(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 February 2019 (28.02.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/040659 A1

(51) International Patent Classification:
G06T 7/00 (2017.01) *G08B 13/00* (2006.01)
G07G 3/00 (2006.01) *G08B 13/18* (2006.01)

(21) International Application Number:
PCT/US2018/047591

(22) International Filing Date:
22 August 2018 (22.08.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/549,368 23 August 2017 (23.08.2017) US

(71) Applicant: **BOSSA NOVA ROBOTICS IP, INC.** [US/US]; 2740 Smallman St, Suite 400, Pittsburgh, PA 15222 (US).

(72) Inventors: **SAVVIDES, Marios**; 2740 Smallman St., Suite 400, Pittsburgh, PA 15222 (US). **LIN, An, Pang**; 2740 Smallman St, Suite 400, Pittsburgh, PA 15222 (US). **VENUGOPALAN, Shreyas**; 2740 Smallman St, Suite 400, Pittsburgh, PA 15222 (US). **THANIKKAL, Ajmal**; 2740 Smallman St, Suite 400, Pittsburgh, PA 15222 (US). **SINGH, Karanhaar**; 2740 Smallman St, Suite 400, Pittsburgh, PA 15222 (US). **ADLER, Gavriel**; 2740 Smallman St, Suite 400, Pittsburgh, PA 15222 (US). **NEBLETT, Kyle**; 2740 Smallman St, Suite 400, Pittsburgh, PA 15222 (US). **PALMER, Christopher**; 2740 Smallman St, Suite 400, Pittsburgh, PA 15222 (US).

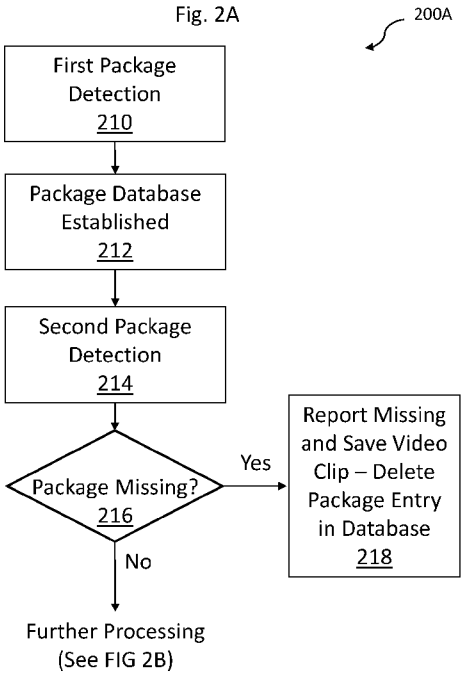
(74) Agent: **STEVENS, David, R.**; Stevens Law Group, 1754 Technology Drive, Suite 226, San Jose, CA 95110 (US).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,

(54) Title: METHOD FOR NEW PACKAGE DETECTION

Fig. 2A



(57) Abstract: A method for processing arrival or removal of packages within the field of view of a video camera includes providing a database for recording packages placed in the field of view. Based on real-time analysis of successive image frames in the camera, a human person's entry and exit from the field of view of the camera is also detected. Delivery or removal of objects is recorded in the database. In one embodiment, the method also determines whether or not a newly arrived package is placed alongside or on top of an existing package.

WO 2019/040659 A1

[Continued on next page]

UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

METHOD FOR NEW PACKAGE DETECTION

Cross -Reference to Related Patent Application

[0001] The present disclosure is part of a non-provisional patent application claiming the priority benefit of U.S. Patent Application No. 62/549,368, titled "Method For New Package Detection", filed August 23, 2017.

Field of the Invention

[0002] This invention relates generally to security surveillance technology. In particular, the present invention relates to security surveillance system which efficiently uses a video camera for object recognition locally and remote network resources for registering and monitoring events and other security applications.

Background

[0003] Existing camera-equipped home security systems are expensive, requiring both remote servers to process the captured images and unnecessary human intervention. For example, in a typical object detection application, an object is detected in an image by classifying image patches across various shifts and across a variety of image scales. Specifically, if a pedestrian is to be detected, a typical system first generates a scale-space image pyramid to obtain images at various resolutions and to detect the pedestrian in all possible fixed-size windows across these images. Typically, great computational cost must be incurred to allow a system to robustly determine if an object is in an image and to classify objects at all locations and at all resolutions.

[0004] In many applications, it is desired that the video camera carry out local object detection functions to more efficiently use computational resources, and thereby to decrease costs and to increase system capacity.

Summary

[0005] According to one embodiment of the present invention, a method for processing arrival or removal of an object within the field of view of a video camera includes: (a) providing a database for recording objects residing in the field of view; (b) based on real-time analysis of successive image frames in the camera, detecting a human person's entry into the field of view of the camera; (c) saving a first set of image frames, corresponding image frames captured by the video camera immediately prior to the detection of the human person's entry into the field of view of the camera; (d) based on real-time analysis of successive image frames in the camera, detecting a human person's exit from the field of view of the camera; (e) saving a second set of image frames, corresponding to image frames captured by the video camera immediately after the detection of the human person's exit from the field of view of the camera; (f) comparing the first set of image frames with the second set of image frames to identify one or more objects that appear or disappear from the field of view based on difference between the first and the set of image frames; (g) determining and eliminating objects identified by comparing the first and second sets of image frames as artifacts; (h) determining from the objects remaining and the objects residing in the database: (1) whether or not one of the objects residing in the database has been removed; and (2) whether or not an object has newly arrived; and (i) entering any newly arrived objects into the database and removing any objects residing in the database that have been determined to have been removed. In one embodiment, the method also determines whether or not the newly arrived object is placed alongside or on top of an existing object residing in the database. The objects may be, for example, packages that are delivered or collected.

[0006] According to one embodiment of the present invention, the artifacts may correspond, for example, to change in lighting conditions, change in lighting conditions due to the opening or shutting of a door or change in other conditions.

[0007] According to one embodiment of the present invention, a block-based discrete cosine transform on each of the first set of image frames and the second set of image frames is computed to determine therefrom a change in spatial frequencies.

[0008] According to one embodiment of the present invention, differences in light intensities between corresponding pixels of the first and the second sets of image frames are computed. In one embodiment, accumulated differences over predetermined corresponding regions are computed from the computed differences in light intensity.

[0009] According to one embodiment of the present invention, determining from the objects remaining and the objects residing in the database includes computing one or more intersections between each object remaining and each object residing in the database. The intersections may be, for example, an intersection in shape between objects or an intersection in location between objects.

[0010] In one embodiment, further identification between the newly identified objects and the existing objects in the database comprises comparing color histograms between the currently identified objects and the existing objects in the database.

[0011] In another embodiment, a method for package detection includes the steps of providing a database for recording packages in the field of view. Based on real-time analysis of successive image frames in the camera, a human person's entry and exit from a field of view of the camera can be detected. After the human person's entry and exit, whether packages have been removed or added is determined by computing one or more intersections between each

package remaining and each package residing in the database. Added packages are entered into the database and information relating to removed packages residing in the database are removed. In some embodiments, a change detection module and a bounding box filtering module can be used to artifacts and/or eliminate bounding boxes unlikely to surround packages.

[0012] In another embodiment, a system for processing arrival or removal of an object within the field of view of a video camera can include a database module for recording objects residing in the field of view. A detection module is used to determine a human person's entry into the field of view of the camera based on real-time analysis of successive image frames in the camera, with the detection module saving a first set of image frames corresponding to image frames captured by the video camera immediately prior to the detection of the human person's entry into the field of view of the camera and saving a second set of image frames corresponding to image frames captured by the video camera immediately after the detection of the human person's exit from the field of view of the camera. A comparison module can be used to compare the first set of image frames with the second set of image frames to identify one or more objects that appear or disappear from the field of view based on differences between the first and the set of image frames. An artifact elimination module can be used to determine and eliminate objects identified by comparing the first and second sets of image frames as artifacts. In operation, the database module records (a) whether or not one of the objects residing in the database has been removed; (b) whether or not an object is newly arrived; and c) enters any newly arrived objects into the database and removes any objects residing in the database that have been determined to have been removed.

Brief Description of the Drawings

[0014] The specific features, aspects and advantages of the present invention will become better understood with regard to the following description and accompanying drawings where:

[0015] FIG. 1 illustrates a camera system suitable for identification of packages left or removed; and

[0016] FIGS. 2A, 2B, and 2C collectively form a flow chart illustrating the operations of the processor in the camera for identification of packages left or removed

DETAILED DESCRIPTION

[0017] A security surveillance system which efficiently uses the computational power of a processor in a video camera to perform object detection and identification is disclosed in U.S. Patent 9,928,708, serial no. 14/968,714, entitled "Real-time Video Analysis for Security Surveillance," filed on December 14, 2015, the disclosure of which is hereby incorporated by reference in its entirety. In that security surveillance system, the local video camera system performs real-time detection and identification of persons and objects (e.g., a package left by a courier) within the field of view of the video camera.

[0018] As seen in FIG. 1, a camera system 100 that can be used to implement the disclosed system and method can be an "IP camera" (i.e., a camera capable of communicate using the Internet Protocol), a USB camera (i.e., a camera having a Universal Serial Bus interface), or a camera capable of communicating with another device by any suitable means. As shown in FIG. 1, camera 100 includes processor 101, memory 102, image sensor 103, communication interfaces 104, camera optical and actuator system 105, and storage 106. Processor 101 controls the overall operations of the camera, such as operating camera optical and actuator system 105, and communication interfaces 104. Camera optical and actuator system 105 controls the operations of the camera, such as exposure control for image captured at image sensor 103. Camera optical and actuator system 105 may include a fixed lens system or an adjustable lens system (e.g., zoom and automatic focusing capabilities). Camera can be fixedly or movably mounted to point downwardly or upwardly. In certain embodiments a wide field or fish-eye camera can be used. Alternatively, a narrow field, zoomable telephoto camera or a high resolution, tilt controllable, height adjustable camera can be used.

[0019] In some embodiments, camera optical and actuator system 105 may include or be associated with an infrared (IR) illumination system (e.g., IR light from light-emitting diodes (LED)) or any suitable illumination system for supplementing light when an environment does not provide sufficient illumination (e.g., at night)). In certain embodiments multiple cameras and/or LEDs can be used to reduce glare from highly reflective surfaces, since multiple cameras pointed in slightly different directions can result in at least one image with little or no glare. Lights can be mounted along with, or separately from, the sensors, and can include monochromatic or near monochromatic light sources such as lasers, light emitting diodes (LEDs), or organic light emitting diodes (OLEDs). Broadband light sources may be provided by multiple LEDs of varying wavelength (including infrared or ultraviolet LEDs), phosphor supported white LEDs, halogen lamps or another suitable conventional light source. Various spectral filters that may include narrowband, wideband, or polarization filters and light shields, lenses, mirrors, reflective surfaces, diffusers, concentrators, or other optics can provide wide light beams for area illumination or tightly focused beams for improved local illumination intensity.

[0020] According to some embodiments, both cameras and lights can be movably mounted. For example, hinged, rail, electromagnetic piston, or other suitable actuating mechanisms used to programmatically rotate, elevate, depress, oscillate, or laterally or vertically reposition cameras or lights.

[0021] Image sensor 103 captures images under control of processor 101 from light from the environment entering camera 100. The camera is capable of capturing the images from the environment. Sensor 103 may provide a color image or a gray-scale image. In certain embodiments, conventional RGB CMOS or CCD sensors can be used, alone or in combination

with spectral filters that may include narrowband, wideband, or polarization filters. Embodiments can also include sensors capable of detecting infrared, ultraviolet, or other wavelengths to allow for hyperspectral image processing. This can allow, for example, monitoring and tracking of markers, labels or guides that are not visible to people, or using flashing light in the invisible spectrum to reduce energy consumption and motion blur.

[0022] Communication interfaces 104 typically include one or more communication interfaces (e.g., a network interface, a USB interface) which allows image data to be transferred from storage 106 to a communicating external device (e.g., a computer). Storage 106 provides non-volatile storage (e.g., archived images and software). Memory 102 provides run-time memory support for processor 101, such as frame buffers for image processing operations

[0023] According to one embodiment of the present invention, memory 102 may be allocated to include multiple memory spaces. such as a manufacturer's memory space, a developer's memory space, and a user memory space. The manufacturer's memory space may be provided for system software provided by the camera manufacturers, such as firmware for operating camera system 100. The user memory space may be used, for example, for allocating frame buffers for image processing. Frame buffers are typically allocated for holding image data captured by image sensor 103. Such image data may include, for example, frame buffers holding consecutive frames of images. In one embodiment, at least 50 megabytes are allocated for frame buffers in the user memory space for two more consecutive frames of images. The developer's memory space may be used, for example, for holding software modules executed by processor 101 for carrying out a system or a method of the present invention.

[0024] According to one embodiment of the present invention illustrated with respect to FIGS 2A, 2B, and 2C, a database in the video camera stored in a memory module accessible by

the processor is initialized to record packages that have been detected within the video camera's field of view. Initially, the database contains no packages. Upon initialization, using the techniques disclosed in U.S. Patent 9,928,708, the processor detects the arrival of the first package. For example, video frames from camera system 100 can be sent to a change detection module. Bounding boxes are created to surround potential objects. Instead of immediately classifying objects in the bounding boxes, a bounding box filtering module can be used to eliminate bounding boxes unlikely to surround objects of interest. The remaining bounding boxes can then have contained objects classified and/or identified in a filtered detection step.

[0025] In operation, the change detection module takes in a raw frames and subjects them to image preprocessing that can include, but is not limited to, blurring, sharpening, downsizing, upsizing, morphological operations including rotation or stretching, and color or wavelength filtering. In some embodiments, color depth of video frame data in the frame buffers can be reduced, for example by conversion color to gray-scale. Such a conversion may be carried out in one of numerous ways. For example, image data may be represented by the RGB, the RGBA, or one of many industry standard image data formats under various color space representations. Typically, these data formats are expressed in multiple data channels, with one of the channels used to represent intensity. Alternatively, the intensity of each pixel may be derived from the image data in two or more of the channels. For example, under the RGB format, the intensity is a known weighted average of all three channels. Conversion from image data in industry standard image data to grey-scale data based on intensity is known to those skilled in the art. (Some cameras also provide a separate gray-scale channel for their color pictures.) In some embodiments, down-sampling by 4 (e.g., reduction by a factor of 2 in each dimension of the 2-

dimensional image) is achieved by simply sampling every other pixels in each dimension, which enables a potential approximate speedup by a factor of 4.

[0026] After preprocessing of the video frames, bounding boxes corresponding to recent changes in the scene can be determined. These changes correspond to both valid moving objects and false detections or noise. In one embodiment, an object of interest segmentation algorithm can use a background differentiation approach in order to estimate new objects that have entered the scene. Such an algorithm utilizes the difference between consecutive frames to identify moving objects in the scene. This difference image is then thresholded to determine bounding boxes for potential objects. Since the algorithm does not need to model the background directly, it responds quickly to changes.

[0027] The bounding box filtering module performs filtering based on the bounding box properties to remove false detections and keep valid detections. Bounding box classification is necessary in order to determine whether the detected region corresponds to a valid detected object or to irrelevant changes not caused by moving objects (e.g. lighting changes). The algorithm can be trained off-line to learn the decision boundary between valid and invalid objects based on training samples. The features used to discriminate can include combinations of size, aspect ratio, and location of the bounding box within the frame. The feature vector is then fed into a decision tree classification component, which rejects or accepts the bounding box.

[0028] Advantageously, computation of any features from actual pixel data is not required by the bounding box filtering module, allowing the filtering to be decoupled from the change detection module. This feature, along with use of a tree classification procedure that only requires object detection if needed, greatly improves overall computation speed. In effect, valid moving objects can be detected, identified, and tracked against a variety of background by first

filtering out nearly all invalid detections such as plant motions, environmental noise, and sudden lighting changes. The described system requires no on-device learning time or adjustment period, unlike conventional background modeling-based approaches.

[0029] After eliminating invalid detections and/or bounding boxes, valid detections can be assumed to include packages left or removed by human beings coming into and leaving the camera's field of view. Accordingly, the processor related to the video camera replaces a set of image frames that includes a background model or a "region of interest" at the rate decided by the camera capacity and program pre-set (such as by a non-limiting example of 2-5 frames per second). The set of images are analyzed in real-time using known techniques (e.g., characteristic speeds and aspect ratios) to detect the appearance of a human person.

[0030] Upon detecting a human appearing in the captured images, (e.g., at time "t-entry"), the processor saves the set of one or more image frames ("M-entry") corresponding to the background model captured immediately prior to entry of the human. M-entry serves as the set of baseline image frames that allow further processing. The processor then analyzes successive captured images under a "pedestrian detection model" to determine whether or not the previously detected human has left the scene. Within a predetermined time interval immediately after detecting the human leaving the scene (e.g., at time "t-exit"), a set of one or more image frames ("M-exit") of the same scene is captured.

[0031] The processor computes an intensity difference ("M-difference") between the sets of image frames M-entry and M-exit. A thresholding step may then be performed to ensure insignificant intensity differences are eliminated from M-differences. The processor may also sum all the individual M-differences within predetermined regions (e.g., 10 pixels by 10 pixels blocks) to generate the resulting "accumulated M-difference". The processor then filters the

accumulated M-differences to remove the intensity differences due to change in lighting. In one embodiment, a block-based discrete cosine transform (DCT) is computed for each pixel block in M-entry and M-exit. Spatial frequency changes obtained by subtracting corresponding blocks after DCT in M-entry and M-exit are examined. Accumulated M-differences where the spatial frequency changes are less than a predetermined value may be zeroed.

[0032] In embodiments where the camera's field of view includes an entry way of a building, lighting changes resulting from door opening and shutting conditions (e.g., a person leaving the building or entering the building) affect the image frames in M-entry and M-exit. Effects due to door conditions must be learned, identified and eliminated from the accumulated M-differences. For example, in one embodiment, a detected package candidate (described below) having 15% or more of its expected image area outside the region of the interest is removed as a door condition artifact.

[0033] After filtering for known artifacts, the accumulated M-differences are further processed to identify package candidates in a first step 210 of method 200A as illustrated in FIG. 200. A package is established in a database (step 212) and second or additional packages can also be established in the database (step 214). Where no similarity in shape is found matching the newly identified package with an existing package in the database, the processor may recognize a newly arrived package and new entry is added to the database. The user may also be appropriately notified about the newly arrived package.

[0034] Missing package detection is also possible (step 216). If a package in the database is removed without authorization or by identified users, the package can be reported as missing and relevant video clips saved for analysis or report to police or other parties. A missing package can be deleted from the database (step 218). In one embodiment the entry in the database is removed

and the user is notified of the removal using the same means as that used for informing the user of an arrival of a package. In another embodiment, a video clip of the event (i.e., the image sequence between times t-entry and t-exit) can be saved locally at the camera or remotely in a file server connected to the camera by a computer network. The video clip can also be transmitted to the user. If the package is not missing, further processing is possible as seen with respect to FIG. 2B.

[0035] In some embodiments, package candidates are identified from all the regions having a high frequency texture change between M-exit and M-entry. Each package candidate may be identified by a set of spatial coordinates. For example, in one embodiment, each package candidate is represented by a sequence of spatial coordinates (e.g. (x1, y1), (x2, y2), (x3, y3), ….). The processor calculates the size of the region corresponding to the package candidate from the sequence of spatial coordinates. The process may remove package candidates having sizes that do not exceed a threshold. The processor may also eliminate any package candidate having aspect ratio and occupancy ratios that are outside the respective threshold values to eliminate the effects of shadows and the effects of irregular shape packages, respectively. The package candidates remaining after the various filters above are then recorded in the local database (i.e., inside the camera) as packages. If the database is empty, the first package identified is presumed the first detected package since initialization. The arrival of the package may also be announced through any appropriate means (e.g., through an email server accessed over the internet or a wireless communication network).

[0036] In one embodiment, each package is represented by its sequence of spatial coordinates, the associated image frame sets M-entry and M-exit, and a unique assigned identification number.

[0037] For each package that is currently identified at a time when the database is not empty, each currently identified package is matched to the package or packages already in the database. Specifically, as seen with respect to method 200B of FIG. 2B, the processor looks for intersection between currently identified packages with packages already recorded in the database based on their respective spatial coordinates (step 220). Each currently identified package is examined under two different types of intersection: intersection in shape (i.e. similarity in shape "SIS") and intersection in location. An intersection in shape pertains to size and shape of the package. An intersection in location pertains to the space occupied by the comparing packages (i.e. significant in overlap "SIO"). In addition to the intersection, the processor may also compare their respective color histograms, and the spatial frequency changes between M-exit with the M-entry. Other features may also be used to help match previously recorded packages with the currently identified packages. System parameters governing the intersection (e.g., thresholds of the intersection) are specified by the user or the administrator or learned empirically using any appropriate learning technique (step 222).

[0038] When a currently identified package has zero (or below threshold) SIS or SIO intersections with the packages already in the database (step 226), the package is determined to be a new arrival. The package will be entered into the database in the same manner as described above for the arrival of the first package, after being assigned the unique identification number. The user may also be informed appropriately. When some degree of intersection in shape or in location is found with a package already in the database, the processor calculates a figure-of-merit for each of the intersection types. When both shape and location intersections between a currently identified package and a package in the database are found to be significant (i.e., significant similarity in shape and significant overlap in location), other similarity features are

compared (e.g., their respective color histograms and spatial frequency changes). If these features are consistent and only a change in location is found, the processor may determine that an existing package has shifted in location (step 226). The location for the existing package can examined to ensure that the previous package is no longer associated with the previous sequence of spatial coordinates. In that case, the sequence of spatial coordinates in the database associated with the existing package is updated.

[0039] Alternatively or in addition, as seen with respect to method 200C of FIG. 2C, a determination can be made as to whether a database registered package has been left (step 230) when SIS is below threshold and SIO is above threshold. A determination can be made as to whether an old package has shifted location (step 232), in which case the database is updated for the shifted location (step 234).

[0040] Alternatively, based on verified features or differences found (step 236), the processor may determine that a newly arrived package is placed on top of an existing package (step 238). The newly arrived package is entered in the database in substantially the same manner as the first arrival package. In one embodiment, a convex hull of the combination of regions occupied by the newly identified and the existing package is also recorded in the database. For both the newly identified package and the existing package, the associated M-entry is the set of images frames saved at time t-entry of the existing package, and the M-exit is the set of image frames saved at time t-exit of the newly identified package. Alternatively, in one embodiment, the two packages are represented as a single package and the database updated (step 240). The user can be appropriately notified about the newly arrived package.

[0041] As will be understood, the camera system and methods described herein can operate locally or in via connections to either a wired or wireless connect subsystem for interaction with

devices such as servers, desktop computers, laptops, tablets, or smart phones. Data and control signals can be received, generated, or transported between varieties of external data sources, including wireless networks, personal area networks, cellular networks, the Internet, or cloud mediated data sources. In addition, sources of local data (e.g. a hard drive, solid state drive, flash memory, or any other suitable memory, including dynamic memory, such as SRAM or DRAM) that can allow for local data storage of user-specified preferences or protocols. In one particular embodiment, multiple communication systems can be provided. For example, a direct Wi-Fi connection (802.11b/g/n) can be used as well as a separate 4G cellular connection.

[0042] Connection to remote server embodiments may also be implemented in cloud computing environments. Cloud computing may be defined as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned via virtualization and released with minimal management effort or service provider interaction, and then scaled accordingly. A cloud model can be composed of various characteristics (e.g., on-demand self-service, broad network access, resource pooling, rapid elasticity, measured service, etc.), service models (e.g., Software as a Service ("SaaS"), Platform as a Service ("PaaS"), Infrastructure as a Service ("IaaS"), and deployment models (e.g., private cloud, community cloud, public cloud, hybrid cloud, etc.).

[0043] In certain embodiments, processing related to one or more of video frame preprocessing, motion detection, change detection, bounding box creation, filtering of potential objects, or classification/identification can be performed at a remote server. For example, local processing can be used for video frame preprocessing, motion detection, change detection, bounding box creation, filtering of potential objects, package database detection and updates.

Computationally difficult package or person classification/identification not locally processed can be done using the remote server or other cloud computing facility.

[0044] Many modifications and other embodiments of the invention will come to the mind of one skilled in the art having the benefit of the teachings presented in the foregoing descriptions and the associated drawings. Therefore, it is understood that the invention is not to be limited to the specific embodiments disclosed, and that modifications and embodiments are intended to be included within the scope of the appended claims. It is also understood that other embodiments of this invention may be practiced in the absence of an element/step not specifically disclosed herein.

CLAIMS

1. A method for processing arrival or removal of an object within the field of view of a video camera, comprising:

providing a database for recording objects residing in the field of view;

based on real-time analysis of successive image frames in the camera, detecting a human person's entry into the field of view of the camera;

saving a first set of image frames, corresponding image frames captured by the video camera immediately prior to the detection of the human person's entry into the field of view of the camera;

based on real-time analysis of successive image frames in the camera, detecting a human person's exit from the field of view of the camera;

saving a second set of image frames, corresponding to image frames captured by the video camera immediately after the detection of the human person's exit from the field of view of the camera;

comparing the first set of image frames with the second set of image frames to identify one or more objects that appear or disappear from the field of view based on difference between the first and the set of image frames;

determining and eliminating objects identified by comparing the first and second sets of image frames as artifacts;

determining from the objects remaining and the objects residing in the database: (a) whether or not one of the objects residing in the database has been removed; and (b) whether or not an object has newly arrived; and

entering any newly arrived objects into the database and removing any objects residing in the database that have been determined to have been removed.

2. The method of claim 1, wherein the objects comprise packages that are delivered or collected.

3. The method of claim 1, wherein the artifacts correspond to change in lighting conditions.

4. The method of claim 1, wherein the artifacts correspond to change in lighting conditions due to the opening or shutting of a door.

5. The method of claim 1, further comprising determining whether or not the newly arrived object is placed alongside or on top of an existing object residing in the database.

6. The method of claim 1, wherein comparing the first and second sets of image frames comprises computing a block-based discrete cosine transform on each of the first set of image frames and the second set of image frames to determine therefrom a change in spatial frequencies.

7. The method of claim 1, wherein comparing the first and second sets of image frames comprises computing differences in light intensity between corresponding pixels in the first set of image frames and the second set of image frames.

8. The method of claim 7, further comprising computing accumulated differences for predetermined corresponding regions in the computed differences in light intensity.

9. The method of claim 1, wherein determining from the objects remaining and the objects residing in the database comprises computing one or more intersections between each object remaining and each object residing in the database.

10. The method of claim 9, wherein the computed intersections comprise an intersection in shape between objects.

11. The method of claim 9, wherein the computed intersections comprise an intersection in location between objects.

12. The method of claim 1, wherein the determining from the objects remaining and the objects residing in the database comprises comparing color histograms.

13. A method for package detection, comprising:

providing a database for recording packages in the field of view;

based on real-time analysis of successive image frames in the camera, detecting a human person's entry and exit from a field of view of the camera;

after the human person's entry and exit, determining whether packages have been removed or added by computing one or more intersections between each package remaining and each package residing in the database; and

entering any added packages into the database and removing packages residing in the database that have been determined to have been removed.

14. The method of claim 13, wherein the computed intersections comprise an intersection in shape between objects.

15. The method of claim 13, wherein the computed intersections comprise an intersection in location between objects.

16. The method of claim 13, further comprising determining whether or not the newly arrived object is placed alongside or on top of an existing object residing in the database.

17. The method of claim 13, wherein comparing the first and second sets of image frames comprises computing a block-based discrete cosine transform on each of the first set of image frames and the second set of image frames to determine therefrom a change in spatial frequencies.

18. The method of claim 13, further comprising use of a change detection module to define a bounding box around an object and determine whether it is a package.

19. The method of claim 13, further comprising a change detection module and a bounding box filtering module to eliminate bounding boxes unlikely to surround packages.

20. A system for processing arrival or removal of an object within the field of view of a video camera, comprising:

a database module for recording objects residing in the field of view;

a detection module to determine a human person's entry into the field of view of the camera based on real-time analysis of successive image frames in the camera, with the detection module saving a first set of image frames corresponding to image frames captured by the video camera immediately prior to the detection of the human person's entry into the field of view of the camera and saving a second set of image frames corresponding to image frames captured by the video camera immediately after the detection of the human person's exit from the field of view of the camera;

a comparison module to compare the first set of image frames with the second set of image frames to identify one or more objects that appear or disappear from the field of view based on differences between the first and the set of image frames;

an artifact elimination module to determine and eliminate objects identified by comparing the first and second sets of image frames as artifacts; and

wherein the database module records (a) whether or not one of the objects residing in the database has been removed; and (b) whether or not an object is newly arrived; and c) enters any newly arrived objects into the database and removes any objects residing in the database that have been determined to have been removed.

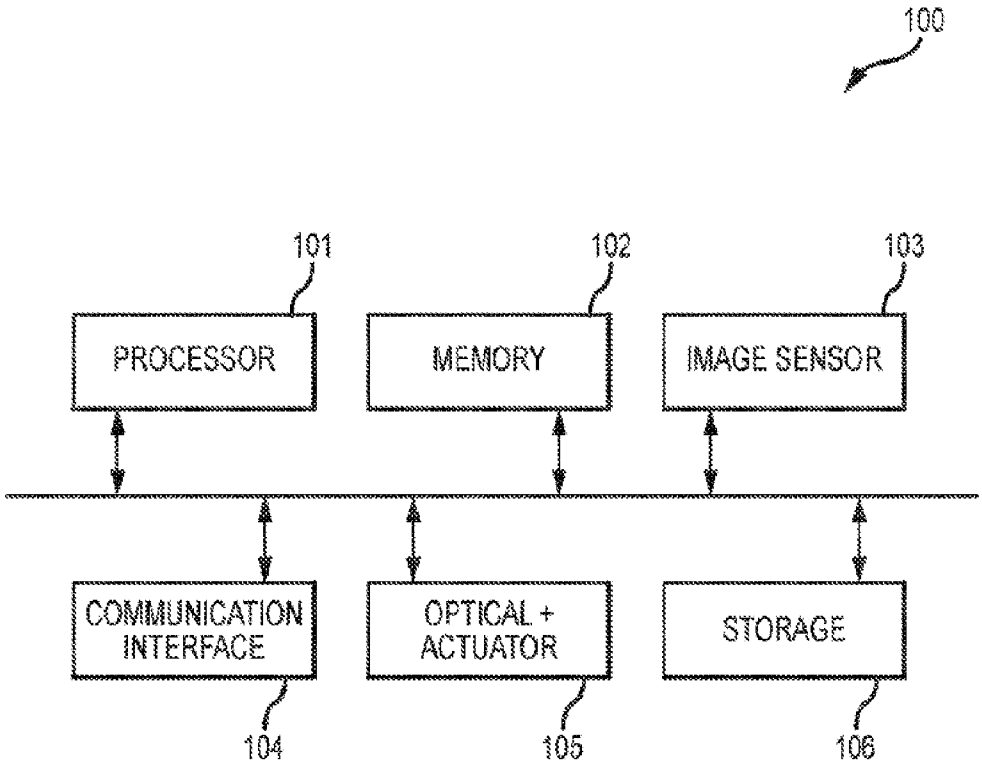

100
101
102
103
PROCESSOR
MEMORY
IMAGE SENSOR
COMMUNICATION INTERFACE
OPTICAL + ACTUATOR
STORAGE
104
105
106

Fig. 1

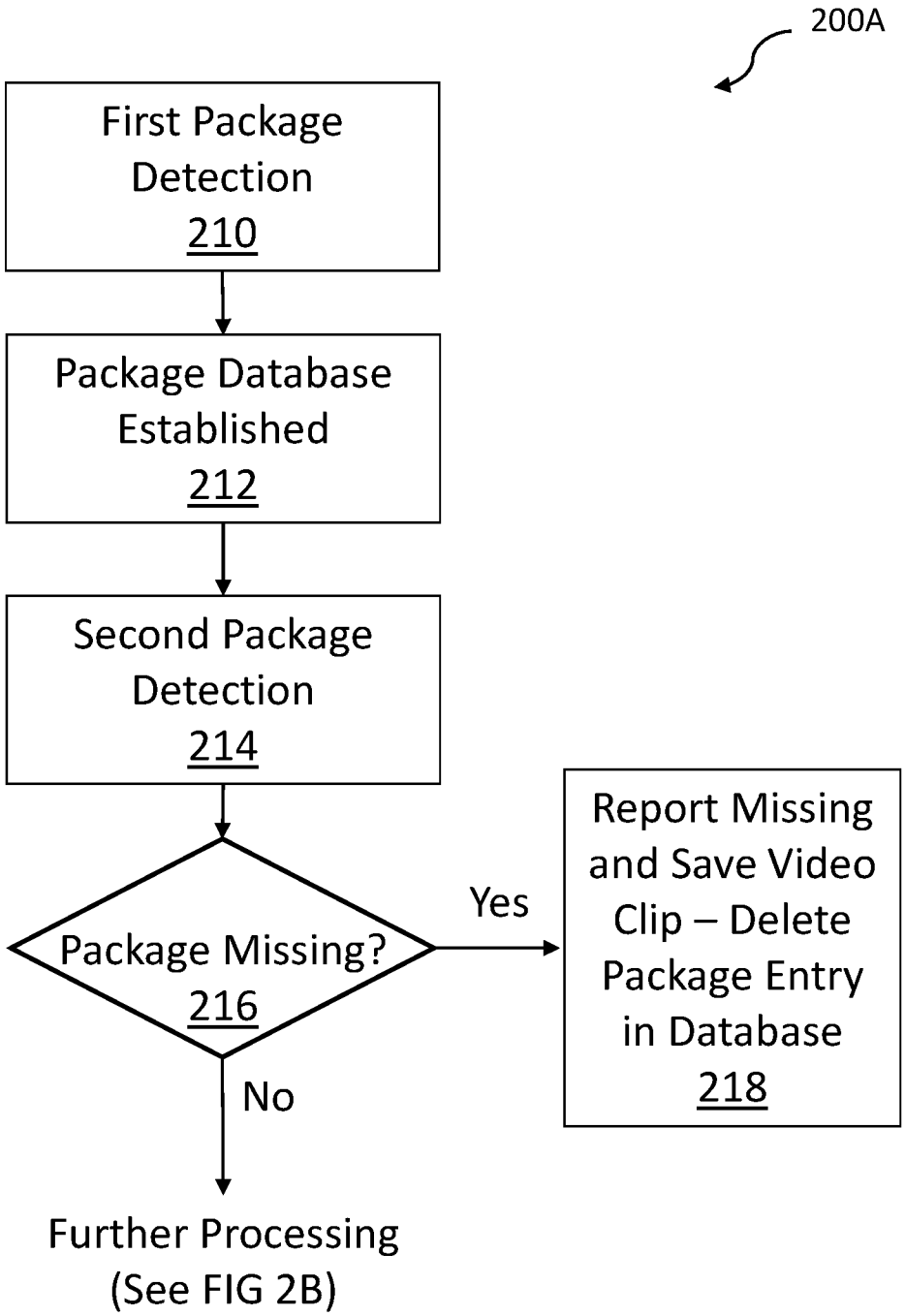

200A
First Package Detection 210
Package Database Established 212
Second Package Detection 214
Package Missing? 216
Yes
Report Missing and Save Video Clip – Delete Package Entry in Database 218
No
Further Processing (See FIG 2B)

Fig. 2A

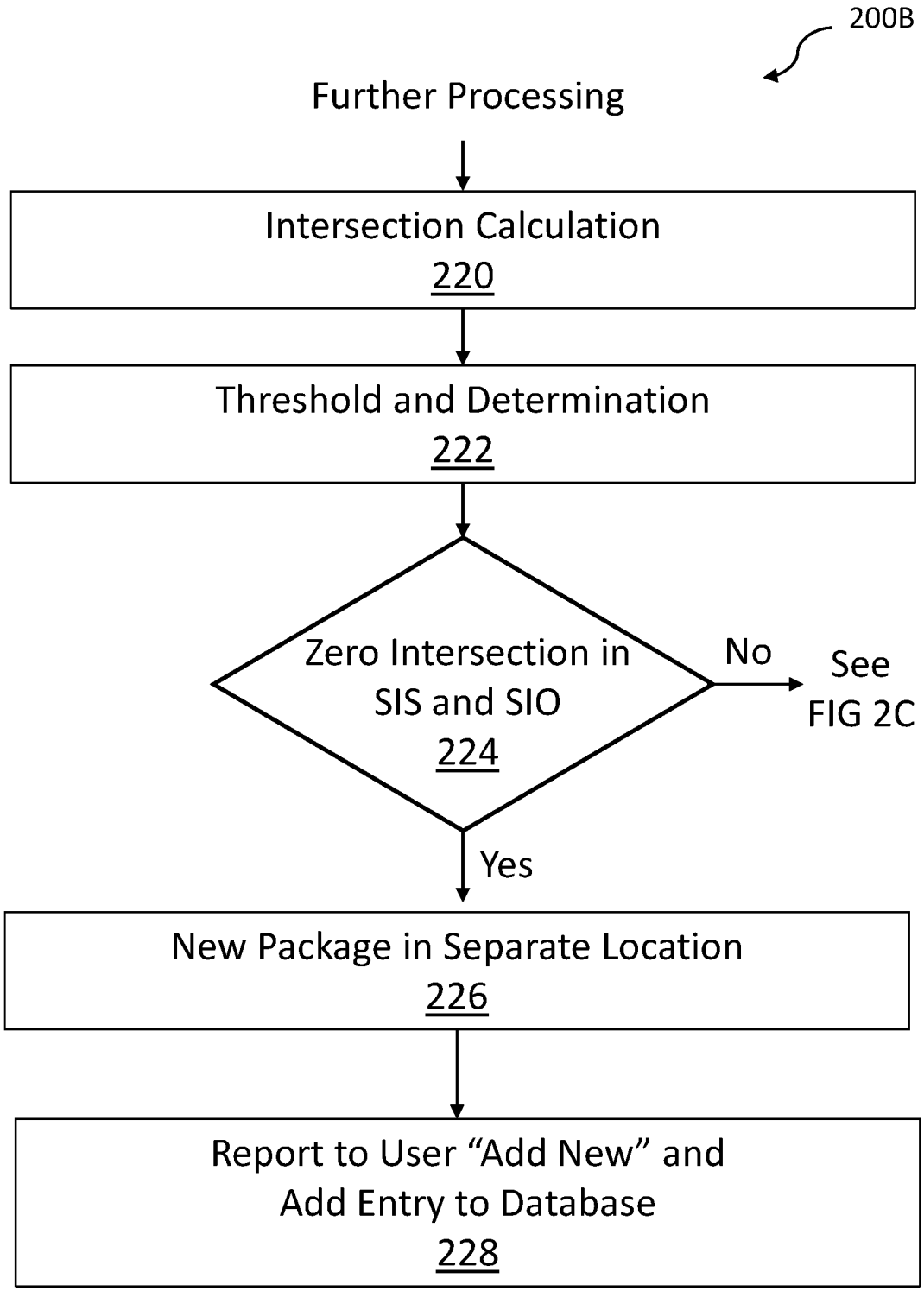

200B
Further Processing
Intersection Calculation
220
Threshold and Determination
222
Zero Intersection in SIS and SIO
224
No
See FIG 2C
Yes
New Package in Separate Location
226
Report to User "Add New" and Add Entry to Database
228

Fig. 2B

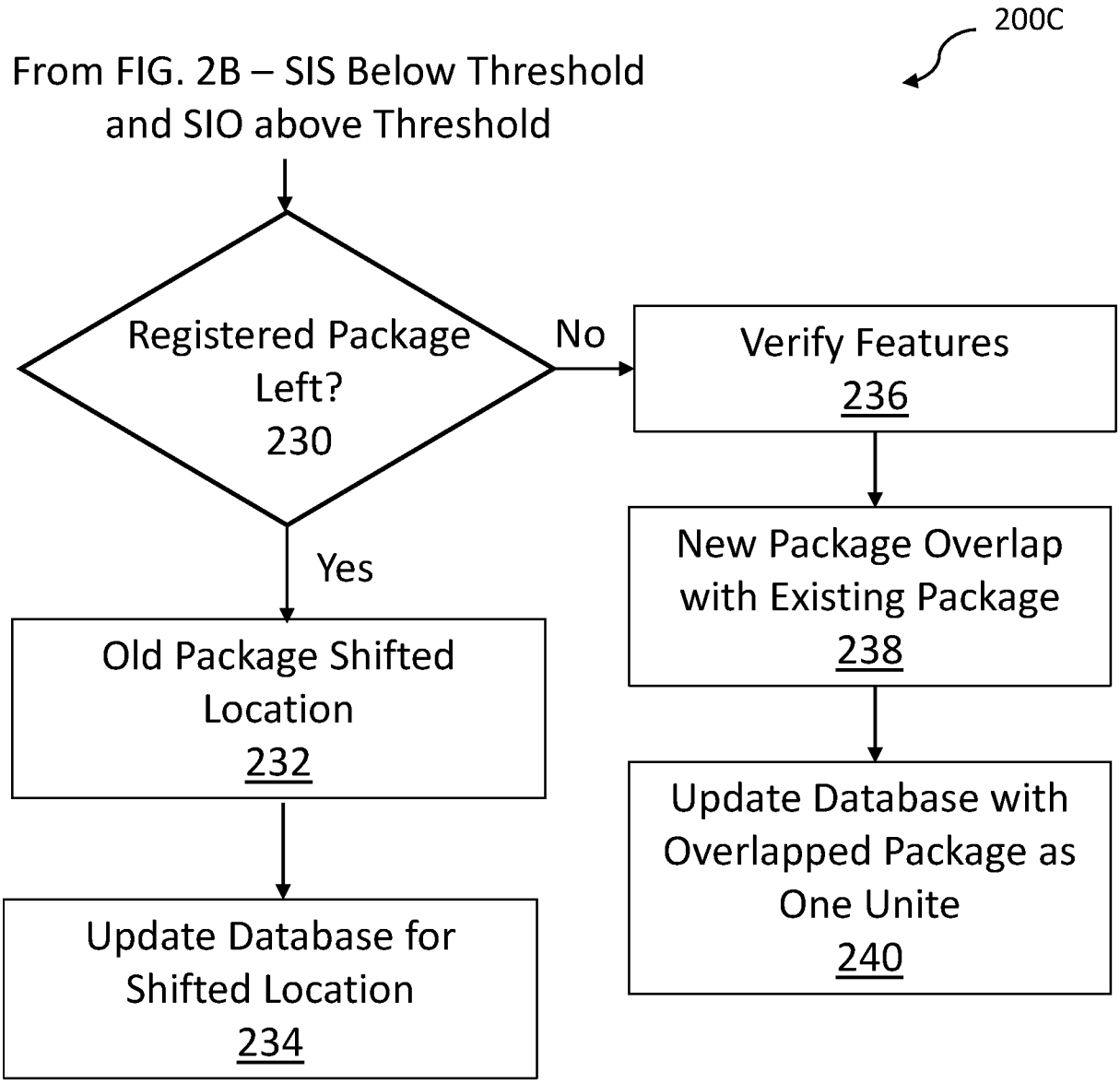

200C
From FIG. 2B – SIS Below Threshold and SIO above Threshold
Registered Package Left? 230
No
Yes
Verify Features 236
New Package Overlap with Existing Package 238
Update Database with Overlapped Package as One Unite 240
Old Package Shifted Location 232
Update Database for Shifted Location 234

Fig. 2C

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2018/047591

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G06T 7/00; G07G 3/00; G08B 13/00; G08B 13/18 (2018.01)

CPC - G06T 7/254; G06T 7/215; G07G 3/003; G08B 13/18; G08B 13/189; G08B 13/194; G08B 13/19608; G08B 13/19613 (2018.08)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

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

USPC - 382/190; 382/191; 382/203; 382/266; 382/276 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2016/0180667 A1 (VIVINT, INC.) 23 June 2016 (23.06.2016) entire document	1-20
Y	US 2006/0227862 A1 (CAMPBELL et al) 12 October 2006 (12.10.2006) entire document	1-20
Y	US 2005/0271277 A1 (LIN et al) 08 December 2005 (08.12.2005) entire document	6, 17
A	US 2016/0044286 A1 (AVO USA HOLDING 2 CORPORATION) 11 February 2016 (11.02.2016) entire document	1-20
A	WO 2009/70662 A1 (CERNIUM CORPORATION) 04 June 2009 (04.06.2009) entire document	1-20
A	US 2012/0213490 A1 (STEINER) 23 August 2012 (23.08.2012) entire document	1-20
A	US 2017/0220872 A1 (VIVINT, INC.) 03 August 2017 (03.08.2017) entire document	1-20

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
08 October 2018	18 OCT 2018

Name and mailing address of the ISA/US	Authorized officer
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)