Neurocomputing 171 (2016) 754–767

Contents lists available at ScienceDirect

Neurocomputing

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

Transition-Aware Human Activity Recognition Using Smartphones

Jorge-L. Reyes-Ortiz [a,b,*], Luca Oneto [c], Albert Samà [b], Xavier Parra [b], Davide Anguita [a]

[a] DIBRIS – University of Genova, Via Opera Pia 13, I-16145 Genova, Italy
[b] CETpD – Universitat Politécnica de Catalunya, Rambla de l'Exposició, 59-69, 08800 Vilanova i la Geltrú, Spain
[c] DITEN – University of Genova, Via Opera Pia 11a, I-16145 Genova, Italy

ARTICLE INFO

Article history:
Received 3 September 2014
Received in revised form
21 June 2015
Accepted 8 July 2015
Communicated by Marco Cristani
Available online 8 August 2015

Keywords:
Activity Recognition
Smartphones
Transitions
Support Vector Machines
Machine learning

ABSTRACT

This work presents the Transition-Aware Human Activity Recognition (TAHAR) system architecture for the recognition of physical activities using smartphones. It targets real-time classification with a collection of inertial sensors while addressing issues regarding the occurrence of transitions between activities and unknown activities to the learning algorithm. We propose two implementations of the architecture which differ in their prediction technique as they deal with transitions either by directly learning them or by considering them as unknown activities. This is accomplished by combining the probabilistic output of consecutive activity predictions of a Support Vector Machine (SVM) with a heuristic filtering approach. The architecture is validated over three case studies that involve data from people performing a broad spectrum of activities (up to 33), while carrying smartphones or wearable sensors. Results show that TAHAR outperforms state-of-the-art baseline works and reveal the main advantages of the architecture.

1. Introduction

Human Activity Recognition (HAR) has nowadays become a prominent research field due to its substantial contributions in human-centered areas of study aiming to improve people's quality of life: Ambient Intelligence, Pervasive Computing and Assistive Technologies [1–3]. These areas make use of HAR systems as an instrument that provides information about people's behavior and actions [4]. This is commonly done by gathering signals from ambient and wearable sensors and processing them through machine learning algorithms for classification. There are currently many applications where HAR systems are used, for instance, the continuous monitoring of patients with motor problems for health diagnosis and medication tailoring [5], and the automated surveillance of public places for crime prevention [6].

In the past decade, several HAR systems have been proposed and surveyed [7–9]. They have encompassed multiple activities from different application domains, including locomotion, daily living activities, transportation and sports [10,11] (e.g. walking, cooking, driving, and running). Regarding their duration and complexity, activities are categorized in three main groups: short events, Basic Activities (BAs) and complex activities. The former group is comprised of brief-duration activities (on the order of seconds) such as postural Transitions (PTs) (e.g. sit-to-stand), and body gestures [8]. Basic activities are instead characterized by a longer duration and can be either dynamic or static (e.g. running and reading) [12]. The latter group, complex activities, is composed of progressions of the aforesaid simpler activities and involve aspects such as interaction with objects and other individuals (e.g. playing sports, social activities) [13]. This research targets the first two categories.

1.1. Wearable sensors and smartphones

Ambient and *wearable* sensors have been actively exploited for HAR [1]. Video cameras, microphones, GPSs, and sensors for measuring proximity, body motion and vital signs are just a few examples. Current research on ambient sensors has mainly focused on video cameras due to the ease of retrieving visual information from the environment. These have also been combined with other sensors (e.g. with accelerometers and microphones [14]) and recently introduced in wearable technologies for novel ubiquitous applications [15]. However, people's privacy is a downside of vision-based technologies that limits their use in every location. In contrast, recent developments in wearable sensing technologies such as inertial and vital signs sensors are offering less invasive alternatives for HAR [16].

* Corresponding author at: DIBRIS – University of Genova, Via Opera Pia 13, I-16145 Genova, Italy.
E-mail addresses: jorge.reyes.ortiz@smartlab.ws (J.-L. Reyes-Ortiz), Luca.Oneto@unige.it (L. Oneto), albert.sama@upc.edu (A. Samà), xavier.parra@upc.edu (X. Parra), Davide.Anguita@unige.it (D. Anguita).

http://dx.doi.org/10.1016/j.neucom.2015.07.085
0925-2312/

The *accelerometer* is the most commonly used sensor for reading body motion signals [8]. This body sensor is generally used either in multi-sensor arrangements (e.g. triaxial accelerometers and Body Sensor Networks (BSN)) or in combination with others (e.g. gyroscopes, magnetometers, temperature, and heart rate sensors) [17]. Bao and Intille [12] proposed one of the earliest HAR systems for the recognition of 20 activities of daily living using five wearable biaxial accelerometers and well-known machine learning classifiers. They achieved reasonably good classification accuracy reaching up to 84% considering the number of activities involved. One evident drawback was related to the number and location of the body sensors used which made the system highly obtrusive. *Gyroscopes* have also been employed for HAR and have demonstrated to improve the recognition performance when used in combination with accelerometers [18,19].

Smartphones have become an alternative for *wearable sensing* due to the diversity of sensors they support. This aspect, along with the device processing and wireless communication capabilities, makes them a robust tool for performing activity recognition [20,21]. Smartphones also have advantages over other *ambient sensing* approaches, such as multi-modal sensors in a home environment or surveillance cameras, because they are ubiquitous and require none or little static infrastructure to operate [1]. Inertial sensors such as accelerometers and gyroscopes are present in modern smartphones as they can be mass produced at a low cost. They are an *opportunistic sensing* resource for retrieving body motion data [22,23].

First smartphone-based approaches worked offline. In [24], the *Centinela* system was presented. It consisted of a chest unit composed of several sensors to measure acceleration data and vital signs (e.g. heart rate, breath amplitude, and respiration rate) and a smartphone wirelessly connected via Bluetooth. Data was later processed and classified offline using different machine learning algorithms. Lee and Cho in [25] developed a HAR system of 5 transportation activities which combines labeled and unlabeled data from smartphone inertial sensors with a mixture-of-expert model for classification. Kwapisz et al. [26] developed an offline HAR system using a smartphone provided with a built-in triaxial accelerometer carried on the pocket. Their recognition model allowed the classification of 6 locomotion activities (2 static postures and 4 dynamic activities). Similarly, we proposed in [27] a HAR system using a waist-mounted smartphone. It used a modified SVM with fixed-point arithmetic prediction aiming to obtain a fast implementation more suitable for battery-constrained devices.

More recently, online smartphone-based HAR systems have been proposed. A Nokia smartphone was used in [28] for the online recognition of 6 activities. In [29], Fuentes et al. presented an online motion recognition system using a smartphone with embedded accelerometer which classified 4 BAs through a One-vs-One (OVO) SVM approach. In the same way, the work presented in [30] used an Android smartphone with an embedded accelerometer for the online classification of 4 activities. It also allowed the adaptation of the learned model for new users by gathering activity samples through a predefined activity protocol.

1.2. Dealing with transitions in HAR systems

In the design of HAR systems there are still some issues that need to be addressed. In most approaches, transitions between activities are usually disregarded since their incidence is generally low and duration is short when compared against other activities. This is pointed out by Lara et al. in [7], nevertheless, the validity of this assumption is application-dependent. Even if the detection of transitions is not required, it is important to notice them in applications where multiple tasks are performed in a short period of time. For instance, activity monitoring during rehabilitation practices, fitness/gymnasium workout activities, equipment assembly and house cleaning. Fluctuations in the prediction during transitions affect the performance of the recognition system if not dealt with properly. A second issue considers that the activities carried out by people are, in real-life situations, more than the ones learned by any HAR system [31]. The remaining activities, unknown to the system, are usually matched as any of the available ones, and this leads to misclassifications. Instead, a better approach would allow the system to tell that it does not predict any of its available classes when its confidence is below certain level. Dealing with these Unknown Activities (UAs) allows more functional HAR systems for a variety of applications.

A number of systems have focused on the detection basic activities and short events. Khan et al. [32] studied 7 basic activities and 7 transitions using three Artificial Neural Networks to separately detect static, dynamic and transitory states. Applications with a large number of classes such as this can give rise to an increase in the false negative rate, especially when the main interest is only on a subset of activities (e.g. basic activities, rather than transitions). In [33], Zhang et al. proposed an offline HAR system that combines basic activities with a joint class of various postural transitions for daily monitoring applications. In [34], Salarian et al. detected *sit-to-stand* and *stand-to-sit* transitions for better distinguishing between *standing* and *sitting*. This was achieved through a fuzzy logic classifier which required for this task, past and future transition information.

Only a few works on HAR have targeted how the presence of transitions between activities impacts system performance. Rednic et al. in [35] performed posture classification of activities for ordnance disposal operations using a multi-accelerometer BSN, while considering the effects of postural transitions in their system using a weighted voting filter in order to improve the classification accuracy of postures by 1%. Moreover, erroneous fluctuations of predicted activities on a classifier can be also dealt in a similar approach. One example of this is also found in [36] where a method called *statistical-hist* was proposed. It processed historical variations of the classifier BA predictions using a voting strategy for spurious classification pruning.

In this work, we propose the TAHAR system architecture for the recognition of human activities using smartphones. It targets the classification of basic activities in real time and pervasively while addressing issues regarding transitions and unknown activities. It offers a flexible and interoperable approach that allows to incorporate new elements (e.g. inertial sensors) into the system and provides an easily exportable output to other ambient intelligent systems that require activity information. Two implementations of the architecture are explored. They differ in the way they deal with transitions that occur in between the activities of interest. In the first case, transitions are treated as unknown activities. Therefore, they are not learned by the machine learning algorithm. Instead, in the second case, transitions are learned by the algorithm as an extra class [33].

We validate the proposed architecture with three case studies: for the most part, we exploit the SBHAR dataset that we have generated from experiments on a group of 30 subjects that performed six locomotion activities while they were carrying a smartphone on their waist. This dataset also contains transition