



- (51) International Patent Classification:
G06T 7/20 (2006.01) *G06K 9/00* (2006.01)
H04N 7/18 (2006.01)
- (21) International Application Number: PCT/EP2012/002661
- (22) International Filing Date: 23 June 2012 (23.06.2012)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
13/194,771 29 July 2011 (29.07.2011) US
- (71) Applicant (for all designated States except US): **TECHNISCHE UNIVERSITÄT BERLIN** [DE/DE]; Straße des 17. Juni 135, 10623 Berlin (DE).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **HERAS EVANGELIO, Ruben** [ES/DE]; Berthelsdorfer Strasse 13, 1204

Berlin (DE). **SIKORA, Thomas** [DE/DE]; Apostel-Paulus-Straße 12, 10823 Berlin (DE). **KELLER, Ivo** [DE/DE]; Friedrich-Ebert-Strasse 42, 14469 Potsdam (DE).

(74) Agent: **KIRICZI, Sven Bernhard**; Schneiders & Behrendt, Mühlthaler Strasse 91 c, 81475 München (DE).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH,

[Continued on next page]

(54) Title: METHOD AND DEVICE FOR VIDEO SURVEILLANCE

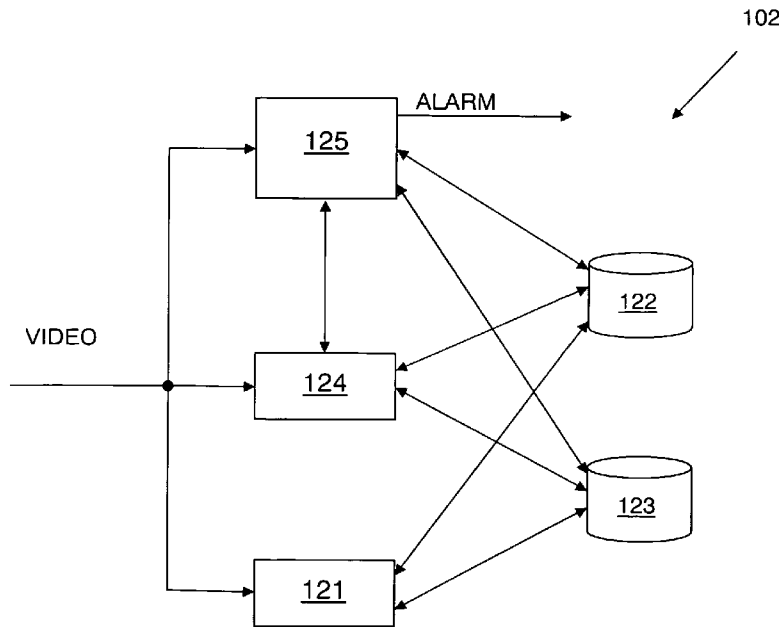


Fig. 2

(57) Abstract: The invention relates to a method and a device for video surveillance, wherein, by means of at least one video camera, an image excerpt of an environment to be monitored in the vicinity of the video camera is recorded, wherein at least one pixel of the image is compared with a corresponding pixel of a short-term background model assigned to the image excerpt and with a corresponding pixel of a long-term background model assigned to the image excerpt

WO 2013/017184 A1

GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ,
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,

SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

METHOD AND DEVICE FOR VIDEO SURVEILLANCE

BACKGROUND AND SUMMARY

[0001] The invention relates to a method and a device for video surveillance, wherein, by means of a video camera, an image of an environment to be monitored in the vicinity of the video camera is recorded.

[0002] EP 1 077 397 A1 discloses a method and a device for the video surveillance of process installations. In that case, a stored first reference image is compared with a first comparison image recorded by a video camera, and an alarm signal is output if the number of differing pixel values is greater than a predetermined threshold value. Furthermore, a second threshold value is provided, which is less than the first threshold value. If the number of differing pixel values lies between these two threshold values, then the associated comparison image is stored as a further reference image and used for subsequent comparisons with newly recorded comparison images.

[0003] WO 98/40855 A1 discloses a device for the video surveillance of an area with a video camera, which optically captures the area from a specific viewing angle, and an evaluation device, wherein video means for optically capturing the same area from a different viewing angle are provided and the evaluation device is suitable for processing the stereoscopic video information originating from the two viewing directions to form three-dimensional video image signal sets and for comparing the latter with corresponding reference signal sets of a three-dimensional reference model.

[0004] US 5 684 898 discloses a method and a device for generating a background image from a plurality of images of a scene and for subtracting a background image from an input image. In order to generate a background image, an image is divided into partial images in order to obtain reference partial images for each position of a partial image, wherein successive partial images are compared with the reference partial image in order to recognize objects between the reference partial image and a video camera that has recorded the image.

[0005] Some known methods for detecting static objects in video sequences are based on the combination of background subtraction methods with tracking information,

-2-

so-called tracking (cf. Bayona, Alvaro, San Miguel, Juan Carlos and Martinez Sánchez, Jose Maria. *Comparative Evaluation of Stationary Foreground Object Detection Algorithms Based on Background Subtraction Techniques*. Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance. 2009, pages 25-30; Guler, S., Silverstein, J.A. and Pushee, L.H. *Stationary objects in multiple object tracking*. Proceedings of the IEEE. International Conference on Advanced Video and Signal Based Surveillance. 2007, 5.248-253; 3. Singh, A., et al. *An Abandoned Object Detection System Based on Dual Background Segmentation*,. Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance. 2009, pages 352-357 and Venetianer, P.L., et al. *Stationary target detection using the object video surveillance system*. Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance, 2007, pages 242-247). What is disadvantageous about systems of this type is that tracking is a problem that is difficult to solve, particularly in scenarios with many moving objects, e.g. airports or stations.

[0006] As an alternative to the use of tracking information, use is made of dual background subtraction methods (cf. Porikli, Fatih, Ivanov, Yuri and Haga, Tetsuji: *Robust abandoned object detection using dual foregrounds*; EURASIP J. Adv. Signal Process. 2008) and methods which interpret the results of a background basic subtraction (cf. Tian, Y., Feris, R.S. and Hampapur, A.: *Real-Time Detection of Abandoned and Removed Objects in Complex Environments*; Proceedings of the IEEE International Workshop on Visual Surveillance; 2008).

[0007] What is disadvantageous about systems of this type is that static objects can also be recorded in the background, which leads to the recognition of “ghost objects” if these objects are removed. In order to solve this problem, the use of finite state machines (FSM) is proposed in Heras Evangelio, Ruben, Senst, Tobias and Sikora, Thomas: *Detection of Static Objects for the Task of Video Surveillance*; IEEE Workshop on Applications of Computer Vision (WACV); 2011 and in Heras Evangelio, Ruben, Pätzold, Michael and Sikora, Thomas: *A system for Automatic and Interactive Detection*

of Static Objects; Proceedings of the IEEE Workshop on Person-Oriented Vision (POV) 2011.

[0008] What is disadvantageous about these systems is that they require knowledge of the empty scenery in advance. In Reddy, V., Sanderson, C. and Lovell, B.C.: A Low-Complexity Algorithm for Static Background Estimation from Cluttered Image Sequences in Surveillance Contexts; EURASIP Journal on Image and Video Processing; 2011 and Gutchess, D., et al.: *A background model initialization algorithm for video surveillance*; IEEE ICCV; 2001 attempts are made to provide a model of the empty scenery. In that case, however, the empty scenery has to be visible at least for a short moment during an initialization phase. This can be difficult particularly in certain situations, such as in public places and in buildings, since the background may not be visible for a long time in these cases.

[0009] It is an object of the invention to improve video surveillance, or make it more robust, particularly with regard to the recognition of static objects. In this case, it is desirable, in particular, for the video surveillance to require no initialization with empty scenery. It is an object of the invention, in particular, to provide video surveillance for recognizing static objects with a high degree of recognition certainty in conjunction with a lower false alarm rate. It is desirable, in particular, to provide video surveillance for recognizing static objects which is suitable particularly for situations with a high proportion of non-static objects. It is desirable, in particular, to provide video surveillance for recognizing static objects which is particularly suitable for airports and stations.

[0010] The object mentioned above is achieved by means of a method for video surveillance, wherein, by means of at least one video camera, an image of an image excerpt of an environment to be monitored in the vicinity of the video camera is recorded, wherein at least one pixel of the image is compared with a corresponding pixel of a short-term background model assigned to the image excerpt and with a corresponding pixel of a long-term background model assigned to the image excerpt, and wherein it is provided, in particular, that a pixel of the image which differs from the corresponding pixel of the short-term background model and from the corresponding pixel of the

-4-

long-term background model is classified as a foreground pixel, wherein a plurality of (adjacent) foreground pixels are advantageously assigned to a region.

[0011] An image excerpt within the meaning of the invention is, in particular, the area which is captured by the video camera. An image excerpt within the meaning of the invention is, in particular, that part of the surroundings of the video camera which is imaged by means of the image.

[0012] A pixel within the meaning of the invention is, in particular, one pixel. However, a pixel within the meaning of the invention can also comprise or be a group of pixels. Such a group of pixels can be, for example, an area of a picture unit.

[0013] A background model can be, for example, a *background model* in accordance with US 5 684 898. Background models can be generated for example in accordance with the methods described in the article by Stauffer, Chris and Grimson, W.E.L.: *Adaptive background mixture models for real-time tracking*; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 1999 wherein any model that can model a multimodal density distribution (cf., for example, Zivkovic: *Improved adaptive Gaussian mixture model for background subtraction*; Proceedings of the International Conference on Pattern Recognition; 2004) can be used. A background model within the meaning of the invention is, in particular, a model of the statistical components of the image which is recorded by the video camera. A short-term background model within the meaning of the invention includes, in particular, pixel values which have a relative statistical relevance with respect to other values observed in the same pixel during a first time interval (values in this sense are colors, in particular). A long-term background model within the meaning of the invention includes, in particular, pixels which have a relative statistical relevance with respect to other values observed in the same pixel during a second time interval. A second time interval within the meaning of the invention contains the first time interval and/or is, in particular depending on the temporal resolution of the video sequence and on the desired sensitivity of the system, longer than a first time interval within the meaning of the invention. A short-term background model within the meaning of the invention is calculated, in particular, from

-5-

the statistical evaluation of a video sequence of a first time window. A long-term background model within the meaning of the invention is calculated, in particular, from the statistical evaluation of a video sequence (= temporal succession of successive images) of a second time window. A second time window within the meaning of the invention is, in particular depending on the temporal resolution of the video sequence and on the desired sensitivity of the system, larger than a first time window within the meaning of the invention.

[0014] It can be provided, in particular, that a predetermined number of modes (colors) is or has been assigned to a pixel of a background model. Thus, by way of example, five different colors and the frequency thereof can be assigned to a pixel of a background model. Within the meaning of the invention, a pixel of an image differs from a corresponding pixel of a background model in particular when the color of the pixel of the image corresponds to none of the predetermined number of colors (modes) of the background model.

[0015] Two pixels are designated as corresponding within the meaning of the invention in particular when they have the same coordinates or lie at the same location. A comparison of background models within the meaning of the invention optionally also encompasses a comparison of variables derived from or dependent on the background models, such as e.g. of foreground masks.

[0016] A region within the meaning of the invention is, in particular, a static region. A region within the meaning of the invention comprises, in particular, adjacent/contiguous pixels having identical features or properties. In this case, a region within the meaning of the invention is regarded as finished in particular when its growth has ended. That is to say, in particular, that a plurality of temporally offset images/frames are employed/used for assessing/defining/ determining a region.

[0017] In a furthermore advantageous configuration of the invention, depending on an area – corresponding to the region – of a subsequent image of the image excerpt, a removed object is recognized or it is determined whether the removal of an object has been recognized. In this case, it is provided, in particular, that the removal of a (static)

-6-

object is regarded as recognized or a removed (static) object is recognized or it is determined whether the removal of an object has been recognized if the pixels of the image within that area of the subsequent image which corresponds to the region can be or are assigned to a single region. An area corresponding to a region within the meaning of the invention is, for example, the contour of the region or a so-called bounding box assigned to the region. In this case, a bounding box is the smallest rectangle that covers the region.

[0018] It may be provided that “subsequent” within the meaning of the invention means “later” and not “directly following”.

[0019] In a furthermore advantageous configuration of the invention, depending on an area – corresponding to the region – of the long-term background model, a removed object is recognized or it is determined whether the removal of an object has been recognized. In this case, it is provided, in particular, that the removal of an object is recognized or determined or regarded as recognized if the corresponding region of the long-term background model correspond to the foreground pixels of the region. Correspondence in this sense can mean that they have or define or establish identical features, such as e.g. identical edges.

[0020] In a furthermore advantageous configuration of the invention, upon recognition of the removal of an object, that area of the long-term background model which corresponds to the region is replaced by that area of the short-term background model which corresponds to the region.

[0021] In a furthermore advantageous configuration of the invention, depending on the/an area - corresponding to the region – of the/a subsequent image of the image excerpt, an added object is recognized or it is determined whether the addition of an object has been recognized, wherein it is provided, in particular, that the addition of a (static) object is recognized or defined as recognized if the pixels of the area – corresponding to the region – of the/a subsequent image of the image excerpt correspond to the foreground pixels of the region. Correspondence in this sense can mean that the regions have or define or establish identical features, such as e.g. identical edges. It is

-7-

provided, in particular, that, upon the determination “addition of a (static) object”, an alarm, a message or a hazard warning message is generated or output. This can be done optically and/or acoustically, for example.

[0022] In a furthermore advantageous configuration of the invention, if a pixel differs from the corresponding pixel of the short-term background model, but not from the corresponding pixel of the long-term background model, the corresponding pixel of the short-term background model is replaced by the corresponding pixel of the long-term background model. In a furthermore advantageous configuration of the invention, if a pixel differs from the corresponding pixel of the short-term background model, but not from the corresponding pixel of the long-term background model, the short-term background model is replaced by the long-term background model.

[0023] In a furthermore advantageous configuration of the invention, if a pixel differs from the corresponding pixel of the long-term background model, but not from the corresponding pixel of the short-term background model, the corresponding pixel of the long-term background model is replaced by the corresponding pixel of the short-term background model, but in a manner reduced by the color of the pixel. In a furthermore advantageous configuration of the invention, this is provided only a single time, however, for a pixel. In a furthermore advantageous configuration of the invention, if a pixel differs from the corresponding pixel of the long-term background model, but not from the corresponding pixel of the short-term background model, the long-term background model is replaced by the short-term background model, but in a manner reduced by the color of the pixel. In a furthermore advantageous configuration of the invention, this is provided only a single time, however, for a pixel.

[0024] The abovementioned object is achieved – in particular in conjunction with features mentioned above – in addition by means of a device for video surveillance, in particular for carrying out a method mentioned above, wherein the device for video surveillance comprises at least one video camera for recording an image of an image excerpt of an environment to be monitored in the vicinity of the video camera, a short-term background model assigned to the image excerpt, a long-term background

-8-

model assigned to the image excerpt, and also an evaluation device for comparing at least one pixel of the image with a corresponding pixel of the short-term background model and with a corresponding pixel of the long-term background model.

[0025] Further advantages and details will become apparent from the following description of exemplary embodiments. In this case, in the figures:

BRIEF DESCRIPTION OF THE DRAWINGS

[0026] Figure 1 shows an exemplary embodiment of a device for video surveillance;

[0027] Figure 2 shows an exemplary embodiment of an evaluation device with evaluation modules;

[0028] Figure 3 shows an exemplary embodiment of a finite state machine;

[0029] Figure 4 shows an exemplary embodiment of a method implemented in a (low-level) evaluation module in accordance with figure 2;

[0030] Figure 5 shows an exemplary embodiment of a method implemented in a (high-level) evaluation module in accordance with figure 2;

[0031] Figure 6 shows an exemplary embodiment of a current image (frame, input frame);

[0032] Figure 7 shows an exemplary embodiment of a region formed from preceding images corresponding to the image section in accordance with figure 6, or a corresponding mask;

[0033] Figure 8 shows an exemplary embodiment of a corresponding long-term background model in the area corresponding to the (analyzed) region in accordance with figure 7;

[0034] Figure 9 shows an exemplary embodiment of recognized edges in the current image (frame, input frame) in accordance with figure 6;

[0035] Figure 10 shows an exemplary embodiment of recognized edges of the region or mask in accordance with figure 7;

-9-

[0036] Figure 11 shows an exemplary embodiment of recognized edges in the corresponding area in accordance with figure 8;

[0037] Figure 12 shows an exemplary embodiment of a current image (frame, input frame);

[0038] Figure 13 shows an exemplary embodiment of a formed from preceding images corresponding to the image section in accordance with figure 12;

[0039] Figure 14 shows an exemplary embodiment of a corresponding long-term background model in the area corresponding to the (analyzed) region in accordance with figure 13;

[0040] Figure 15 shows an exemplary embodiment of recognized edges in the current image (frame, input frame) in accordance with figure 12;

[0041] Figure 16 shows an exemplary embodiment of recognized edges of the region in accordance with figure 13; and

[0042] Figure 17 shows an exemplary embodiment of recognized edges in the corresponding area in accordance with figure 14.

DETAILED DESCRIPTION

[0043] Figure 1 shows an exemplary embodiment of a device 100 for video surveillance, comprising a video camera 101 for recording an image VIDEO of an image excerpt in an environment to be monitored in the vicinity of the video camera 101. The image VIDEO is analyzed by means of an evaluation device 102 in order to recognize static objects such as, for example, bags or suitcases left at an airport or station. If the evaluation device 102 recognizes a static object in the image VIDEO, then it outputs a corresponding message ALARM to an output device 103.

[0044] The evaluation device 102 comprises – as illustrated in figure 2 – a model updating module 121 for updating or generating a short-term background model 122 and a long-term background model 123. The short-term background model 122 and the long-term background model 123 are updated in different time intervals (complementary

-10-

background modeling). A (low-level) evaluation model 124 and a (high-level) evaluation module 125 are additionally provided.

[0045] Figure 3 describes the functioning of the (low-level) evaluation module 124 by means of a finite state machine (FSM). Said state machine is distinguished by the fact that it permits only four states:

[0046] BG: a pixel belongs to the background.

[0047] MO: a pixel belongs to a moving object.

[0048] ST: a pixel belongs to an added (static) object.

[0049] UBG: a pixel belongs to a removed object, that is to say a now visible background (uncovered background).

[0050] The finite state machine (FSM) is defined as a 5-tuple $(I, Q, Z, \delta, \omega)$,

[0051] where I is the possible combinations of the results of a background subtraction,

[0052] where Q denotes the set of states (BG (background), MO (moving), ST (new static) and UBG (uncovered background)) which a pixel (or picture unit) can run through,

[0053] where Z is the set of numbers which is the classification of the pixels (or picture units) ($Z \equiv \{0, 1, \dots | Q\}$),

[0054] where δ is the next/following state in accordance with figure 3,

[0055] and where ω is the output function with output values $Z \in \{0, 1, \dots | Q\}$

in accordance with the state of a pixel (or picture unit).

[0056] Figure 4 shows a method implemented in the evaluation module 124. In this case, reference symbol 21 designates the interrogation as to whether the pixel of an image (frame) is identical to the corresponding pixel of the short-term background model 122. Reference symbols 22 and 23 respectively designate the interrogation as to whether

-11-

the pixel of an image (frame) is identical to the corresponding pixel of the long-term background model 123. If the interrogations 21 and 23 reveal that the pixel differs from the corresponding pixel of the short-term background model, but not from the corresponding pixel of the long-term background model 123, then the interrogation 23 is followed by a step 24, in which the corresponding pixel of the short-term background model 122 is replaced by the corresponding pixel of the long-term background model 123.

[0057] If the interrogations 21 and 22 reveal that the pixel differs from corresponding pixels of the long-term background model 123, but not from the corresponding pixel of the short-term background model 122, then the interrogation 22 is followed by a step 26, in which the corresponding pixel of the long-term background model 123 is replaced by the corresponding pixel of the short-term background model 122, but in a manner reduced by the color of the pixel of the image. If a color/mode is removed, its weight ω_j (statistical frequency) is distributed among the other colors or modes, that is to say that:

[0058]
$$\omega_k = \omega_k + \omega_j / (B - 1), \quad \forall k = 1 \dots B, k \neq j,$$

[0059] where $j \leq B$ denotes the removed color or the removed mode and $B > 1$ denotes the number of colors or modes. B can be 5, for example. Step 26 is provided only a “first time” or once in particular for the corresponding pixel/partial image (that is to say in the first transition from MO to ST from figure 3). In this case, the pixel is marked as a static foreground until it jumps further either to BG or UBG.

[0060] If the interrogations 21 and 23 reveal that the pixel differs from the corresponding pixel of the long-term background model 123 and from the corresponding pixel of the short-term background model 122, then said pixel is classified as foreground or as foreground pixel (step 25). A corresponding notification is effected to the evaluation module 125, in which the method described below with reference to figure 5 is implemented.

[0061] The method implemented in the evaluation module 125 begins with a step 31, in which a region is formed from adjacent/contiguous static foreground pixels (which

-12-

were marked in step 26). (Optionally, only the pixels marked as static foreground pixels over a given time are taken.) Step 31 is followed by an interrogation 32, which involves interrogating whether the determined region grows with respect to a corresponding region at the previous point in time or one or more of the previous points in time. If the growth has ended, the region (also called mask hereinafter) is therefore finished, and so the interrogation 32 is followed by an analysis of the region in a step 33. In this case, the position and size of a bounding box at the point in time of its occurrence and also its disappearance are stored for each new static region.

[0062] An interrogation 34 ensues, which involves interrogating whether a moving object is moving through the region. If no moving object is moving through the region, then the interrogation 34 is followed by an interrogation 35. The interrogation 35 involves interrogating whether the features of a current image in the area corresponding to the bounding box are part of a single region. If this is the case, then the interrogation 35 is followed by a step 36, in which it is assumed that the region belongs to a removed object, such that now the background (empty scenery) is visible with respect to the region. In addition, that area of the long-term background model 123 which corresponds to the region is replaced by that area of the short-term background model 122 which corresponds to the region.

[0063] If the features of a current image in the area corresponding to the bounding box are not part of a single region, then the interrogation 35 is followed by an interrogation 37, which involves interrogating whether (the) features in that area of the long-term background model 123 which corresponds to the region correspond to the corresponding features of the region or mask. If (the) features in that area of the long-term background model 123 which corresponds to the region correspond to the corresponding features of the region or mask, then the interrogation 37 is followed by the step 36.

[0064] If (the) features in that area of the long-term background model 123 which corresponds to the region do not correspond to the corresponding features of the region or mask, then the interrogation 37 is followed by an interrogation 38, which involves

-13-

interrogating whether (the) features in that area of the (current) image/frame which corresponds to the region correspond to the corresponding features of the region or mask. If (the) features in that area of the (current) image/frame which corresponds to the region correspond to the corresponding features of the region or mask, then the interrogation 38 is followed by a step 39, in which the new static region is assigned to an added static object. In addition, the message ALARM is output.

[0065] The functioning of the evaluation device 102 is explained below on the basis of the examples illustrated in figure 6 to figure 17. In this case, figure 6 shows a current image (frame, input frame) and figure 7 shows a region formed from preceding images corresponding to the image section in accordance with figure 6, or a corresponding mask. Figure 8 shows the corresponding long-term background model 123 in the analyzed region. Figures 9, 10 and 11 show the edges corresponding to figures 6, 7 and 8. In this case, the frame illustrated in a dotted fashion in figure 5 is not part of the long-term background model. The edges of the image (see figure 9) correspond to the edges of the region (see figure 10), such that that part of the image which corresponds to the region (see figure 6) is classified as a new static object.

[0066] Figures 12 to 17 describe a further example, wherein figure 12 shows a current image (frame, input frame) and figure 13 shows a region formed from preceding images corresponding to the image section in accordance with figure 12, or a corresponding mask. Figure 14 shows the corresponding long-term background model 123 in the analyzed region. Figures 15, 16 and 17 show the edges corresponding to figures 12, 13 and 14. The edges of the long-term background model 123 in the analyzed region (see figure 17) correspond to the edges of the region (see figure 16), such that that part of the image which corresponds to the region (see figure 12) is classified as a removed static object.

CLAIMS

1. Method for video surveillance, wherein, by means of at least one video camera, an image of an image excerpt of an environment to be monitored in the vicinity of the video camera is recorded, wherein at least one pixel of the image is compared with a corresponding pixel of a short-term background model assigned to the image excerpt and with a corresponding pixel of a long-term background model assigned to the image excerpt.

2. Method according to claim 1, wherein a pixel of the image which differs from the corresponding pixel of the short-term background model and from the corresponding pixel of the long-term background model is classified as a foreground pixel.

3. Method according to claim 2, wherein a plurality of (adjacent) foreground pixels are assigned to a region.

4. Method according to claim 3, wherein, depending on an area – corresponding to the region – of a subsequent image of the image excerpt, it is determined whether the removal of an object has been recognized.

5. Method according to claim 3, wherein, depending on an area – corresponding to the region – of the long-term background model, it is determined whether the removal of an object has been recognized.

6. Method according to claim 4, wherein, upon recognition of the removal of an object, that area of the long-term background model which corresponds to the region is replaced by that area of the short-term background model which corresponds to the region.

7. Method according to claim 3, wherein, depending on the area – corresponding to the region – of a subsequent image of the image excerpt, it is determined whether the addition of an object has been recognized.

8. Method according to claim 1, wherein, if a pixel differs from the corresponding pixel of the short-term background model, but not from the corresponding

-15-

pixel of the long-term background model, the corresponding pixel of the short-term background model is replaced by the corresponding pixel of the long-term background model.

9. Method according to claim 1, wherein, if a pixel differs from the corresponding pixel of the long-term background model, but not from the corresponding pixel of the short-term background model, the corresponding pixel of the long-term background model is replaced by the corresponding pixel of the short-term background model, but in a manner reduced by the color of the pixel.

10. Device for video surveillance, wherein the device for video surveillance comprises at least one video camera for recording an image of an image excerpt of an environment to be monitored in the vicinity of the video camera, a short-term background model assigned to the image excerpt, a long-term background model assigned to the image excerpt, and also an evaluation device for comparing at least one pixel of the image with a corresponding pixel of the short-term background model and with a corresponding pixel of the long-term background model.

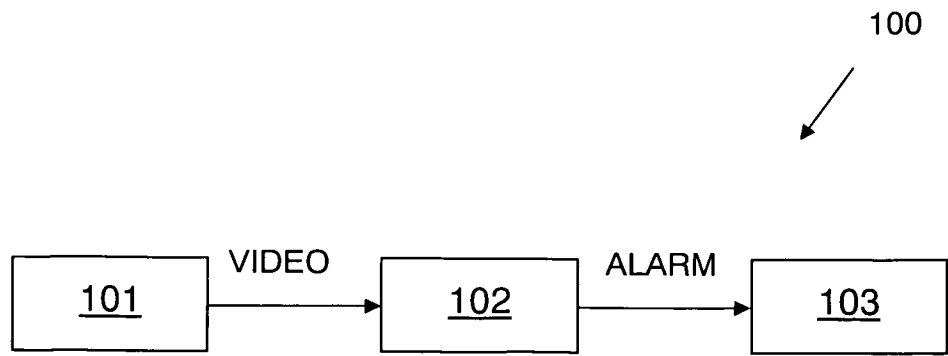


Fig. 1

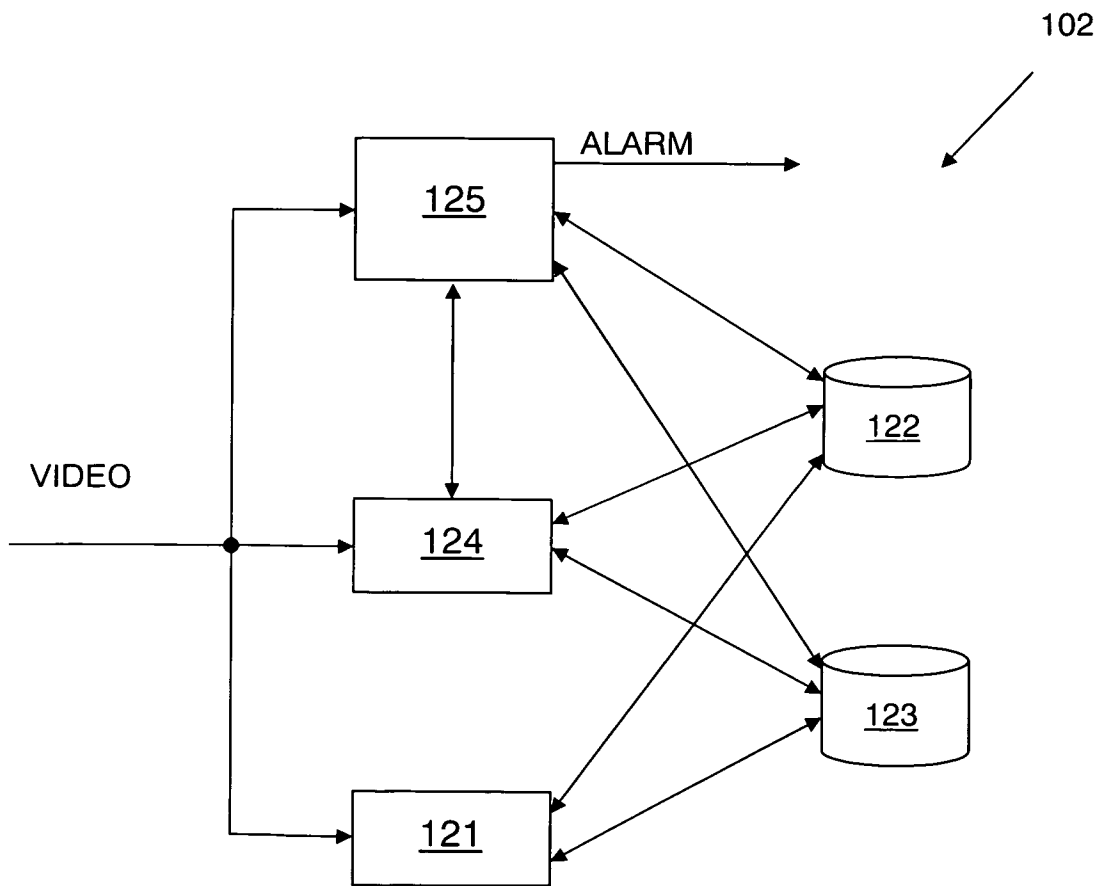


Fig. 2

2/5

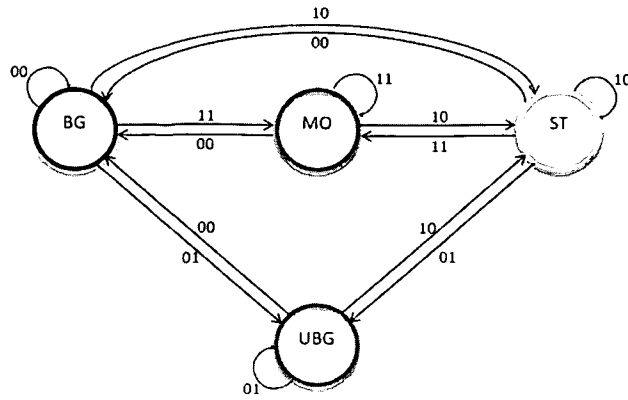


Fig. 3

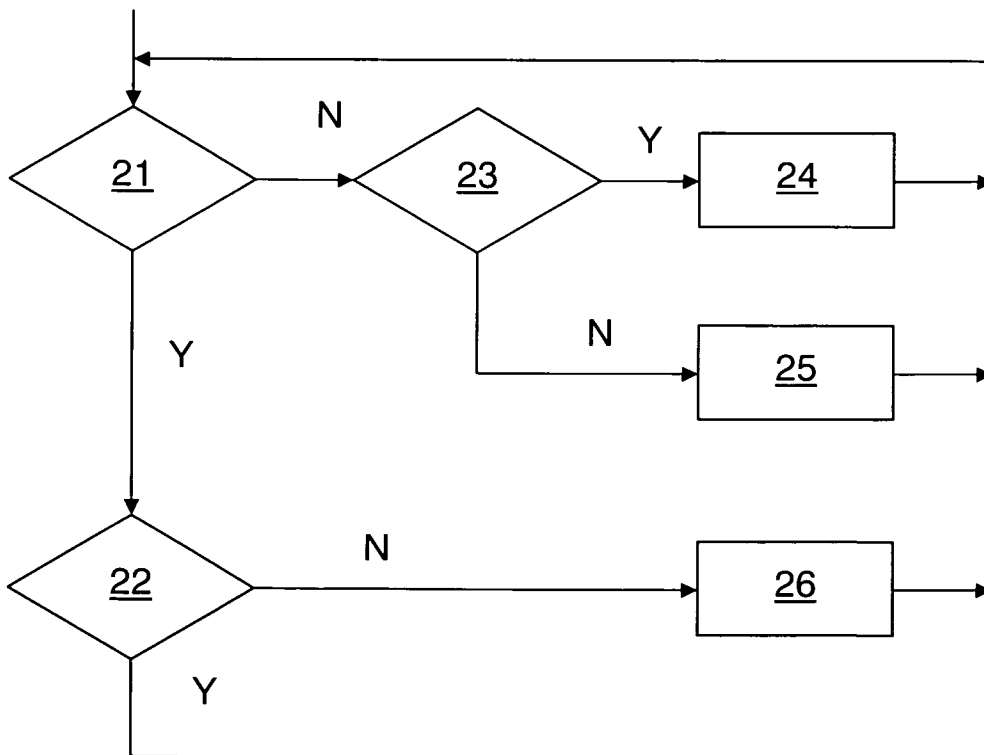


Fig. 4

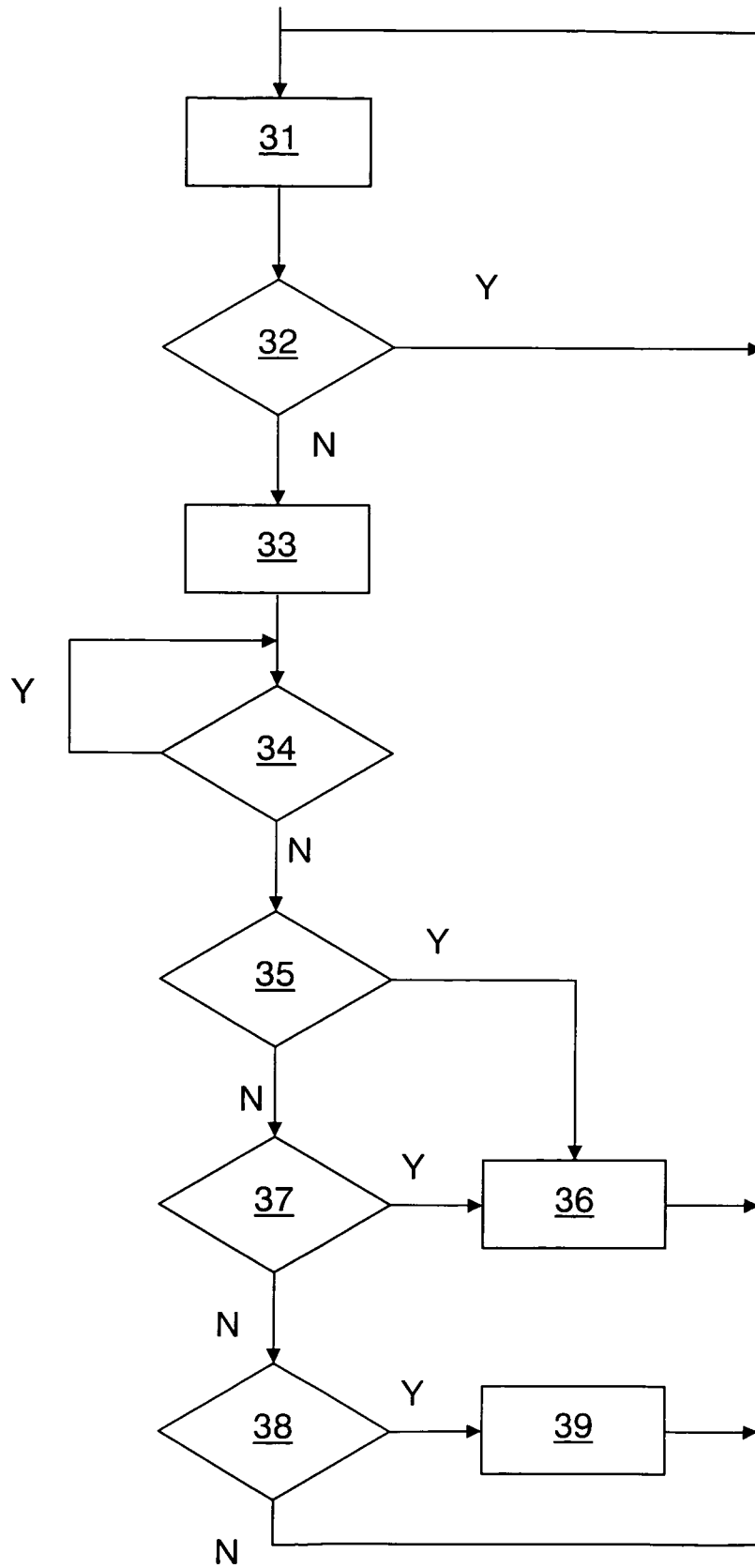


Fig. 5

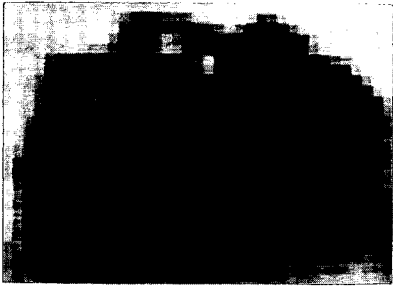


Fig. 6



Fig. 7

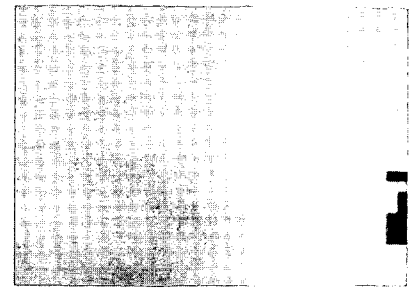


Fig. 8



Fig. 9



Fig. 10

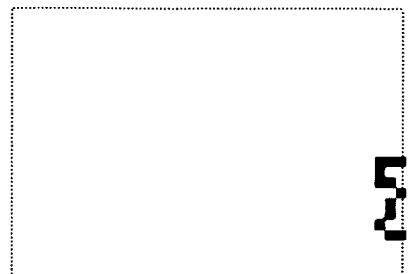


Fig. 11

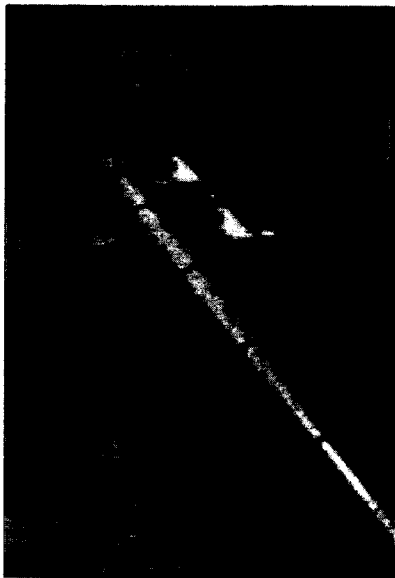


Fig. 12



Fig. 13



Fig. 14

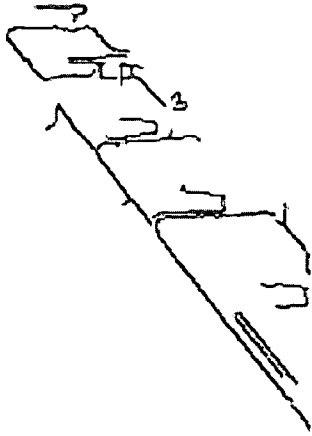


Fig. 15



Fig. 16



Fig. 17

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2012/002661

A. CLASSIFICATION OF SUBJECT MATTER
 INV. G06T7/20 H04N7/18 G06K9/00
 ADD.
 According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
 Minimum documentation searched (classification system followed by classification symbols)
 G06T H04N G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
 EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	RUBEN HERAS EVANGELIO ET AL: "Detection of static objects for the task of video surveillance", APPLICATIONS OF COMPUTER VISION (WACV), 2011 IEEE WORKSHOP ON, IEEE, 5 January 2011 (2011-01-05), pages 534-540, XP031913620, DOI: 10.1109/WACV.2011.5711550 ISBN: 978-1-4244-9496-5 cited in the application the whole document	1-10
A	EP 1 978 470 A1 (MITSUBISHI ELECTRIC CORP [JP]) 8 October 2008 (2008-10-08) abstract paragraph [0016] paragraph [0023] - paragraph [0028] figures 1,2	1-10
	----- -/--	

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E" earlier application or patent but published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 28 August 2012	Date of mailing of the international search report 06/09/2012
---	--

Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Sery, Franck
--	--

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2012/002661

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2007/274402 A1 (IBRAHIM M MOHAMAD [IN] ET AL) 29 November 2007 (2007-11-29) abstract paragraph [0002] - paragraph [0012] -----	1-10
A	US 2009/087086 A1 (EATON JOHN ERIC [US] ET AL) 2 April 2009 (2009-04-02) abstract figure 2 paragraph [0043] - paragraph [0045] -----	1-10

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2012/002661

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
EP 1978470	A1	08-10-2008	CN 101281596 A	08-10-2008
			EP 1978470 A1	08-10-2008
			JP 2008257693 A	23-10-2008
			US 2008247599 A1	09-10-2008

US 2007274402	A1	29-11-2007	NONE	

US 2009087086	A1	02-04-2009	US 2009087086 A1	02-04-2009
			US 2009087093 A1	02-04-2009
			US 2009087096 A1	02-04-2009
