(12) UK Patent (19) GB (11) 2585919 (13) B

(45) Date of B Publication 14.09.2022

(54) Title of the Invention: Method and system for reviewing and analysing video alarms

(51) INT CL: ***G08B 29/18*** (2006.01) ***G08B 13/196*** (2006.01) ***H04N 7/18*** (2006.01)

(21) Application No: **1910561.8**

(22) Date of Filing: **24.07.2019**

(43) Date of A Publication **27.01.2021**

(56) Documents Cited:

WO 2006/101477 A1	**US 20190141297 A1**
US 20190066472 A1	**US 20040201471 A1**
US 10055961	

(58) Field of Search:
As for published application 2585919 A viz:
INT CL **G08B, H04N**
Other: **WPI, EPODOC**
updated as appropriate

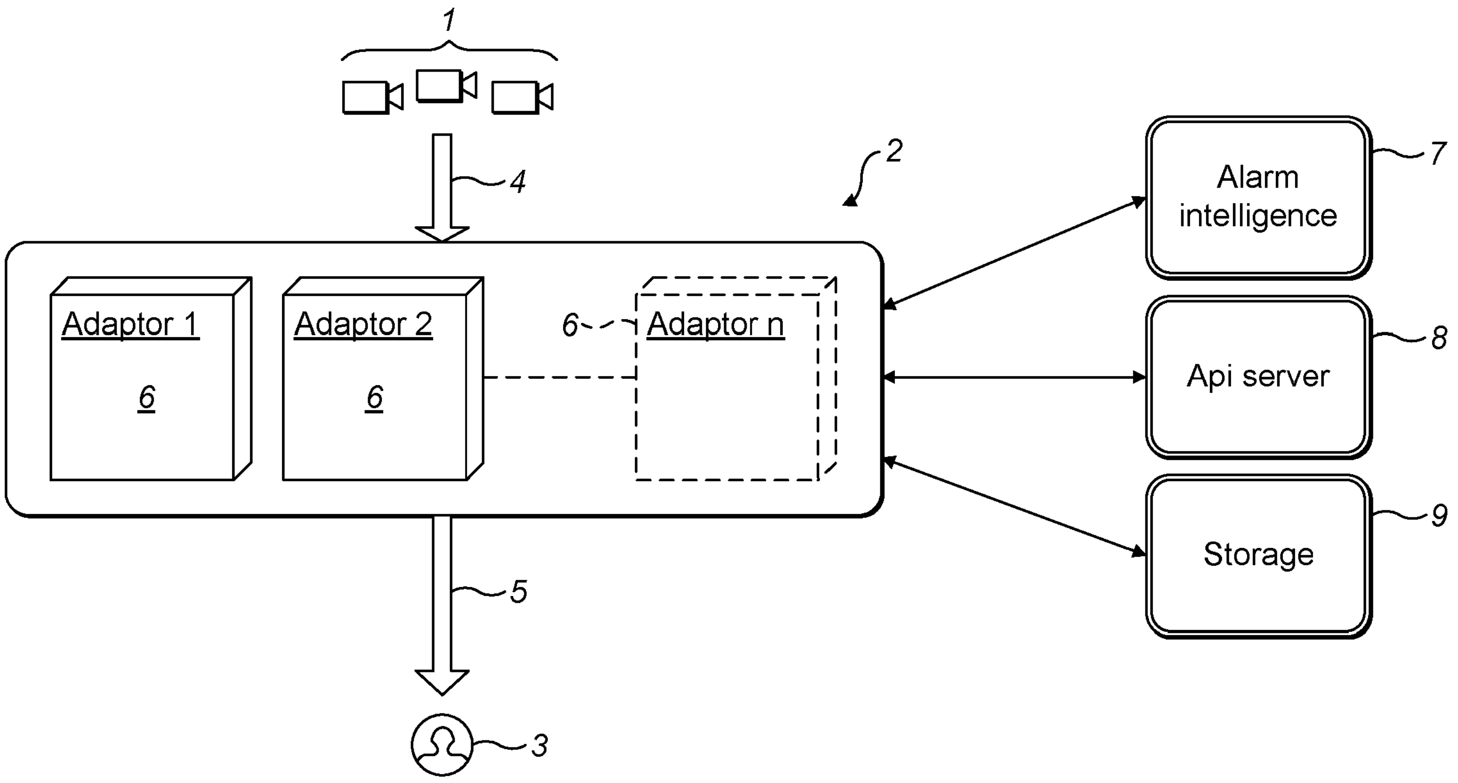
Additional Fields
Other: **None**

(72) Inventor(s):
Boris Ploix
Mohammad Rashid
Siddharth Agrawal
Ashwin D'Cruz
Chris Tegho
Anton Veselev

(73) Proprietor(s):
Calipsa Ltd
8 Duncannon Street, LONDON, WC2N 4JF,
United Kingdom

(74) Agent and/or Address for Service:
Optimus Patents Limited
Peak Hill House, Steventon, BASINGSTOKE,
Hampshire, RG25 3AZ, United Kingdom

12 10 20



1
4
2
Adaptor 1
6
Adaptor 2
6
6
Adaptor n
5
3
Alarm intelligence
7
Api server
8
Storage
9

FIG. 1

The following terms are registered trade marks and should be read as such wherever they occur in this document:

Hikvision
Dahua
Videofied
Adpro

Method and system for reviewing and analysing video alarms

The present disclosure relates to a method and system for reviewing and analysing video alarms, in particular to such a system and method for autonomously reviewing and filtering false alarms in the context of video verification or remote guarding.

A conventional video verification service works in conjunction with burglar alarms to validate an incident. When a locally installed sensor detects an anomaly, a locally installed video camera records a short clip, which is sent to a centralised/remote control room managed by a security company. An operator in the control room manually reviews the clip and, if the operator sees a problem, contacts the police or another relevant emergency response service.

The locally installed sensor may be a hardware sensor (such as an Infrared sensor). There may alternatively or additionally be provided some form of local video analytics performed. Such local video analytics requires a server to be installed locally that processes real-time video. Processed video is sent to the control room for manual review.

There are a number of limitations with the prior art approach:

1. High cost of installation - Each install requires an engineer to visit the site, spend several hours configuring the device and requires the site to be revisited for maintenance. This is in addition to the cost of the server

2. Old fashioned computer vision techniques – Prior art systems rely on hand crafted rules such as size, shape, colour to detect and classify objects. These systems are extremely sensitive to weather and environment and hence lead to a large number of false alarms. Because the system is a local installation, it doesn’t improve or evolve over time.

3. Inability to scale - Traditional analytics systems depend on analysing live video. Due to bandwidth limitations this must be done locally - it is impossible to stream the outputs from 1000s of cameras to a remote location for analysis.

The present invention arose in a bid to provide an improved method and system of reviewing and analysing video alarms capable of reducing false alarms

According to the present invention in a first aspect, there is provided a method for reviewing and analysing video as recited by Claim 1.

By such an arrangement a solution is provided that sits between the alarm device and a manual reviewer (such as an operator at a central monitoring station or a user of a residential security system) and acts like a firewall. Alarms generated at the monitored location are intercepted and automatically reviewed before they are sent for manual review. False alarms may be discarded and true alarms may be forwarded for manual review.

The solution is unique, as compared to the prior art, since automatic analysis is based on the limited number of frames and not on a live stream. Such analysis allows for the offsite/centralised analysis of video and scalability of the system.

Most preferably, upon validation of the alarm, the limited number of frames are automatically forwarded to a further computer for manual review. In the event the alarm is dismissed, which will be the case with a false alarm, the limited number of frames will not be forwarded.

The further computer may be a computer at a central monitoring station (also known as an alarm receiving centre) or may be a personal computing device, including but not limited to a smartphone, tablet computer or other such device. The video source will be linked to the central monitoring station or personal computing device through a user account associated with the video source.

According to the present invention in a further aspect, there is provided an analysis system for implementing the method as detailed above. The analysis system comprising a computer that is arranged to automatically analyse a limited number of frames from the video data and in dependence on the analysis, validate or dismiss an alarm.

Further, preferred, features are presented in the dependent claims.

The present invention will now be described by way of example with reference to the accompanying drawing, in which:

Figure 1 shows, schematically, a system implementing a method for reviewing and analysing video alarms according to an embodiment of the present invention.

In accordance with a first embodiment, there is provided a method for reviewing and analysing video alarms, comprising: receiving video data of an event at a monitored location from a video source associated with an alarm sensor, following an alarm signal; isolating a limited number of frames from the video data in a segment of interest of the video data; sending the limited number of frames via a data connection to a computer that is located geographically remotely to the monitored location; upon receipt by the computer of the limited number of frames, automatically analysing the limited number of frames by the computer; and in dependence on the analysis, validating or dismissing the alarm.

A system implementing the method is shown, schematically, in Figure 1. The system comprises a plurality of video cameras 1, an analysis system 2, which comprises one or more computers, and one or more further computers 3. The arrow 4 indicates the data connection between any one or more of the cameras and the analysis system 2 and the arrow 5 indicates a data connection between the analysis system 2 and any one or more of the further computers 3. The data connections may be wired or wireless, as will be appreciated by those skilled in the art. The analysis system 2 is most preferably cloud based, i.e. implemented using cloud computing.

It is to be noted, as will be readily appreciated by those skilled in the art, that there may be any number of video cameras 1 at any number of locations. The present invention is not to be limited in this regard. There could, for example, be tens of thousands of video cameras at tens of thousands of different locations. Similarly, there may be any number of further computers 3 at any number of locations and the invention is again not to be limited in this regard. A key advantage of the present invention is its scalability. The scalability is possible since limited data is sent from the video cameras to the analysis system 2. There may be 9 or less frames of video sent, more preferably 6 or less frames of video sent and most preferably 3 or less frames of video sent. With a limited number of frames sent, there is no bandwidth issue.

The video cameras 1 will each comprise, or at least be associated with, an alarm sensor that triggers with motion. The alarm sensor may operate on the basis of infrared or otherwise, as will be appreciated by those skilled in the art.

Core to the present invention is the analysis system 2, which may be integrated into existing security camera system infrastructure. That is to say, the analysis system 2 may be configured to receive and send data from/to existing alarm system hardware components. The analysis system 2 may, for example, receive alarm signals from existing/third party video cameras and forward such signals, following validation to existing/third party alarm monitoring systems/user hardware for subsequent manual analysis. As discussed, the analysis system 2 is preferably cloud based.

The analysis system 2 receives the limited number of frames of video from the video cameras 1 and the automatic analysis by the analysis system 2 is based on the limited number of frames (3 or less) and not on a live stream. Such analysis allows for the offsite/centralised analysis of video and scalability of the system, as discussed.

The analysis system 2 is camera agnostic and capable of operating with the cameras of any existing manufacturers. There is presently no common alarm protocol. Each manufacturer generally uses its own protocol to emit alarms. The analysis system 2 preferably allows for integration with any known camera brands/manufacturers.

As shown in Figure 1, the analysis system comprises a plurality of adaptors 6. Each adaptor is arranged to handle one alarm protocol only. The adaptors will be implemented in suitable software. The adaptors are essentially translation modules that are each arranged to receive alarms according to a specific protocol/data format and to convert those alarms into a suitable (common/standard) data format to be understood by the other software components of the analysis system 2. By making the heterogeneous alarms uniform using the adaptors, it becomes possible to decompose the software into micro services, each of which is responsible for one thing only.

All cameras are built around standard protocols which, include SMTP (Simple Email Transfer Protocol), HTTP (HyperText Transfer Protocol) and TCP (Transport Control Protocol). Preferably, whilst different camera brands can usually support more than one of those protocols, the analysis system 2 is configured to associate each brand with only one protocol with such protocol being used for the integration.

The following table, by way of non-limiting example only, illustrates several camera brands against the protocols implemented in the analysis system 2:

Protocol	Description	Camera brands
SMTP	Alarms are sent by email, although the format of those emails can vary greatly.	Hikvision®, Dahua®, Videcon, etc.
HTTP	HTTP request sent with the alarm	Rialto
TCP	Alarms will be streamed bit by bit to the server.	Videofied®
TCP (proprietary)	Custom SDK has been used.	Adpro®

In practice, an alarm will be directed by the analysis system 2 to the appropriate adaptor dependent on its protocol. According to the nature of the alarm, various communications may happen with different services 7, 8, 9 within the analysis system 2.

In the depicted arrangement, as is preferable, the following services are provided:

- Alarm intelligence 7 – a service responsible for analysing the images and determining the validity of the alarm.
- API server 8 – a service responsible for communication with a database of the analysis system 2.
- Storage 9 - cloud storage where the alarm is stored for a predetermined period of time, such as 7 days, 30 days or otherwise, as desired.

Each adaptor preferably perform four tasks:

1. Identify the camera unique identifier.
2. Extract the images from the alarm.
3. Communicate with the services within the analysis system.
4. Forward the alarm to the further computer 3.

The deployment, monitoring and maintenance of the services is preferably done by using the Kubernetes open-source container-orchestration. All of the analysis system 2 services are part of a single cluster and are orchestrated into applications (pods), preferably over multiple machines (nodes). Such an implementation allows the underpinning applications to scale up and down in terms of both function and alarm load, whilst ensuring that no data is lost and costs are saved.

It must be appreciated that in alternative embodiments, the services provided may be different. For example, not all of the above services need be provided and/or the services may be modified.

The system is preferably arranged such that upon validation of an alarm, the analysis system 2 forwards on the respective alarm data in its original data format. Accordingly, the system 2 has no impact on the functioning of the alarm system infrastructure it is integrated into. The existing elements of the alarm system infrastructure

For the automatic analysis of the limited number of frames, advanced machine learning techniques are implemented. A number of different machine learning techniques may be implemented, and the present invention is not be specifically limited in this regard. Several non-limiting examples of possible machine learning techniques are presented below.

In one arrangement, a RCNN (Region proposal based Convolutional Neural Network) object detector is implemented. Such an object detector takes an image as an input and is trained to output a set of bounding boxes which represent the location of objects and the probabilities of the classification of these objects. This algorithm is composed of a neural network divided into three different parts. The first part is a standard classifier preferably trained on a big dataset composed of millions of images. It is used to create an internal representation of the image which is used by the second part, which provides the 300 most likely regions where the objects should be. The third part consists in an evaluation based on a basic classifier of each of these region proposals in order to determine their probability of being an actual object.

This object detector is combined with a basic motion detector and the IOU (Intersection over Union) of this combination is used to determine the validity of an alarm, wherein if the IOU contains at least one element meaning an object has been detected and this object is moving, then the alarm is flagged as true. In the opposite scenario, the alarm is false.

An alternative technique, labelled as a moving object detector for ease of reference, which offers increased accuracy, implements a neural network that is specially designed to overlook static detections and only focus on the moving detections. Instead of getting detections frame by frame, this algorithm takes as its input a pair of consecutive images and returns the bounding boxes of moving objects with a set of probabilities for determining their associated classifications.

The architecture of the neural network is based on the ‘DSOD: Learning Deeply Supervised Object Detectors from Scratch’ paper available at http://openaccess.thecvf.com/content_ICCV_2017/papers/Shen_DSOD_Learning_Deeply_ICCV_2017_paper.pdf.

They are a number of benefits conferred by this architecture, including but not limited to the following:

1. Proposal Free: This network directly predicts detection boxes and their classes. Some other approaches that require a separate or integrated network to propose regions of interest also require pre-training on other datasets. The proposal free method allows for focus exclusively on domain specific data.
2. Deep Supervision: Useful learning signals are provided to multiple layers of the network rather than just the final output layer. Doing so allows the provision of more feedback to the model.
3. Stem Block: This is a simple block that allows efficient compression of information throughout the network. By repeatedly using this block, information loss is kept to a minimum while still ensuring that the network doesn’t get too large (which would increase both training and inference time).
4. Dense Prediction Structure: Here, Dense blocks (as proposed by the work in the ‘Densely Connected Convolutional Networks’ paper available at http://openaccess.thecvf.com/content_cvpr_2017/papers/Huang_Densely_Connected_Convolutional_CVPR_2017_paper.pdf) are used. Dense connections join layers of a network even if they aren’t directly next to each other. These extra connections enable earlier feature maps of the network to have a stronger influence on latter layers, if they are deemed useful. This is especially helpful in an object detection setting as it allows the network to identify objects at a wide range of scales.

In a yet further and preferred technique, the moving object detector is combined with a Convolutional Neural Network (CNN). It takes a pair of images as its input and outputs the pixel locations of where the moving object should be. This class of algorithm is usually called a segmentation neural network and aims at giving a classification of the image pixel by pixel. Hence, instead of having bounding boxes image by image, blobs of pixels are presented where movements are seen.

This model preferably has an encoder-decoder architecture composed of:

1. An encoder which takes as its input two images, corresponding to two consecutive frames, concatenated at the channel level, and normalized. The encoder learns a useful representation of the input, and consists of 8 convolutional layers, increasing the number of filters for each layer.
2. A decoder which takes as its input the representation from the encoder and emits two maps of the same width and height of the input images, with 1s at the pixel locations where the model predicts a movement has occurred, and 0s where the models predicts no movement has occurred. The decoder consists of 5 deconvolutional layers, with a decreasing number of filters for each layer. The last deconvolutional layer is initialized with a prior layer that predicts lower probability for the rare class (foreground) at the start of training. This initialization improves training stability.

The loss function consists of two terms for each image, one for each class in the output (motion present, motion absent). Both the terms are normalized focal loss values where the normalization is over the total number of ground truth pixels belonging to the relevant class. The normalization is done so as to capture every motion blob irrespective of size (if the normalization is absent, the loss will reward the network to get most pixels right by compromising on small blobs of motion).

As will be appreciated by those skilled in the art, alternative machine learning techniques may be implemented than those discussed for automatically detecting motion in the analysis system 2.

Numerous modifications and alterations will be readily appreciated by those skilled in the art, within the scope of the claims that follow.

Claims

1. A method for reviewing and analysing video alarms, comprising:

receiving video data of an event at a monitored location from a video source associated with an alarm sensor, following an alarm signal;

isolating a limited number of frames from the video data in a segment of interest of the video data;

sending the limited number of frames via a data connection to a remote computer that is located geographically remotely to the monitored location;

upon receipt by the remote computer of the limited number of frames, automatically analysing the limited number of frames by the remote computer; and

in dependence on the analysis, validating or dismissing an alarm,

wherein the limited number of frames comprises 3 or less frames,

wherein the remote computer implements object detection machine learning techniques to analyse the limited number of frames, and

wherein the remote computer comprises a plurality of translation modules, each translation module being configured to handle alarm data in a different data format to the remaining translation modules, the translation modules each being configured to convert received alarm signals into a common data format for subsequent analysis by the computer, and:

wherein the remote computer is arranged, upon validation of an alarm, to forward on the respective alarm data in its original data format.

2. A method as claimed in Claim 1, wherein upon validation of the alarm, the limited number of frames are automatically forwarded to a further computer.

3. A method as claimed in Claim 2, wherein the further computer is a computer at a central monitoring station or a personal computing device that is linked to a user account associated with the video source.

4. An analysis system for implementing the method as claimed in any preceding claim, the analysis system comprising a computer that is arranged to automatically analyse a limited number of frames from video data, and, in dependence on the analysis, validate or dismiss an alarm.