Infrared Physics & Technology 76 (2016) 157–167

Contents lists available at ScienceDirect

Infrared Physics & Technology

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

Pedestrian detection from thermal images: A sparse representation based approach

Bin Qi [a,b], Vijay John [c,e], Zheng Liu [d,*], Seiichi Mita [e]

[a] Science and Technology on Underwater Acoustic Laboratory, Harbin Engineering University, Harbin 150001, China
[b] College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China
[c] Intelligent Information Processing Laboratory, Toyota Technological Institute, Nagoya 468-8511, Japan
[d] School of Engineering, The University of British Columbia, Okanagan Campus, Kelowna, BC V1V 1V7, Canada
[e] Research Centre for Smart Vehicles, Toyota Technological Institute, Nagoya 468-8511, Japan

ARTICLE INFO

Article history:
Received 3 September 2015
Revised 22 February 2016
Accepted 22 February 2016
Available online 27 February 2016

Keywords:
Pedestrian detection
Sparse representation
Thermal images
Feature fusion

ABSTRACT

Pedestrian detection, a key technology in computer vision, plays a paramount role in the applications of advanced driver assistant systems (ADASs) and autonomous vehicles. The objective of pedestrian detection is to identify and locate people in a dynamic environment so that accidents can be avoided. With significant variations introduced by illumination, occlusion, articulated pose, and complex background, pedestrian detection is a challenging task for visual perception. Different from visible images, thermal images are captured and presented with intensity maps based objects' emissivity, and thus have an enhanced spectral range to make human beings perceptible from the cool background. In this study, a sparse representation based approach is proposed for pedestrian detection from thermal images. We first adopted the histogram of sparse code to represent image features and then detect pedestrian with the extracted features in an unimodal and a multimodal framework respectively. In the unimodal framework, two types of dictionaries, i.e. joint dictionary and individual dictionary, are built by learning from prepared training samples. In the multimodal framework, a weighted fusion scheme is proposed to further highlight the contributions from features with higher separability. To validate the proposed approach, experiments were conducted to compare with three widely used features: Haar wavelets (HWs), histogram of oriented gradients (HOG), and histogram of phase congruency (HPC) as well as two classification methods, i.e. AdaBoost and support vector machine (SVM). Experimental results on a publicly available data set demonstrate the superiority of the proposed approach.

1. Introduction

Recent advances in science and technology have enabled the growth of endowing machines with the ability to interact with people. This has created an immense opportunity for vision data engineering and research to play an essential role in a wide range of applications from civilian safe driving to national security [1]. Detecting people in images is a fundamental research of human behavior analysis, but plays a paramount role in advanced driver assistant system (ADAS) and autonomous driving, especially with the growth of aging population [2]. For advanced driver assistant systems, pedestrian detection aims to avoid collisions and accidents by offering technologies that alert drivers the potential dangers or implementing safeguards via initiating protective measures, such as taking control of the vehicles. Due to the growth spurt of vehicles, traffic accidents occupy a large part of fatalities with pedestrians being the most vulnerable traffic participants. Investigators revealed that more than 10 million people encountered traffic accidents around the world every year, and 2–3 million of those people were seriously injured [3]. Moreover, pedestrians account for 24% of all traffic fatalities worldwide and pedestrian accidents have the highest fatality rate among all traffic participant groups, with about 8% of all pedestrian accidents being deadly [4].

Generally, pedestrians are detected from gray-scale or color images. However, human bodies have the characteristic of rigidity and flexibility and there is a wide range of possible pedestrian appearance, due to illumination, occlusion, articulated pose and complex background [5]. Sometimes, the use of visible images is not feasible without external artificial illumination, especially at night time. Thus, pedestrian detection is a difficult task with

* Corresponding author.
E-mail address: zheng.liu@ieee.org (Z. Liu).

http://dx.doi.org/10.1016/j.infrared.2016.02.004
1350-4495/

potential challenges for perception in visible spectrum. In contrast, thermal images are captured by thermal sensors which sense emitted radiation from objects of interest such as humans or vehicles [6]. Thermal images use the intensity maps, which provide an enhanced spectral range to make human beings perceptible and highlight the contrast between objects of high temperature variance and the cool background. Fig. 1 shows sample thermal images, which demonstrate the advantages over visible images in these scenarios. Thus, it is possible to detect pedestrians from thermal images with insufficient or over illumination. Moreover, the variability introduced by color, texture, and complex background becomes trivial.

Pedestrian detection is often categorized as a 2-classes object classification problem [7], where learning methods are employed to classify a candidate object or region into pedestrian or non-pedestrian. The pedestrian detection framework, typically, consists of the following steps: candidate selection, feature extraction, and feature classification [1,3,5,8]. For pedestrian detection with thermal images, researchers exploited the thermal contrast in the images to identify the potential positions of candidate pedestrians [9,10]. Discriminative features are extracted and provided as an input to the classification framework from the candidate pedestrians. Based on the sufficient features from prior known pedestrians and non-pedestrians, the representative parameters are generated for the chosen classifier. Then the classifier is used to attach labels to the features for the candidate pedestrians. Significant attention has been paid to extracting discriminative features that enhance the pedestrian detection rate, while reducing the false positive rate. In the work by Olemda et al. [11], to reduce the effect caused by the illumination change in thermal images, image phase congruency feature was used to detect the pedestrians. Fang et al. [9] classified the candidates using multidimensional histograms, inertia, and contrast-based features. The candidates were selected using a projection-based segmentation method. Recently, Liu et al. [10] proposed a feature descriptor named pyramid entropy weighted histograms of oriented gradients to represent the candidate pedestrians. The candidate pedestrians were identified using thermal gradient and threshold-based image segmentation.

While the use of discriminative features significantly improves the detection accuracy, the presence of background noise is still a challenge for pedestrian detection. In Davis et al. [12], a two-stage template-based method is proposed to locate candidate pedestrian locations using contour saliency maps. The hypothesized person locations are validated with an AdaBoost ensemble classifier with an automatically tuned filter. Wang et al. [13], extended this approach by incorporating tracking within their framework. Candidate pedestrians extracted by thermal intensity information are classified using a learned support vector regression model. A pedestrian tracker was used to localize the detected pedestrians and adaptively update the detection parameters. The adaptive framework was shown to increase the robustness of the pedestrian detection. A similar approach is adopted by Dai et al. [14], where detection and tracking were simultaneously implemented. Motion and appearance-based features were used in addition to shape information within an EM-based framework to detect the pedestrians. Multimodal information fusion was also adopted to increase the robustness of the pedestrian detection in varying illumination conditions. Davis et al. [15] combined information from thermal and visible images to perform candidate selection, feature extraction, and feature classification using pedestrian shape information. Contour saliency maps were generated to highlight salient objects with the input and background gradient information. Pedestrians contours were then extracted from the generated saliency maps. Similarly, Ge et al. [8] combined visible and thermal images within a simultaneous pedestrian detection and tracking framework. To account for background noise and illumination variations, the authors trained multiple classifiers on disjoint subsets of different image sizes. The trained classifiers were then arranged in a tree structure to perform detection in a coarse-to-fine manner.

Fig. 1. Human being is not perceptible in the visible image on the left column, but can be detected from thermal images on the right (image courtesy of Infiniti Optics: http://www.infinitioptics.ca).