Signal Processing 89 (2009) 1723–1738

Contents lists available at ScienceDirect

Signal Processing

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

Detecting abnormal human behaviour using multiple cameras

Panagiota Antonakaki *, Dimitrios Kosmopoulos, Stavros J. Perantonis

Computational Intelligence Laboratory, Institute of Informatics and Telecommunications, National Center for Scientific Research "Demokritos", Athens, Greece

ARTICLE INFO

Article history:
Received 6 August 2008
Received in revised form
9 March 2009
Accepted 11 March 2009
Available online 5 April 2009

Keywords:
Behaviour understanding
Trajectory
Hidden Markov Model
Support vector machine
Homography

ABSTRACT

In this paper a bottom-up approach for human behaviour understanding is presented, using a multi-camera system. The proposed methodology, given a training set of normal data only, classifies behaviour as normal or abnormal, using two different criteria of human behaviour abnormality (short-term behaviour and trajectory of a person). Within this system an one-class support vector machine decides short-term behaviour abnormality, while we propose a methodology that lets a continuous Hidden Markov Model function as an one-class classifier for trajectories. Furthermore, an approximation algorithm, referring to the Forward Backward procedure of the continuous Hidden Markov Model, is proposed to overcome numerical stability problems in the calculation of probability of emission for very long observations. It is also shown that multiple cameras through homography estimation provide more precise position of the person, leading to more robust system performance. Experiments in an indoor environment without uniform background demonstrate the good performance of the system.

1. Introduction

Motion analysis in video and particularly human behaviour understanding has attracted many researchers [24], mainly because of its fundamental applications, which include video indexing, virtual reality, human–-computer interaction and smart surveillance. Smart surveillance in itself is one of the most challenging problems in computer vision. Its goal is to automatically model and identify human behaviours, calling for human attention only when a suspicious behaviour is detected. With the increasing number of cameras in many public areas, the related research becomes more appealing and is offered more application possibilities.

This work deals with the classification of behaviours as normal or abnormal. Based on the remark that abnormal behaviour is considered to be rather infrequent (and thus abnormal), we choose to model normal behaviour and define as abnormal any behaviour deviating from that normality model. Our methodology applies two classification criteria:

(1) short-term behaviour;
(2) trajectory.

The short-term behaviour refers to the type of behaviour that can be localized in a spatio-temporal sense, i.e. is brief and within restricted space. Examples of such behaviours are walking, standing still, running, moving abruptly, etc.

In the related literature the aforementioned classification criteria are mostly treated separately and, furthermore, few works concentrate on learning only normal behaviours. The methodology provided herein provides the discrimination of anomaly due to abnormal short-term motion, as happens in the case of abrupt motion, as well as anomaly due to long-term motion, as in the case of abnormal trajectory.

* Corresponding author.
E-mail addresses: ganton@iit.demokritos.gr (P. Antonakaki), dkosmo@iit.demokritos.gr (D. Kosmopoulos), sper@iit.demokritos.gr (S.J. Perantonis).

0165-1684/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2009.03.016

Recently, several researchers have dealt with the problem of anomaly detection, which is the process of behaviour classification as normal or abnormal. A variety of methods, ranging from fully supervised [9,10] to semi-supervised [36] and unsupervised systems [21,22,18], have been proposed in existing literature, which we further review in Section 2. It should be noted, however, that most of the existing approaches do not use multi-camera information, except for [38], where multiple video streams are combined via a coupled Hidden Markov Model.

Our methodology contributes in current research in several ways:

- The presented approach reflects two different criteria of labelling an observed behaviour as normal or abnormal, since the final abnormality decision depends on the output of two different classifiers with independent inputs: short-term behaviour information and trajectory information.
- The behaviours are classified according to the target object's position on the *ground plane*, based on homography (see Section 4) which provides higher accuracy compared to pure image-based techniques.[1]
- We introduce a continuous Hidden Markov Model (cHMM) as an one-class classifier, using the notion of length-normalized log-probability (see Section 6.1).
- A novel algorithm implementing a Forward Backward procedure for the emission probability estimation in HMMs is proposed, handling numerical instability resulting from long sequences (see Section 6.2).

The rest of the paper is organized as follows. In Section 2 recent literature is reviewed, hinting as to the problems the proposed method tackles. Section 3 provides an overview of the proposed architecture. In Section 4 we explain briefly how homography is used to obtain information on the position of target objects on the ground plane. In Section 5 short-term behaviours are defined in terms of a set of extracted features. Section 5.2 describes in detail the classification process which is based on short-term behaviours. In Section 6, on the other hand, trajectories' classification is presented by elaborating how we have used a continuous Hidden Markov Model as an one-class classifier (Section 6.1). As an added value, Section 6.2 contains the description and foundation of a modified algorithm for the Forward Backward procedure of probability estimation tackling long sequences in contemporary computers. Finally, in Section 7 we provide the experimental results and Section 8 concludes this paper through a brief discussion on the lessons learned.

2. Related work

A typical surveillance system is divided into two layers, which include *low level* and *high level* processes, respectively, as depicted in Fig. 1.

[1] An early version of this work has been presented in [20].

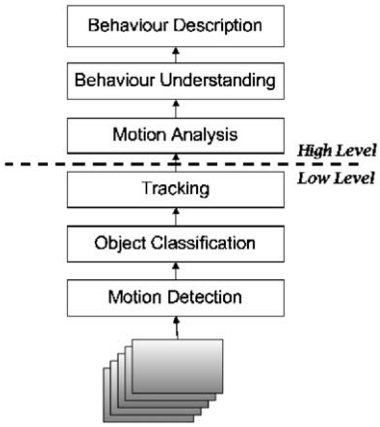

Fig. 1. The main framework for video surveillance systems.

The *low level* contains such methods as motion detection, object classification and tracking. In motion detection research is focused on either static or adaptive background subtraction or temporal differencing algorithms, aiming to isolate the foreground pixels that participate into any kind of motion observed in a given scene. Object classification is the process of classifying detected objects into such classes as humans or vehicles, appearing in a given scene. Following motion detection and object classification, detected objects are located in the course of time and their trajectories are extracted via tracking.

High level processes use motion information from the low level in order to finally identify the type or nature of a moving object's activity. Motion-based techniques are mostly used for short-term activity classification (e.g. walking, running, fighting), and do not take into account object trajectories. These techniques actually calculate features of the motion itself and perform recognition of behaviours based on these features' values. Such methods have been presented by Bobick et al. in [5] where motion energy images (MEIs) and motion history images (MHIs) are used to classify aerobic type exercises. Taking this work another step further, Weinland et al. in [34] focus on the extraction of motion descriptors analogous to MHIs, called *motion history volumes*, from multiple cameras. Then, these history volumes are classified into primitive actions. Efros et al. in [11] compute the optical flow [14] of a given object to recognize short-term behaviours through a nearest-neighbour classification.

Several methods that take into account the object's trajectory for behaviour classification use the centroid of the target object [1,19,27,15] or points of interest in a given image [4]. These methods, however, fail to take into account the short-term actions, for example the case where a man threateningly moves his hands. Most of the existing methods also face problems like view dependency, and occlusion when they extract trajectories from one camera.

HMMs and their variations have been widely applied on trajectory classification, e.g. [7,17,2,32], due to their unsupervised training, their simplicity and computational