Neurocomputing 173 (2016) 406–417

Contents lists available at ScienceDirect

Neurocomputing

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

Saliency detection via a unified generative and discriminative model

Cong Jia, Jinqing Qi *, Xiaohui Li, Huchuan Lu

School of Information and Communication Engineering, Dalian University of Technology, 116024, PR China

ARTICLE INFO

Article history:
Received 20 June 2014
Received in revised form
2 February 2015
Accepted 14 March 2015
Available online 4 September 2015

Keywords:
Saliency detection
Generative model
Discriminative model
Sparse coding

ABSTRACT

In this paper, we propose a visual saliency detection algorithm which incorporates both generative and discriminative saliency models into a unified framework. First, we develop a generative model by defining image saliency as the sparse coding residual based on a learned background dictionary. Second, we introduce a discriminative model by solving an optimization problem that exploits the intrinsic relevance of similar regions for regressing region-based saliency to the smooth state. Third, a weighted sum of multi-scale region-level saliency is computed as the pixel-level saliency, which generates a more continuous and smooth result. Furthermore, object location is also utilized to suppress background noise, which acts as a vital prior for saliency detection. Experimental results show that the proposed algorithm generates more accurate saliency maps and performs favorably against the state-of-the-art saliency detection methods on three publicly available datasets.

1. Introduction

Humans have the capability of quickly localizing salient regions in complex scenes. The task of saliency detection is to identify the most salient pixels or regions in an image by simulating human visual system. In recent years, saliency detection gains much attention and plays an important role in computer vision and graphics applications, such as image segmentation [1,2], content-aware image retargeting [3], image thumbnailing [4], object recognition [5] and video compression [6]. Driven by these applications, salient object detection gives more emphasis on highlighting foreground objects, going far beyond its early goal of predicting human eye fixation on images. In this work, we focus on the salient object detection.

As visual information, images and videos are specially treated as visual big data. How to effectively reduce the dimensionality of data is the key of processing big data efficiently. Saliency detection, which is a branch of image processing, inevitably encounters this problem. In order to reduce dimensionality of data and accelerate the algorithm, most existing saliency models [7–10] merely choose the superpixel or patch instead of pixels as the minimum processing unit. Many other supervised/unsupervised dimensionality reduction algorithms are also explored and applied in computer vision, such as manifold learning [11], sparse representation [12,13] and kernel-based dimensionality reduction [14]. These dimensionality reduction techniques not only overcome the curse of dimensionality, but also save the computation and storage burden. In this work, we exploit the simple linear iterative clustering algorithm [15] and sparse representation to reduce dimensionality of data. And then, sparse residual is utilized to identify salient regions.

2. Related work

Recently many computational models have been proposed for saliency detection. Most of previous methods are generative models, which mainly focus on local and global contrast in images.

Local models measure saliency by various appearance contrast in the local neighborhood of a pixel or patch. Itti et al. [16] define image saliency as the center-surround contrast based on Difference of Gaussians. Liu et al. [17] integrate different regional contrast features into a Conditional Random Field (CRF) framework to detect salient objects. Ma and Zhang [18] and Jiang et al. [19] directly compute center-surrounding color difference in a specified spatial neighborhood. Rahtu et al. [20] compute the saliency likelihood of each pixel based on the center-surround contrast of a sliding window. The local spatial relations are also employed via graph construction for the graph-based saliency detection models [21]. Different from the local ones, global models aim to capture the holistic rarity from an image. Global saliency of the image is computed by the color difference from the average image color in the spatial domain [22], while [23] estimates saliency by the spectral residual in the frequency domain. Perazzi et al. [24] estimate global saliency of a patch by its spatial distribution in addition to color uniqueness, which is also employed as the regional

* Corresponding author. Tel./fax: +86 411 84708971.
E-mail address: jinqing@dlut.edu.cn (J. Qi).

http://dx.doi.org/10.1016/j.neucom.2015.03.122
0925-2312/

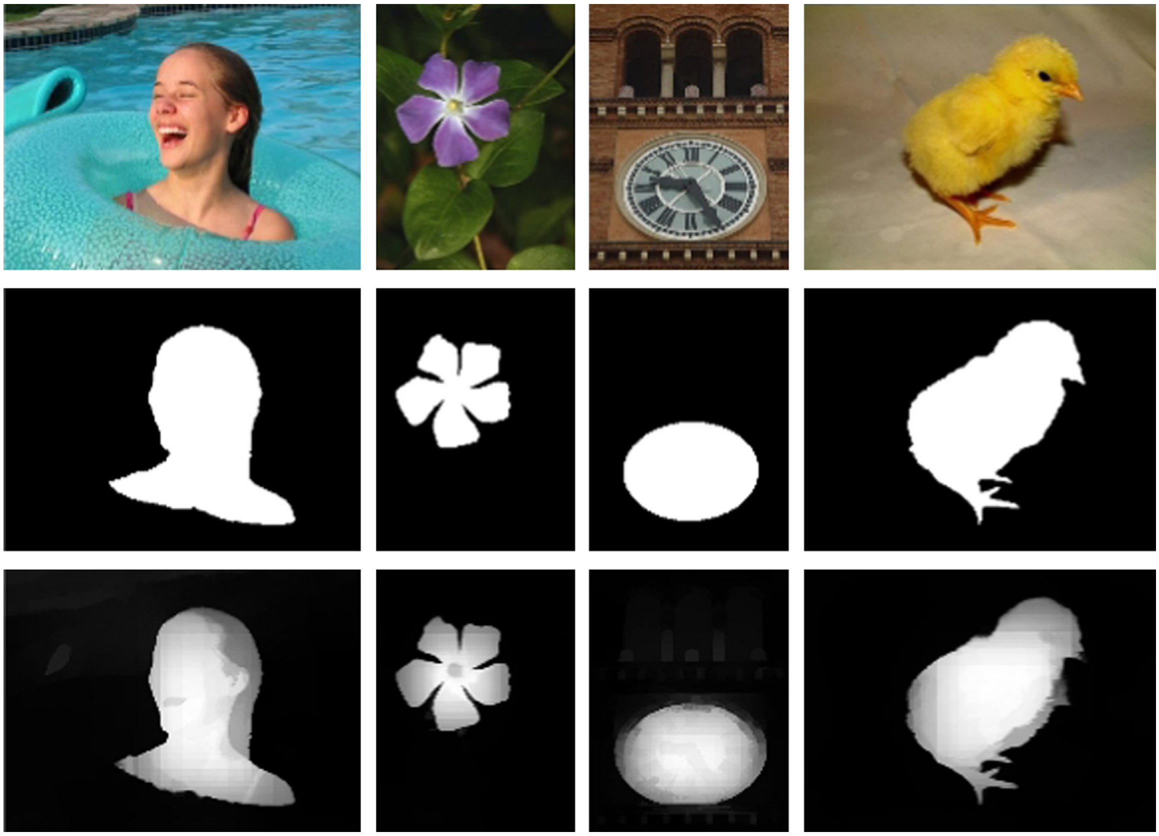

Fig. 1. The visual results of our algorithm. Top: the original images. Middle: the ground truth. Bottom: saliency maps computed by the proposed algorithm.

contrast in [25]. Achanta et al. [26] compute global saliency of each pixel based on its color contrast to the entire image. Goferman et al. [27] propose a context-aware saliency algorithm to detect the image regions that represent the scene based on four principles of human visual attention. Shi et al. [28] compute pixel-wise saliency by aggregating complementary appearance contrast measures with spatial priors. Margolin et al. [9] measure the pattern distinctness of patches using PCA and integrate it with color distinctness and high-level cues to achieve salient object detection. Cheng et al. [29] integrate two global saliency cues by automatically identifying which one is more likely to provide the correct identification of the salient region. Jiang et al. [30] introduce an absorbing Markov chain method where the appearance divergence and spatial distribution between salient objects and the background are considered. From the perspective of reconstruction, Li et al. [31] propose two saliency measures via dense and sparse reconstruction errors and integrate them into an effective Bayesian framework.

Generative models could effectively characterize the visual information for saliency detection. However, they still suffer from great difficulty of separating the salient objects from the background, due to the lack of discriminative prior knowledge. Therefore, the discriminative information is also of great importance to help enhance the difference between foreground and background for robust saliency detection. Several discriminative models have emerged in recent years. Wei et al. [8] measure the saliency of a local patch by its shortest distance to the image borders on a graphical model. Yan et al. [7] resort to hierarchical inference to fuse the multi-layer saliency based on a tree-structure graphical model. To consider the performance gaps and independence among different saliency analysis methods, Mai et al. [32] design various data-driven approaches to achieve individual method-aware and individual image-aware saliency aggregation. Shen and Wu [33] consider salient objects as sparse noises and formulate a low rank matrix recovery problem. Yang et al. [34] rank the similarity of image patches in the graph-based model. The discriminative model [35] regards saliency estimation as a regression problem and utilizes the supervised learning approach to map the regional features to saliency scores. Lu et al. [36] propose a method to learn optimal seeds for object saliency. Kim et al. [10] represent a saliency map of an image as a linear combination of different features in high-dimensional color space. Liu et al. [37] introduce a novel framework to adaptively learn a PDE system from the image for visual saliency detection. However, it is generally difficult for discriminative models to collect sufficient training samples or prior information, since test images processed individually by saliency detection, are of great diversity with multiple image categories.

In a word, generative model captures more expressive and generic descriptions of visual features, whereas discriminative model enforces larger inter-class (i.e., foreground and background) distance due to the useful prior knowledge. However on one hand, a generative model may be less effective in drawing a distinction between the salient objects and background, especially for cluttered scenes. On the other hand, a discriminative model is more dependent on training samples that may be difficult to be collected.

In this work, we incorporate generative and discriminative saliency models into a unified framework to take both advantages of them. In our generative saliency model, we measure saliency by sparse residual based on the background dictionary. Methods of saliency detection based on sparse representation are reviewed as follows. Li et al. [38] use the overlapping surround patches to sparsely reconstruct the center patch and define the ℓ_0 norm of the sparse code as saliency. Sun et al. [39] compute saliency based on the difference-to-average approach using sparse representation