



- (51) **International Patent Classification:**
G06F 17/30 (2006.01) *G06F 15/16* (2006.01)
- (21) **International Application Number:**
PCT/US2014/014293
- (22) **International Filing Date:**
31 January 2014 (31.01.2014)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
13/768,051 15 February 2013 (15.02.2013) US
- (71) **Applicant:** YAHOO! INC. [US/US]; 701 First Avenue,
Sunnyvale, California 94089 (US).
- (72) **Inventors:** ZHONG, Yu; Rochester, New York (US).
GARRIGUES, Pierre; San Francisco, California (US).
CULPEPPER, Benjamin Jackson; Berkeley, California
(US).
- (74) **Agents:** HEILBRUNN, Elise R. et al.; Weaver Austin Vil-
leneuve & Sampson LLP, P.O. Box 70250, Oakland, Cali-
fornia 94612-0250 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

[Continued on next page]

(54) **Title:** REAL TIME OBJECT SCANNING USING A MOBILE PHONE AND CLOUD-BASED VISUAL SEARCH ENGINE

(57) **Abstract:** A system for tagging an object comprises an interface and a processor. The interface is configured to receive an image. The processor is configured to determine a key frame. Determining a key frame comprises determining that the image is stable. The processor is configured to determine a tag for an item in the key frame.

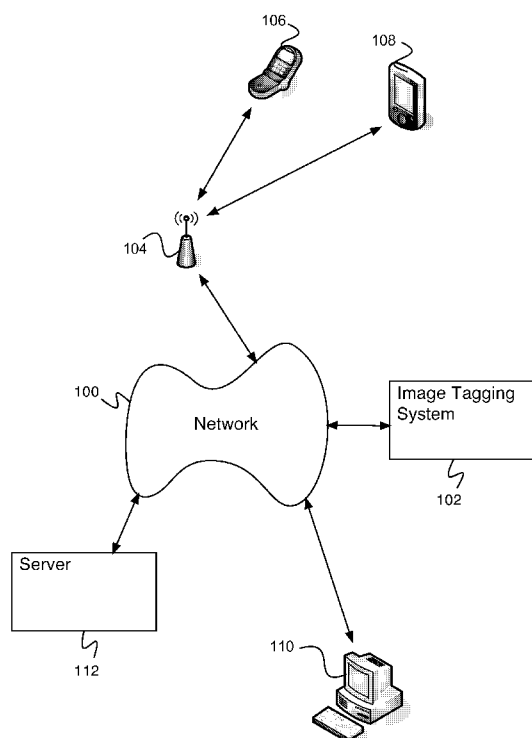


Fig. 1

WO 2014/126735 A1



Published:

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

REAL TIME OBJECT SCANNING USING A MOBILE PHONE AND CLOUD-BASED VISUAL SEARCH ENGINE

RELATED APPLICATIONS

5 **[0001]** This application claims priority from U.S. Patent Application No. 13/768,051 (Atty. Docket No. YAH1P297), entitled “Real Time Object Scanning Using a Mobile Phone and Cloud-Based Visual Search Engine,” by Zhong et al, filed on February 15, 2013, which is incorporated herein by reference in its entirety and for all purposes.

BACKGROUND OF THE INVENTION

10 **[0002]** Mobile phones and smart mobile phones are becoming extremely common. Many people now carry everywhere a device that includes a powerful processor, a high-resolution video camera, and a connection to the Internet. One possible practical application of these devices is image recognition. When a smart
15 phone user is presented with an item that he desires to know more about, he can photograph the item and use image recognition software to identify it. Depending on his preferences he may simply be given the name of the object, or automatically be connected with search results regarding the object, shopping information regarding the object, or any other appropriate information. A particularly useful application of
20 this technology is for visually impaired people; enabling them to identify objects they are not able to identify themselves without assistance. Frequently, when a user takes a photo with a smart phone for the purposes of object identification, the user does not take the photo that best enables the object to be identified – for instance, the image can be blurry or from an angle that makes recognition difficult. This is particularly a
25 problem when assisting blind users, as they cannot see the photo image themselves to verify its quality

BRIEF DESCRIPTION OF THE DRAWINGS

[0003] Various embodiments of the invention are disclosed in the following detailed description and the accompanying drawings.

5 [0004] Figure 1 is a block diagram illustrating an embodiment of a network system.

[0005] Figure 2 is a block diagram illustrating an embodiment of an image tagging system.

[0006] Figure 3 is a block diagram illustrating an embodiment of a mobile device.

10 [0007] Figure 4 is a flow diagram illustrating an embodiment of a process for using a system for real time object scanning.

[0008] Figure 5A is a flow diagram illustrating an embodiment of a process for processing a video stream.

15 [0009] Figure 5B is a flow diagram illustrating an embodiment of a process for processing a key frame.

[0010] Figure 6 is a flow diagram illustrating an embodiment of a process for processing an image frame.

[0011] Figure 7 is a flow diagram illustrating an embodiment of a process for determining if a frame is a key frame.

20 [0012] Figure 8 is a flow diagram illustrating an embodiment of a process for tagging an object.

DETAILED DESCRIPTION

[0013] The invention can be implemented in numerous ways, including as a process; an apparatus; a system; a composition of matter; a computer program product embodied on a computer readable storage medium; and/or a processor, such as a processor configured to execute instructions stored on and/or provided by a memory coupled to the processor. In this specification, these implementations, or any other form that the invention may take, may be referred to as techniques. In general, the order of the steps of disclosed processes may be altered within the scope of the invention. Unless stated otherwise, a component such as a processor or a memory described as being configured to perform a task may be implemented as a general component that is temporarily configured to perform the task at a given time or a specific component that is manufactured to perform the task. As used herein, the term ‘processor’ refers to one or more devices, circuits, and/or processing cores configured to process data, such as computer program instructions.

[0014] A detailed description of one or more embodiments of the invention is provided below along with accompanying figures that illustrate the principles of the invention. The invention is described in connection with such embodiments, but the invention is not limited to any embodiment. The scope of the invention is limited only by the claims and the invention encompasses numerous alternatives, modifications and equivalents. Numerous specific details are set forth in the following description in order to provide a thorough understanding of the invention. These details are provided for the purpose of example and the invention may be practiced according to the claims without some or all of these specific details. For the purpose of clarity, technical material that is known in the technical fields related to the invention has not been described in detail so that the invention is not unnecessarily obscured.

[0015] Real time object scanning using a mobile phone and cloud-based visual search engine is disclosed. A system for tagging an object comprises a processor configured to: determine a key frame, wherein determining a key frame comprises: determining if an image is stable; and in the event that the image is stable: determining if the image has been stable for a predetermined duration; and in the event that the image has been stable for the predetermined duration: identifying the

current frame as a key frame; and determine a tag for an item in the key frame. A memory is coupled to the processor and is configured to provide the processor with instructions. An interface is configured to receive an image (e.g., for processing by the processor).

5 **[0016]** In some embodiments, real time object identification is performed using a mobile phone including a camera and a processor. A mobile phone user performs real time object identification by initiating the object identification process (e.g., by launching an app) and holding their phone steady so the camera captures a steady image of the object of interest. Rather than the user making an indication (e.g.,
10 pushing a button, manually triggering, voice triggering, etc.) that the image is of the desired object, the app determines that the image is steady (e.g., features of interest are not blurry), when the image has been steady for a predetermined duration, and when the visual content is sufficiently different from the previous key frames, the image is captured and processed. By analyzing each frame of the image stream, the
15 app is able to ensure that a high quality image is received for image processing (e.g., in focus, within the frame, etc.). If the app requires the user to push a button to indicate that an image should be captured, it is forced to rely on the user to take a good quality image. This is particularly burdensome when the app is used for assisting a blind user, who will likely have difficulty taking a high quality image.

20 **[0017]** In some embodiments, an image is steady in the event that an object of interest is within a camera view for a predetermined time. In some embodiments, a steady image refers to an image from the perspective of a device with a camera, where the image, or rather stream of images, includes the same subject matter within the view of the camera. In some embodiments, the image is received via an interface to
25 be processed. In some embodiments, the image or stream of images is steady or within the view of the camera for a predetermined period of time. In some embodiments, the predetermined period of time is indicative of a length of time to express the intent of interest in the objects in the view of the camera.

[0018] In some embodiments, when the object identification process is
30 initiated, the processor receives an image stream from the camera. Each image frame is analyzed until a key frame is detected. A key frame comprises an image frame where the image is steady and has been held steady over a predetermined duration,

and where the visual content is different from the previous key frames. When a key frame is detected, the image is provided to object identification software for identifying an object in the image. In some embodiments, the object is identified by an automated object identification process. In some embodiments, the object is identified by a human object identifier. In some embodiments, it is determined whether the object can be identified by an automated object identification process, and if not, the object is identified by a human object identifier.

[0019] In some embodiments, the system comprises a processor for image recognition which performs the expensive computation of analyzing a new object and a key frame processor which performs the less expensive identification of frames that are susceptible frames to having a new object. If the frame is identified as being susceptible by the key frame processor, the frame is sent on to the image recognition processor.

[0020] Figure 1 is a block diagram illustrating an embodiment of a network system. In the example shown, the network system of Figure 1 comprises image tagging system 102 communicating with network 100. The network system of Figure 1 performs real time object scanning using a mobile phone (e.g., smart mobile phone 108) and a cloud-based visual search engine (e.g., image tagging system 102). In some embodiments, the image tagging system is on the mobile device where key frame identification happens. In some embodiments, object recognition is performed on the cloud-based visual search engine due to the power consumption load it requires. Image tagging system 102 comprises a system for assigning tags to objects in images. In various embodiments, network 100 comprises one or more of the following: a local area network, a wide area network, a wired network, a wireless network, the Internet, a cellular network, or any other appropriate network. In the example shown, a user submits an image to be tagged to image tagging system 102 via network 100. In some embodiments, the image to be tagged is created using a mobile device (e.g., camera enabled cellular telephone 106, smart mobile phone 108, or another mobile device, for instance a wireless enabled digital camera, a tablet computer, a laptop, a tablet, a camera, etc.) and conveyed to network 100 via wireless access point 104. In some embodiments, the image to be tagged is located on a user's home computer (e.g., computer 110), either after it was uploaded by the user from a

digital camera or other image creation device or downloaded from the Internet, and is submitted through a wired connection to network 100. In some embodiments, the image tagging system user comprises a web site or other computing system that uses image tagging as part of its service, and images are submitted from the user's system server (e.g., server 112) through network 100, to image tagging system 102.

[0021] Figure 2 is a block diagram illustrating an embodiment of an image tagging system. In some embodiments, image tagging system 200 comprises image tagging system 102 of Figure 1. Image tagging system 200 comprises an image tagging system for receiving an image and providing one or more tags for one or more objects in the image. In the example shown, image tagging system 200 comprises web interface 202, computer vision engines 204, human computation module(s) 206, object databases 208, results processor 210, learning module 212, and dashboard 214. Web interface 202 comprises an interface accessible through the World Wide Web for submitting images to image tagging system 200. In some embodiments, web interface 202 comprises a graphical user interface (GUI), enabling a user to easily submit images to image tagging system 202. In some embodiments, web interface 202 comprises an application programming interface (API), enabling a programmer to create software that submits images and other information to image tagging system 200 and receives tags from image tagging system 200. In some embodiments, web interface 202 uses a communication interface other than the World Wide Web (e.g., a local area network, a wide area network).

[0022] In the example shown, computer vision engines 204 comprises a set of one or more computer vision engines for automatically identifying objects in images received by image tagging system 200. In various embodiments, the computer vision engines use an instance recognition module, a classification module, an optical character recognition module, a face identification module, a barcode recognition module, a clock recognition module, a color detection module, or any other appropriate module. In some embodiments, each different module utilized by computer vision engines 204 is utilized by a separate computer vision engine. In some embodiments, computer vision engines 204 comprise more than one computer vision engine, and the separate computer vision engines execute their modules in parallel. In some embodiments, computer vision engines 204 utilize information stored in object

databases 208 as part of executing computer vision modules. In some embodiments, an instance recognition module utilizes object databases 208 as its instance library. In some embodiments, a classification module utilizes object databases 208 as its class representation library.

5 **[0023]** In the example shown, human computation module(s) 206 comprises one or more human computation modules for image recognition. A human computation module for image recognition comprises one or more humans each capable of recognizing images and providing tags for the images. In some embodiments, a human computation module relays any tags provided by its humans
10 directly to results processor 210. In some embodiments, a human computation module requests tags from human taggers. In some embodiments, a human computation module receives a request for tags from a results processor. In some embodiments, the human computation module provides various processing functionality prior to providing the tags to results processor 210. In various embodiments, processing
15 functionality includes determining a status of a human tagger, determining a validity of a tag based on a status of a human tagger (e.g., valid tag is determined in the event that the tag is received from a tagger indicated as an expert or reliable tagger), confirming a validity of a tag based on receipt from multiple human taggers (e.g., valid tag is determined in the event that the same tag is received from a majority of
20 taggers or a minimum number of taggers or both), marking a tag with a validity status based on a status of a human tagger, or any other appropriate processing functionality. In some embodiments, human computation module(s) 206 comprises more than one human computation module. In various embodiments, human taggers are divided into separate human computation modules based on specialty, experience, age, location,
25 average time to provide results, tagger status, or any other appropriate tagger classification.

[0024] Results processor 210 comprises a results processor for receiving and processing tags from computer vision engines 204 and human computation module(s) 206. In the example shown, results processor 210 ranks the received tags according to
30 a predetermined ranking algorithm and provides the highest ranked tag or tags to the image tagging system user (e.g., the user that originally provided the image). In various embodiments, the predetermined ranking algorithm utilizes a module ranking

(e.g., tags received from a computer vision instance module rank more highly than tags received from a computer vision classification module), a module score (e.g., tags received from the same module are ranked based on a score assigned to them by the module they are received from), or any other appropriate ranking criteria. In some embodiments, results processor 210 requests tags from computer vision engines 204. In some embodiments, results processor 210 requests tags from human computation module(s) 206. In some embodiments, results processor 210 requests tags from human computation module(s) 206 in the event that computer vision engines 204 are not able to provide tags.

[0025] Learning module 212 comprises a module for updating object databases 208 based on the results of an image query. In some embodiments, if an instance recognition module or a classification module are able to determine tags for an image based on information stored in object databases 208 without finding an identical image stored in object databases 208, the learning module then stores the image in object databases 208 associated with the determined tags. In some embodiments, the learning module determines whether adding the image to object databases 208 will broaden the space of images associated with the determined tag. The learning module may then choose to store the image in object databases 208 only if it does broaden the space of images associated with the tag. In some embodiments, if computer vision engines 204 is not able to determine a tag associated with the image but human computation module(s) 206 is, learning module 212 stores the image and associated tag in object databases 208, so that the image can be automatically identified if it is submitted in future queries. In some embodiments, learning module 212 processes an image before storing it and any associated tags in object databases 208. In various embodiments, processing comprises background subtraction, object geometry processing, object class representation creation, or any other appropriate processing.

[0026] Dashboard 214 comprises an interface for an image tagging system administrator to gain information about the current status of the image tagging system. In the example shown, the dashboard system can provide information to a system administrator on image queries, tags that have been applied to images, system

users, image taggers, and system status, as well as allowing the administrator to train the computer vision module by adding images and tags at his discretion.

[0027] Figure 3 is a block diagram illustrating an embodiment of a mobile device. In some embodiments, mobile device 300 comprises smart mobile phone 108 of Figure 1. In various embodiments, mobile device 300 comprises a smart mobile phone, a tablet computer, a laptop computer, an embedded system, or any other appropriate mobile device. Mobile device 300 performs real time object scanning. In some embodiments, mobile device 300 interacts with a cloud-based visual search engine to perform real time object scanning. In the example shown, mobile device 300 comprises camera 302, image processor 304, object databases 306, network interface 308, results processor 310, and user interface 312. Camera 302 comprises a camera for recording images. In some embodiments, camera 302 comprises a still camera (e.g., records only one still image at a time). In some embodiments, camera 302 comprises a video camera (e.g., records images at a given rate, e.g., 5 frames per second, 10 frames per second, 30 frames per second, etc.). Image processor 304 comprises an image processor for processing images. In various embodiments, image processor 304 performs image cropping, image filtering, image transforming, image parameter extraction, object identification, object tagging, key frame detection, or any other appropriate image processing function. Object databases 306 comprise(s) one or more databases of object tags. In some embodiments, object databases 306 comprise(s) a copy of object databases 208. In some embodiments, object databases 306 comprises a subset of object databases 208. Network interface 308 comprises a network interface for communicating with a network (e.g., network 100 of Figure 1). In some embodiments, object databases 306 are updated using network interface 308. In some embodiments, images are uploaded (e.g., to an image tagging system) and tags are downloaded via network interface 308. Results processor 310 comprises a processor for receiving and processing image tagging results (e.g., from object databases 306, image processor 304, or network interface 308). Results processor 310 determines result quality, selects the best results, and processes results for display by user interface 312. User interface 312 receives results from results processor 310 and displays results for a user. In various embodiments, user interface 312 receives an application start command, receives a stream select start command, displays an image

stream to a user, displays an image tag to a user, or performs any other appropriate user interface function.

[0028] Figure 4 is a flow diagram illustrating an embodiment of a process for using a system for real time object scanning. In some embodiments, the process of Figure 4 is executed by a mobile device user using an app for real time object scanning using a mobile phone and cloud-based search engine. In some embodiments, the search engine is on the mobile phone. In the example shown, in 400, the app is launched. In 402, the camera is pointed at an object. In some embodiments, the camera is pointed at the object the user is interested in identifying. The user does not need to make an indication (e.g., click a button, etc.) to the mobile device that the camera is pointed at the object of interest or that the camera should take a photograph. In 404, the device is held stable. In some embodiments, the device is held stable for at least a predetermined period of time (e.g., 0.5 seconds, 5 seconds, etc.). In 406, an object identification is received. In some embodiments, an object identification is determined by the device and presented to the user. In some embodiments, the device interacts with an image tagging system (e.g., communicating via a network) to determine an object identification. In some embodiments, the image tagging system determines an object identification. In some embodiments, the image tagging system interacts with a human to determine an object identification. In some embodiments, object identification comprises an object tag.

[0029] Figure 5A is a flow diagram illustrating an embodiment of a process for processing a video stream. In some embodiments, the process of Figure 5A is executed by a mobile device (e.g., mobile device 300 of Figure 3). The process of Figure 5A is used by a mobile device for real time object scanning without interaction from a user (e.g., without requiring the user to make an indication of the object of interest or manually indicate to take a photograph). In the example shown, in 500, an image frame is processed. In some embodiments, processing an image frame comprises determining if the image frame is a key frame. In some embodiments, an image frame received from a camera is processed. In various embodiments, image frames are captured at 1 frame per second, 5 frames per second, 10 frames per second, 30 frames per second, or at any other appropriate rate. In 502, the process waits for

the next image frame. Once the next image frame has been received, control passes back to 500.

[0030] Figure 5B is a flow diagram illustrating an embodiment of a process for processing a key frame. In some embodiments, the process of Figure 5B is used to indicate an object tag in a key frame to a user. In the example shown, in 520, a key frame is received. In some embodiments, the key frame comprises a key frame identified in 500 of Figure 5A. In some embodiments, a key frame comprises a frame where the image is still (e.g., not blurry). In some embodiments, a key frame comprises a frame where the image is still and the image has been still for a predetermined number of frames before it. In some embodiments, a key frame comprises a frame where the image's visual content is different from the previously detected key frames. In 522, an object is recognized in the key frame. In some embodiments, the object is recognized by the mobile device. In some embodiments, the object is recognized by an image tagging system. In 524, object recognition user interface information is provided. In some embodiments, providing object recognition user interface information comprises displaying an object tag.

[0031] Figure 6 is a flow diagram illustrating an embodiment of a process for processing an image frame. In some embodiments, the process of Figure 6 is executed by image processor 304 of Figure 3. An image processor uses the process of Figure 6 for processing a received image frame and determining if the image frame is locked onto an object. In some embodiments, the process of Figure 6 implements 500 of Figure 5A. In the example shown, in 600, an image frame is received. In some embodiments, the image frame is received from a camera. In some embodiments, the image frame is part of a video stream. In 602, it is determined whether there are points to track. In the event that there are not points to track, in 604 points to track are looked for, and then the process ends. In the event that there are points to track, in 606 displacement for point(s) are found. In 608, it is determined if a frame is a key frame.

[0032] In some embodiments, features are extracted at salient locations such as corners or object contours. A local patch centered around the salient location is encoded. The encoding typically captures histograms of oriented edge energy. The features are compared to the features in the previous frame, and are used to compute

the optical flow. In some embodiments, optical flow is estimated using a Lukas Kanade algorithm. It is determined which feature(s) is/are locked onto. In some embodiments, feature(s) is/are used to determine an object tag. In some embodiments, determining whether feature(s) is/are locked comprises checking to see if feature(s) have been found in a previous frame. In some embodiments, if feature(s) have been locked, the image is steady and a key frame can be detected. In the event feature(s) is/are locked, it is determined if the frame is a key frame. If it is determined that feature(s) have not been locked, feature(s) is/are looked for. In some embodiments, it is determined whether feature(s) have been found. In the event feature(s) have been found, the lock on to feature(s) is indicated. For example, a software module is informed (e.g., by passing a message or indication) that the feature(s) are locked and that the process may proceed (e.g., to process additional images to determine a key frame). Or for example, the lock is stored or recorded (e.g., in a data structure or in memory location(s)). In some embodiments, after the lock on to feature(s) is indicated, the next time the process is executed (e.g., in response to the next image frame received), it will be determined that feature(s) is/ are locked. If it is determined that feature(s) have not been found, the process ends. In some embodiments, a lack of lock on to features is indicated (e.g., stored, recorded, passed as an indication or variable to a software module, etc.).

[0033] Figure 7 is a flow diagram illustrating an embodiment of a process for determining if a frame is a key frame. In some embodiments, the process of Figure 7 implements 608. An image processor uses the process of Figure 7 for processing a received image frame and determining if the image frame is a key frame. In the example shown, in 700, movement of feature(s) (e.g., movement of features since the last frame, blurriness of features, etc.) is determined. In 702, feature(s) with substantial movement (e.g., features that have moved more than a threshold since the last frame, features that are above a blurriness threshold, etc.) are removed. In 704, it is determined whether at least N features remain. In the event that less than N features remain, control passes to 706. In 706, feature(s) is/are reset. In the event that it is determined in 704 that at least N features remain, control passes to 708. In 708, stable frame counter is increased. In some embodiments, the value stored in the stable frame counter represents the duration that the image has been stable. In 710, it is determined whether there is an existing key frame for the feature(s). In the event that

there is an existing key frame for the feature(s), the process ends. In the event that there is not an existing key frame for the feature(s), then in 712 it is determined whether there are enough stable frames. In the event that there are not enough stable frames, the process ends. In the event that there are enough stable frames, in 714 the image is stored as key frame, and the process ends.

[0034] In some embodiments, an accelerometer measurement is used to determine whether there are stable frames. In some embodiments, the accelerometer is associated with a mobile phone.

[0035] In some embodiments, a key frame has different visual content from a previous key frame. In some embodiments, a key frame has the same visual content from a previous key frame.

[0036] Figure 8 is a flow diagram illustrating an embodiment of a process for tagging an object. In some embodiments, the process of Figure 8 is executed by mobile device 300 of Figure 3. The process of Figure 8 is used for real time object scanning using a mobile phone and cloud-based search engine. In some embodiments, the process of Figure 8 is executed on each frame of an incoming image stream. In the example shown, in 800, it is determined whether the image is stable. If it is determined that the image is not stable, the process ends. In some embodiments, the process is then repeated on the next image frame. If it is determined in 800 that the image is stable, control passes to 802. In 802, it is determined if the image has been stable for a predetermined duration. If it is determined that the image has not been stable for the predetermined duration, the process ends. In some embodiments, the process is then repeated on the next image frame. If it is determined in 802 that the image has been stable for the predetermined duration, control passes to 804. In 804, the current frame is identified as a key frame. In 806, a tag is determined for an item in the key frame. In some embodiments, the tag is determined by a machine. In some embodiments, the process additionally determines whether the tag is determined by a machine or by a human. In some embodiments, in the event it is determined that the tag cannot be determined by a machine, the tag is determined by a human.

[0037] Although the foregoing embodiments have been described in some detail for purposes of clarity of understanding, the invention is not limited to the

details provided. There are many alternative ways of implementing the invention.
The disclosed embodiments are illustrative and not restrictive.

WHAT IS CLAIMED IS:

CLAIMS

1. A system for tagging an object, comprising:
an interface configured to receive an image; and
5 a processor configured to:
determine a key frame, wherein determining a key frame comprises
determining that the image is stable; and
determine a tag for an item in the key frame.
2. A system as in claim 1, wherein determining that an image is stable comprises
10 determining a feature.
3. A system as in claim 2, wherein the feature is in a current frame and in a prior frame.
4. A system as in claim 2, wherein the feature is one of a plurality of features.
5. A system as in claim 4, wherein a percentage of the plurality of features is
15 above a threshold for a predetermined number of frames.
6. A system as in claim 4, wherein a number of the plurality of features is above a threshold for a predetermined number of frames.
7. A system as in any of claims 4, wherein each of the plurality of features is in a current frame and a prior frame.
- 20 8. A system as in any of claims 4, wherein movement is determined associated with each of the plurality of features.
9. A system as in any of claims 4-8, wherein in the event that a specific feature of the plurality of features has movement above a threshold, remove the specific feature point from the plurality of features.
- 25 10. A system as in any of claims 1, wherein the image is received from a camera on a mobile device.
11. A system as in claim 10, wherein the mobile device comprises a mobile phone.
12. A system as in claim 10, wherein the mobile device comprises a tablet.

13. A system as in any of claims 1, wherein the tag is determined using a computer vision engine.

14. A system as in any of claims 1, wherein the tag is determined using a human computation module.

5 15. A system as in claim 14, wherein the human computation module requests the tag from one or more human taggers.

16. A system as in claim 15, wherein the human computation module determines validity of the tag.

10 17. A system as in any of claims 14 or 15, wherein the tag is determined using the human computation module in the event that the tag could not be determined using the computer vision module.

18. A system as in any of claims 1, wherein the key frame has visual content different from visual content from a previous key frame.

19. A method for tagging an object, comprising:

15 determining a key frame, wherein determining a key frame comprises determining that an image is stable; and
determining a tag for an item in the key frame.

20. A computer program product for tagging an object, the computer program product being embodied in a tangible computer readable storage medium and
20 comprising computer instructions for:

determining a key frame, wherein determining a key frame comprises determining that an image is stable; and
determining a tag for an item in the key frame.

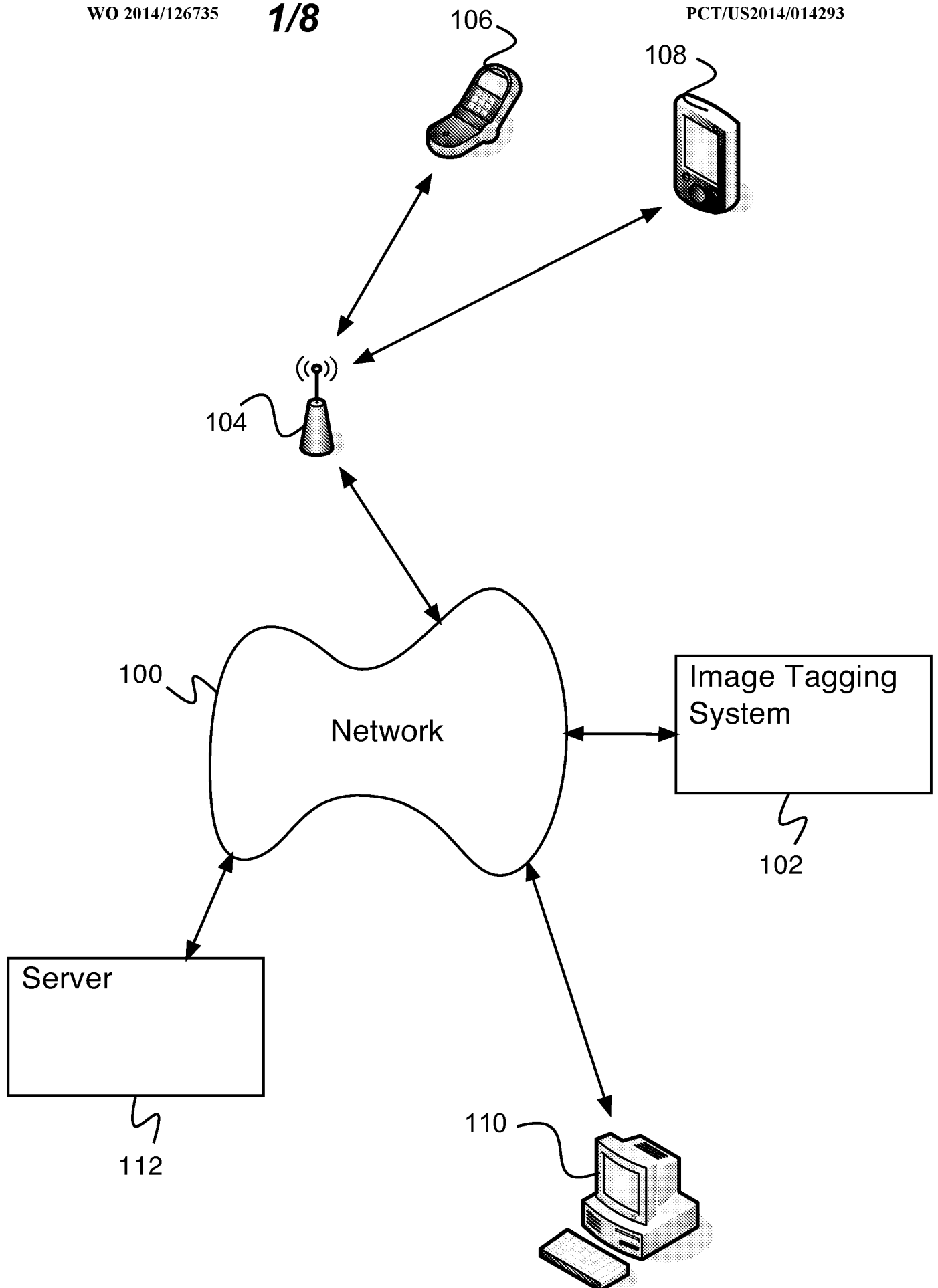


Fig. 1

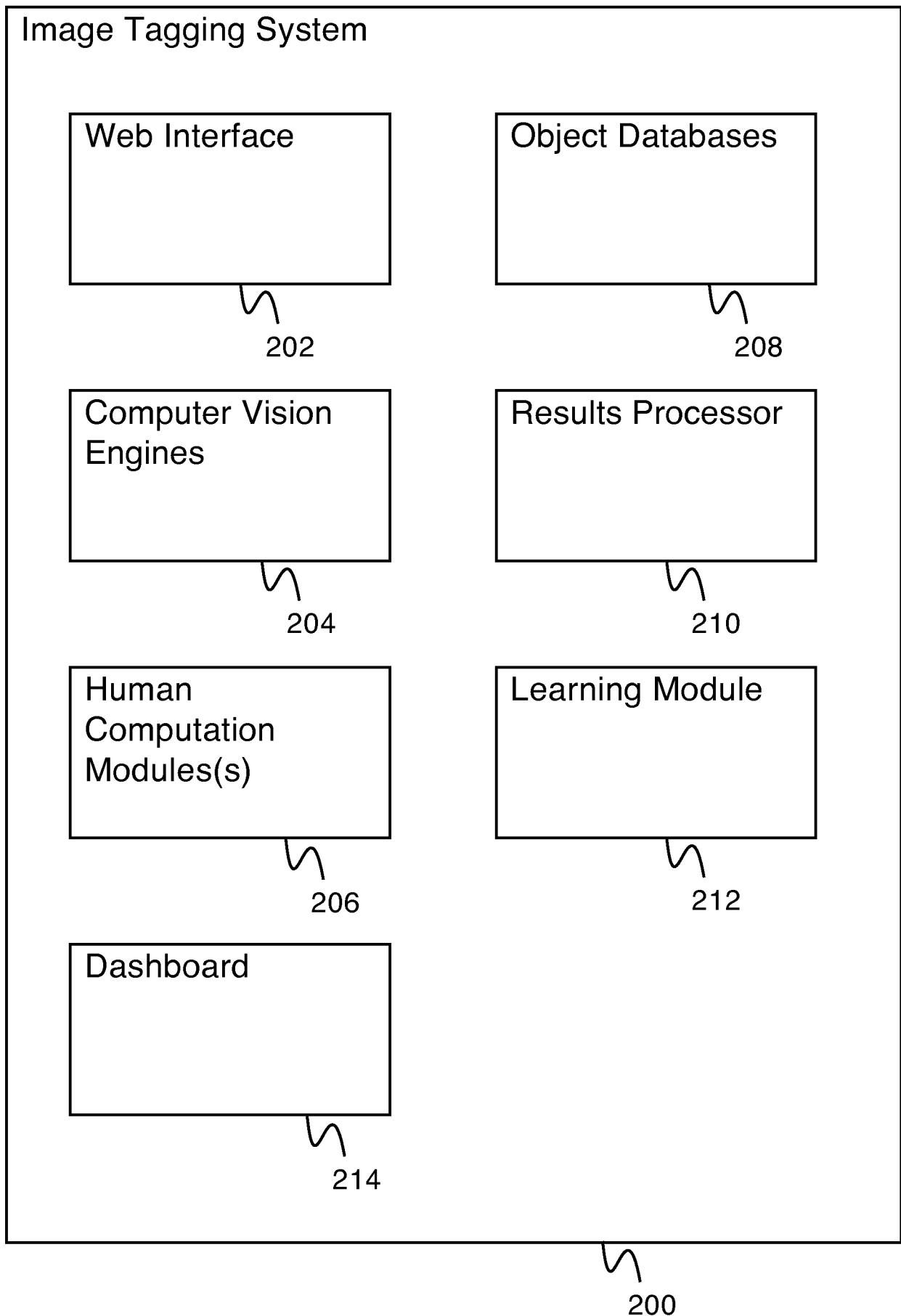


Fig. 2

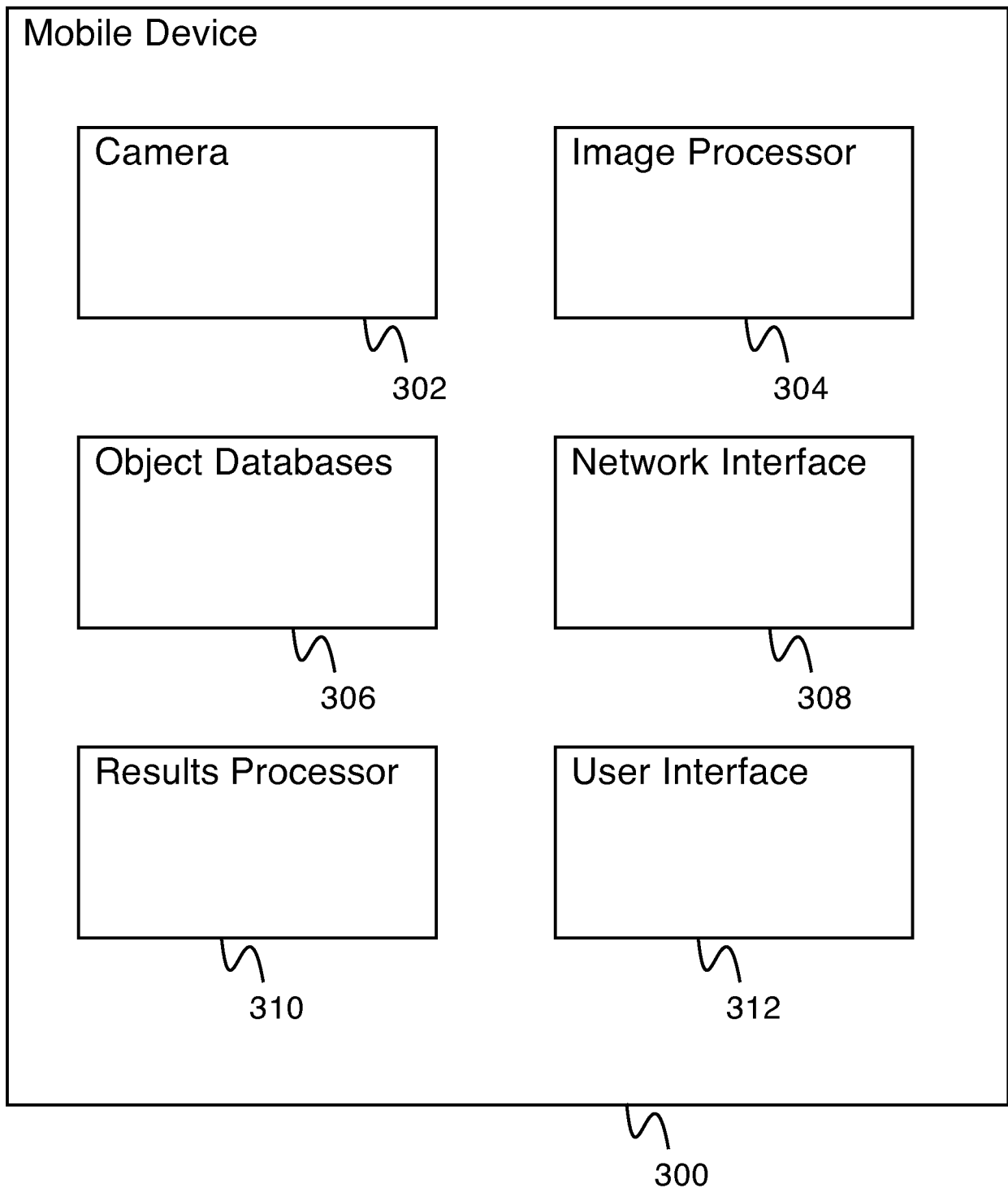


Fig. 3

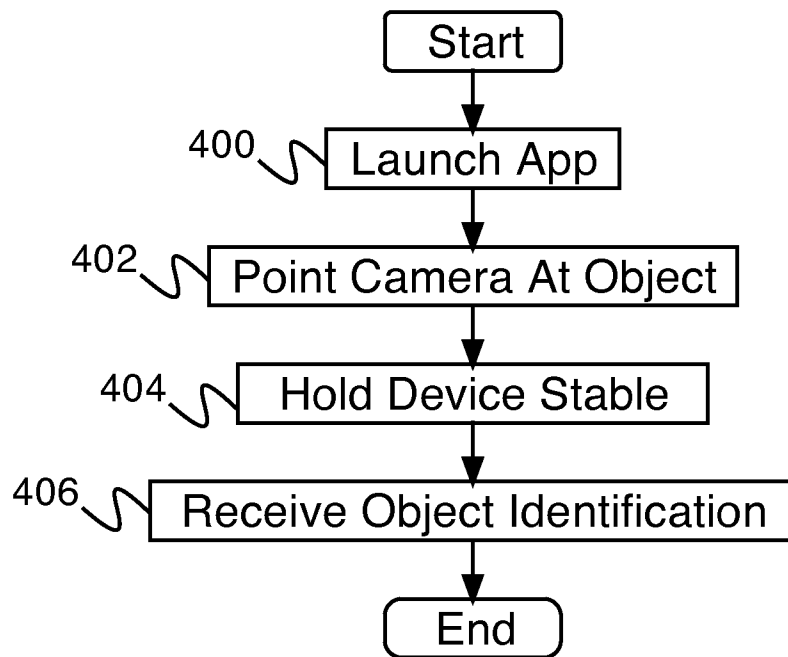


Fig. 4

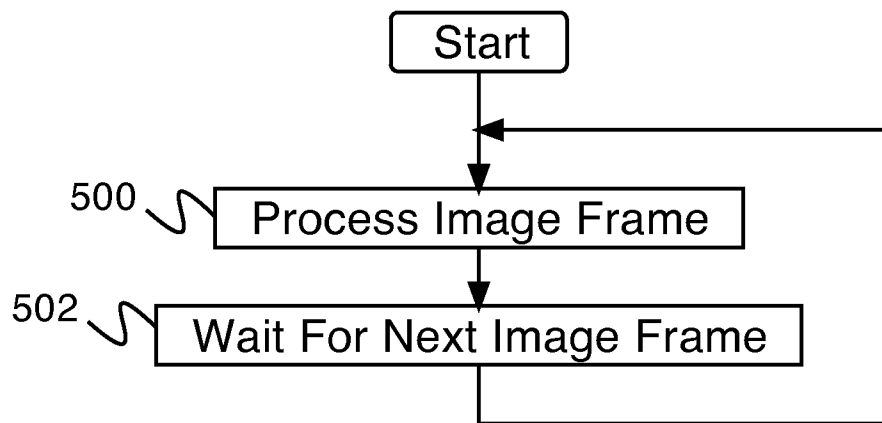


Fig. 5A

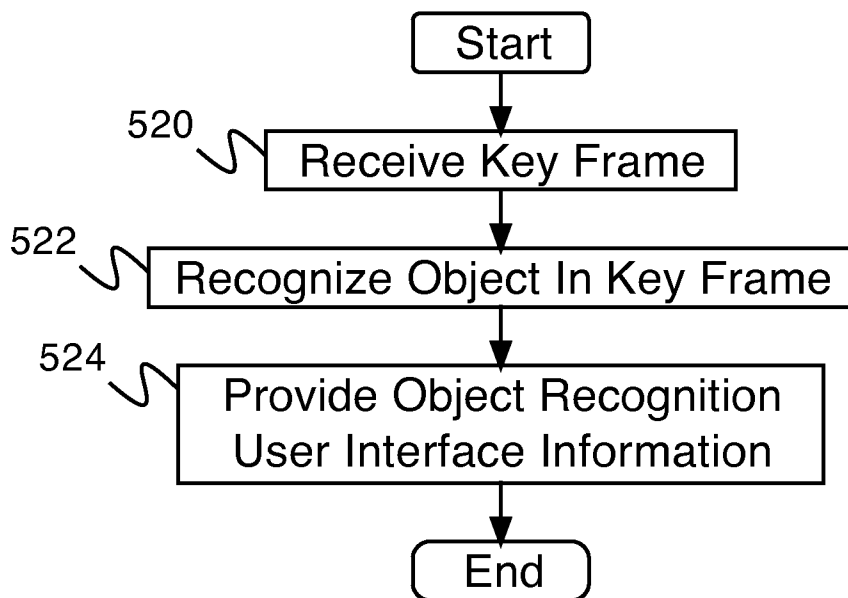


Fig. 5B

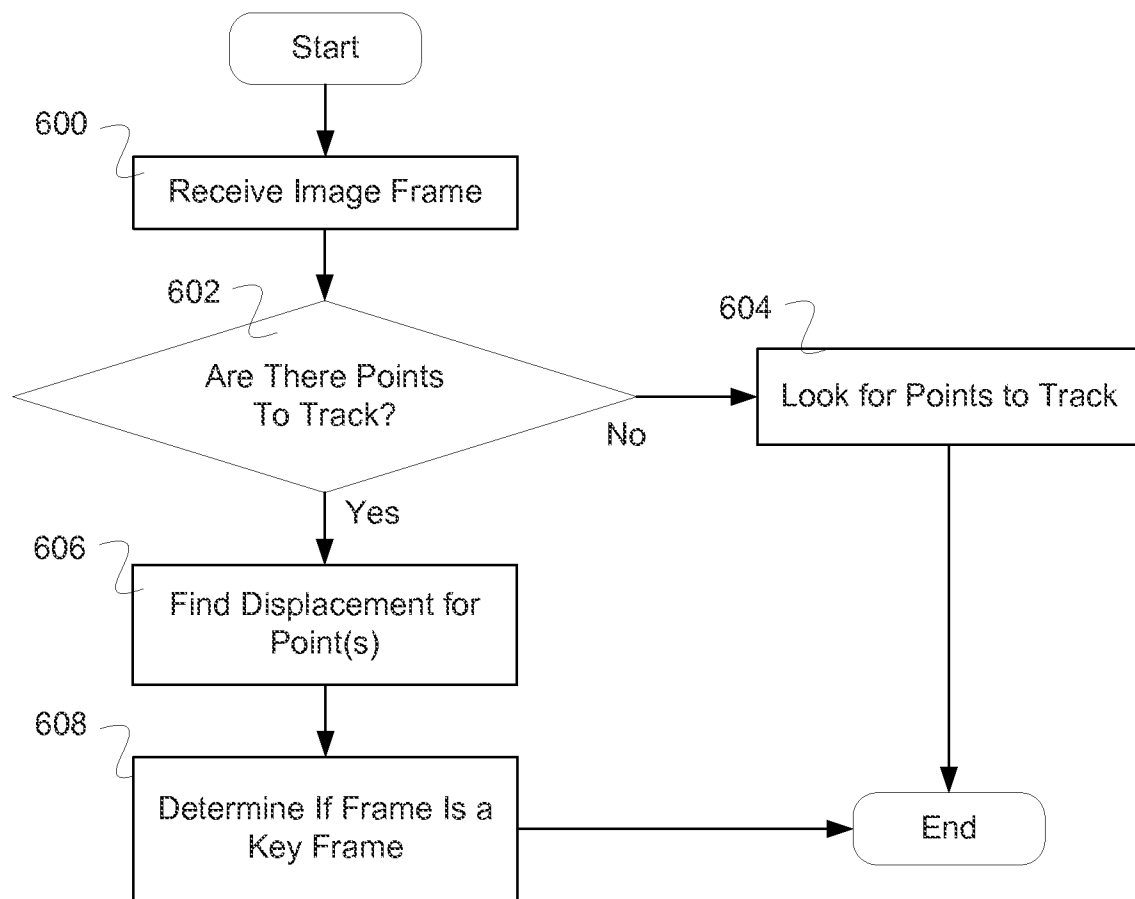


FIG. 6

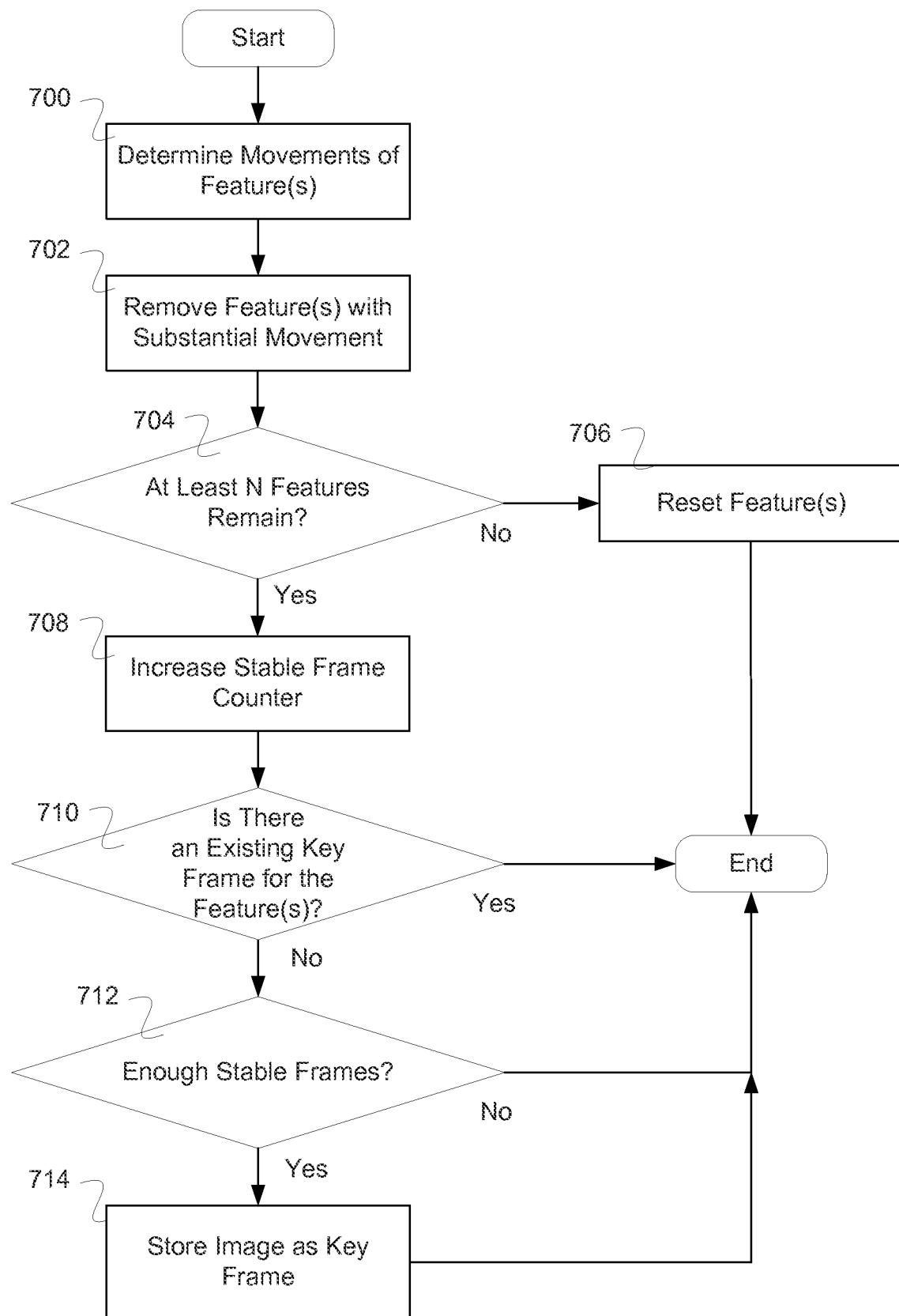


FIG. 7

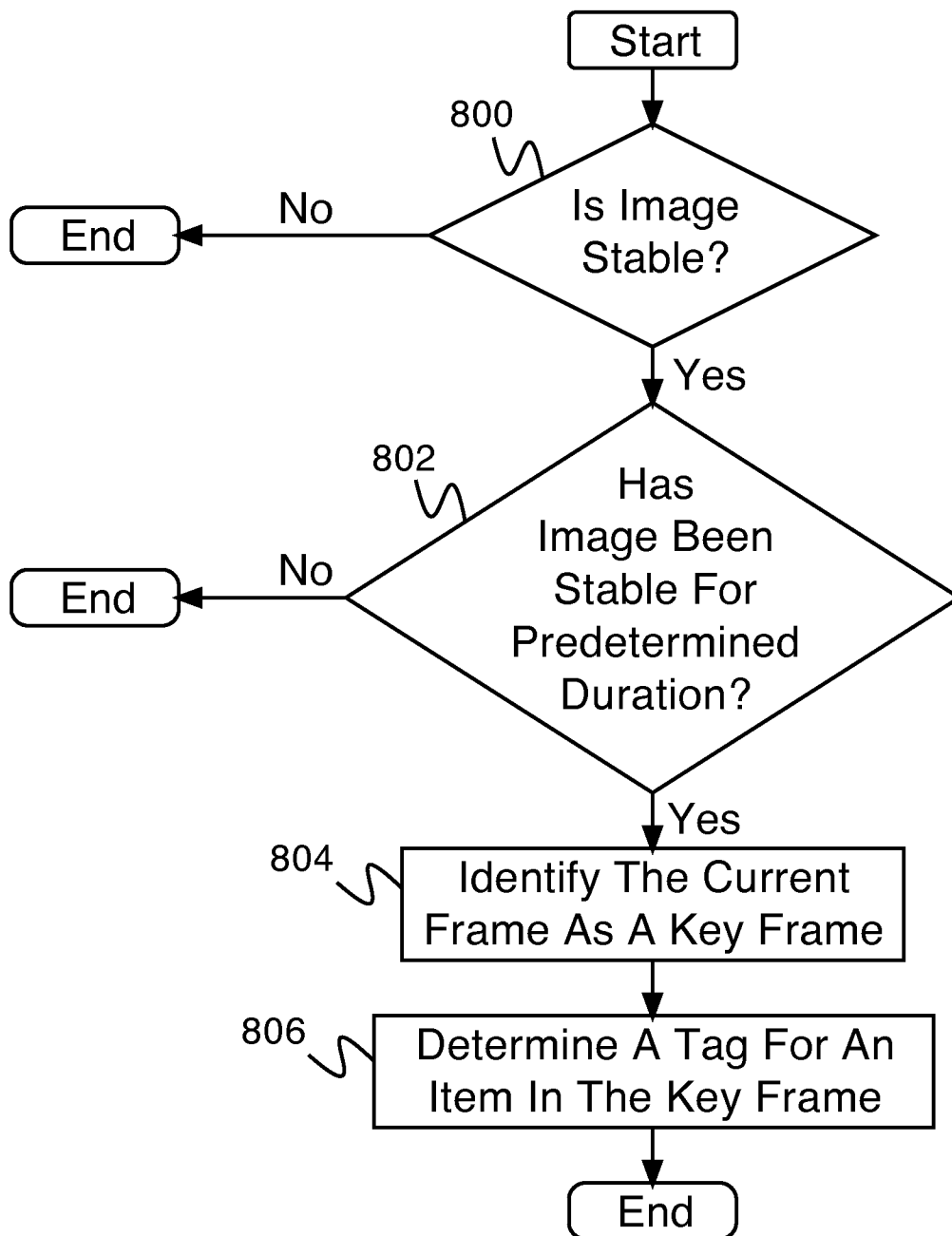


Fig. 8

A. CLASSIFICATION OF SUBJECT MATTER**G06F 17/30(2006.01)i, G06F 15/16(2006.01)i**

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F 17/30; H04N 7/18; H04N 5/225; G06Q 30/00; G06K 9/00; G09G 5/00; G06F 15/16

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

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

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

eKOMPASS(KIPO internal) & keywords: tagging an object, key frame, stable image

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2012-0299961 A1 (GURUMURTHY D. RAMKUMAR et al.) 29 November 2012 See paragraphs [0023]-[0024], [0031]-[0034], [0045]-[0046]; and figures 2, 5-6, 9.	1-20
A	US 2011-0216179 A1 (ORANG DIALAMEH et al.) 08 September 2011 See paragraphs [0092]-[0097]; and figure 5.	1-20
A	US 2012-0308077 A1 (ERICK TSENG) 06 December 2012 See paragraphs [0027]-[0043], [0048]-[0049]; and figures 1-3, 6.	1-20
A	US 8321293 B2 (ALAN LEWIS et al.) 27 November 2012 See column 5, line 60 - column 7, line 3; and figures 5-7.	1-20
A	US 2012-0098981 A1 (SHU-CHING IP et al.) 26 April 2012 See paragraphs [0029]-[0030], [0035], [0040]-[0041]; and figures 2, 5, 8.	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

29 May 2014 (29.05.2014)

Date of mailing of the international search report

30 May 2014 (30.05.2014)

Name and mailing address of the ISA/KR

International Application Division
Korean Intellectual Property Office
189 Cheongsu-ro, Seo-gu, Daejeon Metropolitan City, 302-701,
Republic of Korea

Facsimile No. +82-42-472-7140

Authorized officer

JI, Jeong Hoon

Telephone No. +82-42-481-5688



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2014/014293

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2012-0299961 A1	29/11/2012	WO 2012-166577 A1	06/12/2012
US 2011-0216179 A1	08/09/2011	CN 102906810 A	30/01/2013
		EP 2539883 A1	02/01/2013
		JP 2013-520757 A	06/06/2013
		KR 10-2013-0039720 A	22/04/2013
		US 8605141 B2	10/12/2013
		WO 2011-106520 A1	01/09/2011
US 2012-0308077 A1	06/12/2012	WO 2012-166874 A2	06/12/2012
		WO 2012-166874 A3	07/03/2013
US 8321293 B2	27/11/2012	US 2010-114736 A1	06/05/2010
		US 2011-264549 A1	27/10/2011
		US 2013-091030 A1	11/04/2013
		US 2013-339197 A1	19/12/2013
		US 7991646 B2	02/08/2011
		US 8521609 B2	27/08/2013
US 2012-0098981 A1	26/04/2012	EP 2445193 A2	25/04/2012
		EP 2445193 A3	20/06/2012
		TW 201218088 A	01/05/2012