# DESCRIPTION

## TITLE: Method and system for detecting a static object in an environment monitored by at least one camera

The present invention relates to a method and system for detecting a static object in an environment monitored by at least one camera.

A suspicious item is luggage or any other object that has been abandoned, either deliberately with the intention of committing a malicious act, or unintentionally due to the carelessness of its owner.

Operators of public places (airports, metro stations, subways, trams, etc.) must implement strict security procedures to remove any doubt as to the exact nature of a suspicious package.

Security procedures aim to protect the public, assess the threat posed by the suspicious package and possibly destroy it. Security procedures may vary from context to context and between operators, but generally consist of securing a perimeter around the suspicious package, deploying security personnel with the authority to respond, and possibly implementing appropriate means to destroy the suspicious package.

The launching of these security procedures has major consequences on the smooth running of the venue and the services provided to users (delays, cancellations, etc.). This therefore has significant financial impacts for the operator.

It turns out that the information on the detection of a suspicious package arrives at the decision centre very late in relation to the time when the package was abandoned. The long detection time required for the information to be passed on makes it almost impossible to find the owner in order to qualify the criticality of the event. An alert is therefore systematically triggered and the corresponding security procedure initiated, usually until the suspect package is destroyed.

It is therefore desirable to minimise the time of detection in order to increase the chances of qualifying the criticality of the event and not having to initiate the associated safety procedures and, in any case, to reduce the time between the abandonment of the package and the moment when the situation returns to normal and the place reopens to the public.

Operators often use video surveillance to support the task of detecting suspicious packages in the premises they manage.

The image processing algorithms applied to video surveillance images theoretically allow alerts to be automatically sent to the operators in charge of removing doubts.

However, the level of performance (detection) and sensitivity (false alarms) of existing solutions are still incompatible with use in public places, which are characterised by the presence of a variable number of users (either stationary or moving) and by uncontrolled luminosity (detection of shadows, ground patterns, etc.).

The purpose of the present invention is to solve this problem by providing a solution suitable for public places.

The object of the invention is a method of detecting a static object in an environment monitored by at least one camera, said method comprising: a step of acquisition, by a camera, during a detection time window, of a stream of images, each image of the stream of images being indexed by an acquisition date and corresponding to the observation of a scene in the environment; then, for each image of the stream of images, a pre-processing step consisting of : subtracting from the image in question a background image corresponding to a background of the observed scene, in order to obtain a shape image, which contains shapes present in the observed scene; isolating, in the image in question, a silhouette for each person present in the observed scene, in order to obtain a silhouette image ; and combining the shape image and the silhouette image associated with the image under consideration to obtain a processed image, which comprises only objects present in the observed scene, and finally, on a time succession of processed images corresponding to the observation of the same scene of the environment, a step consisting in tracking from one processed image to the next each object present in the environment and in emitting an alert as soon as an object among the objects present is identified as being static during the detection time window.

This process drastically reduces the time taken to detect a suspicious package, while at the same time offering high reliability, so that false alarms are minimised and security procedures are only initiated when justified.

According to particular embodiments, the assembly comprises one or more of the following features taken in isolation or in any combination that is technically possible:

- the method further comprises a step of automatically identifying alerts received from the plurality of cameras in order to filter out false alerts.

- the identification step comprises a classification sub-step allowing a static object associated with an alert to be assigned one of a plurality of predefined classes corresponding to types of suspect package.

- the identification step includes a sub-step for assigning, to a static object associated with an alert, one of a plurality of dynamic classes, the dynamic classes being defined during the implementation of the method, the defined dynamic classes being stored in a white list.

- a predefined class corresponds to an infrastructure element belonging to the background of the scene whose aspect evolves over time, an alert associated with an object belonging to the said predefined class being suppressed as being considered a false alert.

- a predefined class corresponds to a new type of suspicious package, an alert associated with an object belonging to said predefined class being confirmed as a real alert.

- the identification step consists of applying a classification algorithm of the artificial neural network type.

The invention also relates to a detection system for the implementation of the above detection method.

Preferably, the system comprises at least one camera, the camera being adapted to perform the steps of acquiring an image stream, pre-processing the images of the image stream and tracking the objects during a detection time window so as to issue an alert in the event of detection of a static object.

Preferably, the system comprises a control centre in communication with the at least one camera and adapted to filter the alerts generated by the at least one camera by implementing the automatic identification step.

This system has the ability to evolve based on operator feedback to better filter out false alarms and thus adapt to the scene the camera is actually observing.

The invention and its advantages will be better understood upon reading the following description of a particular embodiment, given only as an example, and with reference to the attached drawings, in which:

- Figure 1 is a schematic representation of an embodiment of the detection system according to the invention ; and,

- Figure 2 is a schematic representation of an embodiment of the detection method according to the invention.

DETECTION SYSTEM

Referring to Figure 1, an embodiment of a detection system according to the invention will be presented.

The detection system 1 comprises a plurality of cameras 10-1, 10-2, ... 10-N, a communication network 20 and a control centre 30.

The function of a plurality of cameras is to monitor an environment, especially a place where the public is present, such as an airport. Each camera observes the environment from a specific point of view. Therefore, the scene observed by a camera is specific to that camera.

The plurality of cameras comprises N cameras, which are referenced 10-1, 10-2, ... 10-N in Figure 1.

The cameras are identical to each other. The structure of a camera, for example the 10-N camera, will now be presented.

A camera comprises an optical sensor 11, acquisition electronics 12, a computer 13 and a communication interface 17.

The optical sensor 11 is suitable for observing a scene in the environment.

The acquisition electronics 12 is adapted to generate an image signal, or simply image in the following, from the light collected by the optical sensor 11.

The computer 13 has computing means, such as a processor 14, and information storage means, such as a memory 15. In particular, the memory 15 stores computer programs, whose instructions, when executed by the processor 14, enable certain functionalities to be implemented. In particular, among these programs, memory 15 stores a pre-processing program 16, the execution of which enables the implementation of part of the detection process according to the invention, as will become apparent hereafter.

The communication interface 17 connects the camera to a network in order to exchange, on the receiving end, camera parameterisation information (for example, parameters for adjusting the optical properties of the sensor 11, or parameters for configuring the program 16) and, on the transmitting end, all or part of the raw images or pre-processed images and/or suspicious package detection alerts.

The plurality of cameras is connected to a communication network 20. For example, this is a Local Area Network (LAN), using the Ethernet protocol for example.

Un centre de contrôle 30 est par exemple un ordinateur. It comprises a communication interface 32, a computer 33, a peripheral interface 37 and a plurality of peripherals 38.

The communication interface 32 connects the control centre 30 to the network 20, in particular to allow communication with the plurality of cameras, in reception, of image pre-processing information delivered by the various cameras and, in transmission, of camera control information.

The computer 33 comprises computing means, such as a processor 34, and information storage means, such as a memory 35. In particular, the memory 35 stores computer programs, whose instructions, when executed by the processor 34, enable certain functionalities to be implemented. In particular, among these programs, memory 35 stores an alert filtering program, the execution of which enables the implementation of part of the detection process according to the invention, as will become apparent below.

The memory 35 also stores a white list 39 of objects that have led to a false alarm, but which are not suspicious, for example belonging to the background of the observed scene although changing in appearance over time.

The plurality of peripherals 38 allow an operator to interact with the control centre 30. For example, the plurality of peripherals 38 comprises, for example, a screen that can be used to view images acquired by a particular camera, a mouse that can be used to select a particular area of an image displayed on the screen, and a keyboard that can be used to update the white list 39, for example by entering a descriptor of the selected area on the screen.

The skilled person will note that the detection system 1 is part of a distributed architecture in which certain tasks are deported so as to be carried out in the cameras or at least as close to the cameras as possible (this concerns in particular the tasks of pre-processing the images acquired by each camera individually) whereas other tasks are centralised in the control centre (this concerns the tasks centralising all the alerts raised by the cameras).

## DETECTION METHOD

Referring to figure 2, a mode of implementation of the detection process according to the invention will be presented in detail.

The detection process 100 starts with a step 110 of acquiring an image stream by a camera (for example the camera 10-1 in figure 1), more precisely by the optical sensor and the associated acquisition electronics.

Each image in the stream is indexed with the time t at which it was acquired. The image at time t is thus denoted I(t). Since the camera operates with a sampling time $\Delta t$, the image that follows the image I(t) is the image $I(t+\Delta t)$.

An image corresponds to the observation of a scene, defined as the observation of a zone of the environment according to a certain pose of the camera, i.e. according to an orientation of the optical axis of the camera and taking into account the optical properties of the latter (in particular the angle of aperture of the camera).

As the cameras are fixed, the pose of each camera is static. Therefore, the images in the image stream delivered by the same camera all correspond to the same observed scene.

The second step 120 of the process 100 consists of applying pre-processing to each image I(t) acquired. Preferably, the computer 13 performs this step by running the pre-processing program 16.

In a first substep 122 a background image I0 is subtracted from the image I(t). At the end of this first sub-step, an image of shapes at time t is obtained, noted If(t).

Background subtraction is a known treatment. An example is given in the article by Z. Zivkovic, “Improved adaptive Gaussian mixture model for background subtraction”, in Proceedings of the 17th International Conference on Pattern Recognition, August 2004, ICPR 2004, Vol. 2, pp. 28-31, IEEE.

The background image I0 is an image of the background of the scene observed by the 10-1 camera. Advantageously, it is obtained by means of the 10-1 camera that enabled the acquisition of the image I(t) in question.

The background image I0 corresponds to the observation of the scene when it is empty, i.e. when there is no person or object (other than the objects belonging to the scene background). The background image I0 is obtained, for example, by acquiring images at night, when the location is not frequented. It can also be obtained by combining portions of images that remain invariant over a long period of observation of the scene.

Preferably, the shape image If(t) is a binary image, a pixel taking for example the value zero when the corresponding pixel of the image I(t) is in the background image I0, and the value unity otherwise.

The shape image if(t) contains a priori different shapes, which correspond to the people and objects, static or moving, present in the observed scene at time t.

This first sub-step is not sufficient because, in public environments with a high density of people, the immobility of a user, present next to a piece of luggage, makes it difficult to distinguish the object from the user: in the image of forms there will be no border between these two forms, which will be merged into a single form.

In addition, in an environment where the brightness is not controlled, false alarms are often generated by shadows, patterns on the floor, objects in the background that are

not static (such as escalators, lifts, automatic doors, advertising signs). These infrastructure elements, coming from the background of the scene, cause artefacts in the If(t) shape image.

The pre-processing step 120 thus comprises a second sub-step 124, conducted in parallel with the first sub-step 122 and consisting of isolating, in the image I(t), a silhouette for each person present in the observed scene. An image of silhouettes at time t, denoted Is(t), is obtained at the output.

Preferably, an algorithm of the person segmentation type is applied to the image I(t).

Algorithms such as person segmentation are known. An example of such an algorithm is presented in the paper by Cao, Z., Martinez, G. H., Simon, T., Wei, S. E., & Sheikh, Y. A., “OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields”, 2019, IEEE transactions on pattern analysis and machine intelligence.

A person segmentation type of algorithm is used to identify whether a person is present in the observed scene. If so, a silhouette is associated with this person.

The silhouette image groups together the silhouettes associated with each of the people in the image I(t).

Preferably, the silhouette image Is(t) is a binary image, with a pixel taking, for example, the value zero when the corresponding pixel in the image I(t) is within the silhouette of a person in the observed scene, and the value unity otherwise.

The third sub-step 126 of the second step 120 consists in combining the images obtained from the first and second sub-steps 122 and 124 from the same image I(t) of the stream in order to obtain an image of objects at time t, noted Io(t).

The object image Io(t) now only contains those shapes in the observed scene that correspond to objects, the shapes corresponding to people having been eliminated.

For this purpose, the shape image and the silhouette image at time t, If(t) and Is(t), are combined pixel by pixel, for example, by applying a logical “AND” operator. A pixel in the object image is set to unity, for example, when the corresponding pixel in the shape image If(t) is inside a shape and when the corresponding pixel in the silhouette image Is(t) is outside a silhouette, and to zero otherwise.

This combination step is particularly interesting since it limits the pixels of interest to a small number of pixels, which makes downstream processing, such as object tracking, easier.

Advantageously, in a fourth sub-step 128, the pixels of the image of objects Io(t) that can be associated with the same object are brought together in a region delimited by a contour (“blob”).

A region detection algorithm is applied to the image Io(t) to define areas whose pixels have substantially constant properties, but which differ from the pixels in the surrounding areas. This is a classical approach based on edge detection and for example illustrated by the article by S. Suzuki, “Topological structural analysis of digitized binary images by border following”, Computer vision, graphics, and image processing, 30(1), 32-46, 1985.

The output is a processed image at time t, denoted I*(t).

After pre-processing the image at time t, I(t), the next image in the stream, I(t+Δt), is processed in the same way.

The third step 130 of the method 100 consists in analysing a temporal succession of processed images I*(t). For example, images within a detection time window of a few seconds are taken into account.

This involves tracking an object in the observed scene from one processed image to the next in order to identify static objects in the detection time window.

A tracking algorithm is used here. This processing step is known as such. It is explained for example in Maggio, E., & Cavallaro, A, “*Video tracking: theory and practice*. John Wiley & Sons” published in 2011.

An alarm Ai is issued for each object Oi among the objects in the observed scene that is identified as stationary throughout the detection time window.

Advantageously, a vignette Zi is isolated in the image I(t) around the object Oi. Only this sticker Zi is transmitted, together with the alert Ai, to the control centre 30 for further processing 100.

The detection method 100 includes a fourth step 140 of automatically identifying alerts to filter the alerts to limit the number of false alerts and confirm the true alerts among all the alerts Ai received from the plurality of cameras.

Step 140 comprises a first classification sub-step 142. It consists in running a classification algorithm on a thumbnail Zi to determine the class to which the corresponding object Oi belongs.

The classification algorithm is configured to assign one of a plurality of predefined classes. The plurality of predefined classes includes for example the classes “suitcases”, “handbag”, “backpack”, etc.

The classification algorithm used is for example a suitably trained deep neural network. Preferably, this classification algorithm is the one presented in the paper by He, K., Zhang, X., Ren, S., & Sun, J., “Deep residual learning for image recognition”, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.

The output of substep 142 is a score S1 of membership in one of the plurality of predefined classes that the classifier can distinguish

In parallel to the first sub-step 142, step 140 comprises a second sub-step 144 of determining membership of the white list 29.

This step implements a whitelist membership determination algorithm such as the one presented in the paper Chen, Y., Zhou, X. S., & Huang, T. S., “One-class SVM for learning in image retrieval”, in Proceedings 2001 International Conference on Image Processing, Vol. 1, pp. 34-37, October 2001, IEEE.

The white list 29 contains infrastructure elements that belong to the background of the scene but whose appearance changes over time: shadows of static equipment, automatic doors that open and close, advertising signs that change, moving escalators, etc. Such changing infrastructure elements cannot be removed from the image in pre-processing, since it is not possible to know which elements are in the background when the camera is installed. Such changing infrastructure elements cannot be removed from the image during pre-processing, since it is not possible, at the time of camera installation, to know which objects in the background of the scene actually observed might lead to artefacts in the processed image, and therefore to false alarms.

The white list then contains the infrastructure elements that belong to the background and their characteristics. It is set dynamically in step 150, as detailed below.

Substep 144 is to determine whether the thumbnail Zi associated with an alert Ai is in fact one of the items in the white list. More precisely, in a first step, a neural network is used to extract characteristic elements from the thumbnail, and then, in a second step, a comparator compares, with the help of a machine learning algorithm, these characteristics with those of characteristic elements from the white list. If it is sufficiently similar to a whitelisted infrastructure element, the sticker is considered to correspond to that whitelisted infrastructure element. This processing will therefore identify image artefacts that are due to known changes in the background of the scene and should therefore not raise any alarm. Sub-step 144 is in fact also a classification step, but on a set of classes (the whitelist infrastructure elements) that evolves dynamically according to the history of use of the detection system.

Sub-step 144 calculates a second score S2.

Finally, step 140 includes a third voting sub-step 146 to compare the first and second thresholds S1 and S2 with each other. If the second threshold is higher than the first for the Ai alert, the alert is removed. If not, the Ai alert is validated and the validated Ai* alert is displayed on the screen for the operator to see and initiate the appropriate response.

False alarms are therefore automatically suppressed.

Advantageously, the method 100 comprises a configuration step 150 allowing the white list 29 to evolve during the use of the detection system and the implementation of the method, allowing the false alarm rejection rate to be increased.

To do this, in a sub-step 152, the operator checks the exact nature of the validated alert Ai*. The images that caused the alert are shown to him.

If it considers this to be a false alarm, in a sub-step 154 it updates the whitelist. The operator adds this false positive to the whitelist of uninteresting objects so that objects similar to those in the whitelist do not return alerts in the future.

Thus, the method 100 identifies in a succession of images of the same scene any static object that does not correspond to an object of a type deemed to be non-interesting.

Note that the whitelist is shared centrally and thus benefits all cameras 10-1 to 10-N, not just the camera from which the so-called “not interesting” object was observed.

In the state of the art, algorithms based on pre-trained classification are not used for the specific case of detecting suspicious packages in public places. They are not because the training must be carried out on a corpus of data containing a predefined set of object classes to be recognised, this set being fixed at the moment of deployment. Thus, if it is initially trained to recognise suitcases, backpacks or handbags, such an algorithm will not be able to recognise objects such as shopping trolleys, prams or plastic bags. This is extremely limiting for effective detection.

The invention solves this problem. Thus, as an alternative to the above-described embodiment, new classes can be created dynamically using the whitelisting mechanism.

For example, a whitelisting mechanism for infrastructure elements is implemented, leading to the rejection of alerts classified as such, and a whitelisting mechanism for new threats leading to the confirmation of alerts classified as such.

Thus, each alert that is not discarded by the white list of infrastructure elements is submitted to an operator, who then determines whether or not the corresponding event constitutes a new kind of threat, which is not part of the set of predefined classes in step 142. If so, it then indicates what the observed object is. This object is included in an

intermediate list, either as a new occurrence of an existing threat category or as the first occurrence of a new threat category. Once a category in this intermediate list has a sufficient number of samples, a new class is automatically created in the new threat whitelist for automatic detection by implementing an algorithm similar to that in step 144.

In relation to the invention, the prior art also does not allow non-invariant objects in the background of the observed scene to be recognised as irrelevant.

The prior art does not implement algorithms based on pre-trained classification, as these algorithms often have high false alarm rates, the reduction of which is achieved at the cost of a drastic decrease in the rate of detected objects. Again this is extremely limiting for the detection of suspicious packages, as it is much better to detect all suspicious packages, even at the cost of false alarms, than to miss one that turns out to be really dangerous.

Finally, these classification algorithms have a very high computational complexity. The pre-processing steps according to the invention allow to limit the computational complexity. In addition, dynamic consideration of the actual situation through the use of a white list reduces the number of predefined classes to be followed.

## CLAIMS

**1.** A method of detecting (100) a static object in an environment monitored by at least one camera, said method comprising :

- a step (110) of acquisition, by a camera (10-1), during a detection time window, of a stream of images, each image of the stream of images being indexed by an acquisition date and corresponding to the observation of a scene of the environment;

then, for each image in the image stream, a pre-processing step (120) consisting of :

- subtracting (122) from the image under consideration, a background image corresponding to a background of the observed scene, to obtain a shape image, which contains shapes present in the observed scene;
- isolating (124) a silhouette for each person in the observed scene in the image under consideration, to obtain an image of silhouettes; and,
- combining (126) the shape image and the silhouette image associated with the image under consideration to obtain a processed image, which comprises only objects present in the observed scene,

and finally, on a time succession of processed images corresponding to the observation of the same scene of the environment, a step (130) consisting in following from one processed image to the next each object present in the environment and in emitting an alert as soon as an object among the objects present is identified as being static during the detection time window.

**2.** A method according to claim 1, further comprising a step (140) of automatically identifying alerts received from the plurality of cameras in order to filter out false alerts.

**3.** Method according to claim 2, in which the identification step (140) comprises a classification sub-step (142) making it possible to assign, to a static object associated with an alert, one of a plurality of predefined classes corresponding to types of suspect package.

**4.** A method according to claim 2 or claim 3, wherein the identification step (140) comprises a substep (144) for assigning, to a static object associated with an alert, one of a plurality of dynamic classes, the dynamic classes being defined during the implementation of the method, the defined dynamic classes being stored in a white list.

**5.** A method according to claim 4, wherein a predefined class corresponds to a piece of infrastructure belonging to the background of the scene whose appearance

changes over time, an alert associated with an object belonging to said predefined class being suppressed as a false alert.

**6.** A method according to claim 4 or claim 5, wherein a predefined class corresponds to a new type of suspect package, an alert associated with an object falling within said predefined class being confirmed as a true alert.

**7.** A method according to any one of claims 2 to 5, wherein the identification step comprises applying a classification algorithm of the artificial neural network type.

**8.** A detection system (1) for implementing a detection method (100) according to any of the preceding claims.

**9.** System according to claim 8, comprising at least one camera (10-1), the camera being suitable for carrying out the steps of acquiring (110) a stream of images, pre-processing (120) the images of the stream of images and tracking (130) the objects during a detection time window so as to issue an alert in the event of detection of a static object.

**10.** System according to claim 9, comprising a control centre (30) in communication with said at least one camera (10-1) and able to filter the alerts generated by said at least one camera by implementing the automatic identification step (140).