Information Sciences 310 (2015) 130–148

Contents lists available at ScienceDirect

Information Sciences

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

A hierarchical neuro-fuzzy architecture for human behavior analysis

Giovanni Acampora [a,*], Pasquale Foggia [b], Alessia Saggese [b], Mario Vento [b]

[a] *School of Science and Technology, Nottingham Trent University, UK*
[b] *Dept. of Information and Electrical Engineering and Applied Mathematics, University of Salerno, 84084 Fisciano, SA, Italy*

ARTICLE INFO

Article history:
Received 9 September 2014
Received in revised form 26 February 2015
Accepted 10 March 2015
Available online 20 March 2015

Keywords:
Behavior understanding
Trajectories analysis
Neuro-Fuzzy Modelling

ABSTRACT

Analysis and detection of human behaviors from video sequences has became recently a very hot research topic in computer vision and artificial intelligence. Indeed, human behavior understanding plays a fundamental role in several innovative application domains such as smart video surveillance, ambient intelligence and content-based video information retrieval. However, the uncertainty and vagueness that typically characterize human daily activities make frameworks for human behavior analysis (HBA) hard to design and develop. In order to bridge this gap, this paper proposes a hierarchical architecture, based on a tracking algorithm, time-delay neural networks and fuzzy inference systems, aimed at improving the performance of current HBA systems in terms of scalability, robustness and effectiveness in behavior detection. Precisely, the joint use of the aforementioned methodologies enables both a quantitative and qualitative behavioral analysis that efficiently face the intrinsic people/objects tracking imprecision and provide context aware and semantic capabilities for better identifying a given activity. The validity and effectiveness of the proposed framework have been verified by using the well-known CAVIAR dataset and comparing our system's performance with other similar approaches working on the same dataset.

1. Introduction

Human behavior analysis (HBA) is a new and hot research topic in the area of computer vision and artificial intelligence that is attracting the attention of several researchers, due its potential applications such as smart video surveillance, ambient intelligence and content-based information retrieval. The main task of HBA frameworks is to provide images and video with a semantic interpretation for trying to bridge the gap between their low-level representation in terms of pixels, and the high-level, natural language description that a human would give about them. In several application domains, this semantic-based approach may support human beings to overcome a well-defined *psychological overcharge* issue [37] that causes a decrease in human capabilities to analyze raw data flows from multiple sources of multimedia information (e.g. video surveillance) [3,13]; indeed, as stated by a study conducted by Security Solutions magazine, "After 12 min of continuous video monitoring, a guard will often miss up to 45% of screen activity. After 22 min of video, up to 95% is overlooked". Different analysis paradigms have been proposed to face this fascinating challenge [6,33,25,31,9,7,20] and, in particular, the approaches based on *bottom-up* and *top-down* analysis can be considered as the most commonly adopted in the current HBA literature. Precisely, bottom-up approaches analyze and interpret a human behavior based on low-level features of the video or image scene and

* Corresponding author.
E-mail address: giovanni.acampora@ntu.ac.uk (G. Acampora).

http://dx.doi.org/10.1016/j.ins.2015.03.021
0020-0255/

try to semantically refine them by using a chain of methodologies from artificial intelligence and machine learning fields; vice versa, top down approaches try to recursively split a whole scene into a collection of semantically-defined events, whose joint analysis may enable the identification of the behaviors of humans populating the scene.

However, in spite of the growing interest in HBA, this research topic is still in its embryonic phase and, indeed, there are several challenges to be faced towards a fully suitable framework for detecting human activities. For instance, several approaches try to analyze a fixed collection of human behaviors and, as a consequence, they lack the scalability feature that could make the system robust enough to deal with novel and not yet considered human activities. Other proposals only analyze pixel-based information, such as trajectories, without taking into account the semantics of events, the scenario where human activity occurs and the dynamic relationships existing among different actors populating this scenario. Moreover, the uncertainty and vagueness that typically characterize human daily activities make HBA frameworks hard to design and develop.

For all these reasons, we propose an effective and scalable HBA architecture capable of understanding human behaviors by analyzing motion information, through a context-awareness framework based on the joint exploitation of an enhanced tracking algorithm [11,10], neural computation and fuzzy logic theory. In particular, similarly to other approaches focusing on different application domains [27,21,22], we introduce a hierarchical architecture based on the integration of different computational intelligence techniques capable of providing benefits both from a *quantitative* and a *qualitative* point of view. In fact, the lower layer of the proposed architecture, based on suitably trained *time delay neural networks (TDNNs)*, analyzes the raw kinematic data obtained by a tracking algorithm, and provides a set of primitive behaviors (e.g. walking, running, stopping, loitering, etc.) performed by human beings moving in the scene under analysis; higher layers, based on a collection of *fuzzy inference engines (FIs)*, act as a context-aware process that semantically enriches the set of primitive behaviors and identify a collection of refined and context depending behaviors (e.g. meeting, walking together, money withdrawing, danger situation, etc.).

This architectural organization provides the resulting HBA system with high levels of scalability. Indeed, at low level, TDNNs can be opportunely re-trained in order to learn novel tracking information and, as a consequence, identify new primitive human behaviors. At high level, the modular approach used for designing the collection of FISs, enables our architecture to enhance its human behavior understanding capabilities by considering semantic relationships corresponding to new human activities. The separation between low and high level behaviors provides our system with improved capabilities from the point of view of behavioral learning. Indeed, only the lowest layer of our architecture, dealing with raw data, is depending upon some environmental features such as the camera view in terms of angle and distance. Consequently, our system can be trained by using less data and a reduced computational effort than other learning-based HBA approaches.

Moreover, another significant benefit provided by the proposed architecture is related to its robustness to tracking imprecision; this feature is due to the exploitation of theories such as fuzzy logic and neural networks that naturally deal with vague or approximate information. The validity and efficiency of the proposed framework have been verified by using the well-known CAVIAR dataset and comparing our system's performance with other similar approaches working on the same dataset.

In the following, we will provide in Section 2 a panoramic summary of related works in the general area of human behavioral analysis; Section 3 introduces our proposal of hierarchical neuro-fuzzy HBA architecture; Section 4 will describe an experimental validation, performed on the standard CAVIAR database, and will discuss the obtained results; finally, in Section 5, we will present our conclusions and will outline some future developments.

2. Related works

The challenging problem of human behavior analysis and understanding has been addressed by a large number of researchers in the last years. However, most of the research activities focused on the recognition of low level human actions [1,34,18] without taking into account how semantic relationships between a human and its surroundings can influence the detection of a complex behavior. For this reason, lastly, different analysis paradigms have been proposed to extend the conventional human behavior understanding paradigms, based on a low-level analysis, with enhanced context-aware capabilities for performing complex behavioral understanding. One of the first work on human behavioral analysis was introduced in 2004 by [19]; they present a framework, based on fuzzy self organizing maps, for learning patterns of object activities in image sequences for anomaly detection and activity prediction, and show that their method obtains a higher efficiency than Kohonen self organizing feature maps. Another pioneering HBA work is presented in [12] where the authors address the problem of learning and recognizing human activities of daily living by introducing the Switching Hidden Semi-Markov Model (S-HSMM), a two-layered extension of the hidden semi-Markov model (HSMM) for the modeling task. The aim of the HBA framework presented in [2] is to classify the speed of moving objects as normal or abnormal in order to detect anomalous events, taking into account the object class and spatio-temporal information such as locations and movements; this approach is based on a fuzzy knowledge base automatically generated through a learning algorithms based on a 3D description of the environment in which the system is installed. The research work presented by [16] deals with the idea of jointly modeling simple and complex behaviors to report local and global human activities in natural scenes; this system uses a state machine approach for activity representation incorporating knowledge about the problem domain in order to provide the systems with additional context-aware capabilities. In [29] the authors decompose a complex behavior pattern