Signal Processing 110 (2015) 37–45

Contents lists available at ScienceDirect

Signal Processing

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

Flexible sliding windows with adaptive pixel strides

Xiaoheng Jiang [a], Yanwei Pang [a,*], Jing Pan [b], Xuelong Li [c]

[a] School of Electronic Information Engineering, Tianjin University, Tianjin 300072, PR China
[b] School of Electronic Engineering, Tianjin University of Technology and Education, Tianjin 300222, PR China
[c] Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, PR China

ARTICLE INFO

Article history:
Received 11 March 2014
Received in revised form
19 June 2014
Accepted 3 August 2014
Available online 23 August 2014

Keywords:
Object detection
Sliding window
Adaptive strides
Sparse sampling

ABSTRACT

Detecting humans, faces, and other objects is important for pose estimation, behaviour analysis, and recognition. Most of the existing detection methods exploit sliding-window technique that typically evaluates image patches at uniform grids, with constant pixel strides along horizontal and vertical directions in an image. This approach can lead to excessively redundant computation since target object instances are relatively rare events compared to non-objects of the background. Therefore, most of the computation is unnecessarily wasted on the background. In this paper, to deal with this problem, we propose to scan the image using adaptive instead of fixed (unchanged) strides. That is, a high response of current sliding window generates small strides for the subsequent sliding windows in both horizontal and vertical directions, which means an object instance is likely to exist around. Otherwise, a low response gives large pixel strides and results in sparse sampling on the background. As a result, the areas of potential object instances are densely sampled and the areas of non-objects are sparsely sampled, leading to the reduction of total computation cost without loss of detection accuracy. Experimental results on face detection demonstrate the high efficiency and effectiveness of the proposed method.

1. Introduction

Object detection is a key problem in computer vision [11,17,30]. An object detection system typically consists of three stages: image patch generation, feature extraction, and feature classification. Our work mainly focuses on the first part since the number of image patches directly determines the detection speed. Nowadays, the emergence of portable devices makes the demand for real-time object detection increasingly urgent. We intend to speed up the object detector by decreasing the number of image patches. Suppose the total number of image patches is N, and the cost of feature generation and classification for one image patch is denoted by c_f and c_c, respectively. And then the total cost of detecting object instances in an image is denoted by

$$c = N \times (c_f + c_c) \tag{1}$$

There are many promising object detection systems that build on different combinations of features and classifiers. However, the most widely used image patch generation technique adopted by these systems is sliding window approach [5,16,22]. This approach exhaustively searches image patches at all possible positions and sizes. For each layer of an image pyramid with scale stride Δs, the number of sliding windows (image patches) is determined by

$$N_{SW} = \frac{w}{\Delta x} \cdot \frac{h}{\Delta y} \tag{2}$$

where w and h are the width and the height of an image layer, respectively, and Δx and Δy are the pixel strides.

* Corresponding author.
E-mail addresses: jiangxiaoheng@tju.edu.cn (X. Jiang), pyw@tju.edu.cn (Y. Pang), jingpan23@gmail.com (J. Pan), xuelong_li@opt.ac.cn (X. Li).

http://dx.doi.org/10.1016/j.sigpro.2014.08.004
0165-1684/

From Eq. (2) it is clear to see that the number of sliding windows is inversely proportional to pixel strides. The impact of pixel strides on the detection results is twofold: on the one hand, large pixel strides can reduce the amount of computation; on the other hand, large pixel strides can result in the decrease of detection accuracy. The problem has become a stumbling block preventing many promising object detection systems from being applied in real-time applications.

It appears difficult to step out of the above dilemma, but is it really impossible? Our answer is No. In this paper, we deal with this problem by means of adopting flexible sliding windows with adaptive (instead of unchanged) pixel strides. We call the proposed method as ASW (adaptive sliding window). ASW is able to decrease the number of sliding windows without degrading detection accuracy since it can intelligently grow the sliding strides at obviously non-object region, whereas shrink the strides at the possible object region. As a result, the image patches are densely sampled at potential object regions, but are sparsely sampled at non-object regions. Considering that object instance is relatively rare event in natural scene, the total number of image patches can be greatly reduced. According to Eq. (1), the decrease of N can reduce the total computation, resulting in a faster object detector. Furthermore, detection accuracy is guaranteed because ASW can intelligently evaluate more image patches at potential object regions. However, during traditional sliding window searching, the corresponding values for Δx and Δy of each layer of an image pyramid are constant. Constant pixel strides lead to that image patches are equally sampled both at regions containing potential object instances and regions of background. So equal sampling falls into the dilemma between detection speed and detection accuracy. In either case, it can result in a large waste of computation at non-object regions.

In summary, the main contributions of the paper are as follows: (1) We propose to generate image patches with flexible pixel strides in accord with the outputs of object detector on the former sliding windows. (2) The proposed ASW approach can promote many sliding-window based object detection systems, making it possible for them to be applied in real-time applications.

The rest of the paper is organized as follows: Section 2 reviews some related works. Section 3 presents the proposed ASW method in detail. Experimental results are reported in Section 4. Finally, conclusions are presented in Section 5.

2. Related work

Many researchers have made contributions to the progress of object detection [1–3,5–8,10,15,16,18,19,22–27,31]. When it comes to how to generate image patches, we generally cast the object detection methods into three categories.

The first class of methods utilizes sliding window technique [5,6,8,16,22] to generate image patches. Suppose the object detector is $\mathbf{g}$ and the feature extracted from current image patch is $\mathbf{x}$, and then the sliding window based decision function can be denoted by

$$\mathrm{sgn}(f(\mathbf{g},\mathbf{x},\Delta x,\Delta y,\Delta s)) = \begin{cases} 1 & \text{object} \\ 0 & \text{background} \end{cases} \tag{3}$$

where $\mathrm{sgn}(\cdot)$ is the sign function. Eq. (3) presents that image patches are sampled uniformly with scaling factor Δs in scale domain and corresponding constant pixel strides $(\Delta x, \Delta y)$ in space domain. The values for Δs, Δx and Δy are usually obtained through experimental results. It is easy to generate image patches using uniform sliding window approach. However, the tradeoff between detection speed and detection accuracy becomes the bottleneck. One can improve detection accuracy by using small pixel strides at the cost of detection speed. Or one can accelerate detection speed by using large pixel strides and sacrificing detection accuracy.

The second class of methods generates image patches by efficiently exploring the sub-window space through optimal solution algorithms [4,13,14,20]. Authors of [4] presented a principled model of visual research that simulates a digital fovea and scans the image so as to maximize the expected amount of information obtained about the location of the target. Efficient Subwindow Search (ESS) proposed in [14] is of high efficiency in generating image patches. The reason lies in the fact that it finds the global optimal window by hierarchically splitting the parameter spaces into disjoint spaces and using quality functions to reject large parts of the parameter space until the left space can no longer be split. However, ESS can find at most one optimal image patch at once, which is not convenient when lots of object instances exist.

Another image patch generation approach is multistage particle-window (MS-PW) proposed by Gualdi et al. [10]. This paradigm castes object detection into a statistical-based search using Monte Carlo sampling for estimating the distribution of target objects. In each stage, the proposal distribution is updated based on the outputs of the classifiers. As a result, sub-windows are sampled progressively closer to the target objects with high probability. As presented in their experiments, the total number of image patches needing to be evaluated is much fewer than that of traditional sliding window method. However, the detection performance depends not only on the number of stages employed for detection but also on the number of particle windows of each stage. This approach is prone to miss object instances if the total number of particle windows is not large enough. Moreover, compared with traditional sliding window method, the MS-PW approach spends much more time on estimating the distribution of target objects.

From the above analysis, we can find that traditional sliding window method generally generates uniform image patches and is easy to implement, whereas MS-PS approach generates fewer non-uniform image patches in a complex way. Inspired by these two methods, we conceive a new image patch generation method – ASW. ASW generates image patches with flexible pixel strides in accord with the outputs of object detector on the former sliding windows. In fact, there are already some study about adaptive object detection and adaptive image