Pattern Recognition Letters 72 (2016) 91–99

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Learning to rank salient segments extracted by multispectral Quantum Cuts☆

Çağlar Aytekin[a,*], Serkan Kiranyaz[b], Moncef Gabbouj[a]

[a] Tampere University of Technology, Department of Signal Processing, FI-33720 Tampere, Finland
[b] Qatar University, Department of Electrical Engineering, P.O. Box 2713 Doha, Qatar

ARTICLE INFO

Article history:
Available online 25 December 2015

Keywords:
Quantum Cuts
Saliency detection
Learning to rank
Salient object segmentation
Multispectral analysis

ABSTRACT

In this paper, a learn-to-rank algorithm is proposed and applied over the segment pool of salient objects generated by an extension of the unsupervised Quantum-Cuts algorithm. Quantum Cuts is extended in a multiresolution approach as follows. First, superpixels are extracted from the input image using the simple linear iterative k-means algorithm; second, a scale space decomposition is applied prior to Quantum Cuts in order to capture salient details at different scales; and third, multispectral approach is followed to generate multiple proposals instead of a single proposal as in Quantum Cuts.

The proposed learn-to-rank algorithm is then applied to these multiple proposals in order to select the most appropriate one. Shape and appearance features are extracted from the proposed segments and regressed with respect to a given confidence measure resulting in a ranked list of proposals. This ranking yields consistent improvements in an extensive collection of benchmark datasets containing around 18k images. Our analysis on the random forest regression models that are trained on different datasets shows that, although these datasets are of quite different characteristics, a model trained in the most complex dataset consistently provides performance improvements in all the other datasets, hence yielding robust salient object segmentation with a significant performance gap compared to the competing methods.

1. Introduction

Salient object extraction is a well-studied research area with rapidly growing interest. The motivation of this interest is due to the extensive applications of this particular topic such as, video surveillance [1,2], saliency based compression [3], image manipulation [4] and image cropping [5], to name a few.

Visual saliency literature can be divided into two: eye-gaze prediction and salient object extraction. Eye-gaze prediction is an effort to estimate the sparse-set of points that humans tend to show attention in a scene [6–12]. Salient object extraction, on the other hand, focuses on the problem of detecting an object which is appealing to the eye, as a whole [13–27].

In this paper, we will focus on the salient object extraction track. There are several cues that are extensively used for this task such as contrast cue which assumes that a salient object should exhibit local [16] and/or global contrast [18]. The boundary cue [23, 27] on the other hand assumes that a salient object is likely to differentiate from the image boundary pixels in appearance. Other priors include assumptions about the location and shape of the object [28].

Recently, several successful methods were proposed exploiting these cues such as: Graph-based manifold ranking (GBM) [25], absorbing Markov chain (MARKOV) [26], geodesic saliency (GS) [23], saliency filters (SF) [19], and robust background detection (RBD) [27]. In [19], first, the image is represented as a set of superpixels, and then global contrast cue is exploited by fusing a global element uniqueness measure and the spatial distribution of the elements across the image. The approach doesn't make use of the background cue, and exploits the local contrast only up to a certain extent. The study in [23] treats the image boundary as definite background and uses a geodesic-distance based saliency measured simply by assigning high saliency scores to regions that have high geodesic distance from the image boundaries. The main drawback of this work is the fact that the algorithm relies heavily on the boundary cue and exploits the other cues in a weaker way. In [25], the dissimilarity of the image compared to all four boundaries of the image was detected by graph based manifolding and merged. Finally, this fused map was refined by recalculating the similarity of the image regions according to the high-scored regions in this map. This approach too, mainly relies on the boundary cue and the refinement step can only contribute to the contrast cue up to a

☆ This paper has been recommended for acceptance by Denis Laurendeau

* Corresponding author. Tel.: +358 41 478 6220; fax: +358 3 364 1352.
E-mail address: caglar.aytekin@tut.fi (Ç. Aytekin).

http://dx.doi.org/10.1016/j.patrec.2015.12.005
0167-8655/

certain extent allowed by the first step. In [26], a Markov Chain based saliency model was used. Boundaries of the image were defined as absorbing nodes in a Markov-Chain and the saliency estimation is based on the absorption times of transient nodes. The objects that are touching the image boundaries were given specific attention in [27]. The authors made the assumption that the salient object touching the boundaries are more likely to have a small boundary length to area ratio. By this assumption foreground regions that are touching the image boundary were detected with a considerable success. Next, a measure of global contrast and difference from the boundaries that correspond to estimated background were used together to find a saliency map. The main drawback of [23, 25] and [26] is that they focus only on the background cue and do not strongly exploit the local and global contrast cues. On the other hand, [27] has integrated the global contrast with the background cue. However, the local contrast cue, which may be the strongest cue in visual saliency and object extraction [29], remains weakly exploited.

Recently in [31] we have proposed a novel automatic saliency map generation method, the Quantum-Cuts (QCUT). QCUT can strongly exploit background and contrast cues in a cascaded manner, i.e., first we have defined an absolute background (the image boundaries) and searched for the optimal region that is strongly separated from its surrounding. The contrast cue was considered as minimizing the cost of cutting the region from the rest of the image. Furthermore, within the saliency optimization, the area of the region that is cut was also maximized. The solution of this joint optimization is based on the foundations of quantum mechanics, which proved to be useful in several other computer vision algorithms [30].

Although QCUT achieved a considerable accuracy over a wide selection of datasets, there is still a lot of room for improvements. Since QCUT is an unsupervised method, erroneous results can occur due to the lack of knowledge of ‘objectness’, which can be obtained only by learning. However, there are indeed very few works on learning to detect salient objects [22,13,33]. These methods follow an approach to exploit learning during the saliency map extraction. However, the results remain poor when compared to more generic approaches. We believe that the main reason for this outcome is that, in both approaches, the objective was to learn the visual saliency of certain over-segmented regions [13] or pixels [22]. However, a definition of saliency is meaningful when the salient region is concerned all-together. In other words, combining several salient regions or patches does not necessarily correspond to a large salient object.

To overcome this issue, in this study, we propose a different approach. We first extract a number of salient object proposals by a multispectral analysis of QCUT. Next, these salient object proposals are ranked according to a learn-to-rank model. Hence, instead of integrating learning to the process of bottom-up saliency extraction, we use it as a ranking step. Limiting the search space to the set of proposals also provides a certain extent of generality since this set is constructed by a generic criteria. Indeed, we observe that this approach can even provide a performance improvement on the salient object extraction for a (test) dataset that has quite different characteristics than that of the training dataset. In order to increase the quality of the segment pool, several extensions are provided on top of QCUT. These extensions were briefly discussed in [32]. In this paper, we provide a detailed presentation, analysis and interpretation of the different extensions proposed. The extensions include a superpixel abstraction followed by a multi-resolution approach, a double-neighborhood approach and a novel affinity measure between graph nodes. We call this novel method multiresolution QCUT (M-CUT). By extensive experiments, we prove that with these improvements, we achieve a saliency representation that exceeds the performance of competing methods. By incorporating learn-to-rank step, we achieve even better results with significant performance gap to other methods.

The rest of the paper is organized as follows: the extensions to the existing Quantum Cuts algorithm are explained in Section 2. In Section 3, the learning methodology to rank salient segments approach is explained in detail. In Section 4, both visual and numerical experimental results are presented and the performance of the M-QCUT and proposed ranking method is analyzed in detail. Finally, Section 5 concludes the paper and suggests topics for future research.

2. Improvements on Quantum Cut

2.1. Quantum Cuts

In an earlier work, [31], we have discovered an interesting and useful link between quantum mechanics and graph theory. We observed that the ground state solution of a quantum mechanics energy operator and the energy minimization of the spectral graph-cut problems exhibit a high degree of similarity. Moreover, the unique concepts in quantum mechanics such as potential field and Hadamard product of wave-functions bring additional useful aspects to spectral graph-cut algorithms which are observed to be advantageous for certain computer vision problems. With this motivation, we have investigated an automatic saliency map generation and salient object segmentation application exploiting this link. The built-in regulations of quantum mechanics yields Quantum-Cuts (QCUT) which is an unsupervised and parameter-free saliency map generation method.

The motivation of QCUT is to reveal how the concepts emerged from quantum mechanics can contribute to visual data processing. The starting point is the well-known, time independent Schrödinger's equation:

$$-\frac{\hbar^2}{2m}\nabla^2\boldsymbol{\psi} = -\boldsymbol{V}\boldsymbol{\psi} + E\boldsymbol{\psi} \tag{1}$$

This equation is a unique representation of the relationship between the Planck's constant $\hbar$, a particle's wave representation $\boldsymbol{\psi}$ (wave-function), the potential field acting on this particle $\boldsymbol{V}$, its energy E and its mass m. We have first mapped this representation into the image domain where the discrete space is defined by the image pixels. The discrete form of the Schrödinger's equation is an eigenvalue problem: $\boldsymbol{H}\boldsymbol{\psi} = E\boldsymbol{\psi}$, where $\boldsymbol{H}$ represents a discrete representation of the Laplacian operator given in Eq. (2).

$$\mathbf{H}(i,j) = \begin{cases} \mathbf{V}(i) + |N_i|\dfrac{\hbar^2}{2m} & i = j \\ -\dfrac{\hbar^2}{2m} & j \in N_i \\ 0 & e.w \end{cases} \tag{2}$$

The main discovery of [31] starts by modifying $\boldsymbol{H}$ to a new operator, $\boldsymbol{H_m}$ to represent a weighted Laplacian where the weights correspond to the color similarities of the image pixels as in Eq. (3).

$$\boldsymbol{H_m}(i,j) = \begin{cases} \boldsymbol{V}(i) + \dfrac{\hbar^2}{2m}\sum\limits_{k\in N_i} w_{i,k} & i = j \\ -\dfrac{\hbar^2}{2m}w_{i,j} & j \in N_i \\ 0 & e.w \end{cases} \tag{3}$$

In Eq. (3), $w_{i,j}$ is the edge weight between nodes i and j, the similarity measure of two nodes in terms of their LAB color values. Note that it is possible to write $\boldsymbol{H_m} = \boldsymbol{D} - \boldsymbol{W} + \boldsymbol{V}$, where $\boldsymbol{D} - \boldsymbol{W}$ is the well-known graph Laplacian matrix.

It is discovered that the Hadamard product of the eigenvector of $\boldsymbol{H_m}$ corresponding to the minimum energy E, with itself, corresponds to the real-value approximation of a binary foreground