Pattern Recognition 47 (2014) 1731–1739

Contents lists available at ScienceDirect

Pattern Recognition

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

# A novel approach to combine features for salient object detection using constrained particle swarm optimization

Navjot Singh *, Rinki Arya, R.K. Agrawal

*School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi 110067, India*

ARTICLE INFO

*Article history:*
Received 4 April 2013
Received in revised form
11 September 2013
Accepted 7 November 2013
Available online 20 November 2013

*Keywords:*
Salient object detection
Particle swarm optimization
Multi-objective function

ABSTRACT

Despite significant amount of research works, the best available visual attention models still lag far behind human performance in predicting salient object. In this paper, we present a novel approach to detect a salient object which involves two phases. In the first phase, three features such as multi-scale contrast, center-surround histogram and color spatial distribution are obtained as described in Liu et al. model. Constrained Particle Swarm Optimization is used in the second phase to determine an optimal weight vector to combine these features to obtain saliency map to distinguish a salient object from the image background. To achieve this, we defined a simple fitness function which highlights a salient object region with well-defined boundary and effectively suppresses the background regions in an image. The performance is evaluated both qualitatively and quantitatively on a publicly available dataset. Experimental results demonstrate that the proposed model outperforms existing state-of-the-art methods in terms of precision, recall, F -measure and area under curve.


## 1. Introduction

Salient object detection [1,2] is one of the key problems in computer vision, having received continuous attention since its birth. Visual saliency refers to the ability to locate the relevant information (object) in an image quickly and efficiently. The yield of the salient object detection process is a saliency map [1] where each pixel is assigned a measure of relevance [2]. This can be achieved by giving high score to the interesting information and low score to the irrelevant information.

Salient object detection provides fast solutions to many complex processes real-time applications such as surveillance systems [3] to track vehicle(s), pedestrian(s) or any object. It is also used in remote sensing [4] and image retrieval [5,6]. Additionally, it is used for automatic target detection such as finding traffic signs [1,7] along the road or military vehicles in a savanna [7], in robotics to find salient objects in the environment as navigation landmarks. It can also be applied in the area of image and video compression [7] by giving higher quality to salient objects at the expense of degrading background clutter, automatic cropping/centering [8] of images for display on small portable screens [9]. It also finds its applications in detecting tumors in mammograms [10], advertising a design [7], image collection browsing [11], image enhancement [12] and many more.

Several approaches have been suggested to model visual saliency based on neurobiological concepts, computational and mathematical methods. They can be broadly classified into two major categories [13]: bottom-up and top-down. In bottom-up models, multiple low-level visual features (such as intensity, color, orientation, and texture) are extracted from the image. Then these features are normalized and combined into a saliency map. Salient locations are identified using winner-take-all [1] and inhibition-of-return [1] operations. On the contrary, the top-down models are task-dependent and use a priori knowledge of the visual system. They are always integrated with the bottom-up models to generate saliency maps for localizing objects of interest.

Recently, Liu et al. [14] proposed a salient object detection model based on the combination of bottom-up and top-down approach [13]. It combined multi-scale contrast, center-surround histogram and color spatial distribution with conditional random field under maximum likelihood estimation (MLE) criteria. MLE is a well-known parameter estimation technique with many advantages [15]. It provides a consistent and asymptotically efficient approach for parameter estimation. It gives unbiased variance when sample size is large. It has approximate normal distributions and approximate sample variances that can be used to generate confidence bounds and hypothesis tests for the parameters, and has a lower variance in comparison to other methods. However, it is overshadowed by certain disadvantages: MLE can be heavily biased for small samples and is highly sensitive to the choice of

* Corresponding author. Tel.: +91 9650506400.
*E-mail addresses:* navjot.singh.09@gmail.com (N. Singh), rinki.arya89@gmail.com (R. Arya), rkajnu@gmail.com (R.K. Agrawal).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.11.012

starting values. MLE is a derivation based approach where the function should have an analytical form. Also, the solution does not converge all the time and is usually non-trivial for the numerical estimation. Also, Liu et al. [14] used a common linear weight vector, obtained by MLE, to combine the feature maps for all test images. This weight vector may not give better saliency results for images which are significantly different from the training set. The weight vectors for such images must be learned in such a way that they give better saliency results for their corresponding images.

In this paper, we used a modified form of Particle Swarm optimization (PSO) [16], a commonly used optimization method, to obtain weight vector in order to optimally combine the features extracted from the image. PSO utilizes the fitness function to obtain the optimal solution. For this we have proposed a new fitness function to obtain better saliency results. To check the efficacy of our proposed model, the performance is evaluated in terms of precision, recall, F -measure, area under curve and computation time. Experiments are carried out on a publicly available image dataset and performance is compared with Liu et al. [14] model and 10 other popular state-of-the-art models.

The paper is organized as follows. Section 2 includes the state-of-the-art methods to obtain visual salient object. The proposed model is discussed in Section 3. The experimental setup and results are included in Section 4. Conclusion and future work are presented in Section 5.

## 2. Related work

### 2.1. Bottom-up methods

Itti et al. [1] proposed a biologically plausible model that computes saliency map by combining intensity, color and orientation features at multiple scales. Walther and Koch [17] extended the Itti et al. model to detect proto object regions. Harel et al. [18] modeled the graph theoretic ideas to determine activation maps from the raw features. The model gives high saliency values to the nodes which are at the center of the image. Han et al. [19] integrated the Itti's model with Markov random field and region growing techniques to extract attention objects. Meur et al. [20] employed visibility, perception and perceptual grouping in their model. However, the model considered the achromatic structure in general and gave unclear boundaries. Bruce and Tsotsos [21] proposed a neurally plausible bottom-up overt attention model based on the principle of information maximization sampled from the scene. Yu and Wong [22] used a grid cell level real time clustering algorithm for image segmentation. It gives good initial centers of the clusters in lesser number of iterations but highly depends on the image segmentation accuracy. Achanta et al. [23] used an image subtraction technique to generate a frequency tuned saliency model. Cheng et al. [24] detects salient objects based on spatial coherence and global contrast.

### 2.2. Integration of top-down and bottom-up methods

Goferman et al. [25] presented a context-aware saliency detection algorithm based on four psychological principles- local low level, global, visual organization rules and high-level factors. Zhang et al. [13] utilized the position, area and intensity saliencies and a maximum saliency difference technique to classify a pixel into salient object or background object using Bayesian framework. Shen and Wu [26] incorporated the low-level features with the high-level knowledge to detect the salient object in the image represented as a low rank matrix and a sparse noise in some feature space. Liu et al. [14] used a supervised approach to separate the salient object from the image background. The features are extracted at local, regional and global level. The local feature is obtained by computing the contrast information of a pixel in a given neighborhood at different levels of details. The regional feature is made up of a center-surround histogram map. The global feature is represented in terms of a color spatial distribution map using Gaussian mixture models (GMM). To combine these features into a saliency map, linear weights are determined using conditional random field learning under maximum likelihood estimation criteria.

In Liu et al. [14] model, a common linear weight vector, determined using training images, is applied on all test images in order to combine the features into saliency maps. The above procedure is not only computationally expensive, but the obtained weight vector may not be appropriate for a given test image which is significantly different from training images. Hence there is a need to determine weight vector to combine the features for better saliency result for such an image. To achieve this, we used a constrained particle swarm optimization method (C-PSO) and proposed a new fitness function to obtain weight vector for the combination of features to improve saliency result.

## 3. Proposed salient object detection framework

Liu et al. [14] extracted multi-scale contrast, center-surround histogram and color spatial distribution feature maps from the image. These features can be combined in many ways to obtain a saliency map. One possible way is to give equal weightage to all the three features. However, there can be an image which is salient in terms of only singleton feature or combination of two features with different weights or combination of all the three features with different weights. So an appropriate weight vector is required in order to combine the features optimally for generating salient objects.

In this paper, this optimal weight vector for a given image is calculated using C-PSO. The C-PSO utilizes a fitness function, which in our case is given in terms of two components; one of the components is represented in terms of the attention pixels and the other in terms of the background pixels. For object to be salient, the fitness component obtained in terms of saliency values of attention pixels should be maximized and the other fitness component in terms of saliency values of background pixels be minimized. The optimal weight vector so determined is used to combine the features into the final saliency map. We describe C-PSO, mathematical formulation of the objective function and the steps involved to determine the saliency map in sub-sections given below.

### 3.1. Overview of constrained particle swarm optimization

Constrained Particle swarm optimization (C-PSO) is a modified form of Particle swarm optimization (PSO) [16]. PSO is a stochastic population-based evolutionary method inspired by the flocking behavior of birds [27]. The algorithm was developed by Kennedy and Eberhart in 1995 [16] for optimizing continuous multidimensional functions. It adopts an intelligent procedure to iteratively approach the desired optimal solution. The algorithm starts with a predefined number of swarms (or particles) in the population, also known as the candidate solutions. These particles are initialized with randomized positions and velocities. Each particle $i$ has a position $\mathbf{x}_i^d$ and a velocity $\mathbf{v}_i^d$ in the $D$-dimensional search space, where $d = 1, \ldots, D$. Each particle evaluates its fitness value. The movements of the particles are guided by their own best known position $\mathbf{pbest}_i$ in the search-space as well as the entire swarm's best known position $\mathbf{gbest}$. After few iterations of this process