US 20180322750A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2018/0322750 A1**

Venetianer et al. (43) **Pub. Date: Nov. 8, 2018**

(54) **VIDEO SURVEILLANCE SYSTEM EMPLOYING VIDEO PRIMITIVES**

(71) Applicant: **AVIGILON FORTRESS CORPORATION**, Vancouver (CA)

(72) Inventors: **Peter L. Venetianer**, McLean, VA (US); **Alan J. Lipton**, Austin, TX (US); **Yongtong Hu**, Herndon, VA (US); **Andrew J. Martone**, San Francisco, CA (US); **Weihong Yin**, Herndon, VA (US); **Li Yu**, Sterling, VA (US); **Zhong Zhang**, Great Falls, VA (US)

(21) Appl. No.: **16/035,942**

(22) Filed: **Jul. 16, 2018**

Related U.S. Application Data

(63) Continuation of application No. 15/044,902, filed on Feb. 16, 2016, now Pat. No. 10,026,285, which is a continuation of application No. 14/203,065, filed on Mar. 10, 2014, now Pat. No. 9,378,632, which is a continuation of application No. 11/300,581, filed on Dec. 15, 2005, now Pat. No. 8,711,217, which is a continuation-in-part of application No. 11/057,154, filed on Feb. 15, 2005, now abandoned, which is a continuation-in-part of application No. 09/987,707, filed on Nov. 15, 2001, now abandoned, which is a continuation-in-part of application No. 09/694,712, filed on Oct. 24, 2000, now Pat. No. 6,954,498.

Publication Classification

(51) **Int. Cl.**

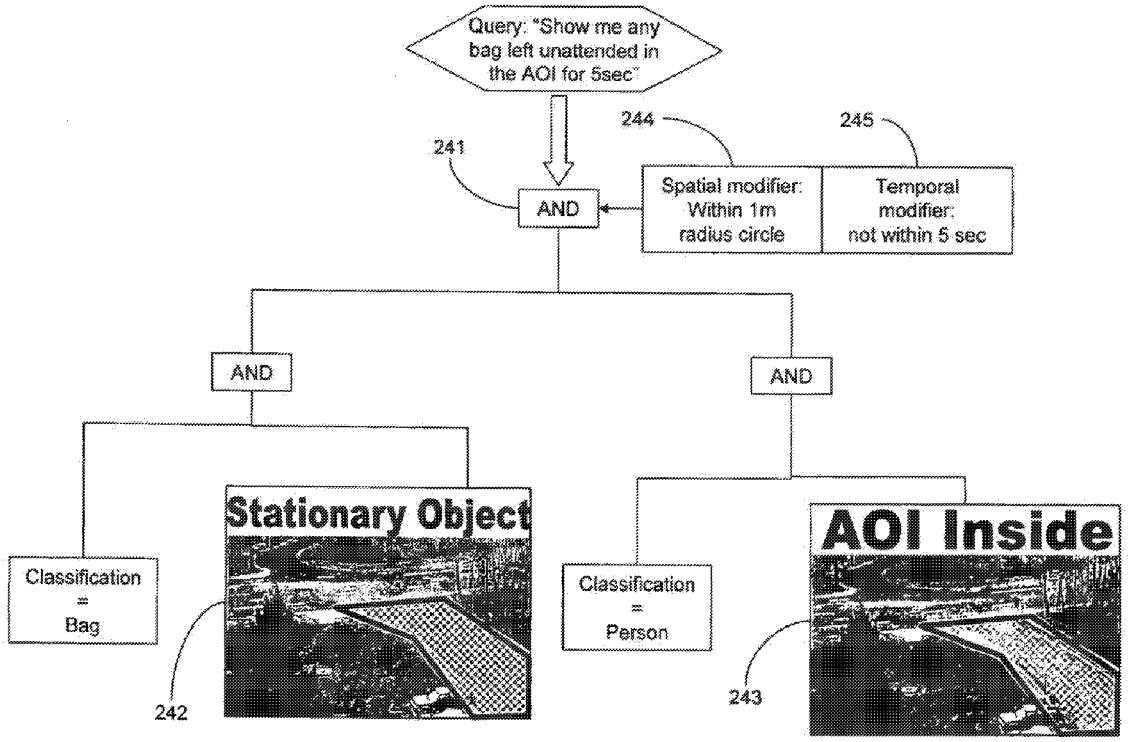
G08B 13/196	(2006.01)
H04N 7/18	(2006.01)
G08B 31/00	(2006.01)
G08B 29/18	(2006.01)
G08B 21/04	(2006.01)
G06F 17/30	(2006.01)
G06K 9/00	(2006.01)

(52) **U.S. Cl.**

CPC ***G08B 13/19615*** (2013.01); ***H04N 7/18*** (2013.01); ***G08B 31/00*** (2013.01); ***G08B 29/188*** (2013.01); ***G08B 21/0476*** (2013.01); ***G08B 13/19697*** (2013.01); *G08B 13/196* (2013.01); ***G06F 17/3079*** (2013.01); ***G06K 9/00778*** (2013.01); ***G06K 9/00771*** (2013.01); ***G06F 17/30811*** (2013.01); ***G06F 17/30805*** (2013.01); ***G08B 13/19604*** (2013.01); ***G06F 17/30799*** (2013.01)

(57) **ABSTRACT**

A video surveillance system is set up, calibrated, tasked, and operated. The system extracts video primitives and extracts event occurrences from the video primitives using event discriminators. The system can undertake a response, such as an alarm, based on extracted event occurrences.



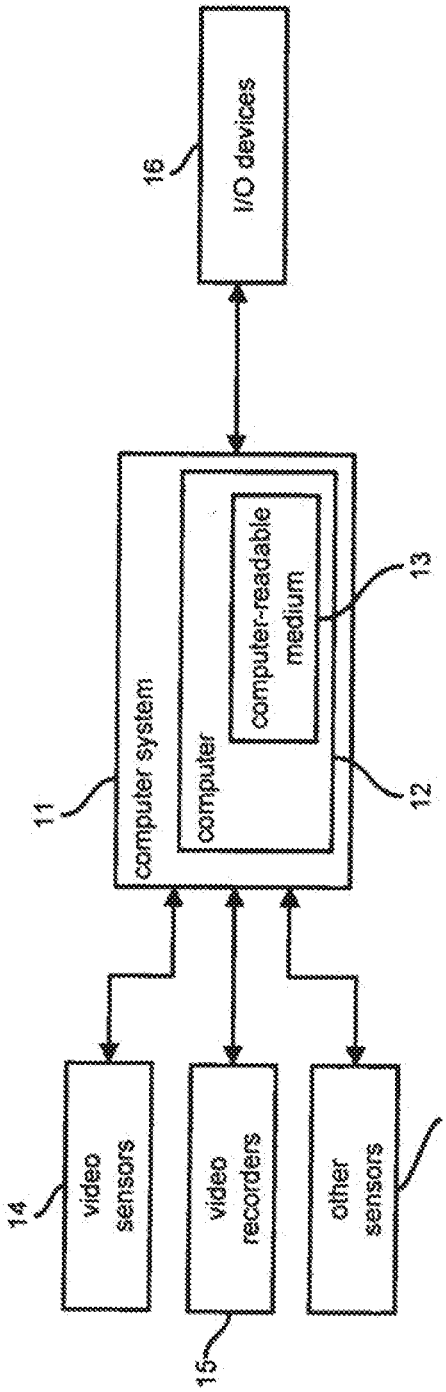

16
I/O devices
11
computer system
computer
computer-readable medium
13
12
14
video sensors
15
video recorders
other sensors
17

FIG. 1

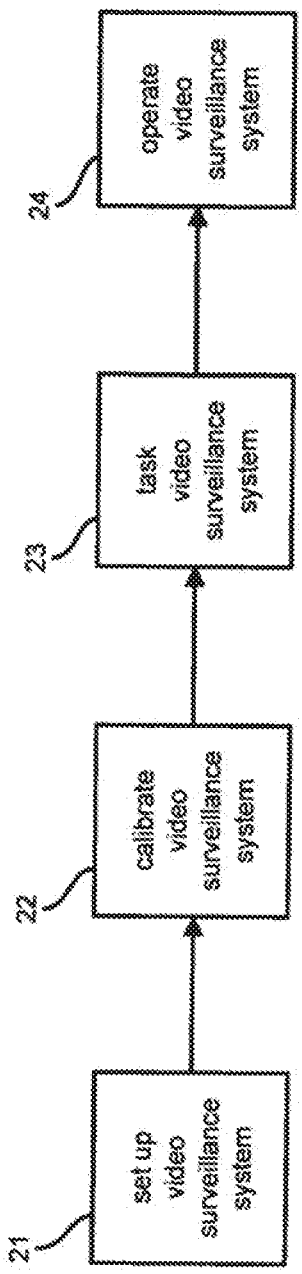

24
operate video surveillance system
23
task video surveillance system
22
calibrate video surveillance system
21
set up video surveillance system

FIG. 2

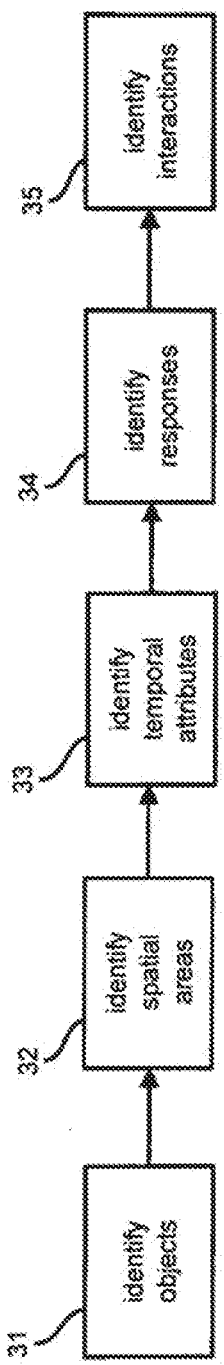


FIG. 3

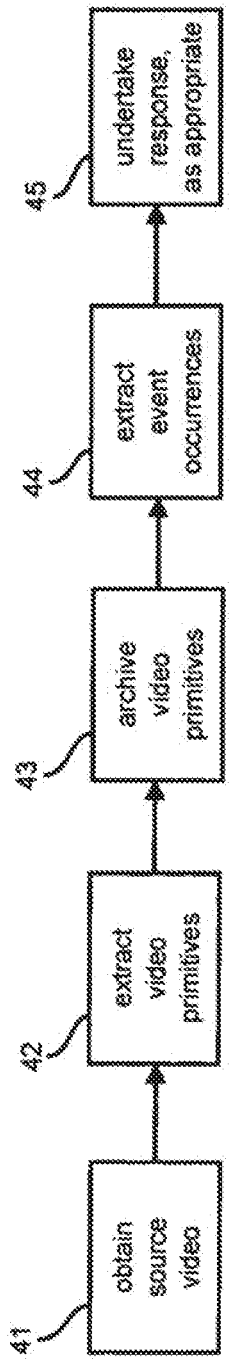


FIG. 4

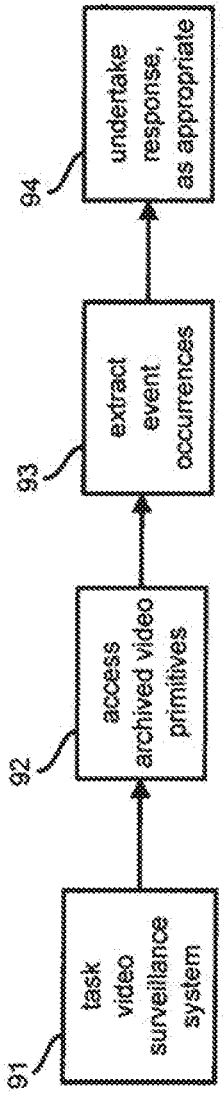


FIG. 9

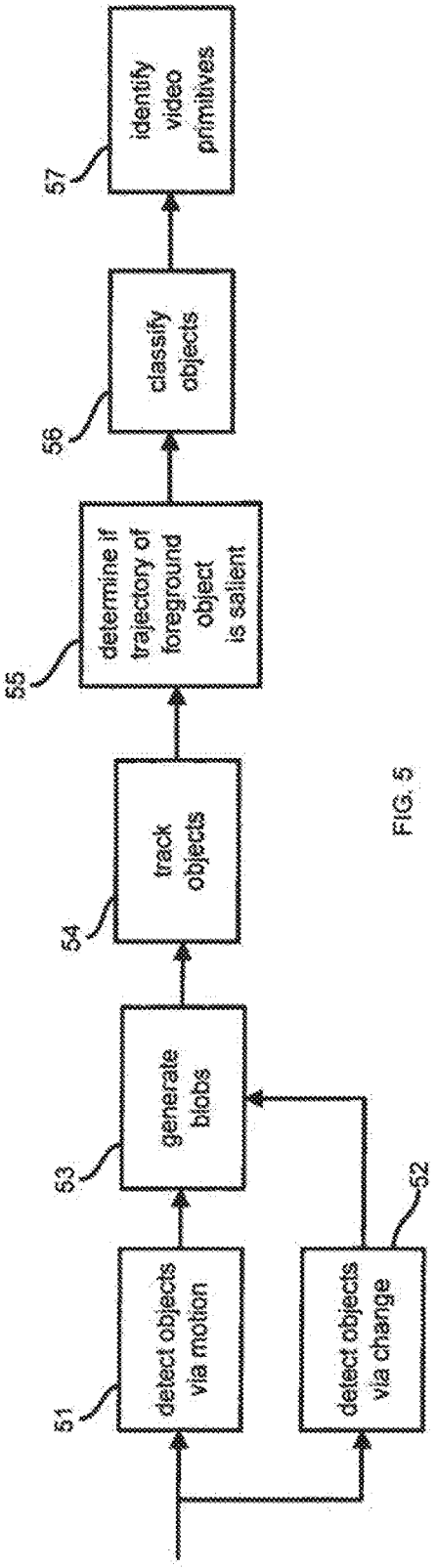

detect objects via motion
51
detect objects via change
52
generate blobs
53
track objects
54
determine if trajectory of foreground object is salient
55
classify objects
56
identify video primitives
57

FIG. 5

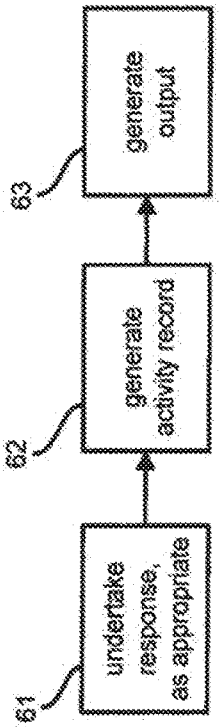

undertake response, as appropriate
61
generate activity record
62
generate output
63

FIG. 6

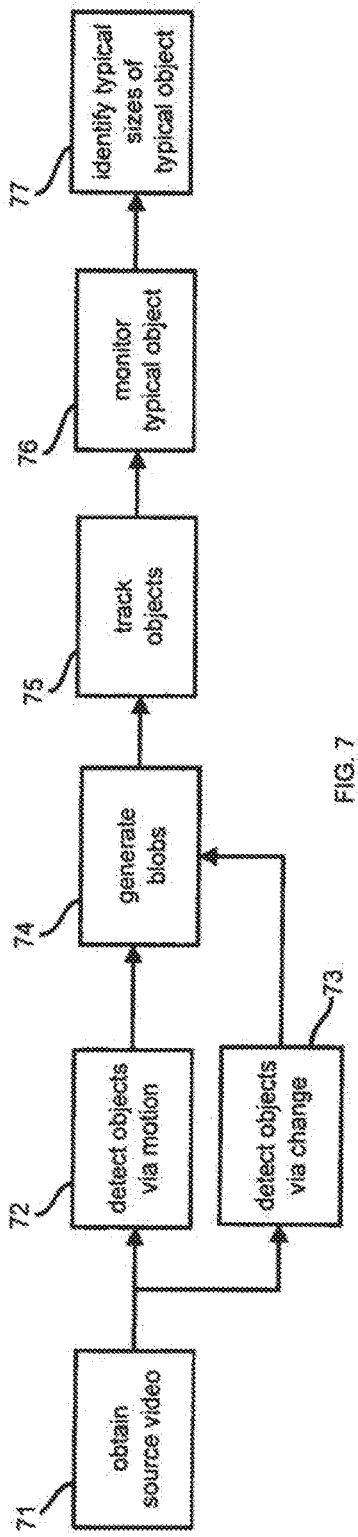

71
obtain source video
72
detect objects via motion
73
detect objects via change
74
generate blobs
75
track objects
76
monitor typical object
77
identify typical sizes of typical object

FIG. 7

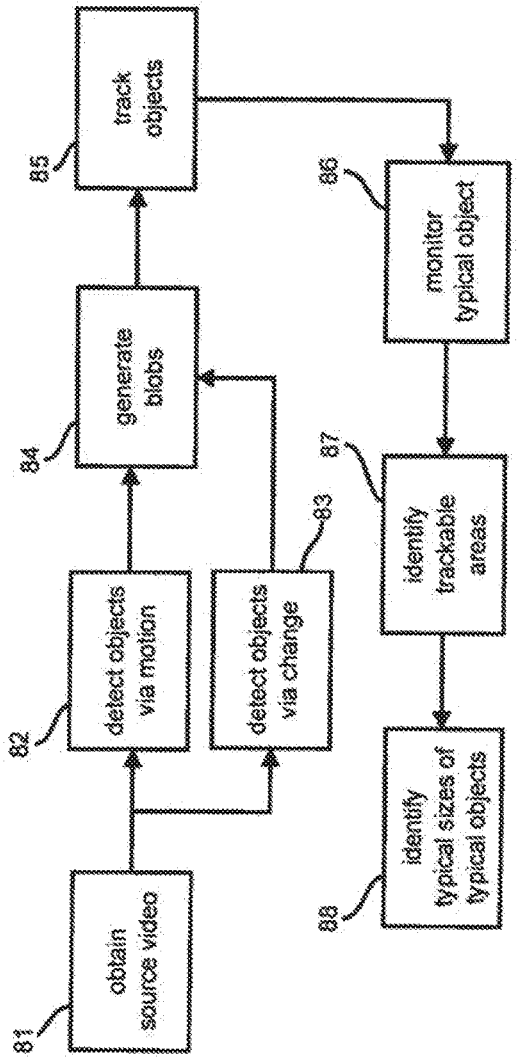

81
obtain source video
82
detect objects via motion
83
detect objects via change
84
generate blobs
85
track objects
86
monitor typical object
87
identify trackable areas
88
identify typical sizes of typical objects

FIG. 8

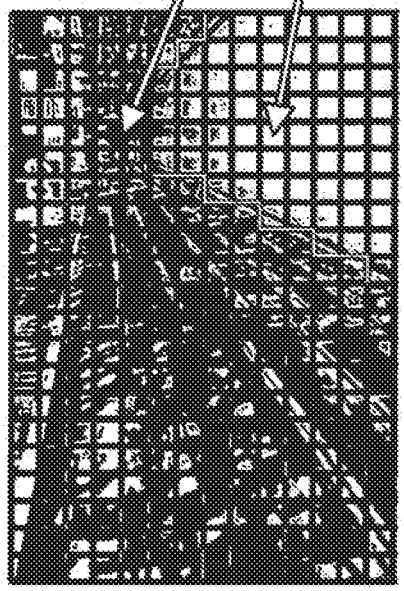

Can't get reliable tracks in this region
This region is fair game for good tracks

FIG. 10

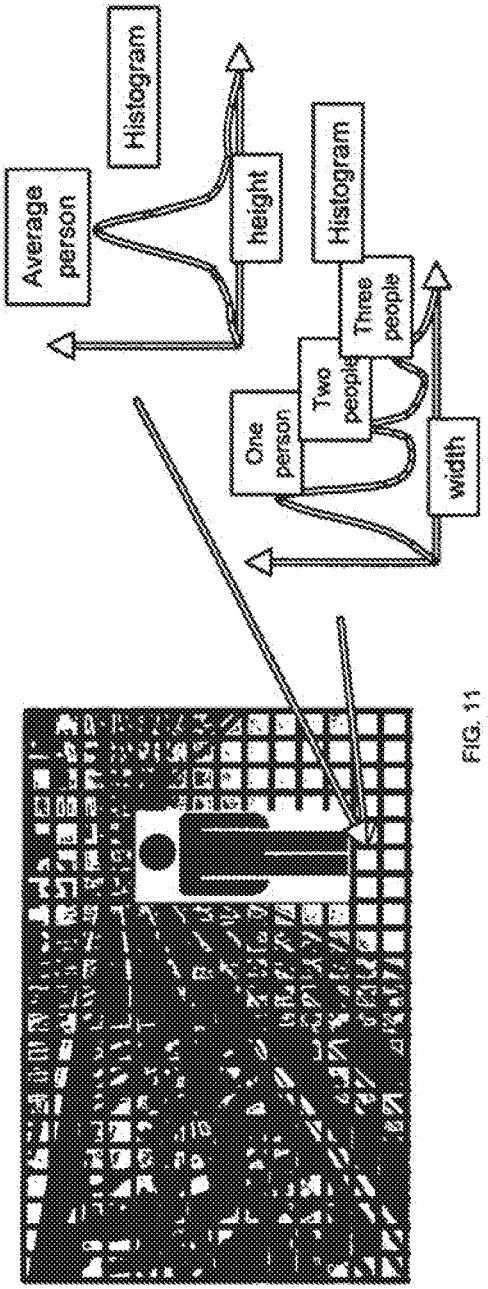

Histogram
Average person
height
Histogram
Three people
Two people
One person
width

FIG. 11

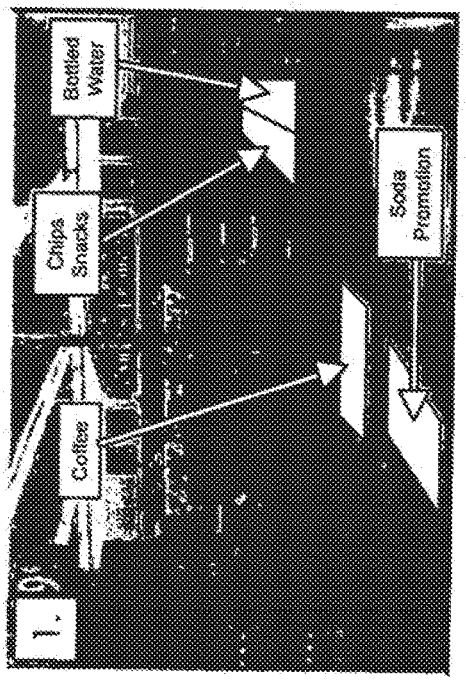


FIG. 12

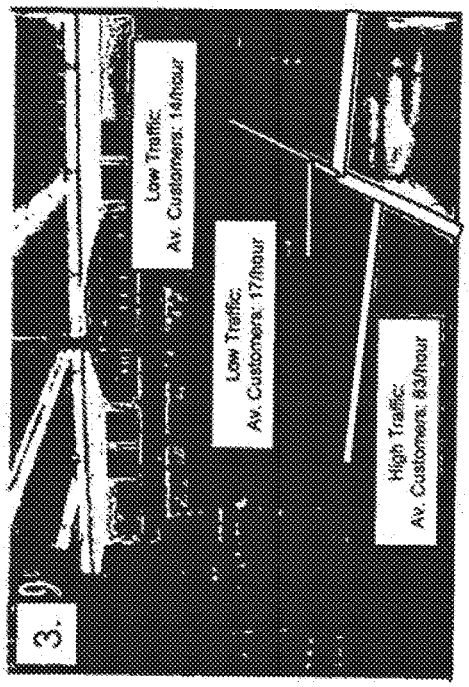


FIG. 14

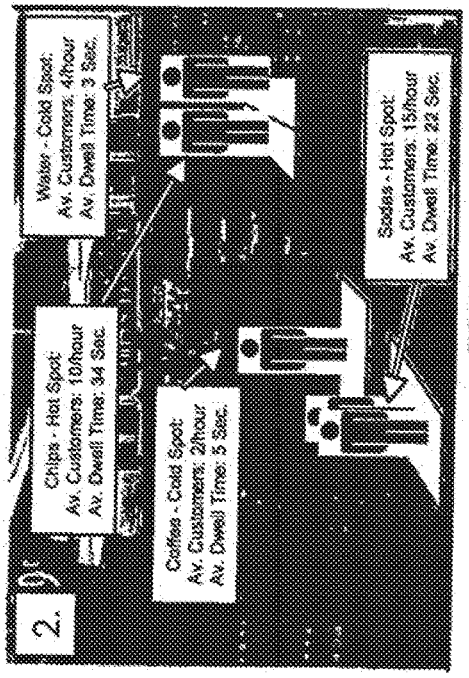


FIG. 13

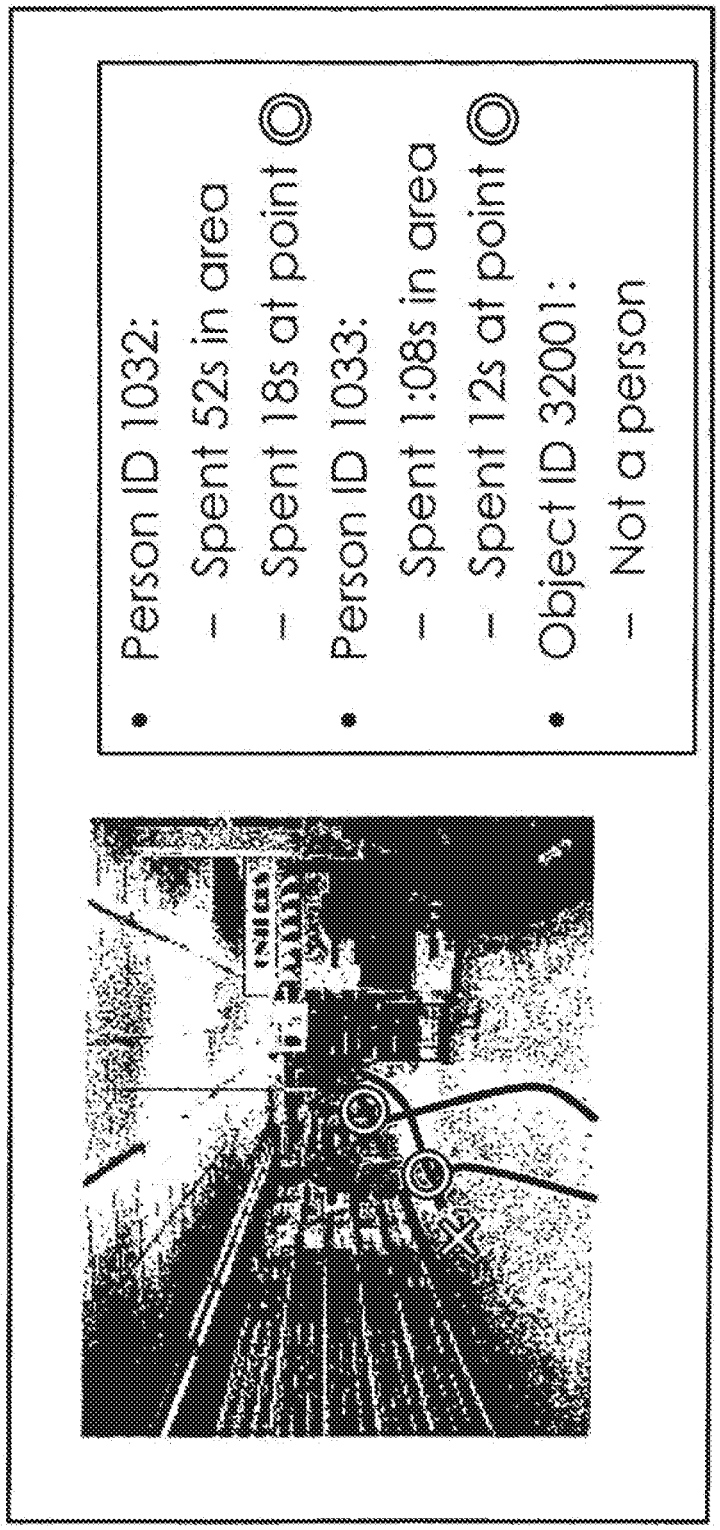

• Person ID 1032:
– Spent 52s in area
– Spent 18s at point ◎
• Person ID 1033:
– Spent 1:08s in area
– Spent 12s at point ◎
• Object ID 32001:
– Not a person

FIG. 15

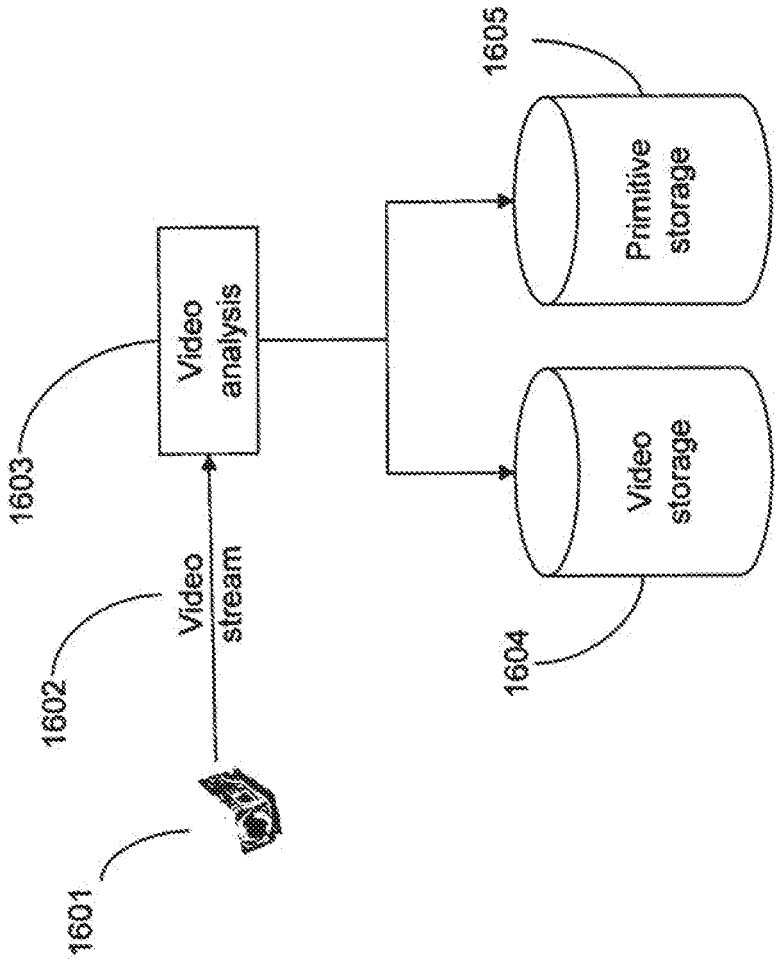

1605
Primitive storage
Video analysis
1603
Video stream
1602
Video storage
1604
1601

Figure 16a

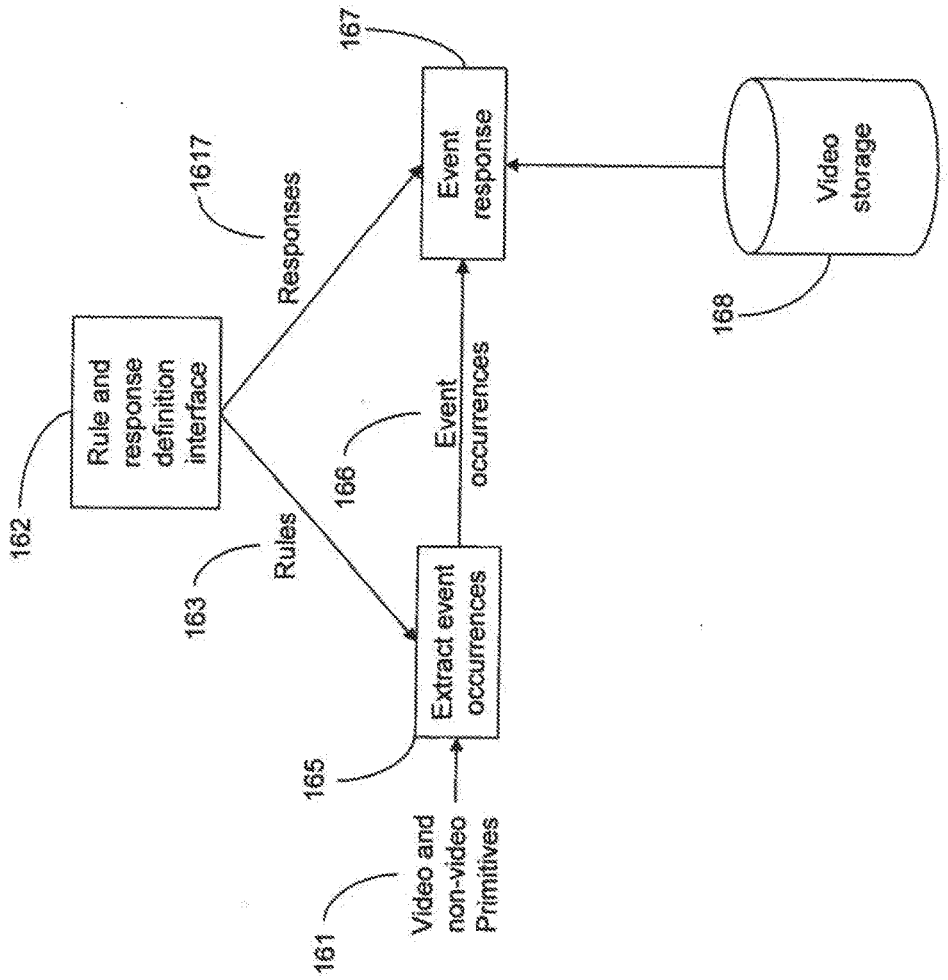

Rule and response definition interface
162
Rules
163
Responses
1617
Extract event occurrences
165
Event occurrences
166
Event response
167
Video storage
168
Video and non-video Primitives
161

Figure 16b

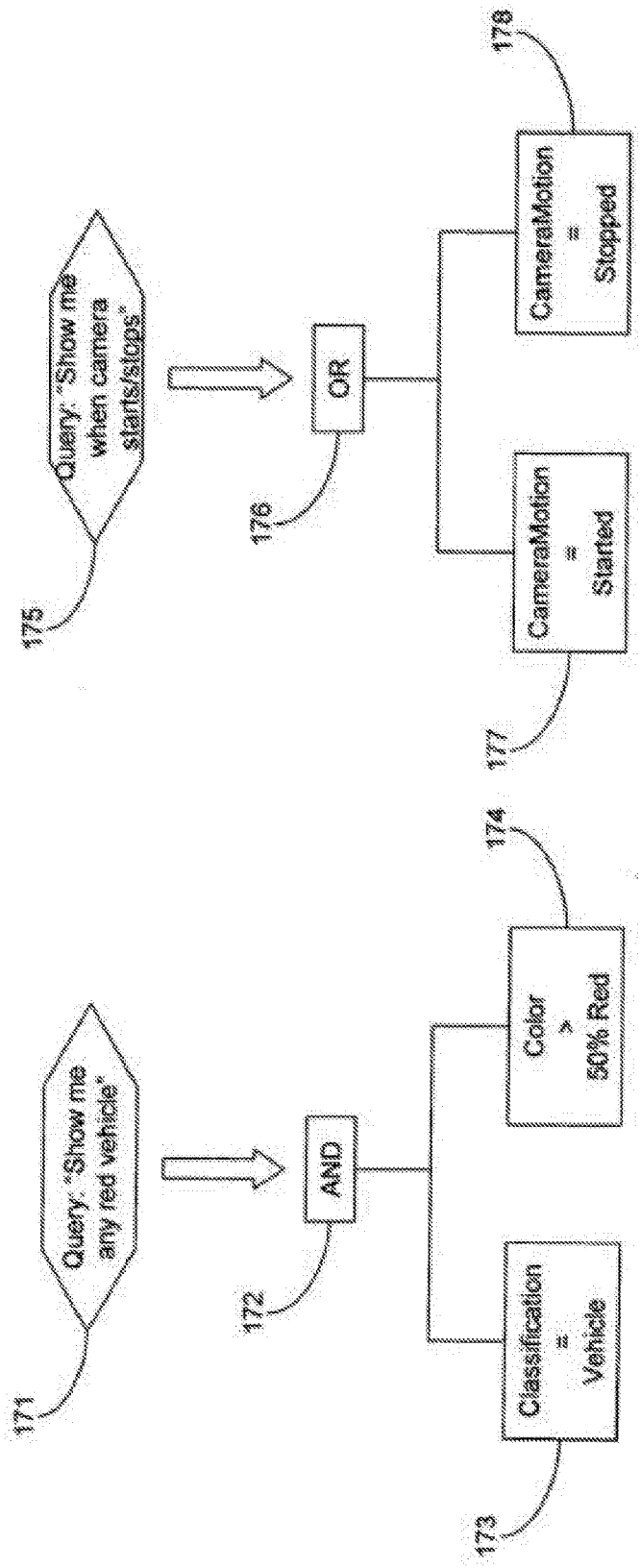


Figure 17a

Figure 17b

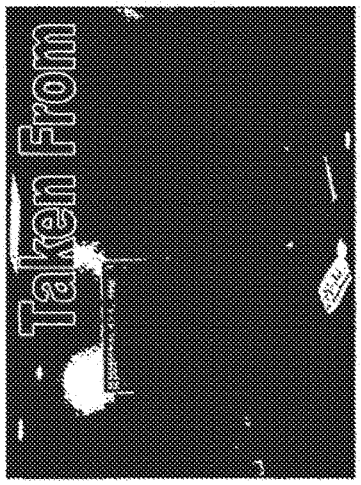


Figure 18c

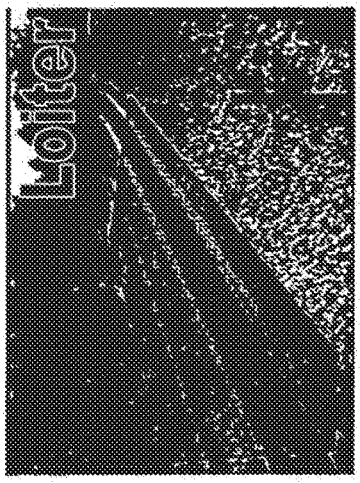


Figure 18b

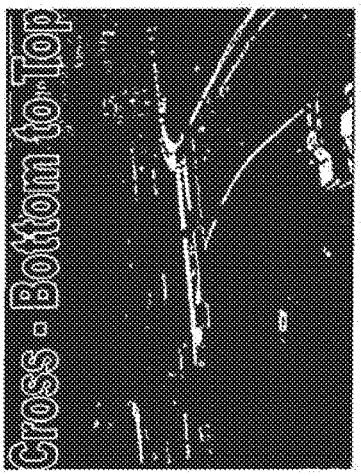


Figure 18a

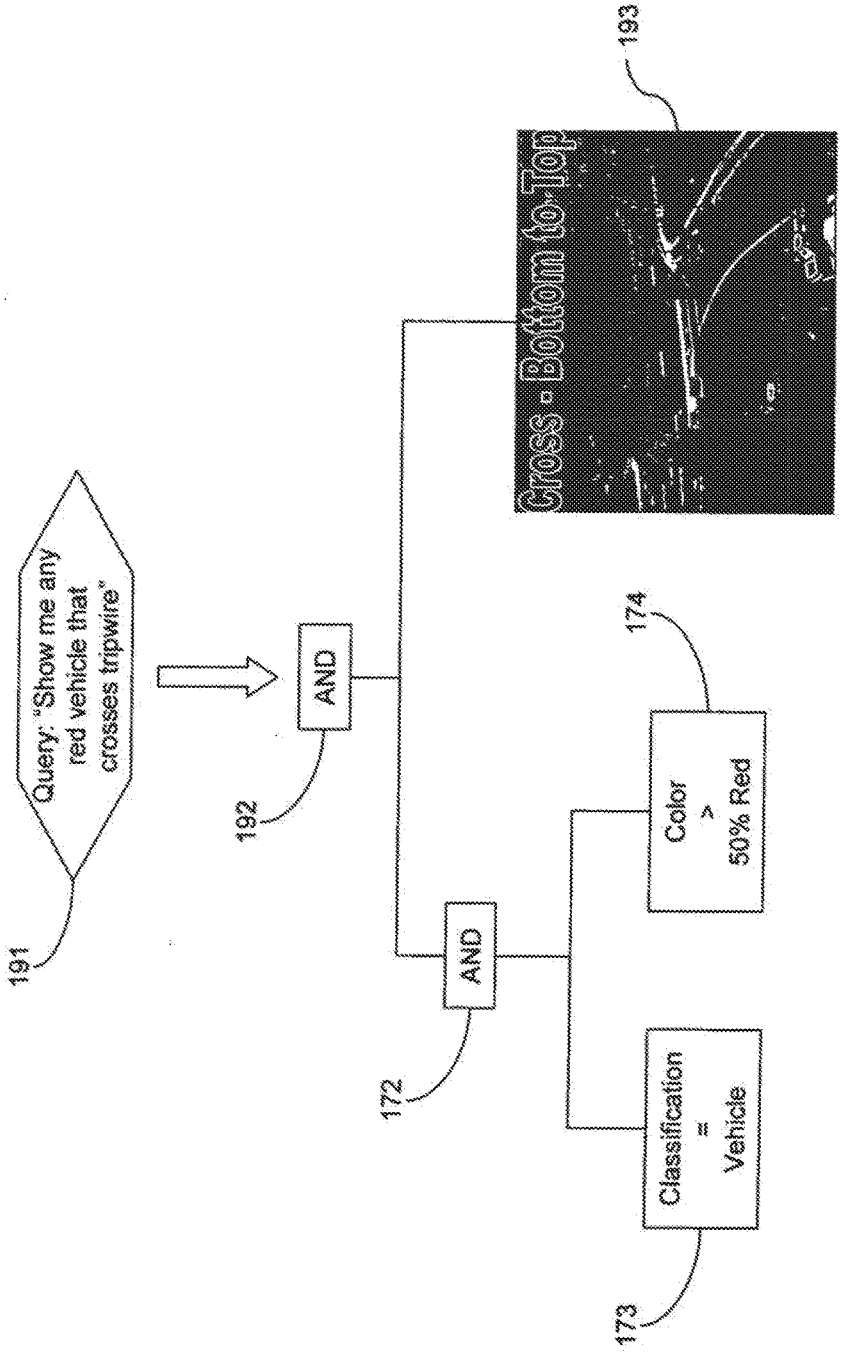

Query: "Show me any red vehicle that crosses tripwire"
191
AND
192
193
Cross - Bottom to Top
AND
172
Color > 50% Red
174
Classification = Vehicle
173

Figure 19

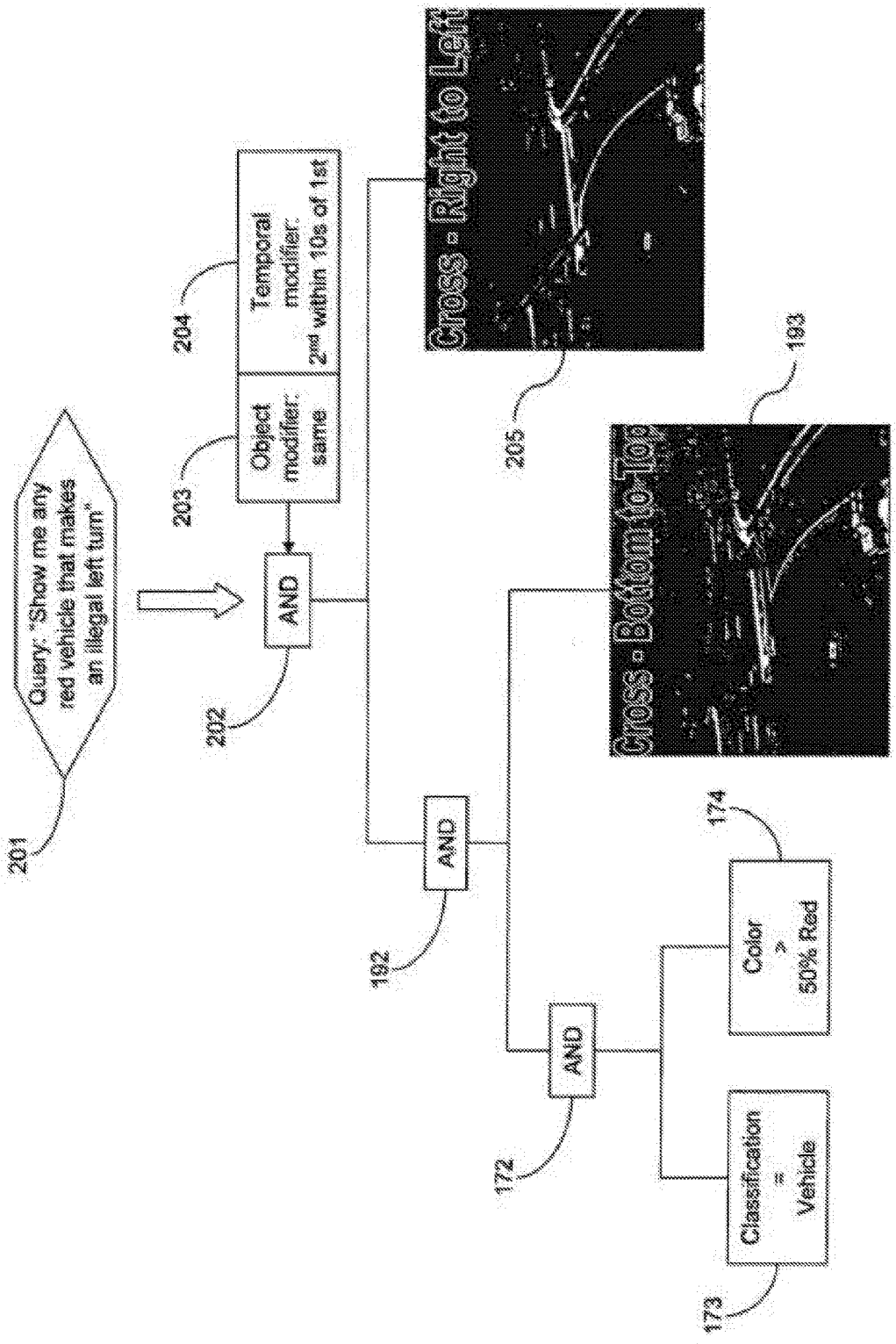

Query: "Show me any red vehicle that makes an illegal left turn"
201
AND
202
Object modifier: same
203
Temporal modifier: 2nd within 10s of 1st
204
Cross - Right to Left
205
Cross - Bottom to Top
193
AND
192
AND
172
Color > 50% Red
174
Classification = Vehicle
173

Figure 20

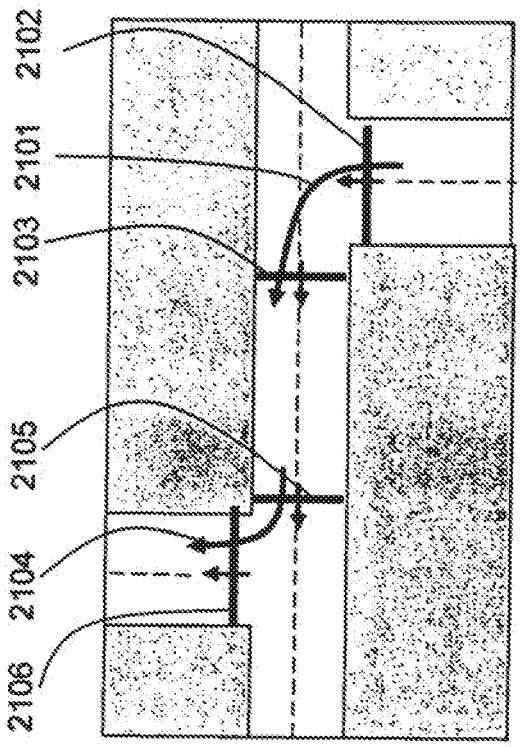


Figure 21a

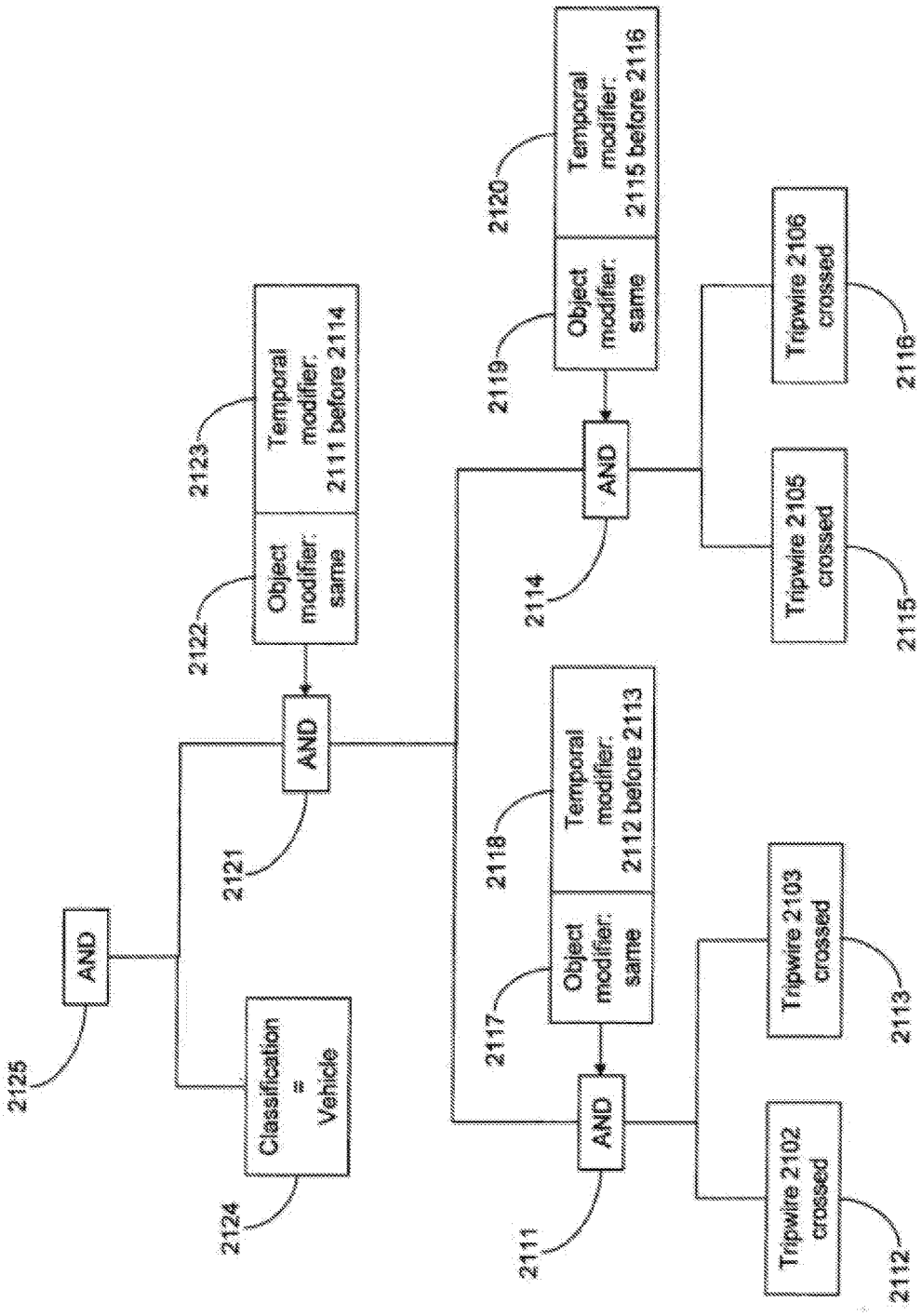

AND
2125
Classification = Vehicle
2124
AND
2121
Object modifier: same
2122
Temporal modifier: 2111 before 2114
2123
AND
2111
Object modifier: same
2117
Temporal modifier: 2112 before 2113
2118
AND
2114
Object modifier: same
2119
Temporal modifier: 2115 before 2116
2120
Tripwire 2102 crossed
2112
Tripwire 2103 crossed
2113
Tripwire 2105 crossed
2115
Tripwire 2106 crossed
2116

Figure 21b

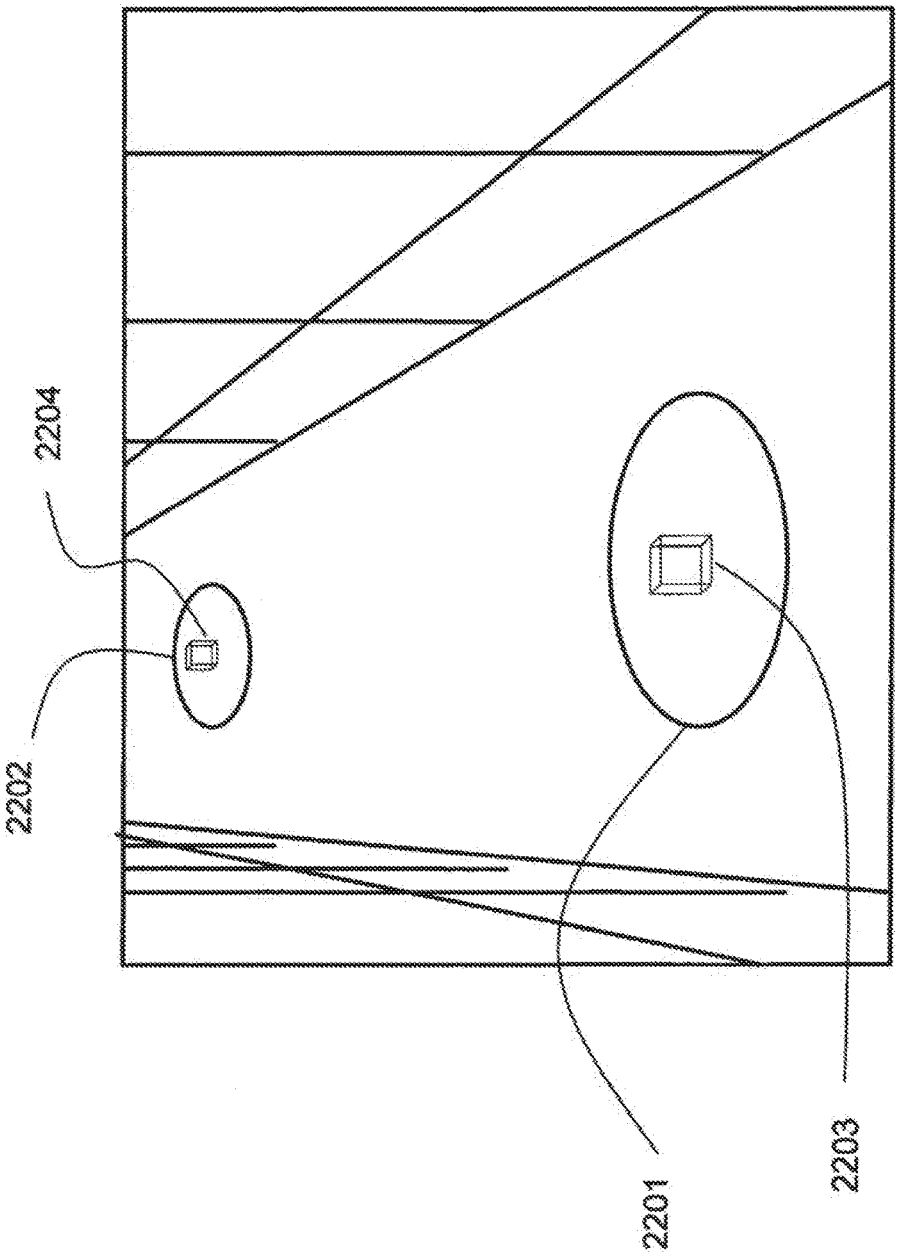


Fig 22.

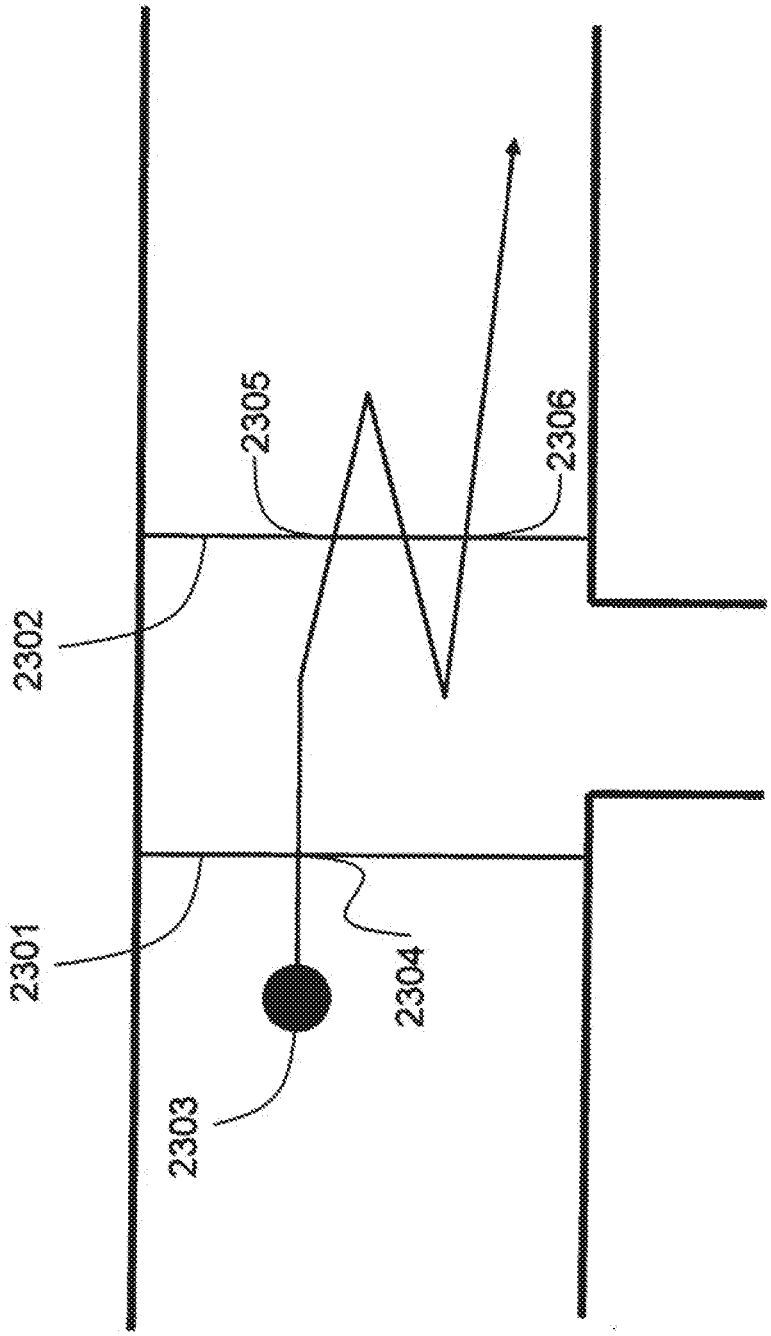


Fig 23.

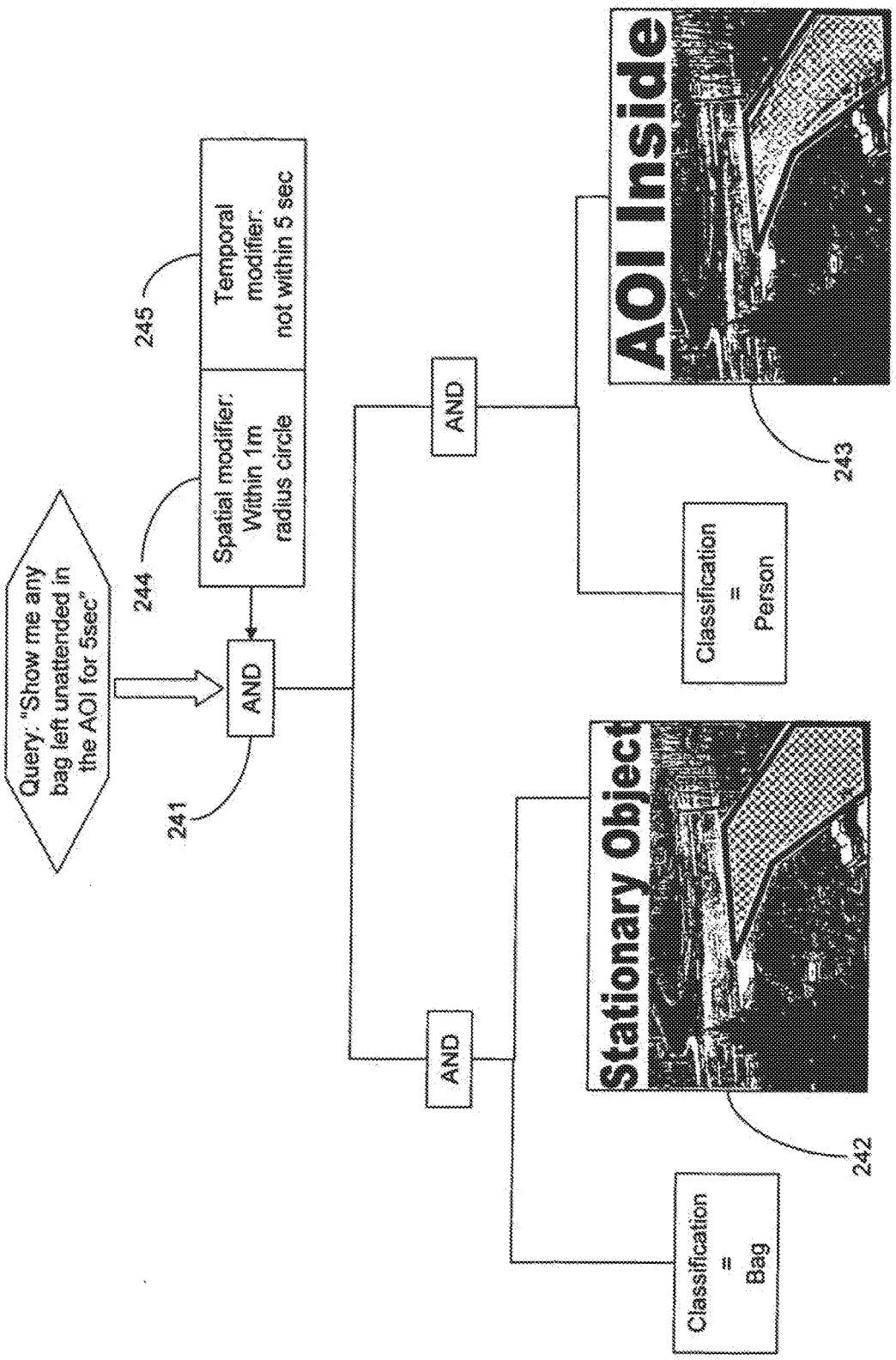

Query: "Show me any bag left unattended in the AOI for 5sec"
241
AND
244
Spatial modifier: Within 1m radius circle
245
Temporal modifier: not within 5 sec
AND
AND
Classification = Bag
Stationary Object
242
Classification = Person
AOI Inside
243

Figure 24

VIDEO SURVEILLANCE SYSTEM EMPLOYING VIDEO PRIMITIVES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to and is a continuation of U.S. patent application Ser. No. 15/044,902, which is a continuation of U.S. patent application Ser. No. 14/203,065 filed Mar. 10, 2014, which is a continuation of U.S. patent application Ser. No. 11/300,581 filed on Dec. 15, 2005, which is a continuation-in-part of U.S. patent application Ser. No. 11/057,154, filed on Feb. 15, 2005, which is a continuation-in-part of U.S. patent application Ser. No. 09/987,707, filed on Nov. 15, 2001, which is a continuation-in-part of U.S. patent application Ser. No. 09/694,712, filed on Oct. 24, 2000 (based on which U.S. Pat. No. 6,954,498 issued on Oct. 11, 2005), all of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

Field of the Invention

[0002] The invention relates to a system for automatic video surveillance employing video primitives.

REFERENCES

[0003] For the convenience of the reader, the references referred to herein are listed below. In the specification, the numerals within brackets refer to respective references. The listed references are incorporated herein by reference.

[0004] The following references describe moving target detection:

[0005] {1} A. Lipton, H. Fujiyoshi and R. S. Patil, "Moving Target Detection and Classification from Real-Time Video," Proceedings of IEEE WACV '98, Princeton, N.J., 1998, pp. 8-14.

[0006] {2} W. E. L. Grimson, et al., "Using Adaptive Tracking to Classify and Monitor Activities in a Site", CVPR, pp. 22-29, Jun. 1998.

[0007] {3} A. J. Lipton, H. Fujiyoshi, R. S. Patil, "Moving Target Classification and Tracking from Real-time Video," *IUW*, pp. 129-136, 1998.

[0008] {4} T. J. Olson and F. Z. Brill, "Moving Object Detection and Event Recognition Algorithm for Smart Cameras," *IUW*, pp. 159-175, May 1997.

[0009] The following references describe detecting and tracking humans:

[0010] {5} A. J. Upton, "Local Application of Optical Flow to Analyse Rigid Versus Non-Rigid Motion," *International Conference on Computer Vision*, Corfu, Greece, September 1999.

[0011] {6} F. Bartolini, V. Cappellini, and A. Mecocci, "Counting people getting in and out of a bus by real-time image-sequence processing," *IVC*, 12(1):36-41, January 1994.

[0012] {7} M. Rossi and A. Bozzoli, "Tracking and counting moving people," *ICIP*94, pp. 212-216, 1994.

[0013] {8} C. R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-time tracking of the human body," *Vismod,* 1995.

[0014] {9} L. Khoudour, L. Duvieubourg, J. P. Deparis, "Real-Time Pedestrian Counting by Active Linear Cameras," *JEI* 5(4):452-459, October 1996.

[0015] {10} S. Ioffe, D. A. Forsyth, "Probabilistic Methods for Finding People," *IJCV*, 43(1):45-68, June 2001.

[0016] {11} M. Isard and J. MacCormick, "BraMBLe: A Bayesian Multiple-Blob Tracker," *ICCV,* 2001.

[0017] The following references describe blob analysis:

[0018] {12} D. M. Gavrila, "The Visual Analysis of Human Movement: A Survey," *CVIU,* 73(1):82-98, January 1999.

[0019] {13} Niels Haering and Niels da Vitoria Lobo, "Visual Event Detection," *Video Computing Series*, Editor Mubarak Shah, 2001.

[0020] The following references describe blob analysis for trucks, cars, and people:

[0021] {14} Collins, Lipton, Kanade, Fujiyoshi, Duggins, Tsin, Tolliver, Enomoto, and Hasegawa, "A System for Video Surveillance and Monitoring: VSAM Final Report," Technical Report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May 2000.

[0022] {15} Lipton, Fujiyoshi, and Patil, "Moving Target Classification and Tracking from Real-time Video," 98 *Darpa IUW*, Nov. 20-23, 1998.

[0023] The following reference describes analyzing a single-person blob and its contours:

[0024] {16} C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland. "Pfinder: Real-Time Tracking of the Human Body," *PAMI*, vol 19, pp. 780-784, 1997.

[0025] The following reference describes internal motion of blobs, including any motion-based segmentation:

[0026] {17} M. Allmen and C. Dyer, "Long-Range Spatiotemporal Motion Understanding Using Spatiotemporal Flow Curves," *Proc. IEEE CVPR*, Lahaina, Maui, Hi., pp. 303-309, 1991.

[0027] {18} L. Wixson, "Detecting Salient Motion by Accumulating Directionally Consistent Flow", IEEE Trans. Pattern Anal. Mach. Intell., vol 22, pp. 774-781, August 2000.

BACKGROUND OF THE INVENTION

[0028] Video surveillance of public spaces has become extremely widespread and accepted by the general public. Unfortunately, conventional video surveillance systems produce such prodigious volumes of data that an intractable problem results in the analysis of video surveillance data.

[0029] A need exists to reduce the amount of video surveillance data so analysis of the video surveillance data can be conducted.

[0030] A need exists to filter video surveillance data to identify desired portions of the video surveillance data.

SUMMARY OF THE INVENTION

[0031] An object of the invention is to reduce the amount of video surveillance data so analysts of the video surveillance data can be conducted.

[0032] An object of the invention is to filter video surveillance data to identify desired portions of the video surveillance data.

[0033] An object of the invention is to produce a real time alarm based on an automatic detection of an event fern video surveillance data.

[0034] An object of the invention is to integrals data from surveillance sensors other than video for improved searching capabilities.

[0035] An object of the invention is to integrate data from surveillance sensors other than video for improved event detection capabilities.

[0036] The invention includes an article of manufacture, a method, a system, and an apparatus for video surveillance.

[0037] The article of manufacture of the invention includes a computer-readable medium comprising software for a video surveillance system, comprising code segments for operating the video surveillance system based on video primitives.

[0038] The article of manufacture of the invention includes a computer-readable medium comprising software for a video surveillance system, comprising code segments for accessing archived video primitives, and code segments for extracting event occurrences from accessed archived video primitives.

[0039] The system of the invention includes a computer system including a computer-readable medium having software to operate a computer in accordance with the invention.

[0040] The apparatus of the invention includes a computer including a computer-readable medium having software to operate the computer in accordance with the invention.

[0041] The article of manufacture of the invention includes a computer-readable medium having software to operate a computer in accordance with the invention.

[0042] Moreover, the above objects and advantages of the invention are illustrative, and not exhaustive, of those that can be achieved by the invention. Thus, these and other objects and advantages of the invention will be apparent from the description herein, both as embodied herein and as modified in view of any variations which will be apparent to those skilled in the art.

Definitions

[0043] A "video" refers to motion pictures represented in analog and/or digital form. Examples of video include: television, movies, image sequences from a video camera or other observer, and computer-generated image sequences.

[0044] A "frame" refers to a particular image or other discrete unit within a video.

[0045] An "object" refers to an item of interest in a video. Examples of an object include: a person, a vehicle, an animal, and a physical subject.

[0046] An "activity" refers to one or more actions and/or one or more composites of actions of one or more objects. Examples of an activity include: entering; exiting; stopping; moving; raising; lowering; growing; and shrinking.

[0047] A "location" refers to a space where an activity may occur. A location can be, for example, scene-based or image-based. Examples of a scene-based location include: a public space; a store; a retail space; an office; a warehouse; a hotel room; a hotel lobby; a lobby of a building; a casino; a bus station; a train station; an airport; a port; a bus; a train; an airplane; and a ship. Examples of an image-based location include: a video image; a line in a video image; an area in a video image; a rectangular section of a video image; and a polygonal section of a video image.

[0048] An "event" refers to one or more objects engaged in an activity. The event may be referenced with respect to a location and/or a time.

[0049] A "computer" refers to any apparatus that is capable of accepting a structured input, processing the structured input according to prescribed rules, and producing results of the processing as output. Examples of a computer include: a computer; a general purpose computer; a supercomputer; a mainframe; a super mini-computer; a mini-computer; a workstation; a micro-computer; a server; an interactive television; a hybrid combination of a computer and an interactive television; and application-specific hardware to emulate a computer and/or software. A computer can have a single processor or multiple processors, which can operate in parallel and/or not in parallel. A computer also refers to two or more computers connected together via a network for transmitting or receiving information between the computers. An example of such a computer includes a distributed computer system for processing information via computers linked by a network.

[0050] A "computer-readable medium" refers to any storage device used for storing data accessible by a computer. Examples of a computer-readable medium include: a magnetic hard disk; a floppy disk; an optical disk, such as a CD-ROM and a DVD; a magnetic tape; a memory chip; and a carrier wave used to carry computer-readable electronic data, such as those used in transmitting and receiving e-mail or in accessing a network.

[0051] "Software" refers to prescribed rules to operate a computer. Examples of software include: software; code segments; instructions; computer programs; and programmed logic.

[0052] A "computer system" refers to a system having a computer, where the computer comprises a computer-readable medium embodying software to operate the computer.

[0053] A "network" refers to a number of computers and associated devices that are connected by communication facilities. A network involves permanent connections such as cables or temporary connections such as those made through telephone or other communication links. Examples of a network include: an internet, such as the Internet; an intranet; a local area network (LAN); a wide area network (WAN); and a combination of networks, such as an internet and an intranet.

BRIEF DESCRIPTION OF THE DRAWINGS

[0054] Embodiments of the invention are explained in greater detail by way of the drawings, where the same reference numerals refer to the same features.

[0055] FIG. **1** illustrates a plan view of the video surveillance system of the invention.

[0056] FIG. **2** illustrates a flow diagram for the video surveillance system of the invention.

[0057] FIG. **3** illustrates a flow diagram for tasking the video surveillance system.

[0058] FIG. **4** illustrates a flow diagram for operating the video surveillance system.

[0059] FIG. **5** illustrates a flow diagram for extracting video primitives for the video surveillance system.

[0060] FIG. **6** illustrates a flow diagram for taking action with the video surveillance system.

[0061] FIG. **7** illustrates a flow diagram for semi-automatic calibration of the video surveillance system.

[0062] FIG. **8** illustrates a flow diagram for automatic calibration of the video surveillance system.

[0063] FIG. **9** illustrates an additional flow diagram for the video surveillance system of the invention.

[0064] FIGS. **10-15** illustrate examples of the video surveillance system of the invention applied to monitoring a grocery store.

[0065] FIG. **16***a* shows a flow diagram of a video analysis subsystem according to an embodiment of the invention.

[0066] FIG. **16***b* shows the flow diagram of the event occurrence detection and response subsystem according to an embodiment of the invention.

[0067] FIGS. **17***a* and **17***b* show exemplary database queries.

[0068] FIGS. **18***a*, **18***b* and **18***c* show three exemplary activity detectors according to various embodiments of the invention: detecting tripwire crossings (FIG. **18***a*), loitering (FIG. **18***b*) and theft (FIG. **18***c*).

[0069] FIG. **19** shows an activity detector query according to an embodiment of the invention.

[0070] FIG. **20** shows an exemplary query using activity detectors and Boolean operators with modifiers, according to an embodiment of the invention.

[0071] FIGS. **21***a* and **21***b* show an exemplary query using multiple levels of combinators, activity detectors, and property queries.

[0072] FIG. **22** shows an exemplary indication of exclusion zone size, according to some embodiments of the invention.

[0073] FIG. **23** shows an exemplary detection of a particular activity according to some embodiments of the invention.

[0074] FIG. **24** shows an exemplary implementation of a query regarding stationary object detection with an exclusion zone according to some embodiments of the invention.

DETAILED DESCRIPTION OF THE INVENTION

[0075] The automatic video surveillance system of the invention is for monitoring a location for, for example, market research or security purposes. The system cars be a dedicated video surveillance installation with purpose-built surveillance components, or the system can be a retrofit to existing video surveillance equipment that piggybacks off the surveillance video feeds. The system is capable of analyzing video data from live sources or from recorded media. The system is capable of processing the video data in real-time, and storing the extracted video primitives to allow very high speed forensic event detection later. The system can have a prescribed response to the analysis, such as record data, activate an alarm mechanism, or activate another sensor system. The system is also capable of integrating with other surveillance system components. The system may be used to produce, for example, security or market research reports that can be tailored according to the needs of an operator and, as an option, can be presented through an interactive web-based interface, or other reporting mechanism.

[0076] An operator is provided with maximum flexibility in configuring the system by using event discriminators. Event discriminators are identified with one or more objects (whose descriptions are based on video primitives), along with one or more optional spatial attributes, and/or one or more optional temporal attributes. For example, an operator can define an event discriminator (called a "loitering" event in this example) as a "person" object in the "automatic teller machine" space for "longer than 15 minutes" and "between 10:00 p.m. and 6:00 a.m." Event discriminators can be combined with modified Boolean operators to form more complex queries.

[0077] Although the video surveillance system of the invention draws on well-known computer vision techniques from the public domain, the inventive video surveillance system has several unique and novel features that are not currently available. For example, current video surveillance systems use large volumes of video imagery as the primary commodity of information interchange. The system of the invention uses video primitives as the primary commodity with representative video imagery being used as collateral evidence. The system of the invention can also be calibrated (manually, semi-automatically, or automatically) and thereafter automatically can infer video primitives from video imagery. The system can further analyze previously processed video without needing to reprocess completely the video. By analyzing previously processed video, the system can perform inference analysis based on previously recorded video primitives, which greatly improves the analysis speed of the computer system.

[0078] The use of video primitives may also significantly reduce the storage requirements for the video. This is because the event detection and response subsystem uses the video only to illustrate the detections. Consequently, video may be stored at a lower quality. In a potential embodiment, the video may be stored only when activity is detected, not all the time. In another potential embodiment, the quality of the stored video may be dependent on whether activity is detected: video can be stored at higher quality (higher frame-rate and/or bit-rate) when activity is detected and at lower quality at other times. In another exemplary embodiment, the video storage and database may be handled separately, e.g., by a digital video recorder (DVR), and the video processing subsystem may just control whether data is stored and with what quality.

[0079] As another example, the system of the invention provides unique system tasking. Using equipment control directives, current video systems allow a user to position video sensors and, in some sophisticated conventional systems, to mask out regions of interest or disinterest Equipment control directives are instructions to control the position, orientation, and focus of video cameras. Instead of equipment control directives, the system of the invention uses event discriminators based on video primitives as the primary tasking mechanism. With event discriminators and video primitives, an operator is provided with a much more intuitive approach over conventional systems for extracting useful information from the system. Rather than tasking a system with an equipment control directives, such as "camera A pan 45 degrees to the left," the system of the invention can be tasked in a human-intuitive manner with one or more event discriminators based on video primitives, such as "a person enters restricted area A."

[0080] Using the invention for market research, the following are examples of the type of video surveillance that can be performed with the invention: counting people in a store; counting people in a part of a store; counting people who stop in a particular place in a store; measuring how long people spend in a store; measuring how long people spend in a part of a store; and measuring the length of a line in a store.

[0081] Using the invention for security, the following are examples of the type of video surveillance that can be performed with the invention; determining when anyone enters a restricted area and storing associated imagery; determining when a person enters an area at unusual times;

determining when changes to shelf space and storage space occur that might be unauthorized; determining when passengers aboard an aircraft approach the cockpit; determining when people tailgate through a secure portal; determining if there is an unattended bag in an airport; and determining if there is a theft of as asset.

[0082] An exemplary application area may be access control, which may include, for example: detecting if a person climbs over a fence, or enters a prohibited area; detecting if someone moves in the wrong direction (e.g., at an airport, entering a secure area through the exit); determining if a number of objects detected in an area of interest does not match an expected number based on RFID tags or card-swipes for entry, indicating the presence of unauthorized personnel. This may also be useful in a residential application, where the video surveillance system may be able to differentiate between the motion of a person and pet, thus eliminating most false alarms. Note that in many residential applications, privacy may be of concern; for example, a homeowner may not wish to have another person remotely monitoring the home and to be able to see what is in the house and what is happening in the house. Therefore, in some embodiments used in such applications, the video processing may be performed locally, and optional video or snapshots may be sent to one or more remote monitoring stations only when necessary (for example, but not limited to, detection of criminal activity or other dangerous situations).

[0083] Another exemplary application area may be asset monitoring. This may mean detecting if an object is taken away from the scene, for example, if an artifact is removed from a museum. In a retail environment asset monitoring can have several aspects to it and may include, for example: detecting if a single person takes a suspiciously large number of a given item; determining if a person exits through the entrance, particularly if doing this while pushing a shopping cart; determining if a person applies a non-matching price tag to an item, for example, filling a bag with the most expensive type of coffee but using a price tag for a less expensive type; or detecting if a person leaves a loading dock with large boxes.

[0084] Another exemplary application area may be for safety purposes. This may include, for example: detecting if a person slips and falls, e.g., in a store or in a parking lot; detecting if a car is driving too fast in a parking lot; detecting if a person is too close to the edge of the platform at a train or subway station while there is no train at the station; detecting if a person is on the rails; detecting if a person is caught in the door of a train when it starts moving; or counting the number of people entering and leaving a facility, thus keeping a precise headcount, which can be very important in case of an emergency.

[0085] Another exemplary application area may be traffic monitoring. This may include detecting if a vehicle stopped, especially in places like a bridge or a tunnel, or detecting if a vehicle parks in a no parking area.

[0086] Another exemplary application area may be terrorism prevention. This may include, in addition to some of the previously-mentioned applications, detecting if an object is left behind in an airport concourse, if an object is thrown over a fence, or if an object is left at a rail track; detecting a person loitering or a vehicle circling around critical infrastructure; or detecting a fast-moving boat approaching a ship in a port or in open waters.

[0087] Another exemplary application area may be in care for the sick and elderly, even in the home. This may include, for example, detecting if the person falls; or detecting unusual behavior, like the person not entering the kitchen for an extended period of time.

[0088] FIG. **1** illustrates a plan view of the video surveillance system of the invention. A computer system **11** comprises a computer **12** having a computer-readable medium **13** embodying software to operate the computer **12** according to the invention. The computer system **11** is coupled to one or more video sensors **14**, one or more video recorders **15**, and one or more input/output (I/O) devices **16**. The video sensors **14** can also be optionally coupled to the video recorders **15** for direct recording of video surveillance data. The computer system is optionally coupled to other sensors **17**.

[0089] The video sensors **14** provide source video to the computer system **11**. Each video sensor **14** can be coupled to the computer system **11** using, for example, a direct connection (e.g., a firewire digital camera interface) or a network. The video sensors **14** can exist prior to installation of the invention or can be installed as part of the invention. Examples of a video sensor **14** include: a video camera; a digital video camera; a color camera; a monochrome camera; a camera; a camcorder, a PC camera; a webcam; an infra-red video camera; and a CCTV camera.

[0090] The video recorders **15** receive video surveillance data from the computer system **11** for recording and/or provide source video to the computer system **11**. Each video recorder **15** can be coupled to the computer system **11** using, for example, a direct connection or a network. The video recorders **15** can exist prior to installation of the invention or can be installed as part of the invention. The video surveillance system in the computer system **11** may control when and with what quality setting a video recorder **15** records video. Examples of a video recorder **15** include: a video tape recorder; a digital video recorder; a video disk; a DVD; and a computer-readable medium.

[0091] The I/O devices **16** provide input to and receive output from the computer-system **11**. The I/O devices **16** can be used to task the computer system **11** and produce reports from the computer system **11**. Examples of I/O devices **16** include: a keyboard; a mouse; a stylus; a monitor; a printer; another computer system; a network; and an alarm.

[0092] The other sensors **17** provide additional input to the computer system **11**. Each other sensor **17** can be coupled to the computer system **11** using, for example, a direct connection or a network. The other sensors **17** can exit prior to installation of the invention or can be installed as part of the invention. Examples of another sensor **17** include, but are not limited to: a motion sensor; an optical tripwire; a biometric sensor; an RFID sensor; and a card-based or keypad-based authorization system. The outputs of the other sensors **17** can be recorded by the computer system **115** recording devices, and/or recording systems.

[0093] FIG. **2** illustrates a flow diagram for the video surveillance system of the invention. Various aspects of the invention are exemplified with reference to FIGS. **10-15**, which illustrate examples of the video surveillance system of the invention applied to monitoring a grocery store.

[0094] In block **21**, the video surveillance system is set up as discussed for FIG. **1**. Each video sensor **14** is orientated to a location for video surveillance. The computer system **11** is connected to the video feeds from the video equipment **14**

and **15**. The video surveillance system can be implemented using existing equipment or newly installed equipment for the location.

[0095] In block **22**, the video surveillance system is calibrated. Once the video surveillance system is in place from block **21**, calibration occurs. The result of block **22** is the ability of the video surveillance system to determine an approximate absolute size and speed of a particular object (e.g., a person) at various places in the video image provided by the video sensor. The system can be calibrated using manual calibration, semi-automatic calibration, and automatic calibration. Calibration is further described alter the discussion of block **24**.

[0096] In block **23** of FIG. **2**, the video surveillance system is tasked. Tasking occurs after calibration in block **22** and is optional. Tasking the video surveillance system involves specifying one or more event discriminators. Without tasking, the video surveillance system operates by detecting and archiving video primitives and associated video imagery without taking any action, as in block **45** in FIG. **4**.

[0097] FIG. **3** illustrates a flow diagram for tasking the video surveillance system to determine event discriminators. An event discriminator refers to one or more objects optionally interacting with one or more spatial attributes and/or one or more temporal attributes. An event discriminator is described in terms of video primitives (also called activity description meta-data). Some of the video primitive design criteria include the following: capability of being extracted from the video stream in real-time; inclusion of all relevant information from the video; and conciseness of representation.

[0098] Real-time extraction of the video primitives from the video stream is desirable to enable the system to be capable of generating real-time alerts, and to do so, since the video provides a continuous input stream, the system cannot fall behind.

[0099] The video primitives should also contain all relevant information from the video, since at the time of extracting the video primitives, the user-defined rules are not known to the system. Therefore, the video primitives should contain information to be able to detect any event specified by the user, without the need for going back to the video and reanalyzing it.

[0100] A concise representation is also desirable for multiple reasons. One goal of the proposed invention may be to extend the storage recycle time of a surveillance system. This may be achieved by replacing storing good quality video all the time by storing activity description meta-data and video with quality dependent on the presence of activity, as discussed above. Hence, the more concise the video primitives are, the more data can be stored. In addition, the more concise the video primitive representation, the faster the data access becomes, and this, in turn may speed up forensic searching.

[0101] The exact contents of the video primitives may depend on the application and potential events of interest. Some exemplary embodiments are described below

[0102] An exemplary embodiment of the video primitives may include scene/video descriptors, describing the overall scene and video. In general, this may include a detailed description of the appearance of the scene, e.g., the location of sky, foliage, man-made objects, water, etc; and/or meteorological conditions, e.g., the presence/absence of precipitation, fog, etc. For a video surveillance application, for example, a change in the overall view may be important. Exemplary descriptors may describe sudden lighting changes; they may indicate camera motion, especially the facts that the camera started or stopped moving, and in the latter case, whether it returned to its previous view or at least to a previously known view; they may indicate changes in the quality of the video feed, e.g., if it suddenly became noisier or went dark, potentially indicating tampering with the feed; or they may show a changing waterline along a body of water (for further information on specific approaches to this latter problem, one may consult, for examples co-pending U.S. patent application Ser. No. 10/954,479, filed on Oct. 1, 2004, and incorporated herein by reference).

[0103] Another exemplary embodiment of the video primitives may include object descriptors referring to an observable attribute of an object viewed in a video feed. What information is stored about an object may depend on the application area and the available processing capabilities. Exemplary object descriptors may include generic properties including, but not limited to, size, shape, perimeter, position, trajectory, speed and direction of motion, motion salience and its features, color, rigidity, texture, and/or classification. The object descriptor may also contain some more application and type specific information: for humans, this may include the presence and ratio of skin tone, gender and race information, some human body model describing the human shape and pose; or for vehicles, it may include type (e.g., truck, SUV, sedan, bike, etc.), make, model, license plate number. The object descriptor may also contain activities, including, but not limited to, carrying an object, naming, walking, standing up, or raising arms. Some activities, such as talking, fighting or colliding, may also including, but not limited to, face or gait.

[0104] Another exemplary embodiment of the video primitives may include flow descriptors describing the direction of motion of every area of the video. Such descriptors may, for example, be used to detect passback events, by detecting any motion in a prohibited direction (for further information on specific approaches to this latter problem, one may consult, for example, co-pending U.S. patent application Ser. No. 10/766,949, filed on Jan. 30, 2004, and incorporated herein by reference).

[0105] Primitives may also come from non-video sources, such as audio sensors, heat sensors, pressure sensors, card readers, RFID tags, biometric sensors, etc.

[0106] A classification refers to an identification of an object as belonging to a particular category or class. Examples of a classification include: a person; a dog; a vehicle; a police car; an individual person; and a specific type of object.

[0107] A size refers to a dimensional attribute of an object. Examples of a size include: large; medium; small; flat; taller than 6 feet; shorter than 1 foot; wider than 3 feet; thinner than 4 feet; about human size; bigger than a human; smaller than a human; about the size of a car; a rectangle in an image with approximate dimensions in pixels; and a number of image pixels.

[0108] Position refers to a spatial attribute of an object. The position may be, for example, an image position in pixel coordinates, an absolute real-world position in some world coordinate system, or a position relative to a landmark or another object.

[0109] A color refers to a chromatic attribute of an object. Examples of a color include: white; black; grey; red; a range of HSV values; a range of YUV values; a range of RGB values; an average RGB value; an average YUV value; aid a histogram of RGB values.

[0110] Rigidity refers to a shape consistency attribute of an object The shape of non-rigid objects (e.g., people or animals) may change from frame to frame, while that of rigid objects (e.g., vehicles or houses) may remain largely unchanged from frame to frame (except, perhaps, for slight changes due to turning).

[0111] A texture refers to a pattern attribute of an object. Examples of texture features include: self-similarity; spectral power; linearity; and coarseness.

[0112] An internal motion refers to a measure of the rigidity of an object. An example of a fairly rigid object is a car, which does not exhibit a great amount of internal motion. An example of a fairly non-rigid object is a person having swinging arms and legs, which exhibits a great amount of internal motion.

[0113] A motion refers to any motion that can be automatically detected. Examples of a motion include: appearance of an object; disappearance of an object; a vertical movement of an object; a horizontal movement of an object; and a periodic movement of art object.

[0114] A salient motion refers to any motion that can be automatically detected and can be tracked for some period of time. Such a moving object exhibits apparently purposeful motion. Examples of a salient motion include: moving from one place to another; and moving to interact with another object.

[0115] A feature of a salient motion refers to a property of a salient motion. Examples of a feature of a salient motion include: a trajectory; a length of a trajectory in image space; an approximate length of a trajectory in a three-dimensional representation of the environment; a position of as object in image space as a function of time; at approximate position of an object in a three-dimensional representation of the environment as a function of time; a duration of a trajectory; a velocity (e.g., speed and direction) in image space; an approximate velocity (e.g., speed and direction) in a three-dimensional representation of the environment; a duration of time at a velocity; a change of velocity in image space; an approximate change of velocity in a three-dimensional representation of the environment; a duration of a change of velocity; cessation of motion; and a duration of cessation of motion. A velocity refers to the speed and direction of an object at a particular time. A trajectory refers a set of (position, velocity) pairs for an object for as long as the object can be tracked or for a time period.

[0116] A scene change refers to any region of a scene that can be detected as changing over a period of time. Examples of a scene change include: an stationary object leaving a scene; an object entering a scene and becoming stationary; an object changing position in a scene; and an object changing appearance (e.g. color, shape, or size).

[0117] A feature of a scene change refers to a property of a scene change. Examples of a feature of a scene change include: a size of a scene change in image space; an approximate size of a scene change in a three-dimensional representation of the environment; a time at which a scene change occurred; a location of a scene change in image space; and an approximate location of a scene change in a three-dimensional representation of the environment.

[0118] A pre-defined model refers to an a priori known model of an object. Examples of a pre-defined model may include: an adult; a child; a vehicle; and a semi-trailer.

[0119] FIG. **16***a* shows an exemplary video analysis portion of a video surveillance system according to an embodiment of the invention. In FIG. **16***a*, a video sensor (for example, but not limited to, a video camera) **1601** may provide a video stream **1602** to a video analysis subsystem **1603**. Video analysis subsystem **1603** may then perform analysis of the video stream **1602** to derive video primitives, which may be stored in primitive storage **1605**. Primitive storage **1605** may be used to store non-video primitives, as well. Video analysis subsystem **1603** may further control storage of all or portions of the video stream **1602** in video storage **1604**, for example, quality and/or quantity of video, as discussed above.

[0120] Referring now to FIG. **16***b*, once the video, and, if there are other sensors, the non-video primitives **161** are available, the system may detect events. The user tasks the system by defining rules **163** and corresponding responses **164** using the rule and response definition interface **162**. The rules are translated into event discriminators, and the system extracts corresponding event occurrences **165**. The detected event occurrences **166** trigger user defined responses **167**. A response may include a snapshot of a video of the detected event from video storage **168** (which may or may not be the same as video storage **1604** in FIG. **16***a*). The video storage **168** may be part of the video surveillance system, or it may be a separate recording device **15**. Examples of a response may include, but are not necessarily limited to, the following: activating a visual and/or audio alert on a system display; activating a visual and/or audio alarm system at the location; activating a silent alarm; activating a rapid response mechanism; locking a door; contacting a security service; forwarding data (e.g., image data, video data, video limited to, the Internet; saving such data to a designated computer-readable medium; activating some other sensor or surveillance system; tasking the computer system **11** and/or another computer system; and/or directing the computer system **11** and/or another computer system.

[0121] The primitive data can be thought of as data stored in a database. To detect event occurrences in it, an efficient query language is required. Embodiments of the inventive system may include an activity inferencing language, which will be described below.

[0122] Traditional relational database querying schemas often follow a Boolean binary tree structure to allow users to create flexible queries on stored data of various types. Leaf nodes arc usually of the form "property relationship value," where a property is some key feature of the data (such as time or name); a relationship is usually a numerical operator (">", "<", "**", etc); and a value is a valid state for that property. Branch nodes usually represent unary or binary Boolean logic operators like "and", "or", and "not".

[0123] This may form the basis of an activity query formulation schema, as in embodiments of the present invention. In case of a video surveillance application, the properties may be features of the object detected in the video stream, such as size, speed, color, classification (human, vehicle), or the properties may be scene change properties. FIG. **17** gives examples of using such queries. In FIG. **17***a*, the query, "Show me any red vehicle," **171** is posed. This may be decomposed into two "property relationship value" (or simply "property") queries, testing whether the classifi-

cation of an object is vehicle **173** and whether its color is predominantly red **174**. These two sub-queries can combined with the Boolean operator "and" **172**. Similarly, in FIG. **17***b*, the query, "Show me when a camera starts or stops moving," may be expressed as the Boolean "or" **176** combination of the property sub-queries, "has the camera started moving" **177** and "has the camera stopped moving" **178**.

[0124] Embodiments of the invention may extend this type of database query schema in two exemplary ways: (1) the basic leaf nodes may be augmented with activity detectors describing spatial activities within a scene; and (2) the Boolean operator branch nodes may be augmented with modifiers specifying spatial, temporal and object interrelationships.

[0125] Activity detectors correspond to a behavior related to an area of the video scene. They describe how an object might interact with a location in the scene. FIG. **18** illustrates three exemplary activity detectors. FIG. **18***a* represents the behavior of crossing a perimeter in a particular direction using a virtual video tripwire (for further information about how such virtual video tripwires may be implemented, one may consult, e.g., U.S. Pat. No. 6,696,945). FIG. **18***b* represents the behavior of loitering for a period of time on a railway track. FIG. **18***c* represents the behavior of taking something away from a section of wall (for exemplary approaches to how this may be done, one may consult U.S. patent application Ser. No. 10/331,778, entitled, "Video Scene Background Maintenance—Change Detection & Classification," filed on Jan. 30, 2003). Other exemplary activity detectors may include detecting a person falling, detecting a person changing direction or speed, detecting a person entering as area, or detecting a person going in the wrong direction.

[0126] FIG. **19** illustrates an example of how an activity detector leaf node (here, tripwire crossing) can be combined with simple property queries to detect whether a red vehicle crosses a video tripwire **191**. The property queries **172**, **173**, **174** and the activity detector **193** are combined with a Boolean "and" operator **192**.

[0127] Combining queries with modified Boolean operators (combinators) may add further flexibility. Exemplary modifiers include spatial, temporal, object, and counter modifiers.

[0128] A spatial modifier may cause the Boolean operator to operate only on child activities (i.e., the arguments of the Boolean operator, as shown below a Boolean operator, e.g., in FIG. **19**) that are proximate/non-proximate within the scene. For example, "and—within 50 pixels of" may be used to mean that the "and" only applies if the distance between activities is less than 50 pixels.

[0129] A temporal modifier may cause the Boolean operator to operate only on child activities that occur within a specified period of time of each other, outside of such a time period, or within a range of times. The time ordering of events may also be specified. For example "and—first within 10 seconds of second" may be used to mean that the "and" only applies if the second child activity occurs not more than 10 seconds after the first child activity.

[0130] An object modifier may cause the Boolean operator to operate only on child activities that occur involving the same or different objects. For example "and—involving the same object" may be used to mean that the "and" only applies if the two child activities involve the same specific object

[0131] A counter modifier may cause the Boolean operator to be triggered only if the condition(s) is/are met a prescribed number of times. A counter modifier may generally include a numerical relationship, such as "at least n times," "exactly n times," "at most n times," etc. For example, "or—at least twice" may be used to mean that at least two of the sub-queries of the "or" operator have to be true. Another use of the counter modifier may be to implement a rule like "alert if the same person takes at least five items from a shelf."

[0132] FIG. **20** illustrates an example of using combinators. Here, the required activity query is to "find a red vehicle making an illegal left turn" **201**. The illegal left turn may be captured through a combination of activity descriptors and modified Boolean operators. One virtual tripwire may be used to detect objects coming out of the side street **193**, and another virtual tripwire may be used to detect objects traveling to the left along the road **205**. These may be combined by a modified "and" operator **202**. The standard Boolean "and" operator guarantees that both activities **193** and **205** have to be detected. The object modifier **203** cheeks that the same object crossed both tripwires, while the temporal modifier **204** checks that the bottom-to-top tripwire **193** is crossed first, followed by the crossing of the right-to-left tripwire **205** no more than 10 seconds later.

[0133] This example also indicates the power of the combinators. Theoretically it is possible to define a separate activity detector for left turn, without relying on simple activity detectors and combinators. However, that detector would be inflexible, making it difficult to accommodate arbitrary turning angles and directions, and it would also be cumbersome to write a separate detector for all potential events. In contrast, using the combinators and simple detectors provides great flexibility.

[0134] Other examples of complex activities that can be detected as a combination of simpler ones may include a car parking and a person getting out of the car or multiple people forming a group, tailgating. These combinators can also combine primitives of different types and sources. Examples may include rules such as "show a person inside a room before the lights are turned off;" "show a person entering a door without a preceding card-swipe;" or "show if an area of interest has more objects than expected by an RFID tag reader," i.e., an illegal object without an RFID tag is in the area.

[0135] A combinator may combine any number of sub-queries, and it may even combine other combinators, to arbitrary depths. An example, illustrated in FIGS. **21***a* and **21***b*, may be a rule to detect if a car turns left **2101** and then turns right **2104**. The left turn **2101** may be detected with the directional tripwires **2102** and **2103**, while the right turn **2104** with be directional tripwires **2105** and **2106**. The left turn may be expressed as the tripwire activity detectors **2112** and **2113**, corresponding to tripwires **2102** and **2103**, respectively, joined with the "and" combinator **2111** with the object modifier "same" **2117** and temporal modifier "**2112** before **2113**" **2118**. Similarly, the right turn may be expressed as the tripwire activity detectors **2115** and **2116**, corresponding to tripwires **2105** and **2106**, respectively, joined with the "and" combinator **2114** with the object modifier "same" **2119** and temporal modifier "**2115** before **2116**" **2120**. To detect that the same object turned first left then right, the left turn detector **2111** and the right turn detector **2114** are joined with the "and" combinator **2121** with the object modifier "same"

2122 and temporal modifier "**2111** before **2114**" **2123**. Finally, to ensure that the detected object is a vehicle, a Boolean "and" operator **2125** is used to combine the left-and-right-turn detector **2121** and the property query **2124**.

[0136] As an example to illustrate a use of combinators, consider the detection of inserted/left behind objects. Examples of inserted/left behind objects may include, e.g., an unattended bag, a parked car, graffiti, etc. U.S. patent application Ser. Nos. 10/354,096, 11/139,600, and 11/288,200, which are incorporated by reference herein, discuss various methods that may be applied to the detection of a stationary object in video. In certain applications simply detecting the stationary object is sufficient, e.g., in some areas, the existence of any parked vehicle or any unattended package may provide sufficient reason to generate an alert. In some applications, however, such an indiscriminate approach would generate a large number of false alerts. Hence, additional tests and filters may be needed to eliminate such false alerts. A basic stationary object detection algorithm may provide an alert any time a bag is put down and not moved for some period of time. However, for example, in a public waiting area, like a train platform or an airport terminal, putting a bag down and not touching it is normal behavior. The real event of interest in such a scenario may be detection of unattended luggage.

[0137] An exemplary embodiment of defecting unattended luggage is to define an exclusion zone around the stationary object and alerting only if no object of interest (e.g., a person) is inside the exclusion zone. The exclusion zone may mean that there should be nothing in the zone. The exclusion zone may also be interpreted so as to not contain any object of a certain type. For example, one way to establish an exclusion zone may be that there should be no people in the exclusion zone, but there may be other objects, like another bag or a cart. The exclusion gone may also have a time associated with it so that an alert is generated only if the exclusion zone has no object of interest for a prescribed time—this way a person stepping away from the bag just briefly may still be allowed. Conversely, if an object of interest is in the exclusion zone for only a brief period of time, e.g., a person walks by the unattended bag, an alert may still be generated.

[0138] The size of the exclusion zone may be determined in multiple ways. In one exemplary embodiment, if camera calibration information is available, the exclusion zone may be defined as a shape with fixed real-world size around the object, e.g., a one-meter radius circle. As another exemplary embodiment, not requiring calibration, as illustrated in FIG. **22**, the user may specify a larger shape (e.g., a circle, an ellipse or a rectangle) in the near field (**2201**) and a similar smaller shape (i.e., of the same type, e.g., a circle, an ellipse or a rectangle) in the far field of the camera (**2202**), indicating the exclusion zone for an object near the camera (**2203**) and far away from the camera (**2204**), respectively. This may be done, e.g., via a user interface, and may be done, for example, graphically (for example, via a user interface that permits a user to graphically superimpose shapes on video images) or by entering parameters (e.g., shape, size, etc.). The system may then interpolate and extrapolate the exclusion zone for every point in the image from the near field and far field shapes. The interpolation aid extrapolation may be simple linear interpolation, or they may comprise more complex types of interpolation and/or extrapolation. The interpolation and/or extrapolation may also reflect more complex camera projection models, e.g., a quadratic model. The size of the exclusion zone may also be adjusted dynamically, depending on factors like the density of a crowd in the scene (e.g., a smaller exclusion zone for a larger crowd) or the security threat level (e.g., a smaller exclusion zone for higher threat levels). The dynamically resizable exclusion zone can be defined in combination with either of the above-described fixed-size methods. In such a combination, a fixed size exclusion zone may be defined as the base size, and that exclusion zone size may be dynamically increased or decreased according to the scenario.

[0139] In another exemplary embodiment, the size of the exclusion zone may be learned. The learning may be supervised or unsupervised. In an exemplary embodiment, the system may observe and analyze a given video scene for an extended period of time, and may detect all stationary objects and the distances of other objects to the stationary object to learn the typical normal distance between a stationary object and a person putting the object down. Similarly, the duration of time that a person may spend outside of the exclusion zone may also be learned.

[0140] Embodiments of the invention may perform video surveillance in two steps: in the first step the video is analyzed and primitives are generated; in the second step, activity inferencing is performed on the primitives. This allows the user to experiment with the exclusion zone and duration on a prerecorded primitive database to minimize false alerts, while still providing timely alerts.

[0141] Stationary object detection with an exclusion zone may be implemented using a modified "and" operator **241**, combining a stationary target detection activity detector **242** and an area of interest inside activity detector **243** with spatial **244** and temporal **245** modifiers, as illustrated in FIG. **24**.

[0142] The stationary object detection algorithm may also detect the person responsible for leaving the object behind. In such a case, the system may continue to track the person after the stationary object detection. Hence, the system may detect the unattended package when the person is outside of the exclusion zone for a sufficient amount of time, independent of whether there are other people around the package. This approach may even detect a suspicious exchange of a bag or other similar object, e.g., when a first person stops, puts a bag down, a second person stops next to the bag, and the first person walks away, leaving the bag with the second person.

[0143] Another example when the results of stationary object detection may have to be filtered is when a car cannot be left unattended in a certain area. The system may detect an unattended vehicle if a vehicle stops, a person gets out of the vehicle from the driver's seat and that person leaves the area of the vehicle for longer than a predefined time period. In contrast, if nobody gets of the car, or if nobody gets out from the driver's seat, there may be no alert generated, or an alert may be generated only after a longer period of time.

[0144] Another example when the results of stationary object detection may have to be filtered is when the stationary event has to be detected in response to an external event. For example, the user may want to detect stationary bags on a subway platform only when the train leaves. This may be accomplished using a modified "and" operator combining the train departure event and the stationary object event with a temporal modifier. The train departure event may be detected using an external event, e.g., the subway system

sending a "door closing" or a "train departing" signal to the surveillance system, or it may be detected using the video surveillance system and detecting when the train starts moving again or when people stop entering or exiting the train. The external event may be integrated into the system in different ways. In one exemplary embodiment, the external events may be represented using non-video primitives. This information may be stored in the primitive storage **1605**, thus making it available for later off-line processing. In another exemplary embodiment, the external event may be a filter on the alert generations, i.e., it may enable or disable the event response **167**. In another exemplary embodiment, the external event may generate event occurrences, and the event response **167** may combine the occurrences and determine whether to generate an alert.

[0145] A modified "and" operator stores alerts from its sub-queries and combines the results according to the various modifiers. How and which safe-queries are stored may also affect the detection results. Additional modifiers may determine how the sub-queries are stored and used in event detection. One modifier may allow a sub-query result to be part of multiple events or to be part of only one event. Another modifier may limit sub-query storage to only one sub-query event altogether, or to just one sub-query event per target. These modifiers may also determine whether that single sub-query event is the first or the last overall or per target

[0146] Certain applications may require all sub-query results of the detection by the modified "and" operator to be unique, i.e., each sub-event may contribute to only a single event. For example, the goal may be to detect people going straight instead of turning. As illustrated in FIG. **23**, this can be achieved using a rule detecting when the same target crosses tripwire **1** (**2301**) and later tripwire **2** (**2302**) within a prescribed amount of time, i.e., using a modified **2303** should generate an alert when crossing tripwire **2** (**2302**) the first time at **2305**, but when the same target **2303** crosses again at **2306** there should be no alert, even though the second crossing may also be within the prescribed time window from the time of crossing tripwire **1** (**2301**) at **2304**. In contrast, when, e.g., detecting piggybacking, one may want to allow the same sub-query to participate in multiple detections. The rule may detect, e.g., when a card is swiped and subsequently more than one person enters. Such a rule should then generate an alert tor the card swipe and the second person entering, but it should also generate another alert for the same card swipe and the third person entering, etc.

[0147] An application may be to verify that within some predetermined time period of a person exiting an area through a door, the door locks automatically. This may be detected using a modified "and" combinator of an area of interest exit activity detector and a door not locking detector with a temporal modifier. Since several people may leave the area close to each other, it is not guaranteed that the door locks alter every single person exiting. This means that a modifier may be used to store only the last area of interest exit activity detector result

[0148] All these detectors may optionally be combined with temporal attributes. Examples of a temporal attribute include: every 15 minutes; between 9:00 pm and 6:30 am; less than 5 minutes; longer than 30 seconds; and over the weekend.

[0149] In block **24** of FIG. **2**, the video surveillance system is operated. The video surveillance system of the invention operates automatically, detects and archives video primitives of objects in the scene, and detects event occurrences in real time using event discriminators. In addition, action is taken in real time, as appropriate, such as activating alarms, generating reports, and generating output. The reports and output can be displayed and/or stored locally to the system or elsewhere via a network, such as the Internet. FIG. **4** illustrates a flow diagram for operating the video surveillance system.

[0150] In block **41**, the computer system **11** obtains source video from the video sensors **14** and/or the video recorders **15**.

[0151] In block **42**, video primitives are extracted in real time from the source video. As an option, non-video primitives can be obtained and/or extracted from one or more other sensors **17** and used with the invention. The extraction of video primitives is illustrated with FIG. **5**.

[0152] FIG. **5** illustrates a flow diagram for extracting video primitives for the video surveillance system. Blocks **51** and **52** operate in parallel and can be performed in any order or concurrently. In block **51**, objects are detected via movement. Any motion detection algorithm for detecting movement between frames at the pixel level can be used for this block. As an example, the three frame differencing technique can be used, which is discussed in {1}. The detected objects are forwarded to block **53**.

[0153] In block **52**, objects are detected via change. Any change detection algorithm for detecting changes from a background model can be used for this block. An object is detected in this block if one or more pixels in a frame are deemed to be in the foreground of the frame because the pixels do not conform to a background model of the frame. As an example, a stochastic background modeling technique, such as dynamically adaptive background subtraction, can be used, which is described in {1} and U.S. patent application Ser. No. 09/694,712 filed Oct. 24, 2000. The detected objects are forwarded to block **53**.

[0154] The motion detection technique of block **51** and the change detection technique of block **52** are complimentary techniques, where each technique advantageously addresses deficiencies in the other technique. As an option, additional and/or alternative detection schemes can be used for the techniques discussed for blocks **51** and **52**. Examples of an additional and/or alternative detection scheme include the following: the Pfinder detection scheme for finding people as described in {8}; a skin tone detection scheme; a face detection scheme; and a model-based detection scheme. The results of such additional and/or alternative detection schemes are provided to block **53**.

[0155] As an option, if the video sensor **14** has motion (e.g., a video camera that sweeps, zooms, and/or translates), an additional block can be inserted before blocks between blocks **51** and **52** to provide input to blocks **51** and **52** for video stabilization. Video stabilization can be achieved by affine or projective global motion compensation. For example, image alignment described in U.S. patent application Ser. No. 09/609,919, filed Jul. 3, 2000, now U.S. Pat. No. 6,738,424, which is incorporated herein by reference, can be used to obtain video stabilization.

[0156] In block **53**, blobs are generated. In general, a blob is any object in a frame. Examples of a blob include: a moving object, such as a person or a vehicle; and a consumer

product, such as a piece of furniture, a clothing item, or a retail shelf item. Blobs are generated using the detected objects from blocks **32** and **33**. Any technique for generating blobs can be used for this block. An exemplary technique for generating blobs from motion detection and change detection uses a connected components scheme. For example, the morphology and connected components algorithm can be used, which is described in {1}.

[0157] In block **54**, blobs are tracked. Any technique for tracking blobs can be used for this block. For example, Kalman filtering or the CONDENSATION algorithm can be used. As another example, a template matching technique, such as described in {1}, can be used. As a further example, a multi-hypothesis Kalman tracker can be used, which is described in {5}. As yet another example, the frame-to-frame tracking technique described in U.S. patent application Ser. No. 09/694,712 filed Oct. 24, 2000, can be used. For the example of a location being a grocery store, examples of objects that can be tracked include moving people, inventory items, and inventory moving appliances, such as shopping carts or trolleys.

[0158] As an option, blocks **51-54** can be replaced with any detection and tracking scheme, as is known to those of ordinary skill. An example of such a detection and tracking scheme is described in {11}.

[0159] In block **55**, each trajectory of the tracked objects is analyzed to determine if the trajectory is salient. If the trajectory is insalient, the trajectory represents an object exhibiting unstable motion or represents an object of unstable size or color, and the corresponding object is rejected and is no longer analyzed by the system. If the trajectory is salient, the trajectory represents an object that is potentially of interest. A trajectory is determined to be salient or insalient by applying a salience measure to the trajectory. Techniques for determining a trajectory to be salient or insalient are described in {13} and {18}.

[0160] In block **56**, each object is classified. The general type of each object is determined as the classification of the object. Classification can be performed by a number of techniques, and examples of such techniques include using a neural network classifier {14} and using a linear discriminatant classifier {14}. Examples of classification are the same as those discussed for block **23**.

[0161] In block **57**, video primitives are identified using the information from blocks **51-56** and additional processing as necessary. Examples of video primitives identified are the same as those discussed for block **23**. As an example, for size, the system can use information obtained from calibration in block **22** as a video primitive. From calibration, the system has sufficient information to determine the approximate size of an object. As another example, the system can use velocity as measured from block **54** as a video primitive.

[0162] In block **43**, the video primitives from block **42** are archived. The video primitives can be archived in the computer-readable medium **13** or another computer-readable medium. Along with the video primitives, associated frames or video imagery from the source video can be archived. This archiving step is optional; if the system is to be used only for real-time event detection, the archiving step can be skipped.

[0163] In block **44**, event occurrences are extracted from the video primitives using event discriminators. The video primitives are determined in block **42**, and the event discriminators are determined from tasking the system in block **23**. The event discriminators are used to filter the video primitives to determine if any event occurrences occurred. For example, an event discriminator can be looking for a "wrong way" event as defined by a person traveling the "wrong way" into an area between 9:00 a.m. and 5:00 p.m. The event discriminator checks all video primitives being generated according to FIG. **5** and determines if any video primitives exist which have the following properties: a timestamp between 9:00 a.m. and 5:00 p.m., a classification of "person" or "group of people", a position inside the area, and a "wrong" direction of motion. The event discriminators may also use other types of primitives, as discussed above, and/or combine video primitives from multiple video sources to detect event occurrences.

[0164] In block **45**, action is taken for each event occurrence extracted in block **44**, as appropriate. FIG. **6** illustrates a How diagram for taking action with the video surveillance system.

[0165] In block **61**, responses are undertaken as dictated by the event discriminators that detected the event occurrences. The responses, if any, are identified for each event discriminator in block **34**.

[0166] In block **62**, an activity record is generated for each event occurrence that occurred. The activity record includes, for example: details of a trajectory of an object; a time of detection of an object; a position of detection of an object, and a description or definition of the event discriminator that was employed. The activity record can include information, such as video primitives, needed by the event discriminator. The activity record can also include representative video or still imagery of the object(s) and/or area(s) involved in the event occurrence. The activity record is stored on a computer-readable medium.

[0167] In block **63**, output is generated. The output is based on the event occurrences extracted in block **44** and a direct feed of the source video from block **41**. The output is stored on a computer-readable medium, displayed on the computer system **11** or another computer system, or forwarded to another computer system. As the system operates, information regarding event occurrences is collected, and the information can be viewed by the operator at any time, including real time. Examples of formats for receiving the information include: a display on a monitor of a computer system; a hard copy; a computer-readable medium; and an interactive web page.

[0168] The output can include a display from the direct feed of the source video from block **41**. For example, the source video can be displayed on a window of the monitor of a computer system or on a closed-circuit monitor. Further, the output can include source video marked up with graphics to highlight the objects and/or areas involved in the event occurrence. If the system is operating in forensic analysts mode, the video may come from the video recorder.

[0169] The output can include one or more reports for an operator based on the requirements of the operator and/or the event occurrences. Examples of a report include: the number of event occurrences which occurred; the positions in the scene in which the event occurrence occurred; the times at which the event occurrences occurred; representative imagery of each event occurrence; representative video of each event occurrence; raw statistical data; statistics of event occurrences (e.g., how many, how often, where, and when); and/or human-readable graphical displays.

[0170] FIGS. **13** and **14** illustrate an exemplary report for the aisle in the grocery store of FIG. **15**. In FIGS. **13** and **14**, several areas are identified in block **22** and are labeled accordingly in the images. The areas in FIG. **13** match those in FIG. **12**, and the areas in FIG. **14** are different ones. The system is tasked to look for people who stop in the area.

[0171] In FIG. **13**, the exemplary report is an image from a video marked-up to include labels, graphics, statistical information, and an analysis of the statistical information. For example, the area identified as coffee has statistical information of an average number of customers in the area of 2/hour and an average dwell time in the area as 5 seconds. The system determined this area to be a "cold" region, which means there is not much commercial activity through this region. As another example, the area identified as sodas has statistical information of an average number of customers in the area of 15/hour and an average-dwell time in the area as 22 seconds. The system determined this area to be a "hot" region, which means there is a large amount of commercial activity in this region.

[0172] In FIG. **14**, the exemplary report is an image from a video marked-up to include labels, graphics, statistical information, and an analysis of the statistical information. For example, the area at the back of the aisle has average number of customers of 14/hour and is determined to have low traffic. As another example, the area at the front of the aisle has average number of customers of 83/hour and is determined to have high traffic.

[0173] For either FIG. **13** or FIG. **14**, if the operator desires more information about any particular area or any particular area, a point-and-click interface allows the operator to navigate through representative still and video imagery of regions and/or activities that the system has detected and archived.

[0174] FIG. **15** illustrates another exemplary report for an aisle in a grocery store. The exemplary report includes an image from a video marked-up to include labels and trajectory indications and text describing the marked-up image. The system of the example is tasked with searching for a number of areas: length, position, and time of a trajectory of an object; time and location an object was immobile; correlation of trajectories with areas, as specified by the operator; and classification of an object as not a person, one person, two people, and three or more people.

[0175] The video image of FIG. **15** is from a time period where the trajectories were recorded. Of the three objects, two objects are each classified as one person, and one object is classified as not a person. Each object is assigned a label, namely Person ID **1032**, Person ID **1033**, and Object ID **32001**. For Person ID **1032**, the system determined be person spent **52** seconds in the area and 18 seconds at the position designated by the circle. For Person ID **1033**, the system determined the person spent 1 minute and 8 seconds in the area and 12 seconds at the position designated by the circle. The trajectories for Person ID **1032** and Person ID **1033** are included in the marked-up image. For Object ID **32001**, the system did not further analyze the object and indicated the position of the object with an X.

[0176] Referring back to block **22** in FIG. **2**, calibration can be (1) manual, (2) semi-automatic using imagery from a video sensor or a video recorder, or (3) automatic using imagery from a video sensor or a video recorder. If imagery is required, it is assumed that the source video to be analyzed by the computer system **11** is from a video sensor that obtained the source video used for calibration.

[0177] For manual calibration, the operator provides to the computer system **11** the orientation and internal parameters for each of the video sensors **14** and the placement of each video sensor **14** with respect to the location. The computer system **11** can optionally maintain a map of the location, and the placement of the video sensors **14** can be indicated on the map. The map can be a two-dimensional or a three-dimensional representation of the environment. In addition, the manual calibration provides the system with sufficient information to determine the approximate size and relative position of an object.

[0178] Alternatively, for manual calibration, the operator can mark up a video image from the sensor with a graphic representing the appearance of a known-sized object, such as a person. If the operator can mark up an image in at least two different locations, the system can infer approximate camera calibration information.

[0179] For semi-automatic and automatic calibration, no knowledge of the camera parameters or scene geometry is required. From semi-automatic and automatic calibration, a lookup table is generated to approximate the size of an object at various areas in the scene, or the internal and external camera calibration parameters of the camera are inferred.

[0180] For semi-automatic calibration, the video surveillance system is calibrated using a video source combined with input from the operator. A single person is placed in the field of view of the video sensor to be semi-automatic calibrated. The computer system **11** receives source video regarding the single person and automatically infers the size of person based on this data. As the number of locations in the field of view of the video sensor that the person is viewed is increased, and as the period of time that the person is viewed in the field of view of the video sensor is increased, the accuracy of the semi-automatic calibration is increased.

[0181] FIG. **7** illustrates a flow diagram for semi-automatic calibration of the video surveillance system. Block **71** is the same as block **41**, except that a typical object moves through the scene at various trajectories. The typical object can have various velocities and be stationary at various positions. For example, the typical object moves as close to the video sensor as possible and then moves as far away from the video sensor as possible. This motion by the typical object can be repeated as necessary.

[0182] Blocks **72-25** are the same as blocks **51-54**, respectively.

[0183] In block **76**, the typical object is monitored throughout the scene. It is assumed that the only (or at least the most) stable object being tracked is the calibration object in the scene (i.e., the typical object moving through the scene). The size of the stable object is collected for every point in the scene at which it is observed, and this information is used to generate calibration information.

[0184] In block **77**, the size of the typical object is identified for different areas throughout the scene. The size of the typical object is used to determine the approximate sizes of similar objects at various areas in the scene. With this information, a lookup table is generated matching typical apparent sizes of the typical object in various areas in the image, or internal and external camera calibration parameters are inferred. As a sample output, a display of stick-sized figures in various areas of the image indicate

what the system determined as an appropriate height. Such a stick-sized figure is illustrated in FIG. **11**.

[0185] For automatic calibration, a learning phase is conducted where the computer system **11** determines information regarding the location in the field of view of each video sensor. During automatic calibration, the computer system **11** receives source video of the location for a representative period of time (e.g., minutes, hours or days) that is sufficient to obtain a statistically significant sampling of objects typical to the scene and thus infer typical apparent sizes and locations.

[0186] FIG. **8** illustrates a flow diagram for automatic calibration of the video surveillance system. Blocks **81-86** are the same as blocks **71-76** in FIG. **7**.

[0187] In block **87**, trackable regions in the field of view of the video sensor are identified. A trackable region refers to a region in the field of view of a video sensor where an object can be easily and/or accurately tracked. An untrackable region refers to a region in the field of view of a video sensor where an object is not easily and/or accurately tracked and/or is difficult to track. An untrackable region can be referred to as being an unstable or insalient region. An object may be difficult to track because the object is too small (e.g., smaller than a predetermined threshold), appear for too short of time (e.g., less than a predetermined threshold), or exhibit motion that is not salient (e.g., not purposeful). A trackable region can be identified using, for example, the techniques described in {13}.

[0188] FIG. **10** illustrates trackable regions determined for an aisle in a grocery store. The area at the far end of the aisle is determined to be insalient because too many confusers appear in this area. A confuser refers to something in a video that confuses a tracking scheme. Examples of a confuser include: leaves blowing; rain; a partially occluded object; and an object that appears for too short of time to be tracked accurately. In contrast, the area at the near end of the aisle is determined to be salient because good tracks are determined for this area.

[0189] In block **88**, the sizes of the objects are identified for different areas throughout the scene. The sizes of the objects are used to determine the approximate sizes of similar objects at various areas in the scene. A technique, such as using a histogram or a statistical median, is used to determine the typical apparent height and width of objects as a function of location in the scene. In one part of the image of the scene, typical objects can have a typical apparent height and width. With this information, a lookup table is generated matching typical apparent sizes of objects in various areas in the image, or the internal and external camera calibration parameters can be inferred.

[0190] FIG. **11** illustrates identifying typical sizes for typical objects in the aisle of the grocery store from FIG. **10**. Typical objects are assumed to be people and are identified by a label accordingly. Typical sizes of people are determined through plots of the average height and average width for the people detected in the salient region. In the example, plot A is determined for the average height of an average person, and plot B is determined for the average width for one person, two people, and three people.

[0191] For plot A, the x-axis depicts the height of the blob in pixels, and the y-axis depicts the number of instances of a particular height, as identified on the x-axis, that occur. The peak of the line for plot A corresponds to the most common height of blobs in the designated region in the scene and, for this example, the peak corresponds to the average height of a person standing in the designated region.

[0192] Assuming people travel in loosely knit groups, a similar graph to plot A is generated for width as plot B. For plot B, the x-axis depicts the width of the blobs in pixels, and the y-axis depicts the number of instances of a particular width, as identified on the x-axis, that occur. The peaks of the line for plot B correspond to the average width of a number of blobs. Assuming most groups contain only one person, the largest peak corresponds to the most common width, which corresponds to the average width of a single person in the designated region. Similarly, the second largest peak corresponds to the average width of two people in the designated region, and the third largest peak corresponds to the average width of three people in the designated region.

[0193] FIG. **9** illustrates an additional flow diagram for the video surveillance system of the invention. In this additional embodiment, the system analyzes archived video primitives with event discriminators to generate additional reports, for example, without needing to review the entire source video. Anytime after a video source has been processed according to the invention, video primitives for the source video are archived in block **43** of FIG. **4**. The video content can be reanalyzed with the additional embodiment in a relatively short time because only the video primitives are reviewed and because the video source is not reprocessed. This provides a great efficiency improvement over current state-of-the-art systems because processing video imagery data is extremely computationally expensive, whereas analyzing the small-sized video primitives abstracted from the video is extremely computationally cheap. As an example, the following event discriminator can be generated: "The number of people stopping for more than 10 minutes in area A in the last two months." With the additional embodiment, the last two months of source video does not need to be reviewed. Instead, only the video primitives from the last two months need to be reviewed, which is a significantly more efficient process.

[0194] Block **91** is the same as block **23** in FIG. **2**.

[0195] In block **92**, archived video primitives are accessed. The video primitives are archived in block **43** of FIG. **4**.

[0196] Blocks **93** and **94** are the same as blocks **44** and **45** in FIG. **4**.

[0197] As an exemplary application, the invention can be used to analyze retail market space by measuring the efficacy of a retail display. Large sums of money are injected into retail displays in an effort to be as eye-catching as possible to promote sales of both the items on display and subsidiary items. The video surveillance system of the invention can be configured to measure the effectiveness of these retail displays.

[0198] For this exemplary application, the video surveillance system is set up by orienting the field of view of a video sensor towards the space around the desired retail display. During tasking, the operator selects an area representing the space around the desired retail display. As a discriminator, the operator defines that he or she wishes to monitor people-sized objects that enter the area and either exhibit a measurable redaction in velocity or stop for an appreciable amount of time.

[0199] After operating for some period of time, the video surveillance system can provide reports for market analysis. The reports can include: the number of people who slowed

down around the retail display; the number of people who stopped at the retail display; the breakdown of people who were interested in the retail display as a function of time, such as how many were interested on weekends and how many were interested in evenings; and video snapshots of the people who showed interest in the retail display. The market research information obtained from the video surveillance system can be combined with sales information from the store and customer records from the store to improve the analysts understanding of the efficacy of the retail display.

[0200] The embodiments and examples discussed herein are non-limiting examples.

[0201] The invention is described in detail with respect to various embodiments, and it will now be apparent from the foregoing to those skilled in the art that changes and modifications may be made without departing from the invention in its broader aspects, and the invention, therefore, as defined in the claims is intended to cover all such changes and modifications as fall within the true spirit of the invention.

1-20. (canceled)

21. A method of video surveillance, comprising:

receiving, by a computer system, a video comprising video images from a video sensor,

the computer system performing the steps of:

analyzing the video to detect stationary objects in the video;

analyzing the video to detect people in the video;

upon detecting a first stationary object in the video, defining a zone as a portion of the video around the first stationary object, defining the zone being responsive to a location of the first stationary object, the zone being defined to be larger than an outer boundary of the first stationary object and smaller than a field of view of the video, a size of the zone allowing detection of multiple non-overlapping objects;

tracking a duration that the first stationary object remains stationary; and

issuing an alert in response to determining that the duration of time the first stationary object has remained stationary exceeds a threshold while no person of interest detected in the video has been inside the zone.

22. The method of claim **21**, wherein the person of interest comprises any person detected in the video.

23. The method of claim **21**, further comprising detecting a person leaving the first stationary object as the person of interest.

24. The method of claim **21**, wherein the size of the zone is determined prior to the detecting of the first stationary object.

25. The method of claim **21**, wherein the person of interest comprises a person detected as leaving the first stationary object and detected as having been in the zone longer than a first period of time.

26. The method of claim **21**, wherein the size of the zone is defined in an image space that varies as a function of the location of the first stationary object.

27. The method of claim **21**, wherein the zone is defined as a shape in the real-world of the video independent of detecting the first stationary object.

28. The method of claim **21**, wherein the size of the zone is determined by performing an interpolation or an extrapolation of a first shape.

29. The method of claim **21**, further comprising permitting a user to set, via a user interface, at least one parameter selected from the group consisting of: the size of the zone, a shape of the zone, and the duration of time.

30. A method of video surveillance, comprising:

receiving, by a computer system, a video comprising video images from a video sensor,

the computer system performing the steps of:

analyzing the video to detect stationary objects in the video;

analyzing the video to detect people in the video;

upon detecting a first stationary object in the video, defining a zone as a portion of the video around the first stationary object, defining the zone being responsive to a location of the first stationary object, the zone being defined to be larger than an outer boundary of the first stationary object and smaller than a field of view of the video, a size of the zone allowing detection of multiple non-overlapping objects;

tracking a duration that the first stationary object remains stationary;

issuing an alert in response to the determining that the duration of time the first stationary object has remained stationary exceeds a threshold while at least one of the following occurs:

no person of interest detected in the video has been detected as being in the zone, and

no person of interest detected in the video has been detected as being in the zone longer than a first period of time.

31. The method of claim **30**, wherein the zone is one of a circle, ellipse or a rectangle.

32. The method of claim **30**, wherein the size of the zone is responsive to a selection by a user.

33. The method of claim **30**, further comprising permitting a user to set, via a user interface, at least one parameter selected from the group consisting of: the size of the zone, a shape of the zone, and the duration of time.

34. A method of video surveillance, comprising:

receiving, by a computer system, a video comprising video images from a video sensor,

the computer system performing the steps of:

analyzing the video to detect stationary objects in the video;

analyzing the video to detect people in the video;

upon detecting a first stationary object in the video, defining a zone as a portion of the video around the first stationary object, defining the zone being responsive to a location of the first stationary object, the zone being defined to be larger than an outer boundary of the first stationary object and smaller than a field of view of the video, a size of the zone allowing detection of multiple non-overlapping objects;

tracking a duration that the first stationary object remains stationary;

issuing an alert in response to determining that the duration of time the first stationary object has remained stationary exceeds a threshold while no person of interest detected in the video has been inside the zone; and

determining a security threat level,

wherein the size of the zone is responsive to the determined security threat level.

35. A method of video surveillance, comprising:
receiving, by a computer system, a video comprising video images from a video sensor,
the computer system performing the steps of:
analyzing the video to detect stationary objects in the video;
analyzing the video to detect people in the video;
upon detecting a first stationary object in the video, defining a zone as a portion of the video around the first stationary object, defining the zone being responsive to a location of the first stationary object, the zone being defined to be larger than an outer boundary of the first stationary object and smaller than a field of view of the video, a size of the zone allowing detection of multiple non-overlapping objects;
tracking a duration that the first stationary object remains stationary;
determining an occurrence of an external event; and
issuing an alert in response to determining that the duration of time the first stationary object has remained stationary exceeds a threshold while no person of interest detected in the video has been inside the zone and the occurrence of the external event,
wherein the size of the zone is responsive to detection of other stationary objects prior to the detecting of the first stationary object.

36. The method of claim **35**, further comprising:
analyzing the video to determine distances between other stationary objects and people putting down a corresponding one of the other stationary objects prior to the detecting of the first stationary object,
wherein the size of the zone is responsive to the analyzing of the video to determine distances between the other stationary objects and the people putting down the corresponding one of the other stationary objects.

37. The method of claim **35**, further comprising:
analyzing the video to determine durations of time people put down other objects prior to the detecting of the first stationary object,
wherein the duration of time is determined in response to the analyzing of the video to determine the durations of time people put down the other objects.

38. The method of claim **35**, further comprising:
tracking a person responsible for leaving behind the first stationary object.

39. A method of video surveillance, comprising:
receiving, by a computer system, a video comprising video images from a video sensor,
the computer system performing the steps of:
analyzing the video to detect stationary objects in the video;
analyzing the video to detect people in the video;
upon detecting a first stationary object in the video, defining a zone as a portion of the video around the first stationary object, defining the zone being responsive to a location of the first stationary object, the zone being defined to be larger than an outer boundary of the first stationary object and smaller than a field of view of the video, a size of the zone allowing detection of multiple non-overlapping objects;
tracking a duration that the first stationary object remains stationary;
determining an occurrence of an external event;
issuing an alert in response to the determining of the occurrence of the external event and determining that the duration of time the first stationary object has remained stationary exceeds a threshold while, during a first period of time, a person of interest detected in the video has been detected as being in the zone only for part of the first period of time; and
wherein the size of the zone is determined dynamically and is responsive to a crowd density.

40. The method of claim **39**, further comprising permitting a user to set, via a user interface, at least one parameter selected from the group consisting of: the size of the zone, a shape of the zone, and the duration of time.

* * * * *