Optik 154 (2018) 22–32

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Original research article

Pedestrian abnormal event detection based on multi-feature fusion in traffic video

Xuan Wang*, Huansheng Song, Hua Cui

School of Information Engineering, Chang'an University, Xi'an, China

ARTICLE INFO

Article history:
Received 21 September 2016
Received in revised form 29 August 2017
Accepted 28 September 2017

Keywords:
Abnormal behavior event
Multi-feature fusion
Pedestrian trajectory analysis
Event classification

ABSTRACT

Pedestrian abnormal event detection is an active research area to improve traffic safety for intelligent transportation systems (ITS). This paper proposes an efficient method to automatically detect and track far-away pedestrians in traffic video to determine the abnormal behavior events. Firstly, pedestrian features are extracted by the multi-feature fusion method. Then, the similar features in current frame of all candidate objects are matched with the characteristic information of pedestrians in the previous frame which is considered as a template. Finally, pedestrian trajectory analysis algorithms are employed on the tracking trajectories and the motion information is attained, which can realize the early classification warning of pedestrian events. Experimental results on different traffic scenes in practice demonstrate that this method has good robustness in complex traffic. Moreover, the proposed method performs better compared with some other methods.

1. Introduction

In the past several years, the traffic abnormal events, such as illegal parking, abandoned object, illegal pedestrian, speeding, overloading, have been an increasing problem in traffic safety. Obviously, pedestrian abnormal behavior is one of the main reasons for the traffic accidents. In order to detect and manage illegal pedestrian events as soon as possible, this paper focuses on the issue of pedestrian abnormal behavior detection, which contains pedestrian detection, pedestrian tracking, and pedestrian trajectories analysis.

Pedestrian detection has attracted the attention of numerous researchers in recent years. It is an important issue in intelligent transportation systems (ITS). It requires preprocessing which consists of the recognition of the pedestrian target in video sequences and distinction of pedestrians from other video objects. The features of pedestrian is quite different from other objects, thus many researchers focus on the feature extraction to detect pedestrian. Gavrila and Giebel [1] presented a generic system for shape-based object recognition, which use the texture features to train a neural network to detect the pedestrian. Fang et al. [2] proposed a shape-independent pedestrian detection method in infrared images. It defines multi-dimensional histogram-, inertial-, and contrast-based classification features, which are shape-independent, complementary to one another, and capture the statistical similarities of image patches containing pedestrians with different poses. To avoid large areas of search, the system requires a pedestrian template and use brightness similarity to match candidate images and templates. Mohan et al. [3] presented a general example-based framework for detecting pedestrian by four components of the human body: the head, legs, left arm, and right arm. Xu et al. [4] adopted some local features of key parts of human

* Corresponding author.
E-mail address: jessica036@126.com (X. Wang).

https://doi.org/10.1016/j.ijleo.2017.09.104
0030-4026/

body to assist pedestrian detection. It used both appearance and motion global features of human body to select candidates, and then used local features of head and leg to do further confirmation. By combining Histograms of Oriented Gradients (HOG) and Local Binary Pattern (LBP) as the feature set, Wang et al. [5] proposed a pedestrian detection approach capable of handling partial occlusion, which use two kinds of detectors, i.e., global detector for whole scanning windows and part detectors for local regions, to learn from the training data by linear SVM.

Pedestrian tracking which locates the position of each detected target in the frames of analysed video sequences has a wide range of applications. In recent years, loads of methods for pedestrian tracking including active contour tracking [6], model tracking [7], feature tracking [8], etc., have been proposed by vision researchers. However, these approaches have some disadvantages. For example, active contour tracking is difficult to locate the contour of target accurately and mode tracking has a trouble in establishing pedestrian models which requires a large amount of computation. Because the objects of this paper are traffic video sequences, which are easily influenced by external environment, and traffic video incident detection system adopts mononuclear DSP processing chip which need to deal with a mass of image data in a short time. All of these require a high real-time ability, thus both of the first two methods are not suitable. However, features such as Gabor, color, edge, centroid, area and corner have a good performance in real-time tracking with less computation. Lim and Kim [9] tracked pedestrians with possibly partial occlusions with block matching method using color information. Barbu [10] used a robust pedestrian tracking method based on template matching process of Histogram of Oriented Gradient (HOG). Dinh et al. [11] presented a high throughput FPGA architecture for detecting corner features on traffic images.

Pedestrian trajectories analysis involves the analysis and recognition of motion patterns, and the production of high-level description of actions and interactions [12]. In general, it can be classified into two main categories of behavior description methods: statistical models and formalized reasoning. The Bayesian network model [13] is one of the statistical model. This model classifies certain events and behaviors by analysis of time sequences and statistical modeling. These methods need high-level reasoning based on a large amount of prior knowledge. Formalized reasoning [14] uses symbol systems to represent behavior patterns, and classified methods such as predication logic to recognize and classify events. For instance, Kojima et al. [15] proposed a method for generating natural language descriptions of pedestrian behaviors appearing in the image sequences.

In this paper, a method based on multi-feature fusion is used to detect and track pedestrians in traffic video so as to classify the pedestrian event. Motivated by that single feature can not solve the problem of segmentation very well, we combine several features of pedestrian and construct a mathematical model of matching function to detect pedestrian targets in traffic video. The proposed method has been tested on real-time video sequences and obtains good performance. The main contributions of this paper are concluded as follows:

1. We combine three different features of pedestrian to extract the pedestrian target. It avoids the limitations of a single feature, which has good robustness in complex traffic scene.
2. We establish the discriminant mode of pedestrian abnormal event to alarm these events timely in traffic scene.
3. The proposed method has better performance compared with some other method in the discrimination of pedestrian abnormal event.

The rest of the paper is organized as follows. Sections 2 and 3 describe the method of multi-feature extraction and fusion and the implementation of tracking, respectively. In Section 4, the motion information of pedestrian trajectory are analysed. Experiment results are given in Section 5. The final section draws the conclusion of this paper.

2. Multi-feature extraction and fusion

The pedestrian target of adjacent frames can be matched with certain characteristic information. In this paper, we bring some geometrical characteristics of pedestrian target together to establish the matching function as

$$F(x) = \sum_{i=1}^{n} w_i f_i(x), \tag{1}$$

where $f_i(x)$ is the entry data of $F(x)$, w_i is the corresponding weights, and $F(x)$ is defined as matching function. It is obvious that entries are the decisive parameters of the matching function. Weight coefficients represent the importance of each decisive parameter. The result of the matching function is used to measure the similarity between targets. In this paper, there are three entries including degree of location closeness, degree of area similarity and degree of corner point matching, and the weight coefficients are obtained by machine learning.

2.1. Centroid position

According to research statistics, the speed of pedestrian is generally 1.4 m/s and the shooting view of the traffic surveillance cameras is usually far, so pedestrian obtained in video image is quite small. Assuming that the speed of pedestrians motion is uniform so that the movement distance of pedestrians is short during adjacent frames. What is more, the altitude