Pattern Recognition 51 (2016) 148–175

Contents lists available at ScienceDirect

Pattern Recognition

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

Human detection from images and videos: A survey

Duc Thanh Nguyen [a,b,*], Wanqing Li [a], Philip O. Ogunbona [a]

[a] School of Computer Science and Software engineering, University of Wollongong, NSW 2522, Australia
[b] Information Systems Technology and Design, Singapore University of Technology and Design, Singapore

ARTICLE INFO

Article history:
Received 10 January 2015
Received in revised form
14 July 2015
Accepted 30 August 2015
Available online 13 September 2015

Keywords:
Human detection
Human description
Object detection
Object description

ABSTRACT

The problem of human detection is to automatically locate people in an image or video sequence and has been actively researched in the past decade. This paper aims to provide a comprehensive survey on the recent development and challenges of human detection. Different from previous surveys, this survey is organised in the thread of human object descriptors. This approach has advantages in providing a thorough analysis of the state-of-the-art human detection methods and a guide to the selection of appropriate methods in practical applications. In addition, challenges such as occlusion and real-time human detection are analysed. The commonly used evaluation of human detection methods such as the datasets, tools, and performance measures are presented and future research directions are highlighted.

1. Introduction

The problem of detecting humans can be simply stated as: *given an image or video sequence, localise all subjects that are human*. This problem corresponds to determining regions, typically the smallest rectangular bounding boxes, in the image or video sequence that enclose humans. Fig. 1 shows some examples of human detection.

During the last decade, human detection has attracted considerable attention in computer vision and pattern recognition largely due to the variety of applications it enables. In visual content management, a typical task is to tag (or label) the objects, especially humans, in the images and videos. Such tagging will enable subsequent annotation, search, and retrieval. Human detection is an essential component of automatic tagging. In video-based surveillance, one of the key tasks is to detect, identify, and monitor humans in crowded and public scenes such as airports, train stations, and supermarkets. Human detection is also found to be crucial in autonomous vehicles. It detects the presence of pedestrians on streets to alert the driver of dangerous situations. Examples include the ARGO vehicle[1] developed by the University of Parma and the Chamfer system[2] released by the University of Amsterdam and Daimler Chrysler. Recently, Mobileye[3] launched the first vision-based collision warning system with full auto brake and pedestrian detection for use in the Volvo S60 cars. Fig. 2 illustrates various applications of human detection.

Human beings are capable of detecting humans by only using subtle clues. Automated algorithms are instead still far from matching or even just approaching this ability. This is, in part, due to the intrinsic difficulties associated with the human body and the environment in which it is located. The non-rigid nature of the human body produces numerous possible poses. It is also challenging to model simultaneously view (orientation) and size variations arisen from the change of the position and direction (e.g. tilt angle) of the camera. Unlike other types of objects, humans can be clothed with varying colours and texture, which adds another dimension of complexity. In addition, the environment can make humans less visually noticeable. For example, the ambient illumination could either enhance or degrade the visual appearance of human objects. A cluttered background, often encountered in outdoor scenes, could camouflage humans. Furthermore, whether a single human is in activities or multiple humans interact with each other in crowded scenes, occlusions where the body of the human object is not completely observed are inevitable.

Human detection has been studied and progressed substantially in the past. The review of human detection has also been conducted in several works such as [1–10]. However, our survey is different from those in the following dimensions:

- Human detection in existing surveys was reviewed so far either as a component of a human motion analysis system, e.g. [1–5], or in the context of specific applications, for instance, pedestrian protection in a driving assistance system [6–10]. In contrast, this paper considers human detection as an object

* Corresponding author. Tel.: +61 2 4221 3103; fax: +61 2 4221 4170.
E-mail addresses: dtn156@uowmail.edu.au, ducthanh_nguyen@sutd.edu.sg (D.T. Nguyen), wanqing@uow.edu.au (W. Li), philipo@uow.edu.au (P.O. Ogunbona).

[1] http://www.argo.ce.unipr.it/argo/english/index.html
[2] http://www.gavrila.net/Research/Chamfer_System/chamfer_system.html
[3] http://www.mobileye.com/

http://dx.doi.org/10.1016/j.patcog.2015.08.027
0031-3203/

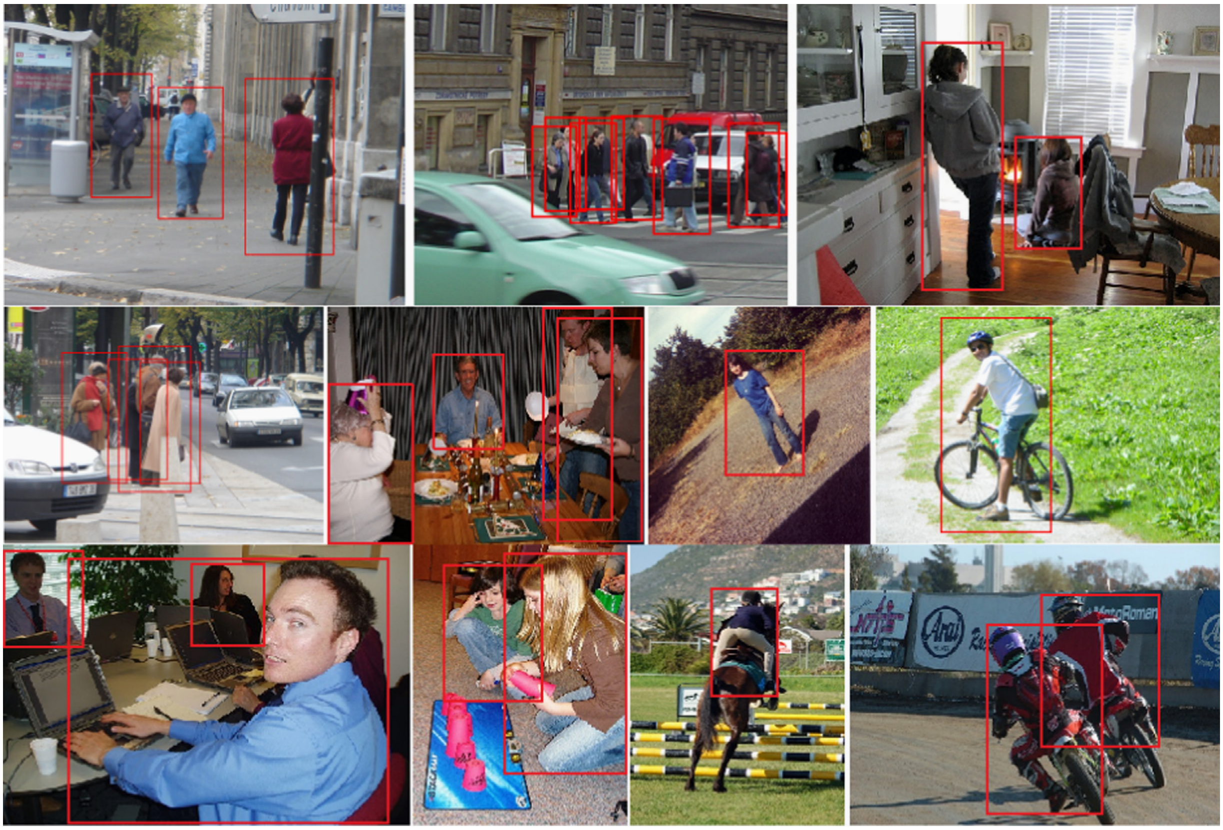

Fig. 1. Some examples of human detection results.

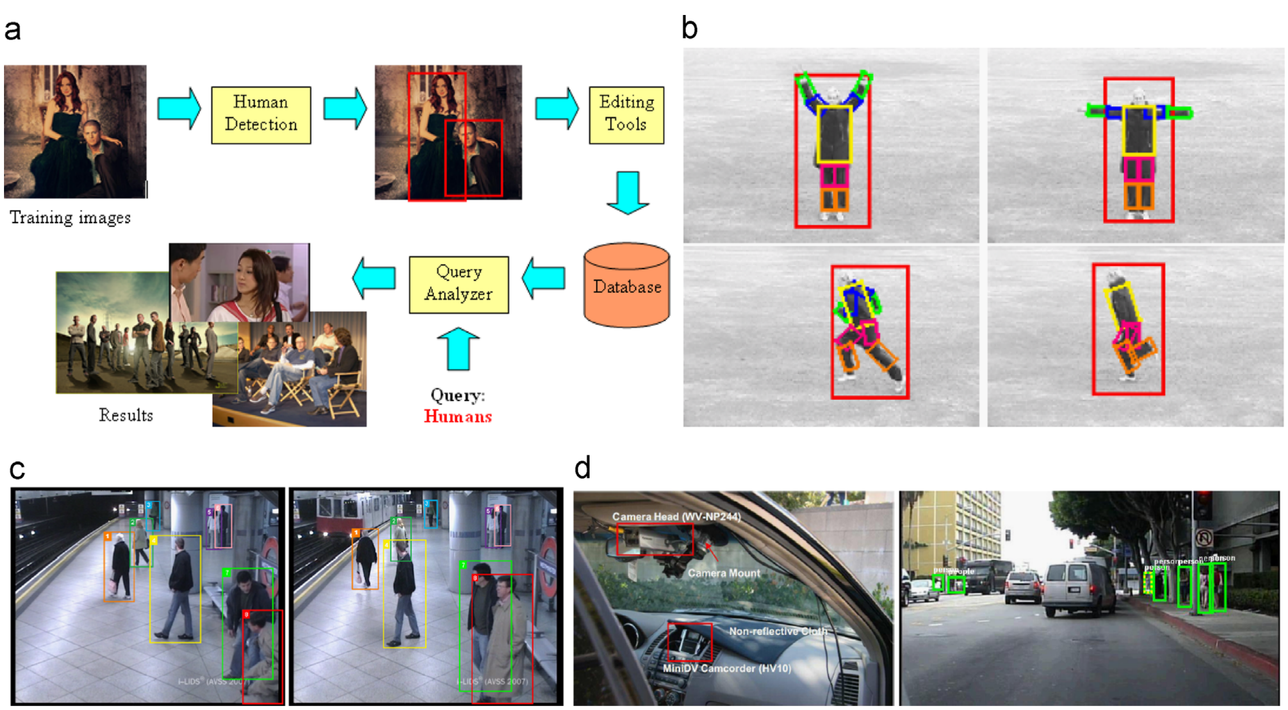

Fig. 2. Applications of human detection. (a) Image/video retrieval. (b) Human activity recognition. Different body parts are labelled with different colours. (c) Surveillance systems (with detection and tracking). Humans are labelled using different colours. (d) Driving assistance system (from [9]).

detection problem with an emphasis on the specific challenges posed by the articular nature and versatile visual appearance of the human body.

- Existing surveys decompose a human detection method into two components: features and classifiers. However, we have found that given the same feature, different ways of constructing the object descriptor from the feature could gain different performance. In addition, this scheme cannot well explain the recent developments of human detection, e.g. in [11–14], the structure of human objects was modelled to cope with deformations. Our