J. Vis. Commun. Image R. 26 (2015) 168–181

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Joint global–local information pedestrian detection algorithm for outdoor video surveillance ☆

Hai-Miao Hu [a,b,*], Xiaowei Zhang [a], Wan Zhang [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 28 May 2014
Accepted 12 November 2014
Available online 21 November 2014

Keywords:
Pedestrian detection
Small-scale pedestrian
Global information
LBP
Histogram equalization pre-process
Outdoor video surveillance
Gaussian low-pass filtering
Global-local integrated model

ABSTRACT

The pedestrian size is usually small in practical outdoor surveillances. The small-scale pedestrian detection for outdoor surveillances is an important but difficult issue due to the limited information and the background interference. According to human cognition, the global information is important for the pedestrian detection. Therefore, a joint global–local information pedestrian detection algorithm is proposed to fully exploit and utilize the global information. The LBP feature is explicitly extracted from the low-frequency component of original images, which are utilized as the global information to suppress the background interference and enrich the description of pedestrian. Moreover, a structure-LBP is proposed to apply the inherent topology structure of human body to LBP. The structure-LBP feature extracted from original images can achieve a more discriminative description of pedestrians compared with the original LBP. The experimental results demonstrate that the proposed algorithm can improve the overall recognition performance for the small-scale pedestrians.

1. Introduction

Pedestrian detection is one of the active research areas in recent years and has been widely applied in video surveillances [1], collision avoidance on vehicles [2], robotics applications and advanced assistive technology for the visually impaired. Its major purpose is to automatically detect persons in one video and it is a fundamental technology for flow statistics of people in station and human activities analysis [3], which are important for outdoor video surveillance. However, for the outdoor video surveillance, the size of pedestrians is usually small and it is not as large as the size which the current pedestrian detection researches focus on (as shown in Fig. 1). This limitation makes a challenge for the current pedestrian detection algorithms [4,7,8,10,13–25,44–48,52,53] to be directly implemented for the practical applications.

Recently, many pedestrian detection algorithms have been proposed [4,7,8,10,13–25] and they can be divided into three categories, namely the human contour modeling algorithms [13,26], the template matching algorithms [14], and the statistical classification algorithms [4,7,8,10,13,22]. Compared with the other two algorithms, the statistical classification algorithms utilize machine learning theory to obtain a pedestrian classifier with a low computational complexity and good generalization. The current literature of statistical classification on pedestrian detection is typified by feature extraction, followed by a trainable classifier such as SVM [5,9], boosted classifiers [10]. Note that the proposed algorithm in this paper belongs to the statistical classification algorithms.

Feature extraction is one of the key issues in the statistical classification algorithms. The common features include the gray-scale feature (e.g., Haar [19]), the texture feature (e.g., the Local Binary Pattern, LBP [29,34]), and the shape feature [4,28,32] (e.g., the Histogram of Oriented Gradients, HOG [4]). Considering the fact that different features contain different information [33], these features are usually combined to achieve effective feature description over single feature [7,8,27,29].

However, the current researches mainly focus on the public dataset (e.g., MIT [12], INRIA [4], Caltech [11]), wherein the size of pedestrians is relatively big compared to the actual size of pedestrians in practical surveillance systems. Note that the pedestrian bigger than $18 * 36$ pixels and smaller than $25 * 50$ pixels is considered as the small-scale pedestrian and the pedestrians bigger than $25 * 50$ pixels are called the normal-scale pedestrians in this paper. The pedestrian size comparison is shown in Fig. 2. Although the current researches can be directly implemented for the small-scale pedestrians [6–8], the direct implementation fails

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author at: Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China. Fax: +86 010 82316225.

E-mail address: frank0139@163.com (H.-M. Hu).

http://dx.doi.org/10.1016/j.jvcir.2014.11.009
1047-3203/

to achieve the appropriate performance. According to the experimental result of [22], the best feature detector can achieve 21% mean miss-rate for the normal-scale pedestrians in Caltech Pedestrian Benchmark [11], while the mean miss-rate will rapidly increase to 73% for the small-scale pedestrians.

Therefore, the performance decline with decreasing scale is the major bottleneck for current pedestrian detection researches [22,56]. There are two major reasons, which will be analyzed as follows.

Firstly, in practical surveillance systems, the small-scale pedestrian is captured at a long distance. The small-scale pedestrians incline to be blurry due to the long distance and the background interference. Thus, a small amount of features will be extracted when the conventional methods aiming at the normal-scale pedestrians are directly implemented for the small-scale pedestrians. This can be observed from Fig. 2, wherein the amount of extracted LBP features [29,34] on pedestrians of two different scales is compared. It is obvious that the extracted LBP features on the normal-scale pedestrians are plentiful and discriminative when compared with the extracted LBP features on the small-scale pedestrian. Therefore, it is necessary to further exploit the limited information and make full use of it to improve the recognition performance of the small-scale pedestrian.

Secondly, the conventional features are proposed for the normal-scale pedestrians. However, different features are specifically designed to represent different characteristics and not all kinds of features can be directly applied to describe the characteristic of the small-scale pedestrians. As shown in Fig. 3, the HOG feature is good at describing the normal-scale pedestrians, while it fails to describe the small-scale pedestrians well due to degradation of the edge/local shape information. According to our experimental results, the LPB feature can achieve a relatively better performance for the small-scale pedestrians by comparing with other features. However, the direct implementation of LBP feature for small-scale pedestrian detection is still far from the requirement of practical applications. The detailed analysis will be given in the next section.

Note that human can recognition pedestrians from a relatively long distance regardless of the background interference, which provide an inspiration to improve the recognition performance of the small-scale pedestrians. According to the current researches, human cognition process is a process of starting from the global scope to the local part [39–41]. To be specific, a coarse analysis is made through a wide range of global information and a fine analysis is carried out by capturing the local details. This is because the global information (e.g., structures, contours, and topologies) possesses the salient characteristics of pedestrian and plays an important role in the pedestrian detection [39].

However, the global information is underutilized in the current researches. On the one hand, there is no generally accepted high-level feature which can accurately represent the structures, contours, and topologies. In most current researches, the high-level features are indirectly described based on the low-level features (e.g., HOG, LBP) and the extraction of the high-level features from the low-level features is implicitly finished through the classifier training.

On the other hand, there are several algorithms [5,6,20,24,30] using part model based on inherent structural information to classify pedestrians. And also, the head-shoulder structure [57] is a widely-used global feature due to its stability. But these algorithms tend to resolve the issue of the normal-scale pedestrian detection and fail on the small-scale pedestrians. Moreover, in terms of the small-scale pedestrians, the original images are usually blurred and they easily mix the global and the local information. Then the global information may be submerged due to the noise and blurring. It cannot discriminate the global information to directly extract features from the original images.

Therefore, it is important to fully exploit and efficiently utilize the global information. This paper proposes a joint global–local information pedestrian detection algorithm aiming at improve the performance of the small-scale pedestrian detection. The proposed algorithm mainly contains two contributions. Firstly, the global–local integrated pedestrian detection model is proposed to separate the low-frequency component from the original images through the Gaussian low-pass filtering and to extract LBP feature from the low-frequency component. The extracted features are explicitly utilized as the global information and they are integrated into local features to enrich feature description of the small-scale pedestrians. The proposed global–local integrated pedestrian

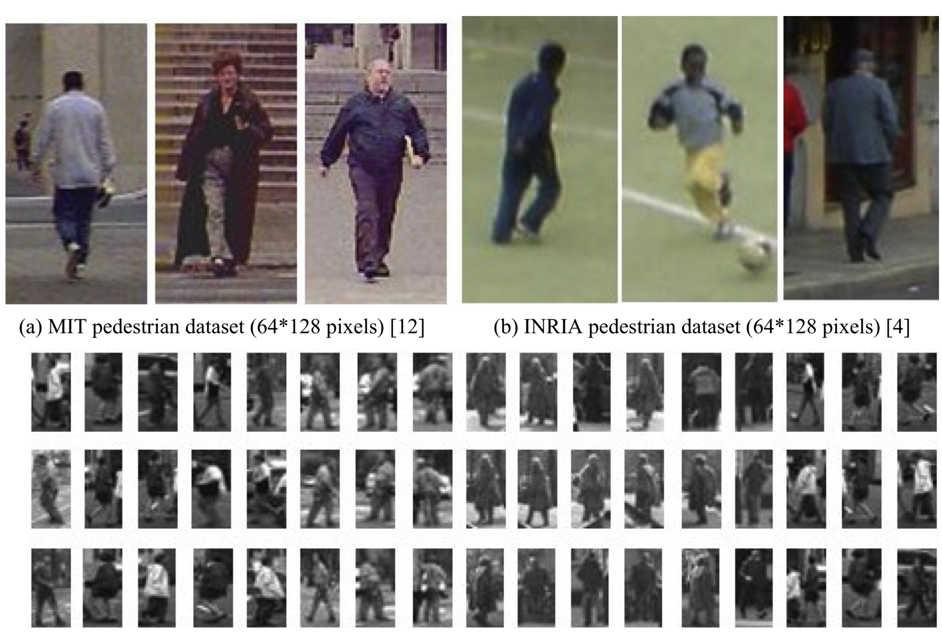

Fig. 1. The pedestrian size comparison among different datasets. (See above-mentioned references for further information.)