ABSTRACT

As computing becomes ubiquitous in our modern society, automated recognition of human activities emerges as a crucial topic where it can be applied to many real-life human-centric scenarios such as smart automated surveillance, human computer interaction and automated refereeing. Although the perception of activities is spontaneous for the human visual system, it has proven to be extraordinarily difficult to duplicate this capability into computer vision systems for automated understanding of human behavior. Motion pictures provide even richer and reliable information for the perception of the different biological, social and psychological characteristics of the person such as emotions, actions and personality traits of the subject. In spite of the fact that there is a considerable body of work devoted to human action recognition, most of the methods are evaluated on datasets recorded in simplified settings. More recent research has shifted focus to natural activity recognition in unconstrained scenes with more complex settings.
INTRODUCTION

Much research within the computer vision community is dedicated towards the analysis of and understanding of human motion. The perception of human motion is one of the most important skills people possess, and our visual system provides particularly rich information in support of this skill. Yet, attempts and efforts to understand the human visual system or to devise an artificial solution for visual perception have proven to be a difficult task. Human motion analysis has received much attention from researchers in the last two decades due to its potential use in a plethora of applications. This field of research focuses on the perception and recognition of human activities. As computing becomes ubiquitous in our modern society, the recognition of human activities emerges as a crucial topic where it can be applied to many real-life human-centric scenarios (Aggarwal & Ryoo, 2011). Furthermore, given the immense expansion of video data being recorded in everyday life from security surveillance cameras, movies production and internet video uploads, it becomes an essential need to automatically analyse and understand video content semantically. This is to ease the process of video indexing and fast retrieval of data when dealing with large multimedia content and big data. Hence, the importance of automated systems for human activity recognition is central to the success of such applications (Turaga, Chellappa, Subrahmanian, & Udrea, 2008). Further, due to the proliferating number of crimes and terror attacks as well as the vital need to provide safer environment, it becomes a necessary requirement to improve current state of surveillance systems via the use of computer vision methods to automate procedures of detecting suspicious human activities.

Human activity recognition aims to automatically infer the action or activity being performed by a person or group of people. For instance, recognizing whether someone is walking, raising hands or performing other types of activities. This usually involves the analysis and recognition of different motion patterns in order to produce a high-level semantic description for the human activities or interaction between people. This is vital to apprehend the human behavior and to determine whether their behavior is abnormal or normal via the use of automated methods (Ko, 2008). There have been considerable amount of work by the computer vision community dedicated to activity recognition with numerous approaches and methods being proposed to address different aspects and contexts of this area of research (Aggarwal & Ryoo, 2011; Poppe, 2010). Many of the early approaches have considered the use of video sequences recorded using a single camera with people being asked to perform basic actions in simplified settings and conditions. Various low-features have been proposed for encoding the human activity either at a temporal or spatial level such as edges, curvatures or complex features such as interest point descriptors. The detection of human motion is considered as a rudimentary component for constructing the activity descriptor in the majority of approaches either explicitly or implicitly for recovering other high level features. In fact, it is infeasible to detect human action from a still frame as even though achieving pose recovery can be possible from a single image, the perception of human activity can be challenging. Vishwakarma and Agrawal (2013) grouped the methods object detection through motion estimation into six conventional methods: background subtraction, statistical methods, temporal differencing and optical flow. Various recent surveys can be found in the literature on the representation of different features for human activity recognition (Poppe, 2010; Turaga et al., 2008; Vishwakarma & Agrawal, 2013). Interestingly, a new trend of research has emerged on activities recognition through the use of wearable sensors mounted to the human body (Lara & Labrador, 2013).