Pattern Recognition Letters 34 (2013) 2033–2042

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

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

The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition

Ricardo Chavarriaga [a,*], Hesam Sagha [a], Alberto Calatroni [b], Sundara Tejaswi Digumarti [a], Gerhard Tröster [b], José del R. Millán [a], Daniel Roggen [b]

[a] *Chair on Non-Invasive Brain-Machine Interface, Center for Neuroprosthetics, School of Engineering, École Polytechnique Fédérale de Lausanne, Station 11, 1015 Lausanne, Switzerland*

[b] *Wearable Computing Laboratory, ETH Zürich, Gloriastrasse 35, 8092 Zürich, Switzerland*

ARTICLE INFO

Article history:
Available online 3 January 2013

Keywords:
Activity recognition
Machine learning
Body-sensor networks
Performance evaluation
Metrics
ROC analysis

ABSTRACT

There is a growing interest on using ambient and wearable sensors for human activity recognition, fostered by several application domains and wider availability of sensing technologies. This has triggered increasing attention on the development of robust machine learning techniques that exploits multimodal sensor setups. However, unlike other applications, there are no established benchmarking problems for this field. As a matter of fact, methods are usually tested on custom datasets acquired in very specific experimental setups. Furthermore, data is seldom shared between different groups. Our goal is to address this issue by introducing a versatile human activity dataset recorded in a sensor-rich environment. This database was the basis of an open challenge on activity recognition. We report here the outcome of this challenge, as well as baseline performance using different classification techniques. We expect this benchmarking database will motivate other researchers to replicate and outperform the presented results, thus contributing to further advances in the state-of-the-art of activity recognition methods.

1. Introduction

Multiple applications require human activity recognition systems ranging from health care and assistive technologies (Tentori and Favela, 2008) to manufacturing (Stiefmeier et al., 2008) or gaming. New sensing technology allows the use of multimodal setups involving on-body, object-placed or ambient sensors. From a machine learning perspective activity recognition is a challenging problem as it typically deals with high-dimensional, multimodal streams of data characterised by a large variability (e.g. due to changes in the user's behaviour or as a result of noise). Moreover, real-life deployments are required to detect when no relevant action is performed (i.e. *Null* class) (Stiefmeier et al., 2008). For these reason robust methods are required tackling issues ranging from the feature selection and classification (Preece et al., 2009), to decision fusion and fault-tolerance (e.g., Chavarriaga et al., 2012; Sagha et al., 2011b; Zappi et al., 2007).

However, the comparison of different approaches is often not possible due to the lack of common benchmarking tools and datasets that allow for replicable and fair testing procedures across several research groups. Currently, each research group assess the performance of their algorithms using experimental setups specially conceived for a narrow purpose. This contrasts with other application fields where publicly available datasets allow the independent assessment of different algorithms in the very same conditions. This is common practice in the machine learning community covering applications like computer vision, biometrics or speech recognition (e.g., the UCI machine learning repository). Furthermore, methods are often evaluated in the frame of open competitions or challenges providing a fair comparison of them. Recent examples of these competitions have focused on computer vision, bioinformatics, or brain–computer interfaces (e.g., Everingham et al., 2010; Guyon and Athitsos, 2011; Blankertz et al., 2006).

Considering this, we believe that there is a need for publicly available databases on human activity recognition. This will allow the replication of the testing procedures for different approaches. Ideally, these databases should reflect the variability of real-world activities, and be flexible enough to emulate different experimental setups and recording modalities. In order to address these issues we recorded a large recording of realistic daily life activities in a sensor-rich environment, i.e. The *Opportunity* activity recognition dataset (Roggen et al., 2010).

We used a subset of this dataset to organise an open challenge where different classification methods contributed by different research groups were compared. The selected benchmarking dataset is publicly available and contains recordings of on-body sensors

* Corresponding author.
E-mail address: ricardo.chavarriaga@epfl.ch (R. Chavarriaga).

0167-8655/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2012.12.014

while subjects perform activities of daily living, ranging from simple motion primitives to complex gestures. Thus, this dataset offers a rich playground to assess methods for sensor selection, feature extraction, classifier calibration and adaptation, multimodal data fusion, automatic segmentation, among others. It also captures the challenges common to many other activity recognition scenarios. Thus, methods proved to be robust on this dataset can likely be successfully translated to other activity recognition problems.

This paper provides an overview of the Opportunity dataset (Section 3) and the activity recognition challenge (Sections 4 and 5). Furthermore, we also compare different recognition systems using four well-known classification techniques, namely *k*-NN, NCC, LDA and QDA classifiers (Section 6). The performance of these methods and the contributed techniques are then reported (Section 7), followed by the conclusions (Section 8).

2. Related work

Several datasets for activity recognition are currently available. However, they tend to be specific to an activity recognition purpose. Widely popular in the pervasive computing community is the *PlaceLab* dataset. It contains ambient and object sensing of subjects recorded over several days (up to a week) in an environment with multimodal sensors (Intille et al., 2006). Its main strength is to provide long-term recordings although it does not include a high number of activity instances. Another dataset recorded by van Kasteren et al. (2008) features longer recordings (month-long) but fewer sensors. It uses digital or binary sensors (e.g. reed switches) to record interactions with objects of interest, but does not include information about modes of locomotion or body posture. The *Darmstadt routine dataset* – used to study unsupervised activity pattern discovery (Huynh et al., 2008) – is a long recording from body activity collected by the Porcupine system (Van Laerhoven et al., 2006). The *TUM Kitchen dataset* focuses on video-based activity recognition (Tenorth et al., 2009), and also contains RFID and reed switch data, but it does not include on-body sensors. A more recent database focuses on fine grained human activities in the kitchen, but it is more suitable for computer vision techniques (Rohrbach et al., 2012). As it can be seen, these databases – although useful – are limited due to the reduced number of recorded sensors and activity instances as well as the fact that they were conceived for very specific purposes.

An exceptional effort to collect a large scale human activity corpus, termed *HASC corpus*, was led by Nagoya University (Kawaguchi et al., 2011). It is the result of a collaboration among 20 teams that gathered data from 116 subjects. Data from each subject contains a set of six activities (stay, walk, jogging, skip, stair-up and stair-down) recorded with a single commercially available accelerometer. There was no constraint on the location of the sensor. The main strength of this corpus lies in the large amount of subjects available. However, the fact that there is only one sensor that may be located at any place, effectively limits its use.

In addition, most of the previous activity recognition challenges have focused on isolated gestures using video (Guyon and Athitsos, 2011). One exception to this is the open contest organised at the 2011 Body sensor network conference (<http://bsncontest.org>) (Giuberti and Ferrari, 2011). In this contest, the organisers provided three datasets provided from different groups. Datasets differ in the number, arrangement and type of sensors used, as well as the number of subjects. Participants were asked to provide methods for the recognition of several actions mainly focusing on modes of locomotion.

The lack of more general databases can be explained by the difficulty to conceive and record a dataset that reflects the complexity and variability of daily life situations. Moreover, proper comparison of machine learning techniques requires these datasets to provide a reasonable amount of instances for the different recorded actions and to include several subjects in order to allow the assessment of inter-subject variability. In addition, if the database is used to emulate changes in the sensor network, then activities should be recorded by a large and diverse set of sensors. These aspects were taken into account for the database described in the next section.

3. The Opportunity dataset

The challenge was based on a subset of the *Opportunity activity recognition dataset* (Roggen et al., 2010), a dataset of complex naturalistic activities with a particularly large number of atomic activities (more than 27,000) collected in a sensor rich environment (c.f. Fig. 1). Overall, it comprises recordings of 12 subjects using 15 networked sensor systems, with 72 sensors of 10 modalities, integrated in the environment, in objects, and on the body (Fig. 1(b)). These characteristics make it well suited to benchmark various activity recognition approaches (Sagha et al., 2011a). An illustrative video of the recording and database is provided as supplementary material.

We designed the activity recognition environment and scenario to generate many activity primitives, yet in a realistic manner. We purposely did not record human behaviour in daily life to favour

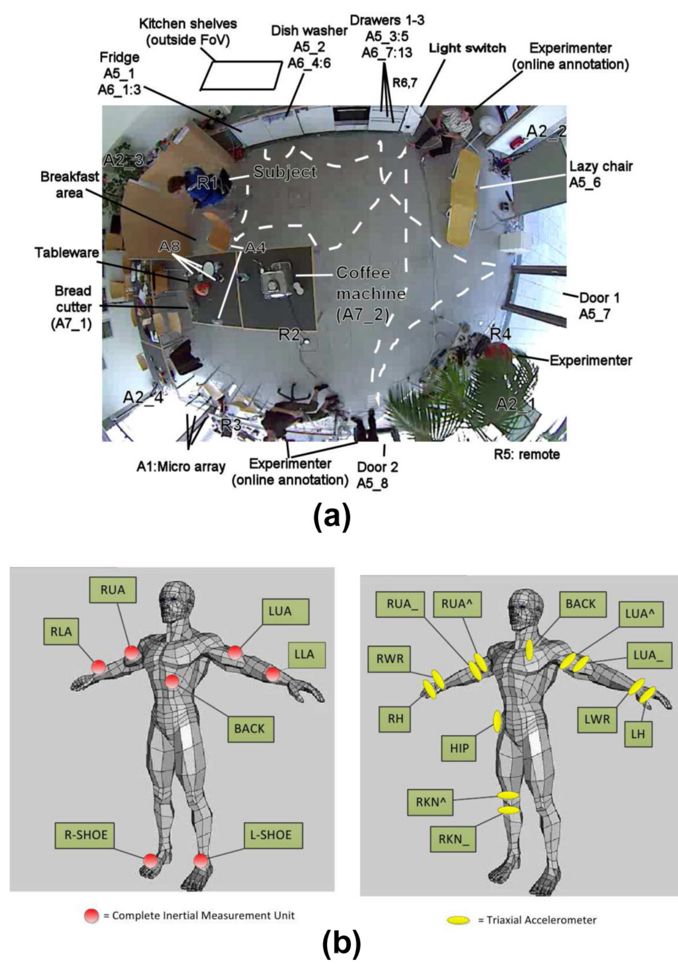

Fig. 1. Opportunity dataset setup. (a) Top view of the recording room. The dashed line shows a typical user trajectory in the drill run. (b) On-body sensors used for the activity recognition challenge (red: IMU sensors; yellow: 3-axis accelerometers). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)