Expert Systems with Applications 39 (2012) 10873–10888

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

A review on vision techniques applied to Human Behaviour Analysis for Ambient-Assisted Living

Alexandros André Chaaraoui, Pau Climent-Pérez, Francisco Flórez-Revuelta *

Department of Computing Technology, University of Alicante, P.O. Box 99, E-03080 Alicante, Spain

ARTICLE INFO

Keywords:
Human behaviour
Ambient-Assisted Living
Computer vision
Motion analysis
Action recognition
Activity recognition
Activities of daily living (ADLs)

ABSTRACT

Human Behaviour Analysis (HBA) is more and more being of interest for computer vision and artificial intelligence researchers. Its main application areas, like Video Surveillance and Ambient-Assisted Living (AAL), have been in great demand in recent years. This paper provides a review on HBA for AAL and ageing in place purposes focusing specially on vision techniques. First, a clearly defined taxonomy is presented in order to classify the reviewed works, which are consequently presented following a bottom-up abstraction and complexity order. At the motion level, pose and gaze estimation as well as basic human movement recognition are covered. Next, the mainly used action and activity recognition approaches are presented with examples of recent research works. Increasing the degree of semantics and the time interval involved in the HBA, finally the behaviour level is reached. Furthermore, useful tools and datasets are analysed in order to provide help for initiating projects.

1. Introduction

Human Behaviour Analysis—and Understanding—(HBA, HBU) involves a wide range of investigation fields from motion detection and background extraction to expert systems and high-level abstraction behaviour models. This paper targets two purposes: On the one hand, researchers need to categorise existing works assuming a common taxonomy and a clear differentiation basis. On the other hand, as the application areas of these fields grow constantly; stable areas, like Video Surveillance, are covered thoroughly; while other more recent areas, like Ambient-Assisted Living (AAL) and ageing in place at smart home scenarios, present a lack of unifying works and recent state-of-the-art reviews. This makes initiation in these areas difficult, also because of the involvement of a wide variety of pure research areas from artificial intelligence to natural language processing.

For this reason, this paper deals with the state-of-the-art of HBA/HBU from an Ambient Intelligence (AmI) point of view, focusing especially on indoor scenarios and techniques which are designed for AAL purposes. This way, recognition of activities of daily living (ADLs) covers the main interest of this paper. Nevertheless, it is necessary to first face a classification of HBA levels, and to deal with all the necessary previous tasks.

To avoid the common difficulties present in vision-based systems (such as occlusions, view-dependent features, lightning conditions, etc.), occasionally these systems are enhanced with other sensors; mostly binary sensors and RFID labels. Therefore, although vision will be focused on mainly, other complementary sensors involved will be discussed briefly too.

The remainder of this paper is organised as follows: Section 2 goes through taxonomies which are applied by other authors and presents an abstraction, degree of semantics and time-oriented classification which is used in the rest of the paper. Section 3 deals with the lowest level, i.e. pose, gaze and motion estimation. These elements are used as *action primitives* in Section 4 where human actions are recognised based on video data and other sensor data fusion (RFID tags, accelerometers, etc.). Section 5 focuses on activity recognition methods which are of special interest in AAL: ADLs in indoor environments, like cooking and grooming, are recognised with different approaches detailed in that section. Finally, Section 6 deals with behaviour recognition methods that establish the highest degree of abstraction. Section 7 summarises some of the most used datasets and tools in the reviewed works that are available.

2. HBA taxonomies

In this section, different Human Behaviour Analysis taxonomies from some of the most recent and relevant research works are discussed in order to point out differences and converge at a well-defined classification of the works analysed in following sections.

Moeslund, Hilton, and Krüger (2006) defined an action taxonomy which has been adopted in later works and subsequent surveys. From lower to higher degree of abstraction three levels are defined:

* Corresponding author. Tel.: +34 965903681; fax: +34 965909643.
E-mail addresses: alexandros@dtic.ua.es (A.A. Chaaraoui), pcliment@dtic.ua.es (P. Climent-Pérez), florez@dtic.ua.es (F. Flórez-Revuelta).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.03.005

- Basic motion recognition derives in so called *action* or *motor primitives* representing the atomic entities out of which actions are built. Therefore, as stated in Poppe (2010), an action primitive is an atomic movement that can be described at the limb level.
- A set of different or repetitive *action primitives* make up an *action*.
- Involving a larger scale of events, the context of the environment and the interacting objects or humans it is possible to recognise the actual *activity*.

This way, when making a cup of tea, single movements of arms and hands would be *action primitives*; placing the kettle on the stove or grabbing a cup from the cupboard would be *actions*; and finally, the whole process would make up an *activity* as different actions and interaction with several objects are involved.

Although this taxonomy is clearly defined and quite often referenced in HBA-related papers, most researchers use their own taxonomy, as usefulness depends on research goals and application areas. Since this classification is particularly focused on actions, it is difficult to adapt to higher level approaches, where the main targets are ADLs and behaviour analysis.

In Wu, Osuntogun, Choudhury, Philipose, and Rehg (2007), *activities* are defined as the combination of *actions* and *objects*. Whereas actions are recognised by a set of *verbs*, objects or places are recognised by a set of *nouns* which are targets of actions. Instead of recognising the actions, object recognition is tackled in order to infer human activities. Turaga, Chellappa, Subrahmanian, and Udrea (2008) distinguish between actions and activities by defining that activities involve coordinated actions among a small number of humans.

Regarding behaviour analysis, Ji, Liu, Li, and Brown (2008) define behaviours as human motion patterns involving high-level description of actions and interactions. In contrast to Moeslund et al. (2006), dependence on the context of the environment, objects and human interaction are taken into account at the behaviour level. In Monekosso and Remagnino (2010), behaviours are understood as patterns in a sequence of observations of activities or events. Activities such as *cooking*, *eating*, *watching TV* or *no detectable activities*; and events from the environment, emitted by binary sensors installed in smart homes, enable to recognise repeatable patterns and detect anomalies.

In this paper, HBA tasks are classified into *motion, action, activity* or *behaviour* levels regarding the degree of semantics and the amount of time involved in the analysis. Therefore, Fig. 1 shows that both the time frame taken into account and the degree of semantics (DoS) involved in the recognition and classification process grow as we reach a higher level of the pyramid.

At the *motion* level, tasks such as movement detection, and background extraction and segmentation are faced (Hu, Tan, Wang, & Maybank, 2004; Moeslund et al., 2006; Porle, Chekima, Wong, & Sainarayanan, 2009). Using a time frame in units of frames, a lot of research is done in the field of gaze and head-pose estimation (Launila & Sullivan, 2010; Ozturk, Yamasaki, & Aizawa, 2009; Reale, Hung, & Yin, 2010, 2010; Rybok, Voit, Ekenel, & Stiefelhagen, 2010; Shimizu & Poggio, 2003).

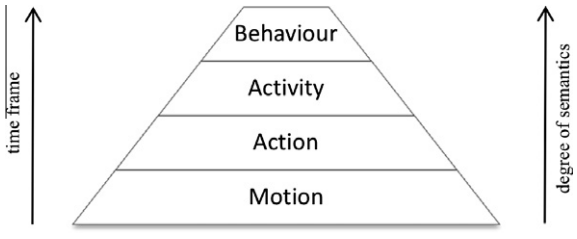

Fig. 1. Human behaviour analysis tasks – classification.

At the *action* level, human motion is not only detected, but also recognised in order to establish what a person is doing or with which objects the person is interacting. In a time frame in units of seconds, simple human activities; like sitting, standing or walking (Bao & Intille, 2004; Chung & Liu, 2008; Liu, Chung, & Chung, 2010; Lester, Choudhury, & Borriello, 2006; Zhou et al., 2008); can be recognised; as well as location changes in indoor and outdoor environments (Nait-Charif & McKenna, 2004).

At the *activity* level, a set of multiple actions is classified in order to understand human behaviour in a time frame from tens of seconds to units of minutes. ADLs are recognised; like *cooking*, *taking a shower* or *making the bed*; as those require tracking and classification of a sequence of actions in a particular order. This way, the sets of actions are understood as activities, where these activities are either the goals or the results of their involving human actions.

At the *behaviour* level, highly-semantic comprehension comes into play. Within a time frame ranging from days to weeks; ways of living, personal habits, and timetables and routines of ADLs can be analysed. At this point, abnormal behaviours and anomalies can be detected, for instance, in order to be able to detect senile dementia prematurely (Karaman et al., 2010; Mihailidis, Boger, Craig, & Hoey, 2008; Mihailidis, Carmichael, & Boger, 2004).

Table 1 summarises the different degrees of semantics considered by the taxonomy, along with some examples. Not only time frame and semantic degree grow at higher levels of this hierarchy, but also complexity and computational cost lead to heavy and slow recognition systems, as each level requires most of the previous level tasks to be done too. For this reason, level abstraction is key in order to analyse only the necessary parts and avoid redundant processes. Human tracking is the best example because it can be approached at least at the first three levels, having different tracking targets and using different kinds of features from the underlying levels. Therefore, tracking will not be discussed in this paper on its own, but tracking approaches from the analysed works will be mentioned when significant.

3. Pose, gaze and motion estimation

Motion recognition is the basis for estimation of human pose and gaze direction (also referred to as focus of attention) and for further HBA tasks. Motion can be seen as a series of poses along the time; the human body is an articulated system of rigid segments connected by joints (as models used in Andriluka, Roth, & Schiele (2009) and Sapp, Toshev, & Taskar (2010) assume); and human motion is often considered as a continuous evolution of the spatial configuration of the segments or body posture (as stated in Li, Zhang, & Liu (2010) and exploited in Andriluka et al. (2009) and Sapp et al. (2010)). On the other hand, the gaze can either be seen as a line in the 3D space or a cone; or, if working only in the horizontal plane (as some works do, as seen later on); a direction and an angle.

3.1. Pose estimation

There are handfuls of previous surveys which analyse and describe "human motion" or "human behaviour understanding" (Hu et al., 2004; Jaimes & Sebe, 2007; Poppe, 2007, 2010; Wang, 2003); earlier works, as is logic, review lower level techniques (e.g. the work by Gavrila (1999)); and later works review also further abstraction levels, approaching more to what is understood as *behaviour* by the taxonomy employed in this review.