



(51) International Patent Classification:

G06V 20/52 (2022.01) G06V 10/34 (2022.01)

G06V 20/40 (2022.01) G06V 10/75 (2022.01)

G06V 10/62 (2022.01)

(21) International Application Number:

PCT/US2022/029490

(22) International Filing Date:

16 May 2022 (16.05.2022)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/189,515 17 May 2021 (17.05.2021) US

17/485,221 24 September 2021 (24.09.2021) US

63/342,022 13 May 2022 (13.05.2022) US

(71) Applicant: APPLE INC. [US/US]; One Apple Park Way, Cupertino, California 95014 (US).

(72) Inventors; and

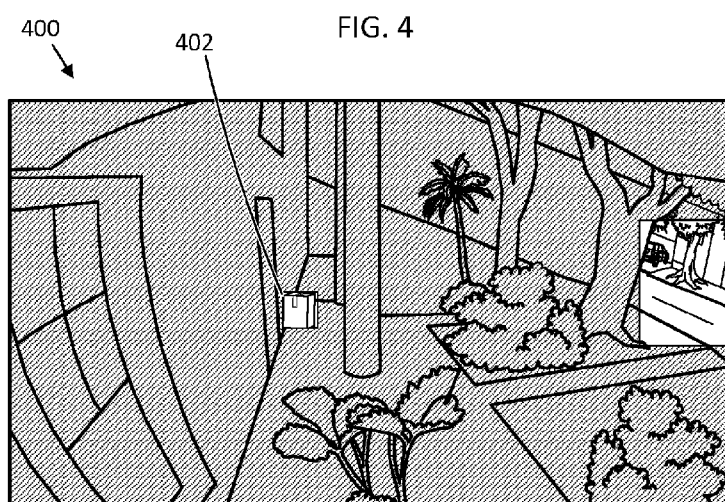
(71) Applicants: KHALILI, Mey; One Apple Park Way, Cupertino, California 95014 (US). DAHLKAMP, Hendrik; One Apple Park Way, Cupertino, California 95014 (US). BEBENITA, Michael; One Apple Park Way, Cupertino, California 95014 (US). JIN, Jonghoon; One Apple Park Way, Cupertino, California 95014 (US).

(74) Agent: MARLINK, Jeffrey S. et al.; Kilpatrick Townsend & Stockton LLP, Mailstop: IP Docketing - 22, 1100 Peachtree Street, Suite 2800, Atlanta, Georgia 30309 (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, IQ, IR, IS, IT, JM, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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, 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).

(54) Title: TECHNIQUES FOR DETECTION/NOTIFICATION OF PACKAGE DELIVERY AND PICKUP



(57) Abstract: Systems, computer-readable media, methods, and approaches described herein may identify delivery and/or pickup of packages. For example, packages may be identified within the areas captured by images and/or video. Based on the identification of the packages, it may be determined whether the package was delivered or picked up. A notification may be initiated that indicates that a package has been delivered and/or picked up.



Published:

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

TECHNIQUES FOR DETECTION/NOTIFICATION OF PACKAGE DELIVERY AND PICKUP

CROSS-REFERENCES TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 63/189,515, filed on May 17, 2021, U.S. Provisional Patent Application No. 63/342,022, filed on May 13, 2022, and U.S. Nonprovisional Patent Application No. 17/485,221, filed on September 24, 2021, the contents of which are incorporated herein by reference in their entireties for all purposes.

BACKGROUND

[0002] As digital camera technology has become cheaper and more prevalent, it has become commonplace for cameras to be positioned outside of structures (such as residential buildings, commercial buildings, manufacturing buildings, etc.) to monitor activities occurring outside the structures, or in an unsecured portion of the structures to monitor activities occurring within the unsecured portion of the structures. For example, cameras have been implemented in door bells and security systems to monitor for activities occurring outside the structure. Additionally, the increase in online sales of goods has resulted in more packages being delivered to structures. The delivery of packages to the structures has led to challenges, including theft of the packages.

BRIEF SUMMARY

[0003] The approaches described herein may utilize cameras to determine when packages have been delivered to or picked up from structures. In particular, a system may have one or more cameras, or may be coupled to one or more cameras, that are directed to capture video and/or images of areas where packages may be delivered or placed for pickup, such as outside of a structure, within an unsecured portion of a structure, and/or within another portion of a structure where packages may be intended to be placed for delivery and/or pickup. The cameras may be directed to areas where packages are expected to be delivered by a delivery person or placed for pickup by a delivery person. The system may identify delivery and/or pickup of packages within the video or images captured by the cameras. Based on the system identifying

delivery and/or pickup of the packages, the system may determine notification settings associated with the cameras and provide notification of the delivery and/or pickup of the packages.

[0004] In some embodiments, a method may include detecting motion within video received from a camera, capturing, in response to detecting the motion, a representative image from the video, and retrieving a canonical image that represents an average of one or more frames received from the camera before detecting the motion. The method may further include determining a difference between the canonical image and the representative image, the difference identifying a set of pixels of the representative image that are different from corresponding pixels in the canonical image. Further, the method may include removing low density pixels from the set of pixels, removing high frequency pixels from the set of pixels that have a high frequency of change, and executing a flood-fill process on the set of pixels, the flood-fill process adding additional pixels causing a first subset of disjointed pixels to become contiguous. The method may further include determining a candidate age for each pixel in the set of pixels, and removing, based at least in part on the candidate age of each pixel in the set of pixels, young pixels of the set of pixels that have a candidate age that is less than a threshold. Further, the method may include executing a classifier using the set of pixels, the classifier providing a prediction of an identification of an object represented by a second subset of the set of pixels, and outputting the prediction of the identification of the object.

[0005] In some embodiments, one or more computer-readable media having instructions stored thereon, wherein the instructions, when executed by a system, may cause the system to detect motion within video received from a camera, capture, in response to detecting the motion, a representative image of an area from the video, and retrieve a canonical image of the area that represents an average of one or more frames received from the camera before detecting the motion. The instructions, when executed by the system, may further cause the system to determine a difference between the canonical image and the representative image, the difference identifying a set of pixels of the representative image that are different from corresponding pixels in the canonical image. Further, the instructions, when executed by the system, may cause the system to remove low density pixels from the set of pixels, remove high frequency pixels from the set of pixels that have a high frequency of change, and execute a flood-fill process on the set

of pixels, the flood-fill process adding additional pixels causing a first subset of disjointed pixels to become contiguous. The instructions, when executed by the system, may further cause the system to determine a candidate age for each pixel in the set of pixels, and remove, based at least in part on the candidate age of each pixel in the set of pixels, young pixels of the set of pixels that have a candidate age that is less than a threshold. Further, the instructions, when executed by the system, may cause the system to execute a classifier using the set of pixels, the classifier providing a prediction of an identification of an object represented by a second subset of the set of pixels, output the prediction of the identification of the object.

[0006] In some embodiments, a system may include memory to store images from video received from a camera and one or more processors coupled to the memory. The processors may detect motion within the video, capture, from the stored images, a representative image corresponding to the detected motion, and retrieve, from the stored images, a canonical image that represents an average of one or more frames received from the camera before detection of the motion. The processors may further determine a difference between the canonical image and the representative image, the difference identifying a set of pixels of the representative image that are different from corresponding pixels in the canonical image. Further, the processors may remove low density pixels from the set of pixels, remove high frequency pixels from the set of pixels that have a high frequency of change, and execute a flood-fill process on the set of pixels, the flood-fill process adding additional pixels causing a first subset of disjointed pixels to become contiguous. The system may further determine a candidate age for each pixel in the set of pixels, and remove, based at least in part on the candidate age of each pixel in the set of pixels, young pixels of the set of pixels that have a candidate age that is less than a threshold. The processors may further execute a classifier using the set of pixels, the classifier providing a prediction of an identification of an object represented by a second subset of the set of pixels, and output the prediction of the identification of the object.

[0007] In some embodiments, a method may identifying one or more package candidates from an image. The method may further include determining an analysis frame from the image based at least in part on the one or more package candidates and executing a shared classifier on the analysis frame, the shared classifier to predict identification of packages and objects within

the analysis frame. Further, the method may include outputting a prediction of the identification of the packages and the objects based at least in part on the execution of the shared classifier.

[0008] In some embodiments, one or more computer-readable media may have instructions stored thereon, wherein the instructions, when executed by a system, cause the system to identify one or more package candidates from an image. The instructions may further cause the system to determine an analysis frame from the image based at least in part on the one or more package candidates and execute a shared classifier on the analysis frame, the shared classifier to predict identification of packages and objects within the analysis frame. Further, the instructions may cause the system to output a prediction of the identification of the packages and the objects based at least in part on the execution of the shared classifier.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] FIG. 1 illustrates an example system arrangement in accordance with some embodiments.

15 [0010] FIG. 2 illustrates an example procedure for identifying candidates for packages in accordance with some embodiments.

[0011] FIG. 3 illustrates an example of a canonical image in accordance with some embodiments.

[0012] FIG. 4 illustrates an example of a representative image in accordance with some
20 embodiments.

[0013] FIG. 5 illustrates an example image representation of a set of pixels produced by 208 in accordance with some embodiments.

[0014] FIG. 6 illustrates an example image representation of a set of pixels produced by 210 and 212 in accordance with some embodiments.

25 [0015] FIG. 7 illustrates an example image representation with bounding boxes according to some embodiments.

[0016] FIG. 8 illustrates an example image representation produced by 214 in accordance with some embodiments.

[0017] FIG. 9 illustrates an example procedure for classifying candidates and/or initiating a notification in accordance with some embodiments.

[0018] FIG. 10 illustrates a first portion of an example procedure for identification of an object in accordance with some embodiments.

5 [0019] FIG. 11 illustrates a second portion of the example procedure of FIG. 10 for identification of an object in accordance with some embodiments.

[0020] FIG. 12 illustrates a first portion of another example procedure for identification of an object in accordance with some embodiments.

10 [0021] FIG. 13 illustrates a second portion of the example procedure of FIG. 12 for identification of an object in accordance with some embodiments.

[0022] FIG. 14 illustrates an example procedure for identification of an object in accordance with some embodiments.

[0023] FIG. 15 illustrates an example unified system arrangement in accordance with some embodiments.

15 [0024] FIG. 16 illustrates an example procedure for identifying objects and/or package in accordance with some embodiments.

[0025] FIG. 17 illustrates an example analysis frame arrangement in accordance with some embodiments.

20 DETAILED DESCRIPTION

[0026] Systems, computer-readable media, methods, and approaches described throughout this disclosure may identify delivery and/or pickup of packages. Further, a notification may be provided indicating that the package has been delivered and/or picked up. The notification may allow for a user to be aware of the delivery and/or pickup of the package, which may cause the
25 user to retrieve the package from the area. In some instances, the notification may prevent issues presented by delivery of packages, such as theft of the packages, accidental incorrect deliveries, and/or lack of delivery of the packages.

[0027] The following detailed description refers to the accompanying drawings. The same reference numbers may be used in different drawings to identify the same or similar elements. In the following description, for purposes of explanation and not limitation, specific details are set forth such as particular structures, architectures, interfaces, techniques, etc. in order to provide a thorough understanding of the various aspects of various embodiments. However, it will be apparent to those skilled in the art having the benefit of the present disclosure that the various aspects of the various embodiments may be practiced in other examples that depart from these specific details. In certain instances, descriptions of well-known devices, circuits, and methods are omitted so as not to obscure the description of the various embodiments with unnecessary detail.

I. SYSTEM ARRANGEMENT

[0028] FIG. 1 illustrates an example system arrangement 100 in accordance with some embodiments. In particular, the system arrangement 100 illustrates one example of a system 102 that may utilize one or more cameras to identify delivery and/or pickup of packages. The approaches implemented by the system can accurately and efficiently identify the delivery and/or pickup of packages. Further, the system may indicate to a user when a package has been delivered and/or picked up, which may address some of the issues of packages being placed outside of structures and/or in unsecured portions of structures, such as theft of packages.

[0029] The system arrangement 100 may include the system 102. The system 102 may comprise a computing system that may be able to execute one or more instructions to perform one or more operations. In some embodiments, the system 102 may comprise a smart doorbell, a security system, a computer system, a distributed computer system, a mobile phone (such as a smart phone), a home environment made up of various connected user devices (e.g., a home hub, controller devices, smart television systems, accessories, etc.), or portions thereof.

[0030] The system 102 may include memory 104 and one or more processors 106 coupled to the memory 104. The memory 104 may include one or more computer-readable media. In some embodiments, the computer-readable media may comprise non-transitory computer-readable media. The memory 104 may have one or more instructions stored thereon, wherein the instructions, when executed by the system 102, may cause the system 102 to perform one or

more of the operations disclosed herein. The memory 104 may include any suitable volatile or non-volatile memory such as, but not limited to, dynamic random access memory (DRAM), static random access memory (SRAM), eraseable programmable read only memory (EPROM), electrically eraseable programmable read only memory (EEPROM), flash memory, solid-state memory, any other type of memory device technology, or some combination thereof.

[0031] The processors 106 may be separate from the memory 104 and/or coupled to the memory 104, or the memory 104 may be implemented as part of the processors 106 and may be coupled to the memory 104. The processors 106 may receive the instructions from the memory 104 and execute the instructions to cause the system 102 to perform one or more of the operations described throughout this disclosure. The processors 106 may include processor circuitry such as, for example, baseband processor circuitry, central processor unit circuitry, graphics processor unit circuitry, or some combination thereof.

[0032] The system 102 may further include a detector 108. The detector 108 may comprise computer hardware components, instructions that can be performed by the processors 106, or some combination thereof (which may be referred to as a “detector module” in some instances). The detector 108 may receive images and identify one or more candidates for delivery and/or pickup of a package based on the images. For example, the detector 108 may perform, or cause the system 102 to perform via the processors 106, one or more of the operations related to the identification of candidates from images described further throughout this disclosure.

[0033] The system 102 may further include a classifier 110. The classifier 110 may comprise computer hardware components, instructions that can be performed by the processors 106, or some combination thereof (which may be referred to as a “classifier module” in some instances). The classifier 110 may receive indications of the candidates within the images from the detector 108 and classify the candidates as packages, not packages, or something other than packages. For example, the classifier 110 may perform, or cause the system 102 to perform via the processors 106, one or more of the operations related to the classification of candidates from images described further throughout this disclosure.

[0034] The system arrangement 100 may further include an input device 112. In some embodiments, the input device 112 may comprise one or more cameras, where each of the cameras may capture images and/or video (which may comprise a series of images captured

temporally by the camera) of an area. The areas for which the cameras are to capture images and/or videos may include areas where packages are intended to be placed for delivery and/or pickup of the packages, such as outside of structures, within unsecured portions of structures, within package depository locations, and/or other areas where packages are to be placed for delivery and/or pickup. The input device 112 may be coupled to the system 102 and may provide the captured images and/or videos to the system 102. The system 102 may store the images and/or video received from the input device 112 in the memory 104 of the system 102. In some embodiments, the images and/or portions of the video may be stored by the system 102 in the memory 104 for a period of time, where the images and/or the portions of the video may be removed from the memory 104 at the expiration of the period of time. Removing the images and/or portions of the video from the memory 104 may ensure that there is space within the memory 104 for additional images and/or portions of the video to be stored. In some embodiments, the storage of the images and/or portions of the video may be cyclical, where the oldest images and/or portions of the video may be replaced by new images and/or portions of the video received from the input device 112 (e.g., in a first in, first out (FIFO) manner).

[0035] The system arrangement 100 may further include a notification device 114. The notification device 114 may comprise a computing system such as a smart doorbell, a security system, a computer system, a distributed computer system, a mobile phone (such as a smart phone), an accessory within a home environment, a smart television (e.g., a media streaming device), or portions thereof. The notification device 114 may provide a notification to a user that delivery and/or pickup of a package has been identified or otherwise detected by the system 102. For example, the notification device 114 may be coupled to the system 102 and may receive an indication from the system 102 to provide notification that the system 102 has identified delivery and/or pickup of a package. The notification device 114 may provide a visual notification, a sound notification, a haptic notification, another type of notification, or some combination thereof to a user of the notification device 114. For example, the notification device 114 may display an image of the area with the package on a display of the notification device 114 in response to receiving the indication from the system 102 in some embodiments. In some embodiments, the notification device 114 may present the notification via email, pop-up, an application, sound from a speaker, a motion generator, or some combination thereof.

[0036] While the input device 112 and the notification device 114 are illustrated separate from the system 102 in the illustrated system arrangement 100, it should be understood that the input device 112 and/or the notification device 114 may be included in the system 102 in other embodiments. For example, the system 102 may comprise a smart doorbell or security system in some embodiments where the input device 112 comprises a camera of the smart doorbell or the security system. Further, the notification device 114 may comprise a speaker of the smart doorbell or the security system in embodiments where the system 102 comprises the smart doorbell or the security system, where the notification device 114 may emit a sound to indicate that delivery and/or pickup of a package has been detected by the system 102. For example, the notification device 114 may provide a notification by emitting an audible sound by a smart speaker (such as a home hub) or displaying a notification on a television by a media streaming device (such as a smart television).

II. DETECTOR OPERATION

[0037] A detector of a system may process, or cause the system to process, images and/or video to identify candidates within the images and/or video of package delivery and/or pickup. For example, the system, implementing the detector, may compare a canonical image of an area from prior to detection of motion within the images and/or video with a representative image of the area after the detection of the motion to determine whether any objects have been added or removed from the area captured in the images. The detector may indicate portions of the representative image (such as by removing pixels from the representative image) to be processed by a classifier to determine whether a package has been delivered and/or picked up.

[0038] FIG. 2 illustrates an example procedure 200 for identifying candidates for packages in accordance with some embodiments. The procedure 200 may be performed by a system (such as the system 102 (FIG. 1)). For example, a detector (such as the detector 108 (FIG. 1)) may perform, or cause the system to perform, one or more of the operations of the procedure 200. The system may output candidates for packages based on the performance of the procedure 200. While the procedure 200 is illustrated in a certain order, it should be understood that one or more of the operations within the procedure 200 may be performed in a different order, concurrently with other operations, and/or may be omitted.

[0039] In 202, the system may detect motion within images and/or video received from a camera (such as the input device 112 (FIG. 1)). For example, the system may have received images and/or video from the camera and stored the received images within a memory (such as the memory 104 (FIG. 1)) of the system. Upon receiving an image from the camera, the system

5 may compare the image with a consecutive image captured prior to the image to determine whether there are differences in the area captured within the image and the consecutive image. The system may determine that there are differences in the area based on corresponding pixels within the image and the consecutive image having differences in value or values greater than a threshold amount. For example, the differences in value or values may be value or values
10 indicating a color of the corresponding pixels in the image and the consecutive image. In some embodiments, the system may further determine that a number of contiguous pixels within the image and consecutive image that have difference in value or values is greater than a threshold number of contiguous pixels to determine that there are differences in the area. Further, in response to determining that the image and the consecutive image have corresponding pixels
15 with different values, the system may compare the consecutive image with one or more subsequent images to determine whether the differences, or other related differences, are presented for a certain period of time to determine that there are differences in the area. Based on the system determining that there are differences within the area, the system may determine that motion exists within the images and/or video. Accordingly, the system may detect motion with
20 the images and/or video based on the determination that motion exists within the images and/or video.

[0040] In 204, the system may capture a representative image. The system may capture the representative image from the images and/or video received from camera in response to the detection of the motion in 202. The representative image may be an image captured by the
25 camera after completion of the motion detected in 202. For example, the system may compare consecutive images captured by the camera after the detection of the motion to determine whether the consecutive images are determined to present differences, such as the differences described in relation to 202. Based on a certain number of the consecutive images not presenting differences, the system may determine that the motion detected in 202 has completed. The
30 system may retrieve an image captured after the motion has completed from memory and utilize the retrieved image as the representative image that has been captured. In some embodiments,

the system may cause the camera to capture an image after the motion has completed, receive the captured image from the camera, and utilize the image as the representative image.

[0041] FIG. 4 illustrates an example of a representative image 400 in accordance with some embodiments. In particular, the representative image 400 is an example of a representative image that may be captured in 204. The representative image 400 may have been captured by the camera after motion within an area captured by the camera has been completed. The representative image 400 shows a package 402 in the area captured by the representative image 400 in the illustrated embodiment.

[0042] In 206, the system may retrieve a canonical image. The canonical image may represent an average of one or more frames received from the camera before the detection of the motion in 202. In particular, the frames may comprise one or more images captured by the camera prior to the detection of the motion in 202 and provided to the system. The system may average the values of each of the corresponding pixels within the frames to produce values for each of the pixels within the canonical image. The system may produce the canonical image based on the determined values for each of the pixels and store the canonical image in the memory of the system. The system may retrieve the canonical image from the memory in 206. In other embodiments, the canonical image may be an image captured by the camera prior to the detection of the motion in 202. For example, the system may retrieve an image that was captured prior to the detection of the motion in 202 from the memory of the system and utilize the image as the canonical image. The frames and/or images utilized for retrieving the canonical image may be retrieved by the system from a preroll for the camera. The preroll may comprise one or more images captured within a period of time prior to the detection of motion. In some embodiments, the period of time for the preroll may be four seconds prior to the detection of motion. The images within the preroll are often pretty static with little to no movement with the area captured by the images.

[0043] FIG. 3 illustrates an example of a canonical image 300 in accordance with some embodiments. In particular, the system may have produced the canonical image 300 by the averaging of the values of the correspond pixels with the frames as described in relation to 206. In other embodiments, the canonical image 300 may be an image that was captured prior to the detection of the motion. The area captured by the canonical image 300 may be the same area as

the area captured by the representative image 400 (FIG. 4) in the illustrated embodiment. In other embodiments, the area captured by the canonical image 300 may at least partially overlap with the area captured by the representative image 400. The canonical image 300 shows a truck 302 in the area captured by the canonical image 300 in the illustrated embodiment.

5 [0044] In 208, the system may determine a difference between the canonical image and the representative image. For example, the system may compare the canonical image with the representative image to determine a difference between the canonical image and the representative. In particular, the system may compare values between pixels within the canonical image and corresponding pixels within the representative image to determine differences
10 between the values of the pixels within the canonical image and the corresponding pixels within the representative image. In some embodiments, determining the differences between the values of the pixels may include determining, by the system, distances within a color space using Delta E for each of the corresponding pixels (which may be referred to as related pixels) within the representative image and the canonical image. The Delta E may be the distance metric defined
15 by the International Commission of Illumination. The system may identify pixels as being different based on the determined distances being greater than a threshold distance. In some embodiments, the differences between the distances may be determined in a LAB color space. The LAB color space may be the color space defined by the International Commission of Illumination and may comprise three parameters, the three parameters comprising one parameter
20 defining brightness and two parameters defining color. The difference determined by the system may identify a set of pixels of the representative image that are different from the corresponding pixels in the canonical image. The system may produce an image representation of the set of pixels determined to be different, where the image representation has pixels removed that correspond to corresponding pixels of the canonical image and the representative image that are
25 determined not to be different.

[0045] FIG. 5 illustrates an example image representation 500 of a set of pixels produced by 208 in accordance with some embodiments. In particular, the image representation 500 may indicate a set of pixels that are determined to have differed between the canonical image and the representative image. For example, the image representation 500 may have pixels determined to
30 be different between the canonical image and the representative image shown in the image

representation 500, and may have pixels determined not to be different between the canonical image and the representative image being removed from the image representation 500 (as indicated by shown in black in the image representation 500). In the illustrated embodiment, the image representation 500 includes a first group of pixels 502 and a second group of pixels 504.

- 5 The first group of pixels 502 may correspond to the package 402 (FIG. 4), where the package 402 is shown in the representative image 400 (FIG. 4) and not shown in the canonical image 300 (FIG. 3). The second group of pixels 504 may correspond to the truck 302 (FIG. 3), where the truck 302 is shown in the canonical image 300 and not shown in the representative image 400.

- [0046] In 210, the system may remove low density pixels. In particular, the system may
10 remove the low density pixels from the set of pixels identified in 208. Removing the low density pixels from the set of pixels may include averaging amounts of change in subgroups of the set of pixels. For example, the system may separate the set of pixels into subgroups then average the amounts of change of the each of the pixels within the subgroups to produce an averaged amount of change for each of the subgroups. The subgroups of the set of pixels may be uniform sized or
15 varying sizes in different embodiments. The averaged amount of change for each of the subgroups may be compared to a threshold amount of change. The system may identify a portion of the subgroups that have averaged amounts of change below the threshold amount of change. Based on a portion of the subgroups having averaged amounts of change below the threshold amount of change, the system may determine that the pixels within the portion are low density
20 pixels. The system may remove the identified low density pixels from the set of pixels.

- [0047] In other embodiments of 210, the system may keep high density pixels. In particular, the system may keep the high density pixels while removing other pixels from the set of pixels identified in 208. Keeping the high density pixels may include averaging amounts of change in subgroups of the set of pixels. For example, the system may separate the set of pixels into
25 subgroups then average the amounts of change of each of the pixels within the subgroups to produce an averaged amount of change for each of the subgroups. The subgroups of the set of pixels may be uniform sized or varying sizes in different embodiments. The average amount of change for each of the subgroups may be compared to a threshold amount of change. The system may identify a portion of the subgroups that have averaged amounts of change above the
30 threshold amount of change. Based on a portion of the subgroups having averaged amounts of

change above the threshold amount of change, the system may determine that the pixels within the portion are high density pixels. The system may keep the high density pixels within the set of pixels while removing the other pixels from the set of pixels.

[0048] In 212, the system may remove high frequency pixels. In particular, the system may
5 remove high frequency pixels from the set of pixels, where the high frequency pixels present a high frequency of change. The system may apply temporal averaging to the set of pixels. For example, the system may average values of the pixels within the representative image and previous frames captured by the camera within a time period prior to the representative image to determine average frequencies at which the pixels change among the representative image and
10 the successive images. In some embodiments, the system may perform 208 and 210 for successive images within the time period in addition to performing 208 and 210 for the representative image to produce sets of pixels for the successive images. The system may then average the set of pixels from the representative image and the previous frames to determine average frequencies at which each of the pixels within the set of pixels change. The system may
15 compare the averaged frequencies for the pixels with a threshold frequency to determine whether the pixel has been stable for the time period. For pixels with averaged frequencies above the threshold frequency, the system may determine that the pixels have been unstable within the period of time and, therefore, are high frequency pixels. The system may remove the high frequency pixels indicated by the temporal averaging.

[0049] The system removing the high frequency pixels may result in modifying the set of
20 pixels based on the previous frames to form a modified set of pixels. The modified set of pixels may be smaller than the set of pixels based on the removal of the high frequency pixels. In some embodiments, the modified set of pixels may be utilized in one or more of the operations throughout this disclosure. For example, operations 214, 216, 218, 902 (FIG. 9), 904 (FIG. 9), or
25 some combination thereof may be performed on the modified set of pixels.

[0050] FIG. 6 illustrates an example image representation 600 of a set of pixels produced by
210 and 212 in accordance with some embodiments. In particular, the system may produce the image representation 600 by removing the low frequency pixels or keeping the high frequency
pixels, in accordance with 210, and the high frequency pixels, in accordance with 212, from the
30 image representation 500 (FIG. 5). As can be seen from the image representation 600, the image

representation 600 has more pixels removed as compared to the image representation 500 while maintaining a first group of pixels 602 corresponding to the first group of pixels 502 (FIG. 5) and a second group of pixels 604 corresponding to the second group of pixels 504 (FIG. 5). The first group of pixels 602 and the second group of pixels 604 may be maintained based on the pixels not being determined to be low density pixels or being determined to be high density pixels in 210 and not being determined to be high frequency pixels in 212.

[0051] In 214, the system may perform an execute flood-fill process. The flood-fill process may add additional pixels causing a first subset of disjointed pixels to become contiguous. For example, the system may identify subsets of disjointed pixels from the set of pixels produced by 210 and 212. The system may add additional pixels to each of the subsets of disjointed pixels to cause each of the disjointed pixels to be a contiguous subset. In some embodiments, executing the flood-fill process may include generating bounding boxes around each of the disjointed pixels. For example, the bounding boxes may comprise a square, rectangle, or other shape with edges of the shape corresponding with outer bounds of the disjointed pixels. In other 15 embodiments, the edges of the shape may be outside of the outer bounds of the disjointed pixels, where the edges may be a defined distance from the outer bounds of the disjointed pixels, sizes of the bounding box may be a defined size, sizes of the bounding boxes may be larger than an area defined by the outer bounds of the disjointed pixels by a defined percentage, or some combination thereof. In some of these embodiments, the system may compare the size of the 20 bounding boxes with a size of the image to determine whether to remove one or more of the bounding boxes. For example, if a bounding box for a subset of disjointed pixels is larger than a first threshold percentage of the image or smaller than a second threshold percentage of the image (the second threshold percentage being less than the first threshold percentage), the system may remove the bounding box. This may prevent the system from producing candidates for a package that are too small for a package or too small for a package. This may also prevent the 25 system from erroneously identifying the whole image as a candidate for the package, which may be caused by a change of lighting within the area. The system may add additional pixels to the bounding boxes (or the bounding boxes remaining after the removal of the bounding boxes due to size in some embodiments) to fill the bounding box. In some embodiments, adding the 30 additional pixels may include setting the additional pixels to value of the corresponding pixels

from the representative image. The system adding the additional pixels to the subsets of disjointed pixels may produce candidates for packages.

[0052] FIG. 7 illustrates an example image representation 700 with bounding boxes according to some embodiments. In particular, the image representation 700 shows example bounding boxes that may be generated by 214. For the image representation 700, the system may have performed 214 on the image representation 600 (FIG. 6) to produce the image representation 700. In the illustrated embodiment, the system may have identified a first group of pixels 702 as a first subset of disjointed pixels and a second group of pixels 704 as a second subset of disjointed pixels. The first group of pixels 702 may correspond to the first group of pixels 602 (FIG. 6), which in turn corresponds to the package 402 (FIG. 4). The second group of pixels 704 may correspond to the second group of pixels 604 (FIG. 6), which in turn corresponds to the truck 302 (FIG. 3). As used in relation to the bounding boxes, a subset of disjointed pixels may refer to a group of pixels that are not the shape and/or size of the bounding boxes. The system may generate a first bounding box 706 around the first group of pixels 702, identified as a first subset of disjointed pixels, and a second bounding box 708 around the second group of pixels 704, identified as a second subset of disjointed pixels. In some instances, the system may have identified a third group of pixels 710 as a third subset of disjointed pixels and a fourth group of pixels 712 as a fourth subset of disjointed pixels, but the system may have removed the corresponding bounding boxes based on the system determining that the corresponding bounding boxes were smaller than a threshold percentage of the image.

[0053] FIG. 8 illustrates an example image representation 800 produced by 214 in accordance with some embodiments. In particular, the system may have produced the image representation by adding additional pixels to the bounding boxes. For example, the system may have added additional pixels to fill the first bounding box 706 (FIG. 7) to produce a first candidate 802 and may have added additional pixels to fill the second bounding box 708 (FIG. 7) to produce a second candidate 804.

[0054] In 216, the system may determine candidate ages for pixels within the set of pixels. For example, the system may determine an amount of time that pixels within the set of pixels have been at a current value (or within a defined range of the current value), which may be defined as the candidate ages for the pixels. The system may compare the candidate ages of the

pixels to a threshold to determine if the pixels are young pixels. The system may determine that a pixel is a young pixel based on a candidate age of the pixel being less than the threshold.

[0055] In some embodiments, the system may determine the candidate ages for subsets of disjointed pixels identified in 214 and/or pixels within the bounding boxes generated in 214. In some of these embodiments, the system may compare the candidate ages for each of the pixels within the subsets of disjointed pixels and/or the bounding boxes with the threshold to determine whether each of the pixels are young pixels. In other of these embodiments, the system may average the candidate ages of the pixels within the subsets of disjointed pixels and/or the bounding boxes and then compare the average candidate ages with the threshold to determine whether all the pixels within each of the subsets of disjointed pixels and/or bounding boxes are determined to be young pixels.

[0056] In 218, the system may remove the young pixels from the set of pixels. In particular, the system may remove the pixels determined to be young pixels in 216 from the set of pixels. Accordingly, the system may remove the young pixels from the set of pixels based on the young pixels of the set of pixels having a candidate age that is less than the threshold. In embodiments where 216 and 218 are applied to the candidates produced by 214, candidates having at least a defined percentage (which may be a majority, all, or some other percentage) of young pixels may be removed as candidates. In other embodiments, the individual pixels determined to be young pixels within the candidates may be removed leaving the rest of the pixels as part of a candidate for a package.

[0057] While the procedure 200 is described in a certain order, it should be understood that the operations of the procedure 200 may be performed in different orders and/or concurrently in some embodiments. For example, the operations 210 through 218 may be performed in different orders and/or concurrently in other embodiments.

III. CLASSIFIER OPERATION

[0058] A classifier of a system may analyze the candidates provided from the detector and classify objects captured within the candidates as packages, not packages, or something other than packages. For example, the classifier may, or may cause the system, to apply one or more classification models to the candidates provided from the detector and indicate whether each of

the candidates is a package, not a package, or something other than a package based on the classification models. The system may utilize the classification of the objects within the candidates to determine whether to provide notification that a package has been delivered or picked up.

5 [0059] FIG. 9 illustrates an example procedure 900 for classifying candidates and/or initiating a notification in accordance with some embodiments. The procedure 900 may be performed by a system (such as the system 102 (FIG. 1)), or a combination of a system and a notification device (such as the notification device 114 (FIG. 1)). For example, a classifier (such as the detector 110 (FIG. 1)) may perform, or cause the system to perform, one or more of the
10 operations of the procedure 900. In some embodiments, a notification device may perform one or more of the operation of the procedure 900. The system may output a prediction for the candidates received from a detector and/or initiate a notification based on the performance of the procedure 900. While the procedure 900 is illustrated in a certain order, it should be understood that one or more of the operations within the procedure 900 may be performed in a different
15 order, concurrently with other operations, and/or may be omitted.

[0060] In 902, the system may identify candidates for packages. In particular, the system may identify one or more candidates output by the detector on completion of the performance of the procedure 200. The system utilizing the candidates for the procedure 900 may allow the system to analyze the portions of the representative image and/or the canonical image
20 corresponding to candidates for packages rather than analyzing an entirety of the representative image and/or the canonical image. By analyzing the portions of the representative image and/or the canonical image, the system may operate faster than if the system was to analyze the entirety of the representative image and/or the canonical image, and may provide greater accuracy in classifying packages than if the system was to analyze the entirety of the representative image
25 and/or the canonical image.

[0061] For example, the system may identify the first candidate 802 (FIG. 8) and the second candidate 804 (FIG. 8) for the illustrated image representation 800 (FIG. 8). The first candidate 802 and the second candidate 804 may be output by the detector at the completion of the procedure 200 to the classifier, or another part of the system.

[0062] In 904, the system may execute a classifier on candidates. For example, the system may execute the classifier on the set of pixels within the candidates (which may be the set of pixels with the removed pixels and the added additional pixels from 210 through 218) to determine whether the area captured by the candidate includes a package. The classifier executed by the system may include a classification model that the system utilizes to predict if the areas captured in each of the candidates includes a package. The classification model may comprise a model produced by machine learning. The classification model may have been trained based on a training set and/or variation set to identify packages within the areas captured in each of the candidates. In some embodiments, the training of the classification model by the system may include supervised learning, where desired results for the elements within the training set and/or the variation set may be indicated to the system. In some embodiments, the system may use any machine learning technique (such as random forests, support vector machine, artificial neural networks, or some other machine learning technique) to train the classification model.

[0063] The system may determine, based on the classification model, whether the areas captured in the candidates appear to include a package and may produce predictions for each of the areas as to whether the areas include packages based on the output of the classification model. For example, if the classification model outputs an indication for any of the candidates of a representative image and/or a canonical image appears to include a package, the system may produce a prediction that the area captured in the representative image and/or the canonical image includes a package. If the classification model outputs an indication that an area corresponding to a candidate appears to not include a package, the system may produce a prediction that the area does not include a package.

[0064] In some instances, the system may identify an object within the area captured by a candidate. For example, the system may identify an object based on changes in values of the pixels within the candidate between the canonical image and the representative image. The system may apply the classification model to a subset of the set of pixels corresponding to the candidate to determine an identification of the object. The output of the classification model may indicate whether the object appears to be a package or something other than a package. The system may produce a prediction that the object is a package based on the classification model producing an indication that the object appears to be a package. Alternatively, the system may

produce a prediction that the object is not a package based on the classification model producing an indication that the object appears to be something other than a package.

[0065] In some embodiments, the system may execute the classifier on the candidate (defining the set of pixels) from the representative image provided by the detector and a corresponding set of pixels from the canonical image. The system may determine whether a package appears to be located within the area captured by the candidate in the representative image and whether the package appears to be located within the same area in the canonical image. Based on whether the classification model indicates that a package appears to be located within the area in both the representative image and the canonical image, the system may determine whether a package has been delivered or picked up. For example, if the representative image appears to have a package located within the area based on the classification model and the canonical appears not to have a package located within the area based on the classification model, the system may generate a prediction that a package has been delivered. If the representative image appears not to have a package located within the area based on the classification model and the canonical image appears to have a package located within the area based on the classification model, the system may generate a prediction that a package has been picked up.

[0066] For example, the system, based on the classification model, may determine that the area captured in the first candidate 802 (FIG. 8) appears to have a package and the second candidate 804 (FIG. 8) appears to not have a package. The system may generate a prediction that the area captured by the first candidate 802 has a package located within the area. The system may further generate a prediction that the area captured by the second candidate 804 does not have a package located within the area.

[0067] In some embodiments, the system may further determine whether there are any other objects that overlap with a predicted package when the system predicts that a package has been delivered. For example, the system may analyze the candidate to determine whether there are any other objects, such as individuals and/or animals within an area of the predicted package. The system may identify the individual and/or animals based on the shape of the objects, the movement of the objects, and/or other defining characteristics of the objects detected within the area of the predicted package. In some of these embodiments, the system may identify an

individual within the area of the predicted package, and may determine that the package is being carried by the individual. Based on the determination that the package is being carried by the individual, the system may determine that the package has not been delivered and determine not to initiate a notification in 910 based on the package not having been delivered.

- 5 [0068] In some embodiments, the classifier may include a classification model that the system utilizes to predict if an area within the predicted packages includes an individual or an animal. The classification model may be implemented with the same classification model as the classification model used for predicting the packages or may be a separate classification model from the classification model used for predicting the packages. The classification model may
10 have been trained based on a training set and/or variation set to identify individuals and/or animals within the areas captured in each of the candidates. In some embodiments, the training of the classification model by the system may include supervised learning, where desired results for the elements within the training set and/or the variation set may be indicated to the system. In some embodiments, the system may use any machine learning technique (such as random
15 forests, support vector machine, artificial neural networks, or some other machine learning technique) to train the classification model.

- [0069] In some embodiments, the classifier may further compare a candidate that is predicted to include a package with corresponding pixels from an image captured at a detection of the motion within the video. For example, the system may capture an image at a time when the
20 motion is detected within the images and/or video captured by the camera. The classifier may compare the values of the pixels for the candidate with the values for the corresponding pixels of the image captured at the time when the motion is detected. The classifier may further compensate for changes in lighting between lighting in some embodiments. For example, in comparing the values of the pixels, the system may determine the differences in values for each
25 of the pixels. The system may determine whether the differences for each of the corresponding pixels have changed by amounts corresponding to a change in lighting. For example, if all the pixels change by approximately (within 5%) the same value and/or intensity, the system may determine that a lighting change has occurred and that the predicted package was a false positive due to the lighting changes. Accordingly, the system may predict that the candidate does not
30 include a package based on the lighting change causing a false positive.

[0070] In 906, the system may output a prediction. For example, the system may output the prediction produced in 904. The system may output the prediction to the notification device.

[0071] In 908, the notification device may determine notification settings associated with the cameras and/or the notification device. For example, the notification device may have stored notification settings that indicate when the notification device is to initiate a notification. The notification settings may be presented to a user of the notification device and the user may indicate the events for which the notification device is to initiate a notification. In some embodiments, the notification settings may include one or more selections indicating whether the notification device is to provide a notification for a package being delivered, a package being picked up, or some combination thereof. In some embodiments, the notification settings may include one or more selections indicating a type of notification (such as a visual notification, a sound notification, a haptic notification, another type of notification, or some combination thereof) that the notification device is to utilize for the notification. The user may utilize the selections to indicate whether the notification device is to provide the notification and/or the type of notification that is to be provided. In response to the notification device receiving a prediction from the system, the notification device may review the notification settings and determine whether a notification is to be initiated and/or what type of notification is to be initiated based on the prediction.

[0072] In 910, the notification device may initiate a notification. In particular, the notification system may initiate a notification based at least in part on the predication and the determination if a notification is to be provided in 908. If the notification device determines that a notification is to be provided based on the notification settings in 908, the notification system may initiate the notification in accordance with the notification settings. If the notification device determines that a notification is not to be provided based on the notification settings in 908, 910 may be omitted. Providing the notification may include providing a notification via a smart doorbell, a security system, a computer system, a distributed computer system, a mobile phone (such as a smart phone), an accessory within a home environment, a smart television (e.g., a media streaming device), or portions thereof.

[0073] In 912, the system may identify additional candidates. For example, the system may have previously predicted that a package had been delivered and may identified additional

candidates subsequently in time to the previously predicted package. The system may utilize the same procedure for identifying the subsequent candidates as identifying the candidates in 902..

[0074] In 914, the system may execute the classifier. The execution of the classifier may include one or more of the features of the execution of the classifier in 904. The classifier

5 executed in 914 may further compare a location of the additional candidates to the locations of the previously predicted packages. If the location of the additional candidates is on top of or adjacent to the previously predicted packages, the classifier may further determine whether the additional candidates include an animal and/or an individual. For example, the classifier may utilize the classification model utilized for identifying individuals and/or animals to analyze the
10 additional candidates. If the classification model predicts that there is an animal and/or individual located within the additional candidates, the system may output a prediction of the additional candidates being an animal and/or individual. The system may further determine not to initiate an additional notification in 918 based on the prediction of the animal and/or individual being within the additional candidates. If the system determines that an animal and/or individual is not
15 predicted to be within the additional candidates, the system may further use the classification model for predicting whether a package is within the additional candidates. The system may determine to initiate an additional notification in 918.

[0075] In 916, the system may output a prediction. For example, the system may output the prediction produced in 914. The system may output the prediction to the notification device.

20 [0076] In 918, the notification device may determine notification settings associated with the cameras and/or the notification device. For example, the notification device may have stored notification settings that indicate when the notification device is to initiate a notification. The notification settings may be presented to a user of the notification device and the user may indicate the events for which the notification device is to initiate a notification. In some
25 embodiments, the notification settings may include one or more selections indicating whether the notification device is to provide a notification for a package being delivered, a package being picked up, or some combination thereof. In some embodiments, the notification settings may include one or more selections indicating a type of notification (such as a visual notification, a sound notification, a haptic notification, another type of notification, or some combination
30 thereof) that the notification device is to utilize for the notification. The user may utilize the

selections to indicate whether the notification device is to provide the notification and/or the type of notification that is to be provided. In response to the notification device receiving a prediction from the system, the notification device may review the notification settings and determine whether a notification is to be initiated and/or what type of notification is to be initiated based on the prediction.

[0077] In 920, the notification device may initiate a notification in some instances. In particular, the notification system may initiate a notification based at least in part on the predication and the determination if a notification is to be provided in 918. If the notification device determines that a notification is to be provided based on the notification settings in 918, the notification system may initiate the notification in accordance with the notification settings. If the notification device determines that a notification is not to be provided based on the notification settings in 918, 920 may be omitted. Providing the notification may include providing a notification via a smart doorbell, a security system, a computer system, a distributed computer system, a mobile phone (such as a smart phone), an accessory within a home environment, a smart television (e.g., a media streaming device), or portions thereof.

[0078] FIG. 10 illustrates a first portion of an example procedure 1000 for identification of an object in accordance with some embodiments. The procedure 1000 may be performed by a system, such as the system 102 (FIG. 1). For example, the system may receive an image and output a prediction of an identification of the object. While the procedure 1000 is illustrated in a certain order, it should be understood that one or more of the operations within the procedure 1000 may be performed in a different order, concurrently with other operations, and/or may be omitted.

[0079] In 1002, the system may capture a representative image. For example, the system may capture a representative image from a video provided to the system. The video may be captured by a camera directed to capture a particular area, where the representative image is an image of the particular area.

[0080] In 1004, the system may retrieve a canonical image. For example, the system may retrieve a canonical image that represents one or more frames of the video before the representative image. The canonical image may be an image of the same particular area as the representative image.

[0081] In 1006, the system may determine a difference between the representative image and the canonical image. For example, the system may determine a difference between the canonical image and the representative image, the identifying a set of pixels of the representative image that are different from corresponding pixels in the canonical image. The system may compare the values of the pixels within the representative image to the values of the pixels within the canonical image to determine the differences.

[0082] In some embodiments, determining the difference between the canonical image and the representative image includes determining distances within a color space using Delta E for related pixels within the representative image and the canonical image. The set of pixels that are identified as being different may be based at least in part on the determined distances within the color space. In some of these embodiments, the color space may be a LAB color space.

[0083] In 1008, the system may modify the set of pixels. For example, the system may modify the set of pixels identified in 1006 based at least in part on previous frames of the video before the representative image to form a modified set of pixels. The modified set of pixels may be smaller than the set of pixels identified in 1006.

[0084] In some embodiments, modifying the set of pixels in 1008 may include applying temporal averaging to the set of pixels. The system may identify high frequency pixels from the set of pixels based at least in part on the temporal averaging indicating that the high frequency pixels have been unstable within the previous frames. The temporal averaging may include averaging amounts of change for each pixel within the set of pixels within the previous frames, wherein the averaged amounts of change for the high frequency pixels are below a temporal averaging threshold. The system may determine that the high frequency pixels have been unstable based at least in part on the values of the high frequency pixels continually changing more than a threshold amount. The system may remove the high frequency pixels from the set of pixels based at least in part on the temporal averaging indicating that the high frequency pixels have been unstable within the previous frames.

[0085] In 1010, the system may average amounts of change. For example, the system may average amounts of change in subgroups of the set of pixels. In particular, the system may average the values of the amounts of change in the subgroups. In some embodiments, 1010 may be omitted.

[0086] In 1012, the system may identify a portion of the subgroups having averages amounts of change that are below a threshold amount of change. The pixels within the identified portion of the subgroups may be low density pixels. In some embodiments, 1012 may be omitted.

5 [0087] In 1014, the system may remove the low density pixels. For example, the system may remove the low density pixels identified in 1012 from the set of pixels. In some embodiments, 1014 may be omitted.

[0088] In 1016, the system may identify a subset of disjointed pixels. For example, the system may identify a subset of disjointed pixels from the modified set of pixels produced in 1008. In some embodiments, 1016 may be omitted.

10 [0089] In 1018, the system may generate a bounding box. For example, the system may generate a bounding box around the subset of disjointed pixels identified in 1016. In some embodiments, 1018 may be omitted.

[0090] In 1020, the system may add additional pixels. For example, the system may add additional pixels to the modified set of pixels produced in 1008 to fill the bounding box. The
15 modified set of pixels may be used in execution of a classifier. In some embodiments, 1020 may be omitted.

[0091] FIG. 11 illustrates a second portion of the example procedure 1000 for identification of an object in accordance with some embodiments. The procedure 1000 may proceed from 1022 illustrated in FIG. 10 to 1022 illustrated in FIG. 11.

20 [0092] In 1102, the system may identify a second subset of disjointed pixels. For example, the system may identify a second subset of disjointed pixels from the modified set of pixels produced in 1008. In some embodiments, 1102 may be omitted.

[0093] In 1104, the system may generate a second bounding box. In some embodiments, the system may generate a second bounding box around the second subset of disjointed pixels. In
25 some embodiments, 1104 may be omitted.

[0094] In 1106, the system may determine that the second bounding box is smaller than a first threshold percentage or larger than a second threshold percentage. For example, the system may determine that the second bounding box is smaller than a first threshold percentage of the

representative image or larger than a second threshold percentage of the representative image. The second threshold percentage may be larger than the first threshold percentage. In some embodiments, 1106 may be omitted.

[0095] In 1108, the system may remove pixels corresponding to the second bounding box.

5 For example, the system may remove pixels corresponding to the second bounding box from the set of pixels. A modified set of pixels with the pixels removed may be used in execution of a classifier. In some embodiments, 1108 may be omitted.

[0096] In 1110, the system may execute a classifier. For example, the system may execute a classifier using the modified set of pixels. The classifier may provide a prediction of an

10 identification of an object represented by a subset of the modified set of pixels.

[0097] In 1112, the system may output a prediction. For example, the system may output the prediction of the identification of the object produced in 1110.

[0098] FIG. 12 illustrates a first portion of another example procedure 1200 for identification of an object in accordance with some embodiments. The procedure 1200 may be performed by a
15 system, such as the system 102 (FIG. 1). For example, the system may receive an image and output a prediction of an identification of the object. While the procedure 1200 is illustrated in a certain order, it should be understood that one or more of the operations within the procedure 1200 may be performed in a different order, concurrently with other operations, and/or may be omitted.

20 [0099] In 1202, the system may capture a representative image. For example, the system may capture a representative image from a video provided to the system. The video may be captured by a camera directed to capture a particular area, where the representative image is an image of the particular area.

[0100] In 1204, the system may retrieve a canonical image. For example, the system may
25 retrieve a canonical image that represents one or more frames of the video before the representative image. The canonical image may be an image of the same particular area as the representative image.

[0101] In 1206, the system may determine a difference between the representative image and the canonical image. For example, the system may determine a difference between the canonical image and the representative image identifying a set of pixels of the representative image that are different from corresponding pixels in the canonical image. The system may compare the values of the pixels within the representative image to the values of the pixels within the canonical image to determine the differences.

[0102] In some embodiments, determining the difference between the canonical image and the representative image includes determining distances within a color space using Delta E for related pixels within the representative image and the canonical image. The set of pixels that are identified as being different may be based at least in part on the determined distances within the color space. In some of these embodiments, the set of pixels may be identified as being different based at least in part on distances within the color space corresponding to the set of pixels being greater than a threshold distance.

[0103] In 1208, the system may modify the set of pixels. For example, the system may modify the set of pixels identified in 1206 based at least in part on previous frames of the video before the representative image to form a modified set of pixels. The modified set of pixels may be smaller than the set of pixels identified in 1206.

[0104] In some embodiments, modifying the set of pixels in 1208 may include applying temporal averaging to the set of pixels. The system may identify high frequency pixels from the set of pixels based at least in part on the temporal averaging indicating that the high frequency pixels have been unstable within the previous frames. The temporal averaging may include averaging amounts of change for each pixel within the set of pixels within the previous frames, wherein the averaged amounts of change for the high frequency pixels are below a temporal averaging threshold. The system may determine that the high frequency pixels have been unstable based at least in part on the values of the high frequency pixels continually changing more than a threshold amount. The system may remove the high frequency pixels from the set of pixels based at least in part on the temporal averaging indicating that the high frequency pixels have been unstable within the previous frames.

[0105] In 1210, the system may average amounts of change in subgroups. For example, the system may average amounts of change in subgroups of the set of pixels. In particular, the

system may average the values of the amounts of change in the subgroups. In some embodiments, 1210 may be omitted.

[0106] In 1212, the system may identify a portion of the subgroups having averages amounts of change that are below a threshold amount of change. The pixels within the identified portion
5 of the subgroups may be low density pixels. In some embodiments, 1212 may be omitted.

[0107] In 1214, the system may remove the low density pixels. For example, the system may remove the low density pixels identified in 1212 from the set of pixels. In some embodiments, 1214 may be omitted.

[0108] In 1216, the system may identify a subset of disjointed pixels. For example, the
10 system may identify a subset of disjointed pixels from the modified set of pixels produced in 1208. In some embodiments, 1216 may be omitted.

[0109] In 1218, the system may generate a bounding box. For example, the system may generate a bounding box around the subset of disjointed pixels identified in 1216. In some embodiments, 1218 may be omitted.

15 [0110] In 1220, the system may add additional pixels. For example, the system may add additional pixels to the modified set of pixels produced in 1208 to fill the bounding box. The modified set of pixels may be used in execution of a classifier. In some embodiments, 1220 may be omitted.

[0111] FIG. 13 illustrates a second portion of the example procedure 1200 for identification
20 of an object in accordance with some embodiments. The procedure 1200 may proceed from 1222 illustrated in FIG. 12 to 1222 illustrated in FIG. 13.

[0112] In 1302, the system may execute a classifier. For example, the system may execute a classifier using the modified set of pixels. The classifier may provide a prediction of an identification of an object represented by a subset of the modified set of pixels.

25 [0113] In 1304, the system may output a prediction. For example, the system may output the prediction of the identification of the object produced in 1302.

[0114] In 1306, the system may determine notifications. For example, the system may determine notification setting associated with a camera that captures the video indicate that a

notification is to be provided when the prediction is that the object is a package. In some embodiments 1306 may be omitted.

[0115] In 1308, the system may cause a notification to be provided. For example, the system may cause the notification to be provided based at least in part on the prediction that the object is a package. The notification being provided may cause the system to emit a sound and/or display an image indicating that a package has been placed in or removed from the particular area, and/or transmitting a signal to a device that causes the device to emit a sound and/or display an image indicating that a package has been placed in or removed from the particular area.

[0116] FIG. 14 illustrates an example procedure 1400 for identification of an object in accordance with some embodiments. The procedure 1400 may be performed by a system, such as the system 102 (FIG. 1). For example, the system may receive an image and output a prediction of an identification of the object. While the procedure 1400 is illustrated in a certain order, it should be understood that one or more of the operations within the procedure 1400 may be performed in a different order, concurrently with other operations, and/or may be omitted.

[0117] In 1402, the system may capture a representative image. For example, the system may capture a representative image from stored images stored by the system. The stored images may be captured by a camera directed to capture a particular area, where the representative image is an image of the particular area.

[0118] In 1404, the system may retrieve a canonical image. For example, the system may retrieve a canonical image from the stored images that represents one or more frames of the video before the representative image. The canonical image may be an image of the same particular area as the representative image.

[0119] In 1406, the system may determine a difference between the representative image and the canonical image. For example, the system may determine a difference between the canonical image and the representative image, the difference identifying a set of pixels of the representative image that are different from corresponding pixels in the canonical image. The system may compare the values of the pixels within the representative image to the values of the pixels within the canonical image to determine the differences.

[0120] In some embodiments, determining the difference between the canonical image and the representative image includes determining distances between the canonical image and the representative image in a LAB color space. The difference may be determined based at least in part on the distances.

- 5 [0121] In 1408, the system may modify the set of pixels. For example, the system may modify the set of pixels identified in 1406 based at least in part on previous frames of the video before the representative image to form a modified set of pixels. The modified set of pixels may be smaller than the set of pixels identified in 1406.

10 [0122] In some embodiments, modifying the set of pixels in 1208 may include applying temporal averaging to the set of pixels. The system may identify high frequency pixels from the set of pixels based at least in part on the temporal averaging indicating that the high frequency pixels have been unstable within the one or more frames. The temporal averaging may include averaging amounts of change for each pixel within the set of pixels within the one or more frames, wherein the averaged amounts of change for the high frequency pixels are below a
15 temporal averaging threshold. The system may determine that the high frequency pixels have been unstable based at least in part on the values of the high frequency pixels continually changing more than a threshold amount. The system may remove the high frequency pixels from the set of pixels based at least in part on the temporal averaging indicating that the high frequency pixels have been unstable within the one or more frames.

- 20 [0123] In 1410, the system may execute a classifier. For example, the system may execute a classifier using the modified set of pixels. The classifier may provide a prediction of an identification of an object represented by a subset of the modified set of pixels.

[0124] In 1412, the system may output a prediction. For example, the system may output the prediction of the identification of the object produced in 1410.

25 IV. UNIFIED SYSTEM ARRANGEMENT

[0125] FIG. 15 illustrates an example unified system arrangement 1500 in accordance with some embodiments. For example, the unified system arrangement 1500 may be a system 1502 that can perform both object detection and package detection. For example, the system 1502 may utilize a shared machine learning detector to both detect objects and identify delivery and/or

pickup of packages. Objects, as used herein, may refer to humans, animals, vehicles, and/or any other movable entities other than packages that a machine learning algorithm may be trained to identify. Further, the system 1500 may indicate to a user when a package has been delivered and/or picked up, and/or when an object is detected.

5 [0126] The system arrangement 1500 may include the system 1502. The system 1502 may include one or more of the features of the system 102 (FIG. 1). The system 1502 may comprise a computing system that may be able to execute one or more instructions to perform one or more operations. In some embodiments, the system 1502 may comprise a smart doorbell, a security system, a computer system, a distributed computer system, a mobile phone (such as a smart
10 phone), a home environment made up of various connected user devices (e.g., a home hub, controller devices, smart television systems, accessories, etc.), or portions thereof.

[0127] The system 1502 may include memory 1504 and one or more processors 1506 coupled to the memory 1504. The memory 1504 may include one or more computer-readable media. In some embodiments, the computer-readable media may comprise non-transitory
15 computer-readable media. The memory 1504 may have one or more instructions stored thereon, wherein the instructions, when executed by the system 1502, may cause the system 1502 to perform one or more of the operations disclosed herein. The memory 1504 may include any suitable volatile or non-volatile memory such as, but not limited to, dynamic random access memory (DRAM), static random access memory (SRAM), eraseable programmable read only
20 memory (EPROM), electrically eraseable programmable read only memory (EEPROM), flash memory, solid-state memory, any other type of memory device technology, or some combination thereof.

[0128] The processors 1506 may be separate from the memory 1504 and/or coupled to the memory 1504, or the memory 1504 may be implemented as part of the processors 1506 and may
25 be coupled to the memory 1504. The processors 106 may receive the instructions from the memory 1504 and execute the instructions to cause the system 1502 to perform one or more of the operations described throughout this disclosure. The processors 1506 may include processor circuitry such as, for example, baseband processor circuitry, central processor unit circuitry, graphics processor unit circuitry, or some combination thereof.

[0129] The system arrangement 1500 may further include an input device 1508. In some embodiments, the input device 1508 may comprise one or more cameras, where each of the cameras may capture images and/or video (which may comprise a series of images captured temporally by the camera) of an area. The areas for which the cameras are to capture images and/or videos may include areas where packages are intended to be placed for delivery and/or pickup of the packages, such as outside of structures, within unsecured portions of structures, within package depository locations, and/or other areas where packages are to be placed for delivery and/or pickup.

[0130] The input device 1508 may be coupled to the system 1502 and may provide captured images and/or videos to the system 1502. The system 1502 may store the images and/or video received from the input device 1508 in the memory 1504 of the system 1502. For example, the illustrated embodiment shows a current image 1510 and a first image 1512 stored in the memory 1504 of the system 1502. The current image 1510 and the first image 1512 may have been provided to the system 1502 by the input device 1508. The current image 1510 may be a latest image received by the system 1502 from the input device 1508, where the image may be part of a video in some embodiments. The first image 1512 may be an oldest image stored in a memory 1504 of the device in some embodiments. In other embodiments, the first image 1512 may be an image captured prior to detection of motion or a first image received when motion has been detected. It should be understood that the current image 1510 and the first image 1512 are two examples of images that can be stored by the memory 1504 and that the memory 1504 may store more images from the input device 1508.

[0131] In some embodiments, the images and/or portions of the video may be stored by the system 1502 in the memory 1504 for a period of time, where the images and/or the portions of the video may be removed from the memory 1504 at the expiration of the period of time.

Removing the images and/or portions of the video from the memory 1504 may ensure that there is space within the memory 1504 for additional images and/or portions of the video to be stored. In some embodiments, the storage of the images and/or portions of the video may be cyclical, where the oldest images and/or portions of the video may be replaced by new images and/or portions of the video received from the input device 1508 (e.g., in a first in, first out (FIFO) manner).

[0132] The system 1502 may include an object candidate identifier 1514. The object candidate identifier 1514 may comprise computer hardware components, instructions that can be performed by the processors 1506, or some combination thereof (which may be referred to as an “object candidate module” in some instances). The object candidate identifier 1514 may receive images and identify possible candidates for an object within the images. For example, the object candidate identifier 1514 may perform, or cause the system 1502 to perform via the processors 1506, one or more operations related to identification of possible candidates for objects from images described further throughout this disclosure.

[0133] The system 1502 may include a package candidate identifier 1516. The package candidate identifier 1516 may comprise computer hardware components, instructions that can be performed by the processors 1506, or some combination thereof (which may be referred to as a “package candidate module” in some instances). The package candidate identifier 1516 may receive images and identify possible candidates for a package within the images. For example, the package candidate identifier 1516 may perform, or cause the system 1502 to perform via the processors 1506, one or more operations related to identification of possible candidates for objects from images described further throughout this disclosure.

[0134] The system 1502 may include a shared machine learning classifier 1518. The shared machine learning detector 1518 may comprise computer hardware components, instructions that can be performed by the processors 1506, or some combination thereof (which may be referred to as a “classifier module” in some instances). The shared machine learning classifier 1518 may receive indications of package candidates within an image from the package candidate identifier 1516 and/or indications of object candidates within an image from the object candidate identifier 1514. For example, the shared machine learning classifier 1518 may receive a defined frame from the image that includes the package candidates and/or the object candidates. The shared machine learning classifier 1518 may analyze the package candidates, the object candidates, and/or the defined frame, and may classify the candidates or portions of the define frame as packages, objects, or not packages and objects. For example, the shared machine learning classifier 1518 may perform, or cause the system 1502 to perform via the processors 1506, one or more operations related to classification of candidates and/or portions of defined frames from images described further throughout this disclosure.

[0135] The system arrangement 1500 may include a notification device 1520. The notification device 1520 may comprise a computing system such as a smart doorbell, a security system, a computer system, a distributed computer system, a mobile phone (such as a smart phone), an accessory within a home environment, a smart television (e.g., a media streaming device), or portions thereof. The notification device 1520 may provide a notification to a user that delivery and/or pickup of a package has been identified or otherwise detected by the system 1502, that an object has been identified or otherwise detected by the system 1502, or some combination thereof. For example, the notification device 1520 may be coupled to the system 1502 and may receive an indication from the system 1502 to provide notification that the system 1502 has identified delivery, pickup of a package, and/or detection of an object. The notification device 1520 may provide a visual notification, a sound notification, a haptic notification, another type of notification, or some combination thereof to a user of the notification device 1520. For example, the notification device 1520 may display an image of the area with the package and/or the object on a display of the notification device 1520 in response to receiving the indication from the system 1502 in some embodiments. In some embodiments, the notification device 1520 may present the notification via email, pop-up, an application, sound from a speaker, a motion generator, or some combination thereof.

[0136] While the input device 1508 and the notification device 1520 are illustrated separate from the system 1502 in the illustrated system arrangement 1500, it should be understood that the input device 1508 and/or the notification device 1520 may be included in the system 1502 in other embodiments. For example, the system 1502 may comprise a smart doorbell or security system in some embodiments where the input device 1508 comprises a camera of the smart doorbell or the security system. Further, the notification device 1520 may comprise a speaker of the smart doorbell or the security system in embodiments where the system 1502 comprises the smart doorbell or the security system, where the notification device 1520 may emit a sound to indicate that delivery and/or pickup of a package has been detected by the system 1502. For example, the notification device 1520 may provide a notification by emitting an audible sound by a smart speaker (such as a home hub) or displaying a notification on a television by a media streaming device (such as a smart television).

V. UNIFIED SYSTEM OPERATION

[0137] FIG. 16 illustrates an example procedure 1600 for identifying objects and/or package in accordance with some embodiments. The procedure 1600 may be performed by a system (such as the system 1502 (FIG. 15)). For example, an object candidate identifier (such as the object candidate identifier 1514 (FIG. 15)), a package candidate identifier 1516 (such as the package candidate identifier 1516 (FIG. 15)), and/or a shared machine learning detector (such as the shared machine learning detector 1518 (FIG. 15)) may perform, or cause the system to perform, one or more of the operations of the procedure 1600. The system may output predictions, and/or initiate notifications, of packages and/or objects based on the performance of the procedure 1600.

[0138] In 1602, the system may identify package candidates from a current image (such as the current frame 1510 (FIG. 15)) received from an input device (such as the input device 1508 (FIG. 15)). For example, the system may analyze the current image to determine whether the current image has captured any package candidates. Identifying the package candidates may include one or more of the operations of the procedure 200 (FIG. 2), where the current image may be treated as the representative image. For example, differences between the current image and a canonical image may be determined, low density pixels may be removed, high frequency pixels may be removed, a flood-fill process may be executed, a candidate age may be determined, young pixels may be removed, or some combination thereof. The system may maintain the position of the package candidates within the current image.

[0139] In 1604, the system may identify object candidates from the current image. Object candidates may be identified based on motion within the area captured by the current image. For example, the system may compare the current image to one or more previous images to determine whether there have been changes in pixel values that may indicate motion. In some embodiments, a change in pixel values greater than a threshold may indicate motion. The candidates may include pixels with changes in value that may indicate motion. In some embodiments, an object candidate may be a rectangular (or other defined shape) area of the current image that includes the pixels with changes in value that may indicate motion. The system may maintain the position of the object candidates within the current image.

[0140] In 1606, the system may determine an analysis frame for the first image. For example, the system may determine an area of the current image that includes all of the identified package candidates and object candidates from 1602 and 1604 for the analysis frame. In some embodiments, the analysis frame may be a rectangular (or other defined shape) area of the current image that includes all of the identified package candidates and object candidates.

[0141] FIG. 17 illustrates an example analysis frame arrangement 1700 in accordance with some embodiments. In particular, the analysis frame arrangement 1700 illustrates the determination of an analysis frame for an image.

[0142] In the illustrated embodiment, the analysis frame arrangement 1700 may have a package candidate 1702 and an object candidate 1704 that were identified within the image. The package candidate 1702 may have been identified via 1602 and the object candidate 1704 may have been identified via 1604. The locations of the package candidate 1702 and the object candidate 1704 may be maintained within the image, as is shown by the package candidate 1702 and the object candidate 1704 being shown in their respective locations within the image.

[0143] The system may determine an analysis frame 1706 for the image. The analysis frame 1706 may include both the pixels corresponding to the package candidate 1702 and the object candidate 1704, as is shown by the package candidate 1702 and the object candidate 1704 being located within the analysis frame 1706. The analysis frame 1706 may be predefined to be a certain shape, such as a rectangle in the illustrated embodiment. The analysis frame 1706 may be defined to be a smallest area having the predefined shape and including all the candidates in some embodiments. In other embodiments, the edges of the analysis frame 1706 may be a defined distance from the candidates with the analysis frame 1706 including the object candidates. The analysis frame 1706 may include all the pixels defined by the area of the analysis frame 1706 and not just pixels defined by the package candidates and the object candidates.

[0144] In 1608, the system may perform machine learning classification for the current image. For example, the system may analyze the analysis frame determined in 1606 to determine whether any packages and/or objects are predicted to be located in the area represented by the analysis frame. The analysis of the analysis frame may be performed by a shared machine learning classifier (such as the shared machine learning classifier 1518 (FIG. 15)) that has been

trained to identify packages and objects. By having a shared machine learning classifier perform the analysis to identify both packages and objects, a single invocation of the shared machine learning classifier may be implemented to identify all the packages and/or objects within the area defined by the analysis frame rather than having to invoking a machine learning classifier multiple times when multiple candidates have been identified.

[0145] Further, the shared machine learning classifier may analyze the package candidates and the object candidates for both packages and objects, where reliability scores between a package identification and an object identification may be compared to determine which prediction is more reliable. If the shared machine learning classifier determines that the package identification is more reliable based on the reliability scores, the shared machine learning classifier may predict the candidate as a package regardless of whether the candidate was a package candidate or an object candidate. Conversely, if the shared machine learning classifier determines that the object identification is more reliable based on the reliability scores, the shared machine learning classifier may predict the candidate as an object regardless of whether the candidate was a package candidate or an object candidate. The machine learning classifier may further determine the particular object (such as a human, animal, vehicle, or other object) in instances where the prediction is that the candidate is an object.

[0146] In instances where a package candidate and an object candidate overlap, the shared machine learning classifier may perform the analysis of the object candidate and avoid analysis of the package candidate. In particular, if the shared machine learning classifier determines that the package candidate and the object candidate overlap by a certain amount, the shared machine learning classifier may perform analysis of the object candidate and not the package candidate to determine whether a package and/or an object is to be predicted for the object candidate.

[0147] In instances where the shared machine learning classifier predicts that a package is present in the current image, the procedure 1600 may proceed with 1610 through 1618. If shared machine learning classifier does not predict that a package is present in the current image, 1610 through 1618 may be omitted. Further, 1610 through 1618 may be omitted in other embodiments. 1610 through 1618 may be performed to determine whether any identified packages are newly delivered (for example, the system had not previously detected the package).

[0148] In 1610, the system may identify package candidates from a first image (such as the first image 1512 (FIG. 15)). For example, the system may perform the same operations as 1602, but with the first image rather than the current image. Accordingly, the system may perform one or more of the operations of the procedure 200 to determine whether there are any package candidates within the first frame.

[0149] In 1612, the system may identify object candidates from the first image. For example, the system may perform the same operations as 1604, but with the first image rather than the current image. Accordingly, the system may determine whether any changes in values of pixels within the first image indicate motion, and may identify object candidates for the first image based on the indication of motion.

[0150] In 1614, the system may determine an analysis frame for the first image. For example, the system may perform the same operations as 1606, but with the package candidates and object candidates from the first image rather than the package candidates and object candidates from the current image. The system may determine an area of the first image that includes all of the identified package candidates and object candidates from 1610 and 1616 for the analysis frame of the first image. In some embodiments, the analysis frame may be a rectangular (or other defined shape) area of the first image that includes all of the identified package candidates and object candidates.

[0151] In 1616, the system may perform machine learning detection for the first image. For example, the system may perform the same operations as 1608, but with the analysis frame for the first image rather than the analysis frame for the current image. The system may analyze the analysis frame determined in 1614 to determine whether any packages and/or objects are located in the area represented by the analysis frame. The analysis of the analysis frame may be performed by the shared machine learning classifier that has been trained to identify packages and objects.

[0152] In 1618, the system may compare the predictions from the first image with the predictions from the current image. For example, the system may compare the package predictions from the first image with the package predictions from the current image. The system may compare the locations of the predicted packages within the first image with the predicted packages within the current image to determine whether the predicted packages of the first image

and the predicted packages of the current image are located in same location. In particular, the first image and the current image are images of a same area, such that locations within the first image correspond to the same locations within the current image. The system may determine that a package predicted from the first image located in the same location as a package predicted
5 from the current image are the same package. If the system determines that a package predicted from the first image and a package predicted from the current image are the same package, the system may determine that the package is not newly delivered and may determine not to initiate a notification. In some embodiments, the system may determine that a package predicted from the first image and a package predicted from the current image are the same package based on a
10 certain amount of overlap in locations of the package predicted from the first image and the package predicted from the current image. Further, the system may compare the values of the pixels for the package predicted from the first image and the package predicted from the current image in some embodiments, where a change in the values below a threshold may indicate that the packages are the same package and a change in the values above the threshold may indicate
15 the packages are different packages. In some embodiments, the same comparison of the locations may be performed for the predicted objects to determine whether the objects are newly located within the area.

[0153] In 1620, the system may output a prediction. For example, the system may output the prediction of packages and/or objects located within the area captured by the current image. In
20 some embodiments, the prediction may include the packages and/or objects predicted from 1608. In embodiments where 1610 through 1618 are performed, the prediction may include the newly delivered packages and/or the objects newly located within the area, and may exclude any packages and/or objects that were determined to be same package and/or object in the first image and the current image. The system may output the prediction to a notification device (such as the
25 notification device 1520 (FIG. 15)).

[0154] In 1622, the notification device may determine notification settings associated with the cameras and/or the notification device. For example, the notification device may have stored notification settings that indicate when the notification device is to initiate a notification. The notification settings may be presented to a user of the notification device and the user may
30 indicate the events for which the notification device is to initiate a notification. In some

embodiments, the notification settings may include one or more selections indicating whether the notification device is to provide a notification for a package being delivered, a package being picked up, an object being located within the area, or some combination thereof. In some embodiments, the notification settings may include one or more selections indicating a type of notification (such as a visual notification, a sound notification, a haptic notification, another type of notification, or some combination thereof) that the notification device is to utilize for the notification. The user may utilize the selections to indicate whether the notification device is to provide the notification and/or the type of notification that is to be provided. In response to the notification device receiving a prediction from the system, the notification device may review the notification settings and determine whether a notification is to be initiated and/or what type of notification is to be initiated based on the prediction.

[0155] In 1624, the notification device may initiate a notification in some instances. In particular, the notification system may initiate a notification based at least in part on the predication and the determination if a notification is to be provided in 1622. If the notification device determines that a notification is to be provided based on the notification settings in 1622, the notification system may initiate the notification in accordance with the notification settings. If the notification device determines that a notification is not to be provided based on the notification settings in 1622, 1624 may be omitted. Providing the notification may include providing a notification via a smart doorbell, a security system, a computer system, a distributed computer system, a mobile phone (such as a smart phone), an accessory within a home environment, a smart television (e.g., a media streaming device), or portions thereof.

Examples

[0156] In the following sections, further exemplary embodiments are provided.

[0157] Example 1 may include a method, comprising capturing a representative image from a video, retrieving a canonical image that represents one or more frames of the video before the representative image, determining a difference between the canonical image and the representative image, the difference identifying a set of pixels of the representative image that are different from corresponding pixels in the canonical image, modifying the set of pixels based at least in part on previous frames of the video before the representative image to form a modified set of pixels, the modified set of pixels being smaller than the set of pixels, executing a

classifier using the modified set of pixels, the classifier providing a prediction of an identification of an object represented by a subset of the modified set of pixels, and outputting the prediction of the identification of the object.

[0158] Example 2 may include the method of example 1, wherein determining the difference between the canonical image and the representative image includes determining distances within a color space using Delta E for related pixels within the representative image and the canonical image, and wherein the set of pixels that are identified as being different is based at least in part on the determined distances within the color space.

[0159] Example 3 may include the method of example 2, wherein the color space is a LAB color space.

[0160] Example 4 may include the method of example 1, further comprising averaging amounts of change in subgroups of the set of pixels, identifying a portion of the subgroups having averaged amounts of change that are below a threshold amount of change, wherein pixels within the portion of the subgroups comprise low density pixels, and removing the low density pixels from the set of pixels.

[0161] Example 5 may include the method of example 1, further comprising identifying a subset of disjointed pixels from the modified set of pixels, generating a bounding box around the subset of disjointed pixels, and adding additional pixels to the modified set of pixels to fill the bounding box, wherein the modified set of pixels with the additional pixels is used in the executing of the classifier.

[0162] Example 6 may include the method of example 5, wherein the subset of disjointed pixels is a first subset of disjointed pixels, and wherein the method further comprises identifying a second subset of disjointed pixels from the modified set of pixels, generating a second bounding box around the second subset of disjointed pixels, determining that the second bounding box is smaller than a first threshold percentage of the representative image or larger than a second threshold percentage of the representative image, the second threshold percentage being larger than the first threshold percentage, and removing pixels corresponding to the second bounding box from the set of pixels, wherein the modified set of pixels with the pixels removed is used in the executing of the classifier.

[0163] Example 7 may include the method of example 1, wherein modifying the set of pixels includes applying temporal averaging to the set of pixels, identifying high frequency pixels from the set of pixels based at least in part on the temporal averaging indicating that the high frequency pixels have been unstable within the previous frames, and removing the high frequency pixels from the set of pixels based at least in part on the temporal averaging indicating that the high frequency pixels have been unstable within the previous frames.

[0164] Example 8 may include the method of example 7, wherein applying the temporal averaging includes averaging amounts of change for each pixel within the set of pixels within the previous frames, and wherein the averaged amounts of change for the high frequency pixels are below a temporal averaging threshold.

[0165] Example 9 may include one or more computer-readable media having instructions stored thereon, wherein the instructions, when executed by a system, cause the system to capture a representative image of an area from a video, retrieve a canonical image of the area that represents one or more frames of the video before the representative image, determine a difference between the canonical image and the representative image, the difference identifying a set of pixels of the representative image that are different from corresponding pixels in the canonical image, modify the set of pixels based at least in part on previous frames of the video before the representative image to form a modified set of pixels, the modified set of pixels being smaller than the set of pixels, execute a classifier using the modified set of pixels, the classifier providing a prediction of an identification of an object represented by a subset of the modified set of pixels, and output the prediction of the identification of the object.

[0166] Example 10 may include the one or more computer-readable media of example 9, wherein the prediction of the object includes a prediction that the object is a package, and wherein the instructions, when executed by the system, further cause the system to determine that notification settings associated with a camera that captures the video indicate that a notification is to be provided when the prediction is that the object is a package, and cause the notification to be provided based at least in part on the prediction that the object is a package.

[0167] Example 11 may include the one or more computer-readable media of example 9, wherein to determine the difference between the canonical image and the representative image includes to determine distances within a color space using Delta E for related pixels within the

canonical image and the representative image, and wherein the set of pixels that are identified as being different is based at least in part on the determined distances within the color space.

[0168] Example 12 may include the one or more computer-readable media of example 11, wherein the set of pixels that are identified as being different are identified based at least in part on distances within the color space corresponding to the set of pixels being greater than a threshold distance.

[0169] Example 13 may include the one or more computer-readable media of example 9, wherein the instructions, when executed by the system, further cause the system to average an amount of change in subgroups of the set of pixels, identify a portion of the subgroups having averaged amounts of change that are below a threshold amount of change, wherein pixels within the portion of the subgroups comprise low density pixels, and remove the low density pixels from the set of pixels.

[0170] Example 14 may include the one or more computer-readable media of example 9, wherein the instructions, when executed by the system, further cause the system to identify a first subset of disjointed pixels from the modified set of pixels, generate a bounding box around the first subset of disjointed pixels, and adding additional pixels to the modified set of pixels to fill the bounding box, wherein the modified set of pixels with the additional pixels is used in the execution of the classifier.

[0171] Example 15 may include the one or more computer-readable media of example 14, wherein to modify the set of pixels includes to apply temporal averaging to the set of pixels, identify high frequency pixels from the set of pixels based at least in part on the temporal averaging indicating that the high frequency pixels have been unstable within the previous frames, and remove the high frequency pixels based at least in part on the temporal averaging indicating that the high frequency pixels have been unstable within the previous frames.

[0172] Example 16 may include the one or more computer-readable media of example 15, wherein to apply the temporal averaging includes to average amounts of change for each pixel within the set of pixels within the previous frames, and wherein the averaged amounts of change for the high frequency pixels are below a temporal averaging threshold.

[0173] Example 17 may include a system, comprising memory to store images from video received from a camera, and one or more processors coupled to the memory, the one or more processors to capture, from stored images, a representative image, retrieve, from the stored images, a canonical image that represents one or more frames of the video before the
5 representative image, determine a difference between the canonical image and the representative image, the difference identifying a set of pixels of the representative image that are different from corresponding pixels in the canonical image, modify the set of pixels based at least in part on previous frames of the video before the representative image to form a modified set of pixels, the modified set of pixels being smaller than the set of pixels, execute a classifier using the
10 modified set of pixels, the classifier providing a prediction of an identification of an object represented by a subset of the modified set of pixels, and output the prediction of the identification of the object.

[0174] Example 18 may include the system of example 17, wherein to determine the difference between the canonical image and the representative image includes to determine
15 distances between the canonical image and the representative image in a LAB color space, the difference determined based at least in part on the distances.

[0175] Example 19 may include the system of example 17, wherein to modify the set of pixels includes to apply temporal averaging to the set of pixels, identify high frequency pixels from the set of pixels based at least in part on the temporal averaging indicating that the high
20 frequency pixels have been unstable within the one or more frames, and remove the high frequency pixels based at least in part on the temporal averaging indicating that the high frequency pixels have been unstable within the one or more frames.

[0176] Example 20 may include the system of example 19, wherein to apply the temporal averaging includes to average amounts of change for each pixel within the set of pixels within
25 the one or more frames, and wherein the averaged amounts of change for the high frequency pixels are below a temporal averaging threshold.

[0177] Example 21 may include a method, comprising identifying one or more package candidates from an image, determining an analysis frame from the image based at least in part on the one or more package candidates, executing a shared classifier on the analysis frame, the
30 shared classifier to predict identification of packages and objects within the analysis frame, and

outputting a prediction of the identification of the packages and the objects based at least in part on the execution of the shared classifier.

[0178] Example 22 may include the method of example 21, wherein identifying the one or more package candidates from the image includes determining a difference between the image and a canonical image, wherein the image is from a video, wherein the canonical image represents one or more frames of the video before the image, and wherein determining the difference includes identifying a set of pixels of the image that are different from corresponding pixels in the canonical image.

[0179] Example 23 may include the method of example 22, wherein identifying the one or more package candidates includes modifying the set of pixels based at least in part on previous frames of the video before the image to form a modified set of pixels, the modified set of pixels being smaller than the set of pixels.

[0180] Example 24 may include the method of example 22, wherein determining the difference between the image and the canonical image includes determining distances within a color space using Delta E for related pixels within the image and the canonical image, and wherein the set of pixels that are identified as being different is based at least in part on the determined distances within the color space.

[0181] Example 25 may include the method of example 24, wherein the color space is a LAB color space.

[0182] Example 26 may include the method of example 22, further comprising averaging amounts of change in subgroups of the set of pixels, identifying a portion of the subgroups having averaged amounts of change that are below a threshold amount of change, wherein pixels within the portion of the subgroups comprise low density pixels, and removing the low density pixels from the set of pixels.

[0183] Example 27 may include the method of example 23, further comprising identifying a subset of disjointed pixels from the modified set of pixels, generating a bounding box around the subset of disjointed pixels, and adding additional pixels to the modified set of pixels to fill the bounding box, wherein the modified set of pixels with the additional pixels is used in the executing of the shared classifier.

[0184] Example 28 may include the method of example 23, further comprising applying temporal averaging to the set of pixels, identifying high frequency pixels from the set of pixels based at least in part on the temporal averaging indicating that the high frequency pixels have been unstable within the previous frames, and removing the high frequency pixels from the set of pixels based at least in part on the temporal averaging indicating that the high frequency pixels have been unstable within the previous frames.

[0185] Example 29 may include the method of example 21, further comprising identifying one or more object candidates from the image, wherein determining the analysis frame is further based at least in part on the one or more object candidates.

[0186] Example 30 may include the method of example 29, wherein identifying the one or more object candidates comprises identifying the one or more object candidates based at least in part on motion identified from the image.

[0187] Example 31 may include the method of example 21, wherein the image is a current image, wherein the prediction includes a first package predicted within the current image, wherein the analysis frame is a first analysis frame, and wherein the method further comprises identifying one or more package candidates from a first image, the first image captured prior to the current image, determining a second analysis frame from the first image based on the one or more package candidates from the first image, executing a shared classifier on the second analysis frame, and comparing a first package predicted from the current image with a second package predicted from the first image to determine whether the first package and the second package are a same package.

[0188] Example 32 may include the method of example 31, wherein comparing the first package and the second package comprises comparing a location of the first package and a location of the second package.

[0189] Example 33 may include one or more computer-readable media having instructions stored thereon, wherein the instructions, when executed by a system, cause the system to identify one or more package candidates from an image, determine an analysis frame from the image based at least in part on the one or more package candidates, execute a shared classifier on the analysis frame, the shared classifier to predict identification of packages and objects within the

analysis frame, and output a prediction of the identification of the packages and the objects based at least in part on the execution of the shared classifier.

[0190] Example 34 may include the one or more computer-readable media of example 33, wherein to identify the one or more package candidates from the image includes to determine a
5 difference between the image and a canonical image, wherein the image is from a video, wherein the canonical image represents one or more frames of the video before the image, and wherein to determine the difference includes to identify a set of pixels of the image that are different from corresponding pixels in the canonical image.

[0191] Example 35 may include the one or more computer-readable media of example 34,
10 wherein to identify the one or more package candidates includes to modify the set of pixels based at least in part on previous frames of the video before the image to form a modified set of pixels, the modified set of pixels being smaller than the set of pixels.

[0192] Example 36 may include the one or more computer-readable media of example 34, wherein to determine the difference between the image and the canonical image includes to
15 determine distances within a color space using Delta E for related pixels within the image and the canonical image, and wherein the set of pixels that are identified as being different is based at least in part on the determined distances within the color space.

[0193] Example 37 may include the one or more computer-readable media of example 36, wherein the color space is a LAB color space.

[0194] Example 38 may include the one or more computer-readable media of example 34,
20 wherein the instructions, when executed by the system, further cause the system to average amounts of change in subgroups of the set of pixels, identify a portion of the subgroups having averaged amounts of change that are below a threshold amount of change, wherein pixels within the portion of the subgroups comprise low density pixels, and remove the low density pixels
25 from the set of pixels.

[0195] Example 39 may include the one or more computer-readable media of example 35, wherein the instructions, when executed by the system, further cause the system to identify a subset of disjointed pixels from the modified set of pixels, generate a bounding box around the subset of disjointed pixels, and add additional pixels to the modified set of pixels to fill the

bounding box, wherein the modified set of pixels with the additional pixels is used in the executing of the shared classifier.

[0196] Example 40 may include the one or more computer-readable media of example 35, wherein the instructions, when executed by the system, further cause the system to apply

5 temporal averaging to the set of pixels, identify high frequency pixels from the set of pixels based at least in part on the temporal averaging indicating that the high frequency pixels have been unstable within the previous frames, and remove the high frequency pixels from the set of pixels based at least in part on the temporal averaging indicating that the high frequency pixels have been unstable within the previous frames.

10 [0197] Example 41 may include the one or more computer-readable media of example 33, wherein the instructions, when executed by the system, further cause the system to identify one or more object candidates from the image, wherein to determine the analysis frame is further based at least in part on the one or more object candidates.

[0198] Example 42 may include the one or more computer-readable media of example 41, 15 wherein to identify the one or more object candidates comprises to identify the one or more object candidates based at least in part on motion identified from the image.

[0199] Example 43 may include the one or more computer-readable media of example 33, wherein the image is a current image, wherein the prediction includes a first package predicted within the current image, wherein the analysis frame is a first analysis frame, and wherein the 20 instructions, when executed by the system, further cause the system to identify one or more package candidates from a first image, the first image captured prior to the current image, determine a second analysis frame from the first image based on the one or more package candidates from the first image, execute a shared classifier on the second analysis frame, and compare a first package predicted from the current image with a second package predicted from 25 the first image to determine whether the first package and the second package are a same package.

[0200] Example 44 may include the one or more computer-readable media of example 43, wherein to compare the first package and the second package comprises to compare a location of the first package and a location of the second package.

[0201] Example 45 may include an apparatus comprising means to perform one or more elements of a method described in or related to any of examples 1-44, or any other method or process described herein.

[0202] Example 46 may include one or more non-transitory computer-readable media comprising instructions to cause an electronic device, upon execution of the instructions by one or more processors of the electronic device, to perform one or more elements of a method described in or related to any of examples 1-44, or any other method or process described herein.

[0203] Example 47 may include an apparatus comprising logic, modules, or circuitry to perform one or more elements of a method described in or related to any of examples 1-44, or any other method or process described herein.

[0204] Example 48 may include a method, technique, or process as described in or related to any of examples 1-44, or portions or parts thereof.

[0205] Example 49 may include an apparatus comprising: one or more processors and one or more computer-readable media comprising instructions that, when executed by the one or more processors, cause the one or more processors to perform the method, techniques, or process as described in or related to any of examples 1-44, or portions thereof.

[0206] Example 50 may include a signal as described in or related to any of examples 1-44, or portions or parts thereof.

[0207] Example 51 may include a signal encoded with data as described in or related to any of examples 1-44, or portions or parts thereof, or otherwise described in the present disclosure.

[0208] Any of the above-described examples may be combined with any other example (or combination of examples), unless explicitly stated otherwise. The foregoing description of one or more implementations provides illustration and description, but is not intended to be exhaustive or to limit the scope of embodiments to the precise form disclosed. Modifications and variations are possible in light of the above teachings or may be acquired from practice of various embodiments.

[0209] Although the embodiments above have been described in considerable detail, numerous variations and modifications will become apparent to those skilled in the art once the

above disclosure is fully appreciated. It is intended that the following claims be interpreted to embrace all such variations and modifications.

[0210] Additionally, although embodiments have been described using a particular series of transactions and steps, it should be apparent to those skilled in the art that the scope of the present disclosure is not limited to the described series of transactions and steps. Various features and aspects of the above-described embodiments may be used individually or jointly.

[0211] Further, while embodiments have been described using a particular combination of hardware and software, it should be recognized that other combinations of hardware and software are also within the scope of the present disclosure. Embodiments may be implemented only in hardware, or only in software, or using combinations thereof. The various processes described herein can be implemented on the same processor or different processors in any combination. Accordingly, where components or modules are described as being configured to perform certain operations, such configuration can be accomplished, e.g., by designing electronic circuits to perform the operation, by programming programmable electronic circuits (such as microprocessors) to perform the operation, or any combination thereof. Processes can communicate using a variety of techniques including but not limited to conventional techniques for inter process communication, and different pairs of processes may use different techniques, or the same pair of processes may use different techniques at different times.

[0212] The specification and drawings are, accordingly, to be regarded in an illustrative rather than a restrictive sense. It will, however, be evident that additions, subtractions, deletions, and other modifications and changes may be made thereunto without departing from the broader spirit and scope as set forth in the claims. Thus, although specific disclosure embodiments have been described, these are not intended to be limiting. Various modifications and equivalents are within the scope of the following claims.

[0213] The use of the terms “a” and “an” and “the” and similar referents in the context of describing the disclosed embodiments (especially in the context of the following claims) are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context. The terms “comprising,” “having,” “including,” and “containing” are to be construed as open-ended terms (i.e., meaning “including, but not limited to,”) unless otherwise noted. The term “connected” is to be construed as partly or wholly

contained within, attached to, or joined together, even if there is something intervening. Recitation of ranges of values herein are merely intended to serve as a shorthand method of referring individually to each separate value falling within the range, unless otherwise indicated herein and each separate value is incorporated into the specification as if it were individually
5 recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g., “such as”) provided herein, is intended merely to better illuminate embodiments and does not pose a limitation on the scope of the disclosure unless otherwise claimed. No language in the specification should be construed as indicating any non-
10 claimed element as essential to the practice of the disclosure.

[0214] Disjunctive language such as the phrase “at least one of X, Y, or Z,” unless specifically stated otherwise, is intended to be understood within the context as used in general to present that an item, term, etc., may be either X, Y, or Z, or any combination thereof (e.g., X, Y, and/or Z). Thus, such disjunctive language is not generally intended to, and should not, imply
15 that certain embodiments require at least one of X, at least one of Y, or at least one of Z to each be present.

[0215] The term “computer system” as used herein refers to any type of interconnected electronic devices, computer devices, or components thereof. Additionally, the term “computer system” or “system” may refer to various components of a computer that are communicatively
20 coupled with one another. Furthermore, the term “computer system” or “system” may refer to multiple computer devices or multiple computing systems that are communicatively coupled with one another and configured to share computing or networking resources.

[0216] Throughout the disclosure, the terminology of “removing a pixel” and variations thereof have been used. It is to be understood that the terminology of “removing a pixel” may
25 mean that the pixel is removed from further consideration in processing an image, may mean that the pixel is set to a certain value, may mean that the a value or values related to the pixel (such as a measurement of differences between pixel values or indication of differences between pixel values) is set to a certain value, or some combination thereof.

[0217] Throughout the disclosure, the terminology of “corresponding pixels” between two or
30 more images and variations thereof have been used. It is to be understood that the terminology of

“corresponding pixels” may refer to a first pixel within a first image at a particular position and a second pixel within a second image in the particular position. For example, a first pixel within a first row and first column of the first image may be a corresponding pixel to a second pixel within a first row and a first column of the second image. In other instances, the terminology of “corresponding pixels” may refer to a first pixel within a first image that captures a certain portion of an area captured by the first image and a second pixel within a second image that captures the certain portion of the area captured by the second image. For example, a first pixel that captures a portion of area surrounding a camera that captured may be a corresponding pixel to a second pixel that captures the same portion of the area surround the camera.

[0218] Preferred embodiments of this disclosure are described herein, including the best mode known for carrying out the disclosure. Variations of those preferred embodiments may become apparent to those of ordinary skill in the art upon reading the foregoing description. Those of ordinary skill should be able to employ such variations as appropriate and the disclosure may be practiced otherwise than as specifically described herein. Accordingly, this disclosure includes all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law. Moreover, any combination of the above-described elements in all possible variations thereof is encompassed by the disclosure unless otherwise indicated herein.

[0219] All references, including publications, patent applications, and patents, cited herein are hereby incorporated by reference to the same extent as if each reference were individually and specifically indicated to be incorporated by reference and were set forth in its entirety herein.

[0220] In the foregoing specification, aspects of the disclosure are described with reference to specific embodiments thereof, but those skilled in the art will recognize that the disclosure is not limited thereto. Various features and aspects of the above-described disclosure may be used individually or jointly. Further, embodiments can be utilized in any number of environments and applications beyond those described herein without departing from the broader spirit and scope of the specification. The specification and drawings are, accordingly, to be regarded as illustrative rather than restrictive.

[0221] As described above, one aspect of the present technology is the gathering and use of data (e.g., images of people) to perform facial recognition. The present disclosure contemplates

that in some instances, this gathered data may include personally identifiable information (PII) data that uniquely identifies or can be used to contact or locate a specific person. Such personal information data can include facial and/or non-facial characteristics of a person's body, demographic data, location-based data (e.g., GPS coordinates), telephone numbers, email
5 addresses, Twitter ID's, home addresses, or any other identifying or personal information.

[0222] The present disclosure recognizes that the use of such personal information data, in the present technology, can be used to the benefit of users. For example, the personal information data can be used to identify a person as being a contact (or not known contact) of a user of a user device.

10 [0223] The present disclosure contemplates that the entities responsible for the collection, analysis, disclosure, transfer, storage, or other use of such personal information data will comply with well-established privacy policies and/or privacy practices. In particular, such entities should implement and consistently use privacy policies and practices that are generally recognized as meeting or exceeding industry or governmental requirements for maintaining personal
15 information data private and secure. Such policies should be easily accessible by users, and should be updated as the collection and/or use of data changes. Personal information from users should be collected for legitimate and reasonable uses of the entity and not shared or sold outside of those legitimate uses. Further, such collection/sharing should occur after receiving the informed consent of the users. Additionally, such entities should consider taking any needed
20 steps for safeguarding and securing access to such personal information data and ensuring that others with access to the personal information data adhere to their privacy policies and procedures. Further, such entities can subject themselves to evaluation by third parties to certify their adherence to widely accepted privacy policies and practices. In addition, policies and practices should be adapted for the particular types of personal information data being collected
25 and/or accessed and adapted to applicable laws and standards, including jurisdiction-specific considerations. For instance, in the US, collection of or access to certain health data may be governed by federal and/or state laws, such as the Health Insurance Portability and Accountability Act (HIPAA); whereas health data in other countries may be subject to other regulations and policies and should be handled accordingly. Hence different privacy practices
30 should be maintained for different personal data types in each country.

[0224] Despite the foregoing, the present disclosure also contemplates embodiments in which users selectively block the use of, or access to, personal information data. That is, the present disclosure contemplates that hardware and/or software elements can be provided to prevent or block access to such personal information data. For example, in the case of services
5 related to performing facial recognition, the present technology can be configured to allow users to select to "opt in" or "opt out" of participation in the collection of personal information data during registration for services or anytime thereafter. In addition to providing "opt in" and "opt out" options, the present disclosure contemplates providing notifications relating to the access or use of personal information. For instance, a user may be notified upon downloading an app that
10 their personal information data will be accessed and then reminded again just before personal information data is accessed by the app.

[0225] Moreover, it is the intent of the present disclosure that personal information data should be managed and handled in a way to minimize risks of unintentional or unauthorized access or use. Risk can be minimized by limiting the collection of data and deleting data once it
15 is no longer needed. In addition, and when applicable, including in certain health related applications, data de-identification can be used to protect a user's privacy. De-identification may be facilitated, when appropriate, by removing specific identifiers (e.g., date of birth, etc.), controlling the amount or specificity of data stored (e.g., collecting location data a city level rather than at an address level), controlling how data is stored (e.g., aggregating data across
20 users), and/or other methods.

[0226] Therefore, although the present disclosure broadly covers use of personal information data to implement one or more various disclosed embodiments, the present disclosure also contemplates that the various embodiments can also be implemented without the need for accessing such personal information data. That is, the various embodiments of the present
25 technology are not rendered inoperable due to the lack of all or a portion of such personal information data.

WHAT IS CLAIMED IS:

- 1 1. A method, comprising:
2 capturing a representative image from a video;
3 retrieving a canonical image that represents one or more frames of the video
4 before the representative image;
5 determining a difference between the canonical image and the representative
6 image, the difference identifying a set of pixels of the representative image that are different
7 from corresponding pixels in the canonical image;
8 modifying the set of pixels based at least in part on previous frames of the video
9 before the representative image to form a modified set of pixels, the modified set of pixels being
10 smaller than the set of pixels;
11 executing a classifier using the modified set of pixels, the classifier providing a
12 prediction of an identification of an object represented by a subset of the modified set of pixels;
13 and
14 outputting the prediction of the identification of the object.
- 1 2. The method of claim 1, wherein determining the difference between the
2 canonical image and the representative image includes determining distances within a color
3 space using Delta E for related pixels within the representative image and the canonical image,
4 and wherein the set of pixels that are identified as being different is based at least in part on the
5 determined distances within the color space.
- 1 3. The method of claim 2, wherein the color space is a LAB color space.
- 1 4. The method of claim 1, further comprising:
2 averaging amounts of change in subgroups of the set of pixels;
3 identifying a portion of the subgroups having averaged amounts of change that are
4 below a threshold amount of change, wherein pixels within the portion of the subgroups
5 comprise low density pixels; and
6 removing the low density pixels from the set of pixels.
- 1 5. The method of claim 1, further comprising:

2 identifying a subset of disjointed pixels from the modified set of pixels;
3 generating a bounding box around the subset of disjointed pixels; and
4 adding additional pixels to the modified set of pixels to fill the bounding box,
5 wherein the modified set of pixels with the additional pixels is used in the executing of the
6 classifier.

1 6. The method of claim 5, wherein the subset of disjointed pixels is a first
2 subset of disjointed pixels, and wherein the method further comprises:
3 identifying a second subset of disjointed pixels from the modified set of pixels;
4 generating a second bounding box around the second subset of disjointed pixels;
5 determining that the second bounding box is smaller than a first threshold
6 percentage of the representative image or larger than a second threshold percentage of the
7 representative image, the second threshold percentage being larger than the first threshold
8 percentage; and
9 removing pixels corresponding to the second bounding box from the set of pixels,
10 wherein the modified set of pixels with the pixels removed is used in the executing of the
11 classifier.

1 7. The method of claim 1, wherein modifying the set of pixels includes:
2 applying temporal averaging to the set of pixels;
3 identifying high frequency pixels from the set of pixels based at least in part on
4 the temporal averaging indicating that the high frequency pixels have been unstable within the
5 previous frames; and
6 removing the high frequency pixels from the set of pixels based at least in part on
7 the temporal averaging indicating that the high frequency pixels have been unstable within the
8 previous frames.

1 8. The method of claim 7, wherein applying the temporal averaging includes
2 averaging amounts of change for each pixel within the set of pixels within the previous frames,
3 and wherein the averaged amounts of change for the high frequency pixels are below a temporal
4 averaging threshold.

1 9. One or more computer-readable media having instructions stored thereon,
2 wherein the instructions, when executed by a system, cause the system to:
3 capture a representative image of an area from a video;
4 retrieve a canonical image of the area that represents one or more frames of the
5 video before the representative image;
6 determine a difference between the canonical image and the representative image,
7 the difference identifying a set of pixels of the representative image that are different from
8 corresponding pixels in the canonical image;
9 modify the set of pixels based at least in part on previous frames of the video
10 before the representative image to form a modified set of pixels, the modified set of pixels being
11 smaller than the set of pixels;
12 execute a classifier using the modified set of pixels, the classifier providing a
13 prediction of an identification of an object represented by a subset of the modified set of pixels;
14 and
15 output the prediction of the identification of the object.

1 10. The one or more computer-readable media of claim 9, wherein the
2 prediction of the object includes a prediction that the object is a package, and wherein the
3 instructions, when executed by the system, further cause the system to:
4 determine that notification settings associated with a camera that captures the
5 video indicate that a notification is to be provided when the prediction is that the object is a
6 package; and
7 cause the notification to be provided based at least in part on the prediction that
8 the object is a package.

1 11. The one or more computer-readable media of claim 9, wherein to
2 determine the difference between the canonical image and the representative image includes to
3 determine distances within a color space using Delta E for related pixels within the canonical
4 image and the representative image, and wherein the set of pixels that are identified as being
5 different is based at least in part on the determined distances within the color space.

1 12. The one or more computer-readable media of claim 11, wherein the set of
2 pixels that are identified as being different are identified based at least in part on distances within
3 the color space corresponding to the set of pixels being greater than a threshold distance.

1 13. The one or more computer-readable media of claim 9, wherein the
2 instructions, when executed by the system, further cause the system to:
3 average an amount of change in subgroups of the set of pixels;
4 identify a portion of the subgroups having averaged amounts of change that are
5 below a threshold amount of change, wherein pixels within the portion of the subgroups
6 comprise low density pixels; and
7 remove the low density pixels from the set of pixels.

1 14. The one or more computer-readable media of claim 9, wherein the
2 instructions, when executed by the system, further cause the system to:
3 identify a first subset of disjointed pixels from the modified set of pixels;
4 generate a bounding box around the first subset of disjointed pixels; and
5 adding additional pixels to the modified set of pixels to fill the bounding box,
6 wherein the modified set of pixels with the additional pixels is used in the execution of the
7 classifier.

1 15. The one or more computer-readable media of claim 14, wherein to modify
2 the set of pixels includes to:
3 apply temporal averaging to the set of pixels;
4 identify high frequency pixels from the set of pixels based at least in part on the
5 temporal averaging indicating that the high frequency pixels have been unstable within the
6 previous frames; and
7 remove the high frequency pixels based at least in part on the temporal averaging
8 indicating that the high frequency pixels have been unstable within the previous frames.

1 16. The one or more computer-readable media of claim 15, wherein to apply
2 the temporal averaging includes to average amounts of change for each pixel within the set of

pixels within the previous frames, and wherein the averaged amounts of change for the high frequency pixels are below a temporal averaging threshold.

17. A system, comprising:
memory to store images from video received from a camera; and
one or more processors coupled to the memory, the one or more processors to:
capture, from stored images, a representative image;
retrieve, from the stored images, a canonical image that represents one or more frames of the video before the representative image;
determine a difference between the canonical image and the representative image, the difference identifying a set of pixels of the representative image that are different from corresponding pixels in the canonical image;
modify the set of pixels based at least in part on previous frames of the video before the representative image to form a modified set of pixels, the modified set of pixels being smaller than the set of pixels;
execute a classifier using the modified set of pixels, the classifier providing a prediction of an identification of an object represented by a subset of the modified set of pixels; and
output the prediction of the identification of the object.

18. The system of claim 17, wherein to determine the difference between the canonical image and the representative image includes to determine distances between the canonical image and the representative image in a LAB color space, the difference determined based at least in part on the distances.

19. The system of claim 17, wherein to modify the set of pixels includes to:
apply temporal averaging to the set of pixels;
identify high frequency pixels from the set of pixels based at least in part on the temporal averaging indicating that the high frequency pixels have been unstable within the one or more frames; and
remove the high frequency pixels based at least in part on the temporal averaging indicating that the high frequency pixels have been unstable within the one or more frames.

1 20. The system of claim 19, wherein to apply the temporal averaging includes
2 to average amounts of change for each pixel within the set of pixels within the one or more
3 frames, and wherein the averaged amounts of change for the high frequency pixels are below a
4 temporal averaging threshold.

1 21. A method, comprising:
2 identifying, by a computing device, one or more package candidates from an
3 image;
4 determining, by the computing device, an analysis frame from the image based at
5 least in part on the one or more package candidates;
6 executing, by the computing device, a shared classifier on the analysis frame, the
7 shared classifier to predict identification of packages and objects within the analysis frame; and
8 outputting, by the computing device, a prediction of the identification of the
9 packages and the objects based at least in part on the execution of the shared classifier.

1 22. The method of claim 21, wherein identifying the one or more package
2 candidates from the image includes determining a difference between the image and a canonical
3 image, wherein the image is from a video, wherein the canonical image represents one or more
4 frames of the video before the image, and wherein determining the difference includes
5 identifying a set of pixels of the image that are different from corresponding pixels in the
6 canonical image.

1 23. The method of claim 22, wherein identifying the one or more package
2 candidates includes modifying the set of pixels based at least in part on previous frames of the
3 video before the image to form a modified set of pixels, the modified set of pixels being smaller
4 than the set of pixels.

1 24. The method of claim 22, wherein determining the difference between the
2 image and the canonical image includes determining distances within a color space using Delta E
3 for related pixels within the image and the canonical image, and wherein the set of pixels that are
4 identified as being different is based at least in part on the determined distances within the color
5 space.

1 25. The method of claim 24, wherein the color space is a LAB color space.

1 26. The method of claim 22, further comprising:

2 averaging amounts of change in subgroups of the set of pixels;

3 identifying a portion of the subgroups having averaged amounts of change that are

4 below a threshold amount of change, wherein pixels within the portion of the subgroups

5 comprise low density pixels; and

6 removing the low density pixels from the set of pixels.

1 27. The method of claim 23, further comprising:

2 identifying a subset of disjointed pixels from the modified set of pixels;

3 generating a bounding box around the subset of disjointed pixels; and

4 adding additional pixels to the modified set of pixels to fill the bounding box,

5 wherein the modified set of pixels with the additional pixels is used in the executing of the

6 shared classifier.

1 28. The method of claim 23, further comprising:

2 applying temporal averaging to the set of pixels;

3 identifying high frequency pixels from the set of pixels based at least in part on

4 the temporal averaging indicating that the high frequency pixels have been unstable within the

5 previous frames; and

6 removing the high frequency pixels from the set of pixels based at least in part on

7 the temporal averaging indicating that the high frequency pixels have been unstable within the

8 previous frames.

1 29. The method of claim 21, further comprising:

2 identifying one or more object candidates from the image, wherein determining

3 the analysis frame is further based at least in part on the one or more object candidates.

1 30. The method of claim 29, wherein identifying the one or more object

2 candidates comprises identifying the one or more object candidates based at least in part on

3 motion identified from the image.

1 31. The method of claim 21, wherein the image is a current image, wherein
2 the prediction includes a first package predicted within the current image, wherein the analysis
3 frame is a first analysis frame, and wherein the method further comprises:

4 identifying one or more package candidates from a first image, the first image
5 captured prior to the current image;

6 determining a second analysis frame from the first image based on the one or
7 more package candidates from the first image;

8 executing a shared classifier on the second analysis frame; and

9 comparing a first package predicted from the current image with a second
10 package predicted from the first image to determine whether the first package and the second
11 package are a same package.

1 32. The method of claim 31, wherein comparing the first package and the
2 second package comprises comparing a location of the first package and a location of the second
3 package.

1 33. One or more computer-readable media having instructions stored thereon,
2 wherein the instructions, when executed by a system, cause the system to:

3 identify one or more package candidates from an image;

4 determine an analysis frame from the image based at least in part on the one or
5 more package candidates;

6 execute a shared classifier on the analysis frame, the shared classifier to predict
7 identification of packages and objects within the analysis frame; and

8 output a prediction of the identification of the packages and the objects based at
9 least in part on the execution of the shared classifier.

1 34. The one or more computer-readable media of claim 33, wherein to identify
2 the one or more package candidates from the image includes to determine a difference between
3 the image and a canonical image, wherein the image is from a video, wherein the canonical
4 image represents one or more frames of the video before the image, and wherein to determine the
5 difference includes to identify a set of pixels of the image that are different from corresponding
6 pixels in the canonical image.

1 35. The one or more computer-readable media of claim 34, wherein to identify
2 the one or more package candidates includes to modify the set of pixels based at least in part on
3 previous frames of the video before the image to form a modified set of pixels, the modified set
4 of pixels being smaller than the set of pixels.

1 36. The one or more computer-readable media of claim 34, wherein to
2 determine the difference between the image and the canonical image includes to determine
3 distances within a color space using Delta E for related pixels within the image and the canonical
4 image, and wherein the set of pixels that are identified as being different is based at least in part
5 on the determined distances within the color space.

1 37. The one or more computer-readable media of claim 36, wherein the color
2 space is a LAB color space.

1 38. The one or more computer-readable media of claim 34, wherein the
2 instructions, when executed by the system, further cause the system to:
3 average amounts of change in subgroups of the set of pixels;
4 identify a portion of the subgroups having averaged amounts of change that are
5 below a threshold amount of change, wherein pixels within the portion of the subgroups
6 comprise low density pixels; and
7 remove the low density pixels from the set of pixels.

1 39. The one or more computer-readable media of claim 35, wherein the
2 instructions, when executed by the system, further cause the system to:
3 identify a subset of disjointed pixels from the modified set of pixels;
4 generate a bounding box around the subset of disjointed pixels; and
5 add additional pixels to the modified set of pixels to fill the bounding box,
6 wherein the modified set of pixels with the additional pixels is used in the executing of the
7 shared classifier.

1 40. The one or more computer-readable media of claim 35, wherein the
2 instructions, when executed by the system, further cause the system to:
3 apply temporal averaging to the set of pixels;

4 identify high frequency pixels from the set of pixels based at least in part on the
5 temporal averaging indicating that the high frequency pixels have been unstable within the
6 previous frames; and

7 remove the high frequency pixels from the set of pixels based at least in part on
8 the temporal averaging indicating that the high frequency pixels have been unstable within the
9 previous frames.

1 41. The one or more computer-readable media of claim 33, wherein the
2 instructions, when executed by the system, further cause the system to:

3 identify one or more object candidates from the image, wherein to determine the
4 analysis frame is further based at least in part on the one or more object candidates.

1 42. The one or more computer-readable media of claim 41, wherein to identify
2 the one or more object candidates comprises to identify the one or more object candidates based
3 at least in part on motion identified from the image.

1 43. The one or more computer-readable media of claim 33, wherein the image
2 is a current image, wherein the prediction includes a first package predicted within the current
3 image, wherein the analysis frame is a first analysis frame, and wherein the instructions, when
4 executed by the system, further cause the system to:

5 identify one or more package candidates from a first image, the first image
6 captured prior to the current image;

7 determine a second analysis frame from the first image based on the one or more
8 package candidates from the first image;

9 execute a shared classifier on the second analysis frame; and

10 compare a first package predicted from the current image with a second package
11 predicted from the first image to determine whether the first package and the second package are
12 a same package.

1 44. The one or more computer-readable media of claim 43, wherein to
2 compare the first package and the second package comprises to compare a location of the first
3 package and a location of the second package.

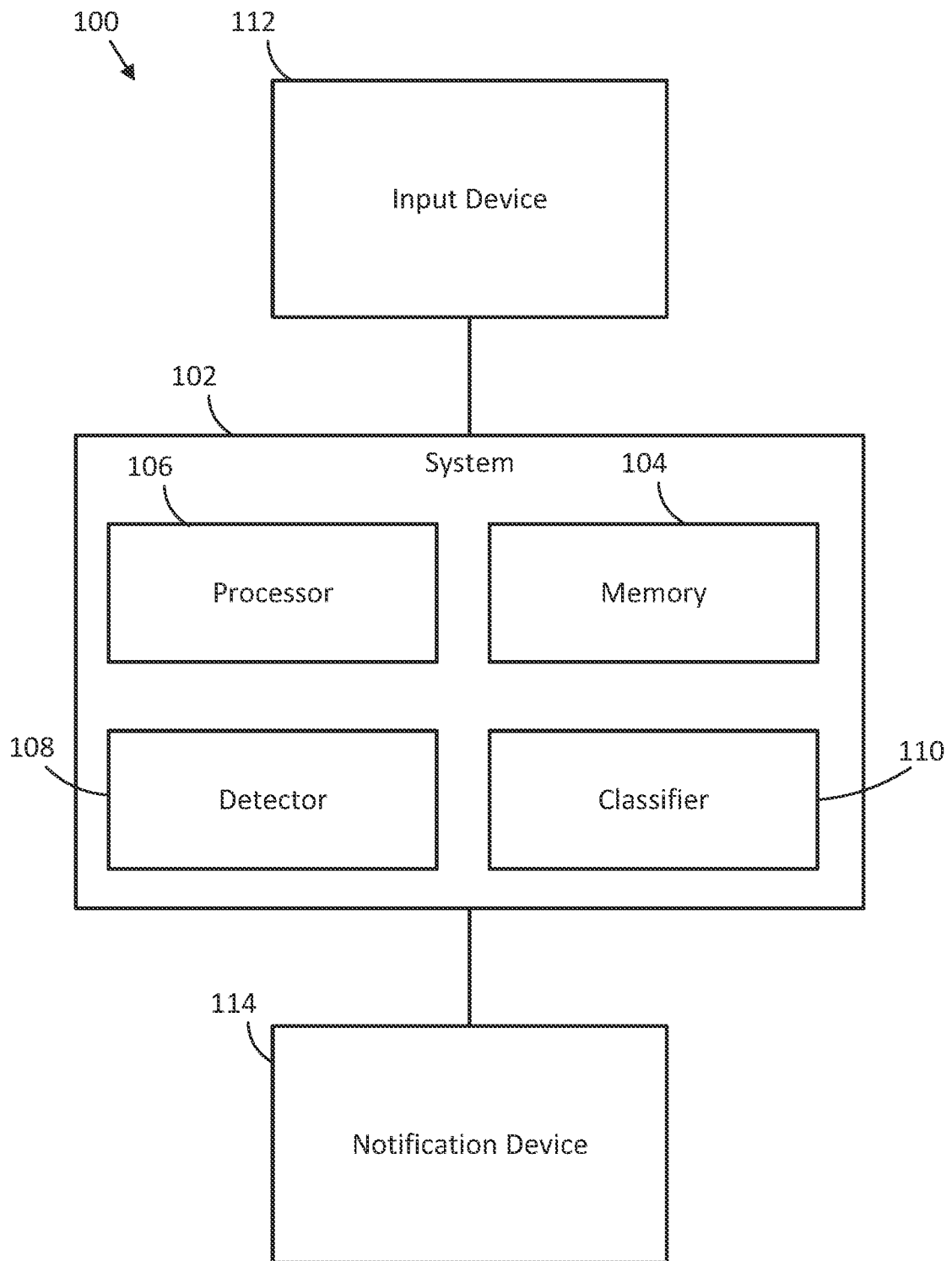


FIG. 1

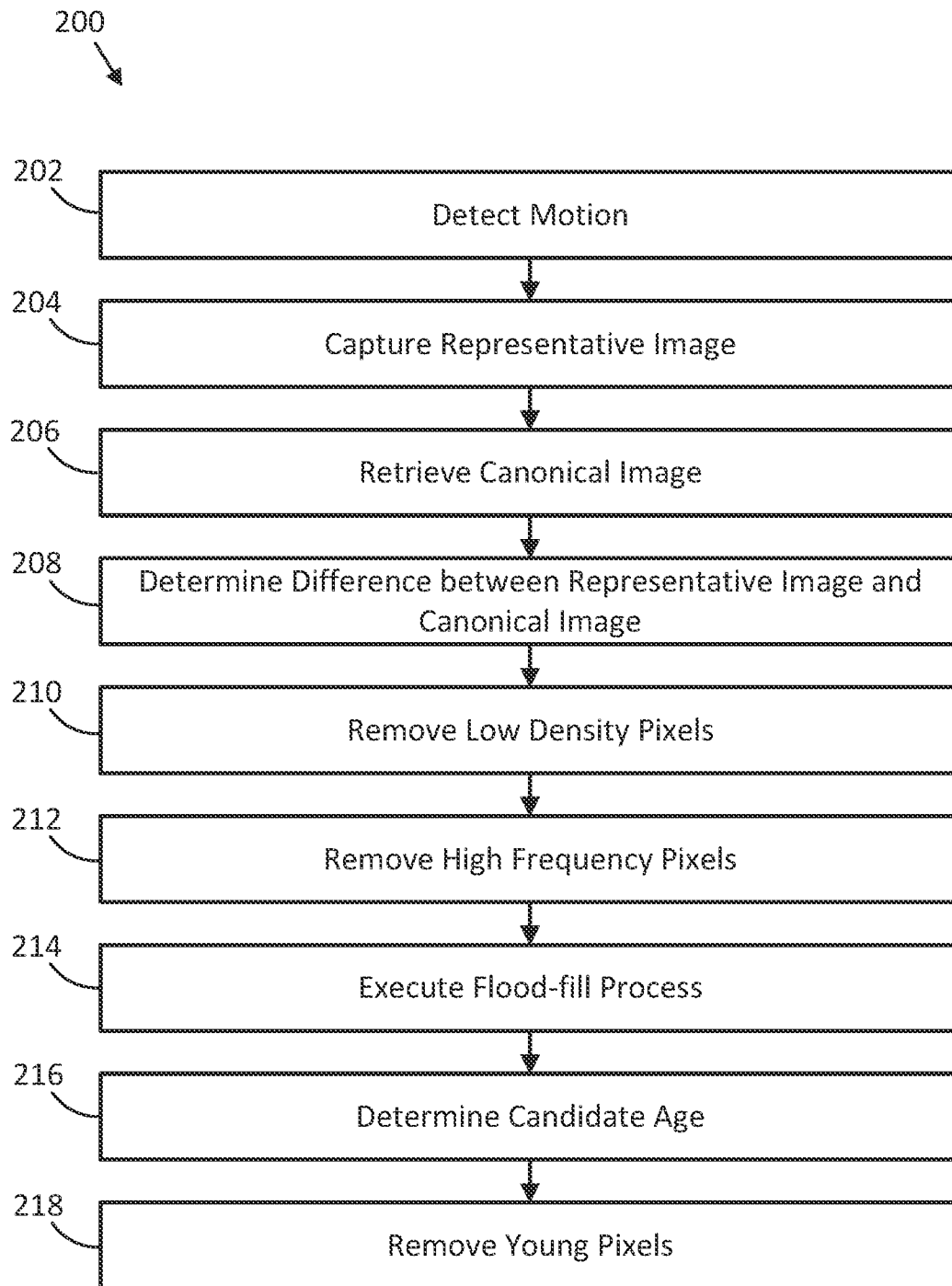


FIG. 2



FIG. 3

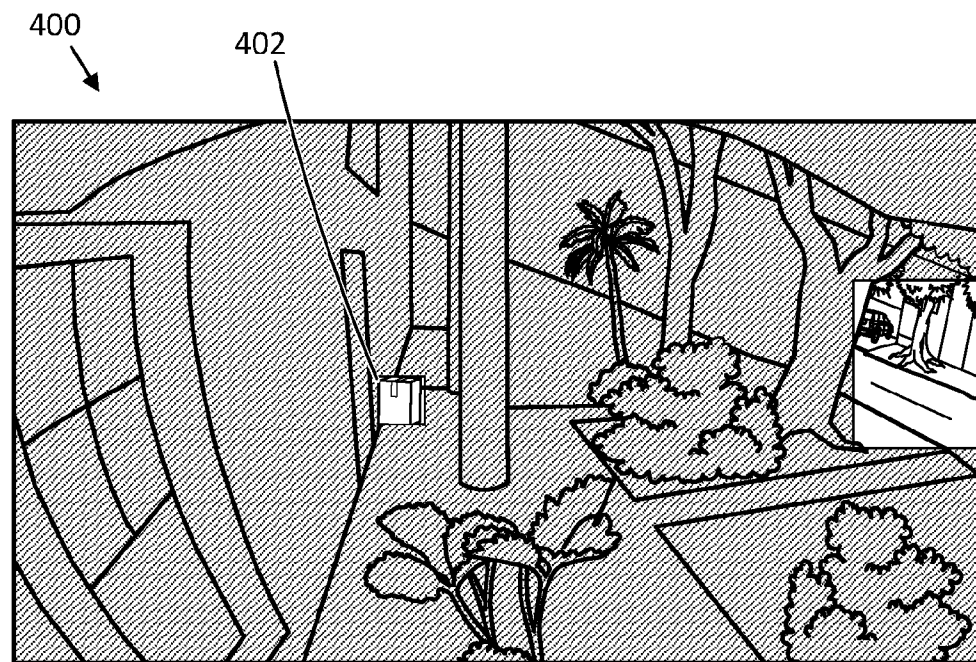


FIG. 4

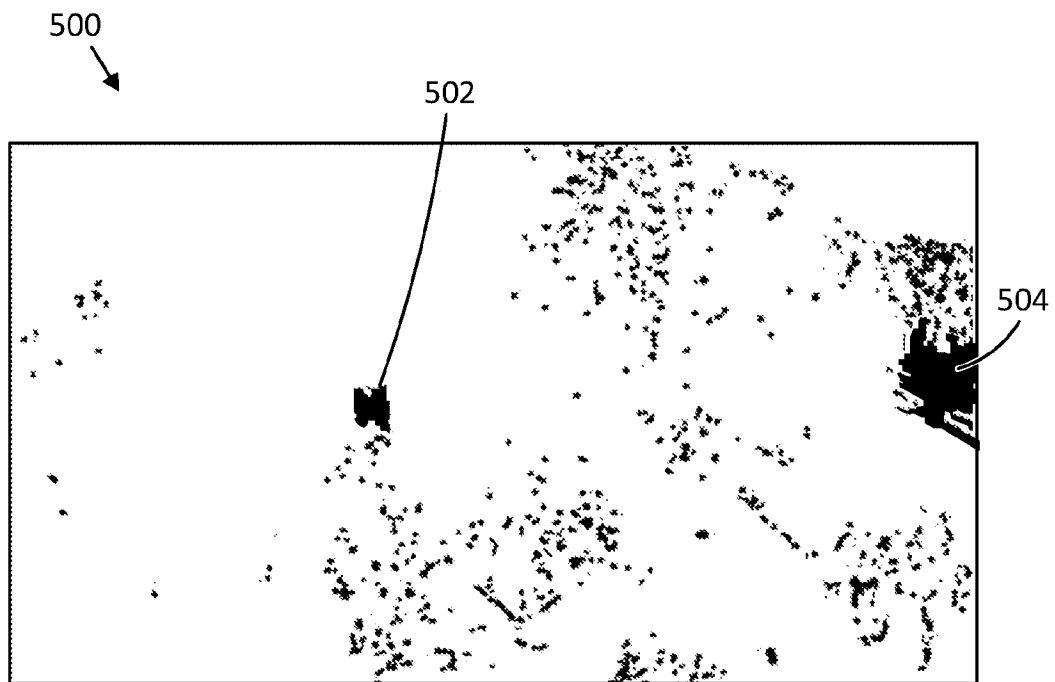


FIG. 5

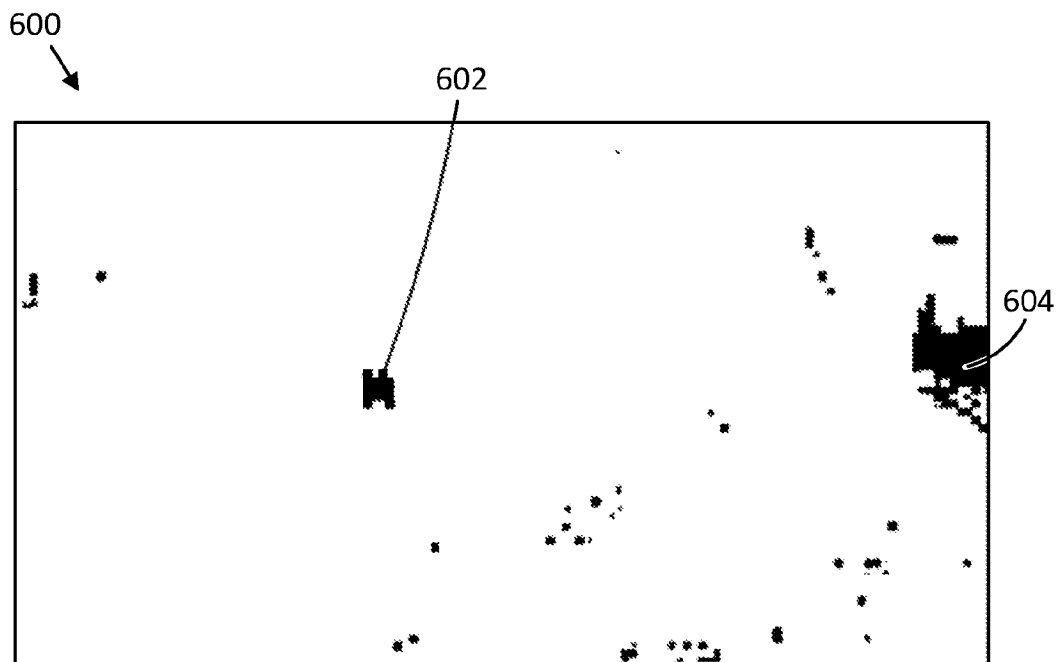


FIG. 6

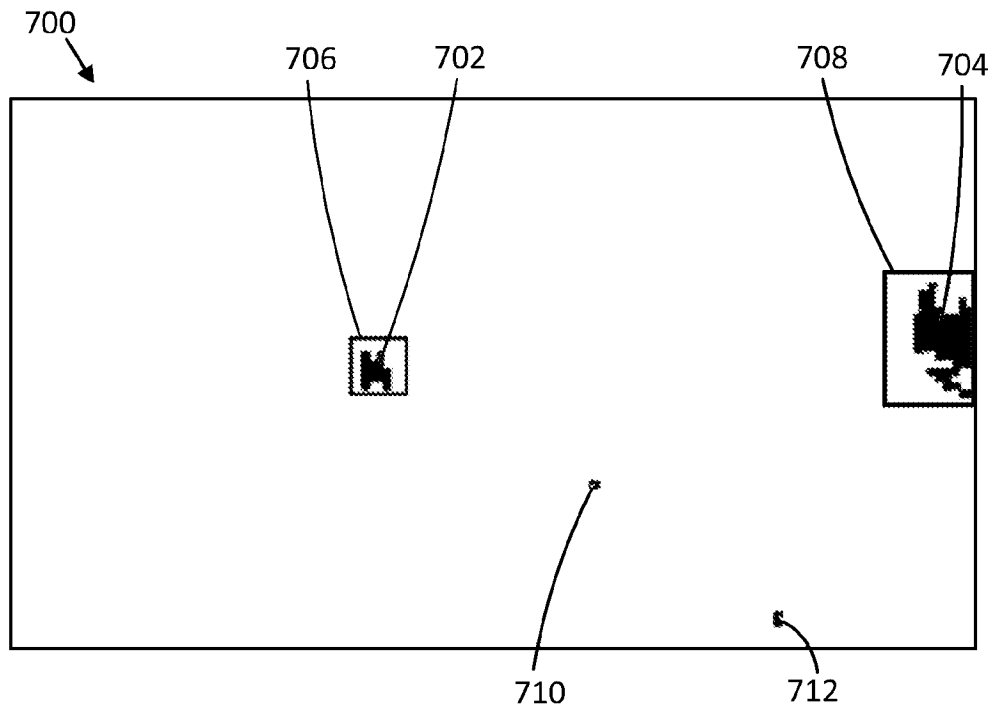


FIG. 7

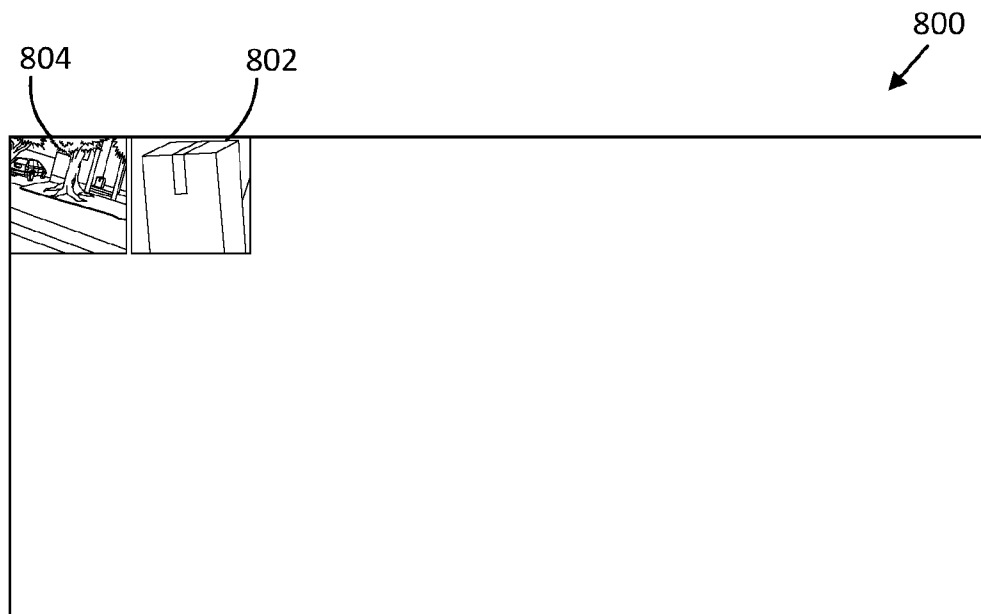


FIG. 8

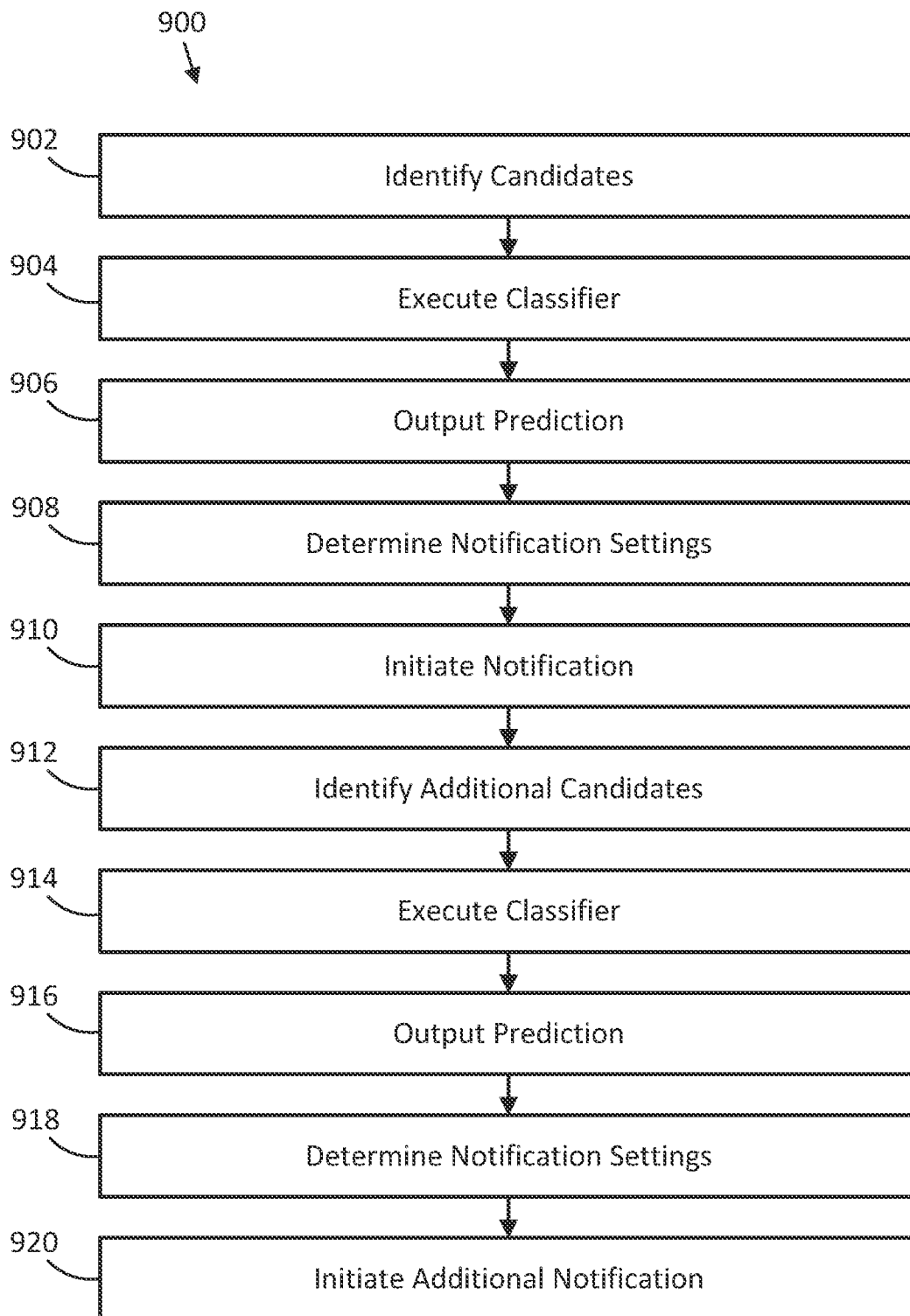


FIG. 9

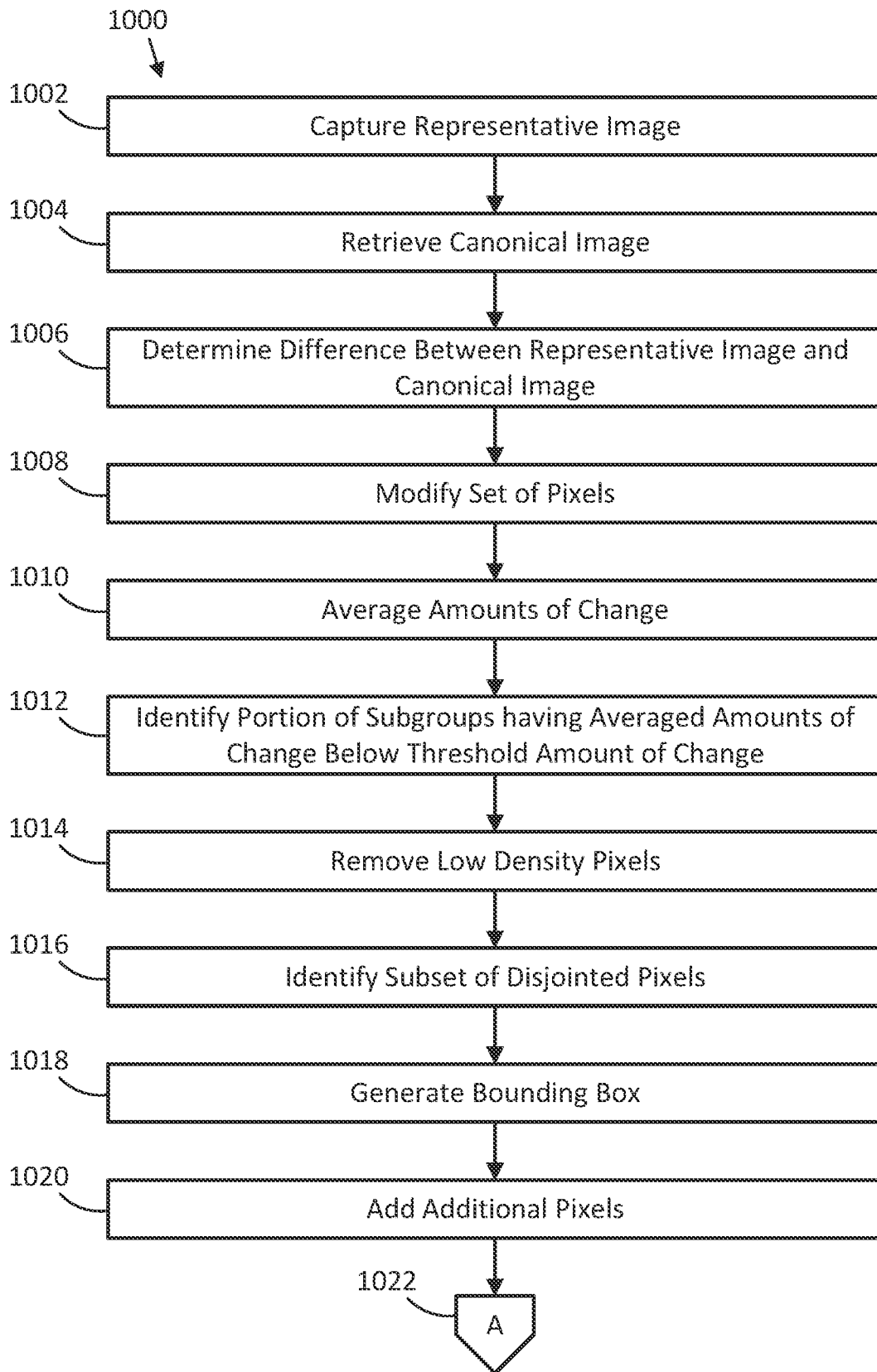


FIG. 10

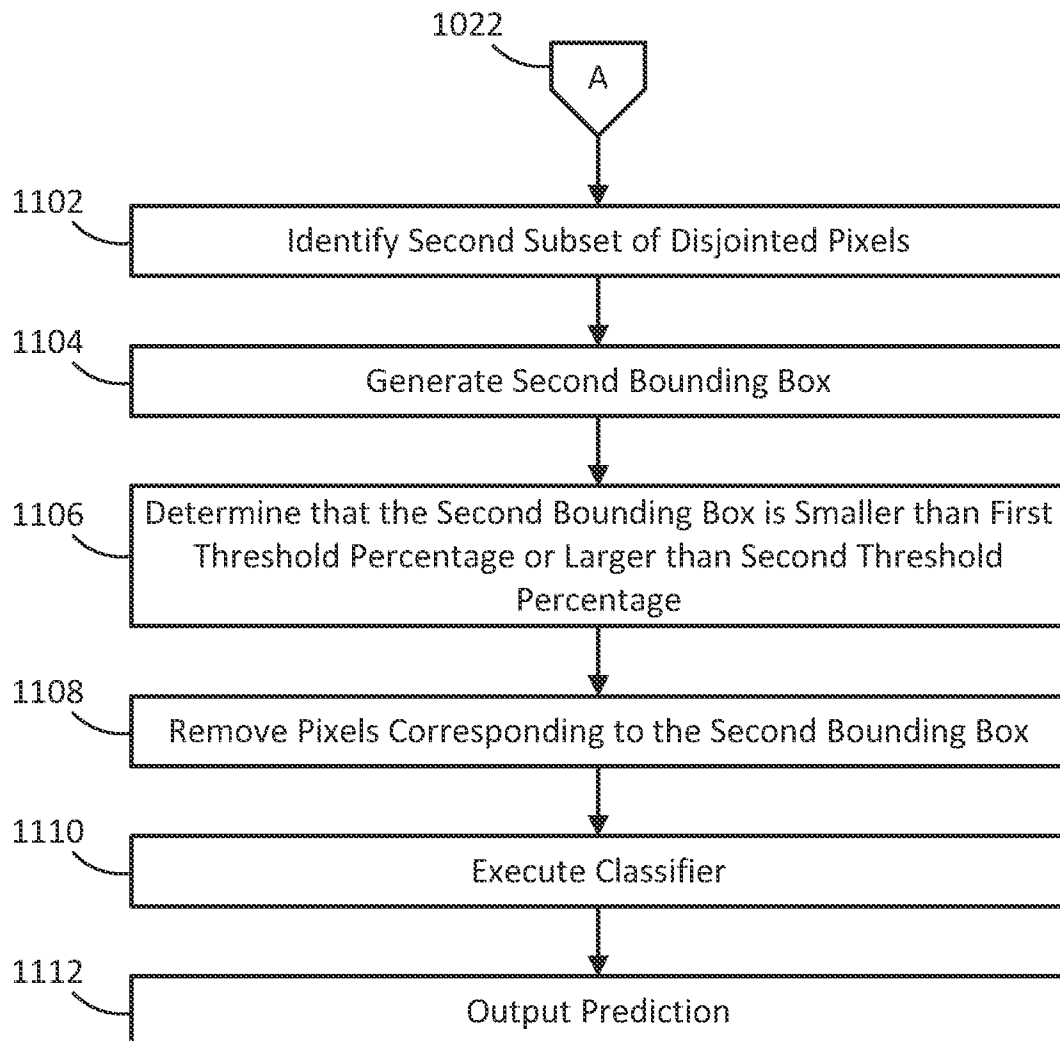


FIG. 11

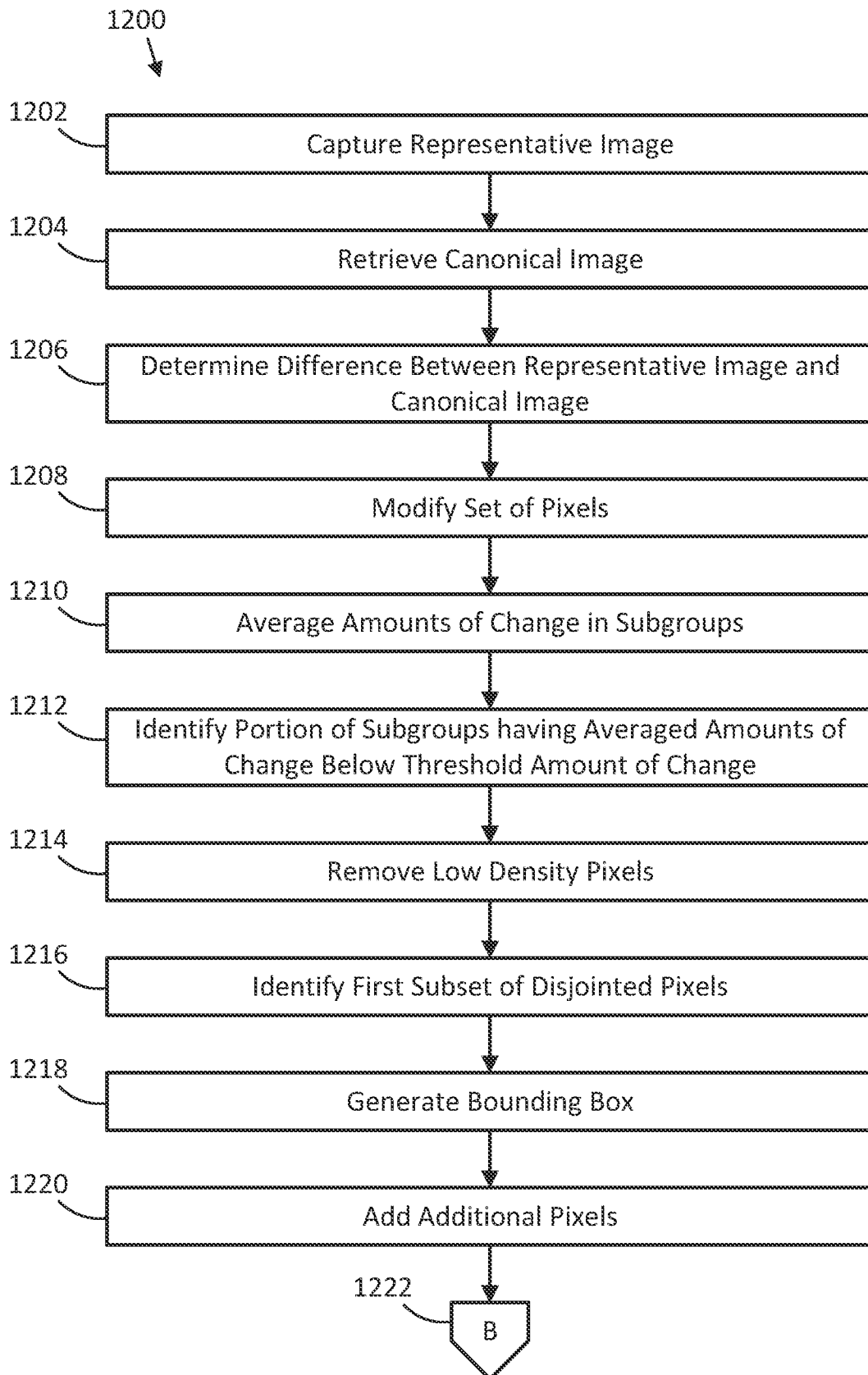


FIG. 12

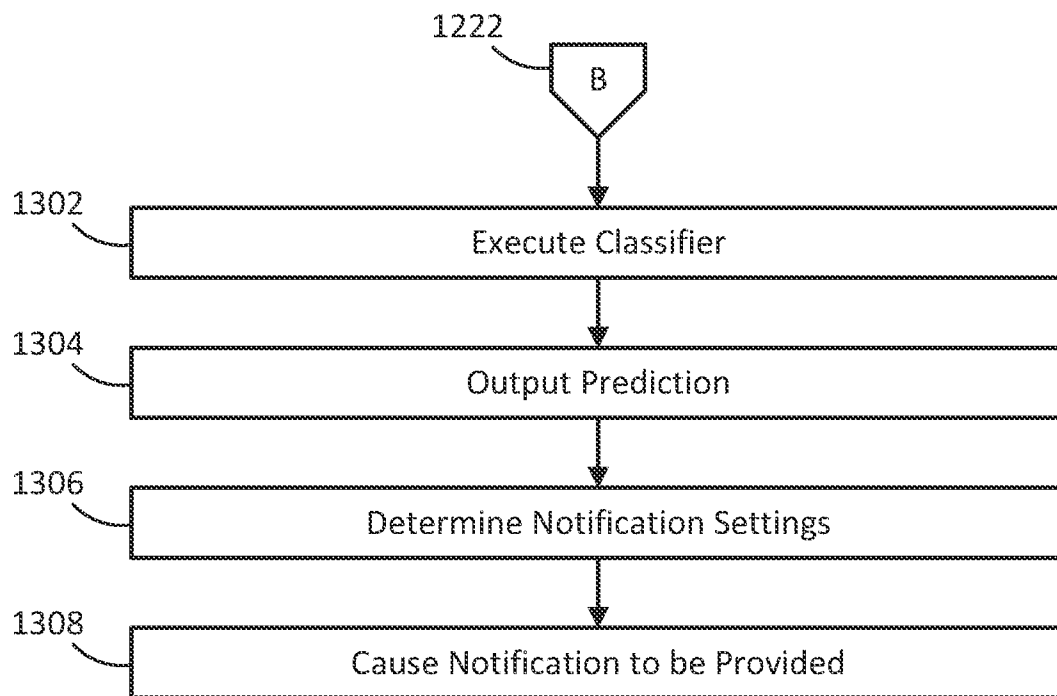


FIG. 13

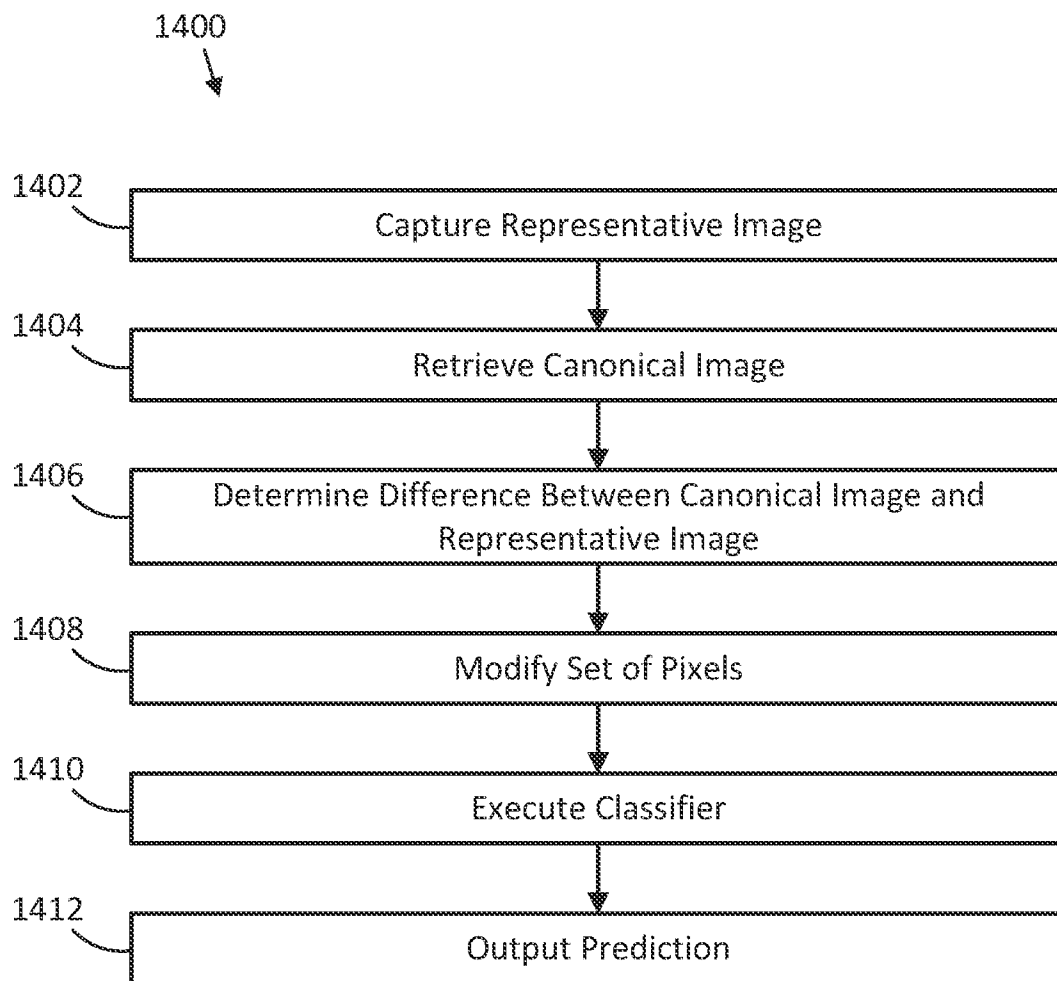


FIG. 14

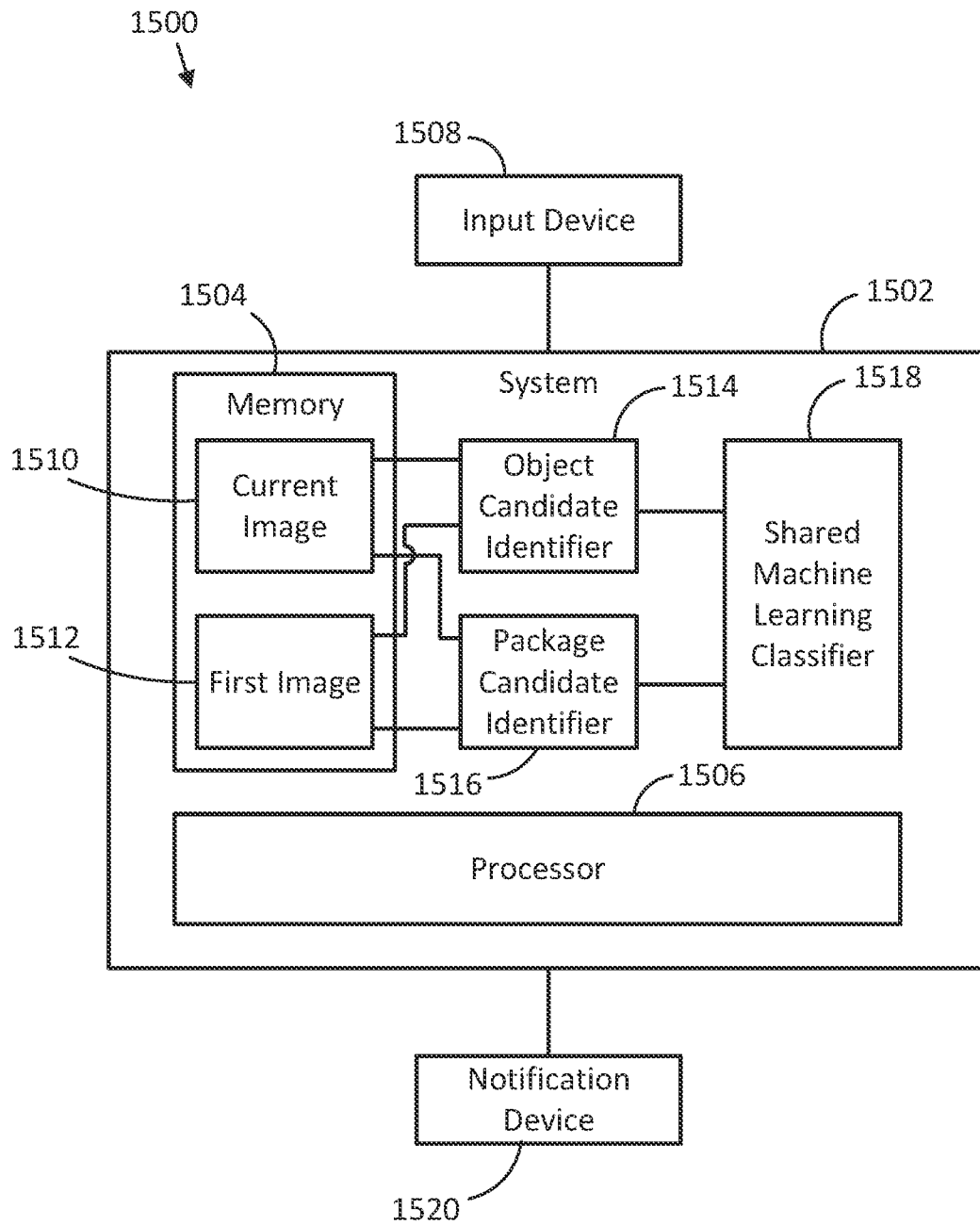


FIG. 15

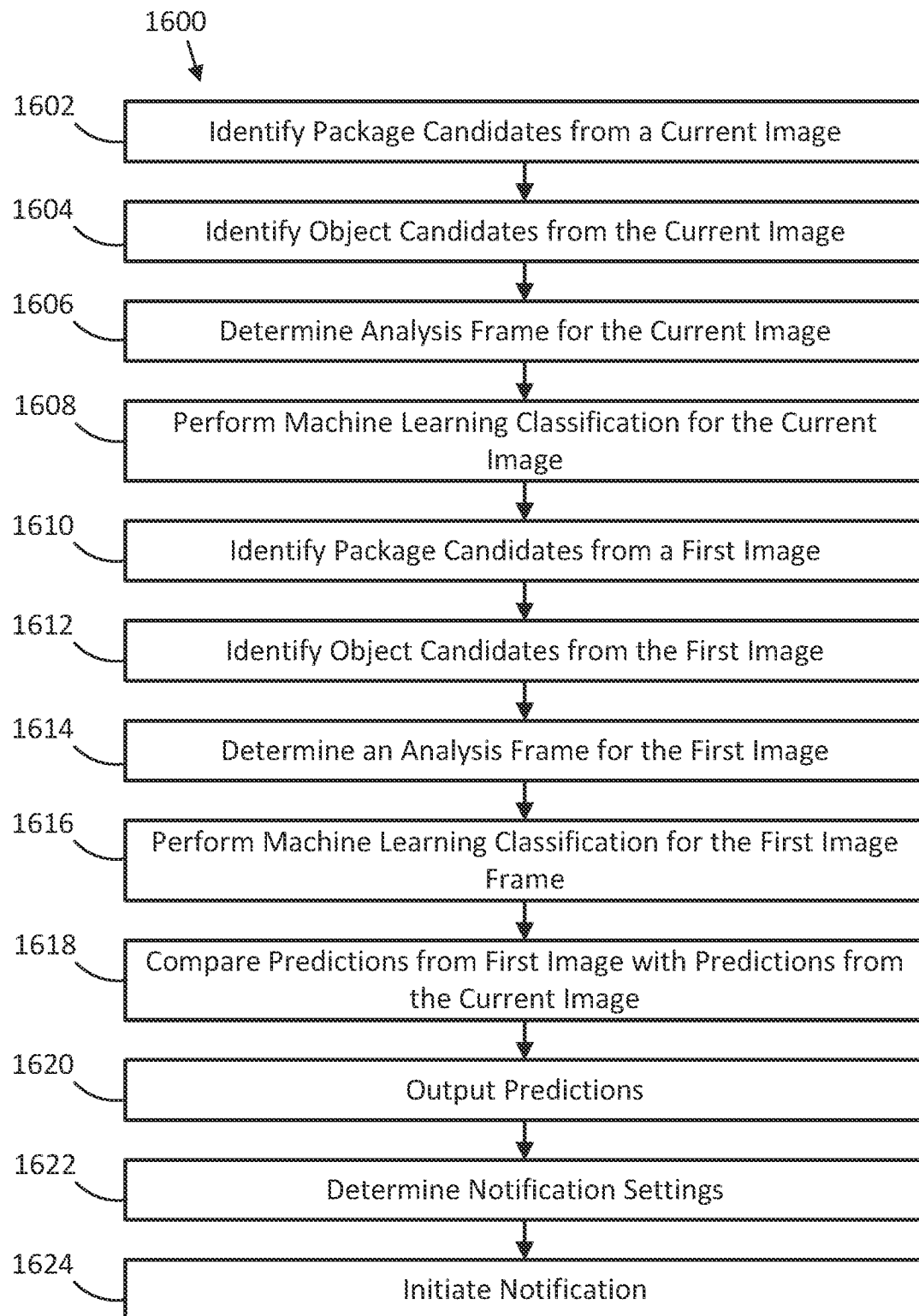


FIG. 16

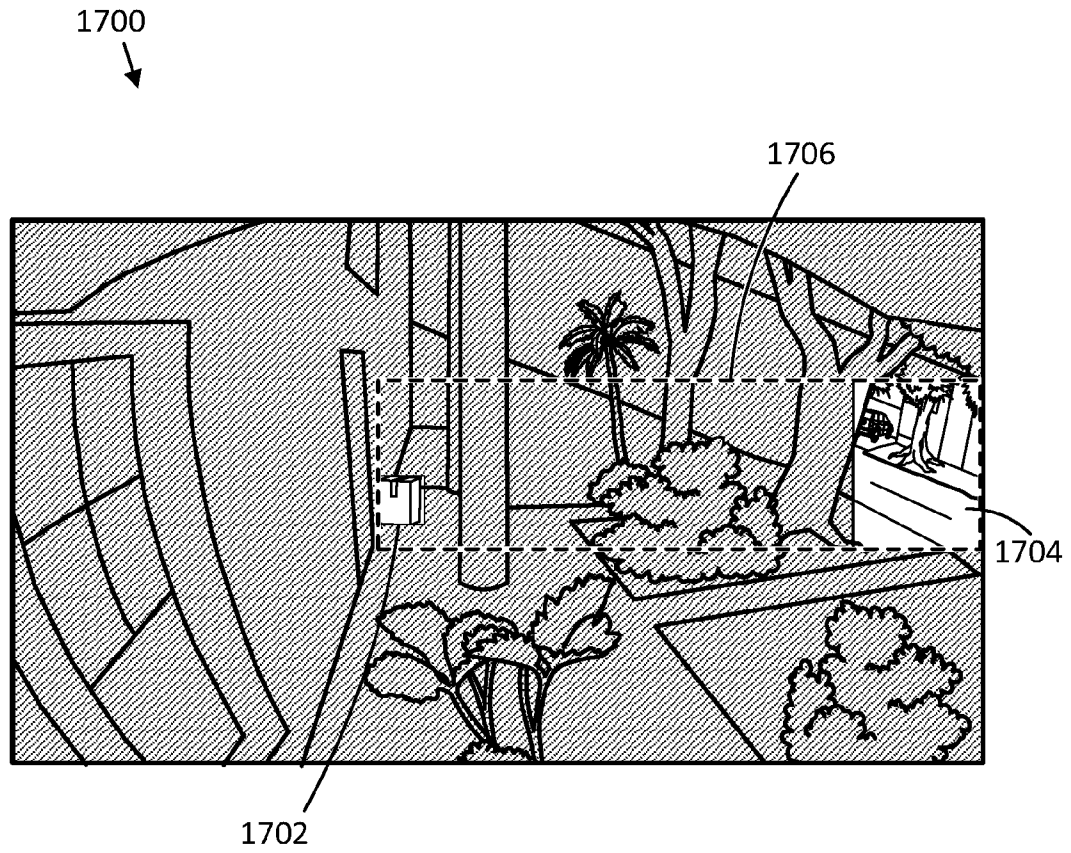


FIG. 17

International application No
PCT/US2022/029490

A. CLASSIFICATION OF SUBJECT MATTER				
INV.	G06V20/52	G06V20/40	G06V10/62	G06V10/34
				G06V10/75
ADD.				

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

Minimum documentation searched (classification system followed by classification symbols)

G06V

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

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

EPO-Internal, WPI Data

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
-----------	--	-----------------------

X	US 6 985 172 B1 (RIGNEY MICHAEL P [US] ET AL) 10 January 2006 (2006-01-10)
---	--

Y figures 2, 3, 5-8; column 6, lines 35-58;
column 8, lines 20-63; column 4, lines
4-5; column 7, 24-28, lines 35-38; column
9, lines 12-17; column 2, lines 38-39;
column 5, line 13; column 1, lines 45-46

1-4,
6-13,
15-20
5, 14

— / —

☒ Further documents are listed in the continuation of Box C.

x 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

12 August 2022

Date of mailing of the international search report

14/10/2022

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

Authorized officer _____

Loza, Artur

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2022/029490

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Anonymous: "Color difference - Wikipedia", / 14 January 2017 (2017-01-14), XP055890890, Retrieved from the Internet: URL:https://web.archive.org/web/20170114140820/https://en.wikipedia.org/wiki/Color_difference [retrieved on 2022-02-13] the whole document -----	2, 3, 11, 18
Y	US 2019/362177 A1 (PIOVANO JEROME [US]) 28 November 2019 (2019-11-28) paragraphs 31-32 -----	5, 14

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2022/029490

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims;; it is covered by claims Nos.:

1-20

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-20

how to identify an object region

2. claims: 21-44

how to improve efficiency and robustness of object
identification

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2022/029490

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 6985172	B1	10-01-2006	NONE	

US 2019362177	A1	28-11-2019	US 10380452 B1	13-08-2019
			US 2019362177 A1	28-11-2019
