Pervasive and Mobile Computing 16 (2015) 251–267

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

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

Multi-scale Conditional Random Fields for first-person activity recognition on elders and disabled patients

Kai Zhan [a,*], Steven Faux [c,b], Fabio Ramos [a]

[a] *School of Information Technologies, University of Sydney, Sydney, Australia*
[b] *Faculty of Medicine, University of New South Wales, Sydney, Australia*
[c] *Sacred Heart Rehabilitation Service, St Vincent's Hospital, Sydney, Australia*

ARTICLE INFO

Article history:
Available online 5 December 2014

Keywords:
Activity recognition
Feature classification
Graphical model
First-person
Feature extraction
Computer vision

ABSTRACT

We propose a novel pervasive system to recognise human daily activities from a wearable device. The system is designed in a form of reading glasses, named 'Smart Glasses', integrating a 3-axis accelerometer and a first-person view camera. Our aim is to classify subject's activities of daily living (ADLs) based on their vision and head motion data. This ego-activity recognition system not only allows caretakers to track on a specific person (such as disabled patient or elderly people), but also has the potential to remind/warn people with cognitive impairments of hazardous situations. We present the following contributions: a feature extraction method from accelerometer and video; a classification algorithm integrating both locomotive (body motions) and stationary activities (without or with small motions); a novel multi-scale dynamic graphical model for structured classification over time. In this paper, we collect, train and validate our system on two large datasets: 20 h of elder ADLs datasets and 40 h of patient ADLs datasets, containing 12 and 14 different activities separately. The results show that our method efficiently improves the system performance (F-Measure) over conventional classification approaches by an average of 20%–40% up to 84.45%, with an overall accuracy of 90.04% for elders. Furthermore, we also validate our method on 30 patients with different disabilities, achieving an overall accuracy up to 77.07%.

1. Introduction

The challenges associated with an ageing population and pressures on carers and nursing services, are opening unprecedented opportunities for pervasive computing. Such systems can improve the quality of life of the aged particularly by making daily activities of living safer and easier to complete. However, most of the current approaches of activity tracking still rely on human, for which can be very costly and time-consuming. For example, at home, disabled subjects generally require supervision of caretakers. In a hospital, patients are often monitored and cared by health professionals. In other cases, subjects are asked to manually update and report their activities by themselves. Both cases have significant inefficiencies in terms of cost, accuracy, scope, coverage and privacy.

A central requirement of pervasive systems is to automatically recognise human activities over time, warning patients of hazardous situations, or reinforcing home-based activities and therapies. According to their activities, the system can be pre-programmed for assistance purposes. For example, patients with memory loss or cognitive disorders can be reminded

* Corresponding author.
E-mail addresses: kai.zhan@sydney.edu.au (K. Zhan), sfaux@stvincents.com.au (S. Faux), fabio.ramos@sydney.edu.au (F. Ramos).

http://dx.doi.org/10.1016/j.pmcj.2014.11.004
1574-1192/

Fig. 1. Senior patients wearing the Smart-Glasses prototype.

to take medicines after having meals or to pick up their walking stick when venturing outdoors. In addition, it is desirable that monitoring solutions also provide online interfaces for patients and carers to remotely access the patients' status, and to derive better treatments in a more effective manner.

Activity recognition is a relatively simple task for observers when watching the activities of other people. With years of social experience, the capacity of understanding and expressing the activities and intentions of others is notable. The recognition of activities by humans involves a series of tasks, from sensing to interpreting which can be complicated and challenging for an automatic system. For an instance, a man in a sitting-posture, on an observer perspective, might be watching TV, reading or just resting. Context generally help with this distinction. However, for a wearable device, it is a difficult challenge to tell what exactly the person is doing despite increasing amounts of sensory data. The main reason for this is that humans learn to interpret sensory data from past experiences, making new judgements significantly easier, in a sensing and learning process. The interpretation of sensory data is accomplished naturally using either the environment or other people to provide feedback. Ultimately, we would also prefer to have a machine with the same capabilities, i.e., automatically learning to interpret sensory data from past experiences and feedback received. The main objective of automatic activity recognition systems is to develop such algorithms combining elements of statistical learning with sensor technologies.

This paper focuses on automatic activity recognition and introduces a novel 'Smart Glasses' system to recognise complex daily living activities, which can be categorised into two major groups in terms of their motion magnitude: *Locomotive* and *Stationary*. A locomotive activity is defined as an activity involving high energy, with specific body movements, such as walking or climbing stairs. A stationary activity involves less or no motion, such as reading books or watching TV.

In this paper, we introduce an automatic activity recognition system integrating accelerometers and a first-person view camera embedded in conventional glasses. The current prototype consists of a smart phone (Android OS) attached on top of safety goggles as shown in Fig. 1. The device collects videos and 3-axis acceleration data. Both are synchronised and collected in parallel. With this data we develop our classification approach that includes the following contributions:

Feature extraction: We carefully select a number of activities following healthcare professionals' directives. Some of the activities have identical static postures. This makes features from the accelerometer less important. Therefore, in this paper, we use the video motion feature as a complementary element. This allows the system to track motion flow from consecutive egocentric images in order to improve the system performance. We design separate feature extraction algorithms for both accelerometer and video data, detailed in Sections 3.2 and 3.3.

Feature integration: From a series of experiments, we select the best classifier and settings for each set of features (from accelerometers and camera), and separate the local classification task into two categories, each associated with weighting parameters obtained during the training process. This allows the model to combine the best set of feature for the different activities.

Multi-scale graphical model: We develop a structured classification approach based on Conditional Random Fields (CRFs) to capture the multi-scale context in a sequence of activities. This model can help to predict user's activities at different temporal scales even when the local classification is significantly noisy or ambiguous. For example, when sitting, reading or drinking, there is a period with no motion features being detected from both accelerometer and camera features but the activity can be recognised from the context.

System validation on elders and patients: At the end of this study, we validate our system's performance on two groups: healthy elderly people and a mixed patient population with recent disabilities. The healthy elderly group includes 5 healthy subjects with an average age of 68 years old. The disabled population group of 30 people, ranged in age from 34 to 89 years old, each is suffering from a variety of disabilities which were loosely categorised as musculoskeletal, neurological and general weakness. Their conditions are detailed in Table 5. We separately evaluate and compare our system performance with conventional approaches.

The paper is organised as follows. We describe related work on automatic activity recognition from wearable and environment sensors in Section 2. A detailed description of our method including feature extraction, feature classification