Infrared Physics & Technology 60 (2013) 288–299

Contents lists available at SciVerse ScienceDirect

Infrared Physics & Technology

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

Robust and fast pedestrian detection method for far-infrared automotive driving assistance systems

Qiong Liu [a,b,*], Jiajun Zhuang [a,b], Jun Ma [b]

[a] School of Software Engineering, South China University of Technology, Guangzhou 510006, China
[b] School of Computer Science and Technology, South China University of Technology, Guangzhou 510006, China

HIGHLIGHTS

- A robust and fast far-infrared automotive pedestrian detection method is presented.
- Estimate potential pedestrian regions using pixel-gradient oriented vertical projection.
- PEWHOG is more effective for far-infrared pedestrian representations.
- Iteratively training procedure is presented to generate more robust classifier.
- Experimental results indicate the presented method is effective and promising.

ARTICLE INFO

Article history:
Received 23 October 2012
Available online 20 June 2013

Keywords:
Pedestrian detection
Far-infrared
ROIs generation
HOG
SVM
Automotive driving assistance systems

ABSTRACT

Despite considerable effort has been contributed to night-time pedestrian detection for automotive driving assistance systems recent years, robust and real-time pedestrian detection is by no means a trivial task and is still underway due to the moving cameras, uncontrolled outdoor environments, wide range of possible pedestrian presentations and the stringent performance criteria for automotive applications. This paper presents an alternative night-time pedestrian detection method using monocular far-infrared (FIR) camera, which includes two modules (regions of interest (ROIs) generation and pedestrian recognition) in a cascade fashion. Pixel-gradient oriented vertical projection is first proposed to estimate the vertical image stripes that might contain pedestrians, and then local thresholding image segmentation is adopted to generate ROIs more accurately within the estimated vertical stripes. A novel descriptor called PEWHOG (pyramid entropy weighted histograms of oriented gradients) is proposed to represent FIR pedestrians in recognition module. Specifically, PEWHOG is used to capture both the local object shape described by the entropy weighted distribution of oriented gradient histograms and its pyramid spatial layout. Then PEWHOG is fed to a three-branch structured classifier using support vector machines (SVM) with histogram intersection kernel (HIK). An off-line training procedure combining both the bootstrapping and early-stopping strategy is introduced to generate a more robust classifier by exploiting hard negative samples iteratively. Finally, multi-frame validation is utilized to suppress some transient false positives. Experimental results on FIR video sequences from various scenarios demonstrate that the presented method is effective and promising.

1. Introduction

Vision-based automatic pedestrian detection has become a hot spot in recent researches because it is widely used in automotive driving assistant systems [1–3], video surveillance [4,5], content-based image/video retrieval [6], etc. Pedestrian-to-vehicle collision happens more frequently after sundown [7] and the total number of traffic fatalities involving pedestrian is several times higher at night than daytime [8]. Therefore, it is necessary to explore reliable night-time pedestrian detection component for automotive driving assistant systems. Cameras in visible spectrum can be strongly influenced by illumination conditions and relatively infrared ones are more suitable to capture information at night, especially far-infrared (FIR) cameras because no active illumination is required. The progress for night-time pedestrian detection is also promoted due to the decreasing cost of FIR cameras in recent years.

Vision-based pedestrian detection method generally consists of two main modules: regions of interest (ROIs) generation and pedestrian recognition. We focus on the advances of monocular

* Corresponding author at: School of Software Engineering, South China University of Technology, Guangzhou 510006, China. Tel.: +86 13378680689; fax: +86 02039380258.
E-mail addresses: liuqiong@scut.edu.cn (Q. Liu), zhuang.jj@mail.scut.edu.cn (J. Zhuang), majun0614@163.com (J. Ma).

1350-4495/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.infrared.2013.06.003

pedestrian detection that share the insights for night-time automotive applications. For more comprehensive survey of recent works on pedestrian detection, the reader is referred to the work of Gerónimo et al. [9] and Dollár et al. [10].

The popular sliding window approach [11] that generates ROIs over multiple scales from the input images is not suitable for real-time automotive pedestrian detection. Inspired by sliding window approach, Sun et al. [12] presented a keypoint-centric based local sliding window technique for ROIs generation. In their method, all the candidate keypoints in images were detected using SUSAN detector and then ROIs were generated within the neighborhood of the detected keypoints. However, pedestrians are usually warmer and hence appear as brighter objects in FIR images from a local perspective and this special clue was not exploited in [12]. Global image thresholding segmentation was presented to generate ROIs by Bertozzi et al. [13], where the threshold was derived from the statistical properties of pre-collected images containing only background targets since the pixels from pedestrians were much brighter than that from background targets in their dataset. But the global thresholding segmentation technique faces difficulty in handling some potential difference in appearance of pedestrians among different image frames because pedestrians may not always be brighter than background from a global perspective. To address this problem, Ge et al. [2] proposed an adaptive local dual-thresholding segmentation algorithm to generate ROIs. But their dense handling within the whole input images is computationally intensive and might generate too many negative ROIs. Unlike the conventional region-growing used in [14], ROIs were generated using feature-based region-growing with high intensity seeds [1] and the algorithm stops when the connected regions' enclosing bounding boxes no longer cover the possible intervals of pedestrians' aspect ratios. Alternatively, Fang et al. [15] estimated ROIs' horizontal location using intensity-based horizontal projection and determined their vertical location through intensity/body-line-based vertical segmentation. Li et al. [4] subsequently proposed a similar method by combining both intensity-based horizontal and vertical projections. The intensity-oriented projection method is flexible, but the accuracy of generated ROIs heavily relies on the quality of infrared images since it requires that intensity of pixels from pedestrians should be higher than the average pixel intensity of the whole input images in a global view.

When a group of ROIs is generated, further validation will be performed in pedestrian recognition module. Following a learning-based discriminative framework, exploring more discriminative descriptors for pedestrian representations and designing more powerful learning algorithms have always been the pursuits. The discriminative descriptors include Haar wavelets [2,16,17], local binary/ternary patterns (LBP/LTP) and their variants [12,18,19], shapelet [20], edgelet [11,21], intensity self similarity (ISS) [22], histograms of oriented gradients (HOG) and its variants [23–26] etc. Then the extracted features are fed to different classifiers. Support vector machines (SVMs) and boosting algorithms are the two occupied learning algorithms and show excellent performance over state-of-the-art methods in pedestrian recognition [9].

One of the first pioneering efforts was the work of Papageorgiou and Poggio [17] where the combination of Haar wavelets and a polynomial SVM was presented. Inspired by the Haar wavelets, Viola et al. [16] introduced Haar-like features for pedestrian representations and proposed a cascade AdaBoost learning framework for both automatic feature selection and efficient pedestrian detection. Ge et al. [2] extended the cascade detection framework and introduced a two-stage (cascade) tree-structure near-infrared pedestrian classifier using Gentle-AdaBoost. Similar tree-structure detection framework was also proven to be effective by Xu et al. [27]. The dense HOG descriptor was designed specifically for pedestrian representations by Dalal and Triggs [23], and has since showed excellent performance for finding pedestrians in visible spectrum. Zhu et al. [24] introduced integral histograms to extract HOG for use in cascade-of-rejectors and they observed that the more informative HOG blocks are those locating in local edge or contour regions of pedestrians. In later work, O'Malley et al. [1] successfully extended HOG to pedestrian recognition on automotive FIR videos. Sun et al. [12] extended LBP using the spatial layout of texture cells to describe the symmetrical characteristic of FIR pedestrians and proposed pyramid binary pattern (PBP). And PBP was performed effectively with an SVM classifier. Inspired by the descriptor for characterizing color self similarity (CSS) [28] in the visible spectrum, Miron et al. [22] proposed ISS based on the relative intensity self similarity within specific FIR pedestrian regions, e.g. the head region of pedestrians shares more similar gray-level intensity.

Despite considerable descriptors have been proposed, those based on image gradients (e.g. HOG-like descriptors) are still probably most effective for pedestrian detection [9,10]. Although the work in [19,29,30] demonstrate that the idea of fusing different descriptors can benefit better recognition performance, we only focus on exploring more discriminative single descriptor in this work, because it can also guarantee superior performance when fusing with other descriptors. By further exploiting the underlying data characteristic of FIR images and following the conclusions of [24], we explore the potential of dense HOG descriptor for FIR pedestrian detection by introducing the entropy of distribution of oriented gradient histograms. Gao [31] presented a close procedure, where the entropy of HOG was calculated to represent the image texture and then entropy thresholding was used to directly filter out the regions with either dense texture or textureless that were regarded as ambiguous for image matching and registration. However, it does not necessarily mean that complete or over-complete feature set is redundant according to the idea of dense HOG. This indicates that the performance of pedestrian recognition would decrease if we filter out the features estimated as less informative directly. In addition, we also consider that the training procedure (often underestimated in literature) is crucial, because the recognition performance usually depends on the adopted training procedure when the initial training data and learning algorithms are fixed [23,28].

This work focuses on developing a robust and efficient night-time pedestrian detection system for automotive applications based on monocular FIR camera. A novel recognition framework is proposed to learn and recognize FIR pedestrians, by taking the advantage of an effective descriptor termed pyramid entropy weighted histograms of oriented gradients (PEWHOG) and a new training procedure. The main contributions of this paper are as follows: (1) Pixel-gradient oriented vertical projection is proposed to estimate the input FIR image and reduce its searching regions efficiently. It can be regarded as a preliminary segmentation procedure with which other accurate ROIs generation techniques (e.g. thresholding segmentation) can be coupled. The combination can generate nearly the same positive ROIs and much fewer negative ROIs with a lower runtime. (2) Novel PEWHOG is proposed for more effective FIR pedestrian representations by capturing both the local object shape using the entropy weighted distribution of oriented gradient histograms and its pyramid spatial layout. Then a three-branch structured classifier based on PEWHOG and SVM is used to address the high within-class variance problem caused by pedestrians' imaging size. (3) An iteratively training procedure combining both the bootstrapping and early-stopping strategy is proposed to generate a more robust classifier with lower prediction error. (4) Extensive experiments including both classifier-level and system-level performance evaluations have been conducted to verify the effectiveness of the presented method, and the results demonstrate that it is robust and suitable for real-time applications.