Expert Systems With Applications 72 (2017) 130–138

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Personness estimation for real-time human detection on mobile devices

Kyuwon Kim[a,b], Changjae Oh[a], Kwanghoon Sohn[a,*]

[a] School of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, South Korea
[b] Mobile Division, Samsung Electronics, Suwon 16677, South Korea

ARTICLE INFO

Article history:
Received 31 August 2016
Revised 2 December 2016
Accepted 9 December 2016
Available online 15 December 2016

Keywords:
Human detection
Detection proposals
Normed gradients
Color attributes
Deformable part-based model

ABSTRACT

One aim of detection proposal methods is to reduce the computational overhead of object detection. However, most of the existing methods have significant computational overhead for real-time detection on mobile devices. A fast and accurate proposal method of human detection called personness estimation is proposed, which facilitates real-time human detection on mobile devices and can be effectively integrated into part-based detection, achieving high detection performance at a low computational cost. Our work is based on two observations: (i) normed gradients, which are designed for generic objectness estimation, effectively generate high-quality detection proposals for the person category; (ii) fusing the normed gradients with color attributes improves the performance of proposal generation for human detection. Thus, the candidate windows generated by the personness estimation will very likely contain human subjects. The human detection is then guided by the candidate windows, offering high detection performance even when the detection task terminates prior to completion. This interruptible detection scheme, called anytime detection, enables real-time human detection on mobile devices. Furthermore, we introduce a new evaluation methodology called time-recall curves to practically evaluate our approach. The applicability of our proposed method is demonstrated in extensive experiments on a publicly available dataset and a real mobile device, facilitating acquisition and enhancement of portrait photographs (e.g. selfie) on widespread mobile platforms.

1. Introduction

Vast numbers of pictures of people are captured and stored daily by mobile devices such as digital cameras and mobile phones. As a result, human detection on mobile devices has attracted significant research interest in recent years. Applications of human detection include human tracking, human segmentation for automatic *backlight compensation* and *selfie* enhancement (Kim, Oh, & Sohn, 2016). Since the introduction of the discriminatively trained part-based model by Felzenszwalb, Girshick, McAllester, and Ramanan (2010), the *deformable part model* (DPM) and its variants have become increasingly popular for human detection (Benenson, Omran, Hosang, & Schiele, 2014; Sadeghi & Forsyth, 2014). However, the practical applicability of DPM human detection is limited by the significant computational overhead on mobile devices.

DPM detectors construct a feature pyramid of multiscale feature maps and search each feature map through a sliding window. DPM detectors also require several mixture models describing various poses and viewpoints. Each mixture model contains one root filter representing the overall object shape at a low resolution and several part filters representing different object parts at a higher resolution. DPM improves the rate of object detection because of these sophisticated procedures and configurations. However, the filter scores computed by DPM require large computational resources because the sliding window approach performs many convolutions between the filters and the feature maps, e.g., for human detection in a (375×500)-pixel image, the OpenCV (Bradski & Kaebler, 2008) implementation of DPM constructs a feature pyramid with 33 scale levels and performs 1,786,962 convolution operations between this feature pyramid and 14 filters. Such convolution operations can dominate the total detection time (approximately 0.75 s or 53.47% of the total detection time on a regular PC).

In order to accelerate the object detection task, many researchers have optimized the DPM procedure by improving the algorithms and using hardware-specific features such as complex instructions and many GPUs on a desktop PC (Benenson, Mathias, Timofte, & Van Gool, 2012; Sadeghi & Forsyth, 2014). However, once the software is developed and submitted to certain

* Corresponding author. Fax: 82-2-2123-2879.
E-mail addresses: q1.kim@yonsei.ac.kr (K. Kim), ocj1211@yonsei.ac.kr (C. Oh), khsohn@yonsei.ac.kr (K. Sohn).

http://dx.doi.org/10.1016/j.eswa.2016.12.017

application stores, the algorithm can be executed on a variety of devices with different specifications. A more significant problem is that the processors in mobile devices are designed for low power consumption and lack high-performance CPUs with complex instructions or a sufficient number of GPU cores to boost the algorithm speed. Therefore, to implement DPM on mobile devices, the target objects should be searched from the most promising windows. Like other algorithms, detection algorithms executing on mobile platforms are time-constrained. Consequently, intensive detection algorithms will deteriorate the performance of the whole system and cause inconvenience to users. Considering the high-resolution imaging and hardware restrictions of mobile devices, the impracticality of an exhaustive sliding window search becomes obvious.

Detection proposal (or the *objectness measure*) has recently emerged as an alternative object detection technique (Hosang, Benenson, Dollar, & Schiele, 2015). A detection proposal method generates person windows that probably contain *generic* objects, avoiding exhaustive searching. Its intention to improve the detection speed appears to be perfectly matched with real-time detection. However, when our DPM implementation consumes approximately 200 ms searching over all multi-scale feature maps on a regular PC, most existing detection proposal methods consume more than 250 ms on the same device (Hosang et al., 2015).[1] In real-time detection, the time required for generating candidate windows at the preprocessing stage should be markedly less than the actual detection time. Therefore, the detection proposal method must be significantly faster than the exhaustive search time of real-time detection.

In the existing methods for detection proposals (Hosang et al., 2015), the generated candidate windows are *generic over categories*. Consequently, these methods extract object segments or well-defined boundaries by solving complex segmentation problems (Alexe, Deselaers, & Ferrari, 2012; Carreira & Sminchisescu, 2012; Chen, Ma, Wang, & Zhao, 2015; Humayun, Li, & Rehg, 2014; Manen, Guillaumin, & Van Gool, 2013; Uijlings, van de Sande, Gevers, & Smeulders, 2013) or by performing sophisticated edge detection (Krähenbühl & Koltun, 2014; Zitnick & Dollár, 2014). However, the computational overhead of exploring unseen categories is too high for real-time processing. Furthermore, a large number of windows are generated for all possible objects, which reduces the speed of the category-specific detectors in the latter stage. To resolve these problems and achieve real-time frame rates on mobile devices, we concentrate on categories that are relevant to the situation. When only person category is relevant, simultaneously considering all possible categories is a substantial waste of computational resources. Therefore, we propose a more efficient and accurate method that estimates person windows in an image, while ignoring category-agnostic candidate windows. The proposed method efficiently utilizes the simple color and edge features, as explained in Section 3. Therefore, our approach shares strong correlation with the human visual system in the sense that the human attentional mechanisms also preferentially notes simple features such as color and orientation when isolating possible candidates in distracting backgrounds (Wolfe & Horowitz, 2004). For convenience, we refer to 'objectness estimation for people' as *personness* estimation. Examples of human detection by personness estimation are presented in Fig. 1.

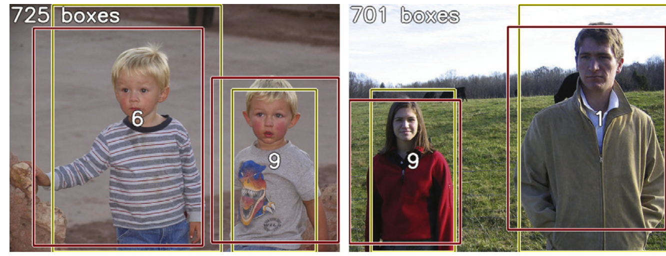

Fig. 1. Examples of DPM human detection by personness estimation in PASCAL VOC 2012 (Everingham et al., 2010) dataset. Yellow bounding boxes (BBs) are the best-matched windows generated in the personness estimation. The numbers in the center of each yellow window are the window indices. For example, in the left image, 725 candidate windows were generated, among which windows 6 and 9 were the best matched. Red BBs are the windows of the detected object in our DPM implementation aided by personness windows. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

1.1. Contributions

1. We present a fast and accurate personness estimation and demonstrate its effectiveness on a low-power mobile processor. The personness estimation rapidly captures the important edge and color features of the person category from the normed gradients (Cheng et al., 2014) and color attributes (Van De Weijer, Schmid, Verbeek, & Larlus, 2009). In this way, our approach generates a limited number of windows using the linear support vector machine (SVM). Evaluated on the person category of the PASCAL VOC dataset (Everingham, Van Gool, Williams, Winn, & Zisserman, 2010), the detection proposals generated by personness estimation allow the DPM detector to obtain more than 50% of its original performance within a 20 ms *window search* on a low-power mobile processor. The window search process includes both window generation and convolution calculation.

2. We show the improved use of detection proposals by the DPM detector. On mobile devices, much importance should be placed on *interruptible* object detection, or *anytime detection*, which yields reasonable results even before all tasks are complete (Karayev, Fritz, & Darrell, 2014; Sadeghi & Forsyth, 2014). To improve the anytime performance (Karayev et al., 2014), our DPM design efficiently computes the filter responses by imposing *time constraints* on the provided candidate windows. The DPM implementation also considers two important factors such as *aspect-ratio threshold* and *patch size* for *pinpoint* to achieve better detection performance using window proposals (see Section 3.4).

3. The detection proposal methods for real-time DPM detection are evaluated by a novel measure called the recall-time curves. As speed is a critical factor in comparing detection proposal methods for anytime detection, it should be considered in the evaluation methodology. Our recall-time graph methodology simultaneously evaluates the speed and quality of detection proposal methods. Specifically, the recall-time curve indicates the extent to which the proposal generator supports the following object-specific detector in a given time. Hence, the recall-time curves identify the proposal generator that best balances the speed and quality of the detection.

The present study introduces several improvements to our preliminary study (Kim & Sohn, 2015). First, the skin color feature is replaced with the color attributes (Van De Weijer et al., 2009), which might generalize the proposed method to categories other than people. Second, our present experiments are performed on a real mobile device (a Samsung Galaxy Note5). Finally, an additional comparison performed with a state-of-the-art detection proposal method, *Edge-Boxes*.

The rest of this paper is organized as follows. In Section 2, we briefly review recent works on proposal generation. Section 3 explains the proposed personness estimation. Our experimental results and conclusions are presented in Sections 4 and 5, respectively.

[1] BING (Cheng, Zhang, Lin, & Torr, 2014) is an exception, requiring only 15 ms on the same PC. BING will be discussed in Sections 2 and 3.1.