Neurocomputing 135 (2014) 357–366

Contents lists available at ScienceDirect

Neurocomputing

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

Hybrid cascade boosting machine using variant scale blocks based HOG features for pedestrian detection

Van-Dung Hoang, My-Ha Le, Kang-Hyun Jo*

School of Electrical Engineering, University of Ulsan, South Korea

ARTICLE INFO

Article history:
Received 13 April 2013
Received in revised form
20 December 2013
Accepted 21 December 2013
Communicated by N.T. Nguyen
Available online 11 January 2014

Keywords:
Cascade boosting
Histograms of oriented gradient
Hybrid machine
Pedestrian detection
Support vector machines
Variant-scale block based HOG features

ABSTRACT

This paper contributes two issues for enhancing the accuracy and speed of a pedestrian detection system. First, it introduces a feature description using variant-scale block based Histograms of Oriented Gradients (HOG) features. By non-restricted block sizes, an extensive feature space that allows high-discriminated features to be selected for classification can be obtained. Second, a classification method based on a hybrid cascade boosting technique and a Support vector machine (SVM) is described. The SVM is known as one of the most efficient learning models for classification. On the other hand, one advantage of cascade boosting structure is to quickly reject most negative examples in the early layers, while retains almost all positive examples for speed up of the system. Because the performance of boosting depends on the kernel of weak classifier, the hybrid algorithms using the proposed feature descriptor is helpful for constructing an efficient classification with low computational time. In addition, an "integral image" method is utilized to support fast computation of the feature. The experimental results showed that performance of the proposed method is higher than the SVM using standard HOG features about 5% and the AdaBoost using variant-scale based HOG features about 4% detection rates, at 1% false alarm rates. The speed of classification using a cascade boosting approach is doubled comparing to that of the non-cascade one.

1. Introduction

Pedestrian detection using vision sensors is a key task for a variety of applications, which have potential influence in modern intelligence systems [1], knowledge integration and management in autonomous systems [2,3]. However, there are many challenges in the detection procedures such as various articulate poses, appearances, illumination conditions and complex backgrounds of outdoor scenes. As a module in the autonomous vehicle project, this task supports the avoidance of pedestrian collision while the vehicle is navigating [4–6]. In the pedestrian detection field, many approaches have been proposed for solving both feature description and detection method. Nowadays, the HOG descriptor is one common feature descriptor. The standard approach proposed by Dalal and Triggs [7] used HOG features and SVM algorithm. This approach is robust under various conditions including illumination, distortion and noise of outdoor environments. However, it requires high computational cost [1,7–10]. Previously, the Haar-like features descriptor and cascade boosting based approach was used in real-time object detection systems. However, the performance of Haar-like features is limited in human detection applications [11,12] due to its sensitivity to a high variety of human appearances and a complex background of scenes. In terms of learning algorithms used in object detection, SVM and boosting methods are the most popular algorithms which have been successfully applied to classification problems. In particular, the boosting technique can reduce the computational time with the use of a cascading structure at the classification stage [8,9,11–14].

This paper investigates the pedestrian detection problem using monocular images from a moving camera under outdoor scenes. There are two main contributions of this paper for constructing a robust pedestrian detection system based on a variant-scale block based HOG features and a cascade boosting SVM learning. First, derived from the original HOG feature descriptor, we proposed the variant-scale block based HOG features using multiple block sizes to construct an extensive set of features, which enables to extract the highly distinguished features between positive examples (pedestrian regions), and negative examples (non-pedestrian regions). An "integral image" method is also used for fast computation of the features. Second, a hybrid of SVM and boosting technique is proposed to take each advantage of SVM and the boosting technique for improving the efficiency of the classification system. SVM is used as a kernel of weak classifier inside of AdaBoost. The set of HOG features within each block (HOGB) is

* Corresponding author. Tel.: +82 52 259 1664
E-mail addresses: hvzung@islab.ulsan.ac.kr (V.-D. Hoang), lemyha@islab.ulsan.ac.kr (M.-H. Le), acejo@ulsan.ac.kr (K.-H. Jo).

http://dx.doi.org/10.1016/j.neucom.2013.12.017
0925-2312/

used to input of SVM. The boosting technique is used to select high discriminative HOGBs, which are used in the classification stage. The experiments showed that the proposed method, which fuses SVM and boosting techniques, achieves more accuracy than AdaBoost, and is faster than SVM. The proposed feature descriptor is more accurate than the standard HOG feature descriptor.

2. Related work and proposed method

In recent years, many successful methods for object detection have been proposed. They can be roughly divided into several categories. The first group focuses on a Haar-like feature description method. The standard approach investigated Haar-like features for object detection [13,15,16]. A method based on SVM and Haar-like features was presented for pedestrian detection [15,16]. On the other hand, a method using AdaBoost and Haar-like features was proposed in [11], which takes advantage of both motion and appearance information for walking person detection. However, Haar-like features are not stable with cluttered backgrounds and illuminative dynamic of outdoor environments. Therefore, other authors proposed a method that combines Haar-like and edge orientation histograms feature descriptor for pedestrian detection [17]. The main approach is based on a key advantage of Haar-like feature, which is fast computed by using a "integral image" method. The "integral image" is based on a cumulative sum of intensities within rectangular regions, which supports computing Haar-like feature with only eight accesses in a rectangle of any size.

The second group focuses on a HOG-based feature descriptor method. The standard approach investigated HOGs for human feature description in [7]. To deal with partial occlusion, Wang et al. [18] combined HOG and Local Binary Pattern (LBP) for feature description. In that system, the authors accumulated both HOG and LBP to construct feature vectors, which are fed to the SVM in both the training and detection stages. Experimental results indicated that the system was capable of handling partial occlusion. Systems based on HOG feature descriptors and SVM learning achieve significant accuracy in human detection field. However, they are usually high in computational cost [10]. To deal with this problem, Zhu et al. [9] and Hoang et al. [19] used the cascade AdaBoost and extended HOG features. In those methods, the authors increase a feature space by using different block sizes and locations, but the block ratios were restricted to be (1:1), (1:2), and (2:1). The block size ranges from 12×12 to 128×64 pixels corresponding to an example with resolution of 128×64 pixels. In the result, the feature vector consists of 5,031 blocks, which is too large for SVM training, and yields similar overlapped regions resulting in redundancy. The problem is solved by using AdaBoost for selecting good features. In another approach, both rectangular HOG and circular HOG feature descriptors were combined to build up a learning-based human detection system [8]. Similar to [9], the authors also restricted the ratios of both rectangular HOG and circular HOG to accumulate a total of 10,062 blocks on an example image with 128×64 pixels resolution. Their experiments showed that the features descriptor are insensitive to various lighting conditions, noise, and can overcome the effects of geometric and rotational variations. However, high time consumption is required to compute both rectangular and circular HOG. That system also encountered the above limitations.

Recently, another group focused on combining the AdaBoost and SVM method [20,21]. They proposed an algorithm combining SVM with boosting techniques to create a better classification benefiting from the desirable properties of both methods. The authors used SVM as a black-box weak learner in AdaBoost. In order to improve the efficiency of classification system, the heuristic technique is added for selecting a subset of training dataset to avoid the duplication examples and emphasizes the certain examples. However, they did not explore the structure of feature description, which effectively supports to select high-discriminated features for the classification stage. The authors in

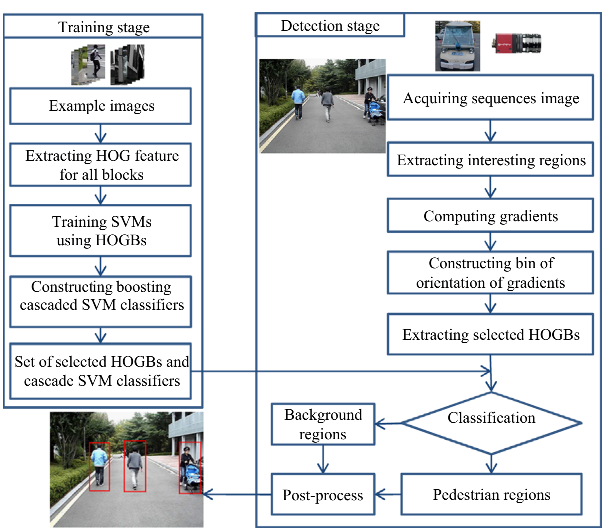

Fig. 1. Pedestrian detection system flowchart.

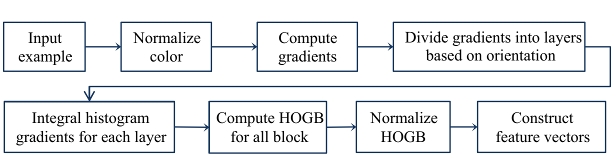

Fig. 2. Feature computation chain.

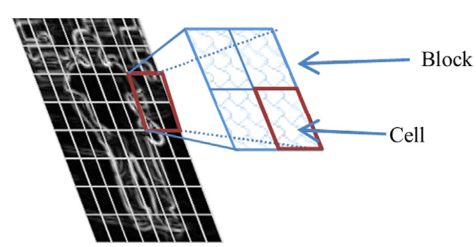

Fig. 4. HOGB features include 2×2 adjacent cells.

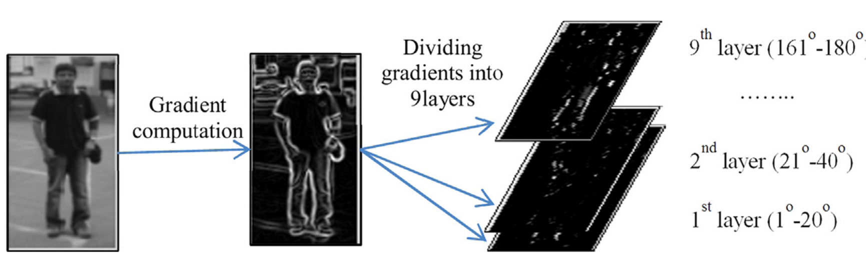

Fig. 3. Construction of 9 layers of gradients based on orientation.