Pattern Recognition Letters 39 (2014) 31–38

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Constraining image object search by multi-scale spectral residue analysis

Grimaldo Silva *, Leizer Schnitman, Luciano Oliveira

Intelligent Vision Research Lab, Federal University of Bahia, Brazil

ARTICLE INFO

Article history:
Available online 5 September 2013

Communicated by S. Sarkar

Keywords:
Fast object detection
Saliency
Multi-scale spectral residue

ABSTRACT

Using an object detector over a whole image can require significant processing time. This is so since the majority of the images, in common scenarios, is composed of non-trivial amounts of background information, such as sky, ground and water. To alleviate this computational load, image search space reduction methods can make the detection procedure focus on more distinctive image regions. In this sense, we propose here the use of saliency information to organize regions based on their probability of containing objects. The proposed method was grounded on a multi-scale spectral residue (MSR) analysis for saliency detection. For better search space reduction, our method enables fine control of search scale, presents more robustness to variations on saliency intensity along an object length, and also a straightforward way to control the balance between search space reduction and false negatives, both being a consequence of region selection. MSR was capable of making object detection three to five times faster compared to the same detector without MSR. A thorough analysis was accomplished to demonstrate the effectiveness of the proposed method using a custom LabelMe dataset of person images, and also a Pascal VOC 2007 dataset, containing several distinct object classes.

1. Introduction

Searching for objects in images can require a substantial amount of time. Although the object of interest will be usually found in only a fraction of the search space, for most real-life systems, much time is dedicated processing recurring background information, such as: sky, earth, water, walls and roads. Many techniques were developed to alleviate the cost of object detection. Some techniques are based on strategies, which, in general, are able to discard non-object regions with a fraction of the runtime cost when compared to processing regions containing the object of interest (Viola and Jones, 2001; Zhu et al., 2006); other approaches rely on contextual information to calculate "regions of interest" where certain object classes are normally found (Wolf and Bileschi, 2006; Perko and Leonardis, 2007), as when searching for cars in roads.

Our hypothesis is that saliency can be used to fast and effectively filter out image regions which do not contain objects, concentrating the use of object detectors only in regions of more interest. In other words, saliency methods can search an image for regions that "stand out". To achieve this goal, several approaches can be used: Local contrast methods define saliency based on a given region contrast in relation to its local neighborhood (Itti et al., 1998; Harel et al., 2007); approaches based on the rarity principle (or global contrast) evaluate how unlike a given region is when compared to the entire image (Achanta et al., 2009; Cheng et al., 2011); other methods rely on information extracted from the frequency domain, such as the Spectral Residue (SR) (Hou and Zhang, 2007).

In Silva et al. (2012), we introduced a method, named multi-scale spectral residue (MSR), which discards image regions considered non-important, by means of saliency detection. That approach differs from context-based region selection, such as Wolf and Bileschi (2006), in its independence from prior object-specific training, once saliency captures information solely based on immediate image information. Furthermore, MSR does not require re-training existing object detectors as required by Zhu et al. (2006). For region selection, our method borrows some core concepts of SR saliency method (Hou and Zhang, 2007), although modifications have been made in the original SR to achieve better search space reduction. The SR method was chosen given its speed, region selection robustness for different scenes (snow, ocean, forest, mountain), and fine control over the size of the searched salient object. In this context, this work extends on our previous one in Silva et al. (2012), by providing a more in-depth description of our method, as well as more experimental analyzes concerning various aspects of MSR performance. Also here, MSR is improved over region selection performance by using different contrast normalization approaches that

* Corresponding author. Tel.: +55 718825 9224; fax: +55 713283 9755.
E-mail addresses: jgrimaldo@ufba.br (G. Silva), leizer@ufba.br (L. Schnitman), lrebouca@ufba.br (L. Oliveira).

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.08.025

allows for a better detection of an object saliency. To provide evidence of MSR applicability, a comparison of different object detectors performance with and without MSR over a custom LabelMe dataset is carried out. Additionally, the impact of our method on runtime speed was measured, and a detailed analysis of region selection performance with over eighteen object classes from Pascal VOC dataset is presented. In this sense, MSR showed evidences to perform best when used to find animals (dog, horse and cat), while over man-made items such as chairs and TV monitors, it did not help so much the actual object detector.

Main contribution: MSR provides a novel way to reduce image search space without incurring in overall runtime speed overhead. The negligible runtime speed of saliency computation does not represent great impact when compared to the number of regions discarded, making the proposed method offer a net gain of processing speed. The proposed method may also reduce false positives by discarding regions before a given object detector would incorrectly regard them as an object. Particularly, this work here provides a more in-depth explanation of MSR, with respect to Silva et al. (2012), also improving its region selection performance by using different contrast normalization techniques.

1.1. Outline of MSR

Our method, coined as multi-scale spectral residue (MSR) after since (Silva et al., 2012), achieves image search space reduction by organizing regions based on their visual importance. To accomplish it, MSR relies on a saliency detection to judge whether image regions require further inspection (by an object detector), or can be safely discarded, avoiding further processing. While borrowing some key ideas from spectral residual (SR) (Hou and Zhang, 2007) saliency detection method, our approach fine tunes saliency detection for region selection over multiple scales. An overview of our approach is presented in Fig. 1.

In step (1) of Fig. 1, the current octave is scaled specifically for saliency detection using the resizing factor β (the rationale for this decision is described in depth in Section 3.2). After β scaling, a contrast normalization is applied to better enhance objects in relation to the background (see Section 3.1). In step (2), the saliency map is calculated using SR, which is described, for completeness, in Section 2. Using this saliency map, in step (3), a sliding window is applied, and, for each unique window position, a quality function is evaluated in step (4), which attempts to select candidate regions by measuring how strong is saliency in that region (see Section 3.3 for more details). Finally, in step (5), each selected window feeds an object detector to be finally labeled. MSR supports most sliding window-based detectors. In this work, however, our proposed method is evaluated using just HOG/SVM (Dalal and Triggs, 2005) and Haar-like/Adaboost (Viola and Jones, 2001) to detect people in images (see Section 5 for experimental evaluations).

The gist of MSR is analyzing a region saliency intensity before object detection. This provides an opportunity to discard regions before further processing is accomplished. Discarding regions before an object detector is used can provide a significant speed-up in image processing.

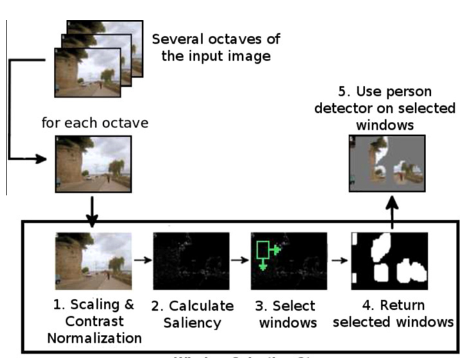

Fig. 1. Overview of MSR. For each image octave (1), the image is specifically scaled for saliency detection using a factor β (detailed in Section 3.2), and then contrast normalized (Section 3.1); in (2), the saliency map is computed (Section 3.2); in (3) and (4), a sliding window is applied to the saliency map, and candidate windows are selected based on a score return by a quality function (Section 3.3). In (5), an object detector is applied only in the candidate windows (for instance, a HOG/SVM person detector) (Section 5 discusses evaluations of other object detectors).

2. Spectral residual analysis

The Fourier transform allows an image to be represented in the frequency domain. This transform gives spectral information that can be used to search for characteristics that are recurrent in salient regions. As properties common to salient objects are hard to conceptualize, the Spectral Residual (Hou and Zhang, 2007) explores properties of the background with respect to the object. For that, the $1/f$ law states that over an ensemble of images, the Fourier spectrum will obey the distribution

$$E\{A(f)\} \propto 1/f, \tag{1}$$

where $E\{A(f)\}$ is the average amplitude over frequency f, in the ensemble of images. As the $1/f$ law is likely not to hold in individual images, parts of the spectrum that generate larger differences between the expected and the actual distribution are likely to contain novel information. These regions with larger differences were interpreted by Hou and Zhang (2007), as composing possible objects in an image. To calculate such differences, represented by $\mathcal{R}$, the frequency domain is then separated into amplitude and phase, in which the following formulation was used:

$$\mathcal{R} = log(A(f)) - h_m * log(A(f)), \tag{2}$$

where A is the amplitude of the Fourier domain, and h_m is a 2D convolution filter of size $m \times m$. This formulation intends to capture regions containing statistical singularities, which jump out of the expected $1/f$ distribution. The residue $\mathcal{R}$ is then re-combined with phase information $P(f)$. In this combination, the inverse Fourier transform, $\mathcal{F}^{-1}$, is applied to generate the final saliency map, S_{SR}, defined as

$$S_{SR}(x,y) = g(x,y) * \mathcal{F}^{-1}[\exp(i \cdot \mathcal{P}(f) + \mathcal{R})]^2, \tag{3}$$

where $g(x,y)$ is a 2D Gaussian filter. From the generated saliency map, salient pixels are selected based on whether they are greater than three times the mean intensity over the saliency map, this value was found based on empirical analysis of images containing salient objects.

3. Multi-scale spectral residue

When searching for image objects, by using a sliding window approach, each unique window position is evaluated by the object detector. However, objects of interest normally appear in only a small fraction of these windows. In this sense, an expensive object detection is applied without consideration to the probability of a region containing the object of interest. To tackle this problem, saliency detection methods can be used in order to cast most likely image regions to find an object.

Saliency detection methods usually associate a measure of global or local importance to a image region. Information about a region importance can be used to choose whether an object detection will be applied in a given window. Discarding windows before detection can reduce the time required to process an image. Yet,