Neurocomputing 144 (2014) 184–197

Contents lists available at ScienceDirect

Neurocomputing

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

Fast object detection based on selective visual attention

Mingwei Guo, Yuzhou Zhao, Chenbin Zhang, Zonghai Chen *

Department of Automation, University of Science and Technology of China, Hefei 230027, PR China

ARTICLE INFO

Article history:
Received 26 October 2013
Received in revised form
9 February 2014
Accepted 24 April 2014
Communicated by Xu Zhao
Available online 9 June 2014

Keywords:
Object detection
Selective visual attention
Deformable part models
Sliding window

ABSTRACT

Selective visual attention plays an important role in human visual system. In real life, human visual system cannot handle all of the visual information captured by eyes on time. Selective visual attention filters the visual information and selects interesting one for further processing such as object detection. Inspired by this mechanism, we construct an object detection method which can speed up the object detection relative to the methods that search objects by using sliding window. This method firstly extracts saliency map from the origin image, and then gets the candidate detection area from the saliency map by adaptive thresholds. To detect object, we only need to search the candidate detection area with the deformable part model. Since the candidate detection area is much smaller than the whole image, we can speed up the object detection. We evaluate the detection performance of our approach on PASCAL 2008 dataset, INRIA person dataset and Caltech 101 dataset, and the results indicate that our method can speed up the detection without decline in detection accuracy.

1. Introduction

Computer vision is a science that tries to make the machine be able to 'see' the world as human beings. Researchers study the mechanisms of human visual system to improve computer vision. One of the mechanisms that make human visual system so effective in acting is the ability to extract the relevant information at an early processing stage, a mechanism called selective visual attention. The information captured by human eyes is too numerous to process on time for human visual system. However, the visual information is not of the same importance. Human brain filters the information and selects the interesting one for further processing by selective visual attention.

Object detection is one of the fundamental problems in computer vision and it focuses on both detecting object from video and static images. This paper mainly discusses object detection based on static image, which is to detect and locate the set class of object from a static image. Today, most object detection methods prefer to simplify object detection into a binary classification problem: determine whether there is the set type of object in the sliding window or not. As a result of sliding window, the speed of object detection is relatively slow which result that these object detection methods have limitation in many applications. The shortage of this kind of object detection methods is derived from the numerous visual information. Computers cannot process all of the visual information in an image on time.

To solve this problem, we naturally think of the selective visual attention. In Fig. 1, column a is the origin image, column b is the detection image and column c is the saliency detection image (the grayscale of the pixel in the saliency map shows the pixel's saliency degree). The origin images come from PASCAL VOC 2008 dataset [8–11]. From the first row of images, we can see that the pixels contained in the bounding box of detection image are of significant saliency. However, with the purpose of getting the bounding box, we have to search all the positions and scales in the image pyramid. It is taken for granted that reducing searching area is a way to speed up the detection. Thus, we consider to get candidate detection area which contains pixels of significant degree saliency as searching area. In row 2, the detection image contains a bounding box without a target object, in which the pixels are of little saliency. So we can eliminate the error detection by getting candidate detection area from the saliency map in order to maintain the detection accuracy.

We simulate the selective visual attention in order to construct an object detection method. From the research of visual psychology, we can summarize visual selective attention mechanism in two aspects [1]: (1) Bottom-up selective visual attention, which is driven by external stimulus, such as strong contrast. This kind of visual attention is a low-level cognitive process; and (2) Top-down visual attention, which is controlled by high-level brain information, such as knowledge, expectation, goal, and so on. Itti et al.[2–4] proposed selective attention model that is a bottom-up, irrelevant to the task, saliency based visual selective attention model, and it can simulate the visual

* Corresponding author. Tel.: +86 55163606104.
E-mail address: chenzh@ustc.edu.cn (Z. Chen).

http://dx.doi.org/10.1016/j.neucom.2014.04.054
0925-2312/

Fig. 1. The contrast of object detection and saliency detection.

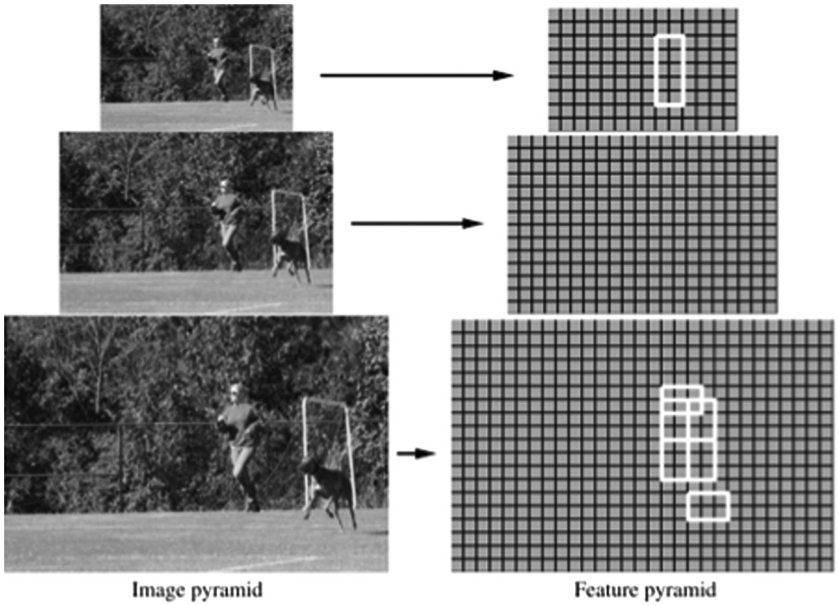

Fig. 2. HOG feature pyramid.

selective attention mechanism of human visual system. Serre and Pogggio et al. [5] simulate the computations performed by the feed-forward path of the ventral stream in visual cortex and the local circuits implementing them to build a model which can extract saliency from images. Moreover, plenty of saliency detection methods are proposed, for example, Zhai and Shah [6] extract color saliency by