

METHOD AND SYSTEM FOR EVENT DETECTION

FIELD OF INVENTION

5 The invention broadly relates to a method and system for event detection.

BACKGROUND

10 The use of closed-circuit television (CCTV) cameras in lifts (elevators) for the prevention of vandalism, crime and other undesirable acts has gained growing popularity in many countries as they can provide a video recording facility for post-incident investigation. This has resulted in more lifts being retrofitted with such cameras.

15 Currently, in addition to CCTV cameras, passenger lifts have also been retrofitted with other types of sensors (e.g. chemical sensors for urine detection) in order to capture vandalism, crime and other undesirable acts. However, some shortcomings include:

- i. Unsatisfactory performance of the chemical sensors.
- ii. The ability of culprits to get around the system once they get acquainted with the sensors.
- 20 iii. More complex events cannot be automatically detected (e.g.: robbery and violence against the elderly, children, and women). In addition, the recorded video may not be able to clearly identify the perpetrators.
- iv. Inability to prevent acts that lead to the damaging of lifts. Acts that cause damage to lifts typically go undetected. Subsequent investigation by reviewing the CCTV video is usually tedious and inconclusive.
- 25 v. In crimes against soft targets (e.g. elderly, children, and women), there are generally no means of taking immediate action. Currently, only post-incident investigation can be done upon the filing of police's report. Thereafter a lengthy information and evidence gathering process may
- 30 follow, coupled with a tedious process to apprehend the culprit.

 Advancements in computational science, coupled with cheaper and more powerful computing platforms, enables information from CCTV cameras to be used to

detect anti-social behaviour such as urinating in the lift. Algorithms can be designed to detect this event in real-time. Currently, a conductive loop is deployed for such detection; however, performance is far from acceptable while maintenance and installation are relatively cumbersome.

5

More recently, there are works involving the combination of video and audio for the purposes of background-foreground separation, localization of humans, and event/behaviour detection. However, these works have separate audio and video processing threads that function independently. The respective classifier results are combined to reach a final decision. In such methods using the aggregation of classifier results, corrupted signals from the other sensors can sway the output and lead to an erroneous result. Useful information can be lost as some types of sensors are more strongly related to the characteristics of certain events than other types of sensors.

10

15

A need therefore exists to provide a method and system for event detection that seeks to address at least one of the abovementioned problems.

SUMMARY

20

According to the first aspect of the present invention, there is provided a system for event detection, comprising: a plurality of sensors; and respective classifier units coupled to the sensors for processing respective sensor signals from the sensors for the event detection; wherein at least one of the classifier units is adapted for detecting an event based on the sensor signal processed in said at least one classifier unit as a main evidence and for issuing sub-evidence queries to the other classifier units for facilitating the event detection.

25

30

The at least one classifier unit may comprise a categorisation unit for categorising the sensor signal into event categories.

The categorisation unit may be adapted to use one or more parameters from the sub-evidence queries to the other classifier units in categorising the sensor signal into the event categories.

- 5 The at least one classifier unit may comprise a dominant signature verification unit for verifying a dominant signature in the sensor signal for the event detection.

10 The dominant signature verification unit may be adapted to use one or more parameters from the sub-evidence queries to the other classifier units in verifying the dominant signature in the sensor signal.

The at least one classifier unit may comprise a machine learning unit for classification processing for the event detection.

- 15 The machine learning unit may be adapted to use one or more parameters from the sub-evidence queries to the other classifier units in the classification processing.

20 According to a second aspect of the present invention, there is provided a method of event detection comprising the steps of: obtaining respective sensor signals from a plurality of sensors; and processing the respective sensor signals for the event detection using respective classifier units coupled to the plurality of sensors, wherein an event is detected based on the sensor signal processed in one classifier unit as a main evidence; and issuing sub-evidence queries to the other classifier units for facilitating the event detection using said one classifier unit.

25 The method may comprise categorising the sensor signal into event categories using a categorisation unit of said one classifier unit.

30 The method may comprise using one or more parameters from the sub-evidence queries to the other classifier units in categorising the sensor signal into the event categories.

The method may comprise verifying a dominant signature in the sensor signal for the event detection using a dominant signature verification unit of said one classifier unit comprises.

5 The method may comprise using one or more parameters from the sub-evidence queries to the other classifier units in verifying the dominant signature in the sensor signal.

10 The method may comprise performing classification processing for the event detection using a machine learning unit of said one classifier unit.

The method may comprise using one or more parameters from the sub-evidence queries to the other classifier units in the classification processing.

15 According to a third aspect of the present invention, there is provided a data storage medium having stored thereon computer program code means for instructing a computer system to execute a method of event detection, and as defined in the second aspect.

20 BRIEF DESCRIPTION OF THE DRAWINGS

Embodiments of the invention will be better understood and readily apparent to one of ordinary skill in the art from the following written description, by way of example only, and in conjunction with the drawings, in which:

Figure 1 is a schematic drawing illustrating the configuration of an audio and video awareness system for lift monitoring, according to an example embodiment of the present invention.

30 Figure 2A is a flow chart illustrating the system architecture and processing steps of an audio and video awareness system for lift monitoring, according to an example embodiment of the present invention.

Figure 2B is a flow chart illustrating the system architecture and processing steps of an audio and video awareness system for lift monitoring, according to another example embodiment of the present invention.

5 Figure 2C is a flow chart illustrating the system architecture and processing steps of an audio and video awareness system for lift monitoring, according to a further example embodiment of the present invention.

10 Figure 3 is a flow chart illustrating a method for event detection, according to an example embodiment of the present invention.

Figure 4 is a schematic of a computer system for implementing the method and system for event detection.

15

DETAILED DESCRIPTION

20 According to embodiments of the present invention, there is provided an audio and video awareness system for lift monitoring that comprises an artificial intelligence (AI) system that fuses the computations of signals from a plurality of sensor types. In the following description, the system is described in relation to a lift (elevator). However, it will be appreciated by a person skilled in the art that the audio and video awareness system described herein may be used in any suitable environment or location.

25 Some portions of the description which follows are explicitly or implicitly presented in terms of algorithms and functional or symbolic representations of operations on data within a computer memory. These algorithmic descriptions and functional or symbolic representations are the means used by those skilled in the data processing arts to convey most effectively the substance of their work to others skilled in the art. An algorithm is here, and generally, conceived to be a self-consistent sequence
30 of steps leading to a desired result. The steps are those requiring physical manipulations of physical quantities, such as electrical, magnetic or optical signals capable of being stored, transferred, combined, compared, and otherwise manipulated.

Unless specifically stated otherwise, and as apparent from the following, it will be appreciated that throughout the present specification, discussions utilizing terms such as "scanning", "calculating", "determining", "replacing", "generating", "initializing", "outputting", or the like, refer to the action and processes of a computer system, or similar electronic device, that manipulates and transforms data represented as physical quantities within the computer system into other data similarly represented as physical quantities within the computer system or other information storage, transmission or display devices.

The present specification also discloses apparatus for performing the operations of the methods. Such apparatus may be specially constructed for the required purposes, or may comprise a general purpose computer or other device selectively activated or reconfigured by a computer program stored in the computer. The algorithms and displays presented herein are not inherently related to any particular computer or other apparatus. Various general purpose machines may be used with programs in accordance with the teachings herein. Alternatively, the construction of more specialized apparatus to perform the required method steps may be appropriate. The structure of a conventional general purpose computer will appear from the description below.

In addition, the present specification also implicitly discloses a computer program, in that it would be apparent to the person skilled in the art that the individual steps of the method described herein may be put into effect by computer code. The computer program is not intended to be limited to any particular programming language and implementation thereof. It will be appreciated that a variety of programming languages and coding thereof may be used to implement the teachings of the disclosure contained herein. Moreover, the computer program is not intended to be limited to any particular control flow. There are many other variants of the computer program, which can use different control flows without departing from the spirit or scope of the invention.

Furthermore, one or more of the steps of the computer program may be performed in parallel rather than sequentially. Such a computer program may be stored on any computer readable medium. The computer readable medium may include

storage devices such as magnetic or optical disks, memory chips, or other storage devices suitable for interfacing with a general purpose computer. The computer readable medium may also include a hard-wired medium such as exemplified in the Internet system, or wireless medium such as exemplified in the GSM mobile telephone system.

5 The computer program when loaded and executed on such a general-purpose computer effectively results in an apparatus that implements the steps of the preferred method.

The invention may also be implemented as hardware modules. More particular, in the hardware sense, a module is a functional hardware unit designed for use with
10 other components or modules. For example, a module may be implemented using discrete electronic components, or it can form a portion of an entire electronic circuit such as an Application Specific Integrated Circuit (ASIC). Numerous other possibilities exist. Those skilled in the art will appreciate that the system can also be implemented as a combination of hardware and software modules.

15

Figure 1 is a schematic drawing illustrating the configuration of an audio and video awareness system for lift monitoring, designated generally as reference numeral 100, according to an example embodiment of the present invention.

20 The system 100 comprises 2 lifts 102/112, an analytical engine 122 (implemented on a computer), a communication module 124 and a lift control module 126. Each of the 2 lifts 102/112 comprise 3 types of sensors: a closed-circuit television (CCTV) camera 108/118, two contact microphones 106a/b / 116a/b embedded on the lower parts of the lift walls, an audio microphone 104/114 embedded on the lift ceiling,
25 and a laminated lift floor 110/120 with a relatively high acoustic signal-to-noise ratio. The audio microphones 104/114 provide acoustic audio signals, the contact microphones 106a/b / 116a/b provide vibration signals and the (CCTV) cameras 108/118 provide video signals. The lift control module 126 is connected and in communication with the analytical engine 122 and/or communication module 124 to facilitate control of the lift's
30 operation. It will be appreciated by a person skilled in the art that although the system 100 comprises 2 lifts 102/112, any suitable number of lifts can be monitored. In other words, sensors from multiple lifts can be fed to the analytical engine 122.

The CCTV cameras 108/118 are preferably analogue CCTV cameras of PAL specification. The analogue signals from the CCTV cameras 108/118, contact microphones 106a/b, 116a/b and audio microphones 104/114 are sent to the analytical engine 122 located, for example, in a lift motor room a distance away, for example, through co-axial cables 128a – d. Appropriate amplifiers (not shown) can be applied along the way to maintain signal quality. The video signals are digitized by hardware encoders (which may be located within the analytical engine 122) into digital streams such as H.264. The audio and contact signals are digitized by audio processors (which may also be located within the analytical engine 122) for digitization into audio streams e.g. in the WAV format. These digital video and audio streams are synchronized using the analytical engine's 122 time stamp at the time of their digitization.

The analogue signals from the 3 types of sensors are processed at the analytical engine 122, wherein information from one of the types of sensors used in a master classifier and information from the other sensors is called by the master classifier from slave classifiers associated with the remaining types of sensors. For example, the system may be primarily driven by audio analytics in that the microphone signal is the main determining factor in deducing the occurrence of a particular event. Upon a suspicious detection from an audio source, computational results from video feeds can be used to verify the validity of the detection from audio signal. This entails an intermediate process involving technical queries to other sensors to build-up the evidence for deducing the occurrence of the particular event. It will be appreciated that more features can be explored in proportion to the capability of the computational platform available.

The system 100 can also comprise an alarm system which is raised when an event occurs and users can simultaneously apply two types of alarm mechanisms:

Distributed mechanism: When an event occurs, the analytical engine 122 directly sends out alarm notifications via the communication module 124 to a designated user's mobile phone through a 3G connection. The user can perform a verification check by playing back the CCTV video feed prior to taking an appropriate action. The system 100 advantageously allows building owners to manage their lifts more comprehensively.

Monitoring and control can be executed remotely, for example, through a 3G based communication system. In contrast, in the prior art, users have to be at designated monitoring stations since connectivity through GPRS can only notify the users through short messages.

Centralized mechanism: When an event occurs, a notification and a respective video segment are sent via the communication module 124 to a web server (not shown) at a central location. This web server can then use artificial intelligence to decide which user(s) to notify based on the rules-of-engagement it is programmed with. The recipient(s) of these alarms can then perform their verifications by connecting to the web server.

Figure 2A is a flow chart illustrating the system architecture/processing steps of an audio and video awareness system for lift monitoring, generally designated as reference numeral 200a, according to an example embodiment of the present invention. The system comprises a plurality of classifiers: an audio classifier 252, a vibration classifier 254 and a video classifier 256. The classifiers 252, 254 and 256 can perform either as a master and/or a slave classifier for different event detection tasks. In example embodiments, depending on the type of event, information from a first type of sensor is used in a master classifier and information from the remaining types of sensors is called by the master classifier from slave classifiers associated with the remaining types of sensors. The master classifier is used to 'suspect' the occurrence of an event using the main evidence from one sensor, while the other classifiers are used to confirm the occurrence of the event using sub-evidences from other sensors (wherein the sub-evidences are not directly linked to the event). Advantageously, the detection of each event is processed independently. However, at any one time, the classifiers 252, 254 and 256 are running concurrently to detect events. In addition, the independent processing steps in each of the classifiers 252, 254 and 256 are preferably synchronized using a common time.

In example embodiments, the master classifier queries the slave classifier on the presence of low-level, mid-level and high-level information.

Examples of low-level queries include:

- Changes in scene or the presence of foreground objects (from video signal)
- The basic shape of the foreground objects (video)
- The pitch-level of the signal (audio/contact)
- The bandwidth of the signal in the frequency domain

Examples of mid-level queries include:

- Presence of human shape/model and number of people (video)
- Presence of human screams/shouts (audio)
- Presence of banging impacts (contact)
- Presence of rattling sound (contact)

Examples of high-level queries include:

- Presence of aggressive motion (video)
- Human posture estimation (video)
- Presence of violent actions (video)

One example of the use of the audio classifier 252, functioning as a master classifier, can be for the detection of crimes against soft targets (e.g. the elderly, children and women) 212. At step 202, an audio sound is detected. At step 204, the audio signal goes through a number of sampling and pre-processing tasks that include the removal of noise and signal re-sampling.

Categorization 206 is performed by the audio classifier 252 wherein the audio signal is categorized into 'silent/non-silent', 'voice/non-voice', and 'normal/abnormal'. Different processing threads are taken from the categorization results. For instance, 'non-voice' audio signals are further investigated into common metadata of normal situations such as footsteps, movement sound, etc. Deviations from the normal situations initiate an identification of parameters that are precursors to an event. Queries (low-level), e.g. indicated at numeral 260 to the video classifier 256 (slave classifier) for foreground object detection 229 and human model detection 230 are issued to assist in the accuracy of this categorization. If the audio signal is "silent" or "normal", or no object is detected, further classification may not be performed. Techniques for foreground

object detection and human model detection are understood in the art and will not be described in further detail. Reference is made, for example, to (i) "Adaptive Background Subtraction with Multiple Feedbacks for Video Surveillance," Intl Symp Visual Computing, (Lake Tahoe, Nevada, USA., Dec. 5-7, 2005), pp. 380-387; and (ii) Liyuan Li, Weimin Huang, Gu, I.Y.H, Leman, K, Qi Tian, "Principal color representation for tracking persons," IEEE International Conference on Systems, Man and Cybernetics, 5-8 Oct. 2003, vol 1, pp. 1007 - 1012), the contents of which are herein incorporated by reference.

Verification of dominant signature 208 is also carried out at the audio classifier 252. The audio signal is analyzed for the presence of sounds characteristic of crimes against soft targets such as shouts and cries (e.g. short duration, non-repetitive high pitch sounds with narrow spectral). Dominant Signatures in the example embodiment are knowledge-based parameters that define the syntax of a particular event. The syntax describes information (e.g. low/mid-level information) of a particular signal.

Machine Learning Classification 210 using weighted features from audio and vibration signals is also performed at the audio classifier 252. The sound is compared to sample sounds characteristic of crimes against soft targets to detect the presence of crimes against soft targets. Machine Learning classification advantageously facilitates a more comprehensive detection of signals compared to the verification of dominant signatures alone. The machine learning classifier is preferably a Support Vector Machine (SVM) classifier with a super feature vector concatenated from MFCC, Prosody, Perceptual Linear Prediction Coding Cepstra (PLPCC) and MPEG-7 descriptor.

A Support Vector Machine (SVM) classifier is understood in the art and will not be described in further detail. Reference is made, for example, to Vladimir Vapnik, *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995. ISBN 0-387-98780-0, the contents of which are herein incorporated by reference.

MFCC is understood in the art and will not be described in further detail. Reference is made, for example, to Fang Zheng, Guoliang Zhang and Zhanjiang Song (2001), "Comparison of Different Implementations of MFCC," *J. Computer Science &*

Technology, 16(6): 582–589), the contents of which are herein incorporated by reference.

Prosody is understood in the art and will not be described in further detail.

- 5 Reference is made, for example, to Hermansky, H. (1990) "*Perceptual linear predictive (PLP) analysis of speech*", J. Acoust. Soc. Am., 87(4), pp.1738-1752, the contents of which are herein incorporated by reference.

- 10 Perceptual Linear Prediction Coding Cepstra (PLPCC) is understood in the art and will not be described in further detail. Reference is made, for example, to Hirschberg, J., Liscombe, J., and Venditti, J. Experiments in Emotional Speech. IEEE Workshop on Spontaneous Speech Recognition. 2003, the contents of which are herein incorporated by reference.

- 15 A MPEG-7 descriptor is understood in the art and will not be described in further detail. Reference is made, for example, to B.S. Manjunath (Editor), Philippe Salembier (Editor), and Thomas Sikora (Editor): Introduction to MPEG-7: Multimedia Content Description Interface. Wiley & Sons, April 2002 - ISBN 0-471-48678-7, the contents of which are herein incorporated by reference.

20

- Machine Learning features are also known as data driven features, in which the system learns the patterns of events by being fed with data samples of the events. In other words, sample sounds of each event are collected. The sample sounds preferably cover a sufficiently wide variation of the conditions in the lift for a particular event. With
25 sufficient good data samples, an engine (also called a classifier) is produced to recognize the events. The engine can be created by a process where signals are fed to the engine and information about each signal is provided. This process is called the training of a classifier. Potential spurious signals that resemble a particular event signal can also be trained into the engine as negative samples. An engine that is created with
30 sufficient trainings of sample data can classify a signal into one of the event categories it is trained. In the example of screaming detection, data from sample events involving screaming are collected, pre-processed, and sent for training using machine learning

tools such as the Neural-Networks, Bayesian Networks, or Support Vector Machine. Mathematical representations can then be used to discriminate different events.

Querying for mid/high level information e.g. indicated at numeral 260 to the video classifier 256 (slave classifier) can be issued to assist in the accuracy of this classification. For example, mid and high level information includes the number of people in the lift and the presence of aggressive behaviour respectively. In the event that there is only one person in the lift, the system may treat it as a false alarm and may only provide a warning to the passenger in the lift. Conversely, if more than one person is in the lift, a high level query to an Aggressive Motion Measurement unit 233 is preferably made. A Neural Network classifier may be used to detect aggressive behaviours based on the shape and motion features extracted from the video. The features can include motion energy image (MEI), motion history image (MHI), and motion rapidity image (MRI). MRI is understood in the art and will not be described in further detail. Reference is made, for example, to Liyuan Li and Maylor K.H. Leung (2001), "Suspicious Human Action Detection for Video Surveillance," in Proc. of 3rd Intl Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS), Germany, the contents of which are herein incorporated by reference. If the system detects the presence of more than one person and aggressive behaviour, it may imply that a crime against a soft target is occurring.

A crime against a soft target can then be detected in the example embodiment through the evidence obtained via the processes of categorization, verification of dominant signature and machine learning classification, assisted / confirmed using sub-evidence obtained from queries from the slave classifier(s).

One example of the use of the vibration classifier 254, functioning as a master classifier, can be for the detection of urination in a lift 242 and acts of vandalism 244. At step 214, a vibration sound is detected. At step 216, the signal goes through a number of sampling and pre-processing tasks that include the removal of noise and signal re-sampling.

Categorization 206 is performed by the vibration classifier 254 wherein the vibration signal is categorized into 'silent/non-silent', and 'normal/abnormal'. Deviations from the normal situations initiate an identification of parameters that are precursors to an event. Queries (low-level), e.g. indicated at numeral 264 to the video classifier 256 (slave classifier) for foreground object detection 229 and human model detection 230, are issued to assist in the accuracy of this categorization. If the audio signal is "silent" or "normal", or no object is detected, further classification may not be performed.

The master classifier 254 queries 264 the slave classifier on the presence of low-level, mid-level and high-level information in this embodiment and examples of such queries have been described above.

Verification of dominant signature 208 is also carried out at the vibration classifier 254. The vibration signal is analyzed for the presence of the characteristic sounds of urination (e.g. low beat and periodic) and/or vandalism (e.g. high pitch scratching sounds).

Machine Learning classification 210 using weighted features from audio and vibration signals is also performed at the vibration classifier 254. The sound is compared to sample urination sounds and/or sample sounds caused by acts of vandalism. As mentioned above, Machine Learning classification advantageously facilitates a more comprehensive detection of signals compared to the verification of dominant signatures.

Querying for mid/high level information e.g. indicated at numeral 264 to the video classifier 256 (slave classifier) can be issued to assist in the accuracy of this classification. For example, with regard to urinating in the lift, if there is more than one person in the lift, the system may treat it as a false alarm and may only provide a warning to the passenger in the lift. Conversely, if only one person is in the lift and is not facing the door, the system may treat it as a true alarm (i.e.: someone is urinating in the lift). A further query can be made to a Water Patch Detection unit 232 to detect the presence of urine or a liquid patch (e.g. vandalism caused by stewing liquids in the lift).

Urinating and vandalism in the lift can then be detected in the example embodiment through the evidence obtained via the processes of categorization, verification of dominant signature and machine learning classification, assisted / confirmed using sub-evidence obtained from queries from the slave classifier(s).

For events such as trash dumping (abandoning objects in the lift) 238 and fire/smoke in the lift 240, an example detection algorithm involves obtaining a video signal from the CCTV camera at step 224. The signal is processed at step 226 to remove noise.

The video classifier 256 performs object segmentation 228. As trash dumping 238 and fire/smoke in the lift 240 do not produce audio or contact signatures, the video classifier 256 is designated as the master classifier. No slave classifier is designated. Event analysis at step 239 can be carried out directly by the video classifier to detect the occurrence of trash dumping 238 or fire/smoke in the lift 240.

In the example embodiment described above, implementation involved both verification of dominant signature and machine learning classification. However, it will be appreciated that alternative example embodiments may be implemented with either verification of dominant signature or machine learning classification only, as will be described below.

Figure 2B is a flow chart illustrating the system architecture/processing steps of an audio and video awareness system for lift monitoring, generally designated as reference numeral 200b, according to another example embodiment of the present invention. Similar to the system 200a, the present system 200b comprises a plurality of classifiers: an audio classifier 252, a vibration classifier 254 and a video classifier 256.

One example of the use of the audio classifier 252, functioning as a master classifier, can be for the detection of crimes against soft targets (e.g. the elderly, children and women) 212. At step 202, an audio sound is detected. At step 204, the audio signal goes through a number of sampling and pre-processing tasks that include the removal of noise and signal re-sampling.

Categorization 206 is performed by the audio classifier 252 wherein the audio signal is categorized into 'silent/non-silent', 'voice/non-voice', and 'normal/abnormal'. Queries (low-level), e.g. indicated at numeral 260 to the video classifier 256 (slave classifier) for foreground object detection 229, and human model detection 230 are issued to assist in the accuracy of this categorization. If the audio signal is "silent" or "normal", or no object is detected, further classification may not be performed.

The master classifier 252 queries 260 the slave classifier on the presence of low-level, mid-level and high-level information in this embodiment and examples of such queries have been described above in relation to the previous embodiment.

Verification of dominant signature 208 is also carried out at the audio classifier 252. The audio signal is analyzed for the presence of sounds characteristic of crimes against soft targets such as shouts and cries (e.g. short duration, non-repetitive high pitch sounds with narrow spectral).

Querying for mid/high level information e.g. indicated at numeral 260 to the video classifier 256 (slave classifier) can be issued to assist in the accuracy of verification. For example, mid and high level information includes the number of people in the lift and the presence of aggressive behaviour respectively. In the event that there is only one person in the lift, the system may treat it as a false alarm and may only provide a warning to the passenger in the lift. Conversely, if more than one person is in the lift, a high level query to an Aggressive Motion Measurement unit is preferably made 233. A Neural Network classifier may be used to detect aggressive behaviours based on the shape and motion features extracted from the video. The features are motion energy image (MEI), motion history image (MHI), and motion rapidity image (MRI). If the system detects the presence of more than one person and aggressive behaviour, it may imply that a crime against a soft target is occurring.

A crime against a soft target can then be detected in the example embodiment through the evidence obtained via the processes of categorization and verification of

dominant signature, assisted / confirmed using sub-evidence obtained from queries from the slave classifier(s).

One example of the use of the vibration classifier 254, functioning as a master classifier, can be for the detection of urination in a lift 242 and acts of vandalism 244. At step 214, a vibration sound is detected. At step 216, the signal goes through a number of sampling and pre-processing tasks that include the removal of noise and signal re-sampling.

Categorization 206 is performed by the vibration classifier 254 wherein the vibration signal is categorized into 'silent/non-silent', and 'normal/abnormal'. Deviations from the normal situations initiate an identification of parameters that are precursors to an event. Queries (low-level), e.g. indicated at numeral 264 to the video classifier 256 (slave classifier) for foreground object detection 229 and human model detection 230, are issued to assist in the accuracy of this categorization. If the audio signal is "silent" or "normal"; or no object is detected, further classification may not be performed.

The master classifier 254 queries 264 the slave classifier on the presence of low-level, mid-level and high-level information in this embodiment and examples of such queries have been described above in relation to the previous embodiment.

Verification of dominant signature 208 is also carried out at the vibration classifier 254. The vibration signal is analyzed for the presence of the characteristic sounds of urination (e.g. low beat and periodic) and/or vandalism (e.g. high pitch scratching sounds).

Querying for mid/high level information e.g. indicated at numeral 264 to the video classifier 256 (slave classifier) can be issued to assist in the accuracy of classification. For example, with regard to urinating in the lift, if there is more than one person in the lift, the system may treat it as a false alarm and may only provide a warning to the passenger in the lift. Conversely, if only one person is in the lift, a high level query is preferably issued to a Human Posture Estimation unit 231 to determine if a passenger is facing the door. If the passenger is not facing the door and there is only one person in

the lift, the system may treat it as a true alarm (i.e.: someone is urinating in the lift). A further query can be made to a Water Patch Detection unit 232 to detect the presence of urine or a liquid patch (e.g. vandalism caused by strewing liquids in the lift).

5 Urinating and vandalism in the lift can then be detected in the example embodiment through the evidence obtained via the processes of categorization and verification of dominant signature, assisted / confirmed using sub-evidence obtained from queries from the slave classifier(s).

10 For events such as trash dumping (abandoning objects in the lift) 238 and fire/smoke in the lift 240, an example detection algorithm involves obtaining a video signal from the CCTV camera at step 224. The signal is processed at step 226 to remove noise.

15 The video classifier 256 performs object segmentation 228. As trash dumping 238 and fire/smoke in the lift 240 do not produce audio or contact signatures, the video classifier 256 is designated as the master classifier. No slave classifier is designated. Event analysis at step 239 can be carried out directly by the video classifier to detect the occurrence of trash dumping 238 or fire/smoke in the lift 240.

20 Figure 2C is a flow chart illustrating the system architecture and processing steps of an audio and video awareness system for lift monitoring, generally designated as reference numeral 200c, according to a further embodiment of the present invention. Similar to the systems 200a/b, the present system 200c comprises a plurality of classifiers: an audio classifier 252, a vibration classifier 254 and a video classifier 256.

30 One example of the use of the audio classifier 252, functioning as a master classifier, can be for the detection of crimes against soft targets (e.g. the elderly, children and women) 212. At step 202, an audio sound is detected. At step 204, the audio signal goes through a number of sampling and pre-processing tasks that include the removal of noise and signal re-sampling.

Categorization 206 is performed by the audio classifier 252 wherein the audio signal is categorized into 'silent/non-silent', 'voice/non-voice', and 'normal/abnormal'. Queries, e.g. indicated at numeral 260 to the video classifier 256 (slave classifier) for foreground object detection 229, and human model detection 230 are issued to assist in the accuracy of this categorization. If the audio signal is "silent" or "normal" ; or no object is detected, further classification may not be performed.

The master classifier 252 queries 260 the slave classifier on the presence of low-level, mid-level and high-level information in this embodiment and examples of such queries have been described above in relation to the previous two embodiments.

Machine Learning classification 210 using weighted features from audio and vibration signals is also performed at the audio classifier 252. Machine Learning Classification advantageously facilitates a more comprehensive detection of signals compared to the verification of dominant signatures. The sound is compared to sample sounds characteristic of crimes against soft targets to detect the presence of crimes against soft targets.

Querying for mid/high level information e.g. indicated at numeral 260 to the video classifier 256 (slave classifier) can be issued to assist in the accuracy of classification. For example, mid and high level information includes the number of people in the lift and the presence of aggressive behaviour respectively. In the event that there is only one person in the lift, the system may treat it as a false alarm and may only provide a warning to the passenger in the lift. Conversely, if more than one person is in the lift, a Neural Network classifier is used then to detect aggressive behaviours based on the shape and motion features extracted from the video. The features are motion energy image (MEI), motion history image (MHI), and motion rapidity image (MRI). If the system detects the presence of more than one person and aggressive behaviour, it may imply that a crime against a soft target is occurring.

A crime against a soft target can then be detected in the example embodiment through the evidence obtained via the processes of categorization and machine learning

classification, assisted / confirmed using sub-evidence obtained from queries from the slave classifier(s).

One example of the use of the vibration classifier 254, functioning as a master classifier, can be for the detection of urination in a lift 242 and acts of vandalism 244. At step 214, a vibration sound is detected. At step 216, the signal goes through a number of sampling and pre-processing tasks that include the removal of noise and signal re-sampling.

Categorization 206 is performed by the vibration classifier 254 wherein the vibration signal is categorized into 'silent/non-silent', and 'normal/abnormal'. Deviations from the normal situations initiate an identification of parameters that are precursors to an event. Queries (low-level), e.g. indicated at numeral 264 to the video classifier 256 (slave classifier) for foreground object detection 229 and human model detection 230, are issued to assist in the accuracy of this categorization. If the audio signal is "silent" or "normal", or no object is detected, further classification may not be performed.

The master classifier 254 queries 264 the slave classifier on the presence of low-level, mid-level and high-level information in this embodiment and examples of such queries have been described above in relation to the previous two embodiments.

Machine Learning classification 210 using weighted features from audio and vibration signals is also performed at the vibration classifier 254. The sound is compared to sample urination sounds and/or sample sounds caused by acts of vandalism.

Querying for mid/high level information e.g. indicated at numeral 264 to the video classifier 256 (slave classifier) can be issued to assist in the accuracy of classification. For example, with regard to urinating in the lift, if there is more than one person in the lift, the system may treat it as a false alarm and may only provide a warning to the passenger in the lift. Conversely, if only one person is in the lift, a high level query is preferably issued to a Human Posture Estimation unit 231 to determine if a passenger is facing the door. If the passenger is not facing the door and there is only one person in the lift, the system may treat it as a true alarm (i.e.: someone is urinating in the lift). A

further query can be made to a Water Patch Detection unit 232 to detect the presence of urine or a liquid patch (e.g. vandalism caused by strewing liquids in the lift).

Uninating and vandalism in the lift can then be detected in the example embodiment through the evidence obtained via the processes of categorization and machine learning classification, assisted / confirmed using sub-evidence obtained from queries from the slave classifier(s).

For events such as trash dumping (abandoning objects in the lift) 238 and fire/smoke in the lift 240, an example detection algorithm involves obtaining a video signal from the CCTV camera at step 224. The signal is processed at step 226 to remove noise.

The video classifier 256 performs object segmentation 228. As trash dumping 238 and fire/smoke in the lift 240 do not produce audio or contact signatures, the video classifier 256 is designated as the master classifier. No slave classifier is designated. Event analysis at step 239 can be carried out directly by the video classifier to detect the occurrence of trash dumping 238 or fire/smoke in the lift 240.

Designating master and slave classifiers in example embodiments of the present invention advantageously results in a more robust outcome since it is more resilient to the effect of noises on the sensors.

The algorithms for the processing of video signals in the example embodiments comprise:

- Dynamic background scene modelling - This is preferably the most fundamental algorithm for the processing of video signals and it functions to extract changes in the background scene of the lift and segments out humans or objects in the lift. The change is called foreground and, for example, can be due to objects/human(s) that come into the scene of a lift's interior. In order for the system to run continuously and accurately without human intervention, the algorithm adapts to changes in the scene. For instance, when the lift door opens

and additional light enters the lift, this does not induce errors in the extraction of changes in the background scene.

- Human model detection - In a confined space of a lift, movement trajectory may not be used to deduce whether a detected object is a human. In embodiments of the present invention, a silhouette of a foreground blob can be analyzed to check the presence of an inverted omega pattern. This pattern signifies the head and shoulders of a human.

- Object basic shape classification - The foreground blob can be analyzed to determine if it is of one of the basic shapes such as oval, rectangular, elongated (vertical/horizontal).

- Aggressive human actions - Human motion is first extracted using computations on successive pictures from a video stream. This motion is further analyzed to match motion patterns of those pertaining to fighting, battering, etc. These classifications are enabled by a machine learning engine (see below) trained with samples of such actions.

Table 1 below summarizes the preferred algorithm components from each sensor for each event, according to example embodiments. Events that involve airborne sounds, such as human voices, are preferably captured by an audio microphone. Events that involve structural sound, such as scribbling on a wall, urinating, are preferably captured by a contact microphone.

	Audio signal	Audio signal	Image processing	Computer vision
	Vibration signal	Vibration signal	Human profile detection	Aggressive emotions
	Vibration signal	Audio signal	Image processing	
		Vibration signal	Human profile detection	
			Posture recognition	NIL
			Scene changes on the floor area	
	NIL	NIL	Object shape	Changes in background scene
	Audio signal	Audio signal	Lift control signal	Computer vision
	Vibration signal	Vibration signal	Image processing	Aggressive emotions
			Scene changes on wall and door areas	

Table 1

In the detection of events the inventor has recognized and the embodiments of the invention exploit the following features:

- The nature of the events is such that the information derived from the signal of one of the audio or contact microphones is generally more informative compared to the information derived from the video signal.
- In crimes against soft targets in lifts, stronger audio signals (such as crying, shouting, screaming, etc) are generated compared to audio signals arising from vigorous physical actions (such as the act of fighting back).

- The act of urinating in the lift can be detected from the sound of liquid dripping due to the contact between the lift structure (e.g.: floor/wall) and urine flow. A liquid patch can also be detected by image processing to confirm the event. In other words, when a dripping sound is detected, a video signal can be concurrently analyzed to see if a human is detected. In order to reduce the number of false alarms, a greater number of concurrent verifications can be performed. This, however, is at the cost of more computational resources.
- Some acts of vandalism (e.g. breaking, scribbling) can be more easily detected by the sounds produced (e.g. when a sharp object is used to carve on a lift wall). Detection that is simply based on image processing is relatively more unreliable.
- Depending on the event, one of the two microphone signals (i.e. contact or audio) will be more accurate for certain events compared to others.
- The confined environment of a lift allows the microphone signals to be more clearly captured for analysis. Conversely, the confined environment may not allow information such as an object's spatial position or rigorous actions to be useful in detecting events. Thus, an AI system based on camera signal alone is generally ineffective.

It will be appreciated by a person skilled in the art that depending on the nature of the event to be detected, different types of sensors that are more suitable than others. For example, in the above description, it is preferable that in order to detect anti social behaviours (e.g. urinating in lifts and dumping of rubbish), vandalism and crimes against soft targets, audio microphones and/or contact microphones are utilized.

Dynamic self-learning

Furthermore, there is additionally provided a dynamic self-learning mechanism that allows the classifiers to automatically learn variations of an event when the system is deployed. The system advantageously improves itself as it is exposed to a greater variety of data in actual use.

The use of multiple sensors in the manner described above advantageously enables the Master Classifier of a particular event to use event data to train the other secondary classifiers. Data pertaining to the event detected by the system can be fed back to the Slave Classifiers that assist the Master Classifier. This incremental learning can widen the exposure of the system to real events that take place at a particular site and makes the overall classifier smarter compared to at the initial period when it is trained with mock-up data.

Embodiments of the present invention utilize multiple sensors wherein each one of the sensors operates in a comprehensive manner (instead of point sensors such as an acid detector which is localized to the point where the sensor is installed), advantageously giving rise to robustness and a relatively greater variety of events that can be detected. In addition, as embodiments of the present invention are software based, there is a potential for expansion to cover other situations (events) that may be required by users.

In addition, embodiments of the present invention advantageously enable better control and management of lifts with a more efficient and economical way for users to respond to events (e.g. via mobile devices, compared to requiring a monitoring centre). Furthermore, embodiments of the present invention provide real-time detection of these events. Alarms can then be forwarded to relevant agencies (e.g. police), if applicable and immediate action can be taken to help the victims.

Further advantages include cost-savings as there is a reduced need for extra cleaning services (e.g. due to urination and trash removal), damages from vandalism and illegitimate operation of the lift. Productivity is decreased due to the typically long and tedious investigation of an event. For instance, when an urination incident occurs, it is first learnt from a user's complaint. This is recorded through a standard administrative process. Investigative action may be taken some days/weeks later. The process involves the localization of the video clips that are retrieved from a centralized recording centre. These videos are then studied frame by frame to locate the incident. Embodiments of the present invention can provide real-time detection where a culprit can be caught

quickly. Thus, investigation is not necessary all the time. In the investigation process of events that are not detected or events that are not configured to be detected but are of interest, users can retrieve recordings that are only to specific conditions (e.g. number of people found in the lift, their action levels, voice or audio profile) rather than manually review voluminous CCTV recordings. In the urination incident for instance, users can employ embodiments of the present invention to filter video segments where there is only a single person in the lift, presumably where there is a higher likelihood of urination.

Figure 3 is a flow chart, designated generally as reference numeral 300, illustrating a method of event detection, according to an example embodiment of the present invention. At step 302, respective sensor signals are obtained from a plurality of sensors. At step 304, the respective sensor signals are processed for the event detection using respective classifier units coupled to the plurality of sensors, wherein an event is detected based on the sensor signal processed in one classifier unit as a main evidence. At step 306, sub-evidence queries are issued to the other classifier units for facilitating the event detection using said one classifier unit.

The method and system of the example embodiment can be implemented on a computer system 400, schematically shown in Figure 4. It may be implemented as software, such as a computer program being executed within the computer system 400, and instructing the computer system 400 to conduct the method of the example embodiment.

The computer system 400 comprises a computer module 402, input modules such as a keyboard 404 and mouse 406 and a plurality of output devices such as a display 408, and printer 410.

The computer module 402 is connected to a computer network 412 via a suitable transceiver device 414, to enable access to e.g. the Internet or other network systems such as Local Area Network (LAN) or Wide Area Network (WAN).

The computer module 402 in the example includes a processor 418, a Random Access Memory (RAM) 420 and a Read Only Memory (ROM) 422. The

computer module 402 also includes a number of Input/Output (I/O) interfaces, for example I/O interface 424 to the display 408, and I/O interface 426 to the keyboard 404.

5 The components of the computer module 402 typically communicate via an interconnected bus 428 and in a manner known to the person skilled in the relevant art.

10 The application program is typically supplied to the user of the computer system 400 encoded on a data storage medium such as a CD-ROM or flash memory carrier and read utilising a corresponding data storage medium drive of a data storage device 430. The application program is read and controlled in its execution by the processor 418. Intermediate storage of program data maybe accomplished using RAM 420.

15 It will be appreciated by a person skilled in the art that numerous variations and/or modifications may be made to the present invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects to
20 be illustrative and not restrictive.

CLAIMS

1. A system for event detection, comprising:
a plurality of sensors; and
5 respective classifier units coupled to the sensors for processing respective
sensor signals from the sensors for the event detection;
wherein at least one of the classifier units is adapted for detecting an event
based on the sensor signal processed in said at least one classifier unit as a main
evidence and for issuing sub-evidence queries to the other classifier units for facilitating
10 the event detection.
2. The system as claimed in claim 1, wherein said at least one classifier unit
comprises a categorisation unit for categorising the sensor signal into event categories.
- 15 3. The system as claimed in claim 2, wherein the categorisation unit is adapted to
use one or more parameters from the sub-evidence queries to the other classifier units
in categorising the sensor signal into the event categories.
4. The system as claimed in any one of the preceding claims, wherein said at least
20 one classifier unit comprises a dominant signature verification unit for verifying a
dominant signature in the sensor signal for the event detection.
5. The system as claimed in claim 4, wherein the dominant signature verification
unit is adapted to use one or more parameters from the sub-evidence queries to the
25 other classifier units in verifying the dominant signature in the sensor signal.
6. The system as claimed in any one of the preceding claims, wherein said at least
one classifier unit comprises a machine learning unit for classification processing for the
event detection.
- 30 7. The system as claimed in claim 6, wherein the machine learning unit is adapted
to use one or more parameters from the sub-evidence queries to the other classifier
units in the classification processing.

8. A method of event detection comprising the steps of:
obtaining respective sensor signals from a plurality of sensors; and
processing the respective sensor signals for the event detection using respective
5 classifier units coupled to the plurality of sensors, wherein an event is detected based on
the sensor signal processed in one classifier unit as a main evidence; and
issuing sub-evidence queries to the other classifier units for facilitating the event
detection using said one classifier unit.

10 9. The method as claimed in claim 8, comprising categorising the sensor signal into
event categories using a categorisation unit of said one classifier unit.

10. The method as claimed in claim 9, comprising using one or more parameters
from the sub-evidence queries to the other classifier units in categorising the sensor
15 signal into the event categories.

11. The method as claimed in any one claims 8 to 10, comprising verifying a
dominant signature in the sensor signal for the event detection using a dominant
signature verification unit of said one classifier unit comprises.

20 12. The method as claimed in claim 11, comprising using one or more parameters
from the sub-evidence queries to the other classifier units in verifying the dominant
signature in the sensor signal.

25 13. The method as claimed in any one any one claims 8 to 12, comprising
performing classification processing for the event detection using a machine learning
unit of said one classifier unit.

14. The method as claimed in claim 13, comprising using one or more parameters
30 from the sub-evidence queries to the other classifier units in the classification
processing.

30

15. A data storage medium having stored thereon computer program code means for instructing a computer system to execute a method of event detection, as claimed in any one of claims 8 to 14.