Computer Vision and Image Understanding 144 (2016) 276–297

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Evidential event inference in transport video surveillance

Xin Hong[a,*], Yan Huang[a], Wenjun Ma[b], Sriram Varadarajan[a], Paul Miller[a], Weiru Liu[a], Maria Jose Santofimia Romero[c], Jesus Martinez del Rincon[a], Huiyu Zhou[a,*]

[a] Centre for Secure Information Technologies, School of EEECS, Queen's University Belfast, Belfast BT3 9DT, UK
[b] Department of Philosophy, East China Normal University, Shanghai, China
[c] Computer Architecture and Networks Group, University of Castilla-La Mancha, Spain

ARTICLE INFO

Article history:
Received 13 January 2015
Accepted 6 October 2015

Keywords:
Transport surveillance
Video events
Event modelling
Reasoning under uncertainty
Spatio-temporal constraint
Minimum conflict optimisation
Event association and recognition

ABSTRACT

This paper presents a new framework for multi-subject event inference in surveillance video, where measurements produced by low-level vision analytics usually are noisy, incomplete or incorrect. Our goal is to infer the composite events undertaken by each subject from noise observations. To achieve this, we consider the temporal characteristics of event relations and propose a method to correctly associate the detected events with individual subjects. The Dempster–Shafer (DS) theory of belief functions is used to infer events of interest from the results of our vision analytics and to measure conflicts occurring during the event association. Our system is evaluated against a number of videos that present passenger behaviours on a public transport platform namely buses at different levels of complexity. The experimental results demonstrate that by reasoning with spatio-temporal correlations, the proposed method achieves a satisfying performance when associating atomic events and recognising composite events involving multiple subjects in dynamic environments.

1. Introduction

Security information and event management systems (SIEMs) are well-established within the field of network security. Physical SIEMs are also well-established within the physical security domain. However, many of the events that they deal with are of a very simple nature with a high degree of certainty, e.g., intrusion alarms, access control. Intelligent analysis and correlation/aggregation of incoming events from different sources represent a challenge to these systems.

Recent developments in the field of video analytics have resulted in a new source of events for PSEIM that can provide rich semantically meaningful information with regard to situational awareness. However, unlike earlier event types, these can have a degree of uncertainty and can conflict with one another. While the video analytics community has been making progress on generating low-level events, typically termed action recognition [1–3], little thought has been given to how one manages events of this nature over a period of time to give higher-level composite events [4]. However, as this technology has started to migrate from the laboratory to the commercial sector, there is a growing realisation of the need to manage the events generated by video analysis software. By manage we mean the representation, storage, reasoning and mining of events.

One of the main tasks of event management systems is that of event composition, whereby patterns of events across a distributed network are detected. Event composition allows us to represent different events and also to instantly infer events of interest by applying rules to combine existing events. In addition, new situations can be captured by simply adding a new rule instead of modifying custom code, hence ensuring a flexible solution for evolving situations. Event composition can either be deterministic, or probabilistic, or both [5,6], however, to date only a few researchers have addressed the problems of imperfect information, or information from different sources that may be conflicting.

For the past decade or so, the deployment of CCTV in major urban centres and cities has become well established. Recently, CCTV technology has begun to be deployed on public transport systems such as buses and trains. The application domain of interest to us is the analysis of people's behaviour as they move into, remain in, and move out of seated areas. While this scenario has received very little attention to date within the computer vision community, seated areas are ubiquitous in many application scenarios. For example, these can be found onboard transport platforms such as buses, trains and planes. They are also to be found in many transport hubs such as train stations and departure lounges in airports. Other sectors where they are to be found include sports stadiums, entertainment venues such as

* Corresponding authors.
E-mail addresses: x.hong@qub.ac.uk (X. Hong), h.zhou@qub.ac.uk (H. Zhou).

http://dx.doi.org/10.1016/j.cviu.2015.10.017
1077-3142/

concert halls, and leisure venues such as restaurants and bars. Of particular interest to us is the bus scenario and detecting anti-social or criminal behaviour on buses. Studies have shown that the vast majority of crime carried out on transport platforms such as buses is by young males [7]. Therefore, having knowledge of the gender of passengers, and how they are moving relative to one another, as well as their seated positions, enables one to infer the degree of threat and likelihood of an anti-social/criminal incident occurring. The vast majority of events on a bus consist of passengers undertaking normal journeys in which nothing untoward happens. This can be decomposed as: a passenger boarding the bus at an entrance, moving into the saloon area along the gangway, and taking a seat, which we classify as atomic events. Similarly, when exiting, a passenger stands, moves along the gangway towards an exit, and then disembarks from the bus. We classify both these as composite events. Less regular composite events include a passenger changing a seat while the bus is moving. This could indicate that one of the passengers is either being intimidated or threatened by another passenger.

Unfortunately, imperfect information frequently occurs in real world applications. For example, in the case of a person entering the bus doorway, the person may be classified as male with a certainty of 85% by the classification analytics, however, the remainder does not imply that the person is female with a 15% certainty, rather, it is unknown. Hence, it can only give imperfect information for the remaining 15%. Imperfect information is usually caused by the unreliability of the information sources. For example, in the classification example above, the camera may have been tampered with, illumination could be poor, or the classifier training set may be unrepresentative. Any or all of these can result in imperfect information which cannot be represented by probability measures.

In this work, we investigate the use of evidential reasoning, for dealing with low-level, or atomic events that are uncertain, and combining them into higher level composite events that have semantic meaning from a security viewpoint. Our main contributions can be summarised as follows:

A. The development of a novel technique for associating identities with atomic events.
B. One of the first attempts at integrating video analytics with an event reasoning framework.
C. First demonstration of the recognition of composite, semantically meaningful, events onboard the challenging environment of a moving transport platform (bus).

The rest of this paper is organised as follows. In Section 2, we review related work. Section 3 provides a preliminary treatment of the Dempster–Shafer theory of evidence and temporal relation representation. We propose a new framework of subject-event association and composite event recognition in Section 4, a case study in Appendix B illustrates how our framework works. In Section 5, the experimental methodology is described and results are presented. Finally, Section 6 concludes this paper and discusses the future work.

2. Related work

During the recent past there has been an extensive amount of work on video action recognition by the computer vision research community ([2,8] and references therein). However, most of this work has been on what we call atomic event recognition, e.g. running, walking etc., which have a unique and enclosed, but limited, semantic meaning in relation to the application context. Less emphasis has been given to the use of reasoning for aggregating atomic events so that high-level semantically rich composite events can be recognised. This straddles the boundary between the computer vision and artificial intelligence communities, and perhaps is the reason why there has been less work in this area [9]. In this section we first review vision-based action recognition and then event reasoning approaches for composite event recognition.

2.1. Action recognition

Given a specific scenario, where interactive elements are known, simple action recognition can be performed by applying human detection to video sequences, and from these generate trajectories which can then be used to describe the actions of the detected subjects. For example, part-based techniques can be used to locate [10] and track [11] human body parts. These trajectories can then be modelled using methods such as Hidden Markov Models (HMMs) [12].

Although these techniques are simple to implement and effective for simple actions and scenarios, they fail to provide richer information of the sort needed to recognise more subtle actions. Extending this methodology to estimate the trajectory of the human pose [13,14], i.e. the trajectory of each body part, allows this. However, current methods have been shown not to be robust for real scenarios and multiple actors [15]. It is also debatable whether such fine grain detail is really necessary for action recognition [16]. In most practical scenarios human detection and pose estimation can be difficult, due to the presence of background clutter and foreground occlusions. Therefore, another approach is to treat a sequence, or part thereof, as a single entity from which low-level spatio-temporal features can be extracted and classified as belonging to a particular action. For example, Klaser et al. [17] proposed calculating the 3D Histogram of Oriented Gradients (HoG) over a space-time volume in order to characterise actions. Similarly, Ke et al. generated over segmented spatio-temporal volumes and optical flow correlation and then used a distance metric to determine the subset of spatio-temporal volumes that best matched a parts-based event template [18]. A common approach is to describe a video sequence as an unordered set of space-time features, e.g., bag of visual words (BoV). Wang et al. proposed a BoV approach to describe videos by dense trajectories [19]. Oneata et al. applied the Fisher vector representation, an extension of BoV, to action classification [20]. By employing approximations to Fisher normalisations they obtained a speed-up of an order of magnitude while maintaining state-of-the-art action recognition performance. In a different approach to BoV, Sadanand and Corso proposed the use of action banks, consisting of a bank of individual template-based action detectors that provided location features by maximum poling of volumetric correlation outputs [21]. The resulting feature vectors generated for different actions and scales were then concatenated and used to train an SVM for classification.

An event scenario that has received considerable attention is the detection of abandoned bags. Tian et al. applied the results of background subtraction for detecting static and foreground regions and, using a novel segmentation algorithm, the former were then classified as being abandoned or removed [22]. Human detection and tracking are also employed in order to reduce false positives. In other work [23], Fan et al. proposed representing abandoned objects alerts by relative attributes, e.g., staticness, foregroundness and abandonment. A ranking function, learnt using low-level spatial and temporal features, was used to determine the relative strengths of these attributes. Their system outperformed other state-of-the-art techniques in terms of precision for the PETS2006 and AVSS-AB datasets.

Another scenario that has been extensively investigated is that of crowd analysis for security and/or safety purposes. As has been noted in [24], techniques developed for non-crowded scenarios tend to fail in crowded scenes. As such, research has focused on addressing those issues, e.g., occlusion and complex collective behaviours, unique to crowds. Idrees et al. proposed identifying prominent individuals within a crowd that are relatively easy to track, and then using the concept of neighbourhood motion occurrence to determine the behaviour of individuals within the crowd [25]. Zhou et al. presented a new mixture model of dynamic pedestrian agents to learn

collective behaviour patterns of pedestrians in crowded scenes [26]. In [27], Yi et al. developed a technique for detecting stationary foreground regions by applying sparse constraints along spatial and temporal dimensions to produce a 3D stationary map. This is then used to detect four types of stationary group behaviours; gathering, relocating, joining and dispersing.

All of the previous work reviewed thus far, assumed prior knowledge of the actions, i.e., they were pre-defined. A more complex problem is the detection of unknown or unusual actions. Leach et al reported on an unsupervised context-aware approach which takes into account scene and social contexts to detect anomalous behaviour [28]. Static and dynamic agents were used by Cho and Kang to model individual and group behaviours as a BoVs [29]. Kittler et al. surveyed the area of anomaly detection and proposed the use of context for anomaly detection in video sequences [30].

2.2. Event reasoning

In this section we review related work on composite event recognition and reasoning, which is the focus of the work reported herein. Composite events have greater semantic meaning to end-users than atomic events, and are high-level semantic interpretations from a set of atomic events [5]. They are not easily identifiable using image features, but, rather, by recognition of their composing events [31].

There are two major approaches to composite event recognition, classification and inference. In the former, one approach is to use actions, scenes and objects as semantic attributes for their classification. Chen et al. [32] started with the identification of candidate concepts for an event by firstly crawling Flickr to search for images with tags related to keywords in the event description. WordNet was then used to filter out noisy tags and each concept verified based on the visual cohesiveness of the images associated with it. This was followed by building a concept visual model using a Support Vector Machine classifier. They found that their approach outperformed others based on low-level visual features for a supervised event modelling task. Li et al. [33] proposed decomposition of a video sequence into short-term segments, which were modelled by a dictionary of attribute dynamics templates using a binary dynamic system. It is common to see probabilistic approaches applied to the recognition of atomic events, due to their limited ability of modelling interrelations between events in both space and time, and representing structural information in event composition.

Inference-based approaches to composite event recognition usually involve development of an event modelling and reasoning mechanism. Composite events consist of a set of atomic events that occur over a considerable time-span and that may have a partial ordering or be concurrent. Thus, one of the main AI-based approaches to composite event recognition is to infer them by reasoning about atomic events. Works on visual event modelling and reasoning tend to follow two major trends; declarative and probabilistic.

In declarative approaches, descriptive templates are used to model events, such as context-free grammar [34] and Petri-Nets [35]. Ryoo and Aggarwal [36] used context-free grammar to model interactions of primitive actions and to recognise composite activities for multi-subject scenarios. Petri-Nets are used as a formalism to model complex logical temporal and spatial relations in event composition [37]. These are derived from semantic descriptions of events in video event ontology languages such as VERL. Recently, ontologies, a semantic web technique, have been used to automate the representation of composite events [38]. In [39], an event ontology for representing complex spatio-temporal events by a composition of simpler ones was proposed. The hierarchy includes primitive events, single-thread composite events and multi-thread composite events. Inferences are made in a bottom-up fashion. Declarative approaches work satisfactorily when describing event semantics. However, major drawbacks include an inability to handle multiple subjects and fragility to uncertainty in sensor measurements, which frequently exist in real applications.

In probabilistic approaches, such as HMMs [6], Dynamic Bayesian Networks [40], and multi-agent methods [41,42], models are constructed to represent events. While these demonstrate impressive robustness to uncertainty in video analytics, they do not define semantically meaningful sub-events. Thus, it is not easy to describe the composition of an event at a semantic level. Though DBNs are more general than HMMs, by considering dependencies between several random variables, the temporal model is still usually Markovian, as is the case for HMMs [5]. Their models can only handle sequential activities and fail to describe complex relations between sub-events. Consequently, they often lack flexibility; hence it is difficult to apply them to dynamic problems in real applications [43].

Recently hybrid approaches have emerged that combine declarative and probabilistic properties. These tend to combine the rich representation ability of declarative approaches with the uncertainty reasoning mechanism of probabilistic approaches. Stochastic grammars [44] have been used for parking lot surveillance in [45]. Tran et al. [43] applied Markov logic networks (MLNs) to probabilistically infer events in video surveillance where noise and missing observations are serious problems. First-order logic production rules are used to represent common sense domain knowledge. A weight is associated to each rule to indicate their confidence. In [46], Kanaujia et al. also proposed the use of MLNs for recognising complex events over a sensor network consisting of four cameras. In their approach, rather than using a single Markov network (MN) for representing all activities, they explicitly partitioned the MN into multiple activity specific networks. They addressed the issue of uncertainty, due to the noisy sensor data and video analytic errors, by generating predicates with an associated probability. Semantic information extracted at each level from the lowest level visual processing is propagated to sub-event detection by each MLN engine and then to a higher-level complex event module to recognise complex events. To tackle the problem of recognising coordinated events in challenging videos with cluttered background and occlusion, Brendel et al. [47] proposed the formulation of probabilistic event logic (PEL) for representing temporal constraints among events. Lavee et al. [31] introduced a certainty score to Petri Nets to cope with uncertain event observations.

Though the majority of previous declarative and probabilistic approaches have been applied to single subject scenarios, a few have tackled the more difficult problem of event recognition in multi-subject videos. Among them, attention has focused on recognising events from understanding the interactions between subjects. These works presume that low-level video analytics can provide sufficient information for the detection of simple semantic events, which often appears untrue in real world applications.

The approach proposed in this paper fits in the hybrid group and focuses on multi-subject video applications. Our solution is different from previous hybrid approaches in several ways. Firstly, we adopt the Dempster–Shafer (DS) theory of evidence [48,49] to handle uncertainty in event recognition, from observations, to event detection and inference. Imperfect information frequently occurs in real world applications. For example, in bus surveillance, when a person enters the bus the camera detects a face and classifies it as female with a certainty of 75%. However, the remainder does not imply that the face is male with a 25% certainty, rather it is deemed to be unknown because the gender classification analysis does not have enough information to distribute the remaining 25% to male or female. In contrast, with probability theory such information can only be represented as $p(female) \geq 0.75$ and $p(male) \leq 0.25$, which is difficult to use for reasoning. Furthermore, the propagation and combination mechanisms of DS theory are superior for composing complex events from simple sub-events and atomic events detected from noisy observations. Hierarchical network templates are used to model the structural semantics of complex event composition. Similar to [36], we use Allen's

temporal interval relation modelling [50] to represent temporal relations between events; however, we go further and deduce the association of events with different subjects in a multi-subject scene. One of challenges for event recognition in multi-subject videos is that video analytics often results in errors, such as missed detections, and broken tracks due to occlusion. To address this, we develop constraint rules, using the temporal relationships between events, and use conflict factors of Dempster's combination rule to measure conflict in event combinations, enabling us to associate events to a particular subject. Part of the current manuscript has been published in conference proceedings [50,51].

3. Preliminaries

In this section, we introduce the main concepts of reasoning under uncertainty and temporal relation representation, which we have relied upon in developing our proposed approach.

3.1. Dempster–Shafer theory of evidence

The fundamental technique of evidential reasoning that this work uses is the Dempster–Shafer theory of evidence (DS theory), which originated from Dempster's work [48] and further extended by Shafer [49]. DS theory is a generalisation of traditional probability theory and describes the propositional space of possible situations for a given problem by a finite, non-empty set called the *frame of discernment*, denoted as Θ. Uncertainty related to propositions of the problem is represented by a *mass function* over the power set 2^{Θ}: the set of all subsets of Θ.

Definition 1. The mapping $2^{\Theta} \rightarrow [0, 1]$ is a basic belief assignment, also called a mass function m, satisfying: (1) $m(\emptyset) = 0$; (2) $\sum_{A\subseteq\Theta} m(A) = 1$.

A mass value can be committed to a subset, A, of Θ with either single or multiple elements. All A are called focal elements if $m(A) > 0$, where $m(A)$ is attributed to A and only A. Due to lack of information this mass value cannot be further distributed amongst specific elements in A, which makes mass functions different with probability functions. When $m(\Theta) = 1$ and $m(A) = 0$ for all $A \neq \Theta$, the mass function represents total ignorance, called a *vacuous mass function*. When all focal elements of a mass function are singletons, the mass function is reduced to a probability function.

When two frames of discernment Θ_G and Θ_H hold relations described by an *evidential mapping* Γ^*, the mass function occurring on Θ_G can be *projected* to Θ_H via Γ^* as follows [51]:

$$m_{\Theta_H}(H_j) = \sum_i m_{\Theta_G}(g_i) f(g_i \rightarrow H_j) \tag{1}$$

$\Gamma^* : \Theta_G \rightarrow 2^{2^{\Theta_H}\times[0,1]}$ assigns an element $g_i \in \Theta_G$ to a set of subset–mass pairs in the following way:

$$\Gamma^*(g_i) = ((H_{i1}, f(g_i \rightarrow H_{i1})), \ldots, (H_{im}, f(g_i \rightarrow H_{im}))),$$

where $H_{ij} \subseteq \Theta_H$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, and f: $\Theta_G \times \Theta_H \rightarrow [0, 1]$ satisfying (a) $H_{ij} \neq \emptyset$, $j = 1, \ldots, m$; (b) $\sum_{j=1}^{m} f(g_i \rightarrow H_{ij}) = 1$; (c) $\Gamma^*(\Theta_G) = ((\Theta_H, 1))$.

When all the $f(g_i \rightarrow H_{ij})$ are either 1 or 0, an evidential mapping Γ^* becomes a multi-valued mapping $\Gamma : \Theta_G \rightarrow 2^{\Theta_H}$. A mass function from frame Θ_G can be *translated* to frame Θ_H as [52]:

$$m(H_j) = \sum_{\Gamma(g_i)=H_j} m(g_i), \tag{2}$$

where $g_i \in \Theta_G$, $H_j \subseteq \Theta_H$.

One advantage of DS theory is that it provides a mechanism of aggregating multiple pieces of evidence from different sources. When mass functions m_1 and m_2 are obtained from two independent sources over the same frame of discernment Θ, the consensus mass function m can be obtained by fusing them using *Dempster's rule of combination* as follows:

$$m(C) = (1 - k)^{-1} \sum_{A\cap B=C} m_1(A)m_2(B), \tag{3}$$

where $k = \sum_{A\cap B=\emptyset} m_1(A)m_2(B) \neq 1$ is considered to be a *conflict factor* that numerically measures the degree of conflict between two pieces of evidence. When $k = 0$, two pieces of evidence are completely consistent. When $k = 1$, the two are completely inconsistent. The combination rule is both commutative and associative.

It is common that information provided by a source may not be completely credible. To reflect the reliability of the source, a discount rate $r \in [0, 1]$ is introduced in [49]. The original mass function m from a source is discounted:

$$m^r(A) = \begin{cases} (1 - r)m(A), & A \subset \Theta \\ r + (1 - r)m(\Theta), & A = \Theta. \end{cases} \tag{4}$$

For decision making, Smets [53] proposed the pignistic transformation of mass functions.

Definition 2. Assume that there exists mass function $m(A)$, $A \subseteq \Theta$. For every element g of Θ, the pignistic probability, denoted *BetP*, can be calculated:

$$BetP(g) = \sum_{g\in A} \frac{m(A)}{|A|}, \tag{5}$$

where $|A|$ is the number of elements of Θ in A.

The pignistic probability is the DS counterpart of the subjective probability that would quantify the agent's beliefs according to the Bayesians [54].

3.2. Temporal relations

Allen proposed a method for modelling temporal relations in [50,55] that enables the representation of multiple subjects' actions over a period of time extending from the present to the future. In Allen's model, temporal information is represented by intervals than points. In this way, real-world events taking place over a time interval can be handled within the same modelling framework as instantaneous events, by treating the latter as occurring over a time interval with the same start and end time. The time of an event can be relative to a reference point rather than being absolute. To describe temporal correlations between two event instances that take place within two time intervals respectively, Allen defined 13 relations as depicted in Table 1.

Allen's temporal model will allow us to enforce constraints in our event inference for both continuous and discrete events.

4. Methodology

This section describes our system for uncertain atomic event management from multiple sensors and composite event inference. The system is proposed in the context of video-surveillance for public transport platforms.

Table 1
The Allen's 13 interval temporal relations on $IX = [IX_a, \quad IX_b]$ and $IY = [IY_a, \quad IY_b]$.

Interval relation	Symbol	Inverse	Endpoint relations
IX before *IY*	b	a	$IX_a < IY_a, IX_a < IY_b, IX_b < IY_a, IX_b < IY_b$
IX meets *IY*	m	mi	$IX_a < IY_a, IX_a < IY_b, IX_b = IY_a, IX_b < IY_b$
IX overlaps *IY*	o	oi	$IX_a < IY_a, IX_a < IY_b, IX_b > IY_a, IX_b < IY_b$
IX starts *IY*	s	si	$IX_a = IY_a, IX_a < IY_b, IX_b > IY_a, IX_b < IY_b$
IX during *IY*	d	di	$IX_a > IY_a, IX_a < IY_b, IX_b > IY_a, IX_b < IY_b$
IX finishes *IY*	f	fi	$IX_a > IY_a, IX_a < IY_b, IX_b > IY_a, IX_b = IY_b$
IX equal *IY*	eq	eq	$IX_a = IY_a, IX_a < IY_b, IX_b > IY_a, IX_b = IY_b$

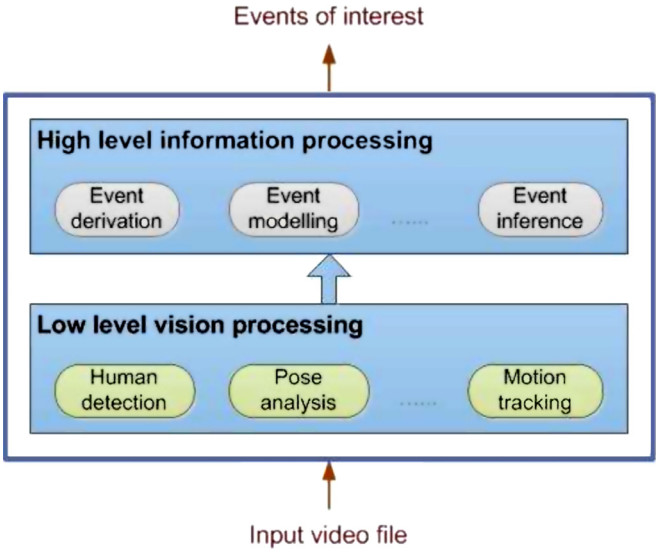


Fig. 1. System of intelligent event management for video surveillance.

4.1. System outline

The main purpose of video surveillance is to provide situational awareness of a specific place over a period of time. In this context, therefore, an *event* is an observation (or collection of observations) that has semantic meaning. An event can be simple or complex depending on the level of relevant semantic information provided. To distinguish these two different concepts, we call the former an *atomic event* and the latter a *composite event*. An atomic event can be directly detected using video analytics and/or sensors. Atomic events can then be aggregated to generate composite events which are more semantically meaningful.

Our system is composed of two main stages, shown in Fig. 1, and integrates computer vision techniques with knowledge representation and reasoning mechanisms. In the first stage, human subjects are detected and video analytics are then generated in order to provide low-level semantic components such as "a female face has been detected" and "a person has moved from the door towards the gangway". The second stage is designed to recognise significant events based on a semantic hierarchy obtained from domain knowledge. At this level, the events of interest are recognised based on the information derived at the lower-level with varying degrees of belief.

First stage modules have been previously developed and presented [56]. In this paper, we concentrate on investigating event inference processing at the upper level of the proposed system.

4.2. Event inference procedure

Knowledge is the main drive behind the proposed event inference approach. Our knowledge base contains frameworks for representing uncertain events, spatio-temporal relations and event network models, which facilitate atomic event detection, event association and composite event recognition, Fig. 2. Event inference starts by deriving atomic events from the outputs of the computer vision analysis modules. Once atomic events are detected, the event association aims to make the correct association of atomic events to specific subjects. Composite event recognition then is performed on the detected atomic events associated to a single subject. The final outputs of the process are the subjects with the composite events they have undertaken. In the following subsections we will describe the proposed methods for the event inference processing.

4.3. Event representation

Uncertainty is intrinsic to event recognition. Video sensors cannot provide complete information of an evolving scenario over time. In other words, the video analysis modules have certain limitations with respect to providing correct visual information about a scene. During information processing, there is uncertainty in representing the relations between two events of interest. Nevertheless, an intelligent event management system should be able to represent and infer useful information in the presence of uncertainty.

We first define a formal representation of atomic events.

Definition 3. In our event inference system, an atomic event E is represented by a tuple:

$$E = (eType, oID, date, time, location, source, reliaR, vFrame, m),$$

where *eType* is the descriptor of an event, e.g. "Female Boards the bus"; *oID* is the identity number, assigned by a video analytics module or sensor, for the detected event, e.g. "track id 12"; *date* is the date of the observed event; *time* is the time-stamp for the observed event; *location* presents location information, e.g. "at seat 3" and "a trajectory"; *source* denotes the source from which the event was detected;

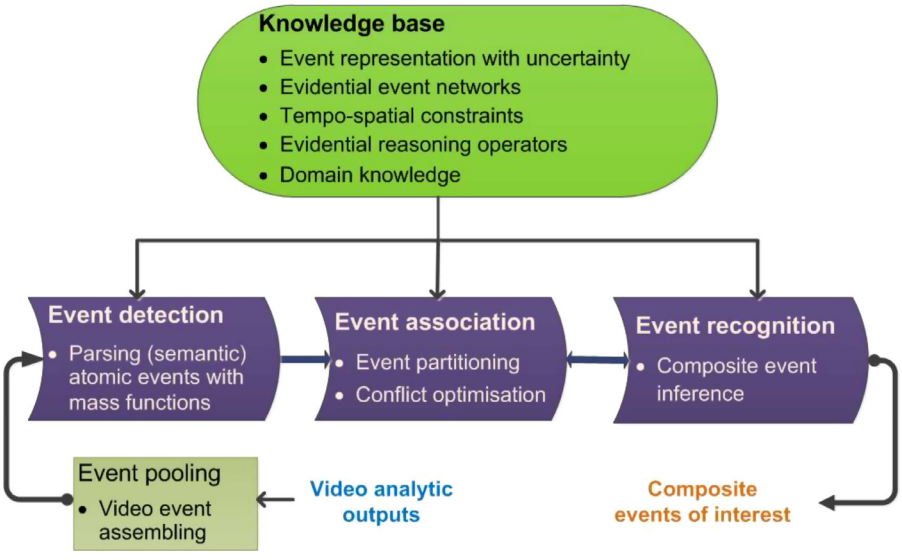


Fig. 2. Event inference components.

reliaR is the degree of reliability of the source; *vFrame* is the frame of discernment that holds all its values; and *m* is a mass function on *vFrame*.

As previously mentioned, in a multi-subject environment, each event, be it atomic or composite, belongs to only one subject. Therefore, to provide a generic framework for the multi-subject scenario that encompasses both atomic and composite events, we introduce the concept of an *event node* which is defined as follows:

Definition 4. An event node n is a tuple:

$$n = (eType, pID, level, oID, date, time, location, source, reliaR, vFrame, m),$$

where *eType*, *oID*, *date*, *time*, *location*, *source*, *reliaR*, *vFrame* and *m* have the same meaning as those in an atomic event; *pID* represents the identity number of the subject who is responsible for the occurrence of the event; *level* indicates whether the event is *atomic* or *composite*.

From the above definition, it can be seen that there are two sorts of event nodes, distinguished by *level*, either be *atomic* or *composite*. For the first type, an event node is an atomic event, except that the event node has an additional element *pID*. For the second type, an event node represents an event deduced from atomic events and/or composite events. *pID* is kept for the same subject through the full sequence and associated to all the event nodes that the subject generates. Therefore, a composite event node has the same *pID* as the atomic/composite events that it consists of. Its *date* and *time* cover the period from the first event starts until the last event ends. For a composite event node, *oID*, *location*, *source* and *reliaR* are omitted.

4.4. Composite event modelling

To represent the hierarchical structure of the relationships between composite and atomic events, and the video analytic outputs, we propose an evidential network model for event composition [57,58].

Definition 5. An evidential event network (*EEN*) is a graph of an upside-down tree $EEN = (ND, EG, MM)$, where:

- **ND** $= \{n_1, \ldots, n_N\}$ is a set of event nodes;
- **EG** is a set of directional lines over *ND*, each of which represents the connection between the nodes at two consecutive layers;
- **MM** is a set of multi-valued mappings Γ, each of which describes compatibility relations between the node at the layer where a line starts and the node at the layer where the connection line ends.

Fig. 3 shows the layout of an example *EEN*, $EEN = (ND, EG, MM)$ where

$$ND = \{AE1, AE2, AE3, AE4, CE1, CE2\},$$
$$EG = \{AE1 \longrightarrow CE1, AE2 \longrightarrow CE1, AE3 \longrightarrow CE2, AE4 \longrightarrow CE2, CE1 \longrightarrow CE2\},$$
$$MM = \{\Gamma : AE1 \rightarrow CE1, \Gamma : AE2 \rightarrow CE1, \Gamma : AE3 \rightarrow CE2, \Gamma : AE4 \rightarrow CE2, \Gamma : CE1 \rightarrow CE2\}.$$

On an *EEN* the nodes are categorised into three levels. The top level contains a root node, and at the bottom level we have many leaf nodes. Between these two levels, the middle level consists of several sub-layers. Over the three levels, there exist two types of nodes that are characterised by the level at which a node sits. A leaf node at the bottom level can be an atomic event, such as *AE*1 in Fig. 3, which is detected by a sensor, e.g. a seat pressure sensor, or a video analytics module, e.g. face detection and a tracker. A leaf node is always connected to the start of an edge. At the other end of the edge, we have nodes from the middle level, such as *CE*1 in Fig. 3. Middle level nodes are composite events, derived from the connected atomic event nodes. Composite event nodes at this sub-level may be further connected together in order to form composite events at higher sub-layers. On the topmost level of the *EEN* tree, there is a composite event node that is formed by atomic and/or composite event nodes below, containing the events of interest to the end users.

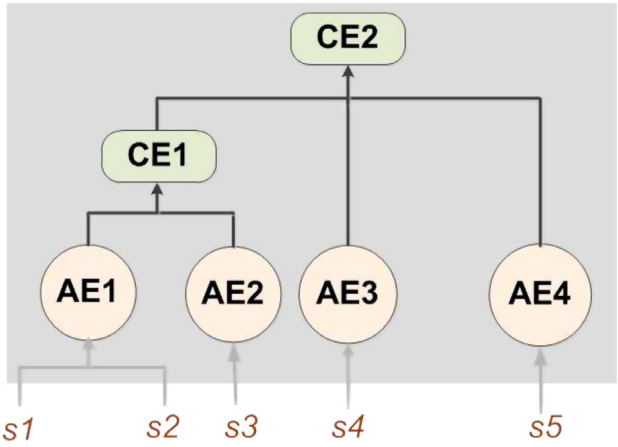


Fig. 3. A simple example of the general layout of evidential event networks: s1–s5 represent sources that provide evidence on atomic events; AE1–AE4 represent event nodes at atomic level; CE1 and CE2 are the event nodes at composite level.

The hierarchical structure of an *EEN* reveals semantic relations between events, which are the foundation of evidential event composition and inference developed below. This paradigm also helps in preventing redundancy by reusing the recognised atomic and composite events across *EEN*s.

Uncertainty associated with each node is defined as a mass function *m*. For an atomic event, denoted as a leaf node of the *EEN*, the mass value can be estimated from the accuracy of the computer vision detection module which is its source. For a composite event, the mass distribution can be derived through a composite event inference process as detailed in the following sub-section.

4.5. Composite event inference

At the bottom level of an *EEN*, the atomic events as leaf nodes are detected from outputs of sensors or video analysis modules. Information on detected atomic event nodes can be used to deduce information on higher-level nodes of composite events by propagating and aggregating evidence of atomic events through the network using evidential reasoning operations.

Composite event inference starts from having detected atomic events from outputs of the computer vision analysis modules and moving up within an *EEN*. The final output of the process is the mass function on the composite event node in concern. Algorithm 1 details the inference process.

4.6. Event–subject association

In multi-subject scenarios, it is usual that several subjects may be present at the same time, resulting in highly ambiguous video analytic output. For example, it is quite common that a single individual is assigned several IDs in complex scenes due to split/erroneous tracks produced by the tracking system. Intuitively arranging all detected atomic events with the same object ID assigned by video analytics into a composite event network *EEN* and directly making inference on the composite event node at *EEN*'s root inevitably produces errors. To solve this problem, we propose an atomic event association method by integrating the use of temporal relation modelling in event composition and evidential reasoning in event inference.

Algorithm 1 Evidential event inference.

Input: an event network *EEN*, mass functions of the detected atomic events
Output: mass function cast on composite event node at the top of the *EEN*

1: start from composite event nodes connected by only atomic event nodes at the start of a connection (so called parent and child nodes);
2: **while** not reach the topmost node of the *EEN* **do**
3: translate mass functions of all child nodes into their parent node using Eq. (2);
4: combine the translated mass functions using Eq. (3);
5: **end while**
6: output the final mass function on the topmost event node.

The event association problem can be seen as the association of all related atomic events with an individual under observation. The problem is two-fold: (i) partitioning a set of atomic events into different groups, and (ii) selecting the most probable set of partitions among many possible sets.

Definition 6. For a set of atomic events $\Xi = \{E_1, \ldots, E_{|\Xi|}\}$, a **partitioning** $S = \{S_1, \ldots, S_{|S|}\}$, satisfies:

$$(1)\ S_1 \cup \cdots \cup S_{|S|} = \Xi \quad (2)\ S_i \neq \emptyset \quad (3)\ S_i \cap S_j = \emptyset$$

where $i, j = 1, \ldots, |S|$ and $i \neq j$.

It is possible that we do not have sufficient information to justify if an atomic event belongs to one subject or another. This results in many possible choices to group those atomic events, i.e. we have many ways for partitioning. In the cases where several possible sets of partitions exist, a partition set may be considered more satisfying than others and is therefore selected as the most optimum partitioning of the atomic events.

4.6.1. Event partitioning

Partitioning atomic events aims to identify subjects who are responsible for the occurrences of the atomic events, in order to infer the composite events undertaken by the subjects. We investigate the intrinsic properties of ID assignments, as well as characteristics of atomic events, in order to determine a possible partitioning. For this purpose, we introduce two functions Φ and Ψ.

Let $PID = \{pID_1, \ldots, pID_P\}$ be a set of subject IDs, $\Xi = \{E_1, \ldots, E_{|\Xi|}\}$ ($|\Xi| \geq P$) be a set of atomic events, and $S = \{S_1, \ldots, S_{|S|}\}$ be a partitioning of Ξ. For Ξ, we have $\Omega = \{e_1, \neg e_1, \ldots, e_{|\Xi|}, \neg e_{|\Xi|}\}$, a set of possible states for all the atomic events related to a subject, whereas e_i means the occurrence of event E_i concerns the subject, and $\neg e_i$ does not.

Definition 7. A function Φ that assigns a partition S_i to a subject ID pID_i is defined as:

$$\Phi(pID_i) = S_i \tag{6}$$

where $S_i \subseteq S, i = 1, \ldots, |S|$.

A mapping function Φ represents the one-to-one mapping relation between a subject ID and a partition of atomic events.

Definition 8. A function Ψ that maps each subject ID pID_i onto possible states of the atomic events is defined as follows:

$$\Psi(pID_i) = \omega_i \tag{7}$$

where $\omega_i \subset \Omega$ and $\nexists E_j \in \Xi, s.t. \{e_j, \neg e_j\} \subseteq \omega_i$.

A mapping function Ψ represents the relation between a subject ID and the occurrence/non-occurrence states of atomic events.

There is a relation between an event E_j and the sate set $\{e_j, \neg e_j\}$ based on the two mapping function Φ and Ψ.

$$\begin{cases} E_j \in S_i & \Leftrightarrow\ e_j \in \omega_i \\ E_j \notin S_i & \Leftrightarrow\ \neg e_j \in \omega_i \end{cases} \quad for\ any\ i = 1, \ldots, P, \quad j = 1, \ldots, |\Xi|.$$

From Definitions 7 and 8, we can see that event partitioning is actually about deciding the state of each atomic event in relation to a subject ID. For a subject pID_i, we can have a set of states of the detected atomic events, where the occurrence/non-occurrence state of an atomic event indicates that pID_i is responsible for the happening of the atomic event. The restrictions of $\omega_i \subset \Omega$ and $\{e_j, \neg e_j\} \not\subseteq \omega_i$ for any $j \in [1, |\Xi|]$ means that, for a given subject ID, either the occurrence or non-occurrence state of an atomic event holds.

To deduce the possible state of an atomic event for a subject, we consider the occurrence constraints on atomic events concerning the subject. For a subject, the occurrence of an atomic event can be affected by and/or has impacts on the occurrence of other atomic events. For example, "I am *reading a book* at home at 9 pm" implies that "I cannot be *playing basketball* at a sports centre at 9 pm on the same day". Identifying the state of an atomic event from the already known states of another atomic events is called *event implication*. This is managed by using *constraint rules*, which determine the possible state of an atomic event with regard to other atomic events in concern.

Definition 9. A **constraint rule** R is expressed as a tuple

$$R = (Statement, Premise, Condition, Result)$$

where:

Statement is the description of the constraint rule that the premise set should obey.
Premise is a set of *eTypes* of which atomic events are prerequisites.
Condition is a conjunction of a set of conditions on the states of some atomic events currently hold.
Result is a set of the states of atomic events in relation to a subject, obtained by applying the constraint rule.

In this work, we consider three types of constraints: temporal, spatial and common knowledge. Since atomic events happen over a period of time, the temporal relation between atomic events usually implies their states in relation to each other. For example, a man cannot play basketball and watch TV at the same time. Similar to temporal constraints, spatial relations between atomic events usually imply their states. For example, a man cannot be in two separate places at the same time. A common knowledge constraint is derived from knowledge about the domain context. Consider a man taking a bus, he cannot exit the bus without boarding the bus first. A constraint rule can include one, two or all three types of constraints.

Condition in the form of formula presents temporal and spatial relations of existing atomic events. In particular Allen's temporal relation models are used to describe temporal relations between two event instances. We abstract the Allen's relations in Table 1, into a small set, $\{b, a, m, mi, ol, eq\}$, as shown in Table 2.

Rules are pre-requisite for finding states of atomic events. Rule R is used to search for events that violate or obey constraints of the three types. Therefore, the state of an atomic event can be identified in relation to a subject. Upon the states of all atomic events have been determined for each subject, the partitions of atomic events can then be obtained.

To show what an event constraint rule looks like, consider two examples from the bus journey scenario. Assume that the atomic events are derived from the video data. From common sense existentialism, we can have the following rules.

Table 2
Mapping of the abstract and original Allen's interval temporal relations.

Abstract relation	Allen's relation(s)
b	*b*
a	*a*
m	*m*
mi	*mi*
ol	*o, oi, s, si, f, fi, d, di*
eq	*eq*

Example 1. A rule, ensuring that one person cannot undertake two different events at a time that are detected in a bus scenario, can be defined as follows:

Rule R1
Statement: a person cannot undertake two events at a time;
Premise: {PB, PM, PSIT, PSTD, PE}[1] is a set of event types;
Condition: $E_i.Time\ eq\ E_j.Time \wedge e_i \in \Psi(pID_p)$;
Result: $\neg e_j \in \Psi(pID_p)$.

Example 2. A rule, describing that a person cannot exit a bus before having boarded the bus, can be defined as follows:

Rule R2
Statement: a person can only exit the bus after having boarded the bus;
Premise = {PB, PE} is a set of event types;
Condition: $E_i.eType = PB \wedge E_j.eType = PE \wedge E_i.Time\ a\ E_j.Time \wedge e_i \in \Psi(pID_p)$;
Result: $\neg e_j \in \Psi(pID_p)$.

Based on the relevance of the events detected by the video analytics and the non-relevance of the events obtained by the implication rule, we attempt to find the optimum partitioning for the set of atomic events. A partitioning of the set of atomic events is to identify the persons under observation, each partition $S_i \subset S$ should satisfy the following principle:

Proposition 1. *Suppose* $PID = \{pID_1, \ldots, pID_P\}$ *is a set of possible person IDs for a set of atomic events* $\Xi = \{E_1, \ldots, E_{|\Xi|}\}$ $(|\Xi| > P)$, $\Omega = \{e_1, \neg e_1, \ldots, e_{|\Xi|}, \neg e_{|\Xi|}\}$ *is a set of possible states for all atomic events in* Ξ, *and* Ψ *is a mapping function that indicates the relation between subject ID and the states of the given atomic events, then we have:*

(*i*) **Uniqueness:** $\nexists e_i \in \Omega, s.t. e_i \in \Psi(pID_u) \cap \Psi(pID_v))$, $u, v = 1, \ldots, P, u \neq v$;
(*ii*) **Completeness:** $\Psi(pID_1) \cup \ldots \cup \Psi(pID_P) = \Omega$.

Proof. (i) Uniqueness: Assume that $\exists e_k \in \Omega, \ s.t. \ e_k \in \Psi(pID_u) \cap \Psi(pID_v), u, v = 1, \ldots P, u \neq v$, by Definition 8, we have $E_k \in S_u$ and $E_k \in S_v$. Thus, $S_u \cup S_v \neq \emptyset$. It violates the definition of partition in Definition 6. Thus, item (i) holds.

(ii) Completeness: We first prove that $(\Psi(pID_1) \cup \cdots \cup \Psi(pID_P)) \subset \Omega$.

By Definition 8, we have $\Psi(pID_i) \subset \Omega$ $(i = 1, \ldots, P)$. Thus, it holds.

Then, we prove that $\Omega \subset (\Psi(pID_1) \cup \cdots \cup \Psi(pID_P))$.

Let $S_1 \cup \cdots \cup S_P$ be a possible partitioning of Ξ that indicates the set of possible person IDs *PID*. By Definition 6, for any $E_i \in \Xi$, we have $E_i \in (S_1 \cup \cdots \cup S_{|S|})$. Moreover, for any state x, $x \in \Omega$, we have $\exists E_k \in \Xi$, such that, $x \in \{e_k, \neg e_k\}$, where $E_k \in (S_1 \cup \cdots \cup S_{|S|})$. Without losing generality, let $E_k \in S_l, S_l \in \{S_1, \ldots, S_P\}$. Then, by Definition 6, for any S_h $(h \neq l)$, we have $S_h \cup S_l = \emptyset$ and $E_k \notin S_h$. Thus, by Definition 8, we have $e_k \in \Psi(pID_l)$ and $\neg e_k \in \Psi(pID_h)$. Clearly, we have $x \in \Psi(pID_l) \cup \Psi(pID_h)$ for any $h \neq l$. So, for any occurrence state x, if $x \in \Omega$, then $x \in (\Psi(pID_1) \cup \cdots \cup \Psi(pID_P))$. Thus, item (ii) holds. □

The completeness states that any atomic event shall be included in a partition. The uniqueness means that an atomic event should be in one and only one partition.

Following Proposition 1, we obtain all the possible partitions for a set of atomic events, indicating the occurrence/non-occurrence states of atomic events that each subject ID holds. The next step is to determine which partition minimises the inferred conflict.

4.6.2. Minimum conflict optimisation

After obtaining a possible set of partitions for all atomic events, we can assign each partition to a possible person ID, i.e. we have a one-to-one mapping from $S = \{S_1^t, \ldots, S_P^t\}$ to $PID^t = \{pID_1^t, \ldots, pID_P^t\}$.[2] Therefore, we can obtain a set of event nodes for each possible person ID defined in Definition 4. Afterwards, we apply *EENs* introduced in Section 4.4 to infer all the composite events related to each possible person ID. However, if we have more than one possible partitioning of the atomic events, how can we choose the best from many possibilities? In this subsection, we will solve this problem using the conflict factor in the Dempster's rule of combination.

After having identified all the atomic events related to a specific subject, we feed the atomic events into the *EEN*s and derive the composite events. This is done by aggregating atomic events through *EEN*s using the Dempster's Rule of Combination as proposed in Section 4.5. When combining atomic event evidence, the conflict factor k in Eq. (3) is a measure of the amount of conflict between the two pieces of evidence as described below.

(1) $k = 0$ *totally agree*;
(2) $0 < k < 1$ *agree to some extent*;
(3) $k = 1$ *totally disagree*.

We use k to select the most probable partition of object IDs. Since each composite event for a possible person ID accompanies a degree of conflict, we need to consider the aggregation effect during the inference process for each possible partitioning.

Definition 10. Let $S_1^t \cup \cdots \cup S_P^t$ be a possible partitioning of the set of atomic events $\Xi = \{E_1, \ldots, E_{|\Xi|}\}$ and P be the total number of persons, where for each S_p^t, we have $S_p^t = \{E_u, \ldots, E_v\}$, each element of which relates to the pth person. We therefore calculate the aggregation effect in terms of a conflict factor when inferring composite events for each possible person, denoted as $\hat{k}_p^t$:

$$\hat{k}_p^t = \frac{\sum_{i=1}^{L} k_i^t}{L}, \tag{8}$$

where L is the total number of the composite events inferred for the pth person. k_i^t is a conflict factor obtained from the inference of the ith composite event, as k in Eq. (3).

For a conflict factor, the smaller its value is, the more confident support evidence has. From this we can have the definition the most probable partitioning.

Definition 11. The dth possible partitioning is the most probable one for the set of object IDs if it satisfies $d = \arg\min_t (k^t)$, $k^t = \sum_{p=1}^{P} \hat{k}_p^t$.

After finding the most possible partition for the set of subject IDs, we retain a set of person IDs based on Definition 7 and the event nodes that have been determined to them by using Definition 4.

Algorithm 2 summarises the event association process.

[1] *PB*—person boarding, *PM*—person moving, *PSIT*—person sitting, *PSTD*—person standing, *PE*—person exiting.

[2] Since there may be more than one possible partitioning for a set of atomic events, we use the superscript t to distinguish them.

Algorithm 2 Event association.

Input: $\Xi = \{E_1, \ldots, E_{|\Xi|}\}$, a set of atomic events;
P the number of persons;
EEN the evidential event networks;
R constraint rules;
Output: $S = \{S_1, \ldots, S_P\}$, a set of atomic event partitions
Begin
1: $\Omega = \{e_1, \neg e_1, \ldots, e_{|\Xi|}, \neg e_{|\Xi|}\}$;
2: initialise $\omega_1 = \cdots = \omega_P = \Omega$;
3: $i = 1$;
4: **while** not reach the end of Ξ **do**
5: Search $\omega_1, \ldots, \omega_P$ to find all ω_j that hold events satisfying the constraints on E_i;
6: **if** possible **then**
7: Delete e_i or $\neg e_i$ accordingly from ω_j;
8: **else**
9: Create the options;
10: **end if**
11: $i++$;
12: **end while**
13: Find all the combinations of elements in ω_j^t by proposition 1;
14: Calculate k for each combination;
15: Select ω^t holding the smallest k as the association;
16: Obtain the partitioning S^t from ω^t;
17: Output the partitioning S^t.
End

5. Experiments

In this section we describe an experiment in which the ability of our system to recognise the following four composite events is measured:

- MBTS: Male boards, moves to a seat and sits down
- FBTS: Female boards, moves to a seat and sits down
- PCS: Person changes seat
- PEX: Person exits

Table 3
Properties of the eight test sequences.

Sequence	No. of passengers	No. of frames	Sequence	No. of passengers	No. of frames
1	1 male and 1 female	2556	5	1 male and 1 female	1902
2	1 male and 1 female	1733	6	2 male and 1 female	5202
3	1 male and 1 female	2667	7	2 male and 2 female	5522
4	1 male and 1 female	2662	8	3 male and 3 female	10,322

We compare the performance of our system to a simple rules-based approach with no reasoning and an adapted Bayesian reasoning system.

5.1. Environmental set-up

We hired a standard single-deck bus from Translink (Northern Ireland), which travelled a defined journey in the Northern Ireland Science Park. Fig. 4a is the aerial view of the local neighbourhood with a red curve outlining the route and six black circles marking six bus stops. The researchers from the ECIT centre were recruited as passengers. The bus saloon and the seat plan are shown in Fig. 4b and c, respectively. In the experiments 20 seats in the first five rows of the bus, numbered C1–C20, were deployed as passenger seats.

Two cameras were used on the bus: a Panasonic camera WV-NP244 (camera A) is used to monitor the front door of the bus, and an AXIS M31-R camera (camera B) is used to monitor the saloon area. Camera A is carefully positioned so that it can capture a passenger's face as s/he boards the bus. Camera B looking at the saloon can record the movements of passengers.

5.2. Dataset

We captured eight sequences of varying complexity, including different numbers of passengers on board, various passenger behaviour patterns, and from simple to difficult scene captures. The properties of the eight sequences are summarised in Table 3. Each sequence is described in detail in Appendix C.

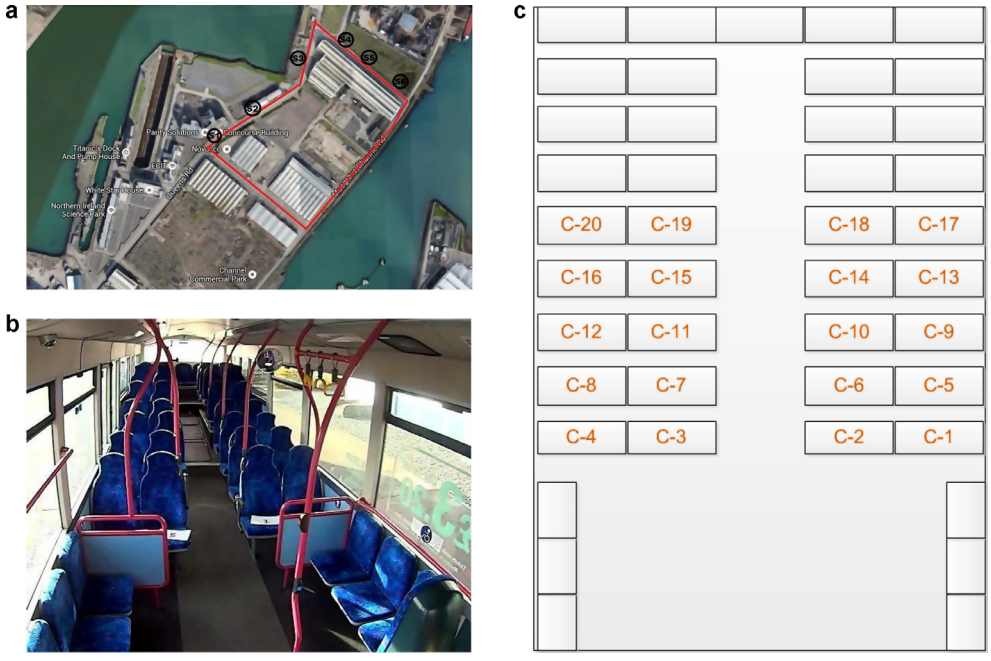


Fig. 4. Experimental environment: (a) route with six designated stops (the red curve highlights the route, the black circles mark the six bus stops); (b) bus saloon; (c) seat layout (numbered seats are used in experiments). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

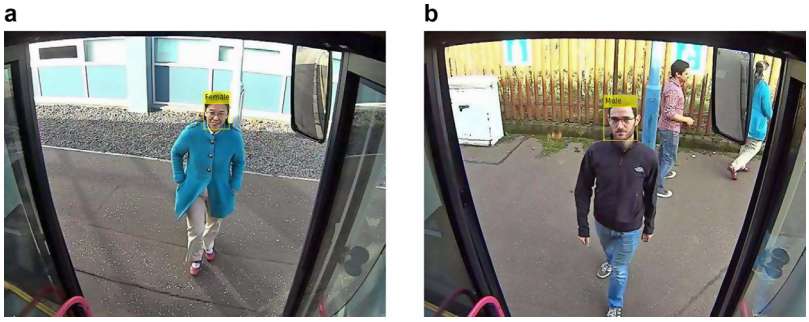


Fig. 5. Gender classification.

Fig. 6. Outputs of the tracker: (a) image with tracker bounding boxes over the person in the scene; (b) the corresponding track plots in real-world coordinates.

5.3. Video processes

For detecting passengers boarding and gender recognition, we employ a camera pointing at the door of the bus. The well-known Jones and Viola face detector is then applied to the acquired video. The output of this detection is then input to a face-based gender classifier. This, firstly, projects the face image onto a subspace derived using a principal component analysis of a training data set of face images. The resulting feature is input to a support vector machine that has been trained on approximately 2000 male and female face images. The resulting output is the credibility of the face as being either female or male (Fig. 5a and b).

For monitoring movements of passengers, we employ a 3D tracker that consists of three stages. Firstly, we apply the Poselet detector to detect instances of humans in the video on a frame-by-frame basis. These detections are then linked together to form tracks using a hierarchical linear assignment procedure. In the first level, detections are linked on a frame-to-frame basis by linear assignment. The resulting tracklets are then subsequently linked into tracks by a second level of linear assignment (see [56] for further details). Fig. 6a shows an example of a male and a female being tracked and their corresponding tracks projected into real-world space, Fig. 6b.

For sitting and standing detection, the height from the top of the head of an individual to the ground plane is calculated and compared to a threshold of 1.4 m. This threshold was empirically determined through trial and error, and is around the lower end of the normal human height distribution. The height can be estimated given the scene calibration and a standard reference height in the scene. When a passenger boards the bus, the passenger is inferred to be standing. For subsequent frames, if the height falls below the threshold at any point, we infer that the passenger has sat down. While sitting, if the height increases beyond the threshold value, we infer that the passenger has just stood up. Fig. 7 shows an example of sitting and standing being detected.

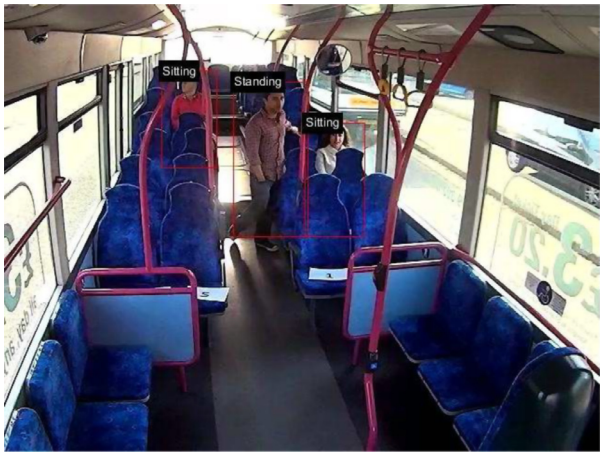


Fig. 7. Sitting and standing detection.

5.4. Evaluation

To evaluate the performance of our system in terms of the association of events with personal IDs and composite event recognition, we use two measurements. The first of these is the accuracy of the event association, A, which is given by

$$A = \frac{C_{AE}}{N_{AE}},$$

where C_{AE} is the number of atomic events in a sequence correctly associated with a personal ID, and N_{AE} is the total number of atomic events in the sequence. The second is the accuracy of the composite event recognition, R, given by

$$R = \frac{1}{N_{CE}} \sum_{i=1}^{N_{CE}} I_i$$

Groundtruth - seq5: part 1

	Gender	Atomic event 1			Atomic event 2			Atomic event 3		
		IDs	Title	End_frame	IDs	Title	End_frame	IDs	Title	End_frame
P1	Female	-2	FB	88	7	PM	265	Seat9	Sit_down	315
		Composite Event 1: FBTS9								
P2	Male	-5	MB	821	70	PM	979	Seat13	Sit_down	1019
		Composite Event 1: MBTS13								

Groundtruth - seq5: part 2

Atomic event 4			Atomic event 5			Atomic event 6		
IDs	Title	End_frame	IDs	Title	End_frame	IDs	Title	End_frame
Seat9	Stand_up	1133	84	PM	1245	Seat3	Sit_down	1246
Composite Event 2: PCS3								
Seat13	Stand_up	1250	105, 133, 136	PM	1442	Seat4	Sit_down	1443
Composite Event 2: PCS13								

Groundtruth - seq5: part 3

Atomic event 7			Atomic event 8			Atomic event 9		
IDs	Title	End_frame	IDs	Title	End_frame	IDs	Title	End_frame
Seat3	Stand_up	1638	162	PM	1700	-15	PE	1765
Composite Event 3: PEX								
Seat4	Stand_up	1725	199	PM	1798	-12	PE	1875
Composite Event 3: PEX								

Fig. 8. Manual ground truth for sequence 5.

Table 4
Association results for the evidential reasoning system.

Sequence	Number of atomic events		A (%)
	Ground truth	Evidentialreasoning system	
1	12 (2 PB, 4 PM, 2 PSIT, 2 PSTD, 2 PE)	12 (2 PB, 4 PM, 2 PSIT, 2 PSTD, 2 PE)	100
2	15 (2 PB, 5 PM, 3 PSIT, 3 PSTD, 2 PE)	13 (2 PB, 5 PM, 2 PSIT, 2 PSTD, 2 PE)	87
3	15 (2 PB, 5 PM, 3 PSIT, 3 PSTD, 2 PE)	15 (2 PB, 5 PM, 3 PSIT, 3 PSTD, 2 PE)	100
4	6 (1 PB, 2 PM, 1 PSIT, 1 PSTD, 1 PE)	6 (1 PB, 2 PM, 1 PSIT, 1 PSTD, 1 PE)	100
5	18 (2 PB, 6 PM, 4 PSIT, 4 PSTD, 2 PE)	16 (2 PB, 6 PM, 3 PSIT, 3 PSTD, 2 PE)	89
6	21 (3 PB, 7 PM, 4 PSIT, 4 PSTD, 3 PE)	21 (3 PB, 7 PM, 4 PSIT, 4 PSTD, 3 PE)	100
7	27 (4 PB, 9 PM, 5 PSIT, 5 PSTD, 4 PE)	22 (4 PB, 6 PM, 4 PSIT, 4 PSTD, 4 PE)	81
8	39 (6 PB, 13 PM, 7 PSIT, 7 PSTD, 6 PE)	24 (6 PB, 8 PM, 5 PSIT, 4 PSTD, 1 PE)	62

where I_i equals one if the ith recognised composite event in the sequence matches the ground truth, and zero if not, and N_{CE} is the total number of composite events in the sequence.

For the purposes of this each sequence is manually ground truthed both in terms of its atomic events and its composite events. Fig. 8 shows the manual ground truth for sequence 5.

From the table we can see that both male and female have nine atomic events each consisting of the sequence: PB, PM, PSIT, PSTD, PM, PSIT, PSTD, PM and PE (*PB = "male or female boards bus", PM = "person moves from X to Y", PSIT = "person sits", PSTD = "person stands" and PE = "person exits"*). Similarly, these correspond to three composite events: MBTS(FBTS), PCS and PEX. Each sequence was then input to our system and the corresponding atomic events, their associated person IDs and the recognised composite events output were recorded. Comparison of these against the ground truth enabled us to calculate both A and R for each sequence.

5.5. Results and analysis

Table 4 shows the variation in A with sequence numbers. Clearly, the event association works very well for almost all the sequences apart from 8, almost above 90% on A for each sequence. The last column in Table 5 shows the R values obtained for each sequence with our evidential reasoning system. The event recognition achieves 100% of R for four sequences, 90% for one sequence, 83% for one sequence, and 80% for one sequence, lower than 50% for one sequence.

Analysis reveals that for sequences 2, 5 and 7 the R values were less than 100%. For sequences 2 and 5 there lacked sitting detections and tracking. This resulted in the PM and PSIT atomic events being incorrect and undetected, which in turn resulted in the composite event P(M/F)BTS being incorrect. Sequence 2 contains five composite events, the resulting R value was 80%. Sequence 5 contains six composite events, the resulting R value was 83%. For sequence 7 there were two atomic events being missed in association to a passenger resulting a composite event, PCS, being undetected. Overall, the sequence contains ten composite events which explains the value of 90% for R.

Sequence 8 performed most poorly with a value of $R = 46\%$. Interestingly, this was also the sequence for which A was lowest at 62%. The ground truth and the system output for this sequence are shown in the tables in Fig. 9.

Here we can see from the ground truth table that there are in total 13 composite events: FBTS11 and PEX for person P1; MBTS19 and PEX for person 2; FBTS9 and PEX for person P3; MBTS18, PCS17, and PEX for person P4; FBTS2 and PEX for person P5; MBTS19 and PEX for person P6. However, only six of the events are correctly recognised, i.e. $R = 6/13 = 46\%$. In the case of the composite event PEX, only for person P6, it is correctly recognised; for others, the composite event was mistakenly mixed up, that is the PEX of P1 was mistakenly assigned to P4, P2 to P1, P3 to P5, P4 to P3, and P5 to P2. Also for person P2 the composite event PCS6 was incorrectly recognised, namely the male was mistakenly recognised as sitting in seat 6 when in fact he

Table 5
Recognition results for rule-based approach, Bayesian approach, and our evidential reasoning approach.

Sequence	Number of composite events—R			
	Ground truth	Rule-based	Bayesian reasoning	Evidential reasoning
1	4 2 PBTS, 2 PEX	3–75% 1 PBTS, 2 PEX	4–100% 2 PBTS, 2 PEX	4–100% 2 PBTS, 2 PEX
2	5 2 PBTS, 1 PCS, 2 PEX	4–80% 1 PBTS, 1 PCS, 2 PEX	2–40% 1 PBTS, 1 PCS, 0 PEX	4–80% 1 PBTS, 1 PCS, 2 PEX
3	5 2 PBTS, 1 PCS, 2 PEX	2–40% 0 PBTS, 0 PCS, 2 PEX	5–100% 2 PBTS, 1 PCS, 2 PEX	5–100% 2 PBTS, 1 PCS, 2 PEX
4	2 1 PBTS, 1 PEX	0–0% 0 PBTS, 0 PEX	2–100% 1 PBTS, 1 PEX	2–100% 1 PBTS, 1 PEX
5	6 2 PBTS, 2 PCS, 2 PEX	5–83% 1 PBTS, 2 PCS, 2 PEX	4–67% 0 PBTS, 2 PCS, 2 PEX	5–83% 1 PBTS, 2 PCS, 2 PEX
6	7 3 PBTS, 1 PCS, 3 PEX	3–43% 0 PBTS, 1 PCS, 2 PEX	7–100% 3 PBTS, 1 PCS, 3 PEX	7–100% 3 PBTS, 1 PCS, 3 PEX
7	10 4 PBTS, 2 PCS, 4 PEX	1–10% 0 PBTS, 0 PCS, 1 PEX	7–70% 3 PBTS, 0 PCS, 4 PEX	9–90% 4 PBTS, 1 PCS, 4 PE
8	13 6 PBTS, 1 PCS, 6 PEX	2–15% 0 PBTS, 0 PCS, 2 PEX	5–38% 4 PBTS, 0 PCS, 1 PEX	6–46% 5 PBTS, 0 PCS, 1 PEX

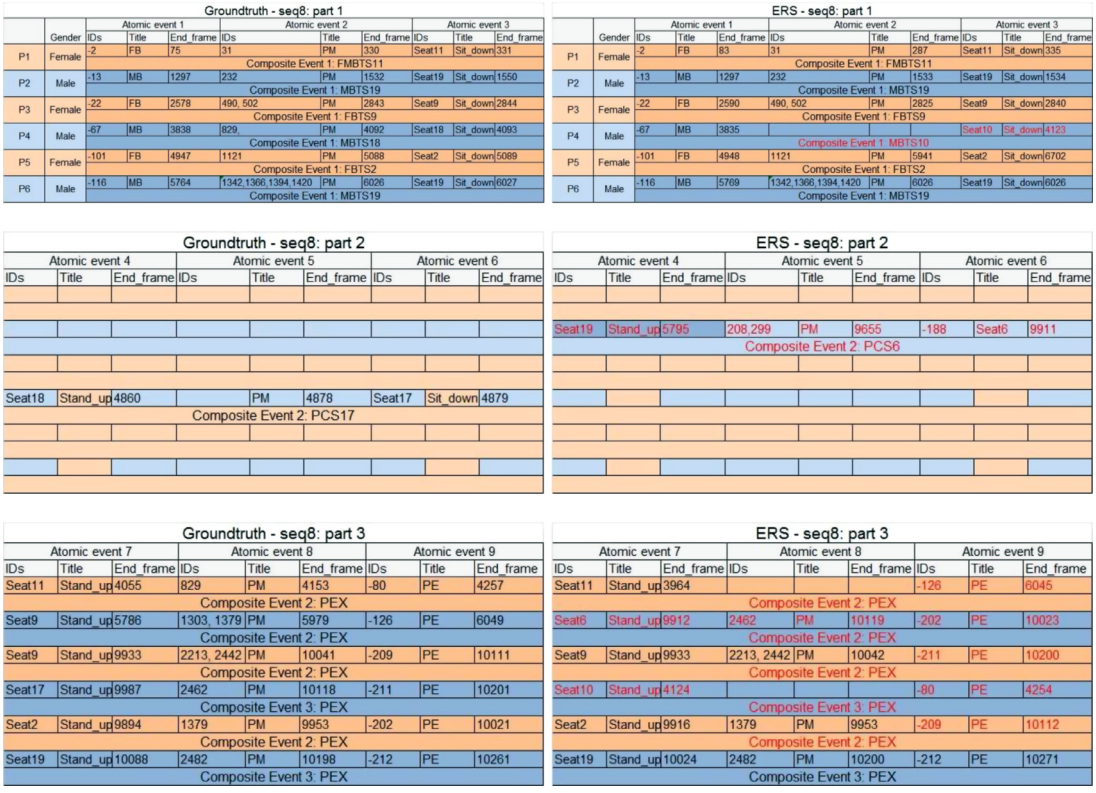


Fig. 9. Ground truth and system output for sequence 8.

had already exited the bus. The most serious mistakes were made on person P4. For the person the composite event MBTS18 was incorrectly recognised, the composite event PCS17 was not detected in addition to mistakenly assigned composite event PEX. To understand this, Table 6 shows a segment of 18 atomic events that were detected.

From Table 6 we can see that event E127 is of type PM, in fact corresponding to the male moving up the gangway towards seat 18, and the female moving back to the exit and exiting the bus. On the left side of Fig. 10 are two images taken from sequence 8 which are a snapshot of the atomic event. Also shown in right side of Fig. 10, are the corresponding track on ground floor, for the PM event E127.

The partitioning at this point is as follows:

$$\omega_1 = \{e_1, \neg e_2, \neg e_3, e_4, \neg e_5, \neg e_6, \neg e_7, e_8, e_9, \neg e_{10}, \neg e_{11}, e_{12}, \neg e_{13}, e_{14}, \ldots, e_{16}, \neg e_{17}, e_{18}, \neg e_{19}, \ldots, \neg e_{22}, e_{23}, \neg e_{24}, \neg e_{25}, e_{26}, e_{27}, \neg e_{28}, \neg e_{29}, e_{30}, \ldots, e_{35}, \neg e_{36}, e_{37}, \neg e_{38}, e_{39}, e_{40}, \neg e_{41}, \ldots, \neg e_{43}, e_{44}, \neg e_{45}, e_{46}, \neg e_{47}, e_{48}, e_{49}, \neg e_{50}, \ldots, \neg e_{76}, e_{77}, e_{78}, \neg e_{79}, \ldots, \neg e_{86}, e_{87}, \neg e_{88}, e_{89}, \neg e_{90}, e_{91}, \neg e_{92}, e_{93}, e_{94}, e_{95}, \neg e_{96}, e_{97}, \neg e_{98}, \neg e_{99}, e_{100}, \neg e_{101}, \ldots, \neg e_{105}, e_{106}, \neg e_{107}, e_{108}, \neg e_{109}, e_{110}, \neg e_{111}, \ldots, \neg e_{113}, e_{114}, \ldots, e_{117}, \neg e_{118}, e_{119}, , \neg e_{120}, \ldots, \neg e_{123}, e_{124}, e_{125}, e_{126}\}$$

$$\omega_2 = \{\neg e_1, \ldots, \neg e_{27}, e_{28}, e_{29}, \neg e_{30}, \ldots, \neg e_{46}, e_{47}, \neg e_{48}, \ldots, \neg e_{51}, e_{52}, e_{53}, \neg e_{54}, \ldots, \neg e_{67}, e_{68}, \neg e_{69}, \ldots, \neg e_{74}, e_{75}, \neg e_{76}, \ldots, \neg e_{126}\}$$

$$\omega_3 = \{\neg e_1, \ldots, \neg e_{53}, e_{54}, e_{55}, e_{56}, \neg e_{57}, e_{58}, \neg e_{59}, e_{60}, \ldots, e_{67}, \neg e_{68}, \ldots, \neg e_{70}, e_{71}, \ldots, e_{73}, \neg e_{74}, \ldots, \neg e_{81}, e_{82},$$

Table 6
Eighteen atomic events detected for sequence 8.

Event idx	Event ID	Event title	Start_frame	End_frame	StartX	StartY	EndX	EndY
E121	−67	gender	3835	3835	0	0	0	0
E122	786	movement id 786	3839	3857	−70	145	−71	144
E123	796	movement id 796	3879	3900	−69	138	−68	138
E124	797	movement id 797	3880	3892	73	130	80	131
E125	808	movement id 808	3911	3919	72	110	70	108
E126	826	movement id 826	3954	3964	73	117	74	116
E127	829	movement id 829	3962	4154	26	−205	27	−195
...								
E139	−72	Seat Sensor ON_10	4123	4123	0	0	0	0

Fig. 10. Sequence 8—top-left: image of male moving close to seat 18; bottom-left: image of female moving away from seat 11, towards the bus door; right: trajectory corresponding to PM event E127 detected by a tracker—TRACK 829.

$$\neg e_{83}, \neg e_{84}, e_{85}, e_{86}, \neg e_{87}, \ldots, \neg e_{95}, e_{96}, \neg e_{97}, \ldots,$$
$$\neg e_{100}, e_{101}, e_{102}, \neg e_{103}, e_{104}, e_{105}, \neg e_{106}, \ldots,$$
$$\neg e_{108}, e_{109}, \neg e_{110}, e_{111}, \ldots, e_{113}, \neg e_{114}, \ldots, \neg e_{119}, e_{120},$$
$$\neg e_{121}, e_{122}, e_{123}, \neg e_{124}, \ldots, \neg e_{126}\}$$
$$\omega_4 = \{\neg e_1, \ldots, \neg e_{120}, e_{121}, \neg e_{122}, \ldots, \neg e_{126}\}$$

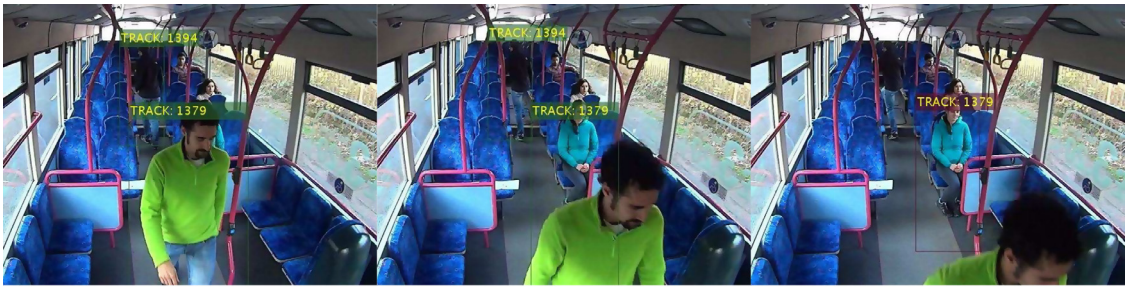
When E127 is detected the system has ruled out partitions ω_2, ω_3 it should be assigned. As its starting half satisfies partition ω_4, the ending half satisfies partition ω_1, the system fails to assign it to any of them. Subsequently at E139 the system incorrectly assigns to ω_4. Continue on, the system fails to correctly assign remaining events to partitions ω_1 and ω_4. Another similar mixed tracker at E214 (Fig. 11), a type of PM, corresponding to person P2 moving to the exit and person P5 staying on seat 2, results more mistakes in event association and consequently incorrectly recognised composite events. For this type of mixed-up atomic events, the system cannot reason to correct assignments. However, when more atomic events are detected, if only the system can revise the beliefs of assigning them, previous incorrect partitions can possibly be corrected.

5.6. Comparison

A simple rule-based approach is chosen as the based line for comparison. In [59] Ma et al. proposed a rule based approach to inferring events of interest by applying rules to combine existing events. Their method employs inference rules to capture new situations, than modifying custom code, hence ensuring a flexible solution for evolving situations. It was initially developed for handling single subject scenarios. It was then adapted by the introduction of linking rules to work on multiple subject environments. The rules are used to link atomic events derived from video analytics by measuring the distance in space or time between two atomic events. Though their inference rules consider imprecision of atomic events derived from video analytics, both inference rules and linking rules make the assumption that the occurrence of each atomic event can be observed, which is not always true considering imperfect video analytics, in particular in a dynamic environment such as on a moving bus platform. When linking atomic events, their involvements in composite event inference are not considered at all. Our evidential reasoning approach powers with the functionality for handling these problems.

We employ DS theory to represent uncertainty in event modelling and event reasoning. DS theory is the generalisation of probability theory, which allows the representation of ignorance due to lack of knowledge. We compared our evidential approach with Bayesian approach, by adapting the evidential reasoning system with probabilities instead of mass functions, in recognising composite events from the set of atomic events having associated to a person. Bayesian approach lacks abilities of handling the problem of incomplete information in event reasoning.

The R values obtained for each sequence with the rule-based approach and Bayesian approach are shown in the third and fourth



(a) frame 5966 (b) frame 5977 (c) frame 5982

Fig. 11. Sequence 8—instances of event E214 detected by a tracker—TRACK 1379.

columns respectively, together with those by our evidential reasoning system in the last columns, in Table 5.

6. Conclusions

In this paper, we propose a novel approach for detection and recognition of composite events on video sequences where multiple subjects present. First, video-analytics and senor measurements are generated in the shape of events. Second, event association and composition are performed by combining the techniques of temporal relation representation, DS theory of evidence and hierarchical network modelling. Our approach can be used to correctly recognise composite events while separating atomic events of multiple subjects with the ability of handling the uncertainty in the video analytics.

Our framework has been evaluated on a real bus environment. The results show the promising performance of the proposed framework. Comprehensive tests on more video data collected from applications and comparison against state-of-art techniques are being performed as future work.

Acknowledgments

This work has been in part supported by UK EPSRC under Grants EP/G034303/1 and EP/N508664/1. Dr. H. Zhou is also supported by UK EPSRC under Grant EP/N011074/1. We would like to thank Fabian Campbell-West and Bhargav Mitra for preparing video data, Niall McLaughlin for his valuable discussions. Thanks also go to the researchers and Ph.D. students in the group for giving their time to participate in the experiments.

Appendix A. Constraint rules

Table A.1
List of rules for partitioning.

Golden rule: R0
Statement: An atomic event cannot be carried out by more than one person.
Premise: $E_i.eType \in \{PB, PM, PSIT, PSTD, PE\}$
Condition: $E_i \in S_m \wedge e_i \in \omega_m$
Result: $E_i \notin S_n, \neg e_i \in \omega_n, n \neq m$

Constraint rule: R1
Statement: If only one person presents in a period of time, all atomic events can only be undertaken by the person.
Premise: $E_i.eType \in \{PB, PM, PSIT, PSTD, PE\}$
Condition: $S = S_1 \wedge \Omega = \omega_1$
Result: $E_i \in S_1, e_i \in \omega_1$

Constraint rule: R2
Statement: A person can only aboard a bus once in a period of time.
Premise: $E_i.eType \in \{PB\} \wedge E_j.eType \in \{PB\}$
Condition: $E_i \in S_m \wedge e_i \in \omega_m$
Result: $E_j \notin S_m, \neg e_j \in \omega_m$

(*Continued*)

Table A.1 (*continued*)

Constraint rule: R3
Statement: A person can only hold one track at a time
Premise: $E_i.eType \in \{PM\}, E_j.eType \in \{PM\}$
Condition: $E_i.time\ ol\ E_j.time \wedge E_i \in S_m \wedge e_i \in \omega_m$
Result: $E_j \in S_n, e_j \in \omega_n, n \neq m$

Constraint rule: R4
Statement: One person can only appear at one place at a time.
Premise: $E_i.eType, E_j.eType \in \{PB, PM, PSIT, PSTD, PE\}$
Condition: $E_i.location \neq E_j.location \wedge E_i.time\ ol\ E_j.time \wedge E_i \in S_m \wedge e_i \in \omega_m$
Result: $E_j \in S_n, e_j \in \omega_n, n \neq m$

Constraint rule: R5
Statement: Two atomic events with the same object ID are carried out by a same person.
Premise: $E_i.eType, E_j.eType \in \{PB, PM, PSIT, PSTD, PE\}$
Condition: $E_i.oID = E_j.oID \wedge E_i \in S_m \wedge e_i \in \omega_m$
Result: $E_j \in S_m, e_j \in \omega_m$

Constraint rule: R6
Statement: Any atomic event happens before a person boards the bus is carried out by other persons.
Premise: $E_i.eType \in PB, E_j.eType \in \{PB, PM, PSIT, PSTD, PE\}$
Condition: $E_j.time\ b\ E_i.time \wedge E_i \in S_m$
Result: $E_j \in S_n, e_j \in \omega_n, n \neq m$

Constraint rule: R7
Statement: Any atomic event happens after a person has exited the bus is carried out by other persons.
Premise: $E_i.eType \in PE, E_j.eType \in \{PB, PM, PSIT, PSTD, PE\}$
Condition: $E_j.time\ a\ E_i.time \wedge E_i \in S_m$
Result: $E_j \in S_n, e_j \in \omega_n, n \neq m$

Constraint rule: R8
Statement: One person cannot carry out two different atomic events at a time.
Premise: $E_i.eType, E_j.eType \in \{PB, PM, PSIT, PSTD, PE\}$
Condition: $E_j.time\ ol\ E_i.time \wedge E_i.eType \neq E_j.eType \wedge E_i \in S_m$
Result: $E_j \in S_n, e_j \in \omega_n, n \neq m$

Appendix B. Case study

To help illustrate how our system works, we describe here an application scenario. In this scenario, two subjects, Alice and Bob, take a bus journey. The bus is a standard single deck bus in use by public transport in Northern Ireland. For recording, two cameras are deployed, one pointing at the front door of the bus, the other at the saloon.

Scenario 1. *At a bus stop, Bob boards the bus and moves to a seat on a row in the middle of the bus saloon, and sits down (Fig. B1a). Alice boards the bus at the next stop and moves to a window seat on the first row, left-hand side, and sits down (Fig. B1b). While Bob stands up and moves to the seat next to Alice and sits down (Fig. B1c). At the following stop, Alice stands up and moves to the door and alights the bus (Fig. B1d). Then Bob stands up and moves to the exit and exits the bus.*

For the purposes of our application scenario, we are interested in the following atomic events: PB = “male or female boards bus”, PM = “person moves from X to Y”, PSIT = “person sits”, PSTD = “person

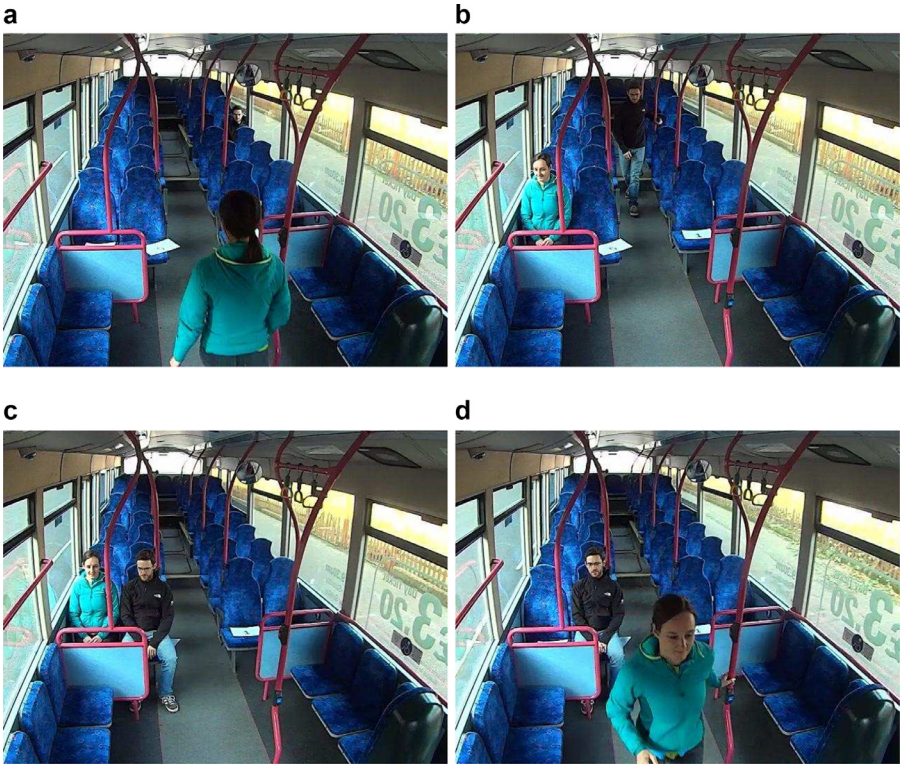


Fig. B1. Four instances of the scenario sequence: (a) female enters; (b) female seats and male stands up, moves; (c) male and female seated; (d) female exits.

stands" and PE = "person exits". The composite events we want to infer from atomic events are: PBTS = "person boards bus and transits to seat", PCS = "person changes seat", and PEX = "person exits bus".

B.1. Atomic event detection

An atomic event E is represented by tuple ($eType$, oID, $date$, $time$, $location$, $source$, $reliaR$, $vFrame$, m) as in Definition 3. $eType$ is the type of the atomic event, such as PB and PM. oID is the identify number assigned by detection. $date$ is the date on which the atomic event has detected. $time$ is an interval of its starting time and ending time. $location$ is the context of spaces the atomic event has covered. $source$ shows which analytic module has provided the detection. $reliaR$ is the reliability of the source. $vFrame$ is the frame of discernment that holds all values that an atomic event of the type can have. For the four types of atomic events, we have

$PB : vFrame = \{MB, FB\};$

$PM : vFrame = \{MS1, \ldots, MS20,\ MGW,\ MDR\};$

$PSIT : vFrame = \{SIT1, \ldots, SIT20,\ \neg SIT\};$

$PSTD : vFrame = \{STD1, \ldots, STD20,\ \neg STD\};$

$PE : vFrame = \{EX, \neg EX\}.$

m is the mass function obtained from a detection.

For the first type of atomic events we employ a camera pointing at the door of the bus. The well-known Jones and Viola face detector is then applied to the acquired video. The output of this is then input to a face-based gender classifier. The resulting output is the probability of the face as being either male or female. Thus, for example, we might have $p(male) = 0.7$ and $p(female) = 0.3$. Based on our training classification accuracy, the module is deemed to have a reliability of $r = 0.9$. Thus, from Eq. (4) we obtain the corresponding mass distribution,

$$m(\{male\}) = 0.7 \times 0.9 = 0.63,\ m(\{female\}) = 0.3 \times 0.9 = 0.27,$$
$$m(\Theta) = 1 - m(\{male\}) - m(\{female\}) = 0.1.$$

As the camera is pointing at the entrance, when we detect a male or female face in its field-of-view, we infer from this either MB or FB, respectively.

$$m(\{MB\}) = 0.63,\ m(\{FB\}) = 0.27,\ m(\Theta) = 0.1.$$

For the PM event we employ a 3d tracker onto the acquired video from the camera pointing at the saloon of the bus. The output of the tracker is a trajectory from which we determine the start-point and the end-point. We then calculate the distance from these points to several schematic locations nearby. These schematic locations consist of all seats, gangway, and door exit. We then use the distance of the tracker to a two closest schematic locations to calculate the mass values for the PM event. For example, for a tracker the distances of its endpoint to the two closest schematic locations, seat 5 and 6, are calculated as $dist(seat5) = 78$ and $dist(seat6,\ \ gangway) = 26$. The corresponding mass functions are then given by

$$m(\{MS5\}) = 26/104 \times 0.8 = 0.2,$$
$$m(\{MS6,\ MGW\}) = 78/104 \times 0.8 = 0.6,\quad m(\Theta) = 0.2$$

where the reliability of 0.8 is derived from the accuracy measurements of the tracker as reported in [56].

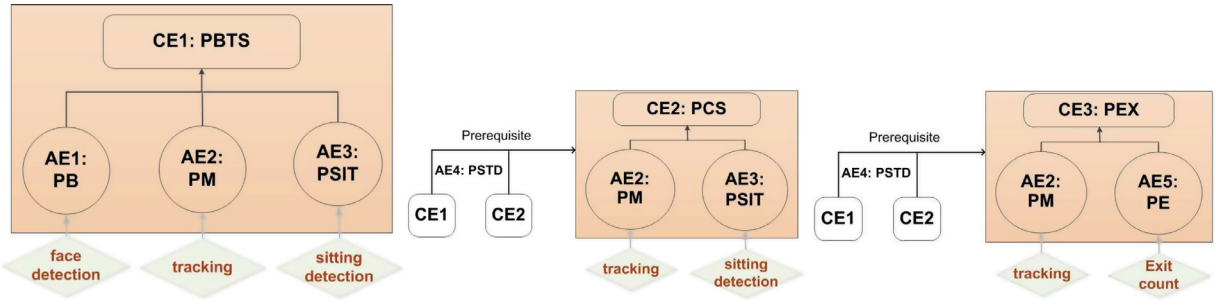
For the PSIT and PSTD events a 3D tracker is used to estimate the shoulder level of an individual in real world space. The resulting output is sitting if the shoulder level goes below a threshold, otherwise standing. PSIT and PSTD are paired together. That means, for example, if there is a SIT9, there should be a STD9 afterwards. Based on our training accuracy, the module is deemed to have a reliability of $r = 0.9$. Thus, from Eq. (4) we obtain the corresponding mass distribution,

$$m(\{SIT9\}) = 1.0 \times 0.9 = 0.9,\ m(\Theta) = 1 - m(\{SIT9\}) = 0.1.$$

From the outputs of video analytics, 26 atomic events are detected for the sequence of scenario 1. The details of oID, $eType$, $time$ in the format of an interval [$Start\ frame$, $End\ frame$], and mass function m, are given in Table B.1. For simplicity, the details of $date$, $location$,

Table B.1
List of atomic events.

Event	oID	eType	Start frame	End frame	Mass function
E1	−2	PB	55	55	$m(\{MB\}) = 0.81, m(\Theta) = 0.19$
E2	3	PM	194	259	$m(\{MS14\}) = 0.43, m(\{MS11, MGW\}) = 0.37, m(\Theta) = 0.2$
E3	−4	PB	769	769	$m(\{MB\}) = 0.09, m(\{FB\}) = 0.81, m(\Theta) = 0.1$
E4	14	PM	894	906	$m(\{MDR\}) = 0.72, m(\{MS3, MGW\}) = 0.08, m(\Theta) = 0.2$
E5	−5	PSIT	896	896	$m(\{SIT2\}) = 0.9, m(\Theta) = 0.1$
E6	−6	PSTD	927	927	$m(\{STD2\}) = 0.9, m(\Theta) = 0.1$
E7	−7	PSIT	948	948	$m(\{SIT2\}) = 0.9, m(\Theta) = 0.1$
E8	18	PM	948	950	$m(\{MS4\}) = 0.64, m(\{MS3, MGW\}) = 0.16, m(\Theta) = 0.2$
E9	−8	PSIT	950	950	$m(\{SIT4\}) = 0.9, m(\Theta) = 0.1$
E10	18	PM	950	1062	$m(\{MS4\}) = 0.45, m(\{MS8\}) = 0.35, m(\Theta) = 0.2$
E11	20	PM	961	1062	$m(\{MS3, MGW\}) = 0.62, m(\{MS4\}) = 0.18, m(\Theta) = 0.2$
E12	−9	PSTD	977	977	$m(\{STD2\}) = 0.9, m(\Theta) = 0.1$
E13	−11	PSIT	1062	1062	$m(\{SIT3\}) = 0.9, m(\Theta) = 0.1$
E14	18	PM	1062	1361	$m(\{MDR\}) = 0.55, m(\{MS3, MGW\}) = 0.25, m(\Theta) = 0.2$
E15	20	PM	1062	1359	$m(\{MS3, MGW\}) = 0.58, m(\{MS4\}) = 0.22, m(\Theta) = 0.2$
E16	44	PM	1170	1184	$m(\{MS7, MGW\}) = 0.64, m(\{MS8\}) = 0.16, m(\Theta) = 0.2$
E17	−13	PSTD	1359	1359	$m(\{STD4\}) = 0.9, m(\Theta) = 0.1$
E18	20	PM	1359	1577	$m(\{MDR\}) = 0.54, m(\{MS3, MGW\}) = 0.26, m(\Theta) = 0.2$
E19	−15	PE	1370	1370	$m(\{EX\}) = 0.8, m(\Theta) = 0.2$
E20	66	PM	1428	1438	$m(\{MS9\}) = 0.47, m(\{MS13\}) = 0.33, m(\Theta) = 0.2$
E21	68	PM	1445	1455	$m(\{MS9\}) = 0.49, m(\{MS13\}) = 0.31, m(\Theta) = 0.2$
E22	−16	PSIT	1448	1448	$m(\{SIT5\}) = 0.9, m(\Theta) = 0.1$
E23	82	PM	1516	1531	$m(\{MS9\}) = 0.45, m(\{MS13\}) = 0.35, m(\Theta) = 0.2$
E24	−19	PSTD	1578	1578	$m(\{STD5\}) = 0.9, m(\Theta) = 0.1$
E25	−18	PSTD	1578	1578	$m(\{STD3\}) = 0.9, m(\Theta) = 0.1$
E26	−20	PE	1586	1586	$m(\{EX\}) = 0.8, m(\Theta) = 0.2$



(a) PBTS (b) PCS (c) PEX

Fig. B2. Three evidential event networks.

source, *reliaR* are not listed. *vFrame* has given at the beginning of this subsection.

B.2. Evidential event networks

Three categories of composite events are concerned: Male/Female Boards bus and Transits to Seat x(PBTS: MBTS/FBTS), Person Changes Seat (PCS), Person EXits bus (PEX). Composite events are consisted of atomic events. For the case study, we can construct three evidential event networks: EEN_{PBTS}, EEN_{PCS} and EEN_{PEX}, presenting the hierarchical structures of the composite events with their atomic events. Fig. B2a –c illustrate three *EEN* respectively.

By Definition 5, we have $EEN_{PBTS} = (ND_{PBTS}, EG_{PBTS}, MM_{PBTS})$, $EEN_{PCS} = (ND_{PCS}, EG_{PCS}, MM_{PCS})$, and $EEN_{PeX} = (ND_{PEX}, EG_{PEX}, MM_{PEX})$. *ND* is a set of event nodes, $ND_{PBTS} = \{AE1, AE2, AE3, CE1\}$, $ND_{PCS} = \{AE2, AE3, CE2\}$, $ND_{PEX} = \{AE2, AE4, CE3\}$. An atomic event node is same as an atomic event in Section B.1, except that it has attribute *pID* indicating to whom it concerns, *level* telling it is an atomic event (or a composite event for a composite event node). For example, $AE1.pID = 1$, $AE1.level =$ '*atomic*'. For a composite event node, its *date* is same as its children at the atomic level, and its *time* interval is decided by the start time of the first child node and the end time of the last child node. *oID*, *location source* and *reliaR* are not required for an composite event node. For the case study, the details of atomic events have been given above. The frame of discernment for a composite event node is as follows:

$$\begin{aligned} PBTS:\ vFrame &= \{MBTS1, \ldots, MBTS20,\ MBTGW,\\ &\qquad FBTS1, \ldots, FBTS20,\ FBTGW,\ \neg PBTS\}\\ PCS:\ vFrame &= \{PCS1, \ldots, PCS20,\ \neg PCS\}\\ PEX:\ vFrame &= \{PEX,\ \neg PEX\} \end{aligned}$$

Each arc of *EG* in an evidential event network represents the relationship between one node to another, which can be represented by a multivalued mapping in *MM*. Table B.2 shows the multivalued mappings for the case study.

B.3. Atomic event association

Now 26 derived atomic events are going to be partitioned into two groups, which are associated to two passengers respectively. Let $\Xi = \{E_1, \ldots, E_{26}\}$ and $\Omega = \{e_1, \neg e_1, \ldots, e_{26}, \neg e_{26}\}$. The goal of event association is to have $S = S_1 \cup S_2, S_1 \cap S_2 = \emptyset$, that also means to have $\omega_1 \subset \Omega$ *and* $\omega_2 \subset \Omega$, satisfying Proposition 1. The association goes through: partitioning Ξ by applying the constraint rules, and if more than two partitionings arise, optimisation by

Table B.2
List of multi-valued mappings.

Relationship	Multivalued mapping.
$AE1 \rightarrow CE1$	$\Gamma(\{MB\}) = \{MBTS1, \ldots, MBTS20, MBTGW\}$, $\Gamma(\{FB\}) = \{FBTS1, \ldots, FBTS20, FBTGW\}$, $\Gamma(\Theta_{AE1}) = \Theta_{CE1}$
$AE2 \rightarrow CE1$	$\Gamma(\{MS1\}) = \{MBTS1, FBTS1\}, \ldots \Gamma(\{MS20\}) = \{MBTS20, FBTS20\}$, $\Gamma(\{MGW\}) = \{MBTGW, FBTGW\}$, $\Gamma(\{MDR\}) = \{\neg PBTS\}$, $\Gamma(\Theta_{AE2}) = \Theta_{CE1}$
$AE3 \rightarrow CE1$	$\Gamma(\{SIT1\}) = \{MBTS1, FBTS1\}, \ldots \Gamma(\{SIT20\}) = \{MBTS20, FBTS20\}$, $\Gamma(\{\neg SIT\}) = \{\neg PBTS\}$, $\Gamma(\Theta_{AE3}) = \Theta_{CE1}$
$AE2 \rightarrow CE2$	$\Gamma(\{MS1\}) = \{PCS1\}, \ldots \Gamma(\{MS20\}) = \{PCS20\}$, $\Gamma(\{MGW\}) = \{\neg PCS\}$, $\Gamma(\{MDR\}) = \{\neg PCS\}$, $\Gamma(\Theta_{AE2}) = \Theta_{CE2}$
$AE3 \rightarrow CE2$	$\Gamma(\{SIT1\}) = \{PCS1\}, \ldots \Gamma(\{SIT20\}) = \{PCS20\}$, $\Gamma(\{\neg SIT\}) = \{\neg PCS\}$, $\Gamma(\Theta_{AE3}) = \Theta_{CE2}$
$AE2 \rightarrow CE3$	$\Gamma(\{MS1\}) = \{\neg PEX\}, \ldots \Gamma(\{MS20\}) = \{\neg PEX\}$, $\Gamma(\{MGW\}) = \{\neg PEX\}$, $\Gamma(\{MDR\}) = \{PEX\}$, $\Gamma(\Theta_{AE2}) = \Theta_{CE3}$
$AE4 \rightarrow CE3$	$\Gamma(\{EX\}) = \{PEX\}$, $\Gamma(\{\neg EX\}) = \{\neg PEX\}$, $\Gamma(\Theta_{AE4}) = \Theta_{CE3}$

selecting the most probable partitioning with a minimum conflict factor.

With domain knowledge, we have constraints to guide the association of the atomic events. The specific constraints being applied to the scenario example are listed in Table A.1 of Appendix A.

Stage 1—Partitioning

Start from E_1 until E_{26}; Golden Rule R0 always applies;

(1–2) $E_1.eType = PB, E_2.eType = PM$
Condition: $S = S_1$
Apply: R1
Results: $e_1, e_2 \in \omega_1$
Partitioning:

$$\omega_1 = \{e_1, e_2, e_3, \neg e_3, \ldots, e_{26}, \neg e_{26}\}$$

(3) $E_3.eType = PB$
Condition: $E_3.eType = PB$; $(e_1, e_2) \in \omega_1$
Apply: R2 and R6
Results: initialise $\omega_2 = \Omega$, $e_3 \in \omega_2, \neg e_3 \in \omega_1$; $(\neg e_1, \neg e_2) \in \omega_2$
Partitioning:

$$\omega_1 = \{e_1, e_2, \neg e_3, e_4, \neg e_4, \ldots, e_{26}, \neg e_{26}\}$$
$$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_4, \ldots, e_{26}, \neg e_{26}\}$$

(4) $E_4.eType = PM$
Condition: $\omega = \omega_1 \cup \omega_2$; $E_4.eType = PM$, $E_4.location - E_2.location > \tau_{location}$, $E_4.location - E_3.location < \tau_{location}$, $e_2 \in \omega_1, e_3 \in \omega_2$
Apply: R3
Results: $\neg e_4 \in \omega_1, e_4 \in \omega_2$
Partitioning:

$$\omega_1 = \{e_1, e_2, \neg e_3, \neg e_4, e_5, \neg e_5, \ldots, e_{26}, \neg e_{26}\}$$
$$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, e_5, \neg e_5, \ldots, e_{26}, \neg e_{26}\}$$

(5) $E_5.eType = PSIT$
Condition: $\omega = \omega_1 \cup \omega_2$; $E_5.eType = PSIT$, $E_5.location - (E_2, E_4).location > \tau_{location}$, $e_2 \in \omega_1, e_4 \in \omega_2$
Apply: R4
Results: $\neg e_5 \in (\omega_1, \omega_2)$
Partitioning:

$$\omega_1 = \{e_1, e_2, \neg e_3, \neg e_4, \neg e_5, e_6, \neg e_6, \ldots, e_{26}, \neg e_{26}\}$$
$$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, e_6, \neg e_6, \ldots, e_{26}, \neg e_{26}\}$$

(6) $E_6.eType = PSTD$
Condition: $\omega = \omega_1 \cup \omega_2$; $E_6.eType = PSTD$, $E_6.location = E_5.location$, $e_5 \notin (\omega_1, \omega_2)$
Apply: R4
Results: $\neg e_6 \in (\omega_1, \omega_2)$
Partitioning:

$$\omega_1 = \{e_1, e_2, \neg e_3, \neg e_4, \neg e_5, \neg e_6, e_7, \neg e_7, \ldots, e_{26}, \neg e_{26}\}$$
$$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, \neg e_6, e_7, \neg e_7, \ldots, e_{26}, \neg e_{26}\}$$

(7) $E_7.eType = PSIT$
Condition: same as in (5)
Apply: R4
Results: $\neg e_7 \in (\omega_1, \omega_2)$
Partitioning:

$$\omega_1 = \{e_1, e_2, \neg e_3, \neg e_4, \neg e_5, \neg e_6, \neg e_7, e_8, \neg e_8, \ldots, e_{26}, \neg e_{26}\}$$
$$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, \neg e_6, \neg e_7, e_8, \neg e_8, \ldots, e_{26}, \neg e_{26}\}$$

(8) $E_8.eType = PM$
Condition: $\omega = \omega_1 \cup \omega_2$; $E_8.eType = PM$, $E_8.location - E_2.location > \tau_{location}$, $E_8.location - E_4.location < \tau_{location}$, $e_2 \in \omega_1, e_4 \in \omega_2$
Apply: R4
Results: $\neg e_8 \in \omega_1, e_8 \in \omega_2$
Partitioning:

$$\omega_1 = \{e_1, e_2, \neg e_3, \neg e_4, \neg e_5, \ldots, \neg e_8, e_9, \neg e_9, \ldots, e_{26}, \neg e_{26}\}$$
$$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, \neg e_6, \neg e_7, e_8, e_9, \neg e_9, \ldots, e_{26}, \neg e_{26}\}$$

(9) $E_9.eType = PSIT$
Condition: $\omega = \omega_1 \cup \omega_2$; $E_8.eType = PM$, $E_9.eType = PSIT$, $E_8.time = E_9.time$, $E_8.location = E_9.location$; $\neg e_8 \in \omega_1$, $e_8 \in \omega_2$
Apply: R4
Results: $\neg e_9 \in \omega_1, e_9 \in \omega_2$
Partitioning:

$$\omega_1 = \{e_1, e_2, \neg e_3, \ldots, \neg e_9, e_{10}, \neg e_{10}, \ldots, e_{26}, \neg e_{26}\}$$
$$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, \neg e_6, \neg e_7, e_8, e_9, e_{10}, \neg e_{10}, \ldots, e_{26}, \neg e_{26}\}$$

(10) $E_{10}.eType = PM$
Condition: $\omega = \omega_1 \cup \omega_2$; $E_8.oID = E_{10}.oID$, $\neg e_8 \in \omega_1, e_8 \in \omega_2$
Apply: R5
Results: $\neg e_{10} \in \omega_1, e_{10} \in \omega_2$
Partitioning:

$$\omega_1 = \{e_1, e_2, \neg e_3, \ldots, \neg e_{10}, e_{11}, \neg e_{11}, \ldots, e_{26}, \neg e_{26}\}$$
$$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, \neg e_6, \neg e_7, e_8, e_9, e_{10}, e_{11}, \neg e_{11}, \ldots, e_{26}, \neg e_{26}\}$$

(11) $E_{11}.eType = PM$
Condition: $\omega = \omega_1 \cup \omega_2$; $E_{11}.eType = PM$; $E_2.eType = PM$, $E_2.location = E_{11}.location$; $E_{10}.eType = PM$, $E_{10}.time\ ol\ E_{11}.time$; $e_2 \in \omega_1, e_{10} \in \omega_2$
Apply: R4 and R5
Results: $e_{11} \in \omega_1, \neg e_{11} \in \omega_2$
Partitioning:

$\omega_1 = \{e_1, e_2, \neg e_3, \ldots, \neg e_{10}, e_{11}, e_{12}, \neg e_{12}, \ldots, e_{26}, \neg e_{26}\}$
$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, \neg e_6, \neg e_7, e_8, e_9, e_{10}, \neg e_{11}, e_{12}, \neg e_{12}, \ldots, e_{26}, \neg e_{26}\}$

(12) $E_{12}.eType = PSTD$
Condition: same as in (6)
Apply: same as in (6)
Results: $\neg e_{12} \in \omega_1, \neg e_{12} \in \omega_2$
Partitioning:

$\omega_1 = \{e_1, e_2, \neg e_3, \ldots, \neg e_{10}, e_{11}, \neg e_{12}, e_{13}, \neg e_{13}, \ldots, e_{26}, \neg e_{26}\}$
$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, \neg e_6, \neg e_7, e_8, e_9, e_{10}, \neg e_{11}, \neg e_{12}, e_{13}, \neg e_{13}, \ldots, e_{26}, \neg e_{26}\}$

(13) $E_{13}.eType = PSIT$
Condition: $\omega = \omega_1 \cup \omega_2$; $E_{11}.eType = PM$, $E_{13}.eType = PSIT$, $E_{11}.time = E_{13}.time$, $E_{11}.location = E_{13}.location$; $e_{11} \in \omega_1, \neg e_{11} \in \omega_2$
Apply: R4
Results: $e_{13} \in \omega_1, \neg e_{13} \in \omega_2$
Partitioning:

$\omega_1 = \{e_1, e_2, \neg e_3, \ldots, \neg e_{10}, e_{11}, \neg e_{12}, e_{13}, e_{14}, \neg e_{14}, \ldots, e_{26}, \neg e_{26}\}$
$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, \neg e_6, \neg e_7, e_8, e_9, e_{10}, \neg e_{11}, \neg e_{12}, \neg e_{13}, e_{14}, \neg e_{14}, \ldots, e_{26}, \neg e_{26}\}$

(14) $E_{14}.eType = PM$
Condition: same as in (10)
Apply: R5
Results: $\neg e_{14} \in \omega_1, e_{14} \in \omega_2$
Partitioning:

$\omega_1 = \{e_1, e_2, \neg e_3, \ldots, \neg e_{10}, e_{11}, \neg e_{12}, e_{13}, \neg e_{14}, e_{15}, \neg e_{15}, \ldots, e_{26}, \neg e_{26}\}$
$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, \neg e_6, \neg e_7, e_8, e_9, e_{10}, \neg e_{11}, \neg e_{12}, \neg e_{13}, e_{14}, e_{15}, \neg e_{15}, \ldots, e_{26}, \neg e_{26}\}$

(15) $E_{15}.eType = PM$
Condition: $\omega = \omega_1 \cup \omega_2$; $E_{15}.oID = E_{11}.oID$, $e_{11} \in \omega_1, \neg e_{11} \in \omega_2$
Apply: R5
Results: $e_{15} \in \omega_1, \neg e_{15} \in \omega_2$
Partitioning:

$\omega_1 = \{e_1, e_2, \neg e_3, \ldots, \neg e_{10}, e_{11}, \neg e_{12}, e_{13}, \neg e_{14}, e_{15}, e_{16}, \neg e_{16}, \ldots, e_{26}, \neg e_{26}\}$
$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, \neg e_6, \neg e_7, e_8, e_9, e_{10}, \neg e_{11}, \neg e_{12}, \neg e_{13}, e_{14}, \neg e_{15}, e_{16}, \neg e_{16}, \ldots, e_{26}, \neg e_{26}\}$

(16) $E_{16}.eType = PM$
Condition: $\omega = \omega_1 \cup \omega_2$; $(E_{14}, E_{15}, E_{16}).eType = PM$, $(E_{14}, E_{15}).time\ ol\ E_{16}.time$, $e_{15} \in \omega_1, e_{14} \in \omega_2$
Apply: R3
Results: $\neg e_{16} \in (\omega_1, \omega_2)$
Partitioning:

$\omega_1 = \{e_1, e_2, \neg e_3, \ldots, \neg e_{10}, e_{11}, \neg e_{12}, e_{13}, \neg e_{14}, e_{15}, \neg e_{16}, e_{17}, \neg e_{17}, \ldots, e_{26}, \neg e_{26}\}$
$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, \neg e_6, \neg e_7, e_8, e_9, e_{10}, \neg e_{11}, \neg e_{12}, \neg e_{13}, e_{14}, \neg e_{15}, \neg e_{16}, e_{17}, \neg e_{17}, \ldots, e_{26}, \neg e_{26}\}$

(17) $E_{17}.eType = PSTD$
Condition: $\omega = \omega_1 \cup \omega_2$; $E_{17}.eType = PSTD$, $E_9.eType = PSIT$, $E_{17}.location = E_9.location$, $\neg e_9 \in \omega_1, e_9 \in \omega_2$
Apply: R4
Results: $\neg e_{17} \in \omega_1, e_{17} \in \omega_2$
Partitioning:

$\omega_1 = \{e_1, e_2, \neg e_3, \ldots, \neg e_{10}, e_{11}, \neg e_{12}, e_{13}, \neg e_{14}, e_{15}, \neg e_{16}, \neg e_{17}, e_{18}, \neg e_{18}, \ldots, e_{26}, \neg e_{26}\}$
$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, \neg e_6, \neg e_7, e_8, e_9, e_{10}, \neg e_{11}, \neg e_{12}, \neg e_{13}, e_{14}, \neg e_{15}, \neg e_{16}, e_{17}, e_{18}, \neg e_{18}, \ldots, e_{26}, \neg e_{26}\}$

(18) $E_{18}.eType = PM$
Condition: same as in (15)
Apply: R5
Results: $e_{18} \in \omega_1, \neg e_{18} \in \omega_2$
Partitioning:

$\omega_1 = \{e_1, e_2, \neg e_3, \ldots, \neg e_{10}, e_{11}, \neg e_{12}, e_{13}, \neg e_{14}, e_{15}, \neg e_{16}, \neg e_{17}, e_{18}, e_{19}, \neg e_{19}, \ldots, e_{26}, \neg e_{26}\}$
$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, \neg e_6, \neg e_7, e_8, e_9, e_{10}, \neg e_{11}, \neg e_{12}, \neg e_{13}, e_{14}, \neg e_{15}, \neg e_{16}, e_{17}, \neg e_{18}, e_{19}, \neg e_{19}, \ldots, e_{26}, \neg e_{26}\}$

(19) $E_{19}.eType = PE$
Condition: $\omega = \omega_1 \cup \omega_2$; $E_{19}.eType = PE$, $E_{18}.eType = PM$, $E_{18}.time\ ol\ E_{19}.time$, $E_{17}.eType = PM$, $E_{17}.location - E_{19}.location < \tau_{location}$; $e_{18} \in \omega_1, e_{17} \in \omega_2$
Apply: R4
Results: $\neg e_{19} \in \omega_1, e_{19} \in \omega_2$
Partitioning:

$\omega_1 = \{e_1, e_2, \neg e_3, \ldots, \neg e_{10}, e_{11}, \neg e_{12}, e_{13}, \neg e_{14}, e_{15}, \neg e_{16}, \neg e_{17}, e_{18}, \neg e_{19}, e_{20}, \neg e_{20}, \ldots, e_{26}, \neg e_{26}\}$
$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, \neg e_6, \neg e_7, e_8, e_9, e_{10}, \neg e_{11}, \neg e_{12}, \neg e_{13}, e_{14}, \neg e_{15}, \neg e_{16}, e_{17}, \neg e_{18}, e_{19}, e_{20}, \neg e_{20}, \ldots, e_{26}, \neg e_{26}\}$

(20) $E_{20}.eType = PM$
Condition: $\omega = \omega_1 \cup \omega_2$; $E_{20}.eType = PM$, $E_{18}.eType = PM$, $E_{18}.time\ ol\ E_{20}.time$, $E_{19}.eType = PE$; $e_{18} \in \omega_1, e_{19} \in \omega_2$
Apply: R3 and R7
Results: $\neg e_{20} \in (\omega_1, \omega_2)$
Partitioning:

$\omega_1 = \{e_1, e_2, \neg e_3, \ldots, \neg e_{10}, e_{11}, \neg e_{12}, e_{13}, \neg e_{14}, e_{15}, \neg e_{16}, \neg e_{17}, e_{18}, \neg e_{19}, \neg e_{20}, e_{21}, \neg e_{21}, \ldots, e_{26}, \neg e_{26}\}$
$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, \neg e_6, \neg e_7, e_8, e_9, e_{10}, \neg e_{11}, \neg e_{12}, \neg e_{13}, e_{14}, \neg e_{15}, \neg e_{16}, e_{17}, \neg e_{18}, e_{19}, \neg e_{20}, e_{21}, \neg e_{21}, \ldots, e_{26}, \neg e_{26}\}$

(21–24) $(E_{21}, E_{23}).eType = PM$, $E_{22}.eType = PSIT$, $E_{24}.eType = PSTD$
Condition: same as in (20)
Apply: R3 and R7
Results: $(\neg e_{21}, \ldots, \neg e_{24}) \in (\omega_1, \omega_2)$
Partitioning:

$\omega_1 = \{e_1, e_2, \neg e_3, \ldots, \neg e_{10}, e_{11}, \neg e_{12}, e_{13}, \neg e_{14}, e_{15}, \neg e_{16}, \neg e_{17}, e_{18}, \neg e_{19}, \ldots, \neg e_{24}, e_{25}, \neg e_{25}, e_{26}, \neg e_{26}\}$
$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, \neg e_6, \neg e_7, e_8, e_9, e_{10}, \neg e_{11}, \neg e_{12}, \neg e_{13}, e_{14}, \neg e_{15}, \neg e_{16}, e_{17}, \neg e_{18}, e_{19}, \neg e_{20}, \ldots, \neg e_{24}, e_{25}, \neg e_{25}, e_{26}, \neg e_{26}\}$

(25) $E_{25}.eType = PSTD$
Condition: $\omega = \omega_1 \cup \omega_2$; $E_{25}.eType = PSTD$, $E_{13}.eType = PSIT$, $E_{25}.location = E_{13}.location$, $e_{13} \in \omega_1$, $E_{19}.eType = PE$, $e_{19} \in \omega_2$
Apply: R4 and R7
Results: $e_{25} \in \omega_1, \neg e_{25} \in \omega_2$

Partitioning:

$$\omega_1 = \{e_1, e_2, \neg e_3, \ldots, \neg e_{10}, e_{11}, \neg e_{12}, e_{13}, \neg e_{14}, e_{15}, \neg e_{16}, \neg e_{17}, e_{18}, \neg e_{19}, \ldots, \neg e_{24}, e_{25}, e_{26}, \neg e_{26}\}$$
$$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, \neg e_6, \neg e_7, e_8, e_9, e_{10}, \neg e_{11}, \neg e_{12}, \neg e_{13}, e_{14}, \neg e_{15}, \neg e_{16}, e_{17}, \neg e_{18}, e_{19}, \neg e_{20}, \ldots, \neg e_{25}, e_{26}, \neg e_{26}\}$$

(26) $E_{26}.eType = PE$
Condition: $\omega = \omega_1 \cup \omega_2$; $E_{26}.eType = PE$, $E_{25}.eType = PSTD$, $E_{25}.location - E_{26}.location < \tau_{location}$, $e_{25} \in \omega_1$, $E_{19}.eType = PE$, $e_{19} \in \omega_2$
Apply: R7
Results: $e_{26} \in \omega_1$, $\neg e_{26} \in \omega_2$
Partitioning:

$$\omega_1 = \{e_1, e_2, \neg e_3, \ldots, \neg e_{10}, e_{11}, \neg e_{12}, e_{13}, \neg e_{14}, e_{15}, \neg e_{16}, \neg e_{17}, e_{18}, \neg e_{19}, \ldots, \neg e_{24}, e_{25}, e_{26}\}$$
$$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, \neg e_6, \neg e_7, e_8, e_9, e_{10}, \neg e_{11}, \neg e_{12}, \neg e_{(13)}, e_{14}, \neg e_{15}, \neg e_{16}, e_{17}, \neg e_{18}, e_{19}, \neg e_{20}, \ldots, \neg e_{26}\}$$

In this scenario, there is no multiple partitionings raised. Therefore, the optimisation does not apply.

The final results of atomic event association are as follows.

$$S = S_1 \cup S_2,$$
$$S_1 = \{E_1, E_2, E_{11}, E_{13}, E_{15}, E_{18}, E_{25}, E_{26}\},$$
$$S_2 = \{E_3, E_4, E_8, E_9, E_{10}, E_{14}, E_{17}, E_{19}\}.$$
$$\omega = \omega_1 \cup \omega_2,$$
$$\omega_1 = \{e_1, e_2, \neg e_3, \ldots, \neg e_{10}, e_{11}, \neg e_{12}, e_{13}, \neg e_{14}, e_{15}, \neg e_{16}, \neg e_{17}, e_{18}, \neg e_{19}, \ldots, \neg e_{24}, e_{25}, e_{26}\},$$
$$\omega_2 = \{\neg e_1, \neg e_2, e_3, e_4, \neg e_5, \ldots, \neg e_7, e_8, e_9, e_{10}, \neg e_{11}, \ldots, \neg e_{13}, e_{14}, \neg e_{15}, \neg e_{16}, e_{17}, \neg e_{18}, e_{19}, \neg e_{20}, \ldots, \neg e_{26}\}.$$

B.4. Composite event recognition

Now the atomic events associated to a passenger are going to be transferred to the evidential event networks and to infer the composite events.

Passenger 1 has associated with the atomic event set $\{E_1, E_2, E_{11}, E_{13}, E_{15}, E_{18}, E_{25}, E_{26}\}$. $E_2.eType = PM$, $E_{11}.eType = PM$, $E_2.m_{end} \cap E_{11}.m_{start} = MS14$, $E_{11}.starttime \gg E_2.endtime$, $E_2.startlocation \neq E_2.endlocation$, $E_{25}.eType = PSTD$, $E_{11}.startlocation \neq E_{11}.endlocation$, therefore E_{11} indicates that a composite event ends and another starts. $E_{25}.eType = PSTD$, E_{25} is used as a point that ends a composite event and starts another composite event. E_{15} and E_{18} take place between E_{13} and E_{25}, their evidence support E_{13} staying at seat 3. Thus E_{15} and E_{18} do not contribute to inference of the composite events.

E_1 and E_2 become the nodes at the lower-level in the network EEN_{PBTS} as shown in Fig. B2 a, are used to infer the composite event $CE1$: $PBTS$ as the node at the higher-level. E_{26} is in the network EEN_{PEX}, Fig. B2 c, and is going to infer $CE3$: PEX.

The inference of composite event $CE1$ starts at translating the mass functions of the nodes at the lower-level into the node at the higher-level, and then combine these together. On the combined mass function, $BetP$ on each single element is calculated. The final decision is made on the element with the highest $BetP$.

On the event network CE1: PBTS,

(i) m_{E_1} and m_{E_2} are transferred onto $CE1$ by using Eq. (2) and applying the multivalued mappings in Table B.2. Therefore, we have m_1 and m_2 along vacuous m_3 representing no knowledge on node AE3, as follows.

$m_1(\{MBTS1, \ldots, MBTS20, MBTGW\})$
$= m_{E_1}(\{MB\}) = 0.81$; $m_1(\Theta) = m_{E_1}(\Theta) = 0.19$.
$m_2(\{MBTS14, FBTS14\}) = m_{E_2}(\{MS14\}) = 0.43$;
$m_2(\{MBTS11, MBTGW, FBTS11, FBTGW\})$
$= m_{E_2}(\{MS11, MGW\}) = 0.37$;
$m_2(\Theta) = m_{E_2}(\Theta) = 0.2$.
$m_3(\Theta) = m_{E_3}(\Theta) = 1$.

(ii) Combining $(m_1 \oplus m_2) \oplus m_3$ by Eq. (3), we have m:

$m(\{MBTS14\}) = 0.35$; $m(\{MBTS11, MTGW\}) = 0.30$;
$m(\{MBTS1, \ldots, MBTS20, MBTGW\}) = 0.16$;
$m(\{MBTS14, FBTS14\}) = 0.08$;
$m(\{MBTS11, MBTGW, FBTS11, FBTGW\}) = 0.07$; $m(\Theta) = 0.04$.

(iii) From m, we can calculate $BetP$ by Eq. (5):

$BetP(\{MBTS14\}) = 0.40$; $BetP(\{MBTS11\}) = 0.18$;
$BetP(\{FBTS14\}) = 0.04$; $BetP(\{FBTS11\}) = 0.02$.

With the highest $BetP(\{MBTS14\})$, we reach the decision that composite event **MBTS14**: the male boards the bus and transits to sit on seat 14, is inferred.

On the event network $CE2$: PCS, E_{11} and E_{13} are used to infer $CE2$. The same steps are gone through to reach the decision that composite event **PCS3**: the person changes to seat 3, with $BetP(\{PCS3\}) = 0.92$, is inferred.

On the event network $CE3$: PEX, E_{25} as $AE5$ is used to infer $CE3$. The decision is that composite event **PEX**: the person exits the bus with $BetP(\{PEX\}) = 0.9$, is inferred.

The same procedure applies to passenger 2 with the associated atomic event set $\{E_3, E_4, E_8, E_9, E_{10}, E_{14}, E_{17}, E_{19}\}$. The composite events inferred are **FBTS4**: the female boards the bus and transits to sit on seat 4, **PEX**: the person exits the bus.

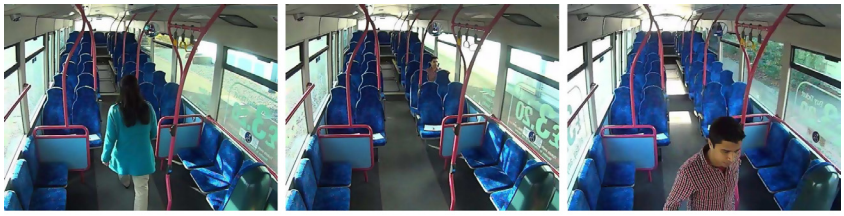
Appendix C. Bus sequences

The first sequence presents a normal bus journey and consists of a male and female boarding the bus, moving into the saloon to a seat and sitting down. After a short period they stand up, move back down the gangway and exit. Fig. C1 shows the example frames of sequence 1.

Sequences 2–3 present a journey in which a passenger changes seat while the bus is moving. This is unusual and is indicative of a passenger who may feel threatened or one who is trying to threaten another passenger. These consist of a male and female entering the saloon and then moving along the gangway to seats and sitting down. After a short period one of them stands and moves to a different seat. At the next bus stop, both passengers stand up and move back down the gangway and exit the bus. With sequence 3, the example frames of the scenario are illustrated in Fig. C2. It is worthy to point out that sequence 2 is used to interpret case studies in the appendices (Appendix B).

Sequence 4 presents a type of threatening behaviour in which one passenger loiters near another who is seated. At a bus stop, a female passenger boards and moves to a seat and sits down. The male at the next stop enters and moves to beside the seat occupied by the female passenger and loiters in the gangway. At the following stop, the female passenger stands up and moves to the exit and exits. The male passenger then follows and moves to the exit and eventually exits. Fig. C3 shows the example frames of sequence 4.

Sequence 5 presents a more threatening behaviour in which both passengers change seat. In this sequence the female passenger enters the bus and moves to seat and sits down. At a different stop, the male passenger enters the bus and moves to the seat right behind the female passenger and sits down. The female passenger then stands



(a) frame 190 (b) frame 1450 (c) frame 2304

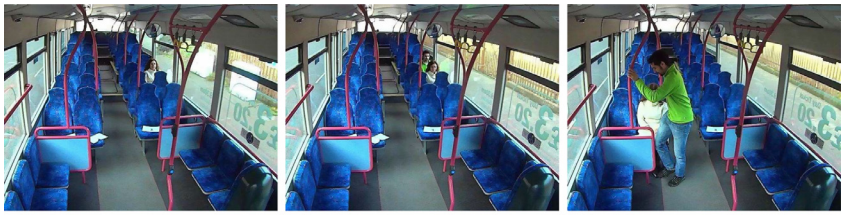
Fig. C1. Example frames of sequence 1.

(a) frame 1133 (b) frame 1458 (c) frame 1672

Fig. C2. Example frames of sequence 3.

(a) frame 459 (b) frame 1426 (c) frame 2423

Fig. C3. Example frames of sequence 4.



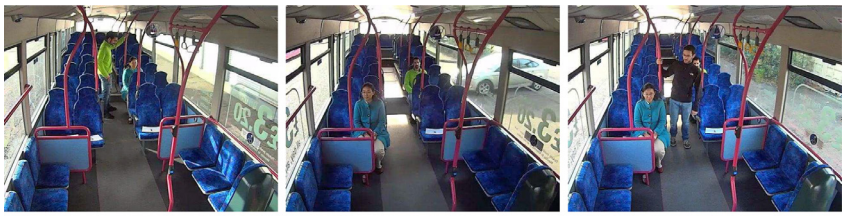
(a) frame 351 (b) frame 1110 (c) frame 1380

Fig. C4. Example frames of sequence 5.

up and moves to a different seat and sits down. The male passenger stands up and moves to the seat beside the female and sits down. The female passenger stands up and moves to the exit and exits the bus. The male stands up and moves to the exit and exits the bus. The example frames of the scenario are shown in Fig. C4.

Sequence 6 consists of three passengers, 2 male and 1 female. Fig. C5 shows the example frames of the sequence. In this experiment, the female passenger enters and moves to seat C-10 and sits down. The first male passenger enters and moves nearby seat C-10 and loiters in the gangway. The female passenger stands up and moves to seat C-3 and sits down. The male passenger sits down on seat C-10, previously occupied by the female passenger. The second male passenger then boards, and moves to beside seat C-3 and loiters in the gangway. The female stands up, moves to the exit and exits the bus. The second male then sits down on seat C-3, vacated by the female passenger. The first male stands up and moves to the exit and exits the bus. Afterwards, the second male passenger stands up and moves to the exit and exits the bus.

Sequence 7 presents a complicated sequence consisting of two male and female passengers with several seat changes and loitering incidents. The scenario is illustrated with the example frames shown in Fig. C6. In this sequence, the first male passenger boards at the first bus stop and moves to seat C-19 and sits down. At the second bus stop, the first female passenger boards and moves to seat C-9 and sits down. At the third stop, the second male passenger boards and moves to the gangway, beside seat C-9, and loiters. The first male passenger stands and moves to gangway. The second male moves to seat C-19 vacated by the first male passenger and sits down. The first male passenger moves to seat C-1 and sits down. At the fourth stop, the second female passenger boards, moves to seat C-2 and sits down. She then stands, moves to seat C-3 and sits down. At the next stop, the first male passenger stands, moves to the exit and exits the bus. The second male passenger stands, moves to the exit and exits the bus. At the last bus stop, the second female passenger then stands, moves to the exit and exits the bus. Lastly, the first female stands, moves to the exit and exits the bus.



(a) frame 1500 (b) frame 2196 (c) frame 2659

Fig. C5. Example frames of sequence 6.

(a) frame 2748 (b) frame 3688 (c) frame 4041 (d) frame 4187

Fig. C6. Example frames of sequence 7.

(a) frame 3358 (b) frame 3554 (c) frame 5917 (d) frame 6054

Fig. C7. Example frames of sequence 8.

The final sequence, 8, is the most complicated one with six people involved, three each of male and female gender. This again involves several seat changes and loitering incidents, and also consists of two passenger passing each other in the gangway. The first female passenger boards, moves to seat C-11 and sits down. At the next stop, the first male passenger boards, moves to seat C-19 and sits down. At the following stop, the second female passenger boards, moves to seat C-9 and sits down. At the fourth stop, the second male passenger boards and moves along the gangway. Meanwhile, the first female passenger stands and exits the bus, and the second male passenger sits down on seat C-18. At the following stop, the third female passenger boards the bus, moves to seat C-2 and sits down. The second male passenger moves to the window seat C-17. At the sixth stop, the third male passenger boards and moves to the gangway. At the same time, the first male passenger stands and passes the third male passenger in the gangway. The first male passenger exits the bus and the third male sits down on seat C-19. At the last stop, the third female stands and exits the bus. The second female moves to the exit and exits the bus, and the second male moves to the gangway. The third male stands. The second male exits the bus, and the third male moves to the gangway, then the exit and exits the bus. Fig. C7 shows the example frames in this video sequence.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.cviu.2015.10.017.

References

[1] T. Moeslund, A. Hilton, V. Kruger, A survey of advances in vision-based human motion capture and analysis, Comput. Vis. Image Understand. 104 (2006) 90–126.

[2] D. Weinland, R. Ronfard, E. Boyer, A survey of vision-based methods for action representation, segmentation and recognition, Comput. Vis. Image Understand. 115 (2011) 224–241.

[3] R. Poppe, A survey on vision-based human action recognition, Image Vis. Comput. 28 (2010) 976–990.

[4] O. Popoola, K. Wang, Video-based abnormal human behavior recognition-a review, IEEE Trans. Syst. Man Cybern. C Appl. Rev. 42 (2012) 865–878.

[5] R. Turaga P.and Chellappa, V. Subrahmanian, O. Udrea, Machine recognition of human activities: a survey, IEEE Trans. Circuits Syst. Video Technol. 18 (2008) 1473–1488.

[6] G. Lavee, E. Rivlin, M. Rudzsky, Understanding video events: a survey of methods for automatic interpretation of semantic occurrences in video, IEEE Trans. Syst. Man Cybern. C Appl. Rev. 39 (5) (2009) 489–504.

[7] A.D. Newton, Crime on public transport, in: Encyclopedia of Criminology and Criminal Justice, 2014, pp. 709–720.

[8] T. deCampos, A survey on computer vision tools for action recognition, crowd surveillance and suspect retrieval, in: XXXIV Congresso da Sociedade Brasileira de Computacao (CSBC), 2014, pp. 1123–1132.

[9] S. Hongeng, R. Nevatia, Multi-agent event recognition, in: Proceedings of ICCV, 2001, pp. 84–91.

[10] I. Atmosukarto, B. Ghanem, N. Ahuja, Trajectory-based fisher kernel representation for action recognition in videos., in: Proceedings of ICPR, 2012, pp. 3333–3336.

[11] D. Ramanan, D. Forsyth, A. Zisserman, Tracking people by learning their appearance, IEEE Trans. Pattern Anal. Mach. Intell. 29 (1) (2007) 65–81.

[12] F. Bashir, A. Khokhar, D. Schonfeld, Object trajectory-based activity classification and recognition using hidden Markov models, IEEE Trans. Image Process. 16 (2007) 1912–1919.

[13] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, A. Blake, Real-time human pose recognition in parts from single depth images, in: Proceedings of CVPR, 2011, pp. 1297–1304.

[14] H. Zhou, H. Hu, H. Liu, J. Tang, Classification of upper limb motion trajectory using shape features, IEEE Trans. Syst. Man Cybern. C Appl. Rev. 42 (6) (2012) 970–982.

[15] L. Bourdev, J. Malik, Poselets: body part detectors trained using 3d human pose annotations, in: Proceedings of ICCV, 2009, pp. 1365–1372.

[16] A. Yao, J. Gall, G. Fanelli, L. Gool, Does human action recognition benefit from pose estimation? in: Proceedings of BMVC, 2011.

[17] A. Kläser, M. Marszalek, C. Schmid, A spatio-temporal descriptor based on 3d-gradients, in: Proceedings of BMVC, 2008, pp. 995–1004.

[18] Y. Ke, R. Sukthankar, M. Hebert, Event detection in crowded videos, in: Proceedings of ICCV, 2007, pp. 1–8.

[19] H. Wang, A. Kläser, C. Schmid, C.-L. Liu, Action recognition by dense trajectories, in: Proceedings of CVPR, 2011, pp. 3169–3176.
[20] D. Oneata, J. Verbeek, C. Schmid, Efficient action localization with approximately normalized fisher vectors, in: Proceedings of CVPR, 2014.
[21] S. Sadanand, J.J. Corso, Action bank: a high-level representation of activity in video., in: Proceedings of CVPR, 2012, pp. 1234–1241.
[22] Y.-L. Tian, R. Feris, A. Hampapur, Real-time detection of abandoned and removed objects in complex environments, in: Proceedings of the Eighth International Workshop on Visual Surveillance, 2008.
[23] Q. Fan, P. Gabbur, S. Pankanti, Relative attributes for large-scale abandoned object detection, in: Proceedings of ICCV, 2013, pp. 2736–2743.
[24] J. Jacques-Jr, S. Mussef, C. Jung, Crowd analysis using computer vision techniques, IEEE Signal Process. Mag. 27 (2010) 66–77.
[25] H. Idrees, N. Warner, M. Shah, Tracking in dense crowds using prominence and neighborhood motion concurrence, Image Vis. Comput. 32 (1) (2014) 14–26.
[26] B. Zhou, X. Wang, X. Tang, Understanding collective crowd behaviors: learning a mixture model of dynamic pedestrian-agents., in: Proceedings of CVPR, 2012, pp. 2871–2878.
[27] S. Yi, X. Wang, C. Lu, J. Jia, L0 regularized stationary time estimation for crowd group analysis, in: Proceedings of CVPR, 2014.
[28] M. Leach, E. Sparks, N. Robertson, Contextual anomaly detection in crowded surveillance scenes, Pattern Recognit. Lett. 44 (2014) 71–79.
[29] S. Cho, H. Kang, Abnormal behavior detection using hybrid agents in crowded scenes, Pattern Recognit. Lett. 44 (2014) 64–70.
[30] J. Kittler, W. Christmas, T. deCampos, D. Windridge, F. Yan, J. Illingworth, M. Osman, Domain anomaly detection in machine perception: a system architecture and taxonomy, IEEE Trans. Pattern Anal. Mach. Intell. 36 (5) (2014) 845–859.
[31] G. Lavee, M. Rudzsky, E. Rivlin, Propagating certainty in petri nets for activity recognition, IEEE Trans. Circuits Syst. Video Technol. 23 (2) (2013) 326–337.
[32] J. Chen, Y. Cui, G. Ye, D. Liu, S.-F. Chang, Event-driven semantic concept discovery by exploiting weakly tagged internet images, in: Proceedings of International Conference on Multimedia Retrieval, 2014, pp. 1:1–1:8.
[33] W. Li, Q. Yu, H. Sawhney, N. Vasconcelos, Recognizing activities via bag of words for attribute dynamics., in: Proceedings of CVPR, IEEE, 2013, pp. 2587–2594.
[34] N. Chomsky, Syntactic Structures, Mouton, 1957.
[35] C. Petri, Communication with automata, Technical Report AD0630125, Defense Tech. Inf. Cntr., 1966.
[36] M. Ryoo, J. Aggarwal, Recognition of composite human activities through context-free grammar based representation, in: Proceedings of CVPR, 2006, pp. 1709–1718.
[37] G. Lavee, A. Borzin, E. Rivlin, M. Rudzsky, Building petri nets from video event ontologies, in: Proceedings of ISVC, Springer-Verlag, 2007, pp. 442–451.
[38] S. Guler, J. Burns, A. Hakeem, Y. Sheikh, M. Shah, M. Thonnat, F. Bremond, N. Maillot, T. Vu, I. Haritaoglu, R. Chellappa, U. Akdemir, L. Davis, An ontology of video events in the physical security and surveillance domain, 2003. http://www.ai.sri.com/~burns/EventOntology.
[39] R. Nevatia, T. Zhao, S. Hongeng, Hierarchical language-based representation of events in video streams, in: Proceedings of the IEEE Workshop on Event Mining, 2003.
[40] R. Romdhane, B. Boulay, F. Bremond, M. Thonnat, Probabilistic recognition of complex event, in: Proceedings of ICCVS, 2011, pp. 122–131.
[41] A. Hakeem, M. Shah, Learning, detection and representation of multi-agent events in videos, Artif. Intell. 171 (8–9) (2007) 586–605.
[42] S. Khokhar, I. Saleemi, M. Shah, Multi-agent event recognition by preservation of spatiotemporal relationships between probabilistic models, Image Vis. Comput. 31 (9) (2013) 603–615.
[43] S.D. Tran, L.S. Davis, Event modeling and recognition using Markov logic networks, in: Proceedings of ECCV, 2008, pp. 610–623.
[44] A. Stolcke, An efficient probabilistic context-free parsing algorithm that computes prefix probabilities, in: Computational Linguistics, MIT Press for the Association for Computational Linguistics, 1995.
[45] Y. Ivanov, A. Bobick, Recognition of visual activities and interactions by stochastic parsing, IEEE Trans. Pattern Anal. Mach. Intell. 22 (8) (2000) 852–872.
[46] A. Kanaujia, T. Choe, H. Deng, Complex events recognition under uncertainty in a sensor network, 2014. arXiv:1411.0085
[47] W. Brendel, A. Fern, S. Todorovic, Probabilistic event logic for interval-based event recognition., in: Proceedings of CVPR, 2011, pp. 3329–3336.
[48] A. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Stat. 28 (1967) 325–339.
[49] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
[50] J. Allen, Maintaining knowledge about temporal intervals, Commun. ACM 26 (11) (1983) 832–843.
[51] W. Liu, J. Hughes, M. McTear, Representing heuristic knowledge in the DStheory, in: Proceedings of UAI, 1992, pp. 182–190.
[52] J. Lowrance, T. Garvey, T. Strat, A framework for evidential-reasoning systems, in: Proceedings of AAAI, 1986, pp. 896–903.
[53] P. Smets, Constructing the pignistic probability function in a context of uncertainty, in: Proceedings of UAI, 1990, pp. 29–40.
[54] H. Xu, Y. Hsia, P. Smets, Transferable belief model for decision making in the valuation-based systems, IEEE Trans. Syst. Man Cybern. A Syst. Hum. 26 (6) (1996) 698–707.
[55] J. Allen, An interval-based representation of temporal knowledge, in: Proc. of IJCAI, 1981, pp. 221–225.
[56] N. McLaughlin, J. Martinez-del Rincon, P. Miller, Online multiperson tracking with occlusion reasoning and unsupervised track motion model, in: Proceedings of AVSS, 2013, pp. 37–42.
[57] X. Hong, Y. Huang, W. Ma, P. Miller, W. Liu, H. Zhou, Video event recognition by Dempster-Shafer theory, in: Proceedings of ECAI, 2014.
[58] X. Hong, W. Ma, Y. Huang, P. Miller, W. Liu, H. Zhou, Evidence reasoning for event inference in smart transport video surveillance, in: Proceedings of ICDSC, 2014.
[59] J. Ma, W. Liu, P. Miller, W. Yan, Event composition with imperfect information for bus surveillance, in: Proceedings of AVSS, 2009, pp. 382–387.