Pattern Recognition Letters 000 (2017) 1–9

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Saliency fusion via sparse and double low rank decomposition

Junxia Li[a,1,*], Jian Yang[b], Chen Gong[b], Qingshan Liu[a]

[a] B-DAT, School of Information and Control, Nanjing University of Information Science and Technology, Nanjing, 210044, China
[b] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Saliency detection
Saliency fusion
Low rank
Sparse noise

ABSTRACT

Video surveillance-oriented biometrics is a very challenging task and has tremendous significance to the security of public places. Saliency detection can support video surveillance systems by reducing redundant information and highlighting the critical regions, e.g., faces. Existing saliency detection models usually behave differently over an individual image, and meanwhile these methods often complement each other. This paper addresses the problem of fusing various saliency detection methods such that the fusion result outperforms each of the individual methods. A novel sparse and double low rank decomposition model (SDLRD) is proposed for such a purpose. Given an image described by multiple saliency maps, SDLRD uses a unified low rank assumption to characterize the object regions and background regions respectively. Furthermore, SDLRD depicts the noises covered on the whole image by a sparse matrix, based on the observation that the noises generally lie in a sparse subspace. After reducing the influence by noises, the correlations among object and background regions can be enhanced simultaneously. In this way, an image is represented as the combination of a sparse matrix plus two low rank matrices. As such, we cast the saliency fusion as a subspace decomposition problem and aim at inferring the low rank one that indicates the salient target. Experiments on five datasets demonstrate that our fusion method consistently outperforms each individual saliency method and other state-of-the-art saliency fusion approaches. Specifically, the proposed method is demonstrated to be effective on the applications of video-based biometrics such as face detection.

1. Introduction

Video surveillance-oriented biometrics has received intensive attentions in computer vision and machine learning for several decades. The main challenge is to develop and deploy reliable systems to detect, recognize and track moving objects, and further to interpret their activities and behaviors to meet the aim of increasing public security. With the rapid development of surveillance cameras, it is becoming more and more difficult for computers to handle the immense amount of video data. Particularly, the high-quality of video frames introduce a great deal of redundant spatial and temporal information which is time-consuming to handle, and there is no doubt that processing useless information deteriorates system performance.

Saliency detection, the task to detect objects attracted by the human visual system in an image or video, has attracted a lot of focused research in computer vision and has resulted in many applications, such as object detection, tracking and recognition, image/video retrieval, retargeting and compression, photo collage, video surveillance and so on. This paper aims to design an effective saliency fusion model to predict salient objects. Using saliency guides the video surveillance systems to reduce the search space for further processing and thus improve the computational efficiency of the whole system. As shown in Fig. 1, we can use the region covered by the red rectangular bounding box instead of the video frame for further object detection, recognition, tracking, *etc.*

With the goal both to achieve a comparable salience detection performance of human visual system and to facilitate various saliency-based applications, a rich number of saliency detection methods have been proposed in the past decade [2,6,16,18,19,27,28,37–39,44,46,51–53,57,60,64–66]. These approaches design a variety of models to simulate the visual attention mechanism or use data-driven methods to calculate a saliency map from an input image. Since different theories lead to different behaviors of saliency models, the saliency maps obtained by different approaches often vary remarkably from each other. Fig. 2 shows a few results produced by several representative saliency detection methods(*i.e.*, CA [27], HS [60], GC [15]). As

* Corresponding author.
E-mail address: junxiali99@163.com (J. Li).
[1] This work was done when the corresponding author was a Ph.D. student in the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China.

http://dx.doi.org/10.1016/j.patrec.2017.08.014
0167-8655/

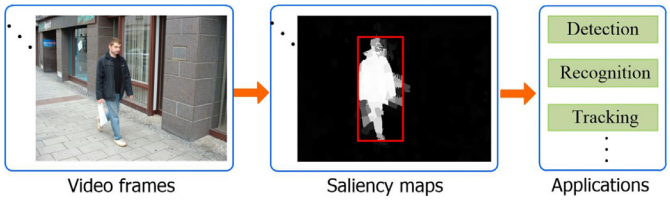

Fig. 1. Illustrating saliency's role of reducing the search space for further processing on the video surveillance systems.

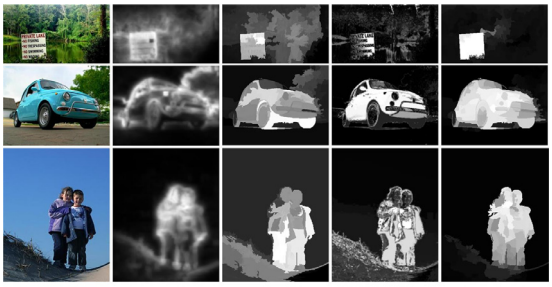

Fig. 2. Saliency fusion results. Individual saliency detection approaches often complement each other. Saliency fusion can effectively combine their results and perform better than each of them.

shown in Fig. 2(b) and (d), the object boundaries are well-defined, but some objects interiors are attenuated. Differently, the results shown in Fig. 2(c) highlight most of the object regions, but some background regions also stand out with the salient regions. Interestingly, these results often can complement each other. This motivates us to combine different saliency maps to achieve better results. Specifically, for a given image, we can first obtain various saliency maps by different saliency detection methods, and then try to find a way to utilize the advantages of these methods, aiming to effectively integrate these saliency maps.

By far, there are few methods attempting to fuse different saliency detection methods. Borji et al. [7] proposed a saliency fusion model using pre-defined combination functions. It treats each individual method equally in the fusion process. This simple strategy may not fully capture the advantages of each saliency detection approach. Mai et al. [48] use a conditional random field (CRF) to model the contribution from each saliency map. Although this method has been shown to be effective, the learnt CRF model parameters are somewhat biased toward the training dataset, due to which it suffers from limited adaptability.

The existing saliency fusion methods are often difficult to produce reliable results for images with diverse properties mainly due to the information contained across multiple saliency maps is not well utilized in the fusion process. To make use of such cross-saliency map information, in our previous work [40,41], we propose two low rank matrix recovery theory based saliency fusion methods, *i.e.*, the robust principle complement analysis (RPCA) model and the double low rank matrix recovery model (DLRMR). However, RPCA assumes that the image object has the sparsity property and hence it does not consider the correlation between object regions. Although DLRMR uses low rank constraint for the object and background regions respectively, it does not consider the noises covered the image in the saliency feature space, and thus leads to the poor robustness.

To address this problem, in this paper we propose a sparse and double low rank decomposition (SDLRD) model for saliency fusion. Fig. 3 is an intuitive illustration on our motivation. Given an image, if we first segment the original image into many homogeneous super-pixels, both object and background contain multiple

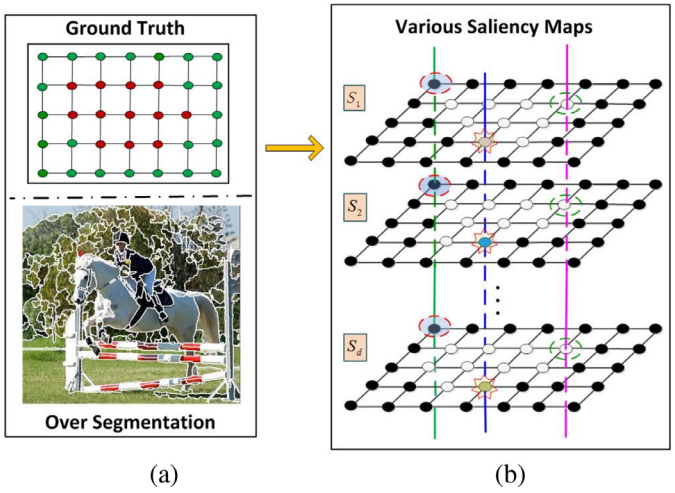

Fig. 3. An example to show the motivation of the proposed SDLRD model. (a) shows the over-segmentation of the original image and its simulated ground truth. In the ground truth image, super-pixels are represented by color nodes: red nodes denote object super-pixels and green ones represent background super-pixels. Clearly, both background and object contain multiple super-pixels. As shown in (b), in all the simulated saliency maps, white nodes denote the super-pixels that are with higher saliency values, while the black ones represent that the corresponding super-pixels are with lower saliency values. The nodes lying on the green (or the pink, blue) line show that they correspond to the same image super-pixel, and we drew a circle over the corresponding node for a visual discrimination. Moreover, there exist some super-pixels that are independent of background and object subspaces and can be considered as noises, *e.g.*, the super-pixel covered by the blue line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

super-pixels. For each super-pixel of object, the corresponding locations of the set of saliency maps are with high probability showing in brighter, indicating that they are with higher saliency values. With image regions being represented by the saliency values of multiple saliency maps, the object super-pixels are highly correlated and the corresponding feature vectors lie in a low-dimensional subspace. Meanwhile, most of background regions tend to have lower saliency values in various saliency maps. They are strongly correlated and lie in a low-dimensional subspace that is independent of the object subspace. Besides, in order to reduce the influence by noises and further to enhance the correlation among the object regions, we assume that the noises covered on the whole image lie in a sparse subspace and can be characterized by using a sparse matrix. Thus, an image can be represented as the combination of a sparse matrix plus two low rank matrices. SDLRD aims at inferring a unified low rank matrix that represents the salient objects. The inference process can be solved efficiently with the alternating direction method of multipliers (ADMM) [8]. Since the correlations within object regions as well as within background regions are well considered, SDLRD can produce more accurate and reliable results than previous saliency fusion models, and also can outperform the performance of each individual saliency detection method.

The contributions of our method mainly include:

1. Our method casts the saliency fusion as a subspace decomposition problem. It provides an interesting perspective for saliency fusion framework.
2. We propose a novel SDLRD model for saliency fusion. Theoretical analysis and experimental results demonstrate the feasibility and effectiveness of the presented method.
3. SDLRD well considers the cross-saliency map information. It performs better than the method which combines saliency maps through pre-defined combination functions.

Please cite this article as: J. Li et al., Saliency fusion via sparse and double low rank decomposition, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2017.08.014