Neural Networks 92 (2017) 47–59

Contents lists available at ScienceDirect

Neural Networks

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

Reprint of "Two-stage sparse coding of region covariance via Log-Euclidean kernels to detect saliency"

Ying-Ying Zhang [a,*], Cai Yang [b], Ping Zhang [a]

[a] Physics & Electronic Engineering College, Nanyang Normal University, Nanyang 473061, People's Republic of China

[b] Computer & Information Technology College, Nanyang Normal University, Nanyang 473061, People's Republic of China

ARTICLE INFO

Article history:
Received 17 March 2016
Received in revised form 5 February 2017
Accepted 13 February 2017
Available online 7 June 2017

Keywords:
Log-Euclidean kernel
Sparse coding
Region covariance
Saliency

ABSTRACT

In this paper, we present a novel bottom-up saliency detection algorithm from the perspective of covariance matrices on a Riemannian manifold. Each superpixel is described by a region covariance matrix on Riemannian Manifolds. We carry out a two-stage sparse coding scheme via Log-Euclidean kernels to extract salient objects efficiently. In the first stage, given background dictionary on image borders, sparse coding of each region covariance via Log-Euclidean kernels is performed. The reconstruction error on the background dictionary is regarded as the initial saliency of each superpixel. In the second stage, an improvement of the initial result is achieved by calculating reconstruction errors of the superpixels on foreground dictionary, which is extracted from the first stage saliency map. The sparse coding in the second stage is similar to the first stage, but is able to effectively highlight the salient objects uniformly from the background. Finally, three post-processing methods – highlight-inhibition function, context-based saliency weighting, and the graph cut – are adopted to further refine the saliency map. Experiments on four public benchmark datasets show that the proposed algorithm outperforms the state-of-the-art methods in terms of precision, recall and mean absolute error, and demonstrate the robustness and efficiency of the proposed method.

1. Introduction

Saliency detection is driven by the special phenomenon that a human visual system often pays more attention to some parts of an image. Saliency detection becomes much more popular in computer vision recently and is widely used in numerous applications, including image retrieval (Hiremath & Pujari, 2008; Jian, Dong, & Ma, 2011; Siagian & Itti, 2007), image resizing (Guo & Zhang, 2010), object detection (Han & Vasconcelos, 2009; Navalpakkam & Itti, 2006; Rutishauser, Walther, Koch, & Perona, 2004), adaptive image compression (Christopoulos, Skodras, & Ebrahimi, 2000) and so on.

Great achievement in saliency detection has been witnessed in recent years. The related work has been generally divided into two categories: top-down and bottom-up approaches. Bottom-up methods (Achanta, Estrada, Wils, & Süsstrunk, 2008; Achanta, Hemami, Estrada, & Susstrunk, 2009; Cheng, Zhang, Mitra, Huang, & Hu, 2011; Duan, Wu, Miao, Qing, & Fu, 2011; Itti, Koch, & Niebur, 1998; Perazzi, Krahenbuhl, Pritch, & Hornung, 2012; Rahtu, Kannala, Salo, & Heikkilä, 2010; Treisman & Gelade, 1980) are data driven and usually extract low-level features such as colors, edges and spatial distances. In contrast, as task-based methods, top-down methods (Alexe, Deselaers, & Ferrari, 2010; Marchesotti, Cifarelli, & Csurka, 2009; Yang & Yang, 2012) mainly involve supervised learning of more representative features from manually labeled ground truth. In addition, it is worth noticing that incorporating high level prior into saliency estimation has been previously investigated by a number of researchers. For example, some latest methods (Jiang et al., 2013; Li, Lu, Zhang, Ruan, & Yang, 2013; Wei, Wen, Zhu, & Sun, 2012; Yang, Zhang, Lu, Ruan, & Yang, 2013; Zhang, Wang, & Lv, 2016) exploit boundary priors, which indicated that regions along the image boundary are more likely to be the background. The boundary prior is proved to be more general than the center prior (Borji, Sihite, & Itti, 2012).

Despite the differences in these methods, they all use the vector feature to compute the saliency measure except Zhang et al. (2016). However, the vector feature considers features equally without considering the relationship of these different features, thereby it becomes invalid when there exist complex textures or various colors in the salient regions. Region covariance matrices naturally

DOI of original article: http://dx.doi.org/10.1016/j.neunet.2017.02.012.
* Corresponding author.
E-mail address: zyyzs226@126.com (Y.-Y. Zhang).

http://dx.doi.org/10.1016/j.neunet.2017.06.001
0893-6080/

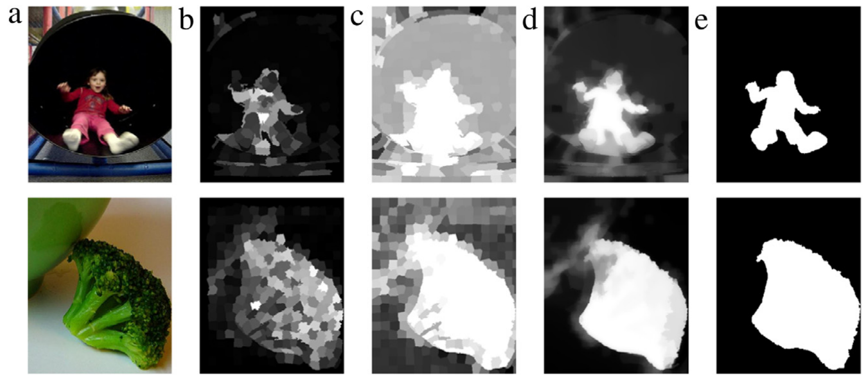

Fig. 1. Saliency maps generated by the proposed method. (a) Input; (b) the background-based saliency map; (c) the foreground-based saliency map; (d) the final saliency map; (e) ground truth.

provide nonlinear integration of different features by modeling their correlations and capture local image structures better than standard vector features (Tuzel, Porikl, & Meer, 2006). Our previous work in Zhang et al. (2016) has exploited sparse coding of covariance matrices for salient region detection. However, using the reconstruction error of background dictionary directly to measure saliency may produce unsatisfying results in complex scenes and influence the overall performance of the method. If the background-based saliency map is weak in highlighting the salient regions uniformly, the method is likely to fail as some salient foreground regions can be missed for the following post-processing steps. Therefore, we propose a two stage mechanism which considers employing foreground dictionary as hints of the foreground to address this problem. In Wang, Lu, Li, Tong, and Liu (2015), both background cues and foreground cues are also taken into consideration and the two saliency maps are integrated by the proposed unified function. However, the foreground cues-based result does not have complementary advantages by comparing with the background-based saliency and only contributes to the final integrated result. In our work, it is observed that the saliency maps based on foreground dictionary have some complementary advantage and disadvantage. The foreground-based saliency mode highlights the salient regions more uniformly and enhances the backgrounds of the saliency maps to some extent at the same time. So we formulate the background-based mode and the foreground-based saliency mode in turn and focus on making the most of the foreground-based saliency mode. We consider the background-based mode to get the rough saliency estimation and provide the foreground dictionary for the foreground-based saliency mode. After obtaining the foreground-based saliency result, we propose some refinements to overcome the disadvantage of the corresponding the foreground-based saliency result and develop the advantage, which is different from the refinements for the corresponding the background-based saliency result (Zhang et al., 2016) and the refinements for the integrated result (Wang et al., 2015). Fig. 1 shows some saliency results generated by the proposed method.

To this end, we put forward a cascade two-stage approach to incorporate sparse coding of region covariance matrices via Log-Euclidean kernels and some refinements. The main steps of the proposed method are shown in Fig. 2. In the first stage, we obtain the background dictionary based on the image border super-pixels in Fig. 2(a), which has been proved to be a good visual cue for background priors in saliency detection. We then calculate the reconstruction error of each super-pixel based on the background dictionary to get the background-based saliency map as shown in Fig. 1(b). But we find that the whole salient foreground region may not be deeply uniformly highlighted. So in the second stage, the foreground dictionary is built based on the first-stage saliency map. We use an adaptive threshold to segment the background-based saliency map and collect the super-pixels which are more likely to be saliency regions as the foreground dictionary in Fig. 2(b). Then the foreground-based saliency map is obtained by performing sparse coding of each super-pixel on the foreground dictionary. This computation of the second stage is similar to the first stage of background-based saliency detection, which can better emphasize the foreground against the background as shown in Fig. 1(c). While the whole salient foreground region is further uniformly highlighted, the background noise is more prominent to some extent than the background-based one. At last, three refinement techniques of highlight-inhibition function, context-based error weighting, and the graph cut are utilized to further refine the second stage result of foreground-based saliency map as shown in Fig. 1(d).

The main contributions of this paper are as follows:

1. We present a cascade two-stage scheme based on the sparse coding of the region covariance matrices via the Log-Euclidean kernels. The background-based mode is formulated to get the rough saliency estimation and followed by the foreground-based mode to overcome the problem of failing to highlight the salient regions uniformly with background-based one. The salient regions can be highlighted more uniformly based on the foreground-based saliency map;
2. We construct three refinement techniques to develop the advantage of the foreground-based saliency detection and overcome the disadvantage, which perform well in uniformly highlighting the salient objects and suppressing the backgrounds simultaneously.

2. Related work

Recently, more and more bottom-up methods have been proposed. The feature integration theory (Treisman & Gelade, 1980) points out that the saliency is dependent on how pixels stand out in various features. Itti–Koch model (Itti et al., 1998) represents the input image with the color, intensity and orientation features in different scales, and proposes the center–around difference mode to get the final saliency map by integrating three feature maps at different scales Since then, plenty of works have been devoted based on the center–around difference framework. Perazzi et al. (2012) measure saliency by fusing the uniqueness and spatial distribution of abstraction elements with the color feature and the position information. Cheng et al. (2011) evaluate global contrast differences and spatial weighted coherence scores simultaneously to obtain regional contrast. Duan et al. (2011) obtain the effective features of every block from the original RGB color features via one-dimensional PCA and then global contrast weighted by spatial