Neurocomputing 168 (2015) 861–870

Contents lists available at ScienceDirect

Neurocomputing

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

Pedestrian detection based on hierarchical co-occurrence model for occlusion handling

Xiaowei Zhang [a], Hai-Miao Hu [a,b,*], Fan Jiang [a], Bo Li [a,b]

[a] Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China
[b] State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China

ARTICLE INFO

Article history:
Received 10 March 2015
Received in revised form
6 May 2015
Accepted 9 May 2015
Available online 19 May 2015

Keywords:
Pedestrian detection
Partial occlusions
Co-occurrence relations
Visibility status

ABSTRACT

In pedestrian detection, occlusions are typically treated as an unstructured source of noise and explicit models have lagged behind those for object appearance, which will result in degradation of detection performance. In this paper, a hierarchical co-occurrence model is proposed to enhance the semantic representation of a pedestrian. In our proposed hierarchical model, a latent SVM structure is employed to model the spatial co-occurrence relations among the parent–child pairs of nodes as hidden variables for handling the partial occlusions. Moreover, the visibility statuses of the pedestrian can be generated by learning co-occurrence relations from the positive training data with large numbers of synthetically occluded instances. Finally, based on the proposed hierarchical co-occurrence model, a pedestrian detection algorithm is implemented to incorporate visibility statuses by means of a Random Forest ensemble. The experimental results on three public datasets demonstrate the log-average miss rate of the proposed algorithm has 5% improvement for pedestrians with partial occlusions compared with the state-of-the-arts.

1. Introduction

Pedestrian detection is an important topic for practical applications, such as video surveillances [9,29], intelligent vehicles [10], and robot sensing. State-of-the-art algorithms have been used for achieving progress on pedestrian detection. However, the presence of partial occlusions causes significant degradation of performance, even for part-based algorithms that are supposed to be robust in that respect [11]. Therefore, pedestrian detection is still a challenge [1–8].

A general part-based hierarchy approach is introduced for detection of partially occluded objects in [15]. The algorithm requires a design for hierarchical object-parts to place parts for sharing weak features. Enzweiler et al. [16] presented a mixture of experts to focus on the unoccluded region applied to depth and motion images for handling partial occlusion. Javier et al. [17] describe a general algorithm for building a robust classifier ensemble by random subspace algorithm against partial occlusions. Zhang et al. [28] proposed a latent hierarchical model with varying structures to represent the behavior with multiple groups, and employ a multi-layer-based inference method to infer the group affiliation. Girshick et al. [6] proposed an extension of the deformable part-based detector [18] with occlusion handling. Specifically, the algorithm tries to place different body parts over the window. However, most previous approaches rely only on the detection score of a part for estimating its visibility and do not consider spatial co-occurrence relations among body parts.

Recently, Duan et al. [19] proposed a structural filter approach to human detection to deal with occlusions and articulated poses. The method manually defines the rules to describe the relationship between the visibility of a part and its overlapping larger and smaller parts. However, the visibility status of a part is obtained by hard thresholding of its detection score. Quyang et al. [20] presented a probabilistic pedestrian detection framework to learn the visibility relationship among overlapping parts at multiple layers. However, this method subjectively designs seven visibility parts to integrate in the last layer and thus is unable to model complex occlusion statuses for pedestrian detection.

The above algorithms have improved the performance for pedestrian detection to some extent. However, these algorithms rely only on their respective detection scores of parts for estimating visibility or depend on spatial consistency among the adjacent visibility parts. These algorithms fail to capture strong correlations among random visible parts, especially complex dependencies among nonadjacent visible parts. Moreover, complex occlusion

* Corresponding author at: Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China.
E-mail address: frank0139@163.com (H.-M. Hu).

http://dx.doi.org/10.1016/j.neucom.2015.05.038
0925-2312/

patterns are often inevitable due to various viewpoint changes. These algorithms just manually design several visibility statuses to integrate adjacent parts and not learn from training data. Thus, the algorithms fail to represent complex occlusion patterns for pedestrian detection.

As shown in Fig. 1, single-part detectors are imperfect, and such visibility estimation is inaccurate. Part detection score is relatively low when its visual cue does not fit the part detector well no matter whether the partial occlusion occurs or not. Furthermore, it is a key issue that how to integrate the inaccurate scores of part detectors and to estimate its location when there is partial occlusion in the sliding window. For example, many part-based deformable models in [6,21] summed the scores of part detectors. A pedestrian existing input window is considered as having a high sum for its score. However, when one part is occluded, the score of its part detector will be relatively low. Consequently, the summed score will be low. If the part-based deformable model [6] is used to detect the image, these occluded pedestrians may be mistaken for negative examples with low summed scores.

Considering the problems faced by the approaches discussed above, this proposes a hierarchical co-occurrence model to automatically learn complex dependencies among different parts for occlusion handling. Spatial consistency is built among the parent–child pairs of nodes from multiple layers, which fully explore complex correlation among visible parts. Moreover, the co-occurrence relations among random visibility parts within the same layer are modeled as latent variables of the structural SVM to generate visibility statuses. Finally, the random forest is used to combine visibility statuses to build a classifier ensemble robust against partial occlusions. Based on the proposed hierarchical co-occurrence model, a pedestrian detection algorithm is implemented for partial occlusion handling. Experimental results on three public datasets demonstrate that the proposed algorithm improves the log-average miss rate by around 5% for pedestrians with partial occlusions compared with the state-of-the-art algorithms.

2. Hierarchical co-occurrence model

2.1. Representation of the hierarchical co-occurrence model

To deal with the variations that cannot be tackled by a monolithic model, approaches to learning multiple parts model have been introduced in [19,20]. Integrating the advantage of part-based detectors in occlusion handling requires solving two key issues for successful detection of partially occluded pedestrians. The first issue is the decision if partial occlusion occurs in a scanning window and which body parts are occluded. The second issue is integrating inaccurate scores of part detectors and estimating their locations if partial occlusion is found in the sliding window. Therefore, the major challenges are the modeling of the correlation of the visibilities of different parts and the proper combinations of the results of part detectors according to the estimation of component visibility.

Fig. 2 shows the proposed hierarchical co-occurrence model with latent variables. The top layer has a wide variety of visibility status, which represents the possibility of the appearance of learned behavior from positive training data with large numbers of synthetically occluded instances. A visibility status is obtained by randomly combining one or more parts in the middle layer. The second layer has 12 part nodes in a 4×3 grid layout. Each of which represents one part of an object. Each node at the second layer has 4 child nodes at the bottom layer that contains 32 block nodes in an 8×4 grid layout. The nodes at lower layers capture more detailed appearance.

Note that in our model, we followed the implementations of [7] to calculate HOG descriptors. Fig. 2 shows the weights of HOG

Fig. 1. Complex occlusion patterns and estimation of visibility of a part from its detection score or from its correlated components. Black regions represent the estimated parts that are not visible.