

Northumbria Research Link

Citation: Zhang, Qiang, Liu, Yi, Zhu, Siyang and Han, Jungong (2017) Salient object detection based on super-pixel clustering and unified low-rank representation. *Computer Vision and Image Understanding*, 161. pp. 51-64. ISSN 1077-3142

Published by: Elsevier

URL: <https://doi.org/10.1016/j.cviu.2017.04.015>
<<https://doi.org/10.1016/j.cviu.2017.04.015>>

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/id/eprint/30911/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

Salient object detection based on super-pixel clustering and unified low-rank representation

Zhang Qiang^{a, b}

Liu Yi^a

Zhu Siyang^a

Jungong Han^{c, *}

jungong.han@northumbria.ac.uk, jungonghan@gmail.com

^aKey Laboratory of Electronic Equipment Structure Design, Ministry of Education, Xidian University, Xi'an, Shaanxi 710071, China

^bCenter for Complex Systems, School of Mechano-Electronic Engineering, Xidian University, Xi'an Shaanxi 710071, China

^cDepartment of Computer and Information Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, U.K

*Corresponding author. Northumbria University, Pandon Building, NE2 1XE, Newcastle upon Tyne, UK.

Abstract

In this paper, we present a novel salient object detection method, efficiently combining Laplacian sparse subspace clustering (LSSC) and unified low-rank representation (ULRR). Unlike traditional low-rank matrix recovery (LRMR) based saliency detection methods which mainly extract saliency from pixels or super-pixels, our method advocates the saliency detection on the super-pixel *clusters* generated by LSSC. By doing so, our method succeeds in extracting large-size salient objects from cluttered backgrounds, against the detection of small-size salient objects from simple backgrounds obtained by most existing work. The entire algorithm is carried out in two stages: region clustering and cluster saliency detection. In the first stage, the input image is segmented into many super-pixels, and on top of it, they are further grouped into different clusters by using LSSC. Each cluster contains multiple super-pixels having similar features (e.g., colors and intensities), and may correspond to a part of a salient object in the foreground or a local region in the background. In the second stage, we formulate the saliency detection of each super-pixel cluster as a unified low-rankness and sparsity pursuit problem using a ULRR model, which integrates a Laplacian regularization term with respect to the sparse error matrix into the traditional low-rank representation (LRR) model. The whole model is based on a sensible cluster-consistency assumption that the spatially adjacent super-pixels within the same cluster should have similar saliency values, similar representation coefficients as well as similar reconstruction errors. In addition, we construct a primitive dictionary for the ULRR model in terms of the local-global color contrast of each super-pixel. On top of it, a global saliency measure covering the representation coefficients and a local saliency measure considering the sparse reconstruction errors are jointly employed to define the final saliency measure. Comprehensive experiments over diverse publicly available benchmark data sets demonstrate the validity of the proposed method.

Keywords: Salient object detection; Laplacian sparse subspace clustering; Unified low-rank representation; Primitive saliency dictionary construction; Super-pixel cluster

1 Introduction

Saliency detection, which is closely related to the selective processing in human visual system, aims to locate eye fixations or interesting areas in images. Such a detector that mimics the human visual attention mechanism has been served as a foundation for many computer vision applications including object classification (Peng and Shao, 2015), image segmentation (Fouquier et al., 2012), image retrieval (Chen and Cheng, 2009), image fusion (Han et al., 2013), and image thumbnailing (Marchesotti et al., 2009).

Most of existing saliency detection works focus on one of the following two specific tasks (Li and Hou, 2014): fixation prediction or salient object detection. The goal of the former is to compute a probabilistic map of an image to simulate the eye movement behaviors of human, while the latter is expected to generate a map that matches the annotated salient object mask. In this paper, we will concentrate on the latter, especially on the detection of large-size salient objects from a cluttered scene, because it can deal with more practical applications.

In the past few years, low-rank matrix recovery (LRMR) techniques, such as low rank representation (LRR) (Liu and Lin, 2013), robust principal component analysis (RPCA) (Cades and Li, 2011) and matrix completion (MC)

(Cades and Tao, 2009), were presented to recover low-rank structures from the corrupted data with sparse but strong noise. Such algorithms have attracted significant attentions in the field of computer vision and image processing due to their super capability to facilitate applications including image segmentation (Cheng and Liu, 2011), object tracking (Zhang and Liu, 2014), image classification (Zhang and Ghanem, 2013), image fusion (Wan et al., 2013), and so on. Not surprisingly, these LRMR techniques have recently been applied to saliency detection (Yan and Zhu, 2010; Lang and Liu, 2012; Shen and Wu, 2012; Rigas et al., 2015; Liu et al., 2015).

Most of the LRMR methods presume that the salient objects only occupy a few parts of the whole image and/or the features of the backgrounds lie in a low-dimensional subspace (Shen and Wu, 2012). Therefore, they first employed some LRMR techniques to decompose the feature matrix, constructed by the local patches from the input image, into a low-rank part plus a sparse noise part (or reconstruction error). Subsequently, they employed the sparse reconstruction errors to indicate the saliency of the local image patches and thus obtained the salient objects or regions within the input image. In general, these methods can well detect salient objects with small size and simple backgrounds, as illustrated in the first row of Fig. 1.

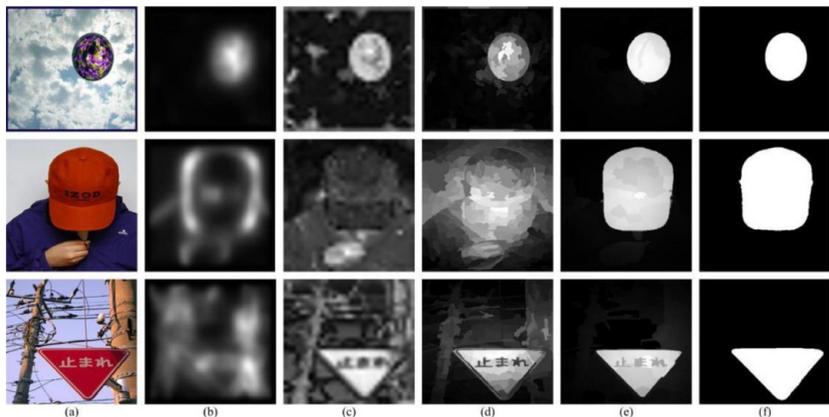


Fig. 1 Fig-1: Saliency detection results by different LRMR based methods. (a) Original images; (b) SR_RPCA (Yan and Zhu, 2010); (c) LRR (Lang and Liu, 2012); (d) ULRR (Shen and Wu, 2012); (e) Proposed; (f) Ground truth.

alt-text: Fig 1

However, such an assumption is no longer reasonable when the input image contains large-size salient objects with complex backgrounds, inevitably giving rise to unsatisfactory results if an LRMR method is straightforwardly employed. For example, as illustrated in the second row of Fig. 1, the traditional LRMR based algorithms mentioned here could not produce uniform saliency values for the whole object when large-size salient objects appear on the image. In addition, as shown in the third row of Fig. 1, most of them mistakenly label a part of the background as the salient region in the case that the backgrounds contain multiple texture regions (e.g., wires and poles).

Aiming to solve the two problems mentioned above, we present a new method based on the Laplacian sparse subspace clustering (LSSC) (Xie et al., 2013) and a unified low-rank representation (ULRR). As can be seen in Fig. 2, the proposed method consists of region clustering and cluster saliency detection. More specifically, the input image is first segmented into many super-pixels, and on top of them, different clusters are formed by grouping them using LSSC. Each cluster contains multiple super-pixels with similar features (e.g., colors and intensities), and may correspond to a part of a salient object in the foreground or a local region in the background. Thus, the detection of salient objects can be converted to the detection of different salient clusters.

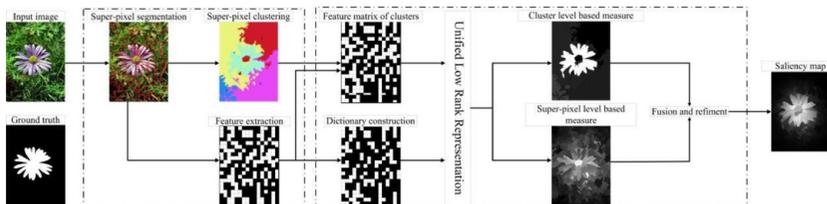


Fig. 2 Fig-2: Diagram of the proposed method.

alt-text: Fig 2

At the later stage, we apply the ULRR, i.e., by integrating a Laplacian regularization term with respect to the sparse error matrix into the traditional LRR model (Liu and Lin, 2013), on the feature matrix of each cluster. For

doing so, a primitive saliency dictionary is first constructed based on the local-global color contrast of each super-pixel. Afterwards, two saliency measures are constructed, in which one is based on the ULRR coefficients for the global saliency detection with respect to the entire image while the other one is based on the sparse reconstruction errors for the local saliency detection with respect to each cluster. Finally, the saliency maps derived by the two measures are fused and reinforced into a full-resolution saliency map. Experimental results demonstrate the superiority of the proposed method over some state-of-the-art methods, including traditional LRMR based and clustering based methods.

In summary, the main contributions of this paper are as follows:

1. We formulate the saliency detection of each super-pixel cluster as a low-rankness and sparsity pursuit problem by using a unified low-rank representation (ULRR) model, i.e., by integrating a Laplacian regularization with respect to the sparse error matrix into the traditional LRR model (Liu and Lin, 2013). This is based on a sensible cluster-consistency assumption that the spatially adjacent super-pixels within the same cluster should have similar saliency values and thus have similar representation coefficients and reconstruction errors. As a result, the whole salient objects are expected to be uniformly highlighted and some isolated regions in the detected result are also expected to be suppressed.
2. We construct a primitive saliency dictionary for the ULRR decomposition on the feature matrices of super-pixel clusters. This clearly differs from the traditional LRR based saliency detection method (Lang and Liu, 2012), in which the data themselves are directly served as the dictionary.
3. We construct two saliency measures by using the ULRR decomposition coefficients and sparse reconstruction errors, respectively. The former is a global cluster-level measure for detecting large objects, while the latter is a local super-pixel-level measure for the detection of small objects or local regions within a large object. The two measures are combined to construct the final saliency measure, thus enabling us to effectively detect both large and small salient objects.

The rest of this paper is organized as follows: Section 2 reviews the related work. In Section 3, the proposed method is described in detail. Experimental results as well as some insightful conclusions are given in Section 4 and Section 5, respectively.

2 Related work

In the past decade, numerous visual saliency detection methods have been presented, which could be generally classified into two categories (Xie et al., 2013): top-down and bottom-up. The former depends on the task at hand whereas the latter is driven by the input image tending to be application agnostic. Here, we limit our review to the bottom-up visual saliency detection methods only due to their relevance to our work.

Many earlier bottom-up visual saliency detection methods (Itti et al., 1998, Ma and Zhang, 2003, Harel et al., 2006, Bruce and Tsotsos, 2005, Hou and Zhang, 2007, Guo and Zhang, 2010, Guo et al., 2008, Li and Martin, 2013, Zhang and Han, 2016) were presented to simulate the eye movement or fixation behaviors of human. For example, as a pioneer, Itti et al. (Itti et al., 1998) derived a bottom-up visual saliency map using center-surround differences across multi-scale image features. In our previous work (Zhang and Han, 2016), we applied deep learning for co-saliency detection, aiming at extracting common salient regions in multiple related images.

In recent years, the research in this field evolves into a new phase. Instead of predicting a few fixation points in an image, new saliency detection methods uniformly highlight the entire salient region in the foreground (Achanta and Hemami, 2009, Cheng and Mitra, 2015, Gong and Tao, 2015, Wang et al., 2016, Chakraborty and Mitra, 2016, Liu and Han, 2016, Kim et al., 2016, Wei and Wen, 2012). For example, Achanta et al. (Achanta and Hemami, 2009) first presented a frequency-tuned based salient region detection method that outputted full resolution saliency maps with well-defined boundaries of salient objects by substantially retaining more spatial frequency contents from the original image. Most of the previous methods mentioned above rely on the assumptions or priors on the objects. In (Wei and Wen, 2012), the authors tackled the problem from a different viewpoint: they focused more on the background rather than the object. Precisely, they exploited two common priors about backgrounds in natural images, i.e., boundary and connectivity priors, to provide more clues for the salient object detection.

More recently, some saliency or salient object detection methods were proposed based on LRMR (Yan and Zhu, 2010, Lang and Liu, 2012, Shen and Wu, 2012, Rigas et al., 2015, Liu et al., 2015). For example, in (Yan and Zhu, 2010), Yan et al. applied the low-rank sparsity matrix decomposition (i.e., RPCA (Cades and Li, 2011)) to the visual saliency detection task, and presented a saliency estimation model for object detection, which directly extracted the saliency information from the sparse matrix obtained by RPCA decomposition. In (Lang and Liu, 2012), a multi-task sparsity pursuit was presented to integrate multiple types of features for image saliency detection. Given an image described by multi-view features, its saliency map is inferred by seeking the consistently sparse elements from the joint low-rank and sparse decomposition of multiple-feature matrices. In (Shen and Wu, 2012), a unified salient object detection model was proposed to incorporate the traditional low-level features with higher-level guidance, in which an image was represented as a low-rank matrix plus sparse noise in a certain feature space. These methods generally work appropriately for the salient objects of small size. However, when detecting the salient objects of large size, these methods tend to only produce higher saliency values on the borders of the salient objects.

3 The proposed salient object detection method

As shown in Fig. 2, the proposed salient object detection method mainly consists of two parts: (1) Super-pixel segmentation and clustering; (2) Super-pixel cluster saliency detection, each being elaborated in the following

subsections.

3.1 Super-pixel segmentation and clustering

This part can be further decomposed into: (1) Super-pixel segmentation using simple linear iterative clustering (SLIC); (2) Super-pixel clustering based on Laplacian sparse subspace clustering (LSSC); (3) Feature extraction.

3.1.1 Super-pixel segmentation

Because of its high computation efficiency and low memory requirement, a simple iterative super-pixel clustering (SLIC) algorithm (Achanta and Shaji, 2012) is adopted to achieve the super-pixel segmentation in this paper. Specifically, given an input image I , a set of super-pixels $\{sp_i | i = 1, 2, \dots, N\}$ can be obtained by using SLIC, where N denotes the total number of super-pixels and is empirically set to 150 in this paper.

3.1.2 Super-pixel clustering

Generally, a super-pixel only denotes a regional atom without any perceptual meaning. As a result, the object and the background can be represented as a group of super-pixels, which is illustrated in Fig. 3(b). When directly performing the saliency detection onto the super-pixels, some super-pixels within the salient object would be mistakenly labeled as non-salient ones, while some super-pixels from the background would be falsely marked as salient ones. This is more likely to occur in the images where the large-size salient objects are coupled with complex backgrounds. To solve this problem, in the proposed method, we will group the super-pixels into different clusters, and perform the saliency detection on the super-pixel *clusters* rather than on the super-pixels. As a result, such a scheme is able to depress the saliency noise caused by the complex background.

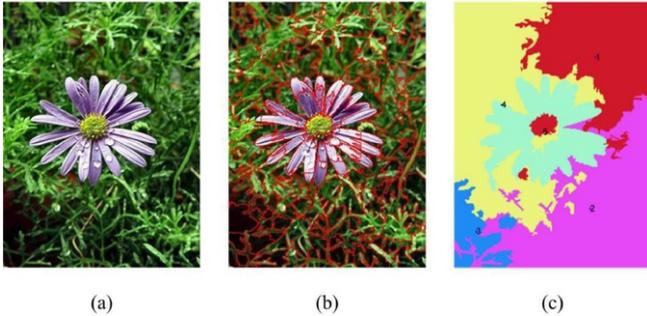


Fig. 3 Fig. 3. Super-pixel segmentation and cluster results. (a) Original image; (b) Super-pixel segmentation result; (c) Cluster result.

alt-text: Fig 3

Regarding the cluster algorithm, we directly use the method reported in by (Xie et al., 2013) by Xie et al., which fed the mid visual information via super-pixels into a Laplacian sparse subspace clustering (LSSC) method. To some extent, LSSC method can be considered as an extension of the sparse subspace clustering (SSC) (Elhamifar and Vidal, 2013) by introducing a Laplacian regularization term, which further enforces similar super-pixels to be clustered into the same group. More details about LSSC can be found in Appendix A.

Given a set of super-pixels $\{sp_i | i = 1, 2, \dots, N\}$ from an input image, a set of super-pixel clusters $\{C_k | k = 1, 2, \dots, K\}$ are obtained using LSSC (Xie et al., 2013), where K denotes the total number of clusters and will be discussed in the experimental part. Here, each cluster C_k contains N_k super-pixels, i.e., $C_k = \{sp_{k,j} | j = 1, 2, \dots, N_k\}$. Fig. 3 shows an example of super-pixel segmentation on an image and its clustering. As shown in Fig. 3(c), the salient object and the background are segmented into only a few number of clusters, thereby facilitating the complete detection of salient object and the suppression of noise from the background.

3.1.3 Feature extraction

Given an image I and a set of its super-pixel clusters $\{C_k | k = 1, 2, \dots, K\}$, the feature extraction (or feature matrix construction) for each super-pixel cluster C_k is described as follows:

- (1) For each pixel p_i in the image I , construct its feature vector $\mathbf{f}_i \in R^d$ of dimension $d = 53$ as suggested in (Shen and Wu, 2012) by: (I) Color feature $\mathbf{v}_{1,i} = [r_i, g_i, b_i, h_i, s_i]^T \in R^5$, where r_i, g_i and b_i denote the red, green and blue color channel components of pixel p_i respectively. h_i and s_i denote its hue and saturation components. (II) Edge feature $\mathbf{v}_{2,i} \in R^{12}$, which is constituted by the absolute values of the outputs of a set of steerable pyramid filters with 3 scales and 4 directions for pixel p_i . (III) Texture feature $\mathbf{v}_{3,i} \in R^{36}$, which is constituted by the absolute values of the outputs of a set of Gabor filters with 3 scales and 12 directions for the current pixel. Thus the feature vector \mathbf{f}_i is constructed by vertically stacking the vectors $\mathbf{v}_{1,i}$, $\mathbf{v}_{2,i}$ and $\mathbf{v}_{3,i}$ i.e.,

$$\mathbf{f}_i = \begin{bmatrix} \mathbf{v}_{1,i} \\ \mathbf{v}_{2,i} \\ \mathbf{v}_{3,i} \end{bmatrix} \in R^{53}. \quad (1)$$

(2) Construct the feature vector $\mathbf{x}_j \in R^d$ for each super-pixel sp_j by averaging all the feature vectors of the pixels contained in the current super-pixel, i.e.,

$$\mathbf{x}_j = \left(\frac{1}{N_{sp_j}} \sum_{p_i \in sp_j} \mathbf{f}_i \right), \quad (2)$$

where N_{sp_j} denotes the number of pixels within the super-pixel sp_j .

(3) Construct the feature matrix $\mathbf{X}_k \in R^{d \times N_k}$ by using all of the feature vectors of the super-pixels grouped into the cluster C_k , i.e.,

$$\mathbf{X}_k = \left[\mathbf{x}_{k,1}, \mathbf{x}_{k,2}, \dots, \mathbf{x}_{k,N_k} \right], \quad (3)$$

where $\mathbf{x}_{k,j}$ denotes the feature vector of the j -th super-pixel $sp_{k,j}$ in the cluster C_k . And N_k refers to the number of super-pixels in the cluster C_k .

3.2 Super-pixel cluster saliency detection based on ULRR

As shown in Fig. 3, each super-pixel cluster corresponds to a part of an object in the foreground or a local region with similar textures in the background. Hence, the salient object detection in an image may be achieved via the saliency detection of different super-pixel clusters, which is similar to that in (Xie et al., 2013). But differently, in the proposed method, we formulate the saliency detection of different super-pixel clusters as a low rankness and sparsity pursuit problem with the ULRR decomposition rather than under a Bayesian framework (Xie et al., 2013) considering the strong correlation among the super-pixels contained in each cluster.

In theory, the feature matrix \mathbf{X}_k obtained in SubSection 3.1-C for each cluster C_k has the intrinsic property of low rankness. However, it may be partially corrupted by some errors or noise in the real application. Given a dictionary $\mathbf{D}_k \in R^{d \times M_k}$ with M_k prototype atoms, the feature matrix \mathbf{X}_k of the super-pixel cluster C_k may be decomposed into a low-rank part plus a sparse error part (Liu and Lin, 2013), i.e.,

$$\mathbf{X}_k = \mathbf{D}_k \mathbf{Z}_k + \mathbf{E}_k, \quad k = 1, 2, \dots, K, \quad (4)$$

where $\mathbf{D}_k \mathbf{Z}_k$ denotes the "intrinsic" low-rank part contained in the matrix \mathbf{X}_k . $\mathbf{Z}_k \in R^{M_k \times N_k}$ refers to the sought after representation coefficient matrix, which is accordingly assumed to have the property of low rankness in this paper. $\mathbf{E}_k \in R^{d \times N_k}$ represents the error or noise part and is assumed to have sparse columns for dealing with the subsequent saliency detection of each super-pixel contained in the cluster C_k .

Eventually, for each super-pixel cluster C_k , its corresponding representation coefficient matrix \mathbf{Z}_k and error matrix \mathbf{E}_k can be obtained via solving the below low-rank representation (LRR) (Liu and Lin, 2013) problem, respectively:

$$\min_{\mathbf{Z}_k, \mathbf{E}_k} \|\mathbf{Z}_k\|_* + \lambda_k \|\mathbf{E}_k\|_{2,1}, \quad s.t. \quad \mathbf{X}_k = \mathbf{D}_k \mathbf{Z}_k + \mathbf{E}_k, \quad k = 1, \dots, K, \quad (5)$$

where $\|\mathbf{Z}_k\|_*$ denotes the nuclear norm of the matrix \mathbf{Z}_k and is defined as the sum of the singular values of the matrix \mathbf{Z}_k . It is a convex relaxation of the rank function (Liu and Lin, 2013). $\|\mathbf{E}_k\|_{2,1}$ denotes the $l_{2,1}$ -norm of the matrix \mathbf{E}_k and is defined as $\|\mathbf{E}_k\|_{2,1} = \sum_j \sqrt{\sum_i (\mathbf{E}_k(i,j))^2}$. $\mathbf{E}_k(i,j)$ is the (i,j) -th entry in the matrix \mathbf{E}_k . Parameter $\lambda_k > 0$ is used to balance the effects of the two parts.

In theory, it might be more reasonable to adaptively construct a local dictionary for each cluster due to large feature variations across the clusters. However, using different dictionaries $\{\mathbf{D}_k | k = 1, 2, \dots, K\}$ in Eq. (5) will give rise to the fact that the representation coefficients and error matrices $\{\mathbf{Z}_k, \mathbf{E}_k | k = 1, 2, \dots, K\}$ are obtained under different conditions. This will affect the fairness of the subsequent saliency measure for different super-pixel clusters or super-pixels. Therefore, we will prefer a global dictionary $\mathbf{D} \in R^{d \times N}$ in the proposed method, which will be constructed by the feature data $\{\mathbf{x}_j | j = 1, 2, \dots, N\}$ of all of the super-pixels in the input image.

Given a dictionary, each cluster will be well represented by solving Eq. (5) independently. However, such a scheme neglects the interrelationships among the spatially adjacent super-pixels (Li and Martin, 2013), thus giving rise to isolated regions in the detected result. In practice, the super-pixels that are spatially adjacent or have similar features are likely to have similar saliency values (Li and Martin, 2013). Therefore, it may be more reasonable to group those super-pixels into the same cluster. Accordingly, these super-pixels in the same cluster will have similar representation coefficients and reconstruction errors. This useful observation inspires us to involve such a cluster-

consistency prior into our proposed method, thereby ensuring the completeness of the segmented salient object. This idea can be implemented by integrating a Laplacian regularization term with respect to the reconstruction error into the LRR model in addition to the low-rankness constraint on the representation matrix.

The problem in Eq. (5) is thus amended to be the following unified LRR (ULRR) one

$$\min_{\substack{\mathbf{Z}_1, \dots, \mathbf{Z}_K \\ \mathbf{E}_1, \dots, \mathbf{E}_K}} \sum_{k=1}^K \|\mathbf{Z}_k\|_* + \lambda_1 \|\mathbf{E}\|_{2,1} + \lambda_2 \text{tr}(\mathbf{E}\mathbf{L}\mathbf{E}^T), \quad \text{s.t.} \quad \mathbf{X}_k = \mathbf{D}\mathbf{Z}_k + \mathbf{E}_k, \quad k = 1, \dots, K, \quad (6)$$

where \mathbf{E} is formed by horizontally concatenating $\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_K$ together along the row, i.e., $\mathbf{E} = [\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_K] \in R^{d \times N}$. $N = \sum_{k=1}^K N_k$ denotes the total number of super-pixels contained in the input image. λ_1 and λ_2 are two positive trade-off parameters and are experimentally set to 0.01 and 0.1, respectively. The Laplacian regularization term $\lambda_2 \text{tr}(\mathbf{E}\mathbf{L}\mathbf{E}^T)$ is computed by

$$\text{tr}(\mathbf{E}\mathbf{L}\mathbf{E}^T) = \frac{1}{2} \sum_{i,j}^N \|e_i - e_j\|_2^2 \omega_{ij} \quad (7)$$

In Eq. (7), e_i denotes the i -th column of the matrix \mathbf{E} . The weight ω_{ij} refers to the similarity between the i -th and j -th super-pixels and is defined as

$$\omega_{ij} = \begin{cases} \exp\left(-\frac{\|p_i - p_j\|_2^2}{2\sigma_p^2}\right) \cdot \exp\left(-\frac{\|x_i - x_j\|_2^2}{2\sigma_f^2}\right), & \text{if } s_i, s_j \text{ belong to the same} \\ & \text{cluster} \\ 0, & \text{otherwise} \end{cases}, \quad (8)$$

where $p_p, p_f \in R^2$ denote the center positions of the super-pixels sp_i and sp_j , $x_p, x_f \in R^m$ are their corresponding feature vectors. σ_p and σ_f are two scaling parameters, and are experimentally set to 0.5 and $\sqrt{0.5}$, respectively. Based on these weights, an affinity matrix $\mathbf{W} \in R^{N \times N}$ with its (i, j) -th entry $\mathbf{W}_{i,j} = \omega_{ij}$ and a diagonal degree matrix $\mathbf{H} \in R^{N \times N}$ with its i -th diagonal element $\mathbf{H}_{i,i} = \sum_j \mathbf{W}_{i,j}$ are constructed. The Laplacian matrix \mathbf{L} is thus defined as $\mathbf{L} = \mathbf{H} - \mathbf{W}$.

As discussed above, the low-rank part $\mathbf{D}\mathbf{Z}_k$ in Eq. (6) contains the "intrinsic" characteristics of each super-pixel cluster C_k . The saliency or difference among different super-pixel clusters $\{C_k | k = 1, 2, \dots, K\}$ could be achieved by comparing their low-rank parts $\{\mathbf{D}\mathbf{Z}_k | k = 1, 2, \dots, K\}$, i.e., their ULRR coefficients $\{\mathbf{Z}_k | k = 1, 2, \dots, K\}$. In addition, the minimization of $l_{2,1}$ norm enables the columns of \mathbf{E} to be near to zeros (i.e., have sparse columns). Here, each column in the matrix \mathbf{E} corresponds to a super-pixel, indicating that the larger (smaller) the magnitude the more salient (non-salient) the super-pixel is (Lang and Liu, 2012). Therefore, in the proposed method, we will jointly employ the representation coefficients $\{\mathbf{Z}_k | k = 1, 2, \dots, K\}$, and the reconstruction errors \mathbf{E} to construct the saliency map.

To sum up, the proposed saliency detection method for super-pixel clusters in this subsection consists of four parts: (1) Primary saliency dictionary construction; (2) ULRR problem solving; (3) Saliency measure by jointly optimizing the representation coefficients and reconstruction errors; (4) Saliency map fusion and refinement. In the following contents, we will describe each part in detail.

3.2.1 Primitive saliency dictionary construction

In this part, we will employ the feature data $\{\mathbf{x}_j | j = 1, 2, \dots, N\}$ of all the super-pixels in the input image to construct the dictionary \mathbf{D} . However, instead of directly employing the feature data as the dictionary, we adopt a two-step approach, where the first step is to coarsely measure the saliency of each super-pixel. Next, a primitive saliency dictionary is constructed with the aim to better exploit the ULRR coefficients in the subsequent saliency measure for each super-pixel cluster. In the primitive saliency dictionary, each feature data \mathbf{x}_j will be still employed as a dictionary atom (or a column in the dictionary \mathbf{D}), but its position (or column number) will be rearranged in terms of its initial saliency score.

For that, given a set of super-pixels $\{sp_i | i = 1, 2, \dots, N\}$, their initial saliency scores $\{CS_i | i = 1, 2, \dots, N\}$ are first obtained by using a local-global color contrast based method (Perazzi and Krahenbull, 2012) in this paper because of its efficiency. Then a set of new numbers $\{s_i | i = 1, 2, \dots, N; 1 \leq s_i \leq N; CS_1 \geq CS_2 \geq \dots \geq CS_N\}$ are obtained according to the initial saliency values. The global dictionary \mathbf{D} is thus constructed as

$$\mathbf{D} = [\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \dots, \mathbf{x}_{s_N}], \quad (9)$$

where \mathbf{x}_{s_i} ($i = 1, 2, \dots, N$) denotes the feature data of the s_i -th super-pixel sp_{s_i} in the input image.

As shown in Fig. 4, the atoms in the dictionary \mathbf{D} can be divided into three groups: (1) the first ρ atoms from the feature data of the potential foreground super-pixels; (2) the last ρ atoms from the feature data of the potential background super-pixels; (3) the rest atoms from the feature data of uncertain super-pixels. ρ is empirically set to 20 in this paper.

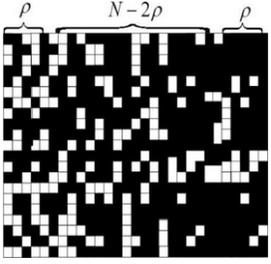


Fig. 4 Primitive saliency dictionary.

alt-text: Fig 4

3.2.2 ULRR problem solving

Solving the problem in Eq. (6) equals a convex optimization, for which there are various methods available. In this paper, we first convert it to the following equivalent problem

$$\min_{\substack{\mathbf{Z}_1, \dots, \mathbf{Z}_K \\ \mathbf{E}_1, \dots, \mathbf{E}_K}} \sum_{k=1}^K \|\mathbf{J}_k\|_* + \lambda_1 \|\mathbf{E}\|_{2,1} + \lambda_2 \text{tr}(\mathbf{E}\mathbf{E}^T), \quad s.t. \quad \mathbf{X}_k = \mathbf{D}\mathbf{Z}_k + \mathbf{E}_k, \\ \mathbf{Z}_k = \mathbf{J}_k, \quad k = 1, \dots, K. \quad (10)$$

To solve it, a linearized alternating direction method with adaptive penalty (LADMAP) (Lin et al., 2011; Zhang et al., 2013) is adopted, which requires the minimization of the following augmented Lagrangian function:

$$L = \lambda_1 \|\mathbf{E}\|_{2,1} + \lambda_2 \text{tr}(\mathbf{E}\mathbf{E}^T) + \sum_{k=1}^K \left(\|\mathbf{J}_k\|_* + \langle \mathbf{Y}_k, \mathbf{X}_k - \mathbf{D}\mathbf{Z}_k - \mathbf{E}_k \rangle \right. \\ \left. + \langle \mathbf{W}_k, \mathbf{Z}_k - \mathbf{J}_k \rangle + \frac{\mu}{2} \|\mathbf{X}_k - \mathbf{D}\mathbf{Z}_k - \mathbf{E}_k\|_F^2 + \frac{\mu}{2} \|\mathbf{Z}_k - \mathbf{J}_k\|_F^2 \right), \quad (11)$$

where $\mathbf{Y}_1, \dots, \mathbf{Y}_K$ and $\mathbf{W}_1, \dots, \mathbf{W}_K$ are Lagrange multipliers used to remove the equality constraint in Eq. (10). $\mu > 0$ is a penalty parameter. $\langle \mathbf{A}, \mathbf{B} \rangle$ denotes the Euclidean inner product of matrices \mathbf{A} and \mathbf{B} . Apparently, this problem for now becomes unconstrained and can be thus minimized with respect to \mathbf{E}_k (or \mathbf{E}), \mathbf{X}_k and \mathbf{J}_k ($k = 1, 2, \dots, K$), respectively. Algorithm 1 summarizes the calculations of the ULRR. More details can be seen in Appendix B.

Algorithm 1 Unified low-rank representation (ULRR) algorithm using IALM.

alt-text: Table 1

Input: Data matrices $\{\mathbf{X}_k\}$, dictionary \mathbf{D} , and parameter λ .
Output: \mathbf{Z}_k ($k = 1, 2, \dots, K$) and \mathbf{E} .
Initialized: $\{\mathbf{Z}_k = \mathbf{0}, \mathbf{J}_k = \mathbf{0}, \mathbf{E}_k = \mathbf{0}, \mathbf{Y}_k = \mathbf{0}, \mathbf{W}_k = \mathbf{0} k = 1, 2, \dots, K\}$, $\mu = 10^{-6}$, $\varepsilon = 10^{-8}$, $\phi = 1.1$, $\mu_{\max} = 10^6$.
While not converged do
(1) Fix the others and update $\mathbf{J}_1, \mathbf{J}_2, \dots, \mathbf{J}_K$ using Eq. (B2);
(2) Fix the others and update $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_K$ using Eq. (B4);
(3) Fix the others and update \mathbf{E} using Eq. (B8);
(4) Update the multipliers \mathbf{Y}_k and \mathbf{W}_k ($k = 1, 2, \dots, K$):
$\mathbf{Y}_k^{j+1} = \mathbf{Y}_k^j + \mu^j (\mathbf{X}_k - \mathbf{D}\mathbf{Z}_k^{j+1} - \mathbf{E}_k^{j+1})$, $\mathbf{W}_k^{j+1} = \mathbf{W}_k^j + \mu^j (\mathbf{Z}_k^{j+1} - \mathbf{J}_k^{j+1})$;
(5) Update μ :
$\mu^{j+1} = \min(\mu^j \phi, \mu_{\max})$;

(6) Check the convergence conditions:

$$\max_k \left\| \mathbf{X}_k - \mathbf{D}\mathbf{Z}_k^{j+1} - \mathbf{E}_k^{j+1} \right\|_\infty < \varepsilon \quad \text{and} \quad \max_k \left\| \mathbf{Z}_k^{j+1} - \mathbf{J}_k^{j+1} \right\|_\infty < \varepsilon ;$$

where $\| \cdot \|_\infty$ denotes the l_∞ -norm of a matrix and is defined as the maximum absolute value of the entries in a matrix.

end while

3.2.3 Saliency measure

In this part, we will propose two saliency measures. One is based on the ULRR coefficients for the saliency detection of each super-pixel cluster, and the other is based on the reconstruction errors for the saliency detection of each super-pixel.

(1) Saliency measure based on the ULRR coefficients

As discussed in the earlier part of [Section 3.2](#), the low-rank part $\mathbf{D}\mathbf{Z}_k$, decomposed by the ULRR in [Eq. \(6\)](#), contains the "intrinsic" characteristics of each super-pixel cluster C_k . Each cluster may correspond to a part of a salient object in the foreground or a local region in the background. Therefore, the detection of a salient object in the foreground could be achieved by using the low-rank part (or information) contained in each cluster.

Consider the i -th column $\mathbf{x}_{k,i}$ which is the feature data of the i -th super-pixel grouped in the cluster C_k . Let $\mathbf{z}_{k,i} \in R^N$ and $\mathbf{e}_{k,i} \in R^d$ denote the i -th column of the matrices \mathbf{Z}_k and \mathbf{E}_k , respectively. Then $\mathbf{x}_{k,i}$ can be represented as

$$\begin{aligned} \mathbf{x}_{k,i} &= \mathbf{D}\mathbf{z}_{k,i} + \mathbf{e}_{k,i} = \begin{bmatrix} \mathbf{x}_{s_1} \\ \mathbf{x}_{s_2} \\ \vdots \\ \mathbf{x}_{s_N} \end{bmatrix} \begin{bmatrix} \mathbf{Z}_k(1,i) \\ \mathbf{Z}_k(2,i) \\ \vdots \\ \mathbf{Z}_k(N,i) \end{bmatrix} + \begin{bmatrix} \mathbf{E}_k(1,i) \\ \mathbf{E}_k(2,i) \\ \vdots \\ \mathbf{E}_k(N,i) \end{bmatrix} \\ &= \mathbf{x}_{s_1}\mathbf{Z}_k(1,i) + \mathbf{x}_{s_2}\mathbf{Z}_k(2,i) + \dots + \mathbf{x}_{s_N}\mathbf{Z}_k(N,i) + \mathbf{e}_{k,i}. \end{aligned} \quad (12)$$

Therefore, as discussed in [\(Zhang et al., 2013\)](#), the coefficient $\mathbf{Z}_k(j,i)$ ($j = 1, 2, \dots, N$) indicates the correlation (or similarity) between the data $\mathbf{x}_{k,i}$ and the j -th atom \mathbf{x}_{s_j} in the dictionary to some extent. Larger absolute value of $\mathbf{Z}_k(j,i)$ indicates higher correlation (similarity) between the data $\mathbf{x}_{k,i}$ and the atom \mathbf{x}_{s_j} .

Moreover, as shown in [Fig. 4](#), the first ρ atoms in the dictionary \mathbf{D} are from the potential foreground super-pixels, and the last ρ atoms are from the potential background super-pixels. As a result, the sum of the absolute values of the first ρ coefficients $\sum_{j=1}^{\rho} |\mathbf{Z}_k(j,i)|$ in the vector $\mathbf{z}_{k,i}$ may reflect the similarity between the super-pixel $sp_{k,i}$ and the potential foreground super-pixels. Accordingly, the sum of the absolute values of the last ρ coefficients $\sum_{j=N-\rho+1}^N |\mathbf{Z}_k(j,i)|$ in the vector $\mathbf{z}_{k,i}$ may reflect the similarity between the super-pixel $sp_{k,i}$ and the potential background super-pixels.

Let \mathbf{Z}_k^{FG} be the first ρ rows of ULRR coefficients \mathbf{Z}_k and \mathbf{Z}_k^{BG} be the last ρ rows of matrix \mathbf{Z}_k . Similarly, the sum of the absolute values of all the entries in the matrix \mathbf{Z}_k^{FG} , denoted by $\|\mathbf{Z}_k^{FG}\|_1$, implies the similarity between the super-pixel cluster C_k and the potential foreground super-pixels. The sum of the absolute values of all the entries in the matrix \mathbf{Z}_k^{BG} , denoted by $\|\mathbf{Z}_k^{BG}\|_1$, refers to the similarity between the super-pixel cluster C_k and the potential background super-pixels. Therefore, the proposed saliency measure $L(C_k)$ for each cluster C_k is defined by

$$L(C_k) = \|\mathbf{Z}_k^{FG}\|_1 - \|\mathbf{Z}_k^{BG}\|_1. \quad (13)$$

(2) Saliency measure based on reconstruction errors

As discussed above, the representation coefficients describe the similarity between each super-pixel cluster and those potential foreground or background super-pixels. Accordingly, the saliency for each super-pixel cluster with respect to the entire image can be computed by using [Eq. \(13\)](#), and the representation coefficient based saliency measure can also be seen as a global saliency measure.

Different from the representation coefficients, each column in the error matrix \mathbf{E} indicates the differences between each super-pixel and its corresponding super-pixel clusters. Those super-pixels that are significantly distinct from their corresponding super-pixel cluster regions actually produce higher reconstruction errors. In other words, the saliency for each super-pixel with respect to its corresponding super-pixel cluster can be measured by the error matrix \mathbf{E} . This is particularly true for those super-pixels corresponding to small salient objects or local parts within the large salient objects but with different features from the salient objects. In view of this fact, we also involve the sparse reconstruction error matrix \mathbf{E} to define a local saliency measure for each super-pixel in this part, which is expected to be complementary with the global representation coefficient based saliency measure for each super-pixel cluster introduced above.

Moreover, many studies have shown that incorporating some priors, such as boundary prior ([Li and Fu, 2014](#)), shape prior ([Jiang and Wang, 2011](#)) and color prior ([Shen and Wu, 2012](#)), could enhance the saliency detection results to some extent. Especially, the color prior is more frequently used in image saliency detection ([Kim et al., 2016](#)). Similarly, we also integrate the color prior ([Shen and Wu, 2012](#)) into the saliency detection of each super-pixel. And the saliency measure $S(sp)$ for the i -th

super-pixel sp_i is computed by

$$S(sp_i) = \|\mathbf{E}(:, i)\|_2 \times \text{prior}(sp_i), \quad (14)$$

where $\mathbf{E}(:, i)$ is the i -th column of the matrix \mathbf{E} , and $\|\mathbf{E}(:, i)\|_2$ denotes its l_2 -norm, i.e., $\|\mathbf{E}(:, i)\|_2 = \sqrt{\sum_j (\mathbf{E}(j, i))^2}$. $\text{prior}(sp_i)$ denotes the color prior of super-pixel sp_i and can be computed as in [Shen and Wu \(2012\)](#).

3.2.4 Pixel-level saliency map

According to the saliency measures $\{L(C_k) | k = 1, 2, \dots, K\}$ and $\{S(sp_i) | i = 1, 2, \dots, N\}$, two pixel-level saliency maps $Sal_L(p)$ and $Sal_S(p)$ with full resolution are obtained by [Eq. \(15\)](#) and [Eq. \(16\)](#), respectively, where p represents a pixel in the input image.

$$Sal_L(p) = L(C_k), \quad \text{if } p \in C_k, \quad (15)$$

$$Sal_S(p) = S(sp_i), \quad \text{if } p \in sp_i. \quad (16)$$

With the two pixel-level saliency maps $Sal_L(p)$ and $Sal_S(p)$, a fused saliency map $S_f(p)$ is obtained using a multiplicative strategy

$$S_f(p) = (Sal_L(p))^\alpha \times (Sal_S(p))^{1-\alpha}, \quad (17)$$

where α is a weight to be experimentally determined.

After that, the final pixel-level saliency map $S(p)$ with full resolution is obtained by integrating the center prior with the above saliency map, which is computed by

$$S(p) = G_o(p) \times S_f(p). \quad (18)$$

Here, the object-biased Gaussian model $G_o(p)$ in [Lu and Li \(2016\)](#), instead of the traditional Gaussian model, is employed as the center prior considering that salient object does not always appear at the image center.

Besides, similar to [Tong and Lu \(2015\)](#), we apply the Max-Flow method [\(Borkov and Kolmogorov, 2004\)](#) to smooth the pixel-level saliency map $S(p)$, and the smoothed saliency map is noted as $S_{smooth}(p)$. Thus, the final full-resolution pixel-level saliency map is formulated as

$$S_{final}(p) = \frac{S(p) + S_{smooth}(p)}{2}. \quad (19)$$

In summary, the main steps of the proposed salient object detection method can be described by [Algorithm 2](#). And [Fig. 5](#) illustrates the results from each component.

Algorithm 2 ~~Algorithm 2~~. Salient object detection based on the LSSC and ULRR.

alt-text: Table 2

Input: Image I ; parameters τ and α .

Output: Pixel-level saliency map $S(p)$.

Begin:

- (1) Super-pixel segmentation using SLIC [\(Achanta and Shaji, 2012\)](#);
- (2) Super-pixel clustering using LSSC [\(Xie et al., 2013\)](#);
- (3) Super-pixel feature extraction using [Eq. \(2\)](#);
- (4) Construction of the feature matrix for each super-pixel cluster using [Eq. \(3\)](#);
- (5) Construction of the primitive saliency dictionary using [Eq. \(9\)](#);
- (6) ULRR decomposition on the super-pixel cluster features using [Eq. \(6\)](#);
- (7) Saliency measure for super-pixel cluster using [Eq. \(13\)](#)
- (8) Saliency measure for each super-pixel using [Eq. \(14\)](#);

(9) Construction of pixel-level saliency map using Eqs. (17) and (18).

(10) Smoothed pixel-level saliency map by using Eq. (19).

End

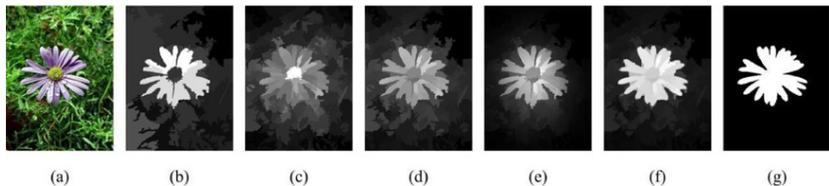


Fig. 5 Fig-5: Illustration of the results obtained by different components of our proposed method. (a) Test image. (b) Result by using the representation coefficient based saliency measure, i.e., Eq. (13). (c) Result obtained by using the reconstruction error based saliency measure, i.e. Eq. (14). (d) Result after fusing (b) and (c), i.e., Eq. (17). (e) Result by performing the center prior on (d), i.e. Eq. (18). (f). Result smooth by using the Max-Flow method. (g). Ground truth.

alt-text: Fig 5

Fig. 5 illustrates the detected results obtained by different steps in the proposed method. As can be seen in Fig. 5(b), the majority of the salient objects could be detected using the representation coefficient based measure only. Meanwhile, several local regions within these salient objects, shown in Fig. 5(c), could be better detected using the reconstruction error based saliency measure. Our idea fusing the two saliency maps allows the whole objects to be extracted completely. After performing the center prior on these saliency maps, the salient objects are further highlighted. In addition, the background noise is well suppressed. This can be viewed in Fig. 5(d) and Fig-5(e), respectively.

3.3 Computational complexity analysis

Closely looking at our system reveals that the LSSC based super-pixel clustering and the ULRR decomposition take up the most time. In this subsection, we theoretically investigate the computational complexity of this algorithmic part.

For LSSC, the major computation is the l_1 minimization in Eq. (A1), whose complexity is about $O(dN^2) + O(r_1N^2)$. Here, N denotes the number of super-pixels to be clustered. d refers to the dimension of each super-pixel feature vector. r_1 is the number of iterations in this step. Hence, the computational complexity of the ULRR model is about $O(d^2N) + O(r_2(d^2N + d^3))$ considering that the global dictionary \mathbf{D} in the proposed ULRR model is constructed by all of the super-pixels in the test image, and d is assumed to be $d \leq N$. Here, r_2 denotes the number of iterations during the ULRR decomposition. Accordingly, the computational complexity of the proposed method is about $O(d^2N) + O(r_1N^2) + O(r_2(d^2N + d^3))$. More specifically, the number of super-pixels N has a greater impact on the computational complexity of the clustering component and the dimension d has a greater impact on the computational complexity of the ULRR component.

4 Experiments and analysis

Several sets of experiments are performed to verify the feasibility of our proposed method (LSSC_ULRR, for short). First, we discuss the impacts of some parameters on the proposed method. Secondly, we test the proposed method on images with large-size salient objects or cluttered backgrounds to verify the claimed contribution. Thirdly, we show the superiority of the proposed method over some state-of-the-art methods using three public datasets. Finally, we show and analyze failure cases for the proposed method.

4.1 Impacts of different parameters

In this subsection, we investigate the effects of several key parameters used in our algorithm based on the MSRA1000 (Achanta and Hemami, 2009) dataset, including cluster numbers, number of potential foreground dictionary atoms and the parameter α in Eq. (17).

First, we discuss the impact of the number of super-pixel clusters K on system performance. Fig. 6 illustrates different detected results when varying the number of super-pixel clusters. For a better comparison, we also provide the detected results using the proposed method but without the clustering operation, i.e., directly performing the ULRR decomposition on feature matrix of the super-pixels. It can be obviously found that the number of clusters has a great impact on the detected results. As shown in the second row of Fig. 6(b), parts of the background and the salient object will be grouped into the same cluster and thus will be mistakenly labeled as the salient ones when K is set to too small value (e.g. $K = 5$). In contrast, when K is set to too large value (e.g. $K = 15$), the object will be segmented into more regions and each region will have different saliency values. This results in the non-uniform detection

results, as shown in the last row of Fig. 6(d). This is particularly true when the clustering operation is not considered during the saliency detection. In addition, some background noise will also be introduced as shown in Fig. 6(e). This is also consistent with the precision versus recall (PR) curves (Li and Hou, 2014), displayed in Fig. 7(a).

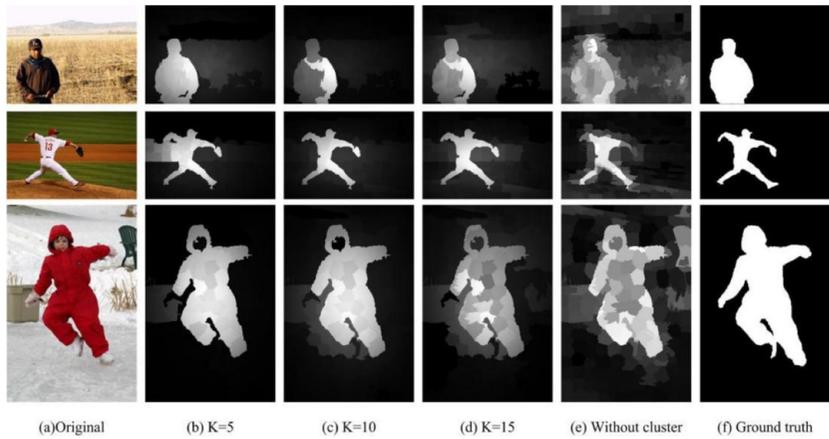


Fig. 6 Fig. 6: Illustrations of the results by using different cluster numbers.

alt-text: Fig 6

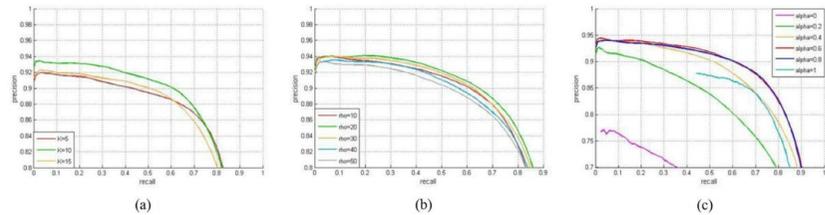


Fig. 7 Fig. 7: PR curves on MSRA1000 dataset by using different parameters. (a) Number of clusters K ; (b) Number of potential foreground dictionary atoms ρ ; (c) Parameter α in Eq. (17).

alt-text: Fig 7

In addition to the number of clusters K , the impacts of the number of potential foreground dictionary atoms ρ and the parameter α in Eq. (17) are also discussed here. Fig. 7(b) and Fig. 7(c) provide the PR curves on the MSRA 1000 dataset obtained using different values ρ and α , respectively. Fig. 7 indicates that the performance of the proposed method achieves the best when the parameters K , ρ and α are set to 10, 20 and 0.8, respectively. Therefore, in the following experiments, these parameters are set to 10, 20 and 0.8, respectively.

4.2 Validity of the proposed method on several types of images

To further verify the claimed contributions, in this subsection, we establish four sub-datasets, in which SO dataset is formed by a set of images with small objects, LO dataset consists of a set of images with large objects, MO dataset contains a set of images with multiple objects and CS dataset constitutes a set of images with complicated structures. Based on those four sub-datasets, we compare our LSSC_ULRR with some traditional LRMR based and clustering based methods, including SR_RPCA (Yan and Zhu, 2010), LRR (Lang and Liu, 2012), ULR (Shen and Wu, 2012) and LSSC_BS (Xie et al., 2013).

Figs. 8-11 illustrate the detected results on these images under different situations, respectively. All the results consistently demonstrate that the proposed method LSSC_ULRR performs the best among the five mentioned methods. In most of cases, it can uniformly detect the whole objects, and meanwhile, well suppress the noise from the background. In summary, the detected results obtained by our proposal are the most near to the ground truth.

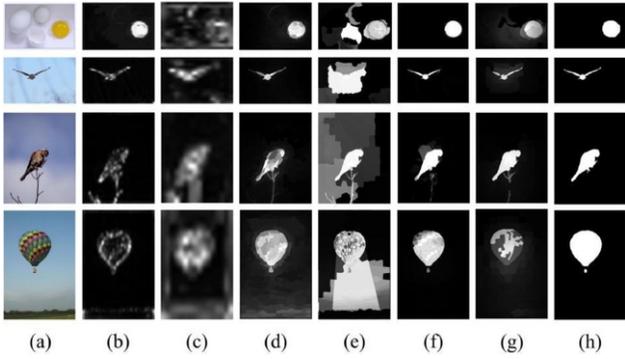


Fig. 8 Fig-8: Illustrations for images with small salient objects. (a) Original images; (b) SR_RPCA (Yan and Zhu, 2010); (c) LRR (Lang and Liu, 2012); (d) ULR (Shen and Wu, 2012); (e) LSSC_BS (Xie et al., 2013); (f) LSSC_ULRR; (g) Ground truth.

alt-text: Fig 8

More specifically, it can be found that all the five methods actually achieve satisfactory results for those images with small salient objects. Especially, ULR and the proposed LSSC_ULRR perform the best in this case. In addition, the two methods also perform better than the others for those images with multiple salient objects. This may be attributable to the usage of ULRR models in these methods. However, as shown in Fig. 9(e) and Fig. 9(f), only LSSC_BS and the proposed LSSC_ULRR perform the best for those images with large salient objects. This may be due to the fact that the two methods carry out the saliency detection on the super-pixel clusters rather than only the local block patches or super-pixels. Finally, it can also be found that the proposed LSSC_ULRR still works well for those images with complicated structures, as shown in Fig. 11(f). In contrast, the other methods unfortunately fail in this case.

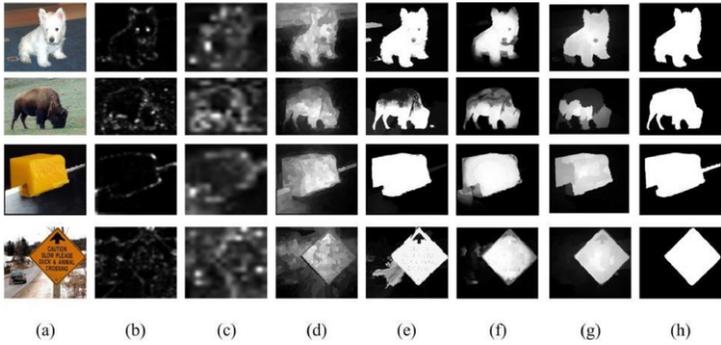
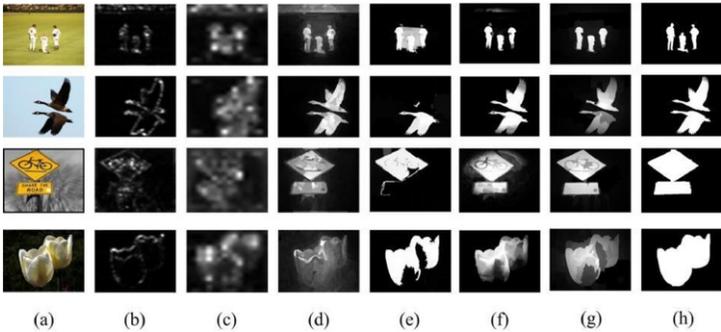


Fig. 9 Fig-9: Illustrations for images with large salient objects. (a) Original images; (b) SR_RPCA (Yan and Zhu, 2010); (c) LRR (Lang and Liu, 2012); (d) ULR (Shen and Wu, 2012); (e) LSSC_BS (Xie et al., 2013); (f) LSSC_ULRR; (g) Ground truth.

alt-text: Fig 9



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 10 Fig. 10: Illustrations for images with multiple salient objects. (a) Original images; (b) SR_RPCA (Yan and Zhu, 2010); (c) LRR (Lang and Liu, 2012); (d) ULR (Shen and Wu, 2012); (e) LSSC_BS (Xie et al., 2013); (f) LSSC_ULRR; (g) Ground truth.

alt-text: Fig 10

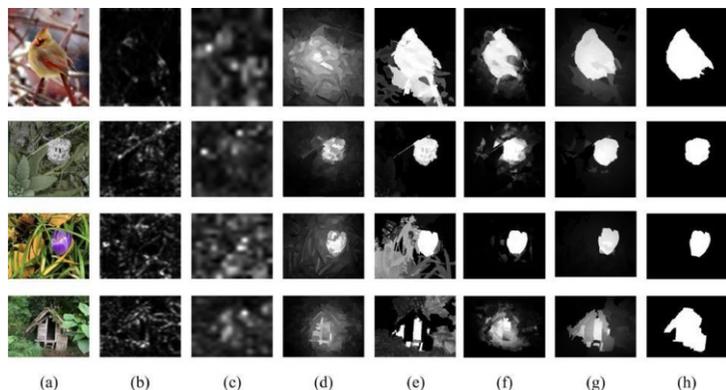


Fig. 11 Fig. 11: Illustrations for images with complicated structures. (a) Original images; (b) SR_RPCA (Yan and Zhu, 2010); (c) LRR (Lang and Liu, 2012); (d) ULR (Shen and Wu, 2012); (e) LSSC_BS (Xie et al., 2013); (f) LSSC_ULRR; (g) Ground truth.

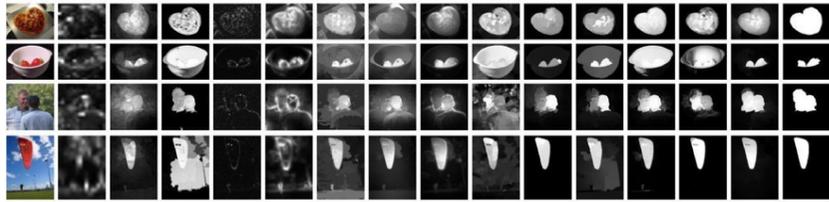
alt-text: Fig 11

4.3 Detection results on some public datasets

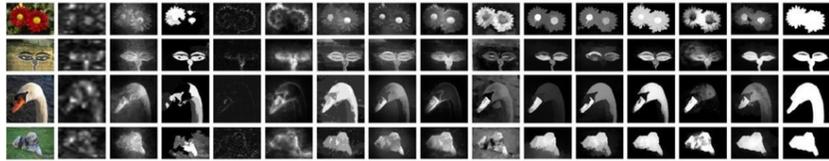
In this subsection, we will employ three public datasets, i.e., MSRA10K (Cheng and Mitra, 2015), ECSSD (Shi et al., 2016) and DUT-OMRON (Yang and Zhang, 2013) to thoroughly test the performance of the proposed method. The MSRA10K dataset contains 10,000 images, most of which have a single object and high contrast between foreground objects and backgrounds. The ECSSD dataset includes 1000 images, in which images are structurally complex and objects cover various categories. The DUT-OMRON dataset contains 5168 images and most of them either involve complex backgrounds or have high contrast with respect to the entire image. Apart from ULR (Shen and Wu, 2012) and LSSC_BS (Xie et al., 2013), some of up to date methods, including SR-LC (Huo and Yang, 2016), DSR (Lu and Li, 2016), RBD (Zhu and Liang, 2014), SF (Perazzi and Krahenbull, 2012), PCA (Margolin et al., 2013), CA (Goferman et al., 2012), SS (Hou et al., 2012), RC (Cheng and Mitra, 2015), DCLC (Zhou and Yang, 2015), and RW_MR (Liu and Cai, 2015), will be compared with our proposed method.

Fig. 12 illustrates the detected results obtained by different methods on the three public datasets. These results demonstrate most methods mentioned here can well detect the salient objects contained in the test images. Especially, DSR, RBD and the proposed LSSC_ULRR perform better than the other methods in most cases. The salient objects are more uniformly highlighted by the three methods. At the same time, the background noise is better suppressed by the three methods. In general, the detected results are closer to the ground truth.

MSRA10K



ECSSD



DUT-OMRON



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l) (m) (n) (o) (p)

Fig. 12 Fig. 12: Illustrations of the results generated by different methods on the three public datasets, i.e., MSRA10K, ECSSD, and DUT-OMRON. (a) Original images; (b) LRR (Lang and Liu, 2012); (c) ULR (Shen and Wu, 2012); (d) LSSC_BS (Xie et al., 2013); (e) SS (Hou et al., 2012); (f) CA (Goferman et al., 2012); (g) RC (Cheng and Mitra, 2015); (h) SF (Perazzi and Krahenbull, 2012); (i) PCA (Margolin et al., 2013); (j) SR-LC (Huo and Yang, 2016); (k) RW_MR (Liu and Cai, 2015); (l) DCLC (Zhou and Yang, 2015); (m) RBD (Zhu and Liang, 2014); (n) DSR (Lu and Li, 2016); (o) LSSC_ULRR; (p) Ground truth.

alt-text: Fig 12

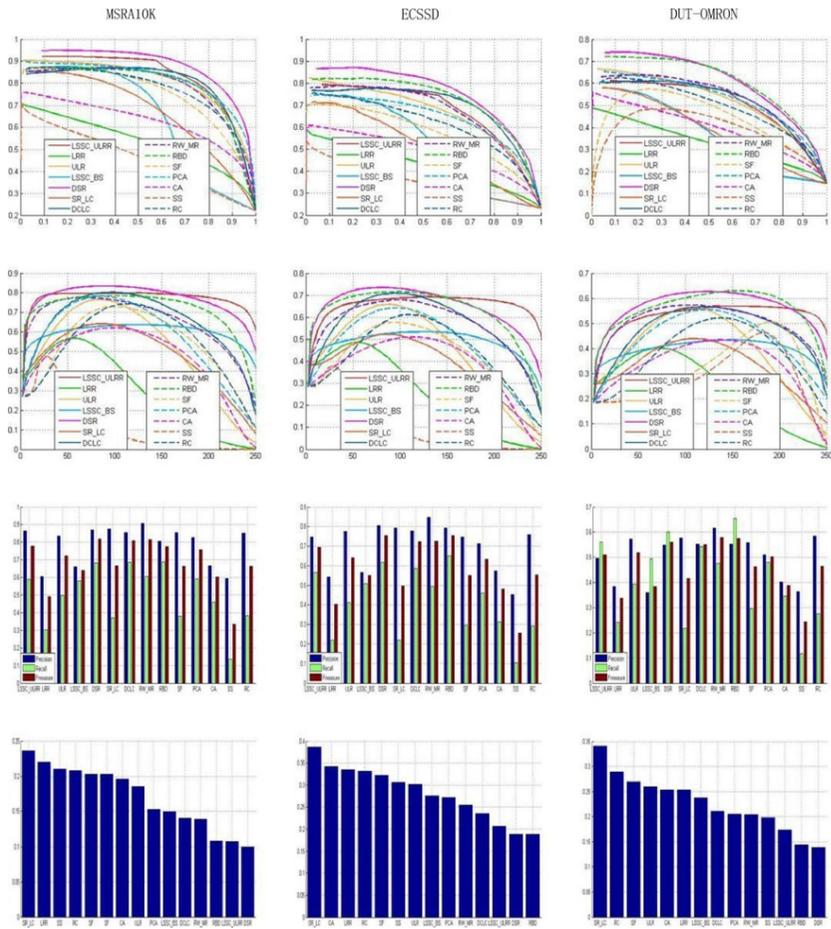


Fig. 13 Fig-13: Quantitative comparisons of different methods on the three public datasets.

alt-text: Fig 13

According to the comparison, the proposed LSSC_ULRR obtains better visual detected results than RBD and DSR in some cases. For example, in the second row for MSRA10K, LSSC_ULRR succeeds in separating the salient objects from background, while RBD and DSR fail. In the third row for ECSSD, RBD and DSR only detect some parts of the salient object, while LSSC_ULRR can detect the whole salient object. In the last row for DUT-OMRON, LSSC_ULRR obtains better suppression of the background noise than RBD and DSR do. Specifically, in the second rows for ECSSD and DUT-OMRON, some background regions are mistakenly labeled as salient regions by DSR, but these regions can be well detected by LSSC_ULRR.

In addition to visual comparison, the quantitative comparisons among different methods on the three datasets are also provided in Fig. 12, including the PR curves (Li and Hou, 2014), the F-measure curves (Li and Hou, 2014), average precision, recall, and F-measure bars (Li and Hou, 2014), and MAE bars (Li and Hou, 2014). Similar conclusions can be drawn according to these quantitative results. In most cases, RBD, DSR and the proposed LSSC_ULRR rank in the top three on the three datasets through comprehensive consideration of the four evaluation metrics. We also find that LSSC_ULRR performs better than RBD on MSRA10K. Especially, the proposed LSSC_ULRR gets the F-measure curve over a wide range for all the three public datasets, meaning that it gets good separation of background and foreground under all thresholds. In other words, the proposed LSSC_ULRR gets background suppressed and meanwhile makes foreground prominent.

As shown in the above quantitative results, DSR generally performs better than the proposed LSSC_ULRR method. However, for some cases, e.g., images with salient objects of greatly large sizes or images in which the salient

objects touch the image boundaries, the proposed LSSC_ULRR method performs better than DSR. For example, as shown in the first two columns of Fig. 14, LSSC_ULRR could obtain more uniform saliency maps for salient objects of greatly large sizes. This may be owing to the proposed cluster-based saliency measure. As shown in the last four columns of Fig. 14, the proposed LSSC_ULRR method could still detect the entire salient objects even if they touch the image boundaries. However, in this case, DSR just detected parts of the salient objects. This may be owing to the different dictionaries employed in the two methods.



Fig. 14 Fig. 14: Visual comparison. (a) Original images; (b) DSR (Lu and Li, 2016); (c) LSSC_ULRR; (d) Ground truth.

alt-text: Fig 14

However, we also found that the proposed method just achieved moderate performance among these mentioned methods in terms of the PR metric. As discussed above, some images in the DUT-OMRON dataset are too complex to be well clustered. The inaccurate clustering results thus degrade the final performance of the proposed method. Fig. 15 illustrates some failure cases of our proposed method. As shown in Fig. 15, the test images are extremely complicated and the clustering results for these images are very inaccurate. Subsequently, the salient objects are not well detected for these images. Exploiting a more effective clustering method may be desirable in this case. This is still a challenging problem, especially for those images with complex structures. We leave this for our future work.

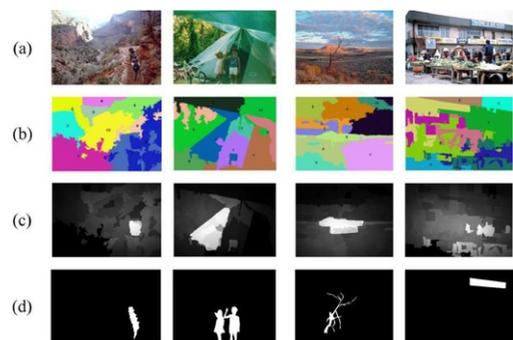


Fig. 15 Fig. 15: Failure cases. (a) Original images; (b) Cluster results; (c) Saliency maps; (d) Ground truth.

alt-text: Fig 15

5 Conclusion

In this paper, we propose a simple but effective salient object detection method based on LSSC and ULRR, in which the salient object detection is achieved via the saliency detection of super-pixel clusters. We first segment the input image into super-pixels and group them with LSSC. Then we formulate the saliency detection of the super-pixels as a unified low-rankness and sparsity pursuit problem by using ULRR. As well, the ULRR coefficients are employed to compute a global saliency measure for each super-pixel cluster with respect to the entire image, and the sparse reconstruction errors are used to construct a local saliency measure for each super-pixel with respect to each super-pixel cluster. Experimental results demonstrate the proposed method performs better than the traditional LRMR based and clustering based methods and is comparable to some current state-of-the-art methods. Especially, it can completely detect the whole salient object with large size in an image in most cases. And the detection results are very close to the ground-truth images. In addition, it can effectively suppress the noise from the backgrounds when the backgrounds of the input images contain multiple different textures.

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 A

In this appendix, we will briefly introduce a spectral clustering algorithm, i.e., Laplacian sparse subspace clustering method (LSSC), presented in [Xie et al., \(2013\)](#), which is used to group super-pixels in our proposed method.

For spectral clustering, one of the main issues is to construct an effective adjacency matrix that describes the similarity between each pairs of super-pixels accurately. In LSSC, a sparse similarity matrix is exploited for spectral clustering, which is motivated by the following two observations ([Cai et al., 2010](#)). One observation is that each data point in a union of subspaces is assumed to belong to a unique subspace and can be represented as a linear or affine combination of other points in the same subspace. Consequently, each point has a sparse representation when entire set of data points is considered. The second observation is that similar super-pixels should have similar sparse coefficients.

Given N points² $\{u_i \in R^d | i = 1, 2, \dots, N\}$ and the constraint (affinity) matrix $\mathbf{S} \in R^{N \times N}$, the sparse representation vector $c_i \in R^{N-1}$ for the point u_i is obtained by solving the following optimization problem

$$\min \|\mathbf{U}_{\hat{i}}^T c_i - \bar{u}_i\|_2 + \lambda_1 \|c_i\|_1 + \lambda_2 / 2 \sum_{ij} \|c_i - c_j\|_2^2 S_{ij} \quad s.t. c_i^T \mathbf{1} = 1, \quad (\text{A1A1})$$

where the basis matrix $\mathbf{U}_{\hat{i}} \in R^{d \times (N-1)}$ is obtained from the matrix $\mathbf{U} = [u_1, u_2, \dots, u_N] \in R^{d \times N}$ by removing the i -th column u_i . λ_1 and λ_2 are two positive trade-off parameters, and are experimentally set to 0.01 and 0.2, respectively. The (i, j) -th element $S_{i,j}$ in the constraint matrix S measures the similarity of the two super-pixels sp_i and sp_j . More details about the computation of the matrix S are seen in [Xie et al., \(2013\)](#). An N -dimensional vector $\hat{c}_i \in R^N$ is obtained by inserting a zero at the i -th row of c_i .

After obtaining the sparse representation vector for each point, a matrix $\mathbf{C} = [\hat{c}_1, \hat{c}_2, \dots, \hat{c}_N] \in R^{N \times N}$ and a corresponding symmetric similarity matrix $\tilde{\mathbf{C}} = (\mathbf{C} + \mathbf{C}^T) \in R^{N \times N}$ are thus constructed. With the matrix $\tilde{\mathbf{C}}$ as the adjacency matrix, a graph $\Upsilon = (V, E)$ is defined, where V are the N points, and $(v_i, v_j) \in E$ if $\tilde{\mathbf{C}}_{i,j}$ is non-zero. The Laplacian matrix \mathbf{A} of the graph Υ is thus formed as $\mathbf{A} = \mathbf{B} - \tilde{\mathbf{C}}$, where \mathbf{B} is a diagonal matrix with $\mathbf{B}_{ii} = \sum_j \tilde{\mathbf{C}}_{ij}$. The clustering result is finally obtained by applying the K -means algorithm to the eigenvector of the Laplacian matrix \mathbf{A} .

Appendix B

In this appendix, the update scheme required for solving [Eq. \(11\)](#) in the text is described in detail.

(1) Update \mathbf{J}_k ($k = 1, 2, \dots, K$)

$$\begin{aligned} \mathbf{J}_k^{j+1} &= \arg \min_{\mathbf{J}_k} \|\mathbf{J}_k\|_* + \left\langle \mathbf{W}_k^j, \mathbf{Z}_k^j - \mathbf{J}_k \right\rangle + \frac{\mu^j}{2} \|\mathbf{Z}_k^j - \mathbf{J}_k\|_F^2 \\ &= \arg \min_{\mathbf{J}_k} \frac{1}{\mu^j} \|\mathbf{J}_k\|_* + \frac{1}{2} \left\| \mathbf{J}_k - \left(\mathbf{Z}_k^j + \frac{1}{\mu^j} \mathbf{W}_k^j \right) \right\|_F^2. \end{aligned} \quad (\text{B1B1})$$

This sub-optimization problem has the following closed-form solution ([Wright et al., 2009](#)):

$$\mathbf{J}_k^{j+1} = SVT_{\frac{1}{\mu^j}} \left(\mathbf{Z}_k^j + \frac{1}{\mu^j} \mathbf{W}_k^j \right), \quad (\text{B2B2})$$

where $SVT_{\delta}(\boldsymbol{\Psi})$ denotes the Singular Value Thresholding (SVT) operation ([Wright et al., 2009](#)) on the matrix $\boldsymbol{\Psi}$ with the threshold δ .

(2) Update \mathbf{Z}_k ($k = 1, 2, \dots, K$)

$$\begin{aligned} \mathbf{Z}_k^{j+1} &= \arg \min_{\mathbf{Z}_k} \left\langle \mathbf{Y}_k^j, \mathbf{X}_k - \mathbf{DZ}_k - \mathbf{E}_k^j \right\rangle + \left\langle \mathbf{W}_k^j, \mathbf{Z}_k - \mathbf{J}_k^{j+1} \right\rangle \\ &\quad + \frac{\mu^j}{2} \left(\|\mathbf{X}_k - \mathbf{DZ}_k - \mathbf{E}_k^j\|_F^2 + \|\mathbf{Z}_k - \mathbf{J}_k^{j+1}\|_F^2 \right) \\ &= \arg \min_{\mathbf{Z}_k} \left\| \mathbf{X}_k - \mathbf{DZ}_k - \mathbf{E}_k^j + \frac{\mathbf{Y}_k^j}{\mu^j} \right\|_F^2 + \left\| \mathbf{Z}_k - \mathbf{J}_k^{j+1} + \frac{\mathbf{W}_k^j}{\mu^j} \right\|_F^2. \end{aligned} \quad (\text{B3B3})$$

This sub-optimization problem has the following closed-form solution:

$$\mathbf{Z}_k^{j+1} = (\mathbf{D}^T \mathbf{D} + \mathbf{I})^{-1} \left(\mathbf{D}^T (\mathbf{X}_k - \mathbf{E}_k^j) + \mathbf{J}_k^{j+1} + \frac{\mathbf{D}^T \mathbf{Y}_k^j - \mathbf{W}_k^j}{\mu^j} \right), \quad (\text{B4B4})$$

where \mathbf{I} denotes an identity matrix.

(B3) Update \mathbf{E}

$$\begin{aligned}
\mathbf{E}^{j+1} &= \arg \min_{\mathbf{E}} \lambda_1 \|\mathbf{E}\|_{2,1} + \lambda_2 \text{tr}(\mathbf{E}\mathbf{L}\mathbf{E}^T) \\
&\quad + \sum_{k=1}^K \left(\left\langle \mathbf{Y}_k^j, \mathbf{X}_k - \mathbf{D}\mathbf{Z}_k^{j+1} - \mathbf{E}_k \right\rangle + \frac{\mu^j}{2} \left\| \mathbf{X}_k - \mathbf{D}\mathbf{Z}_k^{j+1} - \mathbf{E}_k \right\|_F^2 \right) \\
&= \arg \min_{\mathbf{E}} \lambda_1 \|\mathbf{E}\|_{2,1} + \lambda_2 \text{tr}(\mathbf{E}\mathbf{L}\mathbf{E}^T) + \frac{\mu^j}{2} \|\mathbf{E} - \mathbf{G}\|_F^2 \\
&= \arg \min_{\mathbf{E}} \lambda_1 \|\mathbf{E}\|_{2,1} + f(\mathbf{E})
\end{aligned} \tag{B5B5}$$

where $f(\mathbf{E}) = \lambda_2 \text{tr}(\mathbf{E}\mathbf{L}\mathbf{E}^T) + \frac{\mu^j}{2} \|\mathbf{E} - \mathbf{G}\|_F^2$. \mathbf{G} is formed by horizontally concatenating $\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_K$ together along the row, i.e., $\mathbf{G} = [\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_K]$. And \mathbf{G}_k ($k = 1, 2, \dots, K$) is defined by $\mathbf{G}_k = \mathbf{X}_k - \mathbf{D}\mathbf{Z}_k^{j+1} + \frac{\mathbf{Y}_k^j}{\mu^j}$. To solve Eq. (B5), the quadratic term $f(\mathbf{E})$ is replaced by its first order approximation at the previous iteration by adding a proximal term (Wright et al., 2009), i.e.,

$$\begin{aligned}
\mathbf{E}^{j+1} &= \arg \min_{\mathbf{E}} \lambda_1 \|\mathbf{E}\|_{2,1} + \frac{\eta^j}{2} \|\mathbf{E} - \mathbf{E}^j\|_F^2 + \left\langle \nabla_{\mathbf{E}} f(\mathbf{E}^j), \mathbf{E} - \mathbf{E}^j \right\rangle \\
&= \arg \min_{\mathbf{E}} \frac{\lambda_1}{\eta^j} \|\mathbf{E}\|_{2,1} + \frac{1}{2} \left\| \mathbf{E} - \mathbf{E}^j + \frac{1}{\eta^j} \nabla_{\mathbf{E}} f(\mathbf{E}^j) \right\|_F^2,
\end{aligned} \tag{B6B6}$$

where η^j is set to $\eta^j = 1.02 (2\lambda_2 \|\mathbf{L}\|_F^2 + \mu^j)$. $\nabla_{\mathbf{E}} f(\mathbf{E}^j)$ is the partial differential of $f(\mathbf{E})$ with respect to \mathbf{E} , and is computed by

$$\nabla_{\mathbf{E}} f(\mathbf{E}^j) = 2\lambda_2 \mathbf{E}^j \mathbf{L} + \mu^j (\mathbf{E}^j - \mathbf{G}). \tag{B7B7}$$

Thus, the sub-optimization problem has the following closed-form solution (Liu and Lin, 2013):

$$\mathbf{E}^{j+1}(:, i) = \begin{cases} \frac{\left(\|\mathbf{Q}(:, i)\|_2 - \frac{\lambda_1}{\eta^j} \right)}{\|\mathbf{Q}(:, i)\|_2} \mathbf{Q}(:, i), & \text{if } \|\mathbf{Q}(:, i)\|_2 \geq \frac{\lambda_1}{\eta^j}, \\ 0, & \text{otherwise} \end{cases} \tag{B8B8}$$

where $\mathbf{Q} = \mathbf{E}^j - \frac{1}{\eta^j} \nabla_{\mathbf{E}} f(\mathbf{E}^j)$. $\mathbf{E}(:, i)$ and $\mathbf{Q}(:, i)$ denote the i -th column of the matrix \mathbf{E} and \mathbf{Q} , respectively

References

- Achanta R., Hemami S., et al., Frequency-tuned salient region detection, In: *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2009, 1597-1604.
- Achanta R., Shaji A., et al., SLIC superpixels compared to state-of-the-art superpixel methods, *IEEE Trans. Pattern Anal. Mach. Intell.* **34** (11), 2012, 2274-2282.
- Borkov Y. and Kolmogorov V., An experimental comparison of min-cut/max-flow algorithm for energy minimization in vision, *IEEE Trans. Pattern Anal. Mach. Intell.* **26** (9), 2004, 1124-1137.
- Bruce N. and Tsotsos J., Saliency based on information maximization, In: *Proceedings of the Advances in Neural Information Processing Systems*, 2005, 155-162.
- Cades E.J., Li X., et al., Robust principal component analysis?, *J. ACM* **58** (3), 2011, 11-50.
- Cades E.J. and Tao T., The power of convex relation: near optimal matrix completion, *IEEE Trans. Inf. Theory* **56** (5), 2009, 2053-2080.
- Cai J.F., Candes E. and Shen Z., A singular value thresholding algorithm for matrix completion, *SIAM J. Optim.* **20** (4), 2010, 1956-1982.
- Chakraborty S. and Mitra P., A dense subgraph based algorithm for compact salient image region detection, *Comput. Vision Image Understand.* **145**, 2016, 1-14.
- Chen T., Cheng M.M., et al., Sketch2Photo: internet image montage,, *ACM Trans. Graphics* **28** (5), 2009, 1-10.
- Cheng B., Liu G., et al., Multi-task low-rank affinity pursuit for image segmentation, In: *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2011, 2439-2446.
- Cheng M.M., Mitra N.J., et al., Global contrast based salient region detection, *IEEE Trans. Pattern Recog. Mach. Learn.* **37** (3), 2015, 569-582.
- Elhamifar E. and Vidal R., Sparse subspace clustering: algorithm, theory, and applications, *IEEE Trans. Pattern Anal. Mach. Intell.* **35** (11), 2013, 2765-2781.

Fouquier G., Atif J. and Bloch I., Sequential model-based segmentation and recognition of images structures driven by visual features and spatial relations, *Comput. Vision Image Understand.* **116**, 2012, 146-165.

Goferman S., Zelnik M.L. and Tal A., Context-aware saliency detection, *IEEE Trans. Pattern Anal. Mach. Intell.* **34** (10), 2012, 1915-1926.

Gong C., Tao D., et al., Saliency propagation from simple to difficult, In: *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2015, 2531-2539.

Guo C. and Zhang L., A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression, *IEEE Trans. Image Process.* **19** (1), 2010, 185-198.

Guo C., Ma Q. and Zhang L., Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform, In: *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2008, 1-8.

Han J., Pauwels J. and Zeeuw P., Fast saliency-aware multi-modality image fusion, *Neurocomputing* **111**, 2013, 70-80.

Harel J., Koch C. and Perona P., Graph-based visual saliency, In: *Proceedings of the Advances in Neural Information Processing Systems*, 2006, 545-552.

Hou X. and Zhang L., Saliency detection: a spectral residual approach, In: *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2007, 1-8.

Hou X., Harel J. and Koch C., Image signature: highlighting sparse salient regions, *IEEE Trans. Pattern Anal. Mach. Intell.* **34** (1), 2012, 194-201.

Huo L., Yang S., et al., Local graph regularized sparse reconstruction for salient object detection, *Neurocomputing* **194**, 2016, 348-359.

Itti L., Koch C. and Niebur E., A model of saliency-based visual attention for rapid scene analysis, *IEEE Trans. Pattern Anal. Mach. Intell.* **11** (20), 1998, 1254-1259.

Jiang H., Wang J., et al., Automatic salient object segmentation based on context and shape prior, In: *Proceedings of the British Machine Vision Conference*, 2011, 9-21.

Kim J., Han D. and Tai Y., Salient region detection via high-dimensional color transform and local spatial support, *IEEE Trans. Image Process.* **25** (1), 2016, 9-23.

Lang C., Liu G., et al., Saliency detection by multitask sparsity pursuit, *IEEE Trans. Image Process.* **21** (3), 2012, 1327-1338.

Li Y., Fu K., et al., Saliency detection based on extended boundary prior with foci of attention, In: *Proceedings of the IEEE International Conference on Speech and Signal Processing*, 2014, 2798-2802.

Li Y., Hou X., et al., The secrets of salient object segmentation, In: *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2014, 4321-4328.

Li J., Martin D., et al., Visual saliency based on scale-space analysis in frequency domain, *IEEE Trans. Pattern Anal. Mach. Intell.* **35** (4), 2013, 996-1010.

Lin Z., Liu R. and Su Z., Linearized alternating direction method with adaptive penalty for low-rank representation, *Adv. Neural Inf. Process. Syst.* 2011, 612-620.

Liu Y., Cai Q., et al., Saliency detection using two-stage scoring, In: *Proceedings of IEEE International Conference on Image Processing*, 2015, 4062-4066.

Liu N. and Han J., DHSNet: deep hierarchical saliency network for salient object detection, In: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2016, 678-686.

Liu G., Lin Z., et al., Robust recovery of subspace structures by low-rank representation, *IEEE Trans. Pattern Anal. Mach. Intell.* **35** (1), 2013, 171-184.

Liu X., Zhao G. and Yao J., Background subtractions based on low-rank and structured sparse decomposition, *IEEE Trans. Image Process.* **24** (8), 2015, 2502-2514.

Lu H., Li X., et al., Dense and sparse reconstruction error based saliency descriptor, *IEEE Trans. Image Process.* **25** (4), 2016, 1592-1603.

Ma Y.F. and Zhang H.J., Contrast-based image attention analysis by using fuzzy growing, In: *Proceedings of the 11th ACM International Conference on Multimedia*, 2003, 374-381.

Marchesotti L., Cifarelli C. and Csurka G., A framework for visual saliency detection with applications to image thumbnailing, In: *Proceedings of the IEEE International Conference on Computer Vision*, 2009, 2232-2239.

Margolin R., Tal A. and Zelnik-Manor L., What makes a patch distinct?, In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, 1139-1146.

Peng P., Shao L., et al., Saliency-aware image-to-class distances for image classification, *Neurocomputing* **166**, 2015, 337-345.

Perazzi F., Krahenbull P., et al., Saliency filters: contrast based filtering for salient region detection, In: *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2012, 733-740.

- Rigas I., Economou G. and Fotopoulos S., Efficient modeling of visible saliency based on local sparse representation and the use of hamming distance, *Comput. Vision Image Understand.* **143**, 2015, 33-45.
- Shen X. and Wu Y., A unified approach to salient object detection via low rank matrix recovery, In: *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2012, 853-860.
- Shi J., Yan Q. and Jia J., Hierarchical image saliency detection on extended CSSD, *IEEE Trans. Pattern Anal. Mach. Intell.* **38** (4), 2016, 717-729.
- Tong N., Lu H., et al., Salient object detection via bootstrap learning, In: *the Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2015, 1884-1892.
- Wan T., Zhu C. and Qin Z., Multifocus image fusion based on robust principal component analysis, *Pattern Recognit. Lett.* **34** (9), 2013, 1001-1008.
- Wang Q., Zheng W. and Piramuthu R., GraB: visual saliency via novel graph model and background priors, In: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2016, 535-543.
- Wei Y., Wen F., et al., Geodesic saliency using background priors, In: *Proceedings of the European Conference On Computer Vision*, 2012, 29-42.
- Wright S.J., Nowak R.D. and Figueiredo M.A.T., Sparse reconstruction by separable approximation, *IEEE Trans. Signal Process.* **57** (7), 2009, 2479-2493.
- Xie Y., Lu H. and Yang M.H., Bayesian saliency via low and mid level cues, *IEEE Trans. Image Process.* **22** (5), 2013, 1689-1698.
- Yan J., Zhu M., et al., Visual saliency detection via sparsity pursuit, *IEEE Signal Process Lett.* **17** (8), 2010, 739-742.
- Yang C., Zhang L., et al., Saliency detection via graph-based manifold ranking, In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2013, 3166-3173.
- Zhang T., Ghanem B., et al., Low rank sparse coding for image classification, In: *Proceedings of the IEEE International Conference on Computer Vision*, 2013, 281-288.
- Zhang D., Han J., et al., Cosaliency detection based on intrasaliency prior transfer and deep intersaliency mining, *IEEE Trans. Neural Netw. Learn. Syst.* **27** (6), 2016, 1163-1176.
- Zhang T., Liu S., et al., Robust visual tracking via consistent low-rank sparse learning, *Int. J. Comput. Vision* **111** (2), 2014, 171-190.
- Zhang Y., Jiang Z. and Davis L.S., Learning structured low-rank representation for image classification, In: *Proceedings of IEEE Computer Vision and Pattern Recognition*, 2013, 676-683.
- Zhou L., Yang Z., et al., Salient region detection via integrating diffusion-based compactness and local contrast, *IEEE Trans. Image Process.* **24** (11), 2015, 3308-3320.
- Zhu W., Liang S., et al., Saliency optimization from robust background detection, In: *Proceedings of the IEEE Conference on Computer Vision and Patter Recognition*, 2014, 2814-2821.

Footnotes

¹In this case, these super-pixels may be mistakenly grouped into a background super-pixel cluster. For example, the flower center region in [Fig. 3](#) is mistakenly grouped into a background super-pixel cluster denoted by the red color.

²Each point u_i may correspond to the feature vector of the i -th super-pixel sp_i in our proposed method.

Highlights

- A salient object detection method is proposed based on LSSC and ULRR.
 - The issue is converted to super-pixel cluster saliency detection using ULRR.
 - A primitive saliency dictionary is constructed for ULRR decomposition.
 - Representation coefficients and reconstruction errors are used in saliency measures.
 - The method works effectively for large-size objects and complex scenes.
-