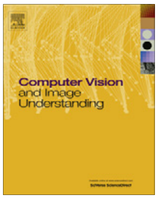
Computer Vision and Image Understanding 116 (2012) 194–209

Contents lists available at SciVerse ScienceDirect

Computer Vision and Image Understanding

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



A suspicious behaviour detection using a context space model for smart surveillance systems

Arnold Wiliem *, Vamsi Madasu, Wageeh Boles, Prasad Yarlagadda

School of Engineering Systems, Faculty of Built Environment and Engineering, Queensland University of Technology, 2 George Street, GPO Box 2434, Brisbane, Queensland 4001, Australia

ARTICLE INFO

Article history:
Received 26 November 2010
Accepted 11 October 2011
Available online 25 October 2011

Keywords:
Suspicious behaviour
Context
Surveillance system
Security

ABSTRACT

Video surveillance systems using Closed Circuit Television (CCTV) cameras, is one of the fastest growing areas in the field of security technologies. However, the existing video surveillance systems are still not at a stage where they can be used for crime prevention. The systems rely heavily on human observers and are therefore limited by factors such as fatigue and monitoring capabilities over long periods of time. This work attempts to address these problems by proposing an automatic suspicious behaviour detection which utilises contextual information. The utilisation of contextual information is done via three main components: a context space model, a data stream clustering algorithm, and an inference algorithm. The utilisation of contextual information is still limited in the domain of suspicious behaviour detection. Furthermore, it is nearly impossible to correctly understand human behaviour without considering the context where it is observed.

This work presents experiments using video feeds taken from CAVIAR dataset and a camera mounted on one of the buildings Z-Block) at the Queensland University of Technology, Australia. From these experiments, it is shown that by exploiting contextual information, the proposed system is able to make more accurate detections, especially of those behaviours which are only suspicious in some contexts while being normal in the others. Moreover, this information gives critical feedback to the system designers to refine the system.

1. Introduction

In the recent years, there is a growing interest in developing automatic detection methods on video surveillance systems for detecting suspicious behaviour. These methods minimise human factors which affects the systems performance to detect security breaches. For example, when fatigue sets in, human observers concentration drops rapidly. In addition, there is a limitation on the number of scenes monitored at the same time by an observer.

Suspicious behaviour has a slightly different notion from anomalous/abnormal behaviour which is generally regarded as an outlier from normal behaviour. An abnormal/anomalous behaviour usually is defined as any behaviour which does not conform with the expected behaviour [1]. On the other hand, suspicious behaviour includes human subjective interpretation. This subjective interpretation renders spotting suspicious behaviour even by human observers is challenging. Generally, human observers rely on their "gut feeling" to make a correct detection. To develop this feeling, they require enormous experience [2]. For instance, a human observer may have "gut feeling" when observing a person loitering in front of a store door late in the night when the store has closed. A possible scenario is that the person might have an intention to commit a crime. So, a normal behaviour could be regarded as suspicious when human subjective interpretation comes into play.

Further understanding of suspicious behaviour requires knowledge of general human behaviour. Some studies in the field of nonverbal behaviour suggest that it is nearly impossible to understand human behaviour without knowing the context in which the behaviour is observed [3]. In other words, a behaviour could only be regarded as suspicious in a particular context, but normal in the other contexts. For example, a person running on a train station platform when there is a train departing is normal, however a person running on a train station platform when the train schedule has finished for that day could be considered as anomalous. So, the notion of normal behaviour needs to be updated over time in order to sufficiently represent the normal behaviour model in the current context. This also renders that it is almost impossible to have labelled training sets which can be used to generate such a normal behaviour model which sufficiently represents normal behaviour in all contexts. Not only that the notion of normal behaviour may change over different contexts, but it is also impossible to have a dataset enumerating all possible human behaviours [4].

* Corresponding author. Current address: National ICT Australia (NICTA), P.O. Box 6020, St Lucia, QLD 4067, Australia.
E-mail address: arnoldw.id@gmail.com (A. Wiliem).

1077-3142/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2011.10.001

From the above discussion, it can be summarised that in order to successfully detect suspicious behaviour, a system needs to have the following capabilities: (a) a capability to continuously extract and learn both contextual and human behavioural information from the incoming stream of video data; (b) a capability to exploit contextual information in making decisions; (c) a capability to incorporate human observers' knowledge in an effective manner.

The existing approaches which can be categorised into misuse detection approaches [5–11], and anomaly detection approaches [4,12–16,11,17–20], in abnormal/anomalous behaviour detection only partially address the posed requirements. Most of them focus on how the system continuously learn from the incoming stream of video data. Only a few papers in the literature that go beyond this by introducing the possibility of using contextual information to get a better system performance [19,17]. Furthermore, human observers' knowledge is limitedly used.

In this paper, we develop a system which attempts to address the posed requirements in order to adequately detect suspicious behaviour. This is achieved by proposing several components. First, we introduce context space model. The model allows not only the system designers to select important information which can be used to describe a context, but also the system to distinguish between two different instances of contexts. Secondly, we introduce the use of data stream clustering algorithm as a means to enable the system to continuously update its knowledge from the incoming video data. In addition, the algorithm also enables efficient and effective knowledge retrieval which is able to retrieve the knowledge learned from a particular context. The third main component of this system is the inference algorithm which combines both contextual information and the system knowledge to make decisions. Furthermore, the algorithm also provides an efficient and effective interface for human observers to feed their knowledge.

Contributions. From the above discussion, this paper proposes the following contributions: (1) a context-based system for detecting suspicious behaviour; (2) a context space model that provides the context features used to describe different contexts; and (3) an inference algorithm to exploit both the system's knowledge and the human observers' knowledge to make correct decisions.

This paper is organised as follows: first, some related work are discussed. Then, in Section 3, we discuss the overall system diagram describing how each element relates to the others. The proposed context space model is discussed in the following section. After that, we present the proposed data stream clustering algorithm used for updating the system's knowledge utilising context information. Then, the inference algorithm used for making decisions is presented. Finally, we present a series of experiments on both CAVIAR dataset and a dataset taken from a real life system deployed in Z-Block building, Queensland University of Technology, Australia.

2. Related work

The problems encountered in detecting suspicious behaviours are not new. Some problems were also investigated in the area of network-based intrusion detection systems, where such systems detect data intrusion by examining packets in the network. Approaches in intrusion detection are classified into those utilising a misuse detection model, and those using an anomaly detection model [21]. The misuse detection model attempts to create attack profiles. An intrusion is detected when there are patterns matched with the created profiles. One of the possible models which follow this line of thinking is the probabilistic model recently proposed by Ryoo and Aggarwal [22]. Technically, each activity/behaviour could be classified using a set of rules. Although this model works very well for the known attacks, it will fail to detect new ones. The anomaly detection model was proposed to overcome this problem. The model creates a long-term usage profile. This profile represents the common users' activities. The short-term profile (i.e. the current user patterns) are compared with the long-term profile. An attack is detected when the short-term profile deviates too far from the long-term profile. As for the surveillance area, both of these models are used and investigated. In addition, the latter approach is becoming more popular as it is able to handle unseen patterns.

Table 1 summaries the current known approaches from the literature in the past 10 years. As we can see from the table, rule-based methods have previously been popular. The anomaly detection model started to get more attention when Stauffer et al. [12] proposed a codebook method which could be used to describe behaviour patterns (i.e. trajectories of pedestrians and vehicles). The code book is constructed by using a vector quantisation method which slowly merges the existing behaviour patterns to k-prototypes. Another well-known method is the one proposed by Vaswani et al. [13]. They proposed the use of shape feature to calculate the common shape of walking paths in an airport scenario. Any walking path which deviated significantly from the common shape is labelled as suspicious. The shape feature is very attractive as it allows the system to calculate the average shapes which represent the common walking paths in the scene. However, averaging behaviour patterns is not always applicable to represent the common patterns. For example, if we average the speed of a person, it would be hard to distinguish between the running and walking actions.

Measuring the deviation of a behaviour pattern from the others could be done in different ways. For example, one may describe the deviation in terms of whether a behaviour can be constructed from the normal behaviour patterns database or not [14,4]. Another way to determine the deviation would be by simply considering a behaviour that cannot be classified into one of the known normal behaviour categories, as the one deviates from normal [18]. These approaches, however, have to have a complete set of normal behaviour. This becomes problematic in real-life scenarios since it is difficult to define all possible normal behaviours. To illustrate the difficulty one may look at what Jianbo et al. coined [4]. They pointed out that the number of suspicious behaviour types are less than the normal ones. In addition, defining all possible suspicious behaviour is considered a difficult one. So, defining all possible normal behaviour is even a more difficult problem.

Table 1
Summary of the known suspicious behaviour detection methods sorted in a chronological order. R: Rule-based model; A: Anomaly detection model; L: by detecting large deviation; C: by detecting that the behaviour pattern cannot be reconstructed from the normal database.

Authors	Year	Detection model	How to detect	Adaptive normal database
Foresti and Pani [5]	1999	R	–	–
Fung and Jerrat [6]	2000	R	–	–
Ivanov and Bobick [7]	2000	R	–	–
Stauffer and Grimson [12]	2000	A	L	No
Vaswani et al. [13]	2003	A	L	No
Zhong et al. [4]	2004	A	C	No
Niu et al. [8]	2004	R	–	–
Boiman and Irani [14]	2005	A	C	No
Makris and Ellis [15]	2005	A	L	No
Piciarelli Foresti [16]	2006	A	L	Yes
Robertson and Reid [9]	2006	R	–	–
Miyanokoshi et al. [10]	2006	R	–	–
Yamamoto et al. [11]	2006	R	–	–
Duque et al. [17]	2007	A	C	Yes
Yue et al. [18]	2007	A	L	No
Xiang et al. [19]	2008	A	L	Yes
Kratz and Nishino [20]	2009	A	L	Yes
Reddy et al. [23]	2011	A	L	No

One solution to address this problem is to design a system which is able to update its normal behaviour patterns database adaptively. One way to do this is by grouping similar behaviour patterns and defining their representation which will be used for classifying a new incoming behaviour pattern into the existing groups. This method is actually similar to the vector quantisation method used by Stauffer and Grimson [12]. The difference is that vector quantisation requires the number of groups to be known initially. Some approaches such as Kratz and Nishino [20] follow this line of thinking.

Xiang et al. [19] model both abnormal and normal behaviours. Their approach maintains both normal and abnormal models. The behaviour classes in both normal and abnormal models are updated whenever an observed behaviour instance is classified into one of them. To detect whether a behaviour instance is abnormal, they use Likelihood Ratio Test (LRT) which calculates the ratio of likelihood of the behaviour instance being normal or abnormal. Both abnormal and normal behaviour patterns can progress to their counterpart (e.g. abnormal becomes normal). This process is governed by the weight given to each pattern and the values assigned to two thresholds (i.e. a threshold governing the minimum weight value for a normal behaviour pattern and a threshold governing maximum weight value for an abnormal behaviour pattern). This work also suggests the use of context to make a more accurate detection. Unfortunately, although the authors stated that the thresholds are not sensitive to their result, the use of these thresholds becomes one of the system's shortcomings. This is because as the behaviour weights are normalised to one, then naturally, the weights decrease when the number of behaviour classes is increasing. Thus, the threshold values need to be readjusted.

Duque et al. [17] classify behaviour into three categories: normal, unusual (e.g. a person running in a hotel lobby) and abnormal (e.g. violation of restricted areas). They use a Dynamic Oriented Graph (DOG) structure in order to model human trajectories. An unusual trajectory is detected when there is no available model that fits the trajectory. The system is then calculate the probability of unusual trajectories. This can be done by calculating the number of unusual trajectories entering the abnormal nodes (i.e. the observers need to define the abnormal nodes which represent restricted areas). The DOG structure is updated over time. Although there are many interesting aspects that the work explored, it depends on the DOG structure. This renders the work inapplicable for other types of features. Furthermore, the approach does not address how to deal with different circumstances which may have totally different interpretation on what abnormality, unusual and normal are.

From previous approaches, there are some common issues which need to be addressed. Basically, to detect suspicious behaviour, one needs to compare it with a set of normal behaviour instances. Most methods rely on an unrealistic assumption which says there is a complete set of normal behaviour. The reason why this assumption is unrealistic is that it is impossible to construct a normal behaviour database which consists of all possible normal behaviour patterns [4]. In other words, the dataset always has insufficient possible types of normal behaviour. This means that any method that relies only on training sets will have over-fitting issues. Furthermore, normal behaviour model needs to be updated so that it represents normality in the current context.

Although some approaches [17,19,24–26] suggest the possibility of using contextual information to further increase the system performance, the use of contextual information in the domain of suspicious behaviour detection is still limited. Contextual information can be used for organising the system's knowledge into groups of knowledge. Each group represents knowledge acquired at the same context. By using only a knowledge group which is relevant to the current context, the system performance can be increased. One of the implications when an approach does not consider contextual information the approach will not be able to detect contextual suspicious behaviour. Contextual suspicious behaviour is defined as the behaviour which is only considered as suspicious in a particular context [1].

3. System description

As depicted in Fig. 1, the proposed system consists of three main components: (a) a context space model; (b) a data stream clustering algorithm; (c) an inference algorithm. The context space model is constructed from selected contextual information that is either fed by the system designers or extracted from sources of contextual information. The context space model provides a set

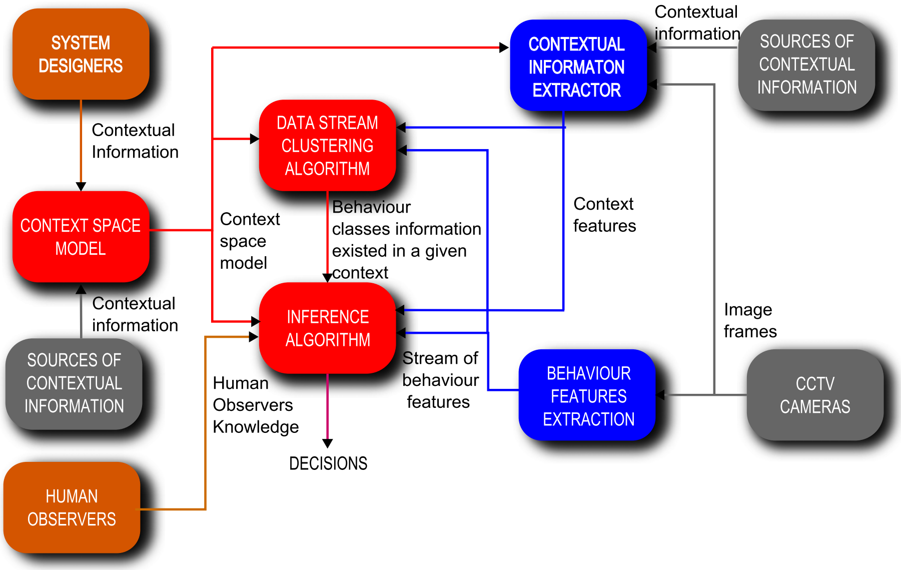

SYSTEM DESIGNERS
Contextual Information
CONTEXT SPACE MODEL
Contextual information
SOURCES OF CONTEXTUAL INFORMATION
HUMAN OBSERVERS
DATA STREAM CLUSTERING ALGORITHM
Context space model
Behaviour classes information existed in a given context
INFERENCE ALGORITHM
Human Observers Knowledge
DECISIONS
Stream of behaviour features
CONTEXTUAL INFORMATON EXTRACTOR
Contextual information
SOURCES OF CONTEXTUAL INFORMATION
Context features
Image frames
BEHAVIOUR FEATURES EXTRACTION
CCTV CAMERAS

Fig. 1. The proposed system diagram. The red boxes are the main system components; the clay boxes represent the sources of information needed by the system; the grey boxes are the sources of information extracted by the system; and the blue boxes represent the system feature extractor components. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of context features used for describing different contexts. This set is extracted by contextual information extractor which takes input either from CCTV cameras or sources of contextual information. By using this set, the system's knowledge can be organised into groups of which the knowledge in each group is contextually related. Then, the knowledge is continuously updated by using data stream clustering algorithm and is used by the inference algorithm to make decision. Finally, the inference algorithm allows an efficient feedback for human observers to feed their knowledge into the system.

From the above discussion, the context space model plays an important role in the system design and operation. It provides a selection of features for describing different contexts. The system then associates any learned knowledge to the context in which the knowledge is acquired. This allows the system to only use knowledge learned from the correct context when making a decision.

4. Context space model

According to the Oxford dictionary, the word 'context' is defined as the parts that immediately precede or follow any particular passage or 'text' and determine its meaning. For instance, the word 'it' in a passage could refer to different subject if we change the preceding parts. Hence, those parts become the context for that word (i.e. the word 'it'). In our case, we broaden this context definition into "any information that would help one to make inferences on the meaning of an object". That information is regarded as independent variables and the subjects whose meaning are influenced by these independent variables are regarded as dependent variables. It is always assumed that both independent and dependent variables are discret variables.

As indicated in Ref. [27] contextual information is one of the important ingredients in the construction of a context. Most context-based approaches describe contextual information as any information that has influence in the understanding of a particular dependent variable [27,28]. For example, the average number of people waiting on a train station platform late at night is less than that in the rush hour. Here, the time influences how one makes an inference on the average number of people. One may draw a conclusion that there is an abnormality by comparing the current average number of people with the average number of people previously extracted from the same period of time. In this case, the current time becomes the contextual information.

When there are two kinds of information influencing each other, each of them can be considered as contextual information for the other. For instance, the information of a keyboard location and the information of a monitor location in a static image could be regarded as contextual information, as keyboards are usually found below a monitor and vice versa [29].

Based on our observations, one of the important properties of contextual information is that the information can be classified as either dependent or independent variables. In other words, the information must have a relationship. Let us suppose that a piece of information is classified as one of the independent variables. This variable then must have a relationship in which it influences the meaning of the dependent variables.

The relationship could either be a one-way relationship, or a mutual relationship. Unlike in a one-way relationship in which the dependent variables do not have influence on the independent variables, the dependent variables in the mutual relationship can have influence on the independent variables. In other words, in mutual relationship, the independent variables can be dependent variables and vice versa. Definition 3 gives the definition of contextual information used in this paper.

Definition 1. (Independent variables) Independent variables constitute information whose meaning is not influenced by other information.

Definition 2. (Dependent variables) Dependent variables constitute information whose meaning is influenced by independent variables.

Definition 3. (Contextual information) A set of information in which its members can be partitioned into two groups. These groups are independent variables and dependent variables groups.

Notice that the above definition considers dependent variables as contextual information. This is because of the possibility that an independent variable can become a dependent variable and vice versa (i.e. in the case of mutual relationships). In addition, although in the case of one way relationships some of the dependent variables may have little influence on independent variables, a group of dependent variables may have stronger influence on the independent variables. In this case, the relationship changes into a two way relationship.

If one selects a couple of different contextual information as independent variables and forms an information space in which these variables are regarded as the base, then such an information space is defined as a context space.

Definition 4. (Context space) Context space is defined as an n-dimensional information space formed by context parameters selected from the contextual information as its bases. The information defined over this information space is referred to as context-sensitive information.

A context space Θ is formally defined as follows. Given a set of contextual information $CI = \{ci_1 \ldots ci_n\}$, there exists subsets CI^1 and CI^2 where $CI = CI^1 \cup CI^2$ and CI^1 and CI^2 have a relationship. Either CI^1 or CI^2 is then chosen as the context space base CSB. The context space Θ defines the sets of context-sensitive information $CSI = \{ci_1, ci_2, ci_3, \ldots, ci_m\}$, where $CSI \in CI^y$, $y = \{1,2\}$ and y depends on the selection of the base (i.e. if CI^1 is chosen as the base, then y is 2 and vice versa). Let us define a mapping function θ as $\theta{:}CSB \to CSI$. In this case, CSB becomes the parameters (or context parameters) of θ and context is defined as the arguments of the function. Fig. 2 presents an illustration of a 3 dimensional context space. A context is then defined as an argument of the mapping function θ.

Here, context space base and context sensitive information can also be regarded as contextual attributes and behavioural attributes. As aforementioned, these attributes are the important ingredients to detect contextual anomalies.

Definition 5. (Context) Given a context space Θ and a function θ: $CSB \to CSI$ (CSB: context space base; CSI: context sensitive information), a context C is defined as an argument of the function θ.

To further clarify the context definition, let us consider the train station example which has been discussed previously. Let f_t be the average number of people waiting on a platform at time t, t be the time, and $train_t$ be the information whether a train is about to arrive or has departed at time t. We know that f_t has one way relationship to t, and $train_t$. For example, there would be more people waiting on the platform when the train is about to arrive than when the train has departed. In this case, t and $train_t$ are selected as context parameters and a two dimensional context space is formed. The f_t becomes the context-sensitive information defined over this space. In other words, the meaning of the value of f_t depends on the given context. Table 2 shows an example of the context space. If the system observes that there are 40 people

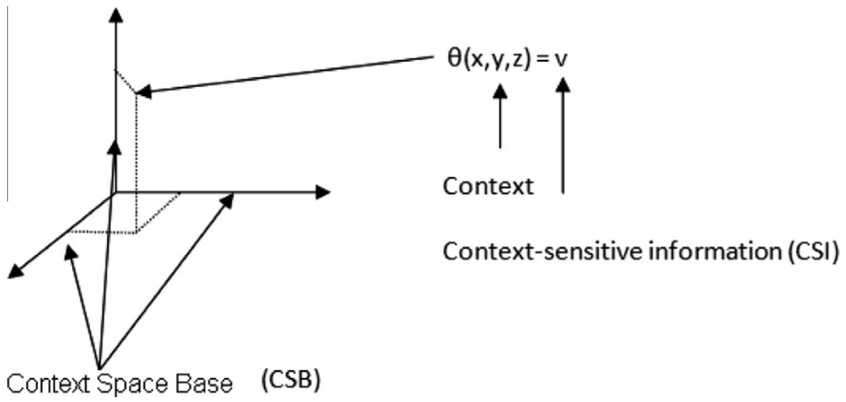


Fig. 2. A graphical illustration of a 3 dimensional context space.

Table 2
An example of context space in the train station platform; CT: current time; AE: arrival event.

	Average number of people on platform	
CT / AE	Train is arriving	Train is not arriving
9.00 am	40	25
11.00 pm	2	0

waiting on the platform at 11.00 pm with no train schedule to arrive, it could flag or alert the human observers that there is an anomaly happening there. In other words, usually in the context of (11.00 pm, train not arriving), the average number of people observed is 2, so if the observed number of people is 40 then it is considered as anomalous.

4.1. Context space model and behaviour model representations

Once the context space model is constructed, the next step is to determine its representation which will be used by the system. Depending on the model complexity, the representations can be as simple as a rule-based representation and increase in complexity to sets representation. For example, if the system is only required to handle a few identified contexts then a rule-based representation can be used. A matrix representation similar to the one in Ref. [30] can be used when the contexts can be enumerated and their number is large. Finally, when not all contexts are identified a priori, then a set representation can be used. By using the set representation the system monitors the values of the context parameters. When at least one of the parameters changes then a context change is detected. In this situation, the system creates a new instance of the context and put it into the set. So, set representation lets the system discover previously unidentified contexts by detecting context change.

In this work, the context space Θ is represented as a set of contexts, $\Theta = \{C_1, C_2, \dots C_m\}$ where $C_i \neq C_j$ if the values of the context parameters are difference (refer to Fig. 2). Each context is represented by context-sensitive information (i.e. dependent variables). For each context C, the context-sensitive information is represented as $C = \{(B_1, f_1), (B_2, f_2), (B_3, f_3), \dots (B_n, f_n)\}$ where B_i is the representation of a behaviour class i and f_i is the frequency of occurrence of behaviour B_i in the context C. The frequency of occurrence of a behaviour class is used by the system in order to measure the commonality level of a behaviour class. The commonality index CV of B_j in the context C_i can be derived using a simple equation as follows:

$$CV(C_i, f_j) = \frac{1}{max(f_k \in F)} f_j \qquad (1)$$

where F is the frequency of occurrence of behaviour classes in context C_i, $F = \{f_1, f_2 \dots, f_n\}$; f_k is the maximum frequency in F. Using this equation, the system measures the commonality level of a behaviour class based on how common a behaviour class is compared with the most often occurring behaviour class. This value is then used to determine the level of how common a behaviour class is.

This work uses three levels of behaviour commonality: significantly common, common and significantly uncommon. These commonality levels are determined by using a simple threshold. ‘Significantly common’ behaviour classes are the ones that have commonality index ranging from T_α to 1, ‘significantly uncommon’ behaviour classes are the ones having value less than T_β or the ones having only one occurrence and the lowest possible commonality index is larger than T_β. The later case assures that every new behaviour class will initially always be detected as a ‘significantly uncommon’ behaviour. Finally, the common behaviour classes are the ones that are neither significantly uncommon nor significantly common (i.e. values between T_α and T_β). Eq. (2) represents how commonality level of a behaviour class B_i in the given context C is determined. The threshold value can either be set by the observers or system designers. In this work, as a default, T_α is set to 0.95 and T_β is set to 0.05. This means that a behaviour class is labelled as significantly common when its frequency of occurrence is only five percent less than the most common class. If its frequency of occurrence is ninety five percent less than the most common class then it is considered as significantly uncommon. These levels will be used in the inference steps.

$$CV\ Level(CV_{B_i}) = \begin{cases} Sig.\ Common & CV_{B_i} \geqslant T_\alpha \\ \quad Common & T_\beta \geqslant CV_{B_i} < T_\alpha \\ Sig.\ Uncommon & CV_{B_i} < T_\beta\ or \\ & \frac{1}{max(f_k \in F)} > T_\beta \\ & and\ CV_{B_i} = \frac{1}{max(f_k \in F)} \end{cases} \qquad (2)$$

There is no specific behaviour representation used in the context space representation. Hence, any behaviour representation can be used in the system as long as a similarity/dissimilarity function which measures the differences between two behaviour instances is provided.

Since contextual information is important in the proposed model, the next logical question is to identify sources from which the information can be extracted. According to Pantic et al. [31], contextual information is usually extracted from various sources. The source of contextual information could vary from one domain to another. One basic guideline is that one may concentrate on the existence of relationships between the context-sensitive information and the contextual information. When a piece of information has a relationship with context-sensitive information then the source of this information could be worth considering.

5. Data stream clustering algorithm

To update the normal behaviour model adaptively, the system requires to have an ability to handle large scale data with limited processing resources and time. In areas where there are many people, several behaviour patterns will emerge in every second. These patterns have to be processed and stored. Since most systems run 24/7, it is impossible to efficiently retrieve the all patterns observed in the past. This is because the historical data may have been removed or archived due to the limited system's storage size. In addition to this work, there is a need for organising the contextually related knowledge into the same group and retrieving only knowledge acquired at a specific context.

In order to address the above requirements, CCTV surveillance system data is modelled as a large scale data stream system. By doing this, a data stream clustering approach can be applied. Specifically, this work uses a data stream clustering algorithm implementation similar to the one developed by Aggarwal [32] which is called Clustream . The Clustream algorithm is chosen as the baseline method because it is able to store and retrieve clustered data stream without requiring any significantly storage size. The algorithm is also able to retrieve information extracted from a given period of time. This ability is very important as the proposed system requires to use only information which is extracted from the same context as the current context.

Clustream algorithm is attractive because of its ability to deal with evolving data streams. This is because in a surveillance system, the context-sensitive information is always evolving depending on the evolution of the context. For instance, in a campus building environment, waiting in front of a class is one of the most common behaviours before a class is scheduled to take place. However, when the class is over, walking would be the behaviour most commonly observed. Clustream algorithm will be able to capture this evolution and store it into snapshots of micro clusters and it will be able to reconcile the snapshots in a time frame to retrieve the context-sensitive information in that time frame.

6. Inference algorithm

When the system has learned the current context, it uses the context information to decide whether an observed instance of behaviour is suspicious or not. The inference algorithm is divided into several steps. Let B is a behaviour class; *Bcur* be the set of observed behaviour classes in the current context $C_{current}$, $Bcur = \{Bcur_1, Bcur_2, \ldots Bcur_n\}$; *Bprev* be the set of behaviour classes previously learned from the same context C_{prev}. Similar to $C_{current}$, $Bprev = \{Bprev_1, Bprev_2, \ldots, Bprev_m\}$; CV_{Bcur_i} be the commonality index of behaviour $Bcur_i$; CV_{Bprev_j} be the commonality index of behaviour $Bprev_j$; *b* be an instance of observed behaviour. The idea is to use the information from the current context $C_{current}$ and the previously learned context C_{prev} to make an inference about an observed behaviour *b*. The same context may reoccur multiple times. Here, the previously learned context C_{prev} and the current context $C_{current}$ are assumed to be the same context appearing at different times. Hence, they must have the same context parameter values. For instance, let us assume that time is the only context parameter for the context space and the current time is 9 o'clock in the morning, so context C_{prev} is the contexts learned by the system at 9 o'clock in the morning yesterday or days before. Then the commonality levels and indices of behaviour b in $C_{current}$ and C_{prev} are derived from the commonality levels and indices from the closest behaviour classes to b in $C_{current}$ and C_{prev} respectively. Fig. 3 illustrates this idea.

The system has five levels of responses when it is given an instance of behaviour. These responses are given based on the information taken from the information of C_{prev} and C_{curr}. The following is the response levels.

- Level I: The behaviour *b* is normal.
- Level II: The information that the system has may not be sufficient to make a decision. The system will not make any decision for a predefined period of time. When *b* is observed again after the predefined period of time and the response level is the same then the response level is increased to III if b_{prev} is common. If b_{prev} is significantly common then the response level is increased to IV.
- Level III: The behaviour b could be suspicious. The observers will be given a low priority notification.
- Level IV: There might be a major unexpected event happening or the context space design may need refining. A moderate priority notification will be given to the observers.
- Level V: There is a high probability that behaviour b is suspicious. The system will give a high priority notification and ask the observers' advice to label the behaviour class B. When the label is available and it is normal then it will not give any notification to the observers.

Fig. 4 presents the inference algorithm flowchart. The flowchart explains two important aspects: determine how much useful context information can be extracted to find out which one of the three cases (refer to Table 3) that the system is in; The second aspect relates to how the system selects the correct the response level given the case and useful context information. The following are the inference steps in detail. The algorithm starts when a new behaviour instance *b* is observed.

1. Classify the observed behaviour *b* to the closest behaviour class $Bcur_i$ in $C_{current}$. Update the behaviour class $Bcur_i$. Since a micro cluster represents each behaviour class, then the update process follows the micro cluster update process.
2. Calculate the commonality index of $Bcur_i$ (CV_{Bcur_i}) using Eq. (1).
3. Determine the commonality level of behaviour $Bcur_i$ ($CVLevel_{Bcur_i}$) using Eq. (2).
4. Find a behaviour class $Bprev_j$ in the learned context C_{prev} which is closest to $Bcur_i$. It is important to know that $Bcur_i$ is a behaviour class in the current context $C_{current}$ and $Bprev_j$ is a behaviour class learned by the system previously in context C_{prev}. The

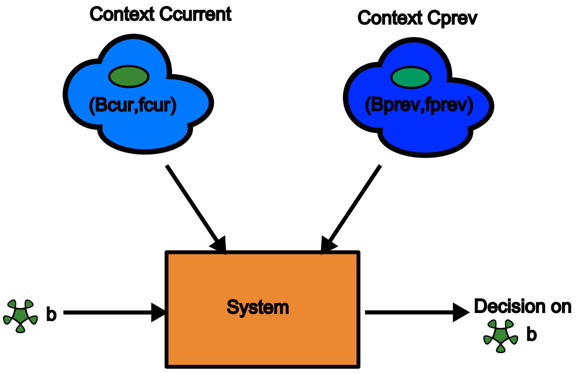


Fig. 3. An illustration of how the system uses the context information for making an inference on the meaning of an observed behaviour *b*. B_{cur} and f_{cur} are the observed behaviours classes in the current context with their frequency of occurrence. B_{prev} and f_{prev} are the observed behaviours classes in C_{prev} with their frequency of occurrence.

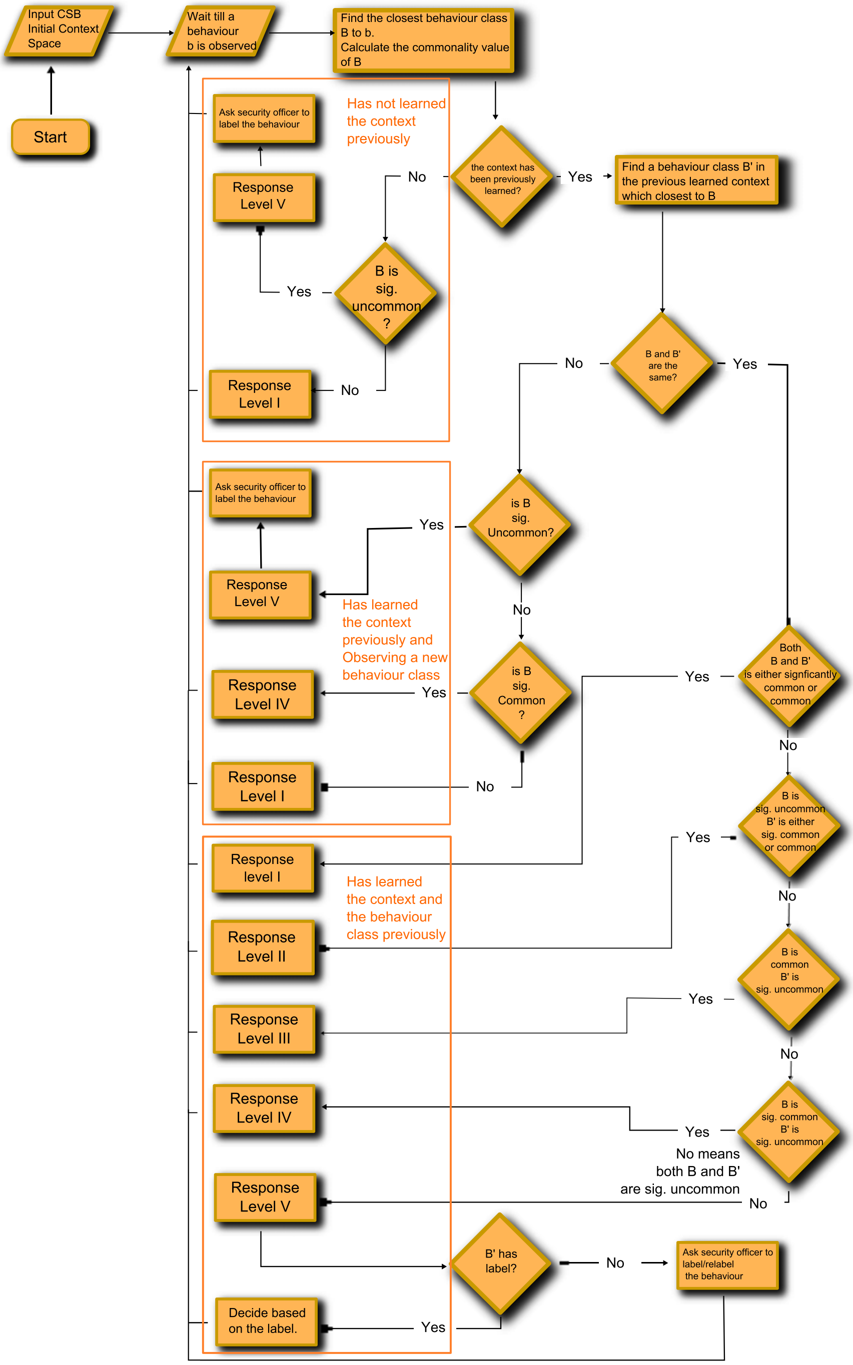


Fig. 4. Inference algorithm flowchart diagram for context-based suspicious behaviour detection system.

Table 3
Possible cases of the information provided by previously learned contexts.

Case	Situation	Possible response levels
I	C_{prev} is not available	Levels I and V
II	C_{prev} is available, and $B_{cur} \notin C_{prev}$	Levels I, IV and V
III	C_{prev} is available, and $B_{cur} \in C_{prev}$	Levels I–V

system initially assumes that context $C_{current}$ and C_{prev} are the same context because they have the same context parameter values. Later in the algorithm, this system will check whether they are the same. If the system has not learned the context previously then it only relies on the $CVLevel_{Bcur_i}$. If it is significantly uncommon, then it will give response level V, otherwise response level I will be given.

5. Determine whether $Bprev_j$ and $Bcur_i$ are significantly different. An inferential statistical test can be used. A z-test or t-test will be used depending on the size of behaviour instance samples in $Bprev_j$ and $Bcur_i$. Let $d(Bprev_j,Bcur_i)$ be the distance between the centroid of $Bprev_j$ and the centroid of $Bcur_i$ then each test measures whether $d(Bprev_j,Bcur_i)$ is significantly different from the distribution between sample distance lengths in one of the behaviour classes. The test has H_0: $d(Bprev_j,Bcur_i)$ is not significantly different in both $Bprev_j$ and $Bcur_i$; H_1: $d(Bprev_j,Bcur_i)$ is significantly different in at least one. Since there are two tests

in this case, the confidence interval τ is determined such that $(1-\tau)^2 = \delta$. δ is set to 0.95. If both $Bprev_j$ and $Bcur_i$ are significantly different then $Bcur_i$ does not exist in C_{prev} then. In this case, the system needs to follow these rules.

(a) If $CVLevel_{Bcur_i}$ is significantly uncommon, it is considered suspicious. The system will give response level V.
(b) If $CVLevel_{Bcur_i}$ is significantly common then it is considered as either normal behaviour which has not appeared in the context previously or the current context $C_{current}$ is different from the context previously learned. The latter case becomes very likely when there is a special event (e.g. fire and terrorist attack) in the current context. The system will give response level IV.
(c) If $CVLevel_{Bcur_i}$ is common then it is considered as normal and the system will give response level I.

6. When both $Bcur_i$ and $Bprev_j$ are considered not significantly different then the system will compare the commonality level of $Bcur_i$ ($CVLevel_{Bcur_i}$) with the commonality level of $Bprev_j$ ($CVLevel_{Bprev_j}$).
   (a) If both are either common or significantly common, the behaviour b is considered as normal. Response level I is given.
   (b) If $CVLevel_{Bcur_i}$ is significantly uncommon and $CVLevel_{Bprev_j}$ is either common or significantly common, then there is a possibility that either the system has not observed $Bcur_i$ sufficiently to make it common or $Bcur_i$ stays significantly uncommon. In either case, the system will wait for a predefined period of time and revaluates when an instance of $Bcur_i$ is observed again. In this case, response level II will be given. When an instance of $Bcur_i$ is observed again after the predefined period of time is over and $Bcur_i$ stays significantly uncommon then if the $Bprev_j$ is significantly common, the system will give response level IV. This is because the posibilites are the same as described in 5.b. Otherwise, response level III will be given.
   (c) If $CVLevel_{Bcur_i}$ is common and $CVLevel_{Bprev_j}$ is significantly uncommon, this means that there is a chance that the behaviour is actually suspicious. The system will give the response level III. In addition, the system also informs the label of $Bprev_j$ when it is available.
   (d) If both $CVLevel_{Bprev_j}$ and $CVLevel_{Bcur_i}$ are significantly uncommon then the system will check $Bprev_j$'s label.
      i. If it is normal then the system will give response level I.
      ii. If it is suspicious then the system will use that label as an indication and notifies the security officers. Additionally it will ask the security officers whether the label is correct.
      iii. If it is unlabelled, the system will notify the security officers that the behaviour could be suspicious and prompts them for their advice (whether or not suspicious). The label is then put into both $Bcur_i$ and $Bprev_j$.

The inference algorithm is able to detect the change in context which might not be captured by the context parameters. This can be seen in Steps 5.b, 6.b and 6.c. The reason why the algorithm can do this is because the composition of common behaviour has a relationship with the context-sensitive information (i.e. commonality level of each behaviour). Hence, this could be one of the context parameter candidates. This work does not put this as one of the context parameters because of its ability to spot unusual/unexpected events (e.g. fire, terrorist attack, etc). Additionally it also can be used for refining the context parameters. When the system is unable to detect a context change, then the selected context parameters need to be evaluated.

Since each context is stored in a snapshot of micro-clusters, it is imperative to define a mechanism to retrieve the micro-clusters efficiently. Let us assume that, the system wants to retrieve the context between 1 pm and 2 pm. In such a scenario, the system first needs to remove information accumulated before 1 pm. Secondly, if the system stores the snapshot every 1 h then the system further needs to reconcile the snapshot stored at 1 pm and the snapshot stored at 2 pm.

6.1. Initialisation

The inference algorithm has to be able to address some issues in initialisation stage. In this stage, most behaviour classes are classified as significantly uncommon regardless of their true commonality level. This is because the system does not have enough information. When the system has previously learned information of the current context then in the first period of time, it will give response level 2, until it has enough information. Furthermore, when the system deals with a new context, then it will notify the observers and wait for a predefined period of time to allow more information to be learned before inferences could be made on the observed behaviour instances.

From the above discussion, it is clear that a behaviour class commonality level changes over time. The inference algorithm gives different meaning on how it progresses. For example, when a behaviour class progressing from significantly common to significantly uncommon and the context does not change, then there could be an unexpected event happening (e.g. terrorist, fight).

7. Experiments and discussions

This experiment section is divided into two parts. The first part is a comparative experiment between the proposed system and the system proposed in [19]. CAVIAR dataset which is available at http://homepages.inf.ed.ac.uk/rbf/CAVIAR/, is used in this comparative experiment. In the second part, the experiments futher validating the benefits of exploiting contextual information are shown. This part uses both CAVIAR dataset and a locally collected dataset (Z-Block). We noted that although experiments with larger datasets (approximately 8 h) have been conducted, this paper focuses on showing the use of context to detect suspicious behaviour.

7.1. Datasets

CAVIAR dataset. – In this work, we only use the first section of the dataset. The first set contains scenarios taken from a wide angle camera lens in the entrance lobby of the INRIA Labs at Grenoble, France. The scenarios are: walking, browsing, resting, slumping or fainting, leaving bags behind, groups of people meeting, walking together and splitting up and two people fighting. Each scenario contains 3–5 clips which lasts 40–60 s. For the experiments, all the scenarios except 'leaving the bags behind' scenarios are used. In total, 23 clips are used in the experiments. The groundtruth in the form of bounding box around each person in each frame is provided by the dataset. Fig. 5 depicts some frames taken from the dataset.

Z-Block dataset. – The Z-Block dataset is a real life video feed taken from one of the CCTV cameras at the foyer of Z Block within the Gardens Point Campus, Queensland University of Technology. The video feeds were collected between September 2009 and November 2009. The dataset consists of two 15 min and 6 min clips taken on the same day in different weeks. The typical activities captured in the feed are students walking, waiting for the next lecture and

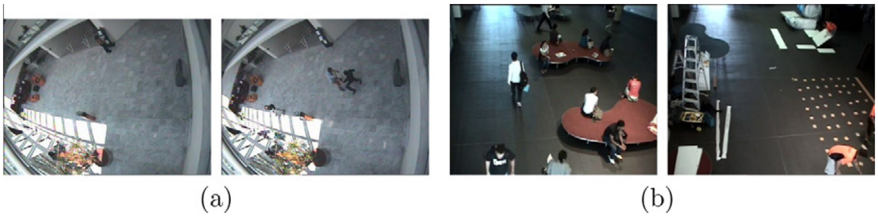


Fig. 5. Some images taken from both the CAVIAR and Z-Block datasets; (a) CAVIAR dataset; (b) Z-Block dataset.

having a group conversations. The frame resolution is 422 × 346 pixels with 25 fps for both parts. In total, the dataset contains 30,000 frames. The groundtruth is obtained via a manual process which was excelled by labelling the groundtruth in sampling rate of 5 fps and interpolating it into a 25 fps sampling rate. Fig. 5 depicts some frames taken from the dataset.

7.2. Human behaviour feature description

All extracted human behaviour in the experiments are represented by using the human behaviour interest point based features proposed in Ref. [33]. Technically, interest point patches are extracted. Then, the tracking information provided by the dataset is used to associate the interest point patches to a person. These patches are then used to represent a person's behaviour.

Each person's behaviour is segmented into behaviour units which have lengths of 1 s. For example, if a person appears in the scene for 3 s, then his/her behaviour will be segmented into three different behaviour units. This segmentation is required in order to avoid making each behaviour too specific. Based on our observation, 1 s behaviour unit is sufficient for the dataset. Then, each behaviour unit is represented by the interest point patches extracted in the duration of the unit. This representation is able to distinguish between basic human actions and their direction of action (i.e. a person walking to the left is considered different from person walking to the right).

As the goal of both approaches is to bring to the attention of human observers a particular scene which is very likely to contain suspicious behaviour, then a correct detection only requires the systems to trigger an alarm at one of the behaviour units. A false negative is counted when the alarm is not triggered for all behaviour units that belong to a suspicious behaviour. Furthermore, false alarm (i.e. false positive) is counted when the approaches trigger an alarm for a behaviour unit which does not belong to a suspicious behaviour.

7.3. Comparative experiment

This section presents a comparative experiment between a system implementing a context space model which we term as "context-based system" and a system which does not use, or implicitly uses context space model which we call "existing system".

7.3.1. Scenario description

We created a simple scenario from the video clips provided in CAVIAR dataset for outlining the advantage of the context-based system over the existing system.

In the scenario, it is assumed that people do not to walk into the hallway at point A (Fig. 6) when after office hours. It is also assumed that this pattern is not identified during the system design. Or in other words, both systems have to discover this by themselves.

In order to create such a scenario, the video clips are organised into two groups of ordered lists. Each group represents a different context. The list of videos are presented in Table 4. Contexts 1 and 2 respectively represent situations during and after office hours. It was assumed that every video clip represents approximately 30 min of scenario time. This means that, the time stamp adds 30 min when a new video clip starts. The time stamp for the first video clips of context 1 and 2 is 8 am and 5.30 pm respectively. The office hours start at 8 am, and finish at 5.30 pm. Meet_Crowd video clip in context 2 contains some people walking into the hallway at point A. These are therefore deemed as anomalies.

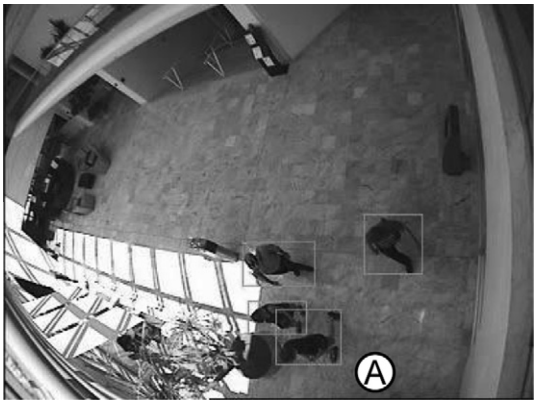


Fig. 6. An illustration of the scenario in CAVIAR dataset.

All the video clips are concatenated into one large video stream. The video stream is then fed into the system being tested. By using this method, all systems would not be aware of the existence of these two different contexts. Apart from the video stream, each system is also given a stream of information about the current time.

7.3.2. Existing system description

The existing system implements the adaptive model approach proposed by Xiang et al. approach [19]. The system has a capability to adapt its normal behaviour model with the current context. A normal behaviour class can be reclassified into abnormal model, and vice versa. Technically, a weight is assigned to each behaviour class. This weight is increased whenever the incoming pattern (i.e. behaviour unit) is classified into the class. The weight also is decreased automatically due to the normalisation of the weights so that their sum must equal one.

We use the same parameter values as in their work [19]. Specifically, Th_{w1}, Th_{w2} and α are set to 0.05, 0.25 and 0.1 respectively. Th_{w1}, Th_{w2} and α are the minimum weight of a normal behaviour model to be still considered as normal, the maximum weight of an abnormal behaviour model, and the learning rate respectively. In order to construct Receiver Operating Characteristics (ROC) plot, we varied $Th_{\mathcal{A}}$, the threshold deciding whether a behaviour pattern is abnormal.

7.3.3. Context space model

According to the scenario, the rate at which person walking into hallway at point A (f_p) and the office hour become the contextual

Table 4
The ordered list of selected video clips for each context.

Context	Video clip names
Context 1	Rest_SlumpOnFloor, Rest_WiggleOnFloor, Meet_Split_3rdGuy, Browse_WhileWaiting1, Browse3, Browse4, Meet_WalkTogether2, Rest_WiggleOnFloor, Split, Walk3, Fight_OneManDown, Meet_WalkSplit, Browse_WhileWaiting2, Browse1, Meet_WalkTogether1, Rest_InChair and Browse2
Context 2	Fight_RunAway1, Fight_Chase, Rest_FallOnFloor and Meet_Crowd

information. The office hour will be represented in terms of time which is discreetised into hour units. So, the context space model can be presented as follows. $CI = \{f_p, \ time\}$, $CSB = \{time\}$ and $CSI = \{f_p\}$.

7.3.4. Results

In order to do comparative analysis between these two systems, we varied the $Th_{\mathcal{A}}$ used in the existing system and T_β in Eq. (2) which is used in the context-based system. The ROC plot is presented in Fig. 7.

The Area Under Curve (AUC) of the proposed system is 0.778 with standard error 0.144. While the AUC of the existing system is 0.460 with standard error 0.153. We use 1-tailed statistic test with confidence interval 95%. The P-value for our and existing systems are 0.027 and 0.602 respectively. This means that our proposed system's AUC is statisticaly greater than 0.5 while the existing system is not statistically greater than 0.5. As we can see from Fig. 7 and the statistical test, our proposed system clearly has a better performance than the existing system. This is because the capability of the context-based system to distinguish these two contexts and exclude the information extracted in context 1 when making decision in context 2.

In the second context, anomalous behaviours only appear in Meet_Crowd video clip. There are four people walking to point A. These people are detected as anomalous by context-based system as soon as they are walking toward point A. The existing system is unsuccessfull in detecting them as anomalies because the behaviour class belongs to these behaviour is classified into the normal behaviour model. This is because the existing system still utilises knowledge learned from context 1 which has a much larger number of occurrence of instances of this behaviour class when making decisions. Unlike the existing system, the context-based system only utilises knowledge learned from the current context.

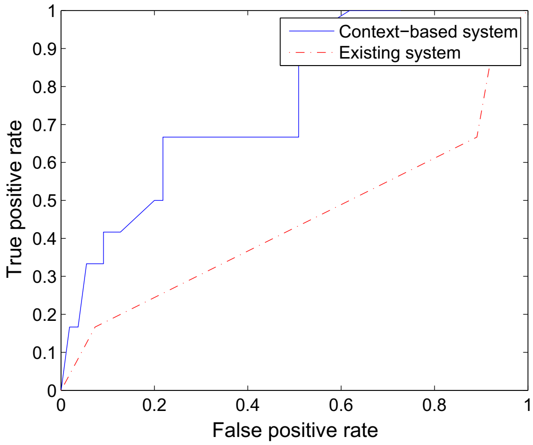


Fig. 7. Receiver Operating Characteristic (ROC) plot of context-based and the existing system proposed in [19]. The Area Under Curve (AUC) and standard error of the proposed and the existing systems are (0.778, 0.144) and (0.460, 0.153) respectively.

Although the existing system employs model adaptation to reflect changes in visual context, knowledge learned in previous context is still considered when making decision. The knowledge will slowly be removed from the system over a period of time. Unlike the existing system, the context-based system automatically excludes the knowledge learned from other different contexts.

7.4. Using information previously learned from the context

In this experiment, we compare the proposed system in two scenarios. The first scenario is when the proposed system has previously dealt with the current context, and the second scenario is when it has not dealt previously. In this experiment the contextually relevant knowledge to the current context is referred as "previously learned context".

Two groups of video clips from the CAVIAR dataset representing two instances of the same context are created. These groups have an equal number of video clips for each scenario. Table 5 presents these groups.

Both contexts contain similar amount of video segments. In here, 'fight' and 'rest' scenarios are considered as suspicious and the rest are considered as normal. The first context is regarded as the context that the system has dealt with previously and the second context is regarded as the current context.

Initially, the proposed system is tested on the second context without using the information extracted from the first context. Then the next experiment is to test the proposed system in the second context by using the information extracted from the first context. To do this, the first context is fed into the proposed system. After that, all the behaviour classes created from this process are either labelled suspicious or normal (i.e. only fight and rest scenarios are considered as suspicious). The labelling process represents the use of human knowledge in the proposed system. Finally, the second context and the information extracted from the first context are fed into the proposed system.

Fig. 8 depicts some examples of behaviour units taken from the common behaviour classes of both contexts.

A quick observation reveals that the micro-clusters between these two contexts can be related. The micro-clusters ID 1, 10, 7 can be associated to the micro-clusters ID 4, 13, 10 depicted in Fig. 8. This signifies that the Clustream is able to create meaningful classes.

Table 5
The list of selected video clips for each context used in the second experiment. These contexts are two different instance of the same context.

Context	Video clip names
Context 1	Meet_WalkSplit, Browse_WhileWaiting2, Browse1, Browse2, Meet_Crowd, Meet_WalkTogether1, Rest_FallOnFloor, Rest_InChair, Walk1, Walk2, Fight_RunAway1, Fight_Chase
Context 2	Meet_Split_3rdGuy, Browse_WhileWaiting1, Browse3, Browse4, Meet_WalkTogether2, Rest_SlumpOnFloor, Rest_WiggleOnFloor, Split, Walk3, Fight_RunAway2, Fight_OneManDown

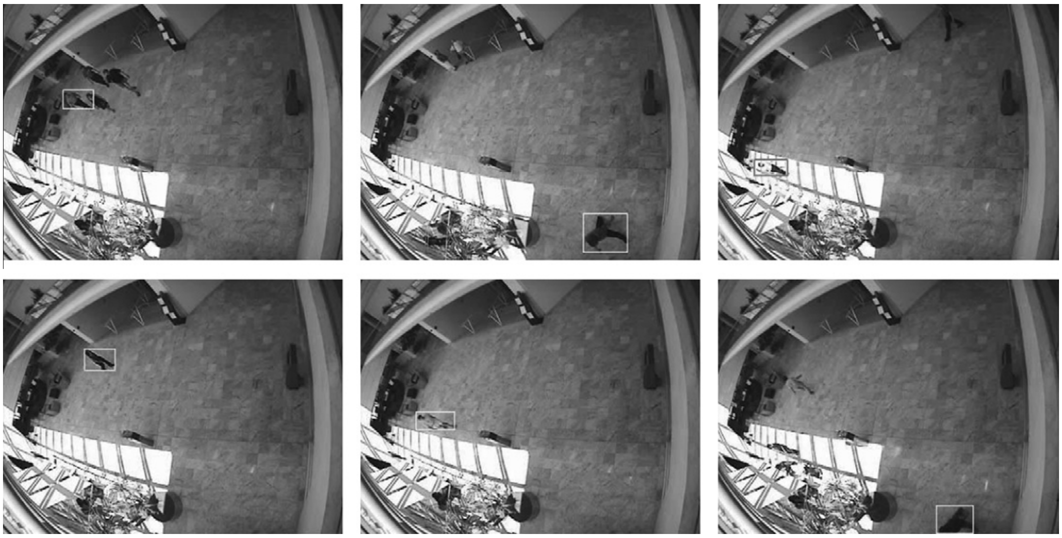


Fig. 8. Example of behaviour units taken from the most common behaviour classes in contexts 1 (first row) and 2 (second row). First row (ordered from the most common class descendingly left to right): micro-cluster ID: 1, 7, and 10; Second row: micro-clusters ID: 4, 13 and 10.

Table 6 presents the results of the experiment. We can clearly observe that false positive rate decreases significantly when the system uses the information extracted from the first context. This is mainly because in this case the proposed system has a better knowledge about the current context. The normal behaviour classes which are not observed before could be observed by the proposed system in the first context. Hence, it will not trigger an alarm when initially these instances appearing.

Fig. 9 depicts decisions made on the proposed systems with the two given scenarios on the behaviour units extracted from Fight_OneManDown. Both the behaviour units of person 4 and person 5 at frame 125–150 are successfully labelled as normal when previously learned context is used. These behaviour units are labelled as suspicious when the previously learned context is not used. When using the previously learned context, the proposed system also labels the behaviour units extracted from person 4 at frame 200–225 and person 5 at frame 175–200 as normal, whilst according to the groundtruth provided by the dataset, these behaviour units should be labelled as suspicious. A quick observation reveals that these behaviour units only contain normal behaviour segments.

Fig. 10 depicts the corresponding image frames taken from the Fight_OneManDown showing the behaviour units which are labelled as suspicious in the groundtruth but detected as normal by the proposed system.

It is also shown in Fig. 9 that both systems have false detections in two behaviour units extracted from person 6 at frame 500–525 and 525–550. This is because the behaviour units are classified as significany uncommon in both current context and previously learned context due to the extreme illumination on the person. Fig. 11 shows some image frames from the behaviour units. The illumination from the sun and the person's white clothing colour disturb the interest point extraction process. This signifies that the feature discriminative power is important to the proposed system performance.

Furthermore, when using previously learned context, the proposed system also successfully detects suspicious behaviour in Rest_WiggleOnFloor video clip which is considered as normal when the proposed system does not use the previously learned context. This can happen because the behaviour instance commonality level in the second context (i.e. current context) is common, however the behaviour is determined as significantly uncommon in the first context (i.e. previously learned context).

Fig. 12 shows the detection results made by both systems at the behaviour units extracted from the Rest_WiggleOnFloor video clip. This video clip contains a person who falls down on the floor. Although according to the provided groundtruth, the behaviour units at frames 650–725 are considered as suspicious, the exact moment when the person falls down is at frames 700–725. When using the previously learned context, the proposed system is able to raise alarm by giving response level III on the behaviour unit at frame 700–725. On the other hand, the proposed system does not raise any alarm when it does not use the previously learned context. From our observations the reason why this happens, could be that because the behaviour unit is miss classified into one of the common behaviour classes. Hence, the behaviour unit is considered as common. However, the system successfully classifies the behaviour unit to the correct behaviour class which is significantly uncommon in the previously learned context. So, when using the previously learned context, the proposed system gives response level III because in the current context, the behaviour unit is considered as common and in the previously learned context it is considered as significantly uncommon. Fig. 13 depicts image frames 701, 705 and 709 of Rest_WiggleOnFloor clip with the response levels made by the proposed system by using the previously learned context. The yellow box means the system gives response level III on the behaviour unit.

Table 6
Confusion matrix representing the results from the second experiment. (a) scenario 1; (b) scenario 2.

	(a)			(b)	
	Suspicious	Normal		Suspicious	Normal
Detected suspicious	5	10	Detected suspicious	6	3
Detected normal	1	88	Detected normal	0	95

7.5. Detecting unexpected events

Unexpected events are detected when the system gives response level IV. This means that the system sees a big difference in the behaviour unit commonality levels between the current context and previously learned context (e.g. significantly uncommon – significantly common or significantly common – significantly uncommon).

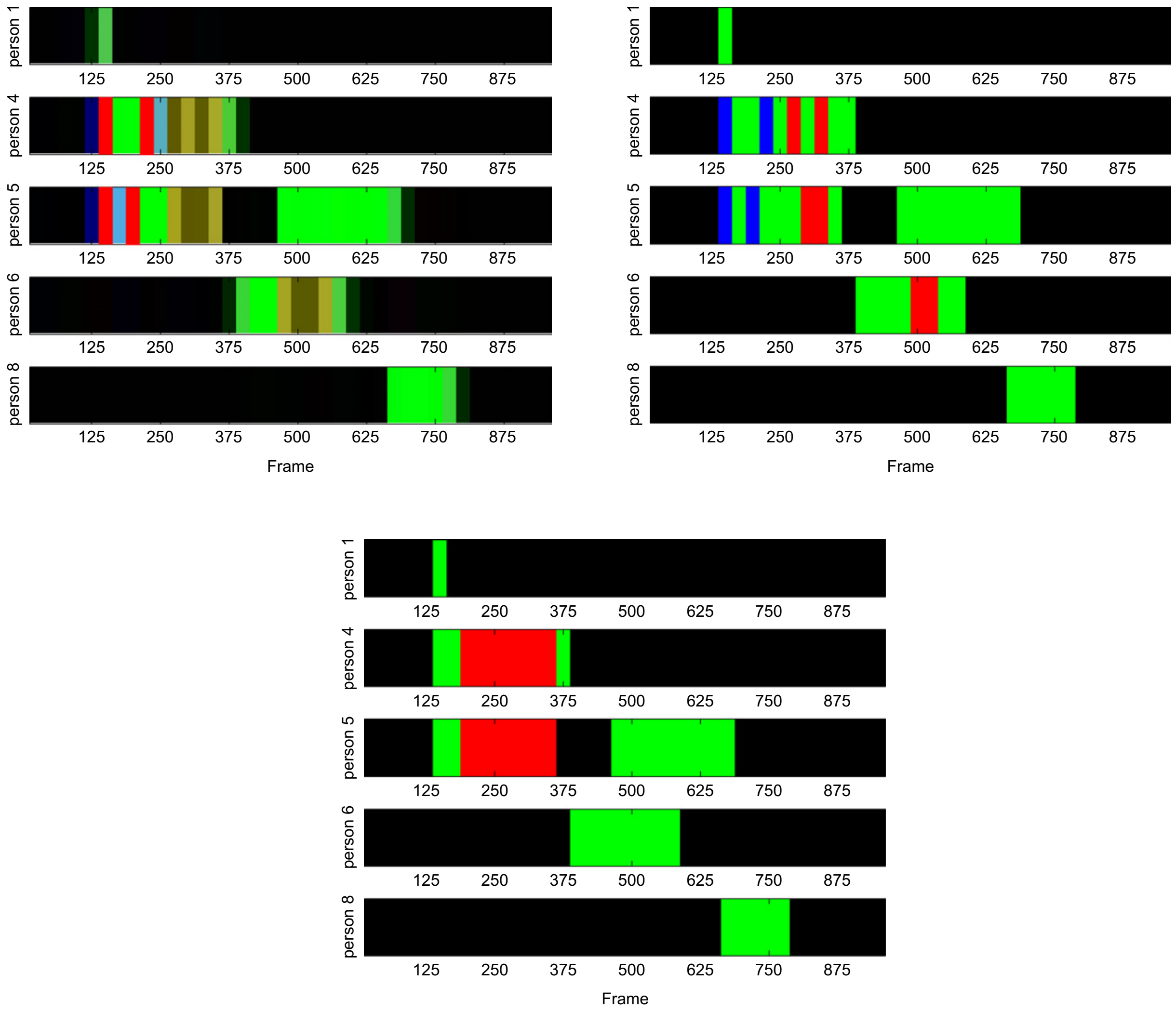


Fig. 9. System response levels on clip Fight_OneManDown. The black colour means the system does not track the person because the person either may not be observed or the number of detected interest points of the person at the given time is less than the threshold. Red colour means response level V. Blue colour means response level II. Left to right: The system response levels without using the previously learned context; Using the previously learned context and the groundtruth. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

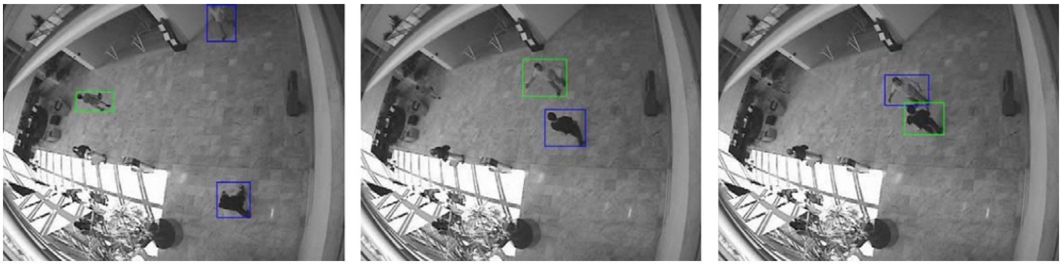


Fig. 10. Image frames taken from Fight_OneManDown with the response levels given by the system using the previously learned context. Left to right: frames 156, 205 and 230.

7.5.1. Context space model

The context space model in this dataset is described as follows. CI = {*Time*, *Lecture schedule*, *commonality level*}, CSB = {*Time*, *Lecture schedule*} and CSI = {*commonality level*}. Different sets of behaviour may appear when there are some lectures scheduled. When a lecture is about to start, there are people queuing in front of the lecture theater. A higher number of people walking are observed when a lecture is finished.

Time is discretised into 30 min time unit. In other words, if the other parameters are not changed, then a context change happens every 30 min.

7.5.2. Dataset

We consider two video clips in Z-Block dataset taken when the context parameters are the same. Specifically, the first clip is taken at noon, Friday, 13th November 2009, while the second clip is

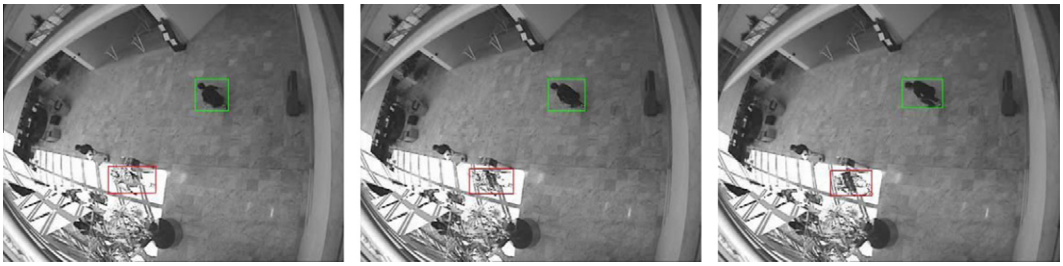


Fig. 11. Image frames taken from Fight_OneManDown with the response levels given by the system. Left to right: frames 521, 525 and 529.

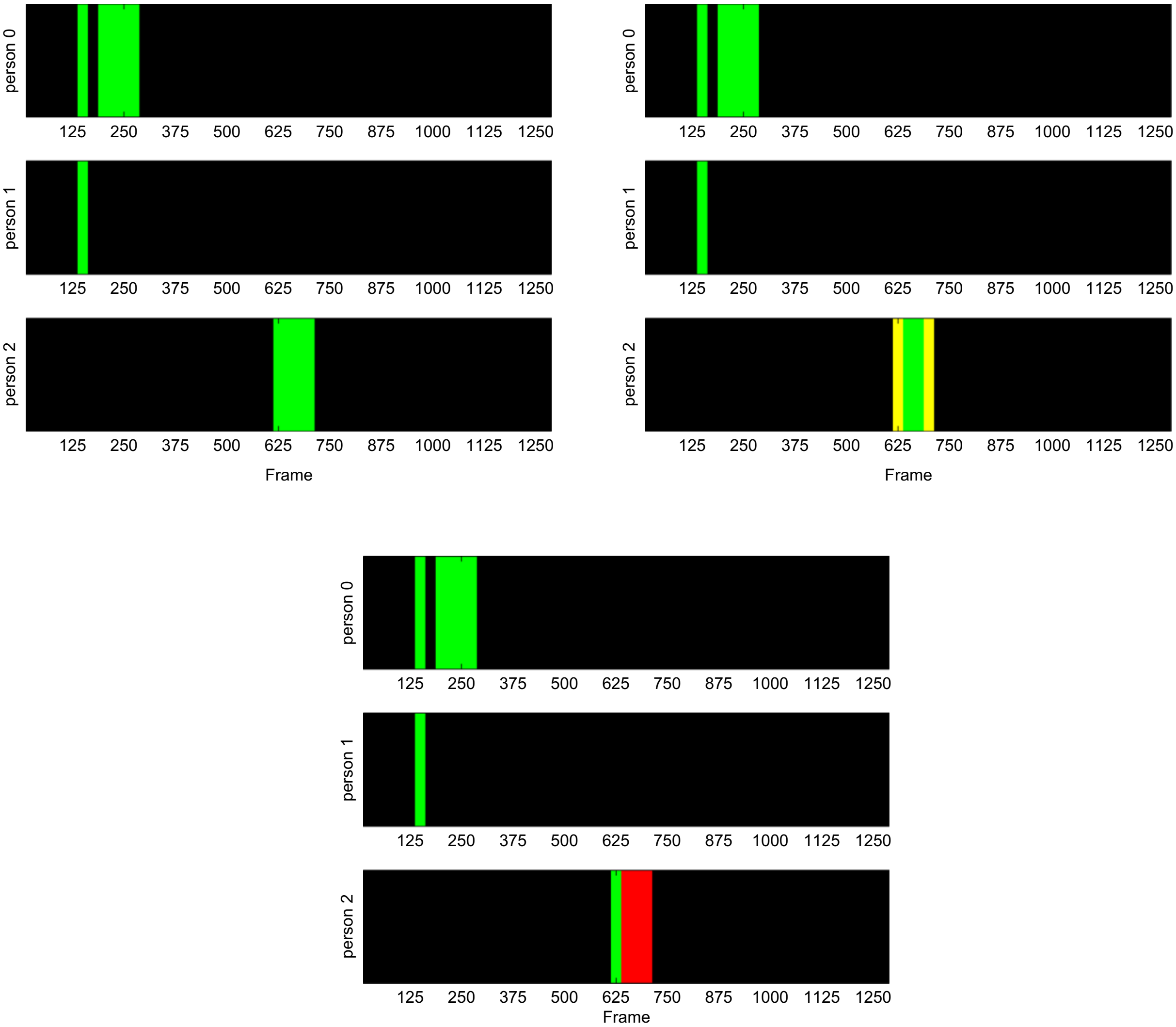


Fig. 12. System response levels on clip Rest_WiggleOnFloor. Left to right: the proposed system without previously learned context; using previously learned context and the groundtruth. The yellow box means the system gives response level III on the behaviour unit.

taken at noon, Friday, 20th November 2009. These days are specifically selected because on the 20th, a construction work takes place for creating temporary stands which to be used for a conference. Fig. 14 depicts some examples taken from both contexts.

Obviously, by monitoring only the context parameters the proposed system will not be aware that there is such activity going on which represents a totally different context. However, since the compositions of the observed behaviour are different between these two contexts, the proposed system can raise response level IV informing that there is an unexpected event happening.

To test whether the proposed system is able to raise response level IV, we consider the first video clip as the previously learned context and the second video clip as the current context. Initially, both the second context and the information extracted from the first context are fed into the system.

7.5.3. Results

We observed that the system is able to raise response level IV. Behaviour depicted in Fig. 15 does not appear in the previously context. However, this behaviour which represents a construction worker working on the ground is quite common in the current context.

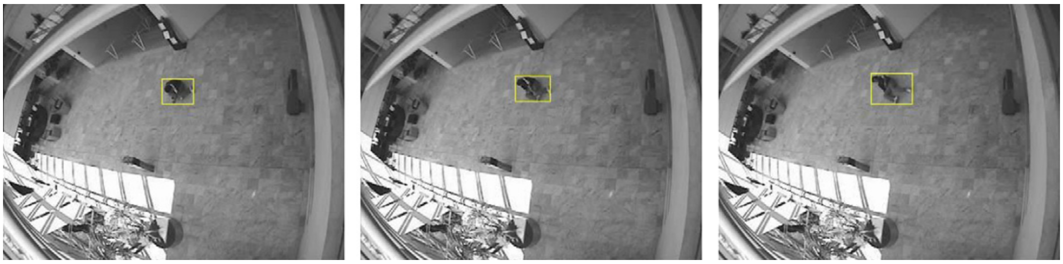


Fig. 13. Image frames taken from Rest_WiggleOnFloor with the response levels given by the proposed system using previously learned context. Left to right: frames 701, 705 and 709.

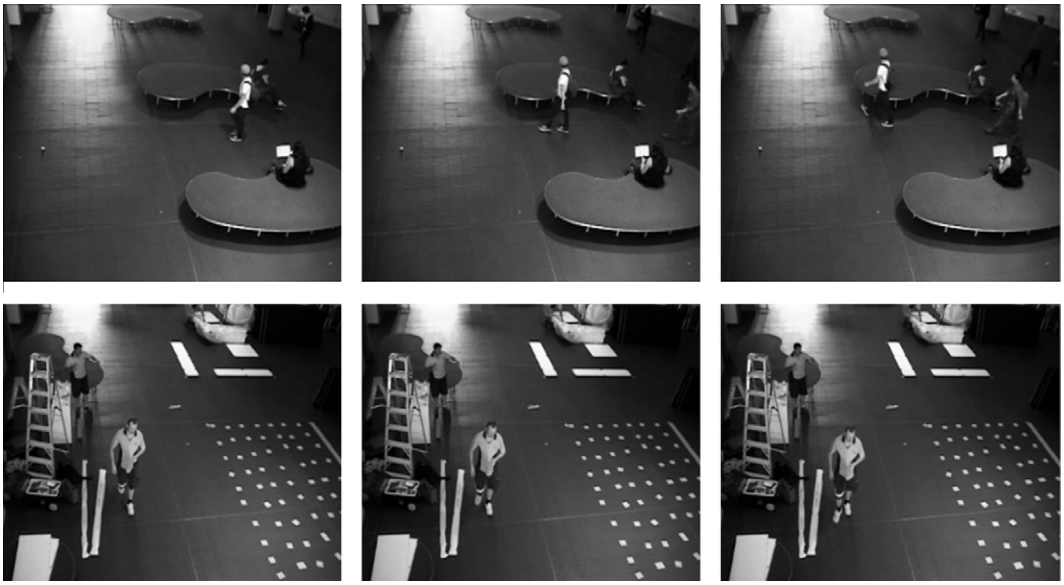


Fig. 14. Example of image feeds taken from Z-Block dataset. (a) context 1 (b) context 2.

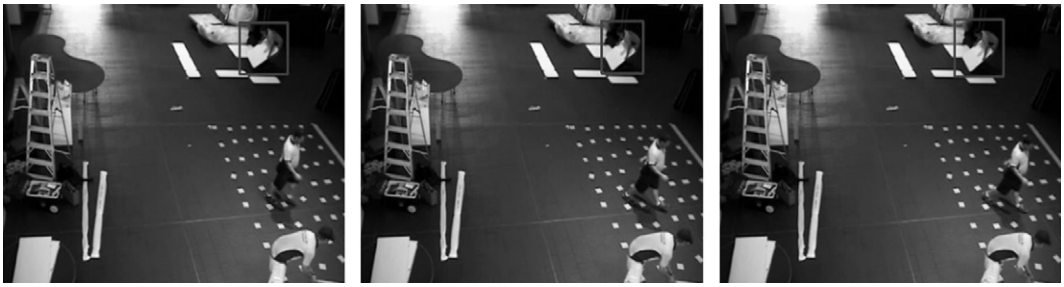


Fig. 15. Some feeds taken when the system raised response level IV, for the first time. The tracked person's behaviour is sig.uncommon in the previously learned context, and sig.common in the current context.

The construction work actually is not an unexpected event. Although it is not often, this event occurs occasionally. So, the response level IV given by the system actually tells the observers that the system does not expect that this event to occur. This is due to the inability of the context space model to differentiate between these two instances of context as different contexts. This means the system designers need to refine the context space model so that the system is able to distinguish them. For example, the arrangement of furniture in the scene could be used as one of the context parameters. Hence, the context space model can be refined as follows. *CI* = {*Time*, *Lecture schedule*, *Furnitures arrangement*, *commonality level*}, *CSB* = {*Time*, *Lecture schedule*} and *CSI* = {*commonality level*}.

The furnitures arrangement information can be captured by maintaining an image containing the background image (i.e. the image which does not contain any person in it). When the background image largely deviates from the previous one, then a new context is detected. By doing this, the system will see those two video clips as two different contexts.

The result is that the system detects a context change in the second video as soon as the difference between the current background and the previous background is over the predefined threshold. This means that the system will not consider the construction work happening in the second video as a major unexpected event because it does not use the information extracted from the first video. The system considers the first and second video as two different contexts due to the difference in the furniture arrangement parameter.

7.6. Discussions

The presented experiments have shown the role of context in suspicious behaviour detection systems. There are several important things in particular that were shown:

- The proposed context space model makes the proposed system possible to make more accurate decisions because the system uses the normal behaviour model relevant to the current context.
- Information extracted from the previously learned context combined with the context space model increases the proposed system performance. This is because this information provides a better knowledge of the current context and it also includes the information fed by the human observers (i.e. the labels given to some behaviour classes).
- Data stream clustering algorithm provides an efficient way to organise and retrieve context related information.
- Information extracted from the previously learned context combined with the context space model helps the human observers to detect major unexpected events which may or may not be harmful. The unexpected events may also be used to refine the context space model. This will be shown in the next section.

7.6.1. How is the previously learned context able to detect unexpected events and/or refine the existing context space model?

It was presented in the third part of the experiments that the information extracted from the previously learned context could be used to detect unexpected events and/or refine the existing context space model. To understand this, we need to revisit the definition of context presented in Definition 5. According to this definition, a context is an argument of a function Θ which maps the values of context space bases (CSBs) to context sensitive information (CSI). As we know that one of the properties of a function is that it assigns a unique value to its inputs. In other words, given a set of function arguments, a function will map this onto a set of output values. This set of output values will always be given by the function if the same set of arguments are used. This property implies that the commonality level of a behaviour will always be the same in instances of the same context. Let us suppose that the commonality level of a behaviour b in context1 is significantly uncommon. Mathematically, it can be expressed as follows. $\Theta(context1) = b\ is\ sig.uncommon$. Let us assume that the system is dealing with context2. Both context1 and context2 have the same CSB values, hence context1 and context2 are considered as the same context. Then, the system uses context1 as the previously learned context. The below expression follows from above: Given context1=context2, $\Theta(context1) = b$ is sig.uncommon $\leftrightarrow \Theta(context2) = b$ is sig.uncommon. Based on the context1, b is expected to be significantly uncommon. However, when b is significantly common in context2, then the rule in the above expression does not hold anymore. In this case, there may be a major unexpected event occurring or the CSB simply needs to be refined so that context1 $\neq$ context2.

Having this ability is very important as the system is able to give feedback to the system designers to refine the context space model. Also, the ability enables the observers to detect major unexpected events which may be harmful.

8. Conclusions

One of the key aspects in evaluating the success of surveillance systems depends on their performance in detecting suspicious human behaviour which could lead to a security breach. Unfortunately, the current surveillane systems rely heavily on human observers. This limits the capability of these systems to become forefront crime fighting tools. The current systems employ various techniques starting from rule-based methods to statistical approaches. However, the use of contextual information in these systems is still limited.

This paper proposed a context based system for detecting suspicious behaviour. There are three main components in the system. The first component is the context space model which provides the features used for describing different contexts. The second component is a data stream clustering algorithm which not only updates the system's knowledge over time, but also organises the knowledge into groups of which each group contains contextually related information. This enables the system to retrieve knowledge acquired in a specific context. Finally, the proposed inference algorithm allows the system to make a detection by combining the knowledge maintained by the proposed data stream clustering algorithm and provides an effective interface for human observers to feed their knowledge.

A comparative experiment between the proposed system and a similar system proposed in the literature was conducted in order to show the effectiveness of the proposed model for detecting suspicious human behaviour. From this experiment it was shown that the proposed system was able to make more accurate detections because it uses only knowledge relevant to the context. The second experiment was conducted in order to further show that the proposed system performance is improved when it has previously acquired knowledge of the current context. In the final experiment it was shown that the system was able to detect unexpected events. From these experiments it is suggested that contextual information plays an important part in detecting suspicious behaviour and should be used more despite its limited utilisation in this domain of research.

References

[1] C. Varun, B. Arindam, K. Vipin, Anomaly detection: a survey, ACM Comput. Surv. 41 (3) (2009) 1–58.

[2] H. Wells, T. Allard, P. Wilson, Crime and cctv in australia: understanding the relationship, Tech. rep., Center for Applied Psychology and Criminology, Bond University, 2006.

[3] P. Bull, Body Movement and Interpersonal Communication, John Wiley and Sons Ltd., 1983.

[4] H. Zhong, J. Shi, M. Visontai, Detecting unusual activity in video, in: Computer Vision and Pattern Recognition. CVPR. IEEE Conference on, vol. 2, 2004, pp. 819–826.

[5] G. Foresti, B. Pani, Monitoring motorway infrastructures for detection of dangerous events, in: International Conference on Image Analysis and Processing. Proceedings, 1999, pp. 1144–1147.

[6] C.C. Fung, N. Jerrat, A neural network based intelligent intruders detection and tracking system using cctv images, in: TENCON. Proceedings, vol. 2, 2000, pp. 409–414.

[7] Y.A. Ivanov, A.F. Bobick, Recognition of visual activities and interactions by stochastic parsing, Trans. Pattern Anal. Mach. Intell. 22 (8) (2000) 852–872.

[8] W. Niu, J. Long, D. Han, Y.-F. Wang, Human activity detection and recognition for video surveillance, in: IEEE International Conference on Multimedia and Expo. ICME, vol. 1, 2004, pp. 719–722.

[9] N. Robertson, I. Reid, A general method for human activity recognition in video, Comput. Vis. Image Understand. 104 (2-3) (2006) 232–248.

[10] Y. Miyanokoshi, E. Sato, T. Yamaguchi, Suspicious behavior detection based on case-based reasoning using face direction, in: SICE-ICASE. International Joint Conference, 2006, pp. 5429–5432.

[11] M. Yamamoto, H. Mitomi, F. Fujiwara, T. Sato, Bayesian classification of task-oriented actions based on stochastic context-free grammar, in: 7th International Conference on Automatic Face and Gesture Recognition, FGR, 2006, pp. 317–322.

[12] C. Stauffer, W.E.L. Grimson, Learning patterns of activity using real-time tracking, Trans. Pattern Anal. Mach. Intell. 22 (8) (2000) 747–757.

[13] N. Vaswani, A. Roy Chowdhury, R. Chellappa, Activity recognition using the dynamics of the configuration of interacting objects, in: Proceedings. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, 2003, pp. 633–640.

[14] O. Boiman, M. Irani, Detecting irregularities in images and in video, in: Tenth IEEE International Conference on Computer Vision. ICCV, vol. 1, 2005, pp. 462–469.

[15] D. Makris, T. Ellis, Learning semantic scene models from observing activity in visual surveillance, IEEE Trans. Syst., Man, Cybern., Part B: Cybern. 35 (3) (2005) 397–408.

[16] C. Piciarelli, G.L. Foresti, On-line trajectory clustering for anomalous events detection, Pattern Recogn. Lett. 27 (15) (2006) 1835–1842.

[17] D. Duque, H. Santos, P. Cortez, Prediction of abnormal behaviors for intelligent video surveillance systems, in: IEEE Symposium on Computational Intelligence and Data Mining. CIDM, 2007, pp. 362–367.

[18] Y. Zhou, S. Yan, T.S. Huang, Detecting anomaly in videos from trajectory similarity analysis, in: Y. Shuicheng (Ed.), IEEE International Conference on Multimedia and Expo, 2007, pp. 1087–1090.

[19] T. Xiang, S. Gong, Incremental and adaptive abnormal behaviour detection, Comput. Vis. Image Understand. 111 (1) (2008) 59–73.

[20] L. Kratz, K. Nishino, Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Fontainebleau Resort, Miami Beach, Florida, 2009.

[21] O. Sang-Hyun, K. Jin-Suk, B. Yung-Cheol, P. Gyung-Leen, B. Sang-Yong, Intrusion detection based on clustering a data stream, in: Third ACIS International Conference on Software Engineering Research, Management and Applications, 2005, pp. 220–227.

[22] M.S. Ryoo, J.K. Aggarwal, Stochastic representation and recognition of high-level group activities, Int. J. Comput. Vis. 93 (2011) 183–200.

[23] V. Reddy, C. Sanderson, B.C. Lovell, Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture, in: Computer Vision and Pattern Recognition Workshops (CVPRW), 2011, pp. 55–61.

[24] B. Yao, L. Fei-Fei, Modeling mutual context of object and human pose in human–object interaction activities, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, USA, 2010.

[25] T. Lan, Y. Wang, W. Yang, G. Mori, Beyond actions: Discriminative models for contextual group activities, in: Advances in Neural Information Processing Systems (NIPS), 2010.

[26] V.I. Morariu, V.S.N. Prasad, L.S. Davis, Human activity understanding using visibility context, in: IEEE/RSJ IROS Workshop: From Sensors to Human Spatial Concepts (FS2HSC), 2007.

[27] M. Marszalek, I. Laptev, C. Schmid, Actions in context, in: IEEE Conference on Computer Vision and Pattern Recognition. CVPR, 2009, pp. 2929–2936.

[28] A. Torralba, Contextual priming for object detection, Int. J. Comput. Vis. 53 (2) (2003) 169–191.

[29] I. Katz, H. Aghajan, Exploring relationship between context and pose: a case study, in: Cognitive Systems and Interactive Sensors, 2007.

[30] T.M. Strat, Employing contextual information in computer vision, in: DARPA93, 1993, pp. 217–229.

[31] M. Pantic, A. Pentland, A. Nijholt, T. Huang, Human computing and machine understanding of human behavior: a survey, in: Artifical Intelligence for Human Computing, 2007, pp. 47–71.

[32] C.C. Aggarwal, J. Han, J. Wang, P.S. Yu, A framework for clustering evolving data streams, in: the 29th VLDB Conference, Berlin, Germany, 2003, pp. 81–92.

[33] P. Dollar, V. Rabaud, G. Cottrell, S. Belongie, Behavior recognition via sparse spatio-temporal features, in: IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005, pp. 65–72.