+1)} = \begin{cases} \min(\rho\mu^{(t)}, \mu_{\max}), & \text{if } \mu^{(t)} \times \max\left(\left\|Z^{(t+1)} - Z^{(t)}\right\|_F, \left\|E^{(t+1)} - E^{(t)}\right\|_F\right) < \varepsilon_1 \\ \mu^{(t)}, & \text{otherwise.} \end{cases},$$

7: Update t: $t = t + 1$.

8: **until** Convergence: $\left\|X - DZ^{(t)} - E^{(t)}\right\|_F / \|X\|_F < \varepsilon_2$

3.2.4. Saliency Map Generation

In this part, we will explain how we compute the saliency measure for each super-pixel as well as how we compute the saliency value for each pixel.

A. Saliency measure for each super-pixel.

Given a background dictionary D, each column of the sparse errors matrix E obtained by solving the model in Eq. (6) may contain the salient information of each super-pixel that is distinct from the background. In addition, the representation coefficients may reflect the similarity between each test super-pixel and the backgrounds to some extent. Therefore, the proposed saliency measure for

each super-pixel consists of two sub-indexes in this paper. One is based on the reconstruction errors while the other is based on the representation coefficients.

Generally, a super-pixel will be more salient if it has larger reconstruction errors with respect to the background dictionary. Considering that, we define the reconstruction errors based saliency measure $Sal_E(s_i)$ for the i-th super-pixel s_i as

$$Sal_E(s_i) = 1 - \exp(-\frac{\|e_i^*\|_2^2}{2\sigma_E^2}) \tag{7}$$

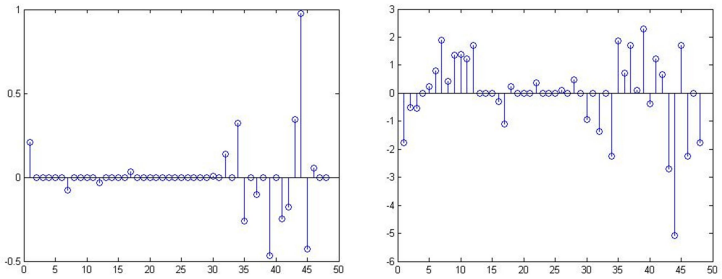
where e_i^* denotes the i-th column vector of the optimal errors matrix E^* obtained by solving Eq. (6) and corresponds to the reconstruction errors of the super-pixel s_i. $\|e_i^*\|_2$ represents the l_2-norm of the vector e_i^* and is defined as $\|e_i^*\|_2 = \sqrt{\sum_j (e_i^*(j))^2}$. σ_E is a scalar parameter and is experimentally set to 4 here.

In addition to the reconstruction errors, the saliency value of each super-pixel can also be determined by its representation coefficients to some extent. For example, as shown in Fig. 4(a), a background super-pixel will be sparsely coded by the predefined background dictionary. In contrast, as shown in Fig. 4(b), a foreground super-pixel will be densely coded by the same background dictionary. In other words, the saliency value of each super-pixel can be determined by the sparsity (also called coding length) of its representation coefficients. As well, the foreground object in an image usually has higher contrast than the background and thus has higher energy. Correspondingly, as shown in Fig. 4, the representation coefficients for a foreground super-pixel have higher magnitudes than those for a background super-pixel.

Based on the above two observations, we define the representation coefficients based saliency measure $Sal_Z(s_i)$ for the i-th super-pixel as

$$Sal_Z(s_i) = \|z_i^*\|_0 \cdot \left(1 - \exp\left(-\frac{\|z_i^*\|_2^2}{2\sigma_Z^2}\right)\right) \tag{8}$$

where z_i^* denotes the i-th column vector of the optimal representation matrix Z^* obtained by solving Eq. (6) and corresponds to the representation coefficients of the super-pixel s_i. $\|z_i^*\|_0$ represents the l_0-norm of the vector z_i^* and defined as the number of nonzero entries in the vector z_i^*. $\|z_i^*\|_0$ and $\|z_i^*\|_2$ indicate the



(a) (b)

Figure 4: Comparisons of representation coefficients of background and foreground super-pixels. (a) Representation coefficients for a background super-pixel; (b) Representation coefficients for a foreground super-pixel.

sparsity and the energy of the representation coefficients corresponding to the super-pixel s_i to some extent, respectively. Similarly, σ_Z is a scalar parameter and is experimentally set to 4 in this paper.

The final saliency measure $Sal(s_i)$ for the super-pixel s_i is defined by integrating the two saliency measures $Sal_E(s_i)$ and $Sal_Z(s_i)$ as

$$Sal(s_i) = Sal_E(s_i)^{\alpha} .* Sal_Z(s_i)^{1-\alpha} \tag{9}$$

where $\alpha \in (0, 1)$ is a scalar to control the contributions of the two terms and is set to 0.8 in the experiment par.

B. Saliency measure for each pixel.

According to the saliency measure defined by Eq. (9), an initial super-pixel level saliency map $M_{sp}(s_i)$, i.e., $M_{sp}(s_i) = Sal(s_i)$, is obtained. After that, a smooth saliency map $M'_{sp}(s_i)$ is obtained by performing the propagation method [16] on the map $M_{sp}(s_i)$. Finally, a pixel-level saliency map $M_{pixel}(p)$ is obtained by mapping the saliency map $M'_{sp}(s_i)$ to the full-resolution image, i.e.,

$$M_{pixel}(p) = M'_{sp}(s_i), \text{ if } p \in s_i \tag{10}$$

Fig. 5 illustrates the saliency detection results obtained by different phases. As shown in Fig. 5(d), the saliency measure defined by Eq. (9) could more

uniformly detect the whole salient object than those measures defined by Eq. (7) or Eq. (8) independently. After the saliency map propagation, the salient object is further uniformly highlighted. Meanwhile, the background is also well suppressed.

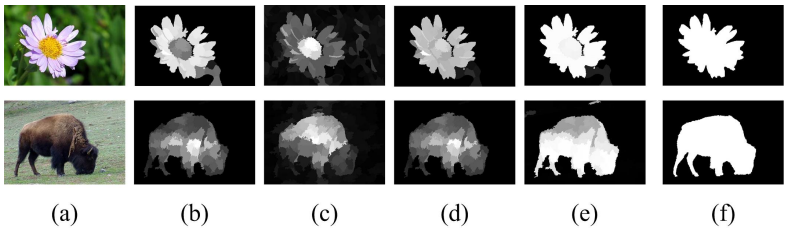


Figure 5: Saliency maps obtained by different saliency measures. (a) Original image; (b) Saliency maps obtained by using the reconstruction errors; (c) Saliency maps obtained by using the representation coefficients; (d) Saliency maps by fusing (b) and (c); (e) Saliency maps after propagation; (f) Ground truth.

3.3. Computational complexity

Suppose the data matrix X and dictionary D have sizes $d \times N$ and $d \times M$, respectively. Then the coefficients matrix Z has size $M \times N$. As discussed in [18], the computational complexity of Algorithm 1 is mainly dependent on the product of three matrices in Eq. (A.3) when the matrix Z is updated. Then the computational complexity of the proposed method is thus $O(rdM^2N)$, where r is the number of iterations needed for convergence. It demonstrates that the number of dictionary atoms M has a greater impact on the computational complexity of the proposed method than other parameters. In the proposed method, M is set to the number of boundary super-pixels (about 49), which is far smaller than the total number of super-pixels N (about 200). This makes the computational cost of the proposed method acceptable.

4. Experiments and analysis

In this section, several sets of experiments are performed to verify the superiority of the proposed salient object detection method.

Table 1: Summary of the public datasets.

Name	Size	Characteristics
MSRA10K [22]	10000	single object, high contrast, simple backgrounds
ECSSD [36]	1000	multiple objects, various object categories, structurally complex scene
DUT-OMRON [28]	5168	multiple objects, different scales and locations, cluttered backgrounds

Table 2: Summary of the evaluation metrics.

Name	Description
Precision-recall (PR) [2]	Precision (P): $\frac{\|S \cap G\|}{\|S\|}$, recall (R): $\frac{\|S \cap G\|}{\|G\|}$
F-measure [2]	$F_\beta = (1+\beta^2)\frac{P*R}{\beta^2 P+R}$, $\beta^2 = 0.3$
precision, recall, and F-measure with an adaptive threshold [20]	$thre = \frac{2}{W \times H}\sum_{i=1}^{W}\sum_{j=1}^{H} S(i,j)$
mean absolute error (MAE) [2]	$MAE = \frac{1}{W \times H}\sum_{i=1}^{W}\sum_{j=1}^{H} \|S(i,j) - G(i,j)\|$

4.1. Experimental setup

4.1.1. Datesets

In our experiments, we employ three public datesets, including MSRA10K [22], ECSSD [36], and DUT-OMRON [28], to test the superiority of our proposed method. The detailed descriptions for these datasets are presented in the Table 1.

4.1.2. Evaluation metrics

In the experiments, we adopt multiple evaluation metrics to verify the superiority of our proposed method. Table 2 shows the summary of the evaluation metrics. In the Table 2, S and G represent a saliency map and the corresponding ground-truth, respectively. $|\cdot|$ computes the number of non-zeros entries in

the mask. $S(i, j)$ represents the saliency value of the (i, j)-th pixel in S. W and H are the width and height of the image, respectively.

4.1.3. Parameters setting

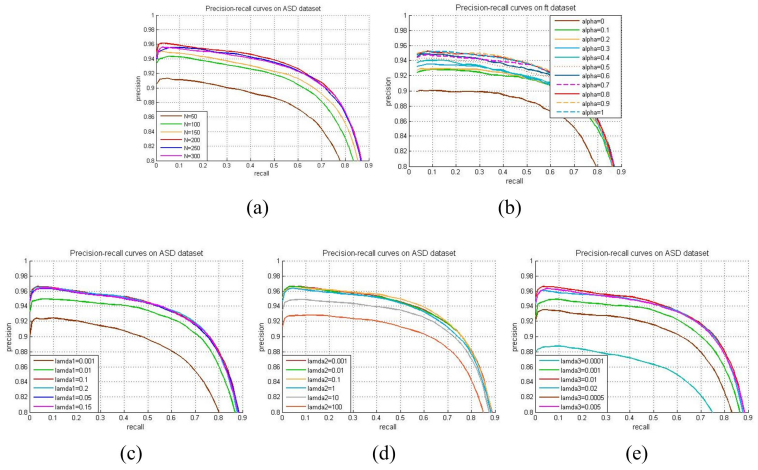


Figure 6: PR curves of the proposed method with different parameters values on the ASD dataset. (a) N; (b) α; (c) λ_1; (d) λ_2; (e) λ_3.

The parameters of our method involve the number of superpixels N, the fusion weight α in Eq. (9), and the balance parameters λ_1, λ_2, and λ_3 in Eq. (2). The setting of these parameters is important to the performance of the proposed method. We experimentally set the values of these parameters based on the performance of the proposed method on the ASD dataset [20], which has 1000 images. This dataset has similar characteristics with the MSRA10K dataset due to the fact that they are both collected from the MSRA database [37]. Each of them is set one by one, respectively, by fixing the others. More specifically, in the experimental part, we test the impacts of the parameters N (with $\alpha = 0.8$, $\lambda_1 = 0.1$, $\lambda_2 = 0.01$, $\lambda_3 = 0.01$), α (with $N = 200$, $\lambda_1 = 0.1$, $\lambda_2 = 0.01$, $\lambda_3 = 0.01$), λ_1 (with $N = 200$, $\alpha = 0.8$, $\lambda_2 = 0.01$, $\lambda_3 = 0.01$), λ_2 (with $N = 200$, $\alpha = 0.8$, $\lambda_1 = 0.1$, $\lambda_3 = 0.01$), and λ_3 (with $N = 200$, $\alpha = 0.8$, $\lambda_1 = 0.1$, $\lambda_2 = 0.01$) on the proposed method, respectively. The PR curves of the proposed method with different parameters are illustrated in Fig. 6.

From Fig. 6(a), it can be found that the performance increases with N, but nearly unchanged when N is greater than or equal to 200. So, we set $N = 200$. The rest of parameters are set in a similar way. In summary, these parameters are set to the values as in Table 3.

Table 3: Some parameters employed in our proposed method.

Parameter	N	α	λ_1	λ_2	λ_3
Value	200	0.8	0.1	0.01	0.01

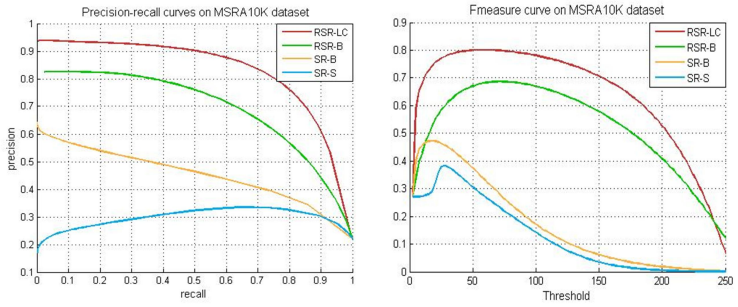
4.1.4. Experiments

First, we employ the public dataset, MSRA10K [22] to illustrate the validity of the robust sparse representation model and local consistency prior when applied to the detection of salient objects. Then, we employ three public datasets, including MSRA10K [22], ECSSD [36], and DUT-OMRON [28], to test the superiority of our proposed method. Finally, we analyze some failure examples of our proposed method.

4.2. Validity of RSR model and local consistency prior

In this part, we will employ the MSRA10K [22] dataset to illustrate the validity of the RSR model and local consistency prior employed in our proposed method. For that, we compare our proposed method (RSR-LC, for short) with another three methods, including a RSR based (RSR-B, for short) and two SR based (SR-S, SR-B, for short) methods mentioned in Fig. 1 in the earlier Introduction part. The RSR-B method can be seen as a special case of our proposed method with $\lambda_2 = \lambda_3 = 0$ in Eq. (3), without considering the local consistency prior. In RSR-B and SR-B methods, a global background dictionary obtained from the image boundary is employed. And in the SR-S method, a local dictionary using the image patches surrounding each test image patch is employed. Some detected results by these four methods can be seen in Fig. 1 in the earlier Introduction part.

Fig. 7 provides the quantitative results of the four methods. It can be found that the RSR-B method significantly outperforms the SR-S and SR-B methods in terms of the PR and F-measure curves. This clearly shows the superiority of the RSR model over the traditional SR model when applied to salient object detection. Moreover, it can also be easily found that the local consistency prior will further improve the performance by comparing the detected results of the RSR-B method and the proposed method.



(a) PR curve (b) F-measure curve

Figure 7: Quantitative comparisons of the RSR-LC, RSR-B, SR-B, SR-S methods.

4.3. Superiority of the proposed method

In addition to the MSRA10K [22] dataset, we will employ another two public benchmark datasets, i.e., ECSSD [36] and DUT-OMRON [28], to further test the superiority of the proposed method. As well, we compare the proposed method with another 9 state-of-the-art ones, including SR-LC [17], BFS [38], HDCT [24], PCA [39], TD [40], GC [41], GS [15], SF [42], SS [43].

4.3.1. Visual comparison

Fig. 8 illustrates some detected results obtained by these methods. As shown in Fig. 8, most of the methods mentioned in this paper could achieve satisfactory results for all of the three datasets. But by a careful comparison, we find that GC, HDCT, BFS and RSR-LC could obtain higher performance in foreground detection and background suppression than the other methods in most cases. Especially, for those images with similar foregrounds and backgrounds (e.g., the

third row in Fig. 8(a) and the second row in Fig. 8(b)), the proposed method RSR-LC could still detect the whole objects. Similar results could also be obtained by the proposed RSR-LC for those images with complex backgrounds (e.g., the last row in Fig. 8(c)). For these images, most of the state-of-the-art methods could not obtain a satisfactory result. In addition, we also find that the proposed method RSR-LC significantly outperforms SR-LC for these images. This further demonstrates the superiority of the RSR model over the traditional SR model.

Figure 8: Visual comparisons on different methods. (a) Results on the MSRA10K dataset; (b) Results on the ECSSD dataset; (c) Results on the DUT-OMRON dataset.

4.3.2. Quantitative comparison

Fig. 9 provides the quantitative results of different methods, complying with those subjective results mentioned in Fig. 8.

For MSRA10K dataset, our proposed method is competitive with HDC-T and better than the other state-of-the-art methods in terms of PR and F-measure curves. Given an appropriate threshold to segment the saliency map, our proposed method will obtain the highest F-measure value among the methods mentioned here. Furthermore, our proposed method obtains a good MAE value.

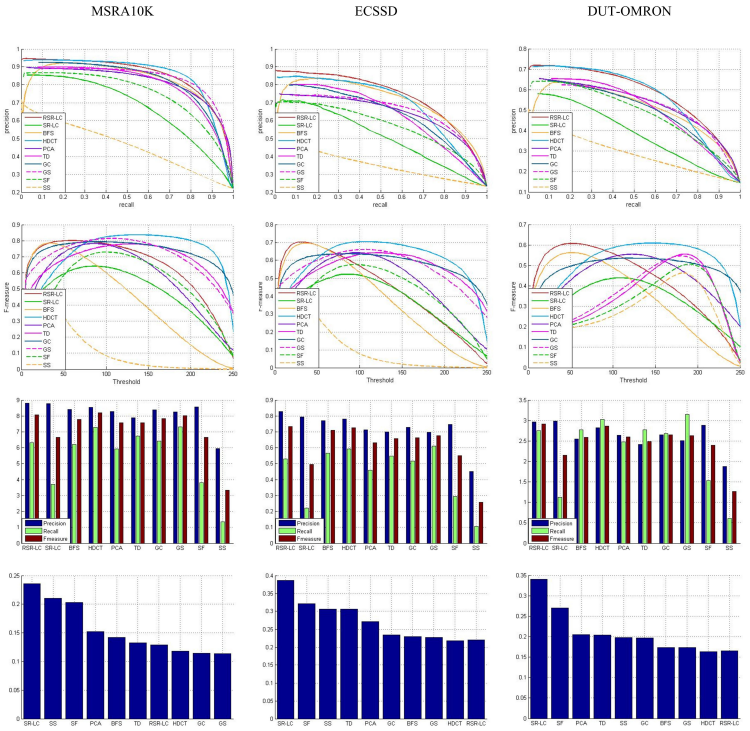


Figure 9: Quantitative comparisons on different methods. From left to right: Results on MSRA10K, ECSSD, and DUT-OMRON datasets, respectively. From top to bottom: PR curves, F-measure curves, precision, recall, and F-measure bars with an adaptive threshold, and MAE bars, respectively.

For ECSSD dataset, our proposed method achieves the best PR and F-

measure performance among the methods. That is to say, compared with the state-of-the-arts, our proposed method is more effective in salient object detection for those images where multiple salient objects exist and belong to various categories. Seen from the F-measure bars with an adaptive threshold, our proposed method also obtains the highest F-measure value among the methods. And our proposed method achieves the least MAE value.

For DUT-OMRON dataset, HDCT and our proposed method perform competitively and are both better than the other state-of-the-art methods in terms of the PR and F-measure curves. That is to say, our proposed method could still achieve satisfactory results for images with more complex background structures. Given an appropriate threshold, our proposed method can obtain the second best F-measure value among the methods. Moreover, our proposed method achieves the best performance in terms of the metric of MAE.

4.3.3. Computational complexity comparison

To demonstrate the computational efficiency of our method, we list the average computational time of some state-of-the-art methods[1] and our proposed method on the ASD dataset [20]. These methods are all run in Matlab 2013 on a personal computer with an Intel(R) Core(TM) i7-4790 3.60 GHz CPU. As shown in Table 4, our proposed method (RSR-LC) is slightly slower than SR-LC but much faster than PCA and BFS.

Table 4: Average running time of some methods (seconds per image).

Methods	Ours	SS	PCA	BFS	SR-LC
Time (s)	1.7649	0.0219	2.4589	6.4427	1.2793

[1]Here we just list those compared methods whose Matlab codes are provided in their corresponding project websites.

4.4. Failure cases

In the proposed method, those super-pixels near the image boundary are selected to construct the background dictionary based on the background prior [15]. In most cases, this may work well. However, in some images, salient objects appear near the image boundary, and thus some foreground regions will be contained in the background dictionary. As a result, these foreground regions will be mistakenly marked as background regions for these images. In addition, if the background regions far from the image boundary and those near the image boundary have obviously distinctive characteristics, some background regions will also be mistakenly marked as foregrounds. Some failure cases are illustrated in Fig. 10. Exploiting more efficient background dictionary construction methods will overcome this problem and we leave this as the future work.

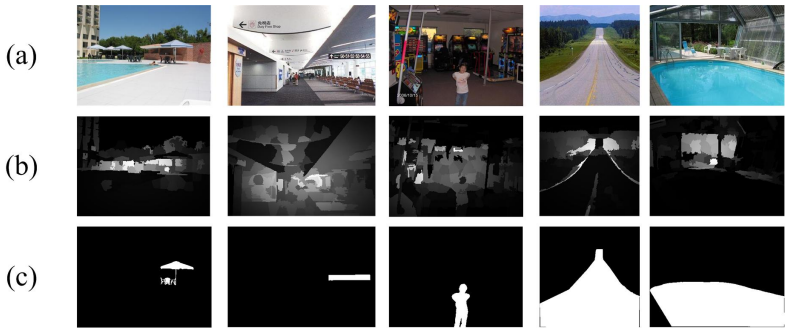


Figure 10: Failure cases. (a) Original images; (b) Saliency maps obtained by the proposed method; (c) Ground truth.

5. Conclusion

In this paper, we have presented a new salient object detection method that incorporates a local consistency prior into the robust sparse representation (RSR) model. By virtue of RSR, salient objects can be seen as strong but sparse "outliers" within an image, which allows to reformulate the salient object detection problem as a sparsity pursuit one. Given the same background dictionary, the RSR based method can better suppress the background noise than those

traditional SR based ones. Moreover, owing to the local consistency prior, the whole salient object could be more uniformly highlighted. Experimental results demonstrate that the proposed method significantly outperforms the traditional SR based methods and is competitive with some current state-of-the-art methods, especially for those images with complex structures.

Acknowledgement

This work is supported by the National Natural Science Foundation of China under Grant No. 61104212, by Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2016JM6008), and by the Fundamental Research Funds for the Central Universities under Grant No. NSIY211416.

Appendix

Appendix A.

In this appendix, the update scheme required for solving Eq. (6) in the text is described in detail.

(1) Update Z:

$$\begin{aligned} Z &= \arg\min_{Z} \|Z\|_1 + \lambda_2 tr(ZLZ^T) + \langle Y, X - DZ - E\rangle + \frac{\mu}{2}\|X - DZ - E\|_F^2 \\ &= \arg\min_{Z} \|Z\|_1 + \lambda_2 tr(ZLZ^T) + \frac{\mu}{2}\left\|X - DZ - E + \frac{1}{\mu}Y\right\|_F^2 \\ &= \arg\min_{Z} \|Z\|_1 + f(Z) \end{aligned} \tag{A.1}$$

where $f(Z) = \lambda_2 tr(ZLZ^T) + \frac{\mu}{2}\left\|X - DZ - E + \frac{1}{\mu}Y\right\|_F^2$. This sub-optimization problem can be solved by using the modified SpaRSA-based method [35] in an iterated way, i.e.,

$$Z^{(t+1)} = \arg\min_{Z} \frac{1}{\eta_Z^{(t)}}\|Z\|_1 + \frac{1}{2}\left\|Z - (Z^{(t)} - \frac{1}{\eta_Z^{(t)}}\nabla_Z f(Z^{(t)}))\right\|_F^2 \tag{A.2}$$

where $\eta_Z^{(t)} = 1.02\left(2\lambda_2 \|L\|_F^2 + \mu^{(t)} \left\|D^T D\right\|_F^2\right)$ [44, 45]. $\nabla_Z f(Z^{(t)})$ is the partial differential of $f(Z)$ with respect to Z in the t-th iteration, and is computed by:

$$\nabla_Z f(Z^{(t)}) = 2\lambda_2 Z^{(t)} L - \mu D^T (X - DZ^{(t)} - E^{(t)} + \frac{1}{\mu^{(t)}} Y^{(t)}) \tag{A.3}$$

Thus, the sub-optimization problem in Eq. (A.2) has the following closed-form solution [34]:

$$Z^{(t+1)} = S_{\frac{1}{\eta_Z^{(t)}}}\left(Z^{(t)} - \frac{1}{\eta_Z^{(t)}} \nabla_Z f(Z^{(t)})\right) \tag{A.4}$$

where the threshold function $S_\tau(x)$ is defined as

$$S_\tau(x) = \begin{cases} x - \tau, & \text{if } x > \tau \\ x + \tau, & \text{if } x < -\tau \\ 0, & \text{otherwise} \end{cases} \tag{A.5}$$

(2) Update E:

$$\begin{aligned} E &= \arg\min_E \lambda_1 \|E\|_{2,1} + \lambda_3 tr(ELE^T) + \langle Y, X - DZ - E\rangle + \frac{\mu}{2} \|X - DZ - E\|_F^2 \\ &= \arg\min_E \lambda_1 \|E\|_{2,1} + \lambda_3 tr(ELE^T) + \frac{\mu}{2} \left\|X - DZ - E + \frac{1}{\mu} Y\right\|_F^2 \\ &= \arg\min_E \lambda_1 \|E\|_{2,1} + f(E) \end{aligned} \tag{A.6}$$

where $f(E) = \lambda_3 tr(ELE^T) + \frac{\mu}{2} \left\|X - DZ - E + \frac{1}{\mu} Y\right\|_F^2$. Similarly, this sub-optimization problem can also be solved by using the modified SpaRSA-based method [33] in an iterated way, i.e.,

$$E^{(t+1)} = \arg\min_E \frac{\lambda_1}{\eta_E^{(t)}} \|E\|_{2,1} + \frac{1}{2} \left\|E - \left(E^{(t)} - \frac{1}{\eta_E^{(t+1)}} \nabla_E f\left(E^{(t)}\right)\right)\right\|_F^2 \tag{A.7}$$

where $\eta_E^{(t)} = 1.02\left(2\lambda_3 \|L\|_F^2 + \mu^{(t)}\right)$ [44, 45]. $\nabla_E f(E^{(t)})$ is the partial differential of $f(E)$ with respect to E in the t-th iteration, and is computed by:

$$\nabla_E f(E^{(t)}) = 2\lambda_3 E^{(t)} L - \mu(X - DZ^{(t)} - E^{(t)} + \frac{1}{\mu} Y^{(t)}) \tag{A.8}$$

The sub-optimization problem in Eq. (A.7) has the following closed-form solution [34]

$$E^{(t+1)}(:,i) = \begin{cases} \frac{\left(\|G(:,i)\|_2 - \frac{\lambda_1}{\eta_E^{(t)}}\right)}{\|G(:,i)\|_2} G(:,i), & \text{if } \|G(:,i)\|_2 \geq \frac{\lambda_1}{\eta_E^{(t)}} \\ 0, & \text{otherwise} \end{cases} \tag{A.9}$$

where $G = E^{(t)} - \frac{1}{\eta_E^{(t)}} \nabla_E f\left(E^{(t)}\right)$. $E(:,i)$ and $G(:,i)$ denote the i-th columns of the matrices E and G, respectively.

References

[1] P. Jiang, H. Ling, J. Yu, J. Peng, Salient region detection by UFO: uniqueness, focusness and objectness, in: IEEE International Conference on Computer Vision, 2013, pp. 1976–1983.

[2] A. Borji, M. M. Cheng, H. Jiang, J. Li, Salient object detection: A benchmark, IEEE Transactions on Image Processing 24 (12) (2015) 5706–5722.

[3] J. Han, K. N. Ngan, M. Li, H. J. Zhang, Unsupervised extraction of visual attention objects in color images, IEEE Transactions on Circuits & Systems for Video Technology 16 (1) (2006) 141–145.

[4] P. Peng, L. Shao, J. Han, J. Han, Saliency-aware image-to-class distances for image classification, Neurocomputing 166 (C) (2015) 337–345.

[5] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma, Robust recovery of subspace structures by low-rank representation, IEEE Transactions on Pattern Analysis & Machine Intelligence 35 (1) (2013) 171–184.

[6] D. Gao, S. Han, N. Vasconcelos, Discriminant saliency, the detection of suspicious coincidences, and applications to visual recognition, IEEE Transactions on Pattern Analysis & Machine Intelligence 31 (6) (2009) 989–1005.

[7] P. Wang, J. Wang, G. Zeng, J. Feng, Salient object detection for searched web images via global saliency, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 3194–3201.

[8] J. Han, E. J. Pauwels, P. De Zeeuw, Fast saliency-aware multi-modality image fusion, Neurocomputing 111 (6) (2013) 70–80.

[9] Y. Li, Y. Zhou, L. Xu, X. Yang, Incremental sparse saliency detection, in: International Conference on Image Processing, 2009, pp. 3093–3096.

[10] B. Han, H. Zhu, Y. Ding, Bottom-up saliency based on weighted sparse coding residual, in: International Conference on Multimedea, 2011, pp. 1117–1120.

[11] J. Yan, M. Zhu, H. Liu, Y. Liu, Visual saliency detection via sparsity pursuit, IEEE Signal Processing Letters 17 (8) (2010) 739–742.

[12] N. Li, B. Sun, J. Yu, A weighted sparse coding framework for saliency detection, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 5216–5223.

[13] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Transactions on Pattern Analysis & Machine Intelligence 31 (2) (2009) 210–227.

[14] C. Li, Y. Ma, X. Mei, C. Liu, Hyperspectral image classification with robust sparse representation, IEEE Geoscience & Remote Sensing Letters 13 (5) (2016) 641–645.

[15] Y. Wei, F. Wen, W. Zhu, J. Sun, Geodesic saliency using background priors, in: European Conference on Computer Vision, 2012, pp. 29–42.

[16] H. Lu, X. Li, L. Zhang, R. Xiang, Dense and sparse reconstruction error based saliency descriptor, IEEE Transactions on Image Processing 25 (4) (2016) 1–1.

[17] L. Huo, S. Yang, L. Jiao, S. Wang, S. Wang, Local graph regularized sparse reconstruction for salient object detection, Neurocomputing 194 (C) (2016) 348–359.

[18] Q. Zhang, M. D. Levine, Robust multi-focus image fusion using multi-task sparse representation and spatial context, IEEE Transactions on Image Processing 25 (5) (2016) 2045–2058.

[19] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Transactions on Pattern Analysis & Machine Intelligence 20 (11) (1998) 1254–1259.

[20] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2009, pp. 1597–1604.

[21] R. Achanta, S. Ssstrunk, Saliency detection using maximum symmetric surround, in: IEEE International Conference on Image Processing, 2010, pp. 2653–2656.

[22] M. M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, S. M. Hu, Global contrast based salient region detection, IEEE Transactions on Pattern Analysis & Machine Intelligence 37 (3) (2015) 569–582.

[23] J. Ren, X. Gong, L. Yu, W. Zhou, M. Y. Yang, Exploiting global priors for rgb-d saliency detection, in: IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015, pp. 25–32.

[24] J. Kim, D. Han, Y. W. Tai, J. Kim, Salient region detection via high-dimensional color transform, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 883–890.

[25] N. Tong, H. Lu, Y. Zhang, X. Ruan, Salient object detection via global and local cues, Pattern Recognition 48 (10) (2015) 3258–3267.

[26] Q. Fan, C. Qi, Saliency detection based on global and local short-term sparse representation, Neurocomputing 175 (2016) 81–89.

[27] W. Zhu, S. Liang, Y. Wei, J. Sun, Saliency optimization from robust background detection, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2814–2821.

[28] C. Yang, L. Zhang, H. Lu, R. Xiang, Saliency detection via graph-based manifold ranking, in: IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3166–3173.

[29] D. A. Klein, S. Frintrop, Center-surround divergence of feature statistics for salient object detection, in: IEEE Conference on Computer Vision, 2011, pp. 2214–2219.

[30] L. Wang, H. Lu, X. Ruan, M. H. Yang, Deep networks for saliency detection via local estimation and global search, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3183–3192.

[31] W. Wang, J. Shen, L. Shao, Consistent video saliency using local gradient flow optimization and global refinement, IEEE Transactions on Image Processing 24 (11) (2015) 4185–96.

[32] Y. Han, Y. Yang, Y. Yan, Z. Ma, N. Sebe, X. Zhou, Semisupervised feature selection via spline regression for video semantic recognition, IEEE Transactions on Neural Networks & Learning Systems 26 (2) (2015) 252–264.

[33] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. SuSstrunk, SLIC superpixels compared to state-of-the-art superpixel methods, IEEE Transactions on Pattern Analysis & Machine Intelligence 34 (11) (2012) 2274–82.

[34] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Foundations & Trends in Machine Learning 3 (1) (2010) 1–122.

[35] S. J. Wright, R. D. Nowak, M. A. T. Figueiredo, Sparse reconstruction by separable approximation, IEEE Transactions on Signal Processing 57 (7) (2009) 3373–3376.

[36] J. Shi, Q. Yan, L. Xu, J. Jia, Hierarchical image saliency detection on extended cssd, IEEE Transactions on Pattern Analysis & Machine Intelligence 38 (4) (2016) 717–729.

[37] T. Liu, J. Sun, N. N. Zheng, X. Tang, Learning to detect a salient object, in: IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.

[38] J. Wang, H. Lu, X. Li, N. Tong, W. Liu, Saliency detection via background and foreground seed selection, Neurocomputing 152 (C) (2015) 359–368.

[39] M. Ran, A. Tal, L. Zelnikmanor, What makes a patch distinct?, in: IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 1139–1146.

[40] C. Scharfenberger, A. Wong, K. Fergani, J. S. Zelek, Statistical textural distinctiveness for salient region detection in natural images, in: IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 979–986.

[41] M. M. Cheng, J. Warrell, W. Y. Lin, S. Zheng, Efficient salient region detection with soft image abstraction, in: IEEE Conference on Computer Vision, 2013, pp. 1529–1536.

[42] P. Krahenbuhl, Saliency filters: Contrast based filtering for salient region detection, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 733–740.

[43] X. Hou, J. Harel, C. Koch, Image signature: Highlighting sparse salient regions, IEEE Transactions on Pattern Analysis & Machine Intelligence 34 (1) (2012) 194–201.

[44] M. Yin, J. Gao, Z. Lin, Laplacian regularized low-rank representation and its applications, IEEE Transactions on Pattern Analysis & Machine Intelligence 38 (3) (2016) 504–517.

[45] D. Tao, J. Cheng, M. Song, X. Lin, Manifold ranking-based matrix factorization for saliency detection, IEEE Transactions on Neural Networks & Learning Systems 27 (6) (2016) 1122–1134.