ARTICLE IN PRESS
IMAVIS-03654; No of Pages 13
Image and Vision Computing xxx (2017) xxx–xxx
Contents lists available at ScienceDirect
Image and Vision Computing
journal homepage: www.elsevier.com/locate/imavis

Salient object detection employing robust sparse representation and local consistency☆

Liu Yi[a,b], Zhang Qiang[a,b,*], Han Jungong[c], Wang Long[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] School of Computing and Communications, Lancaster University, Lancaster, LA1 4YW, U.K.
[d] Center for Systems and Control, College of Engineering, Peking University, Beijing 100871, China

A R T I C L E I N F O

Article history:
Received 28 December 2016
Received in revised form 17 August 2017
Accepted 20 October 2017
Available online xxxx

Keywords:
Salient object detection
Robust sparse representation
Local consistency
Complex structures

A B S T R A C T

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 for those images with complex structures. 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 can be 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 based methods and is competitive with some current state-of-the-art methods, especially for those images with 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 [9–12] 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

☆ This paper has been recommended for acceptance by Nicu Sebe.

* Corresponding author at: P.O. Box 183, Department of Automatic Control, Xidian University, No.2 South TaiBai Road, Xi'an, Shaanxi Province 710071, China.
E-mail address: qzhang@xidian.edu.cn (Z. Qiang).

https://doi.org/10.1016/j.imavis.2017.10.002
0262-8856/© 2017 Elsevier B.V. All rights reserved.

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.

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 [9–11], 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 detection 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 neighboring 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 cannot 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 an $l_{2,1}$-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 encoded 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 of that, the whole salient object can be uniformly highlighted. More importantly, the background can also be well suppressed, as shown in Fig. 1 (e). Eventually, an objective 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

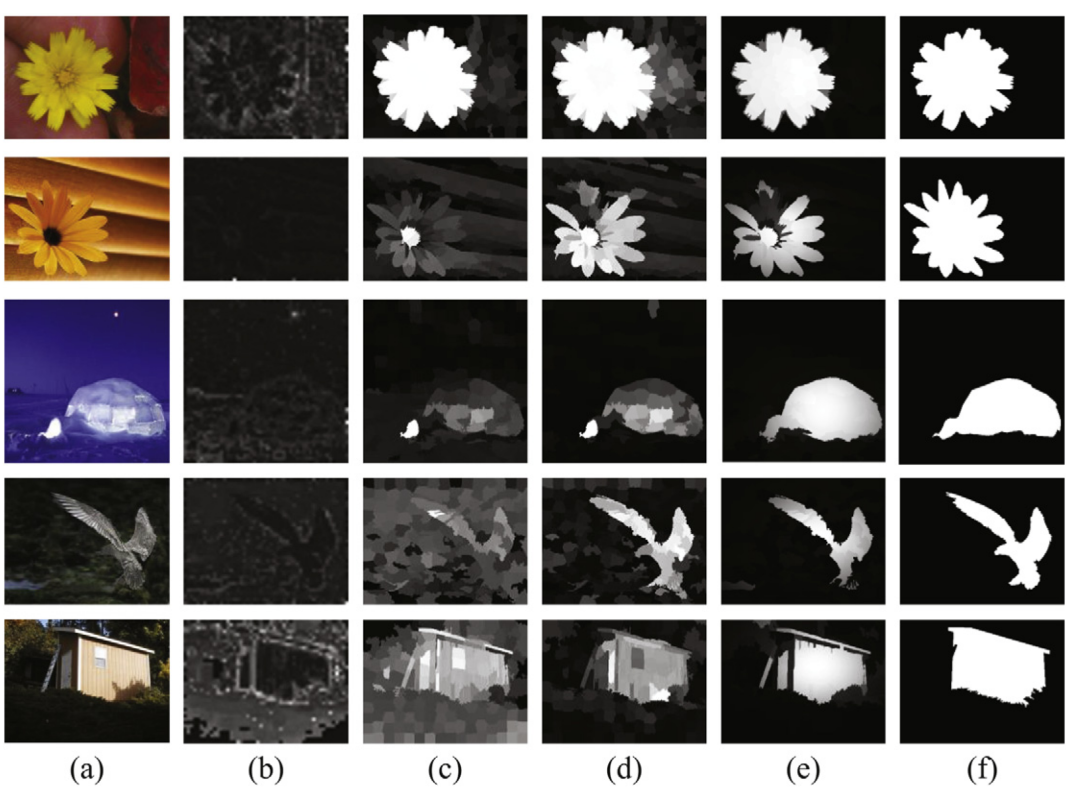

Fig. 1. Typical challenging examples for SR based salient object detection methods. (a) Original images; (b) SR based method with surrounding patches as the dictionary [10]; (c) SR based method with background templates near the image boundary as the background dictionary; (d) Proposed RSR method with 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.

Please cite this article as: L. Yi et al., Salient object detection employing robust sparse representation and local consistency, Image and Vision Computing (2017), https://doi.org/10.1016/j.imavis.2017.10.002