Pervasive and Mobile Computing 6 (2010) 342–350

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

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

Wearable sensor activity analysis using semi-Markov models with a grammar[☆]

O. Thomas [a,*], P. Sunehag [a], G. Dror [b], S. Yun [c], S. Kim [c], M. Robards [a], A. Smola [d], D. Green [e], P. Saunders [e]

[a] Locked Bag 8001, NICTA, Canberra, 2601, ACT, Australia

[b] School of Computer Science, The Academic College of Tel-Aviv-Yaffo, Tel Aviv 61083, Israel

[c] LG Semicon Hall 2106, KAIST, 373-1, Guseong-dong, Yuseong-gu, Daejeon, 305-701, Republic of Korea

[d] Yahoo! Research, Santa Clara, 95050 CA, USA

[e] Department of Physiology, Australian Institute of Sport, Belconnen, 2616, ACT, Australia

ARTICLE INFO

Article history:
Received 9 November 2008
Received in revised form 8 September 2009
Accepted 8 January 2010
Available online 15 January 2010

Keywords:
Activity recognition
Semi-Markov models
Machine learning
Human performance
Accelerometer

ABSTRACT

Detailed monitoring of training sessions of elite athletes is an important component of their training. In this paper we describe an application that performs a precise segmentation and labeling of swimming sessions. This allows a comprehensive breakdown of the training session, including lap times, detailed statistics of strokes, and turns. To this end we use semi-Markov models (SMM), a formalism for labeling and segmenting sequential data, trained in a max-margin setting. To reduce the computational complexity of the task and at the same time enforce sensible output, we introduce a grammar into the SMM framework. Using the trained model on test swimming sessions of different swimmers provides highly accurate segmentation as well as perfect labeling of individual segments. The results are significantly better than those achieved by discriminative hidden Markov models.

1. Introduction

Recently, motivated by ever increasing levels of interaction between humans and computing devices, context aware computing has received extensive research interest. In this broad field, researchers have been developing models and systems for predicting a system's location and status using context aware computing methods [1]. One important facet of this work, and the focus of the research presented in this paper, is that of human activity recognition. The classification of a person's activity, which can be performed using, for example, computer vision [2,3], has significant potential in diverse application domains such as patient care, chronic disease management and promotion of lifelong health and well-being for the aging population [4]. In more recent times, motivated by developments in the underlying technology, the use of *wearable sensors* for activity recognition has received considerable interest. For example, their use has been investigated in the monitoring of rehabilitation of patients in post-ambulatory conditions [5] and in monitoring Parkinson's disease patients [6].

Wearable sensors have also been investigated extensively for the purpose of gait event detection [7–12]. Gait event detection involves detection – often in real time – of the phases of gait during walking, which has been of considerable interest in improving the quality of life in children with cerebral palsy [9,11], and in functional electrical stimulation (FES)

☆ The authors acknowledge the staff and athletes at the AIS who contributed to this research.

* Corresponding author.
E-mail addresses: owen.thomas@nicta.com.au (O. Thomas), peter.sunehag@nicta.com.au (P. Sunehag), gideon@mta.ac.il (G. Dror), yunsungrack@kaist.ac.kr (S. Yun), leehwiso@kaist.ac.kr (S. Kim), matthew.robards@nicta.com.au (M. Robards), alex@smola.org (A. Smola), daniel.green@ausport.gov.au (D. Green), philo.saunders@ausport.gov.au (P. Saunders).

1574-1192/$ – see front matter © 2010 Published by Elsevier B.V.
doi:10.1016/j.pmcj.2010.01.002

Table 1
Training session descriptions for each of the tree swimmers. Each swimmer performed three training sessions each consisting of eight laps, with each lap containing only one stroke type. Here we show the breakdown of each session by the number of laps of each stroke type performed in that session.

Session name	Butterfly	Backstroke	Breaststroke	Freestyle
Medley	2	2	2	2
Freestyle	0	0	0	8
Freestyle/Backstroke	0	4	0	4

walking [7,8]. Of particular interest in this work is the use of accelerometers for the purpose of human activity recognition [7], and the use of machine learning techniques such as the support vector machine — used in [12] for separation of gait phases.

Aside from these applications in health care, sensors have been used in sports and performance arts. This work discusses using wearable sensors technology for the monitoring and analysis of training sessions of professional athletes [13].

The majority of activity recognition research has focused on the task of *recognition*. The sequential data used was either manually segmented or generated in a manner where *segmentation* is not required. For instance, each video sequence represented a single type of activity. In real world scenarios, however, segmentation and recognition are intertwined and therefore they need to be addressed jointly.

In this paper we consider the problem of activity recognition via a combined segmentation and labeling of sports accelerometer data. Given 3-dimensional accelerometer data from a sports training session we describe a method that simultaneously segments the data into atomic actions and labels each action with high accuracy. We further show that the method is robust across multiple athletes; a model trained particular to one athlete performs accurately when applied to another. To perform the labeled segmentation we use a variant of the recent and popular semi-Markov Conditional Random Fields [14] framework. In comparing our results to those achieved by a HMM we demonstrate the superiority of semi-Markov models (SMM) for solving problems of segmentation and by extension, activity recognition.

The paper is structured as follows: In the subsequent section we discuss the data available and segmentation problem in more detail and introduce the HMM and SMM segmentation systems. We then describe the experimental procedure and present the results from experiments comparing the predictive capacity of the HMM and SMM systems. Finally, we finish the paper with a discussion of the two approaches and introduce future work in this area.

2. Method

In this section we will formulate our task, describe our data, introduce the concept of a grammar in this context, provide definitions of two machine learning methods together with the procedures for how to perform training and testing with them. This will provide the components for our general approach to creating an activity recognition system. The approach can be described as follows:

1. Collect video of, and body-worn sensor data from, a person performing the activities of interest.
2. Decide what the exact states of interest are.
3. Formulate a grammar by deciding what state can reasonably follow which other state.
4. Manually create a labeled segmentation of the training data using the video
5. Choose a loss function $\Delta(y, \hat{y})$ which says how much worse a given labeled segmentation y is compared to the truth $\hat{y}$.
6. Train a semi-Markov model using the annotated training data, the grammar and the loss function Δ
7. Now the generated parameters can be used for prediction on new data

2.1. Data

Swimming sensor data was collected from three eight-lap training sessions from three elite female swimmers at the Australian Institute of Sport, in Canberra, Australia. The three training sessions were representative of their training regime, each consisted of eight laps with strokes consistent across a lap, but potentially alternating between laps (Table 1). Each lap consisted of one of four stroke types: Butterfly, backstroke, breaststroke and freestyle. The data was not evenly distributed (Table 1), for example each swimmer performed 10 laps of freestyle to two laps of butterfly. In each session, 3-dimensional acceleration data was sampled at 100 Hz via a Catapult Innovations minimaxX accelerometer attached to the swimmers back. Each session of eight laps comprises approximately 33 000–34 000 samples, which results in approximately 100 000 samples per swimmer. The device is placed in a special pocket on the swim suit that has been designed for the device. This makes the device well attached and, therefore, decreases the noise. To train and evaluate the performance of our method, the data of each was labeled by comparing the sensor read outs to a video of the swimmer in that session. In Fig. 2 we show example accelerometer data used in training.

In order to develop the HMM and SMM formalisms in the following sections we now describe the data in more precise terms. We assume we are given a sequence of observations x of length N, such that each observation $x(t)$ specifies a 3-tuple $x(t)[0..2]$ containing the 3-dimensional acceleration at time t. Associated with x we assume a labeled segmentation y, which, given x, is the quantity we seek to predict. We assume both are generated from a joint distribution $\Pr(x, y)$. To keep notation simple we only present a single sequence x and labeled segmentation y pair, extensions to multiple sequence / segmentation pairs are trivial. We model a labeled segmentation y, for a given observation sequence x, as a sequence of segments, where each segment is represented as a 2-tuple specifying the label of the segment and the coordinate in x at which it begins.