Neurocomputing 181 (2016) 108–115

Contents lists available at ScienceDirect

Neurocomputing

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

From action to activity: Sensor-based activity recognition

Ye Liu *, Liqiang Nie, Li Liu, David S. Rosenblum

School of Computing, National University of Singapore, Republic of Singapore

ARTICLE INFO

Article history:
Received 15 March 2015
Received in revised form
25 June 2015
Accepted 27 August 2015
Available online 10 November 2015

Keywords:
Activity recognition
Temporal pattern mining
Sensor-generated data
Discriminative feature extraction

ABSTRACT

As compared to actions, activities are much more complex, but semantically they are more representative of a human's real life. Techniques for action recognition from sensor-generated data are mature. However, few efforts have targeted sensor-based activity recognition. In this paper, we present an efficient algorithm to identify temporal patterns among actions and utilize the identified patterns to represent activities for automated recognition. Experiments on a real-world dataset demonstrated that our approach is able to recognize activities with high accuracy from temporal patterns, and that temporal patterns can be used effectively as a mid-level feature for activity representation.

1. Introduction

We are living in the era of wearable devices and environmental sensors. In the past decades, great effort in both research and industry has been dedicated to sensor-based human activity recognition, due to its wide application in many fields [1–5]. Consider an application scenario in healthcare as an example. Caregivers use sensors to track and analyze the Activities of Daily Living (ADL) of elderly people in order to provide them with proactive assistance [6], and this could greatly facilitate elderly people's activity tracking and management. Another application scenario is context-aware music recommendation [7], which senses a user's current activity (the context) and recommends a specific type of music suitable for the activity.

Even though great success has been achieved in previous approaches, they mainly worked towards recognition of simple *actions* [1,2,7–11]. In our work, actions are defined as primitives that fulfill a function or simple purpose, such as *walking*, *jumping*, or *opening the fridge*. Ravi et al. [9] designed a classifier to distinguish eight actions, namely *standing, walking, running, stairs up, stairs down, vacuuming, brushing* and *situps*. In another work, Kwapisz et al. [2] took six actions into consideration, *walk, jog, up, down, sit* and *stand*. Rather than investigating *action* recognition, we move one step forward and target the more difficult problem of *activity* recognition. An activity consists of a pattern of multiple actions over time. Typical examples include *cooking*, *basketball playing*, and *coffee time*.

Intuitively, activities are much more complex than actions, but they are also more representative of a human's real life. This is because humans frequently perform multiple actions simultaneously in a variety of temporal combinations. Such temporal combinations produce higher-level semantic activities. Technically, within one activity, the temporal relatedness among actions may be in the form of *sequential, interleaved* or *concurrent* patterns. This makes activity recognition non-trivial. However, few existing sensor-based action recognition methods comprehensively explore the temporal relatedness among actions. They regard the extracted time and frequency features from sensor data as independent data points. Hence they are not applicable or extensible to activity recognition.

It is worth mentioning that various approaches in other research domains have been proposed to handle *sequential, interleaved* and *concurrent* temporal relatedness [12,13]. Most of these approaches are probabilistic methods and handle one or two temporal relations. However, these probabilistic methods are based on the strict assumption that sensor data or feature observations are independent. As such, they lack flexibility in recognizing different levels of activities, since it may not be possible to construct a model incrementally using such methods. Moreover, efficiency is another drawback of these approaches, as the time of complexity grows exponentially with the number of involved actions.

Inspired by discriminative pattern learning in computer vision [14,15] and temporal pattern-based classification in other domains [16–18], we present an efficient pattern-mining algorithm to identify temporal patterns of actions and utilize such temporal signatures to represent activities for further classification and recognition. In particular, we presume that each pattern is a set of temporally interrelated actions, and that each unique activity can

* Corresponding author.
E-mail addresses: liuye@comp.nus.edu.sg (Y. Liu), nieliqiang@gmail.com (L. Nie), dcsliuli@nus.edu.sg (L. Liu), david@comp.nus.edu.sg (D.S. Rosenblum).

http://dx.doi.org/10.1016/j.neucom.2015.08.096
0925-2312/

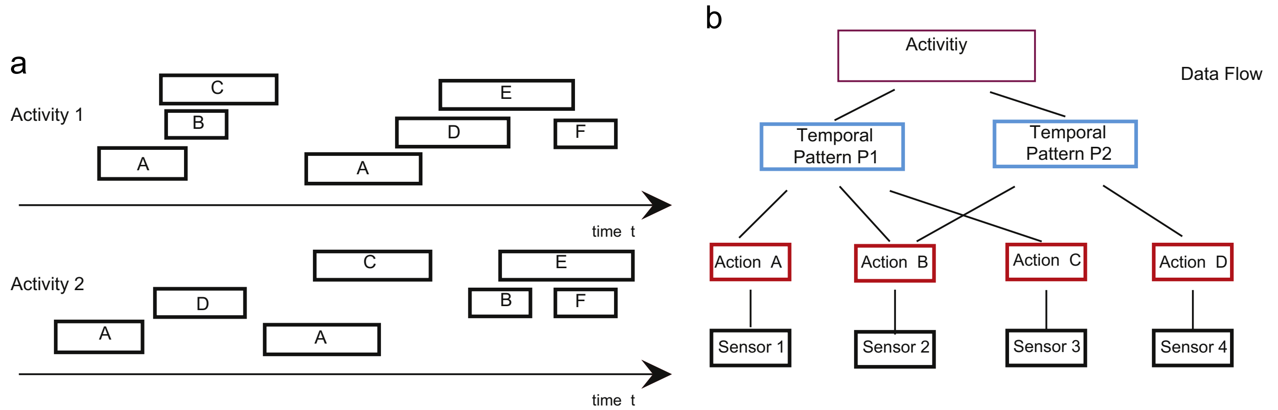

Fig. 1. Examples of activities and overview of our algorithm. In (a), the A–F represent the actions, and each activity consists of a set of temporal actions. The temporal patterns can be a subset of any actions that encode the temporal relationship among actions. For instance, (D before C), (B overlap E) are two possible temporal patterns in Activity 2. The overview of our algorithm is illustrated in (b).

be characterized by a set of patterns. Our algorithm automatically discovers the frequent temporal patterns from these action sets and uses these mined patterns to characterize the high-level activities. Fig. 1 presents an overview of our approach.

The main contributions of our work are threefold:

- As far as we are aware, this is the first work to recognize complex high-level activities from simple low-level actions in the sensor domain. Such recognition can greatly facilitate human activity tracking and management.
- We present a novel approach to identify temporal patterns for activity representation, which are able to capture the signatures of sequential, interleaved and concurrent actions.
- We empirically demonstrate that the accuracy of our approach is high and that it stabilizes at the $\{1,2,3\}$-*pattern* (as defined later), which avoids the curse of feature dimensionality.

The rest of paper is organized as follows. Section 2 discusses related work. Sections 3 and 4 formulate the problem and details, respectively, of the temporal pattern mining algorithm. In Section 5, we present the classification of activities via temporal patterns. Section 6 presents our experimental results. Section 7 concludes the paper and discusses our plans for future work.

2. Related work

Many approaches have been proposed for human activity recognition, and they can be divided roughly into two categories: sensor-based [1–4,6–9,19–21] and vision-based [22–26]. Vision-based methods employ cameras to detect and recognize activities using several different techniques from computer vision, such as object segmentation, feature extraction and feature representation. In contrast, sensor-based methods collect input signals from wearable sensors attached to a humans body, such as accelerometers, gyroscope or physiological signals collectors, and then apply several time-series techniques to give the analytical results. We first briefly discuss vision-based activity recognition methods and then focus our discussion on wearable and environmental sensor-based activity recognition, which is consistent with the concept of *Ubiquitous Sensing* [21,27]. Some of the most important work in sensor-based activity recognition has been recently surveyed by various authors [1,5,21]. We finish the section with a discussion of additional pattern-mining approaches related to our approach.

2.1. Vision-based action recognition

Vision-based methods [24–26] employ video cameras to detect human actions and gestures from video sequences. This is an important area of computer vision research and has a wide range of applications in video surveillance systems, patient monitoring systems and various human–computer interaction systems [24,28–31]. Recognizing actions from video sequences involves some important stages, including pre-processing of images or space–time volume video data, feature extraction with respect to actions, and action modeling based on the extracted features.

The pre-processing step tries to improve the overall performance of systems by removing spurious noise and insignificant features from the video inputs [32], and then feeds higher-level modules with initial outputs, such as human segmentations with body parts [33,34].

The main purpose of feature extraction is to find a set of characteristics such as shape, silhouette, colors, poses and body motions that can properly describe the human actions and then represent the actions by a set of descriptive features. These descriptive features include space–time volume (STV) [35], Lucas–Kanade–Tomasi (LKT) features [36,37], scale-invariant feature transform (SIFT) [38], histogram of oriented gradient (HOG) [39], histogram of optical flow (HOF) [40,41], shape-based features [42] and appearance-based features [43,44,45], which can capture the characteristics of human motions from various aspects.

Based on the extracted features, an action detection or recognition model is applied to detect, recognize or locate the various human actions from videos. These action models can be generally categorized into generative models and discriminative models. Generative models (e.g., HMM [46] and DBN [47]) use probabilistic state-based models to model and recognize human actions, while discriminative models (e.g., SVM [40], RVM [48] and ANN [49]) utilize static classifiers to classify the actions directly.

Although vision-based methods have contributed significantly to research on human activity recognition, they have several limitations. First, they require cameras that are fixed to predetermined points of interest. In addition, the extraction of descriptive features from the captured images or videos requires extensive computational resources, which greatly limits their usage in real-time applications. Moreover, privacy is another important issue, since not everyone is willing to be tracked, monitored and recorded by cameras all the time. Hence, these vision-based techniques provide limited applicability to the problem of activity recognition.