



## (51) International Patent Classification:

*H04N 5/262* (2006.01) *H04N 5/265* (2006.01)  
*G11B 27/34* (2006.01)

## (21) International Application Number:

PCT/US2014/017549

## (22) International Filing Date:

21 February 2014 (21.02.2014)

## (25) Filing Language:

English

## (26) Publication Language:

English

## (30) Priority Data:

294/KOL/2013 14 March 2013 (14.03.2013)

IN

(71) Applicant: **MOTOROLA MOBILITY LLC** [US/US];  
600 North US Highway 45, Libertyville, Illinois 60048  
(US).

(72) Inventors: **RAGHAVAN, Krishnan**,; 109, My Ville,  
Keerthi Riviera Apts. CV Raman Nagar Bangalore,  
Karnataka 560093 (IN). **SHANBHOGUE A., Hari-**  
**prasad**,; Skanda, No. 156, 4th Cross G.K.W. Layout, Vi-  
jayanagar, Karnataka 560040 (IN).

(74) Agents: **PACE, Lalita W.**, et al.; 600 North US Highway  
45, Libertyville, Illinois 60048 (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).

## Published:

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

(54) Title: CONTEXT-BASED TAGGING OF PHOTOGRAPHIC IMAGES BASED ON RECORDED AUDIO AT TIME OF IMAGE CAPTURE

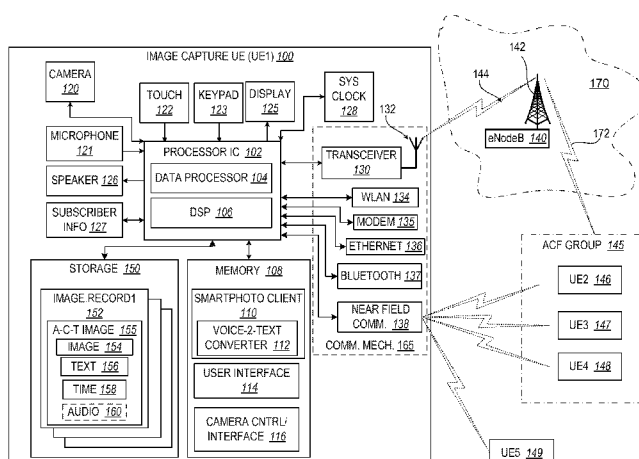


FIG. 1

(57) Abstract: A device (100) includes an image capturing component (120), a timer (260), a communication mechanism (165) that enables the device to communicate with at least one second device (146); and a processor (102). The processor executes an image capture utility (110) that configures the device to: activate (506) the timer to begin tracking a time sequence that extends to an end time following capture of the image; generate (508) an audio capture activation (ACA) signal and transmit the audio capture activation signal via the communication mechanism to the at least one second device; capture the image during the time sequence; receive (518) from the at least one second device content that represents audio that was captured at the at least one second device during the time sequence; link (526) the received content with the image; and store (528) the content linked image as an audio-content-tagged image. During image display, an audio content tag and timeline are displayed as tags in the image.

## **CONTEXT-BASED TAGGING OF PHOTOGRAPHIC IMAGES BASED ON RECORDED AUDIO AT TIME OF IMAGE CAPTURE**

### **BACKGROUND**

#### **1. Technical Field**

[0001] The present disclosure generally relates to electronic image capturing devices and in particular to image capturing devices capable of communicating with other devices. Still more particularly, aspects of the present disclosure relate to a method and device for tagging photographs and/or images taken by an electronic image capturing device.

#### **2. Description of the Related Art**

[0002] Electronic photography has become common place with the inclusion of image capturing technology in a host of portable user devices, such as smart phones, personal digital assistants (PDAs), and tablet computers or tablets. Increasingly, persons taking these photographs publish the photographs on electronic sites, such as social networking sites or other share sites, which can be accessible by the owner of the image, as well as by others. One mechanism by which an image owner makes others aware of a time and location at which the image was taken is by tagging the image. Increasingly, with these online sites, tagging a person in a photograph is common. These tags, however, typically only provide limited information about the identity of the person within the photograph. Very little information is provided about the circumstances occurring at the time the picture was taken or what if anything affected the mood or expression of the person in the photograph. Any such detail has to be manually provided by the person and added as a separate post to the image. Often, the person who uploads the image forgets to place the post or simply forgets important aspects surrounding the timing of the photograph capturing event.

## BRIEF DESCRIPTION OF THE DRAWINGS

[0003] The disclosure will best be understood by reference to the following detailed description of illustrative embodiments when read in conjunction with the accompanying drawings, wherein:

[0004] **FIG. 1** provides a block diagram representation of an example user equipment configured with various functional components that enable one or more of the described features of the disclosure, according to one embodiment;

[0005] **FIG. 2** illustrates an example SmartPhoto execution environment with functional components thereof, according to one or more embodiments;

[0006] **FIG. 3** illustrates an example implementation scenario in which a photographic image is taken of a plurality of subjects, several of whom have personal devices that can record audio and who are speaking before, during and/or after the capture of the image by the photographer;

[0007] **FIG. 4** is a flow diagram illustrating a method by which a first user equipment generates and updates a list of authorized second devices to which an audio capture request can be sent during image capture event occurring at the first UE, according to one or more embodiments;

[0008] **FIG. 5** is a flow diagram illustrating a method by which a first user equipment performs image capture and creates an audio context-tagged image by triggering one or more second UEs to provide detected speech audio feedback data during the image capture event, according to one or more embodiments;

[0009] **FIG. 6** illustrates an example audio context-tagged image in which all audio recorded during the image capture event are presented concurrently as tags on a single frame of the image, according to one embodiment;

[0010] **FIGs. 7-8** provide two different views or frames of the image with source-specific audio context tagging that identifies spoken speech during the image capture event as pop up speech content, according to several embodiments;

[0011] **FIG. 9** provides yet another view of the image with source-specific audio context tagging that identifies the spoken speech during the image capture event as text boxes placed on a periphery of the image, according to several embodiments

[0012] **FIGs. 10-12** provide three different views or frames of the image with time-specific audio context tagging that identifies spoken speech at different times during the image capture event as pop up speech content, according to several embodiments;

[0013] **FIG. 13** provides yet another view of the image with time-specific audio context tagging that identifies the spoken speech at different times during the image capture event as text boxes placed on a periphery of the image, according to one or more embodiments;

[0014] **FIG. 14** is a flow diagram illustrating an example method by which a visual display of an audio-context-tagged image can be presented based on a selected mode of display, according to one or more embodiments;

[0015] **FIG. 15** provides a block diagram representation of an example user equipment configured with various functional components that enable the audio capture and audio context data return features of the disclosure, according to one or more embodiments;

[0016] **FIG. 16** is a flow diagram illustrating a method by which a second user equipment can be configured to allow local audio capture during an image capture event occurring at a first UE, according to one or more embodiments of the disclosure;

[0017] **FIG. 17** is a flow diagram illustrating a method by which a second user equipment initiates audio capture during an image capture event and returns audio context data to a first user equipment performing the image capture event, according to one or more embodiments; and

[0018] **FIG. 18** is a time sequence diagram illustrating the passing of messages and/or signal and data packets between a first user equipment and two second user

equipment during the capturing of an image on the first device within a communication range of the two second devices; and

[0019] FIG. 19 provides another example of an audio-context-tagged image in which all audio recorded during the image capture event, including audio of persons not within the image, are presented concurrently as tags on a single frame of the image, according to one embodiment.

### DETAILED DESCRIPTION

[0020] The illustrative embodiments of the present disclosure provide a method and device that enables audio-context-tagging of a captured electronic image or photograph, where the actual content of what is said by the person in or around the area at the time of image capture is visually associated with the source of the audio output. In one embodiment, the source of the audio output is the person being photographed who spoke before, during, or immediately after the photograph was taken.

[0021] According to a first aspect of the disclosure, a first electronic device performing the various functional aspects related to image capture and tagging includes: an image capturing component; a timer; and a communication mechanism that enables the device to communicate with at least one second device. The electronic device further includes a processor that is communicatively coupled to each of the image capture component, the timer, and the communication mechanism. Additionally, the electronic device includes an image capture utility that executes on the processor and configures the device to: activate the timer to begin tracking a time sequence that extends to an end time following capture of the image; generate an audio capture activation (ACA) signal and transmit the audio capture activation signal via the communication mechanism to the at least one second device; capture the image during the time sequence; receive from the at least one second device content that represents audio that was captured at the at least one second device during the time sequence; link the received content with the image; and store the image along with the linked content as an audio-context-tagged image.

[0022] According to a second aspect of the disclosure, the electronic device or a different viewing device includes a display device, and in response to a selection of the audio-context-tagged image to be displayed on the display device, the image capture utility configures the device to provide, within the display of the image, a visual output of the text content associated with portions of the image as at least one of (a) a time-specific text content and (b) a source-specific text content. The visual output of the source-specific text content identifies the source of the audio and the visual output of the time-specific text content identifies the specific time at which the audio was recorded at the at least one second device relative to the time at which the image was captured by the device.

[0023] According to a third aspect of the disclosure, a second electronic device performing the various functional aspects related to audio capture and transmission includes: a microphone that captures surrounding audio; a timer; a communication mechanism that enables the device to communicate with at least one second device; and a processor that is communicatively coupled to each of the microphone, the timer, and the communication mechanism. The second electronic device further includes an audio capture utility that executes on the processor and configures the device to perform the following list of functions in response to receiving an incoming audio capture activation (ACA) request signal from another device (e.g., the first electronic device): initiate the timer to begin tracking a local time sequence; activate the microphone to begin recording surrounding audio during the local time sequence; store the surrounding audio recorded during the local time sequence; and transmit to the second device an outgoing message packet containing at least one of (a) the surrounding audio recorded and (b) a textual representation of the surrounding audio.

[0024] In the following detailed description of exemplary embodiments of the disclosure, specific exemplary embodiments in which the various aspects of the disclosure may be practiced are described in sufficient detail to enable those skilled in the art to practice the invention, and it is to be understood that other embodiments may be utilized and that logical, architectural, programmatic, mechanical, electrical and other changes may be made without departing from the spirit or scope of the present disclosure. The following detailed description is, therefore, not to be taken in

a limiting sense, and the scope of the present disclosure is defined by the appended claims and equivalents thereof.

**[0025]** Within the descriptions of the different views of the figures, similar elements are provided similar names and reference numerals as those of the previous figure(s). The specific numerals assigned to the elements are provided solely to aid in the description and are not meant to imply any limitations (structural or functional or otherwise) on the described embodiment. It will be appreciated that for simplicity and clarity of illustration, elements illustrated in the figures have not necessarily been drawn to scale. For example, the dimensions of some of the elements are exaggerated relative to other elements.

**[0026]** It is understood that the use of specific component, device and/or parameter names, such as those of the executing utility, logic, and/or firmware described herein, are for example only and not meant to imply any limitations on the described embodiments. The embodiments may thus be described with different nomenclature and/or terminology utilized to describe the components, devices, parameters, methods and/or functions herein, without limitation. References to any specific protocol or proprietary name in describing one or more elements, features or concepts of the embodiments are provided solely as examples of one implementation, and such references do not limit the extension of the claimed embodiments to embodiments in which different element, feature, protocol, or concept names are utilized. Thus, each term utilized herein is to be given its broadest interpretation given the context in which that terms is utilized.

**[0027]** As further described below, implementation of the functional features of the disclosure described herein is provided within processing devices and/or structures and can involve use of a combination of hardware, firmware, as well as several software-level constructs (e.g., program code and/or program instructions and/or pseudo-code) that execute to provide a specific utility for the device or a specific functional logic. The presented figures illustrate both hardware components and software and/or logic components.

[0028] Those of ordinary skill in the art will appreciate that the hardware components and basic configurations depicted in the figures may vary. The illustrative components are not intended to be exhaustive, but rather are representative to highlight essential components that are utilized to implement aspects of the described embodiments. For example, other devices/components may be used in addition to or in place of the hardware and/or firmware depicted. The depicted example is not meant to imply architectural or other limitations with respect to the presently described embodiments and/or the general invention.

[0029] The description of the illustrative embodiments can be read in conjunction with the accompanying figures. It will be appreciated that for simplicity and clarity of illustration, elements illustrated in the figures have not necessarily been drawn to scale. For example, the dimensions of some of the elements are exaggerated relative to other elements. Embodiments incorporating teachings of the present disclosure are shown and described with respect to the figures presented herein.

[0030] When a photograph is being taken, particularly of a group of people, there are typically lots of conversation occurring before, during and/or after the photograph is taken. Some of this conversation actually reflects in the captured image, for example, as a smile, a surprised expression, a look in a different direction, etc. However, with conventional image capture systems, once the image is taken, stored and viewed at a later time, much of the conversation that occurred contemporaneously with the image capture is lost. While a user can manually tag the photographs with comments, the manual process and time required leads most users to avoid adding the information. Also, even if the person later attempts to add the information, the person may have forgotten what actually happened at that earlier time. Additionally, with a large group, the conversations and/or sounds that affect the persons in the image can be localized, such that the photographer and/or another person in the group does not hear and thus cannot re-create what was said.

[0031] As provided herein, general concepts such as tagging of a photograph are presented with the understanding that the term has specific limitations in its conventional usage and refer to a manual process of identifying (or providing a name

or tag to) a person or thing within the image. The present disclosure extends this limited concept to now allow for automatic tagging with what was being said (i.e., of spoken speech) in the general vicinity of the image capture area at the time the image was captured. Thus, rather than present a static image as with conventional tagging, additional contextual data is provided, which yields useful information not previously available from a static image. Thus, as one aspect, the disclosure allows the image to provide information which not only identifies those persons in the image, but also provides the comments made by those people on the visual medium of the image. The disclosure also allows a viewer to know what was being said at the time the image was taken, which allows for a better recollection by the persons in the photograph or by others of the conversation and/or mood and/or activity surrounding the taking of the picture recollect, even after passage of a substantial time after the image was first taken.

**[0032]** As one aspect of the disclosure, an automated capture and tagging process is provided, which eliminates the need for a manual tagging or notation of an image. Accordingly, there is minimal intervention from the user, and the text or audio that gets added as an audio-context tag can automatically double as the tag associated with a specific person or image. According to one aspect, the disclosure enables the real-time, automatic capture of what is being said by the different people when the image is captured and enables that captured information to be represented in the image.

**[0033]** The description of the disclosure provides two different devices, both of which are configured to support different functions, with separate figures illustrating core functional components of the two devices. A first device is presented herein as an image capture device, while a second device is presented as an audio capture device. The devices are described herein as user equipment (UE) to identify that the actual device can be any type of user portable device that supports communication with other devices, and particularly communication via a wireless (i.e., over-the-air) medium. As described below, the image capture device is a master device that initiates both the image capture and the audio capture. The audio capture device is then a secondary device that can, in one embodiment, be triggered to implement the audio capture and content transmission functions in response to receiving activation

signals from the image capture device. The hardware configuration of the two devices illustrate that the two devices can have different functional components, although it is appreciated that a single device can comprise a combination of all of the different components. Descriptions are presented from the perspective of a single one of the two devices. Each of the methods presented by the flow charts relate to specific functionality of a specific one of the devices. For easier flow of the description, the description of the flow charts can be provided out of numerical sequence relative to the figure numbers, and the methods are then described in line with the description of the specific device to which the method functions can be attributed. It is appreciated that the various different functions provided herein can, in one or more embodiments, be functions performed by a single device or functions within a single utility can be selectively performed based on whether the single device is operating as the image capture device or as the audio capture device.

#### **A. SMART PHOTO IMAGE CAPTURE AND GENERATION OF AUDIO-CONTEXT-TAGGED IMAGE**

[0034] Turning now to **FIG. 1**, there is depicted a block diagram representation of an example image capture user equipment (UE) **100**. According to the general illustration, image capture UE **100** is a processing device that is designed to communicate with other devices via one of a wireless communication network, generally represented by base station **140** and antenna **142**, and one or more near field communication (NFC) devices **138**. Image capture UE **100** can be one of a host of different types of devices, including but not limited to, a digital camera, a mobile cellular phone or smart-phone, a laptop, a net-book, an ultra-book, and/or a tablet computing device. These various devices all provide and/or include the necessary hardware and software to enable the capturing of a digital image, such as a photograph. Additionally, image capture UE **100** includes the hardware and software to support both the image capturing functions as well as the wireless or wired communication functions. For simplicity in describing certain of the functional aspects of the disclosure, image capture UE **100** shall be interchangeably referred to

herein as UE1 **100** and/or first device. As will be further appreciated during the description, a second device, namely audio capture UE2 **146** (**FIGs. 1** and **15**) is also introduced herein, and is interchangeably referred to as UE2 **146**, to simplify the description.

[0035] Referring now to the specific component makeup and the associated functionality of the presented components, image capture UE **100** comprises processor integrated circuit (IC) **102**, which connects via a plurality of bus interconnects (illustrated by the bi-directional arrows) to a plurality of functional components of image capture UE **100**. Processor IC **102** can include one or more programmable microprocessors, such as a data processor **104** and a digital signal processor (DSP) **106**, which may both be integrated into a single processing device, in some embodiments. The processor IC **102** controls the communication, image capture, and other functions and/or operations of image capture UE **100**. These functions and/or operations thus include, but are not limited to, application data processing and signal processing.

[0036] Connected to processor IC **102** is memory **108**, which can include volatile memory and/or non-volatile memory. One or more executable applications can be stored within memory for execution by data processor **104** on processor IC **102**. For example, memory **108** is illustrated as containing SmartPhoto client **110**, user interface **114**, and camera controller interface **116**. As shown, SmartPhoto client **110** can include a voice-2-text converter **112**. SmartPhoto client **110** is interchangeably referred to herein as an image capture utility. The associated functionality and/or usage of each of the application software modules will be described in greater detail within the descriptions which follow. In particular, the functionality associated with and/or provided by SmartPhoto client **110** is described in greater details with the description of **FIG. 2** and several of the flow charts and other figures.

[0037] Also shown coupled to processor IC **102** is storage **150** which can be any type of available storage device capable of storing one or more application software and data. It is further appreciated that in one or more alternate embodiments, the device storage can actually be remote storage and not an integral part of the device itself. As

provided, storage **150** contains at least one image record **152**. Image record **152** includes an audio-context-tagged (ACT) image **155**. ACT image **155** includes an actual digital image **154**, such as photographic data, associated text content **156**, and associated time stamp **158**. Also, in one embodiment, ACT image **155** can also include audio content **160**. The specific usage and/or functionality associated with these components are described in greater detail in the following descriptions.

[0038] Image capture UE **100** also comprises one or more input/output devices, including one or more input devices, such as camera **120**, microphone **121**, touch screen and/or touch pad **122**, keypad **123**, and/or one or more output devices, such as display **125**, speaker **126**, and others. Image capture UE **100** can also include a subscriber information module (SIM) **127** which can provide unique identification of the subscriber that owns or utilizes the image capture UE **100**, as well as specific contacts associated with the particular subscriber. In order to allow image capture UE **100** to provide time data, image capture UE **100** also includes system clock **128**.

[0039] According to one aspect of the disclosure and as illustrated by **FIG. 1**, image capture UE **100** supports at least one and potentially many forms of wireless, over-the-air communication, which allows image capture UE to transmit and receive communication with at least one second device. As a device supporting wireless communication, image capture UE **100** can be one of, and be referred to as, a system, device, subscriber unit, subscriber station, mobile station (MS), mobile, mobile device, remote station, remote terminal, user terminal, terminal, communication device, user agent, user device, cellular telephone, a satellite phone, a cordless telephone, a Session Initiation Protocol (SIP) phone, a wireless local loop (WLL) station, a personal digital assistant (PDA), a handheld device having wireless connection capability, a computing device, such as a laptop, tablet, smart phone, personal digital assistant, or other processing devices connected to a wireless modem. To support the wireless communication, image capture UE **100** includes one or more communication components, including transceiver **130** with connected antenna **132**, wireless LAN module **134**, Bluetooth® transceiver **137** and near field communication transceiver module **138**. As further illustrated, image capture UE **100** can also include components for wired communication, such as modem **135** and Ethernet module **136**.

Collectively these wireless and wired components provide a communication means or mechanism **165** by which image capture UE **100** can communicate with other devices and networks.

[0040] The wireless communication can be via a standard wireless network, which includes a network of base stations, illustrated by evolution Node B (eNodeB) **140** and associated base station antenna **142**. A first over-the-air signal **144** is illustrated interconnecting base station antenna **142** with local antenna **132** of image capture UE **100**. Additionally, communication with the at least one second device can be established via near field communication transceiver module **138**. In at least one embodiment, image capture UE **100** can exchange communication with one or more second devices, illustrated as UE2 **146**, UE3 **147**, UE4 **148**, and UE5 **149**. As described in further details below, each of UE2 **146**, UE3 **147**, UE4 **148** are within a pre-established audio capture feedback group **145** that supports audio capture and return of audio context data to image capture UE **100** based on one or more trigger signals communicated from image capture UE **100**. The path of communication between image capture UE **100** and the second devices can be via near field communication or via wireless network **170**, as indicated by the second over-the-air signal **172** between base station antenna **142** and the second devices.

[0041] Turning now to **FIG. 2**, there is illustrated a more detailed diagram of an example SmartPhoto execution environment **200**. SmartPhoto execution environment **200** represents the functional aspects of the SmartPhoto client **110** executing on image capture UE **100** and interacting with specific hardware components of image capture UE **100**. SmartPhoto execution environment **200** consists of storage **150** within which can be maintained a SmartPhoto (SP) image data store **205**, which is a storage location at which SP image records, e.g., image record **152**, are stored. As previously introduced, this storage **150** can be remote from the actual executing device environment, such as cloud storage. Also illustrated within storage **150** is an example of a received message packet **210**, which in one embodiment is received from a second UE within a time period during which or after an image is captured by image capture UE **100**. Received message packet **210** includes a UE identifier (UE-ID) **212** of the second device (e.g., UE2 **146**) from

which the message packet **210** is received. Also included in message packet **210** are text content **214**, a timestamp **216** indicating a time at which the audio was captured, and optionally audio content **218**. It is appreciated that the components within message packet **210** will be associated with or linked to a specific image captured at/by image capture device at a time proximate to the time stamp **216** and that the components would then become the tags associated with an image record (e.g., **152**).

[0042] Also illustrated within storage **150** is contacts **220**, which is a database or list of known persons or second devices or subscribers with which image capture UE **100** can communicate. Each contact is represented by a separate row in contacts **220** and includes a contact ID **222**. Additionally, each contact can have an associated permission data **232**, and optionally photograph or image **242**. Photograph or image **242** is a picture that is associated with the contact. According to one or more embodiments, the photograph or image **242** can be utilized to provide face recognition and association of a person within the captured image and the specific contact within the contact list. In one embodiment, this use of facial recognition and device matching (or device identification) can then be utilized to determine which second devices captured and stored audio corresponding to the particular image and to request or retrieve the audio content captured by a particular second device at a later time. For example, in one implementation, a second device can store the audio content following audio capture, but not immediately communicate the audio content to the image capture UE **100**. When the devices are in communication via NFC or Bluetooth, for example, the second device can move out of communication range of the image capture UE **100** before the second device completes packaging of the audio content for transmission to the image capture UE **100**. When the image is later being analyzed for facial recognition and device matching, the second device can be identified as one associated with the image and which should have captured audio content contemporaneously with the image capture. The audio content can then be pulled from or pushed by the second device at a later time when the two devices are within communication range or via an intermediate server. It is appreciated that not all contacts will have a photograph or image **242** or have SP options **266** set for that contact, and thus not every image capture device can be identified using facial recognition.

[0043] SmartPhoto execution environment **200** also includes SmartPhoto Client **110** and camera controller/interface **116**. As provided by **FIG. 2**, in addition to the components illustrated in **FIG. 1**, SmartPhoto Client **110** can also include SmartPhoto authorized secondary devices list **262**, which can correlate to permission data **232**, and optionally, SmartPhoto authorized primary devices list **264**. SmartPhoto authorized primary devices list **264** is present only in devices that can also operate as a second device that provides audio capture capabilities and then forwards the captured audio to one of the authorized primary devices that requested the audio capture. Rather than present both sets of capabilities within a single device, the disclosure provides a separate description of an audio capture device, which is illustrated by **FIGs. 3 and 4**, described hereafter.

[0044] **FIG. 4** is a flow diagram illustrating a method by which a first user equipment generates and updates a list of authorized second devices to which an audio capture activation (ACA) request can be sent during image capture event occurring at the first UE, according to one or more embodiments. Method **400** is performed at or on image capture UE **100** by execution of specific code segments of image capture utility (or SmartPhoto client **110**) by data processor **104**. Method **400** can thus be described as a process performed by one of image capture UE **100**, data processor **104**, and SmartPhoto client **110**. Method **400** begins at block **402** with data processor **104** opening SmartPhoto client **110** on UE1 **100**. In one implementation, the opening of the SmartPhoto client **100** is triggered by user action. At block **404** UE1 generates and transmits a request to at least one second devices (e.g., UE2 **146**) for UE2 to provide UE1 with authorization to trigger audio capture at UE2 during future image capture events at UE1 that occurs within communication range of UE2. UE1 receives a response to the request at block **406** and determines at decision block **408** whether the response indicates that the audio capture has been authorized by UE2. In one embodiment, a timer function is provided that is triggered when the request is sent out. Then UE1 **100** waits for a pre-set amount of time for a response to the request. When no response is received within that pre-set amount of time, UE1 sets the UE2 response to undetermined. UE1 can then re-transmit the request at a later time to UE2. From decision block **408** if the response indicates that UE2 has rejected the request to permit audio capture, UE1 tags the contact entry of UE2 with a blocked

profile value and adds UE2 to a list of second devices that have blocked audio capture (block **410**). However, when the response indicates that UE2 has authorized audio capture, UE1 **100** adds UE2 to the list of authorized devices for audio capture (i.e., SmartPhoto authorized second devices list **262**) at block **412**. UE2 is then registered within a network of permissive devices. UE1, then stores the updated list of authorized UE2s, and associates the authorization status with the contact list **220** within permission data entry **232** (block **414**). Method **400** then terminates at end block **416**.

[0045] With the above described SmartPhoto execution environment **200** of **FIG. 2** operating within image capture UE **100** of **FIG. 1**, a first aspect of the disclosure provides a device (**100**) comprising: an image capturing component (e.g., camera **120**); a storage **150** within which an image and other data can be stored; a timer (**260**, **FIG. 2**); a communication mechanism (e.g., transceiver **130** or NFC module **138**) that enables the device to communicate with at least one second device (e.g., UE-2 **146**); and a data processor **104** that is communicatively coupled to each of the image capture component, the storage, the timer, and the communication mechanism. The image capture UE **100** further includes an image capture utility, which for purposes of this disclosure can be synonymous with and interchangeably referred to as SmartPhoto client **110**. The image capture utility (**110**) executes on the data processor **104** and configures the device (**100**) to: activate the timer **260** to begin tracking a time sequence that extends to an end time following capture of the image **154**; generate an audio capture activation (ACA) signal and transmit the ACA signal via the communication mechanism to the at least one second device (**146**); capture the image **154** during the time sequence; receive from the at least one second device (**146**) content (e.g., text content **156**) that represents audio that was captured at the at least one second device (**146**) during the time sequence; link the received content (e.g., text content **156**) with the image **154**; and store the image **154** along with the linked content as an audio-context-tagged image. The stored audio-context-tagged image is represented in **FIGs. 1** and **2** as image record **152**. As further illustrated, in at least one embodiment, the received content can include the actual audio content **160**, and the received audio content **160** is stored within the image record **152**. Thus, one or both types of content can be received and stored, in alternate embodiments.

[0046] Returning to FIG. 2, SmartPhoto Client 110 also includes SmartPhoto options settings 266. SmartPhoto options settings 266 represents selectable user and/or device settings for the SmartPhoto client 110. As an example, these settings can include a “transfer ACA requests on shutter opening” setting, which causes the image capture UE 100 to automatically broadcast an ACA request whenever the camera component is activated on the image capture UE 100. Alternatively, this settable option can be set to manual device selection and transfer of ACA requests, where the user of image capture UE 100 can decide one or more of (a) which second devices to transmit the ACA request to and (b) when, before or after activation of the camera, to transmit the ACA request. Another settable option can be related to a timing of transmission of the ACA request, which can be a manual trigger or automatically, following expiration of a pre-set timer, or triggered by detection of the image capture. Permission data 232 can refer to whether or not the particular contact (e.g., the second device associated with the contact’s user ID) has previously sent a response to the image capture UE 100 indicating that audio capture is allowed or authorized for that second device. This setting can be utilized to determine to which second device the ACA request is transmitted among the contact list.

[0047] Thus, according to one or more embodiments, the at least one second device comprises a plurality of second devices. The image capture utility (110) then configures the device (100) to: identify the plurality of second devices within a communication range of the device (100); and transmit the ACA signal to the plurality of second devices within the communication range to trigger the plurality of second devices with audio capture capability to provide audio capture feedback to the device (100). The image capture utility further configures the device (100) to: determine which of the plurality of second devices are configured to provide audio capture feedback to the first device (100). In one or more embodiments, the plurality of second devices that are configured to provide audio capture feedback to the first device are devices that are either (a) pre-configured to be within a permissive network of the first device for audio capture or (b) configured as open devices for audio capture during an image capture event at another device. The utility enables configuration of the first device to only transmit the ACA request to those plurality of

second devices within the communication range that are configured to provide audio feedback.

[0048] In yet another embodiment, the image capture utility further configures the device to: generate a request for one or more of the plurality of second devices to provide authorization to be included within the permissive network; transmit the request to the one or more of the plurality of second devices via a communication means; register within the permissive network those plurality of second devices that provide an affirmative response to the request; store the permissive network file within the storage; and access the permissive network file prior to transmitting the ACA signal to those plurality of second devices registered within the permissive network.

[0049] FIG. 3 illustrates an example implementation scenario in which a photographic image 320 is taken of a plurality of subjects. In FIG. 3, a photographer 305 is shown taking a photographic image 320 with an image capture UE 100, which is configured with SP client 110. Included in the photographic image 320 are several individuals who have personal devices that can record audio. These personal devices are indicated as being UE2 146, UE3 147, UE4 148, UE5 149, and UE6 340. Each of UE2 146, UE3 147, UE4 148, and UE5 149 are within communication range of image capture UE 100. Accordingly, image capture UE 100 is able to communicate with each second device via NFC module 138 or Bluetooth® 137 or some other communication mechanism. In one embodiment, communication can be via wireless communication network 170 (FIG. 1). Of the second devices of the individuals being captured within the photographic image 320, UE2 146, UE3 147, and UE4 148 are within a permissive group 145 that allows locally-captured audio content to be forwarded to one of (a) image capture UE 100 or (b) a secondary computer system at which the photographic image 320 can be tagged with audio content. In embodiments in which image capture UE 100 is communicating via near field communication or Bluetooth®, UE6 340 represents a second device that is out of range of the image capture UE 100 and thus cannot be triggered to initiate audio capture and return of audio content. UE5 149 is within communication range to receive communication from image capture device 100 but is at least one of (a) not configured for audio

capture, (b) not a part of the permissive group accessible to image capture UE **100**, or (c) configured with a block against audio capture. In one or more embodiments, a block against audio capture can be a global block, a location specific block for one or more specific locations, a time-specific block that prevents audio capture at certain times of the day, or some other variation of a partial block against audio capture.

**[0050] FIG. 5** is a flow diagram illustrating a general method **500** by which a first user equipment (UE1 **100**) performs image capture and creates an audio-context-tagged image by triggering one or more second UEs (**146**) to provide speech audio feedback data detected during the image capture event, according to one or more embodiments. Method **500** begins at block **502** at which data processor **104** detects activation of camera **120** and/or selection of camera function of UE1 **100**. At block **504**, in response to user input, data processor **104** opens SmartPhoto client **110** in the background while displaying the camera viewer. In response to activation of the camera and/or concurrently with opening the SmartPhoto client **110** and/or based on a detected user activation of a selectable audio capture timer feature of SmartPhoto client **110**, method **500** includes UE1 **100** activating a timer to begin tracking a time sequence that extends to an end time following capture of the image (block **506**). Method **500** further includes UE1 **100** generating an audio capture activation (ACA) signal and transmitting the ACA signal via a communication mechanism to at least one second device (block **508**). In one embodiment in which UE1 **100** utilizes near field communication, UE1 transmits or broadcasts the ACA request signal into the surrounding space. In one or more alternate embodiments, UE1 **100** can transmit the ACA to a specific set of UE2s within the authorized list of second devices **262** that are detected within communication range of UE1 **100**. In yet another embodiment, the ACA is transmitted via a wireless communication network to the specific UE2, which can, in one implementation, be detected using near field communication to scan the specific area surrounding UE1 **100**. In yet another embodiment, the ACA request can include location information and is broadcasted to all UE2s within the authorized list **262**. However, only those UE2s that are within a specific distance (e.g., 25 feet) of the location of UE1 **100** will be triggered to record audio.

[0051] Thus, in one embodiment, transmitting of the ACA signal includes: identifying at least one second device within a communication range of the first device; determining which of the at least one second device are configured to provide audio capture feedback to the first device. In this context, the second devices that are configured to provide audio capture feedback to the first device comprise (a) second devices that are pre-configured to be within a permissive network of the first device for audio capture and (b) second devices that are configured as open devices for audio capture during an image capture event occurring at another device. Then, method **500** provides transmitting the ACA signal to a group of second devices to trigger those second devices with audio capture capability to provide audio capture feedback to UE1 **100** (step **508**). According to various alternate embodiments, the group of second devices can be one of (1) a first group including each of the at least one second device within the communication range and (2) a second group including only those second devices within the communication range that are configured to provide audio feedback and are one of (a) open devices and (b) second devices within a permissive network of the first device. The ACA request is only transmitted to those second devices that are configured and authorized to provide audio feedback to UE1 **100**.

[0052] Returning to the flow chart, method **500** includes UE1 capturing the image during the time sequence and specifically detecting camera **120** capturing the image (**510**). Method **500** then provides that UE1 monitors the timer for a preset end of audio capture period or receipt of an end input (block **512**). In one embodiment, the preset end of audio capture period can be a set number of seconds after the image capture is completed. In yet another embodiment, the end input can be the actual completion of the image capture or a manually entered selection of an end feature/function provided on UE1 **100**. At decision block **514**, method **500** includes determining if the timer has expired or an input received to end the audio capture, and UE1 **100** continues to monitor for receipt of the end signal or the expiration of the timer, whichever occurs first. In response to the expiration of the timer or receipt of an input to end the audio capturing, method **500** provides at block **516** that UE1 **100** broadcasts the TAC signal into the surrounding area or transmits the TAC signal to the specific second devices in the area.

[0053] At decision block **518**, method **500** includes determining whether one or more message packets have been received from one or more UE2(s) **146** within a specific time following the image capture. The timing of the receipt of the message packets allows for sufficient time for any UE2 that is activated and pre-configured as an audio capture device to provide audio content as feedback to the ACA and/or TAC signals, in one embodiment. In alternate embodiments, the audio content can be received at a later time or be transmitted to a separate server or computing device at which the audio context tagging of the captured image is completed. In response to not receiving a message packet containing audio context data, method **500** includes UE1 **100** storing the image without any audio context tag, i.e., as a regular image (block **520**). However, in response to receiving from the at least one second device a message packet containing audio context data, method **500** includes retrieving the audio context data from the message packet. In one or more embodiments, the audio context data can include one or more of an audio file, text content, and a timestamp (time data) (block **522**). The text content represents text converted from the audio that was captured at the at least one second device during the time sequence of the image capture event.

[0054] With the audio context data retrieved, method **500** further provides UE1 **100** identifying the UE2s **146** from which the message packets **210** were received (block **524**). In at least one embodiment, this identifying includes retrieving device information from the message packet. In another embodiment, this identifying further includes performing a scan of the image for facial recognition to determine who the speaker of the received audio context data was. In yet another embodiment, which can include aspects of the other embodiments, the identifying can also be completed using location based identification of a UE2 relative to the location captured within the image. Use of global positioning system (GPS) coordinates and other location based technologies can be implemented within this latter embodiment. At block **526**, method includes linking with the image the received text content and time data (and optionally the audio data, if received in the message packet). Specifically, the linking involves associating the text content and/or timestamp and/or audio data with the identified source of the specific audio, where the source of the audio can be determined via one or the aforementioned methodologies, including second device

identification and association with a known user and facial recognition of the persons presented within the image. At block **528**, method **500** includes storing the image along with the linked text content as an audio-context-tagged image. In one or more embodiments, the linking and tagging processes are completed at a remote computing device or server. Accordingly, in this and other embodiments, the process of linking or tagging the image with additional audio context data can be completed offline so that users of UE1 **100** can continue to capture additional images, rather than having to wait for the return of audio context data and subsequent tagging of the image in real time.

[0055] As introduced, the message packet can provide time data as a part of or in conjunction with the audio context data. According to one embodiment, the method of linking the text content with the image includes: recording a time at which the image is captured; identifying, from data included within the message packet, a specific time within the time sequence during which the audio was captured at the at least one second device; identifying, within the image, a source of the audio represented within the audio context data (e.g., audio file or text content converted from the captured audio at UE2 **146**); and associating the content and the specific time with the source to provide linked text content.

## **B. SMART PHOTO DISPLAY WITH AUDIO CONTEXT TAGS**

[0056] The first aspect of the disclosure described above covers systems and methods for capturing an image and retrieving audio context data from second devices that can then be utilized to generate an audio-context-tagged image. A second aspect of the disclosure entails the process of actually displaying the image as an audio-context-tagged image on a display device, where the audio context tags can be made visible by superimposing audio context information on the image display. Multiple different embodiments are provided herein, as illustrated by **FIGs. 6 – 13**.

[0057] As shown by the combination of **FIGs. 3** and **6**, several of the persons within the photographic image **320** are speaking at some time before, during, or after the image capture event. A verbal utterance or speech is illustrated in **FIG. 3** using first speech image **315** originating from photographer **305** and second speech image **330**

originating from a bystander **325**. Notably, bystander **325** is shown with a second device, UE7 **335**, which is also in communication range of image capture UE **100**. **FIG. 6** then illustrates an example audio context-tagged image in which audio recorded during the image capture event are presented concurrently as tags on a single frame of the displayed image, according to one embodiment. Specifically, **FIG. 6** presents the audio-context-tagged image **600** with audio-context tags superimposed on the original image **320** that was captured by image capture UE **100** in **FIG. 3**. The audio-context tags include participant tags **602**, **604**, **606**, and **608** associated with the persons whose images are captured within the image **320**. In **FIG. 6**, recorded audio content is presented for all speakers simultaneously. Thus, concurrently displayed are text bubble **602** associated with and containing spoken audio **620** of speaker Krish, text bubbles **604** and **606** associated with and containing spoken audio **620** of speaker Aman, and text bubble **608** associated with and containing spoken audio **620** of speaker Rohit.

[0058] Importantly, when the amount of audio captured is such that the text representation would obscure the entire image, the SmartPhoto image viewing client can default to a mouse over view option or a timeline selection view option to prevent the image from being simply a display of text. Alternatively or in addition, the text content can be located and shown off to the side or periphery of the image or completely separated from the image in a text box that can be scrolled through to read the entire text content.

[0059] As further shown by image **600**, in one embodiment the function of providing the visual output of the time-specific and source-specific text content includes the utility configuring the device to: display the audio-context-tagged image; and associate and display a timeline **605** along with the display of the audio-context-tagged image **600**. The timeline **605** represents a visual depiction of the time sequence, including a time (T) of image capture. The timeline **605** can be superimposed over the image **600**, appended to an end of the image, or displayed proximate to the image but as a separate selectable item. In one embodiment, the timeline is always provided, and the viewer is able to select one of the times and receive a corresponding feedback (e.g., highlighting of a text bubble) of one or more

of the displayed text bubbles to show which specific audio content is associate with the selected time.

[0060] As further illustrated by **FIG. 19**, the audio-context-tagged image can also provide or display the audio context tags of non-participants **1912**, **1914**, **1916**, and **1918** corresponding to the photographer and other bystanders, whose images are not present within the captured photographic image **320 (FIG. 3)**. According to one embodiment, with this capability enabled, the utility (**110**) further configures the device (**100**) to: determine when the source of the audio is not captured within the image; and present an output of the text content along a periphery of the displayed image **1900** along with an identifier of the source in alphanumeric characters to indicate a presence of the source within a surrounding space of the area captured within the displayed image **1900**.

[0061] Notably, included in each tag in this particular representation of audio-context-tagged image **600 (FIG. 6)** and **1900 (FIG. 19)**, are actual text content as well as a timestamp, and an identifier of the audio source. It is appreciated that alternate embodiments can provide a different rendering of the audio content, including with or without the timestamp and/or the audio source identifier. With the participant tags, for example, the audio content can be presented with a physical connection extending to, or some other visual identifier of the source of the particular audio content. Also, alternate embodiments provide that the actual audio recording be played, with the tags not presenting any text based content. Additionally, within the various illustrated and described embodiments, the timestamp is represented by a relative time  $T+x$  or  $T-x$ , where  $T$  is the time of image capture and  $x$  is some offset amount of time ranging from zero (0) to some larger number (e.g., 10 seconds). In alternate embodiments, the actual time can be utilized as the timestamp, where each second device is assumed to be synchronized with image capture UE **100**. Thus, for example, in **FIG. 6**, time  $T$  can represent 11:05.30 AM,  $T-4$  represents 11.05.26AM, and  $T+3$  represents 11:05.33AM. According to one embodiment, audio-context-tagged image **600** can also include a timeline **605** which identifies the time sequence from before image capture (i.e., pre-capture) until after image capture (i.e., post capture). Similarly, audio-context-tagged image **1900** also includes a timeline **1905**. With the above

example of actual time, timeline **605** represents a seven (7) second interval running from 4 seconds before image capture until 3 seconds after image capture. In the described embodiments, the beginning and end time of the time interval represented by timeline **605** can correlate to or be representative of the time of transmission from image capture UE **100** and/or receipt at second device(s) of the ACA and TAC signals.

[0062] According to one embodiment, the function of linking the content with the image includes the utility (**110**) further configuring the device (**100**) to: record a time (*T*) **158** at which the image is captured; identify, from data included within the message packet **210**, a specific time within the time sequence (provided by the timeline **605**) during which the audio was captured at the at least one second device (**146**); identify, within the image **320** a source of the audio that is represented by the received content; and associate the content and the specific time (**158**) with the source to provide the linked content.

[0063] In one embodiment, the received content includes text content **214** converted from the audio that was captured, and the device further includes: a display device **125** communicatively coupled to the data processor **104**. In response to a selection of the audio-context-tagged image to be displayed on the display device **125**, the image capture utility (**110**) further configures the device (**110**) to provide, within the display of the image **320**, a visual output of the text content **214** associated with portions of the image **320** as at least one of (a) a time-specific text content and (b) a source-specific text content. The visual output of the source-specific text content identifies the person speaking as the source of the audio. The visual output of the time-specific text content identifies the specific time at which the audio was recorded at the at least one second device (**146**) relative to the time at which the image **320** was captured by the device, as indicated by the timeline or timestamp.

[0064] According to one embodiment, the utility (**110**) further configures the device (**100**) to transmit a terminate audio capture (TAC) signal via the communication mechanism. The TAC signal is transmitted after one of (a) a pre-set length of time and (b) a received input at the device (**100**) following the capture of the image. The TAC

signal notifies the at least one second device (146) to stop the audio capture. With this embodiment, the end time of the time sequence or timeline corresponds to the transmission and/or receipt of the TAC signal.

[0065] **FIGs. 7 – 8, 9, and 10-13** illustrate a plurality of different methodologies for displaying audio-context tagged images, and specifically audio-context tagging of image 320. Each of the different methodologies provides a different manner for displaying the audio context tag with or within the example image 320, yielding different forms of audio context tagged image, each represented by a different reference numeral. Generally, **FIGs. 6 and 19** provide two examples of an audio-context-tagged image in which all audio recorded during the image capture event, including audio of persons not within the image, are presented concurrently as tags on a single frame of the image, according to multiple embodiments. **FIGs. 7-8** provide two different views or frames of the captured image with source-specific and time-specific audio context tagging that identifies spoken speech during the image capture event as pop up speech content, according to several embodiments; **FIG. 9** then provides another view of the image with the source-specific audio context tagging that identifies the spoken speech during the image capture event as text boxes placed on a periphery of the image; **FIGs. 10-13** provide two different views or frames of the image with time-specific audio context tagging that identifies spoken speech at different times during the image capture event as pop up speech content or a text box, according to several embodiments.

[0066] According to a first embodiment and as represented by **FIGs. 7 and 8**, providing the visual output of the time-specific and source-specific text content comprises the utility configuring the device to: display the image without any of the text content; detect a selection within the image of one source of captured audio that corresponds to a text content linked to the audio-context-tagged image; and in response to detecting the selection, display the specific text content via a text overlay on the image. The utility further configures the device to display the text overlay for one of (a) at least a specific period and (b) an entire period, while the source of the captured audio remains selected. And, in at least one embodiment, the utility further configures the device to concurrently indicate within the visual output a time at which

the audio corresponding to the text content was recorded at the at least one second device.

[0067] As shown by **FIGs. 7 and 8**, while the image **320** is presented on the display device, the utility (**110**) allows a user to mouse over the individual images of the persons within the image **320**. The mouse over occurs as the user moves the mouse pointer **705** across the surface of the moused-over image **700A**, **700B**. When the mouse pointer **705** lands on a specific person **710** who voiced audio content during the time of image capture, a pop up text bubble **715** appears on the display screen near to the person to which the text content is associated. Thus in first mouse-over image **700A (FIG. 7)**, in response to detecting placement of the mouse pointer **705** over Krish, the utility (**110**) generates text bubble **715**, within which text content **720** is presented. In the illustrated embodiment, a timestamp **725** is also provided along with the text content, and in one implementation, the name of the speaker can be indicated as well. Similarly, in second mouse-over image **700B (FIG. 8)**, in response to detecting placement of the mouse pointer **705** over Rohit (**810**), the utility (**110**) generates text bubble **815**, within which text content **820** is presented along with timestamp **825** and the identifier of the speaker. In each example, the text bubble **715** and **815** are located proximate to the identified speaker, such that a recognition of who has spoken the specific words can be determined by simply looking at the placement of the text bubble **715/815**. In more advanced implementations, the utility can determine a best positioning of the text bubble based on the location of people and objects within the photograph.

[0068] According to one embodiment, the device function of providing the visual output of the time-specific and source-specific text content includes the utility configuring the device to concurrently display one or more text content within a single-frame output of the audio-context-tagged image identifying the source of the corresponding audio, with each of the one or more text content being concurrently displayed as an overlay on the image along with a timestamp associated with that text content. **FIG. 9** provides one implementation of concurrent display of multiple audio-context tags on a single image **320**. In this example, Aman has spoken multiple times during the picture taking event. Rather than present text bubbles, which would fill up

a large space in the middle of the mouse-over image **900**, utility (**110**), in response to detecting the mouse over on the image of Aman **910**, generates a plurality of periphery text boxes **915A**, **915B**, and **915C**. Each periphery text box **915A**, **915B**, and **915C** respectively contains text content **920A**, **920B**, and **920C** along with the associated timestamp **925A**, **925B**, and **925C** and identifier of Aman as the speaker.

[0069] **FIGs. 10-13** illustrate a sequence of frames in which the image **1000** is presented for viewing with an associated timeline **1005** providing a movable, select-and-drag time bar **1010**. The functionality presented by the figures are generally described with reference to **FIG. 10**, which represents the first in the series of figures. In at least one implementation, the utility configures the device to provide a moveable time bar **1010** associated with the timeline **1005** that can be selectively moved along the length of the timeline **1005**. The specific period of time is selected when the time bar **1010** is moved to the corresponding point on the timeline **1005** and the text content **1020** is only presented within the visual output of the audio-context-tagged image **1000** when the time bar **1010** is located at that point on the timeline **1005**. Movement of the time bar **1010** can be completed using a mouse pointer **1025**, via one of a select-drag-release operation and simply clicking within a desired location within the timeline **1005**, in one or more embodiments.

[0070] With these embodiments, the image can be initially illustrated without any text bubbles or other audio content displayed depending on a start location of the time bar **1010**. In **FIG. 10**, time bar **1010** is at time T-4 at which time Aman has spoken. Thus, text bubble **1015** pops up on the display screen overlaying a part of the image **1000** and presenting specific text content **1020** of audio spoken by Aman at time T-4 before image capture. According to the embodiment, the utility includes code that enables the device to: select a specific time period on the time line **1005**; and in response to the specific time period selected coinciding with a time of capture of audio, where the captured audio was converted into corresponding text content that is linked to the image, display the corresponding text content (e.g., text content **1020**) as an overlay on the image **1000**, identifying the source of the audio.

[0071] **FIGs. 11-13** similarly present frames of image **1000** at different times during image capture. The specific times are identified by the location of the time bar **1010** which is being moved or scrolled along timeline **1005** under the control of mouse pointer **1025**. In **FIG. 11**, time bar **1010** is placed at time T-2 along timeline **1005**, at which time text bubble **1115** is displayed superimposed on image **1000** with text content **1120** attributed to Krish at time T-2. In **FIG. 12**, time bar **1010** is placed at time T-1 along timeline **1005**, at which time text bubble **1215** is displayed superimposed on image **1000** with text content **1220** attributed to Rohit at time T-1. Finally, in **FIG. 13**, time bar **1010** is placed at time T+1 along timeline **1005**, at which time text box **1315** is displayed superimposed on image **1000** with text content **1320** attributed to Aman at time T+1. In **FIG.13**, the use of text box **1315** located at a periphery of the displayed image **1300** illustrates one or more alternate embodiments of a text bubble located proximate to the actual speaker.

[0072] **FIG. 14** is a flow diagram illustrating an example method by which one of three or four alternative viewing options can be selected for viewing the audio context tags on or with the image. The visual display of an audio-context-tagged image can be presented based on a selected mode of display, according to one or more embodiments. It is appreciated that the image can be displayed either on the display of UE1 **100** or on a display of a secondary device on which the image is loaded, such as a personal computer, tablet, or server computer. Because the audio context tag is embedded in the digital image, the tags can be available for viewing in any one of a plurality of different viewing devices. The described method **1400** can be performed by any electronic device (such as and including UE1 **100** and server **370**) having a processor that is communicatively coupled to a display device and which executes an image/photo viewer application on the device. For simplicity in describing the method, the executing processor is assumed to be data processor **104** of UE1 **100**. Referring to the flow chart, method **1400** begins at start block and proceeds to block **1402** with data processor **104** detecting the selection for opening of an image for visual display on the display device. Method **1400** then includes the data processor **104** retrieving the image and reading the meta-tags associated with the image (block **1404**). And, responsive to detecting that the selected image is an audio-context-tagged image, method **1400** includes data processor **104** opening the SP viewing

client in the background and determining a display setting of the SP viewing client (decision blocks **1406**, **1410**, and **1414**). It is appreciated that the SP viewing client can be a software module that can be an integral part of or separate from the image/photo viewing application being utilized to view the image. SP viewing client can, in one implementation, be an add-on utility to existing image/photo viewing applications. For example, SP viewing client can be a downloaded upgrade to a legacy image/photo viewing application.

[0073] Generally, method **1400** includes, in response to selection of the audio-context-tagged image to display on a display device, providing, within the display of the image, a visual output of the audio context data as text content associated with portions of the displayed image. The audio-context-tagged image can be displayed as at least one of (a) a time-specific text content and (b) a source-specific text content, where the visual output of the source-specific text content identifies the source of the audio and the visual output of the time-specific text content identifies the specific time at which the audio was recorded at the at least one UE2 **146** relative to the time at which the image was captured by UE1 **100**. More specifically, however, method **1400** includes data processor **104** performing a series of determinations including: (1) determining at block **1406** whether the display mode setting for the audio-context-tagged image is that of a fully tagged content display mode in a single frame, such as illustrated by **FIG. 6** and/or **FIG. 19**; (2) determining at block **1410** whether the display mode setting for the audio-context-tagged image is that of a timeline view display mode in multiple frames, such as illustrated by **FIGs. 10-13**; and/or (3) determining at block **1414** whether the display mode setting for the audio-context-tagged image is that of a mouse over content display mode in multiple frames, such as illustrated by **FIGs. 7-9**. It is appreciated that various hybrids or combinations of these display modes can be implemented, without limitation.

[0074] In one embodiment, the fully tagged content display mode is the default mode of the SP viewing client. Also, in at least one implementation, an audio-context-tagged image that is displayed using an image/photo display application that does not include the SP image viewing client, can automatically be displayed in the fully tagged mode in a single frame or provide the audio context tags within the properties

or metadata that are visible when the user opens the properties option on the displayed photo on the stored image file. As provided by **FIGs. 6 and 19**, one or more embodiments provide the concurrent display of the timeline of audio capture as a part of the display of the audio-context-tagged image. In response to the display mode being the fully tagged content display mode, method **1400** includes processor **100** providing the visual output of the time-specific and source-specific text content by concurrent displaying of one or more text content within a single-frame output of the audio-context-tagged image (block **1408**). The displayed one or more text content identifies the source of the corresponding audio. Each of the one or more text content is concurrently displayed as an overlay on the image along with a timestamp associated with that text content.

**[0075]** As generally presented by block **1412**, in response to the display mode being the timeline view mode, method **1400** includes processor **100** providing the visual output of the time-specific and source-specific text content by: displaying the audio-context-tagged image; associating and displaying a timeline along with the display of the audio-context-tagged image, wherein the timeline represents a visual depiction of the time sequence, including a time of image capture; enabling selection of a specific time period on the time line; and in response to the specific time period selected coinciding with a time of capture of an audio, which was converted into corresponding text content that is linked to the image, displaying the corresponding text content as an overlay on the image, identifying the source of the audio.

**[0076]** As generally presented by block **1416**, in response to the display mode being the mouse over content display mode, method **1400** includes processor **100** providing the visual output of the time-specific and source-specific text content by: initially displaying the image from the audio-context-tagged image without any of the text content; detecting a selection within the image of one source of captured audio that corresponds to a specific text content linked to the audio-context-tagged image; in response to detecting the selection, displaying the specific text content via a text overlay on the image, for one of (a) an initial period and (b) an entire period, while the source of the captured audio remains selected; and concurrently indicate within

the visual output a time at which the audio corresponding to the text content was recorded at the at least one second device.

[0077] In response to there not being a SP image viewing client, or the SP viewing client being set to not display any audio context tags within the displayed image, method **1400** includes data processor **104** displaying the image without any audio context tagging or associated features, such as the timeline, visible (block **1418**). Method **1400** further includes enabling toggling of the display modes of the SP image viewing client to change views of the visual display of the image while the image is being displayed (block **1420**). It is appreciated that this option involves presenting a user interface or selectable affordance that allows for selection of one of the display modes supported, such as via a drop down menu item within the image viewing application.

### C. SMART PHOTO AUDIO CAPTURE DEVICE AND FEATURES

[0078] The above description introduces the concepts of one or more second devices that operate to capture audio, including voiced speech, during an image capture event occurring at the image capture UE **100**. In its simplest form, the second device can be an audio recording device that has some means of communicating the detected and stored audio input to another device. The other device can include, but is not limited to one or more of the UE1 **100** from which a trigger signal (e.g., an ACA signal) is received and a server or computer device that executes a SmartPhoto client application capable of linking audio context data to an image. The latter example is described in this section of the disclosure. The recorded audio is locally stored as an audio file, and one or more of the audio file, a text file generated from the audio file, and a timestamp can be communicated to the other device.

[0079] According to one or more aspects of the disclosure, the second devices can be an audio capture UE2 **146**, as provided by **FIG. 15**, which is now described. Audio capture UE2 **146** can be any device that provides both audio capture capabilities and the ability to communicate the captured audio to a next device. In one implementation, the same device can serve as both an image capture and an audio capture UE. This implementation is possible because the functions can be

programmed into a single SmartPhoto client having separate executable modules or be provided as fully functional independent utilities that can be selectively executed on a common processor. The processor can then independently execute the particular module or utility based on whether the device is operating as UE1 or UE2 during an image capture event. The device can even support simultaneous or concurrent, overlapping execution of both utilities, whereby a single device that is being used to capture an image can also be providing audio capture functions in the background based on a trigger received from a next device. Because of the similarities possible in the hardware configuration, the description of **FIG. 15** mirrors that of **FIG. 1** in some parts, and differs only where hardware or functional differences exist between the depiction and/or usage of the two illustrated devices.

[0080] UE2 **146** comprises processor integrated circuit (IC) **1502**, which connects via a plurality of bus interconnects (illustrated by the bi-directional arrows) to a plurality of functional components of UE2 **146**. Processor IC **1502** can include one or more programmable microprocessors, such as a data processor **1504** and a digital signal processor (DSP) **1506**, which may both be integrated into a single processing device, in some embodiments. The processor IC **1502** controls the communication, audio capture, and other functions and/or operations of UE2 **146**. These functions and/or operations thus include, but are not limited to, application data processing and signal processing.

[0081] Connected to processor IC **1502** is memory **1508**, which can include volatile memory and/or non-volatile memory. One or more executable applications can be stored within memory for execution by data processor **1504** on processor IC **1502**. For example, memory **1508** is illustrated as containing SmartPhoto audio capture client **1510**. SmartPhoto audio capture client **1510** is interchangeably referred to herein as an audio capture utility. As shown, SmartPhoto audio capture client **1510** can include a voice-to-text converter **1512**, timer **1514**, message packet generating utility **1516**, audio compression utility **1517**, audio capture authorization and/or blocking options module **1518**, and a user interface **1519**. The associated functionality and/or usage of each of the application software modules will be described in greater detail within the descriptions which follow. In particular, the

functionality associated with and/or provided by SmartPhoto audio capture client **1510** is described in greater details with the following description of several of the flow charts and other figures.

[0082] Also coupled to processor IC **1502** is storage **1550** which can be any type of available storage device capable of storing one or more application software and data. As with the previous references to storage, it is appreciated that the term storage is not limited to a physical component that is within or even physically attached to UE2 **146**. Rather, storage **1550** can be remote storage, such as cloud storage, in one or more embodiments. As provided, storage **1550** contains at least one audio record **1552**. Audio record **1552** includes an audio file **1554** of captured audio and optionally a text content file **1556**, which is the audio content converted into text via voice-2-text converter **1512**. Audio record **1552** also includes associated time data **1558**, such as a timestamp, which is the time at which the audio record **1552** was recorded. Also, in one embodiment, audio record **1552** can also include the target device identifier (ID) **1560**. In the illustrated embodiment, the target device is the next device to which the message packet with audio context data is to be transmitted. Notably as introduced above, the target device ID **1560** can be that of UE1 **100**. However, alternate embodiments can provide a different target device and associated target device ID **1560** than one corresponding to or identifying UE1 **100**. Also, in at least one embodiment, audio record **1552** can also include the ID (not shown) of UE2 **146** to allow for quicker determination of the source of the audio file or other audio context data being transmitted via the message packets generated by message packet generating utility **1516**. The specific usage and/or functionality associated with these components are described in greater detail in the following descriptions.

[0083] UE2 **146** also comprises one or more input/output devices, including one or more input devices, such as microphone **1520**, touch screen and/or touch pad **1522**, keypad **1524**, and/or one or more output devices, such as display **1525**, and others. UE2 **146** can also include a user ID information module **1526** which can provide unique identification of the subscriber that owns or utilizes the UE2 **146**, as well as specific contacts associated with the particular subscriber. In order to allow UE2 **146** to provide time data **1558**, UE2 **146** also includes system clock **1528**.

[0084] According to one aspect of the disclosure and as illustrated by **FIG. 15**, UE2 **146** can support at least one and potentially many forms of wireless, over-the-air communication, which allows UE2 to transmit and receive communication with at least one other device. As a device supporting wireless communication, UE2 **146** can be one of and be referred to as a system, device, subscriber unit, subscriber station, mobile station (MS), mobile, mobile device, remote station, remote terminal, user terminal, terminal, communication device, user agent, user device, cellular telephone, a satellite phone, a cordless telephone, a Session Initiation Protocol (SIP) phone, a wireless local loop (WLL) station, a personal digital assistant (PDA), a handheld device having wireless connection capability, a computing device, such as a laptop, tablet, smart phone, personal digital assistant, or other processing devices connected to a wireless modem. To support the wireless communication, UE2 **146** includes one or more communication components, including transceiver **1530** with connected antenna **1532**, wireless LAN module **1534**, Bluetooth® transceiver **1537** and near field communication transceiver module **1538**. As further illustrated, UE2 **146** can also include components for wired communication, such as modem **1535** and Ethernet module **1536**. The wireless communication can be via a standard wireless network **1570**, which includes a network of base stations, illustrated by evolution Node B (eNodeB) **1564** and associated base station antenna **1566**. A first over-the-air signal **1562** is illustrated interconnecting base station antenna **1566** with local antenna **1532** of UE2 **146**. Additionally, communication with the at least one other device (e.g., UE1 **100**) can be established via near field communication transceiver module **1538**. In at least one embodiment, UE2 **146** can exchange communication with one or more other devices, generally represented by UE1 **100** and audio context tagging server **1575**. As described herein, UE1 **100** can be within a pre-established group that is permitted to trigger audio capture at and receive audio context data in return from UE2 **146**. Audio context tagging server **1575** can be any computing device having the processing capability, software modules, and communication mechanisms to support receipt of an image from UE1 **100**, receipt of corresponding audio context data from UE2 **146**, and subsequent tagging of the received image with audio context data to generate an audio-context-tagged image. The path of communication between UE2

**146** and the other devices represented by UE1 **100** can be via near field communication or via wireless network **1570**, as indicated by the second over-the-air signal **1572** between base station antenna **1566** and the other devices, or via wired connection, as illustrated by the dashed connection to the audio context tagging server **1575**.

[0085] The above description generally provides an audio capture device (UE2 **146**) having: a microphone **1520** that captures surrounding audio; a timer **1514**; a communication mechanism **1540** that enables the device to communicate with at least one other device; and a data processor **1504** that is communicatively coupled to each of the microphone **1520**, the timer **1514**, and the communication mechanism **1540**. The audio capture device further includes an audio capture utility **1510** that executes on the data processor **1504** and configures the device to perform specific functions, in response to receiving an incoming audio capture activation (ACA) request signal from a first device. Specifically, the utility triggers the audio capture device to: initiate the timer to begin tracking a local time sequence; activate the microphone to begin recording surrounding audio during the local time sequence; store the surrounding audio recorded during the local time sequence; and transmit to the other device an outgoing message packet containing at least one of (a) the surrounding audio recorded and (b) a textual representation of the surrounding audio. The audio capture utility further configures the device to: turn off the microphone at one of (a) a preset end time of the local time sequence and (b) receipt of an incoming terminate audio capture (TAC) signal from the first device.

[0086] In one embodiment, the device includes a voice-2-text converter **1512** executing on the processor **1502**, and the audio capture utility further configures the device to: convert into text content the recorded surrounding speech audio; package the text content into an outgoing message packet tagged with a device identifier (ID) that identifies the device UE2 **146**; and forward the outgoing message packet with the text content to the other device (UE1 **100** or server **1575**). In one implementation, the outgoing message packet is forwarded to the UE1 **100** from which the incoming ACA request was received.

[0087] In an alternate embodiment, the audio capture device includes an audio compression component **1517** executing on the data processor **1504**, and wherein the audio capture utility further configures the device to: compress the surrounding audio that was recorded to generate a compressed audio file; create an outgoing message packet tagged with a device identifier (ID) that identifies the device, UE2 **146**; and insert the compressed audio file as additional payload within the outgoing message packet. Further, in one embodiment, the audio capture utility further configures the device to: insert the text content as additional content into the outgoing message packet; and transmit the outgoing message packet with both the text content and the compressed audio file to the other device (e.g., UE1 **100** or server **1575**). According to one aspect of the implementation, transmitting the outgoing message packet comprises the audio capture utility further configuring the device to transmit the outgoing message packet following one of (a) a preset end time of the local time sequence and (b) receipt of an incoming terminate audio capture (TAC) signal from the first device. The recorded audio can then be used to tag the captured image at the first device. It is appreciated that the first device referenced herein and which transmits the ACA and TAC signals is also the image capture device or UE1 **100**.

[0088] According to another embodiment, involving use of a server or other intermediate computing device, the audio capture utility configures the audio capture device to, in response to receiving an incoming audio capture activation (ACA) request signal from a first device: initiate the timer to begin tracking a local time sequence; activate the microphone to begin recording surrounding audio during the local time sequence; store the surrounding audio recorded during the local time sequence; and transmit to the intermediate server an outgoing message packet containing at least one of (a) the surrounding audio recorded and (b) a textual representation of the surrounding audio. According to this embodiment, transmitting the outgoing message packet comprises the audio capture utility further configuring the audio capture device to transmit the outgoing message packet to an intermediate server (e.g., server **1575**). Server **1575** is configured with a processor, memory and/or storage, and includes executable program code that enables the server to perform audio source identification via facial recognition. The server **1575** also tags images

received from the first device (UE1 **100**) with the surrounding audio that was recorded.

[0089] According to one or more embodiments, UE2 **146** can also register itself within permissive network of UE1 **100** and vice-versa. Thus in one implementation, which is also illustrated by the method **1600** of **FIG 16**, the audio capture utility configures the audio capture device to: receive a request from the image capture device to provide authorization to be included within a permissive network of the audio capture device; generate a response message based on user selection and transmit the response message to the image capture device via a communication means; update a locally stored and maintained list of image capture devices within a local permissive network (permissive group **1580**) to indicate whether the response message indicated that the image capture device can be registered within the permissive network of the audio capture device; and store the update to the locally stored list of image capture devices of the local permissive network within the storage **1550**. In one implementation, the audio capture utility configures the audio capture device to: in response to receiving the ACA request signal, access the list of image capture devices within the permissive group (PG) **1580** (i.e., a network of image capture UEs that can also be referred to as a local permissive network) to identify whether the received ACA request signal is from an image capture device (e.g., UE1 **100**) within the list of image capture devices registered within the permissive group **1580**. Then the utility configures the audio capture device to activate the microphone **1520** only in response to receiving the ACA request signal from an image capture device on the list of permissive devices.

[0090] In one embodiment, the audio capture utility further configures the device to: provide a user interface with one or more selectable options to configure the audio capture device for audio capture support during image capture at one or more image capture devices; receive one or more user selections of one or more options ranging from (a) enabling open audio capture and return to all requesting image capture devices, without restrictions, to (b) not enabling audio capture triggered by any image capture device; and configure the audio capture device to implement a device setting for audio capture based on the user selections.

[0091] **FIG. 16** is a flow diagram illustrating a method by which UE2 **146** can be initially configured to allow local audio capture during an image capture event occurring at a first image capture device, e.g., UE1 **100**, according to one or more embodiments of the disclosure. The method **1600** is implemented on UE2 **146** via execution by data processor **1504** of program code segments of SmartPhoto audio capture client **1510**. Method **1600** can be described from the perspective of one of UE2 **146**, data processor **1504**, and/or SmartPhoto audio capture client **1510**. Method begins at block **1602** with UE2 **146** receiving a request from UE1 **100** for registration of UE2 **146** within authorized devices for smart photo audio capture triggered by an image capture event at UE1. The received request asks the user of UE2 **146** to provide authorization to have UE2 **146** be included within a permissive network of UE1 **100**. Method **1600** then includes providing a user interface with one or more selectable options to configure the audio capture device for audio capture support during image capture at one or more other devices. Specifically, this process includes UE2 **146** opening the SmartPhoto client on UE2 (block **1604**) and showing the request on a user interface with specific selectable response options (block **1606**). Method **1600** can then include receiving one or more user selections of one or more options ranging from (a) enabling open audio capture and return to all requesting other devices, without restrictions, to (b) not enabling audio capture triggered by any other device. At decision block **1608**, method **1600** provides determining whether a selected option is received within the user interface. In response to not receiving a selection (e.g., yes, no, limited or restricted authorization) after a pre-established timeout period expires, UE2 **146** closes the interface without sending a response to UE1 **100** (block **1610**). However, once a response is received, method **1600** includes UE2 generating a response message based on user selection and transmitting the response message to the other device via a communication means (block **1612**). Method **1600** then includes updating a locally stored and maintained list of image capture devices within a local permissive network to indicate whether the response message indicated that the UE2 **146** can be registered within the permissive network of the requesting device, UE1 **100**. Method **1600** further includes storing the update to the locally stored list of other devices of the local permissive network within the

storage, and configuring the device to implement a device setting for audio capture based on the user selections.

[0092] According to at least one implementation, method **1600** enables future modification of the selected option for that image capture device (UE1 **100**) at a later time (block **1616**), based on received user input that changes the selections made during the initial selection. Thus, for example, a user can modify a prior selection that allowed automatic audio capture at UE2 based on a trigger received from UE1 to one of several new selections, such as (a) to limit the audio capture to only a specific time, and/or (b) location or (c) to prompt for each received ACA request before commencing audio capture at UE2. Method **1600** then terminates, as provided at end block.

[0093] Once UE2 has been established or set up to provide audio capture capabilities to support image capture events at UE1, a second method **1700** can be implemented on UE2 **146** again via processor execution of SmartPhoto audio capture client **1510**. Turning to **FIG. 17**, there is provided a flow diagram illustrating a method by which an audio capture device, e.g., UE2 **146**, is triggered to capture audio during an image capture event and return the audio as text content to a first user equipment, e.g., UE1 **100**, performing the image capture event, according to one or more embodiments. Method **1700** begins at start block and proceeds to block **1702** at which UE2 **146** receives a signal to start audio capture. As introduced above, the received signal is an incoming audio capture activation (ACA) request signal from an image capture UE, UE1 **100**. In response to receiving an incoming ACA request signal from UE1 **100**, method **1700** provides audio capture processes including opening SmartPhoto audio capture client **1510**, checking the UE2's audio capture permissions setting or authorization list for UE1 **100** (block **1704**), and determining at decision block **1706** whether the UE1 requesting audio capture is authorized or approved at UE2 **146**. In one embodiment, this check also includes checking if UE2 **146** is set for open audio capture, where no restrictions exist and/or no or special permissions are required to trigger UE2 **146** to initiate audio capture.

[0094] In one embodiment, method **1700** provides: in response to receiving the ACA request signal, accessing the list of image capture devices within the locally-stored permissive network to identify whether the received ACA request signal is from an image capture device within the list of image capture devices registered within the local permissive network; and activating the microphone only in response to receiving the ACA request signal from an image capture device on the list of image capture devices.

[0095] In response to UE2 **146** not being open for audio capture, the specific requesting UE1 being on an excluded list of UE1s, or the requesting UE1 not being on a list of authorized UE1s, the method provides closing the SP client and preventing the requested audio capture (block **1708**). The method then ends. However, if the UE2 **146** is set to open audio capture and/or the requesting UE1 is authorized, method **1700** includes initiating the timer to begin tracking a local time sequence and activating a microphone to begin recording surrounding audio during the local time sequence (block **1710**). Method **1700** further includes generating a notification that audio capture is about to commence on UE2 **146** and transmitting the notification to a visible user interface of UE2 **146** (block **1712**). Along with the notification, a selection can be provided that enables the user of UE2 **146** to terminate an audio capture in real time. At decision block **1714**, method **1700** includes determining if the user enters a restriction on audio capture via an input or selection on or at UE2 **146**. If entry of a restriction is detected, method **1700** includes applying the specific restriction to the audio capture (block **1716**). Thus, for example, the restriction can range from an outright refusal to allow an audio capture to allowing the user to control when audio capture is terminated. When the restriction is an outright refusal to allow audio capture, method **1700** includes stopping the audio capture initiated at block **1710** and discarding the recorded audio (**1717**). Method **1700** then ends. However, when the restriction is a partial restriction, method **1700** proceeds to block **1718**.

[0096] Assuming that no restriction input is received within a pre-set amount of time and/or the applied restriction does permit for some audio capture, method **1700** includes determining at decision block **1718** when an end of audio capture condition

occurs, e.g., receipt of a TAC signal from UE1 **100** or an end of timeout period. In response to detection of the end of audio capture condition, method **1700** provides turning off the microphone at one of (a) a preset end time of the local time sequence and (b) receipt of an incoming terminate audio capture (TAC) signal from the other device (block **1720**). Method **1700** then provides converting the recorded surrounding audio into text content (block **1722**); compressing the surrounding audio that was recorded to generate a compressed audio file (block **1724**); packaging the text content into an outgoing message packet tagged with a device identifier (ID) to indicate that the outgoing message packet is being transmitted from the specific audio capture device (block **1726**); and forwarding the outgoing message packet with the text content to another device (block **1728**). As introduced above, the other device can be one of (1) the UE1 **100** from which the incoming ACA request was received or (2) a server or computing device that processes the audio context data and captured image data. However, as shown by block **1719**, method **1700** includes continuing to record the audio so long as the end of audio capture condition is not detected.

[0097] According to one embodiment, method **1700** can include: creating an outgoing message packet tagged with a device identifier (ID) that identifies the audio capture device, inserting the compressed audio file as additional payload within the outgoing message packet; inserting the text content as additional content into the outgoing message packet; and transmitting the outgoing message packet with both the text content and the compressed audio file to the other device. In a first embodiment, transmitting the outgoing message packet includes transmitting the outgoing message packet following one of (a) a preset end time of the local time sequence and (b) receipt of an incoming terminate audio capture (TAC) signal from the first device. In a second embodiment, transmitting the outgoing message packet includes transmitting the outgoing message packet to an intermediate server that performs audio source identification via facial recognition and a tagging of images received from the first device with the surrounding audio recorded.

[0098] At block **1730**, method **1700** optionally provides deleting the audio file following a successful transfer and acknowledgement of the successful transfer of the SP message packet. Method **1700** further includes storing the surrounding audio

recorded during the local time sequence; and transmitting to the other device an outgoing message packet containing at least one of (a) the surrounding audio recorded and (b) a textual representation of the surrounding audio (block 1732).

[0099] **FIG. 18** is a message flow and time sequence diagram illustrating the timing aspects and messaging aspects of an audio capture process triggered by image capture UE (i.e., UE1 100) and involving two audio capture UEs (UE2s 146). The passing of messages and/or signals and data packets between devices occurs during an image capture event on UE1 100, and the UE2s, separately identified as UE2-A and UE2-B are both within a communication range of UE1 100. The timeline, i.e.,  $T$ ,  $T1$ , and  $T2$ , associated with each device is represented by a vertical line running down from each device, and functions occurring at the specific device are then placed on the device's timeline in solid boxes. Messages passed between devices are then illustrated by horizontal arrows, with the associated dashed-line message boxes identifying the specific message. Each message is provided a relative timestamp on the respective timeline of the device from which the message is generated. The time sequence begins with timeline  $T$  of UE1 100 set to zero. A user of UE1 100 opens or activates the camera for image capture at time  $T=0$ , which triggers the start of an image capture period at UE1 100. In at least one embodiment, this action also triggers the start of an audio recording period at UE1 100, since the speech of the photographer is often relevant in providing context to the captured photographic image. This aspect is illustrated by **FIG. 19**. The camera utility, SmartPhoto client 110, automatically obtains a list of supported audio capture devices and identifies which audio capture devices are in communication range within the area in which the image is being captured. The detected devices from the approved list are shown to include at least UE2-A and UE2-B. UE1 100 is thus responsible for identifying compatible devices of people who are associated in some way with the photograph being taken or the surroundings or location of the photograph. In one embodiment, the UE1 100 signals the user of the audio capture devices and instructs the user to initiate audio capture at the user's audio capture device.

[00100] At  $T=t1$ , the camera utility generates and transmits the ACA signals to both UE2-A and UE2-B as soon as the devices are discovered. It is appreciated that some

processing time will be required between the first function at time  $T=0$  and the transmission of the ACA signals. However, these functions can be completed on most UEs within a relatively short period of time (e.g., less than one second). Once the ACA signals are received at UE2-A and UE2-B, both audio capture devices initiate audio capture and set their respective timers,  $T1$  and  $T2$ , to zero. In this scenario,  $T1=0$  and  $T2=0$  corresponds to " $T=t1$  + transmission latency" between devices. One embodiment can involve determining the latency of transmission via Request-Ack pings between devices. However, a simpler embodiment simply utilizes  $T=t1$  as the zero time for both audio capture devices, given the short latency of transmission, even if utilizing cellular based messaging versus NFC messaging.

**[00101]** At  $T=t2$ , UE1 captures the image, and at  $T=t3$ , UE1 **100** stops audio capture. In one embodiment  $t3$  can be the same as  $t2$ , where audio capture is terminated immediately after the image is captured. However, alternate embodiments provide for a delay between  $T=t2$  and  $T=t3$ . At  $T=t4$ , which can again be equal to  $T=t3$ , UE1 **100** generates and transmits an ACT message to each of UE2-A and UE2-B. UE2-A and UE2-B each stops their audio recording at respective times  $T1=M$  and  $T2=N$ , where  $M$  and  $N$  are real values that can be similar, when minimal propagation delay offsets are required for message transfer between devices.

**[00102]** Within the example, UE2-A and UE2-B process the recorded audio differently and thus have different timing for returning audio context data to UE1 **100**. As shown, UE2-A packages the audio file with other details, such as the UE2-A ID and time data, at  $T1=P$  and transmits the message packet to UE1 **100** at  $T1=R$ . Soon thereafter, at  $T=t5$ , UE1 **100** receives the message packet and converts the audio to text before linking the text to the source of the speech in the captured image. UE1 **100** then links or embeds the text in the image as the audio-context tag for the audio source.

**[00103]** At  $T2=Q$ , UE2-B initiates speech-to-text conversion, which takes a variable period of time dependent on how long the recorded speech audio is. It is appreciated that the amount of recorded audio can be as long as the entire period of image capture (less the processing time at UE1 **100** and the transmission latency of the ACA signal)

or with audio capture devices that have audio triggered recording, the recording can last for just the time period during which the user of the device is actually speaking. It is appreciated that having the individual audio capture devices that record the audio also perform the speech to text conversions has the benefit of the audio capture device already being trained to recognize the voice patterns of the device user with its voice recognition and/or audio conversion software. Also, by having the individual audio capture devices tuned to record localized user audio, the influence of noise from the surroundings is reduced as the device is as close to the user as possible. In one example implementation individual users can, with minimal coordination, pair their Bluetooth® headsets and keep the headsets ON as the user moves about in the surroundings thus enabling the audio capture device to clearly record the user's speech and other voiced utterances. This further enhances the quality of the converted text content.

[00104] Returning to the figure, at  $T_2=S$ , UE2-B packages the text content converted from the audio, as well as the other audio context data, such as time data, the audio file, and the device ID into the message packet, and UE2-B transmits the message packet to UE1 100 at  $T_2=V$ . UE1 100 then links the text content to the corresponding source within the captured image, at  $T=t6$ . The audio-context-tagged image is then stored at  $T=t7$ . It is appreciated that the linking of audio-context tags to the image can occur over several cycles, as more and more audio context data is received from different audio capture devices. The image is updated with the relevant audio context tags over a period of time, such that the image can initially be stored without any audio-context tagging and then the image file is subsequently updated with metadata associated with the audio-context tags received.

[00105] As shown in the previous illustrations of the image and in FIG. 19, the metadata are added in such a manner that any image rendering application can associate the audio-context information with the identity of those on the image. As provided in one embodiment, the identities of those who speak during the image capture event can be obtained using face recognition software along with identifying data received in the message packet. In one or more embodiments, a new metadata or existing metadata in file formats such as EXIF can be used to represent the audio-

context tag information. As again illustrated by **FIG. 19**, this audio-context tag information can include (a) the identity of the person in the image who spoke (e.g. Krish), what that person said, presented in text format (e.g., "C'mon guys, Titanic isn't sinking. Smile please"), the overall timeline, e.g.,  $t=-4$  seconds to  $t=3$  seconds, and specific time of the captured audio.

**[00106]** In each of the flow charts presented herein, certain steps of the methods can be combined, performed simultaneously or in a different order, or perhaps omitted, without deviating from the spirit and scope of the described innovation. While the method steps are described and illustrated in a particular sequence, use of a specific sequence of steps is not meant to imply any limitations on the innovation. Changes may be made with regards to the sequence of steps without departing from the spirit or scope of the present innovation. Use of a particular sequence is therefore, not to be taken in a limiting sense, and the scope of the present innovation is defined only by the appended claims.

**[00107]** As will be appreciated by one skilled in the art, embodiments of the present innovation may be embodied as a system, device, and/or method. Accordingly, embodiments of the present innovation may take the form of an entirely hardware embodiment or an embodiment combining software and hardware embodiments that may all generally be referred to herein as a "circuit," "module" or "system."

**[00108]** Aspects of the present innovation are described below with reference to flowchart illustrations and/or block diagrams of methods, apparatus (systems) and computer program products according to embodiments of the innovation. It will be understood that each block of the flowchart illustrations and/or block diagrams, and combinations of blocks in the flowchart illustrations and/or block diagrams, can be implemented by computer program instructions. These computer program instructions may be provided to a processor of a general purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions, which execute via the processor of the computer or other programmable data processing apparatus, create means for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks.

[00109] While the innovation has been described with reference to exemplary embodiments, it will be understood by those skilled in the art that various changes may be made and equivalents may be substituted for elements thereof without departing from the scope of the innovation. In addition, many modifications may be made to adapt a particular system, device or component thereof to the teachings of the innovation without departing from the essential scope thereof. Therefore, it is intended that the innovation not be limited to the particular embodiments disclosed for carrying out this innovation, but that the innovation will include all embodiments falling within the scope of the appended claims. Moreover, the use of the terms first, second, etc. do not denote any order or importance, but rather the terms first, second, etc. are used to distinguish one element from another.

[00110] The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the innovation. As used herein, the singular forms "a", "an" and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms "comprises" and/or "comprising," when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof.

[00111] The corresponding structures, materials, acts, and equivalents of all means or step plus function elements in the claims below are intended to include any structure, material, or act for performing the function in combination with other claimed elements as specifically claimed. The description of the present innovation has been presented for purposes of illustration and description, but is not intended to be exhaustive or limited to the innovation in the form disclosed. Many modifications and variations will be apparent to those of ordinary skill in the art without departing from the scope and spirit of the innovation. The embodiment was chosen and described in order to best explain the principles of the innovation and the practical application, and to enable others of ordinary skill in the art to understand the innovation for various embodiments with various modifications as are suited to the particular use contemplated.

## CLAIMS

What is claimed is:

1. A device comprising:
  - an image capturing component;
  - a timer;
  - a communication mechanism that enables the device to communicate with at least one second device;
  - a processor that is communicatively coupled to each of the image capture component, the timer, and the communication mechanism; and
  - an image capture utility that executes on the processor and configures the device to:
    - activate the timer to begin tracking a time sequence that extends to an end time following capture of the image;
    - generate an audio capture activation (ACA) signal and transmit the audio capture activation signal via the communication mechanism to the at least one second device;
    - capture the image during the time sequence;
    - receive from the at least one second device content that represents audio that was captured at the at least one second device during the time sequence;
    - link the received content with the image; and
    - store the image along with the linked content as an audio-context-tagged image.
2. The device of Claim 1, wherein linking the content with the image comprises the utility further configuring the device to:
  - record a time at which the image is captured;
  - identify, from data included within the message packet, a specific time within the time sequence during which the audio was captured at the at least one second device;

identify, within the image a source of the audio that is represented by the received content; and

associate the content and the specific time with the source to provide the linked content.

3. The device of Claim 2, wherein the received content includes text content converted from the audio that was captured, and the device further comprises:

a display device communicatively coupled to the processor, wherein in response to a selection of the audio-context-tagged image to be displayed on the display device, the image capture utility further configures the device to provide, within the display of the image, a visual output of the text content associated with portions of the image as at least one of (a) a time-specific text content and (b) a source-specific text content, where the visual output of the source-specific text content identifies the source of the audio and the visual output of the time-specific text content identifies the specific time at which the audio was recorded at the at least one second device relative to the time at which the image was captured by the device.

4. The device of Claim 2, wherein the utility further configures the device to transmit a terminate audio capture (TAC) signal via the communication mechanism after one of a pre-set length of time and a received input at the device following the capture of the image, wherein the TAC signal notifies the at least one second device to stop the audio capture and the end time of the time sequence corresponds to the transmission of the TAC signal.

5. The device of Claim 3, wherein providing the visual output of the time-specific and source-specific text content comprises the utility configuring the device to concurrently display one or more text content within a single-frame output of the audio-context-tagged image identifying the source of the corresponding audio, with each of the one or more text content being concurrently displayed as an overlay on the image along with a timestamp associated with that text content.

6. The device of Claim 3, wherein providing the visual output of the time-specific and source-specific text content comprises the utility configuring the device to:

display the image from the audio-context-tagged image without any of the text content;

detect a selection within the image of one source of captured audio that corresponds to a text content linked to the audio-context-tagged image; and

in response to detecting the selection, display the specific text content via a text overlay on the image.

7. The device of Claim 6, wherein the utility further configures the device to display the text overlay for one of (a) at least a specific period and (b) an entire period, while the source of the captured audio remains selected.

8. The device of Claim 6, wherein the utility further configures the device to concurrently indicate within the visual output a time at which the audio corresponding to the text content was recorded at the at least one second device.

9. The device of Claim 3, wherein providing the visual output of the time-specific and source-specific text content comprises the utility configuring the device to:

display the audio-context-tagged image;

associate and display a timeline along with the display of the audio-context-tagged image, wherein the timeline represents a visual depiction of the time sequence, including a time of image capture;

enable selection of a specific time period on the time line; and

in response to the specific time period selected coinciding with a time of capture of audio, which was converted into corresponding text content that is linked to the image, display the corresponding text content as an overlay on the image, identifying the source of the audio.

10. The device of Claim 9, wherein the utility configures the device to provide a moveable time bar associated with the time line that can be selectively moved along the length of the time line, wherein the specific period is selected when the time bar is moved to the corresponding point on the timeline and the text content is only presented within the visual output of the audio-context-tagged image when the time bar is located at that point on the timeline.

11. The device of Claim 2, wherein the utility further configures the device to:  
determine when the source of the audio is not captured within the image; and  
present an output of the text content along a periphery of the image along with an identifier of the source in alphanumeric characters to indicate a presence of the source within a surrounding space of the area captured within the image.

12. The device of Claim 1, wherein the communication mechanism is near field communications, the at least one second device comprises a plurality of second devices, and the image capture utility further configures the device to:

identify the plurality of second devices within a communication range of the device; and

transmit the ACA signal to the plurality of second devices within the communication range to trigger the plurality of second devices with audio capture capability to provide audio capture feedback to the device.

13. The device of Claim 12, wherein the image capture utility further configures the device to:

determine which of the plurality of second devices are configured to provide audio capture feedback to the first device, wherein the plurality of second devices that are configured to provide audio capture feedback to the first device are devices that are either (a) pre-configured to be within a permissive network of the first device for audio capture or (b) configured as open devices for audio capture during an image capture event at another device;

wherein the utility enables configuration of the device to only transmit the ACA request to those plurality of second devices within the communication range that are configured to provide audio feedback.

14. The device of Claim 13, wherein the image capture utility further configures the device to:

- generate a request for one or more of the plurality of second devices to provide authorization to be included within the permissive network;

- transmit the request to the one or more of the plurality of second devices via a communication means;

- register within the permissive network those plurality of second devices that provide an affirmative response to the request;

- store the permissive network within an accessible storage of the device; and

- access the permissive network prior to transmitting the ACA signal to those plurality of second devices registered within the permissive network.

15. The device of Claim 1, wherein the device is one of (a) a camera and (b) a wireless communication device, where the image capturing component comprises a camera.

16. A method for providing audio-context-tagging of an image, the method comprising:

- activating a timer to begin tracking a time sequence that extends to an end time following capture of the image;

- generating an audio capture activation (ACA) signal and transmitting the audio capture activation signal via a communication mechanism to at least one second device;

- capturing the image during the time sequence;

- receiving from the at least one second device text content that represents text converted from audio that was captured at the at least one second device during the time sequence;

- linking the text content with the image; and

storing the image along with the linked text content as an audio-context-tagged image.

17. The method of Claim 16, wherein linking the text content with the image comprises:

recording a time at which the image is captured;

identifying, from data included within the message packet, a specific time within the time sequence during which the audio was captured at the at least one second device;

identifying, within the image, a source of the audio that was converted to the text content; and

associating the text content and the specific time with the source to provide the linked text content.

18. The method of Claim 17, further comprising:

in response to selection of the audio-context-tagged image to display on a display device, providing, within the display of the image, a visual output of the text content associated with portions of the image as at least one of (a) a time-specific text content and (b) a source-specific text content, where the visual output of the source-specific text content identifies the source of the audio and the visual output of the time-specific text content identifies the specific time at which the audio was recorded at the at least one second device relative to the time at which the image was captured by the device.

19. The method of Claim 18, wherein providing the visual output of the time-specific and source-specific text content comprises concurrent displaying of one or more text content within a single-frame output of the audio-context-tagged image identifying the source of the corresponding audio, with each of the one or more text content being concurrently displayed as an overlay on the image along with a timestamp associated with that text content.

20. The method of Claim 18, wherein providing the visual output of the time-specific and source-specific text content comprises:

- initially displaying the image from the audio-context-tagged image without any of the text content;

- detecting a selection within the image of one source of captured audio that corresponds to a specific text content linked to the audio-context-tagged image;

- in response to detecting the selection, displaying the specific text content via a text overlay on the image, for one of (a) an initial period and (b) an entire period, while the source of the captured audio remains selected; and

- concurrently indicate within the visual output a time at which the audio corresponding to the text content was recorded at the at least one second device.

21. The method of Claim 16, wherein providing the visual output of the time-specific and source specific text content comprises:

- displaying the audio-context-tagged image;

- associating and displaying a timeline along with the display of the audio-context-tagged image, wherein the timeline represents a visual depiction of the time sequence, including a time of image capture;

- enabling selection of a specific time period on the time line; and

- in response to the specific time period selected coinciding with a time of capture of an audio, which was converted into corresponding text content that is linked to the image, displaying the corresponding text content as an overlay on the image, identifying the source of the audio.

22. The method of Claim 1, wherein the transmitting of the ACA signal further comprises:

- identifying at least one second device within a communication range of the first device;

- determining which of the at least one second device are configured to provide audio capture feedback to the first device, wherein the second devices that are configured to provide audio capture feedback to the first device comprise (a) second devices that are pre-configured to be within a permissive network of the first device

for audio capture and (b) second devices that are configured as open devices for audio capture during an image capture event occurring at another device; and

transmitting the ACA signal to a group of second devices to trigger those second devices with audio capture capability to provide audio capture feedback to the first device, wherein the group of second devices is one of (1) a first group including each of the at least one second device within the communication range and (2) a second group including only those second devices within the communication range that are configured to provide audio feedback and are one of (a) open devices and (b) second devices within a permissive network of the first device, wherein the ACA request is only transmitted to those second devices that are configured and authorized to provide audio feedback to the first device.

1/15

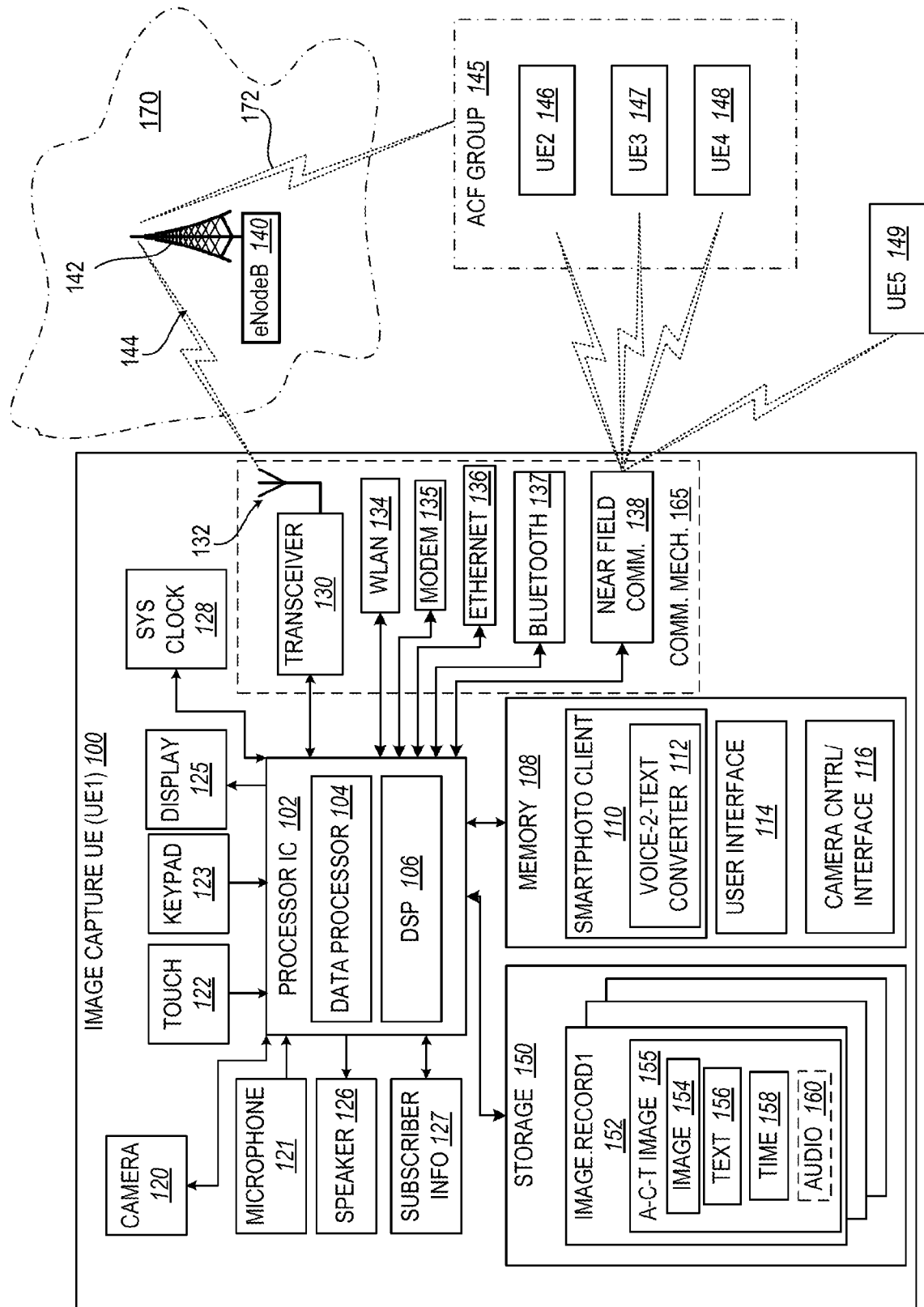


FIG. 1

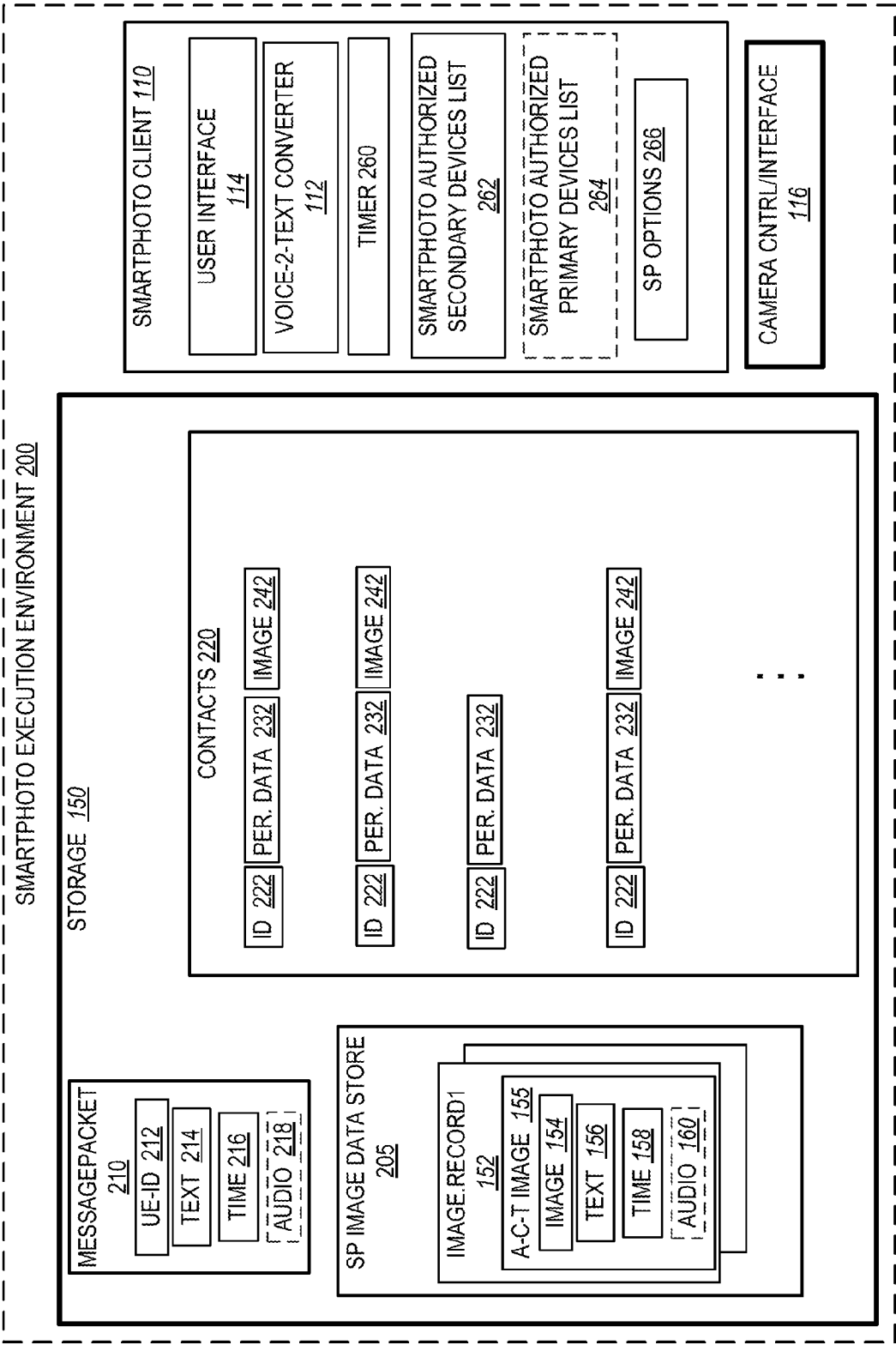
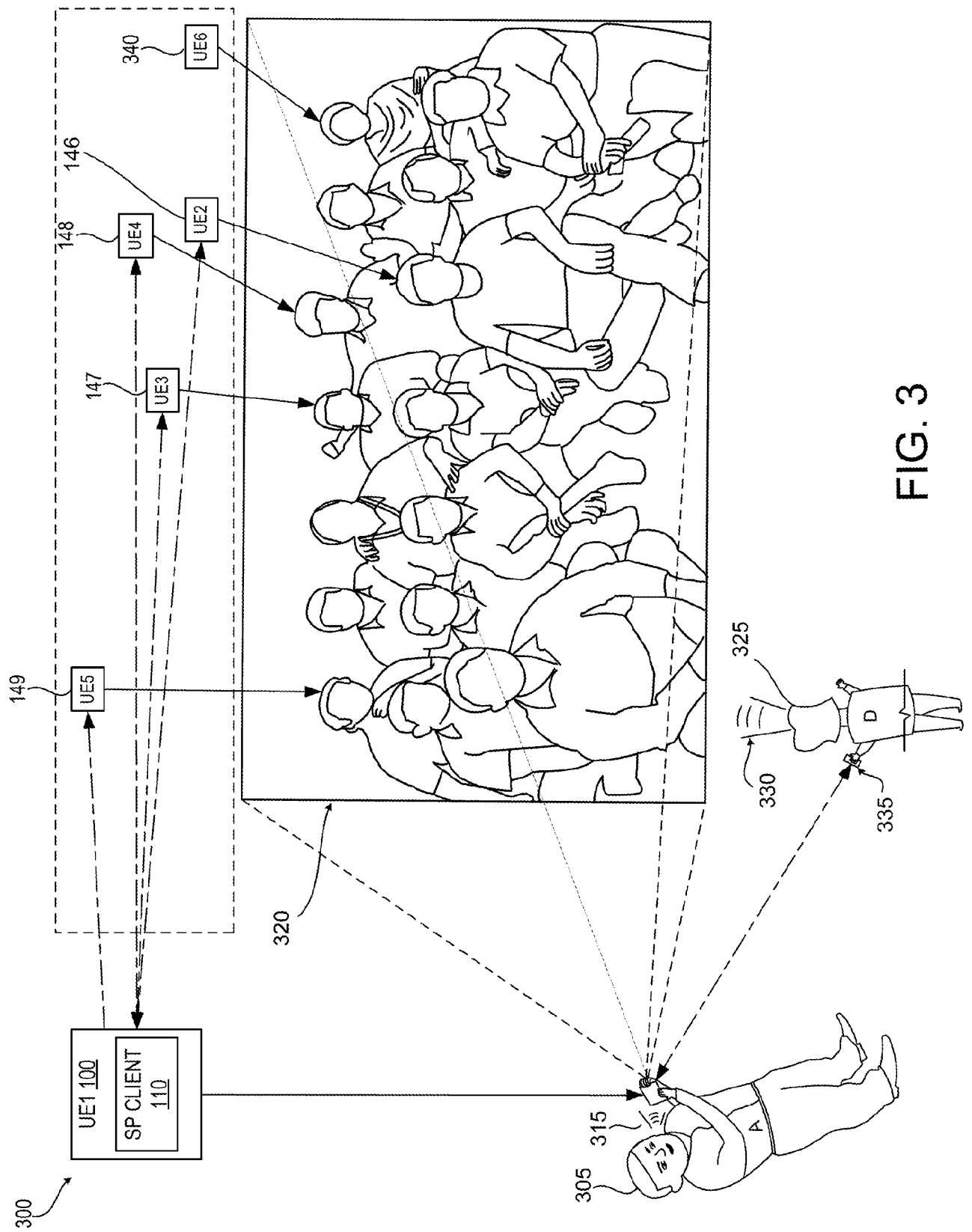


FIG. 2

3/15



4/15

400

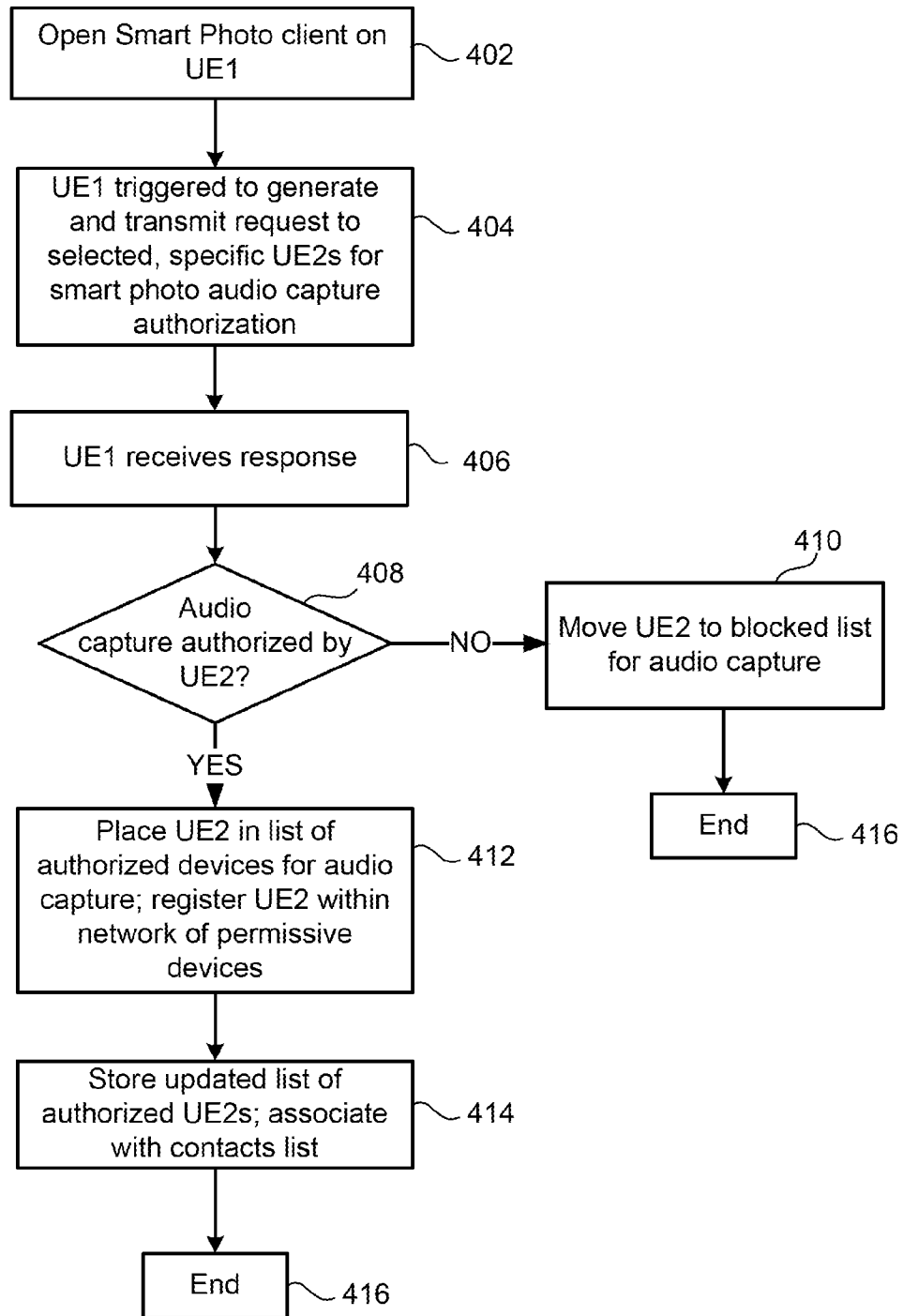


FIG. 4

5/15

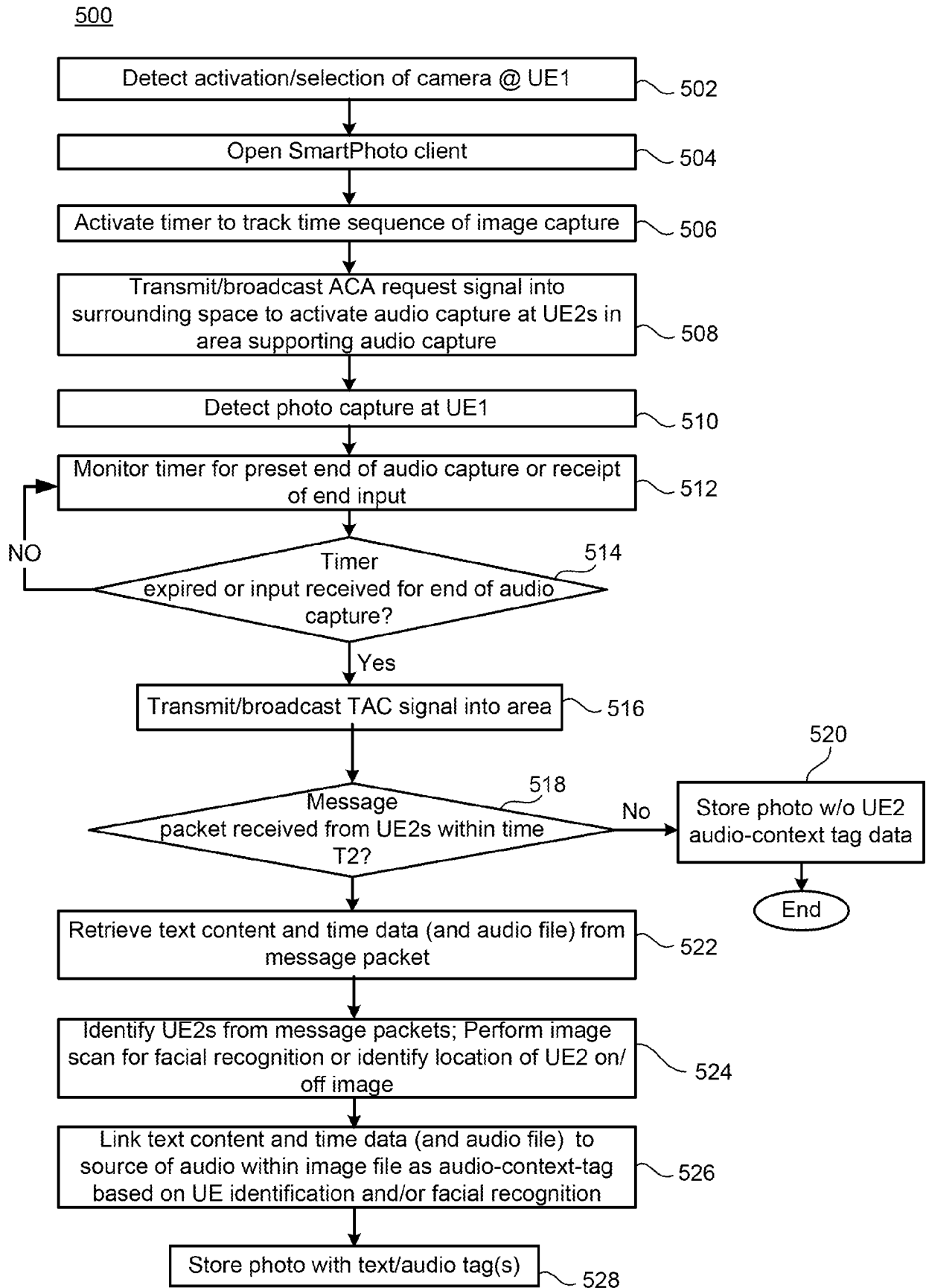


FIG. 5

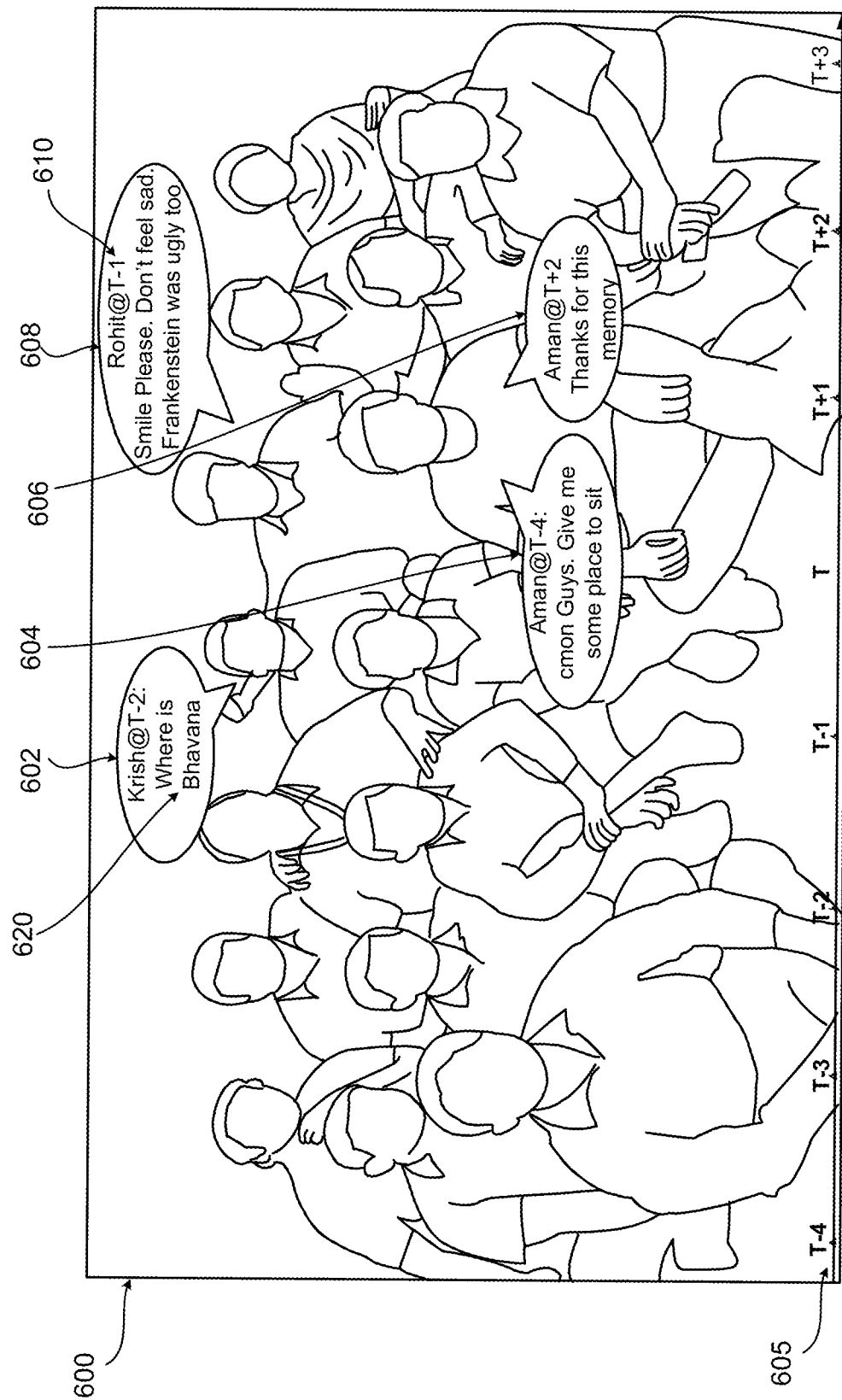


FIG. 6

FIG. 7

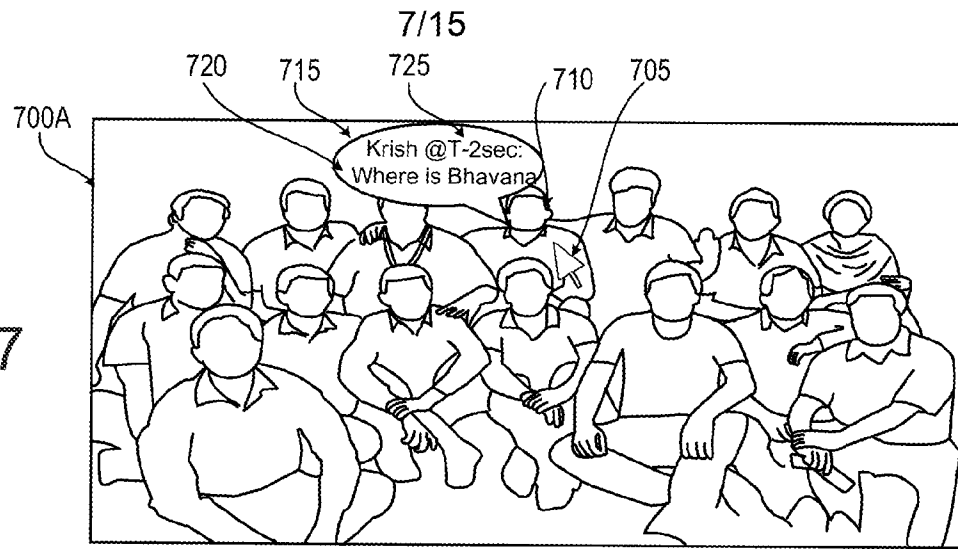


FIG. 8

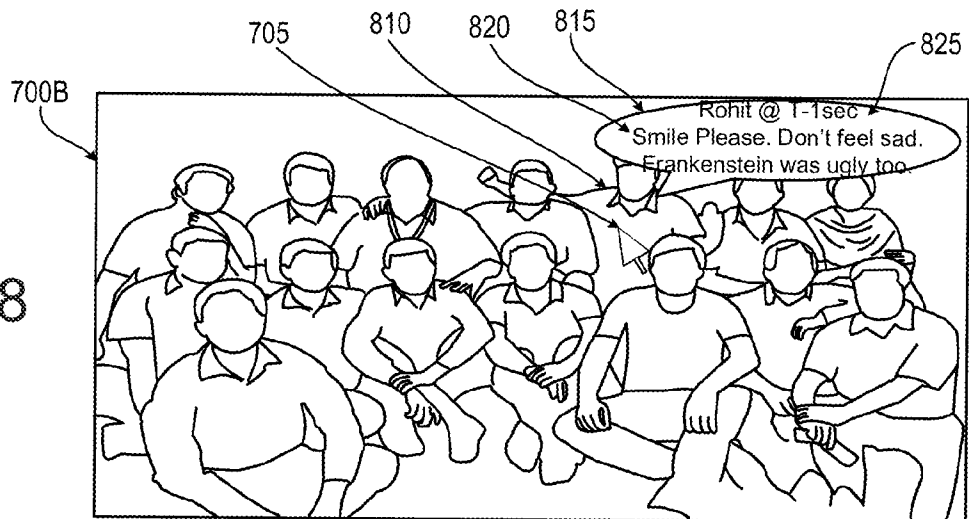
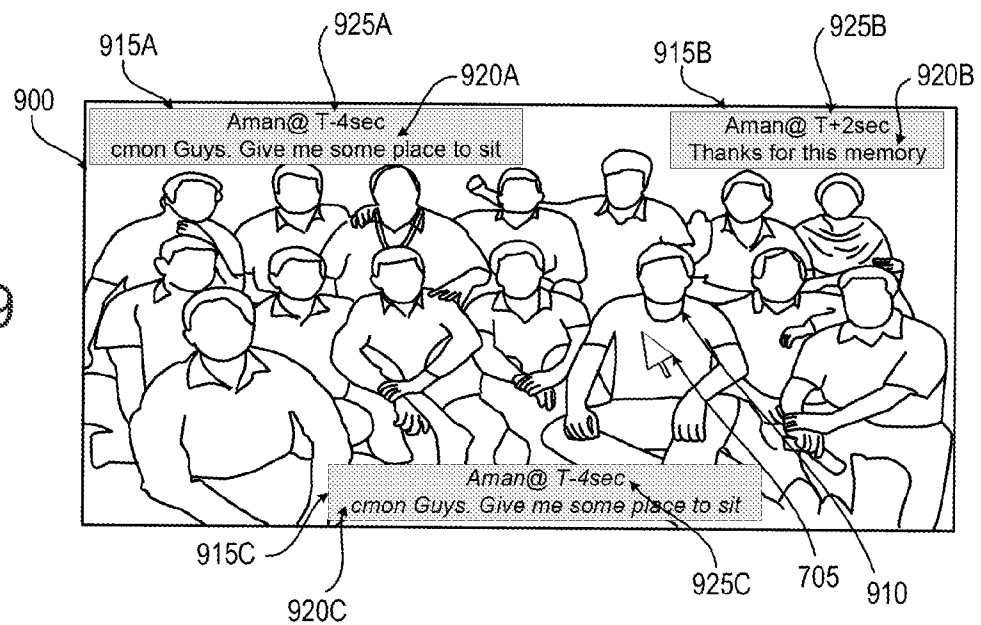


FIG. 9



8/15

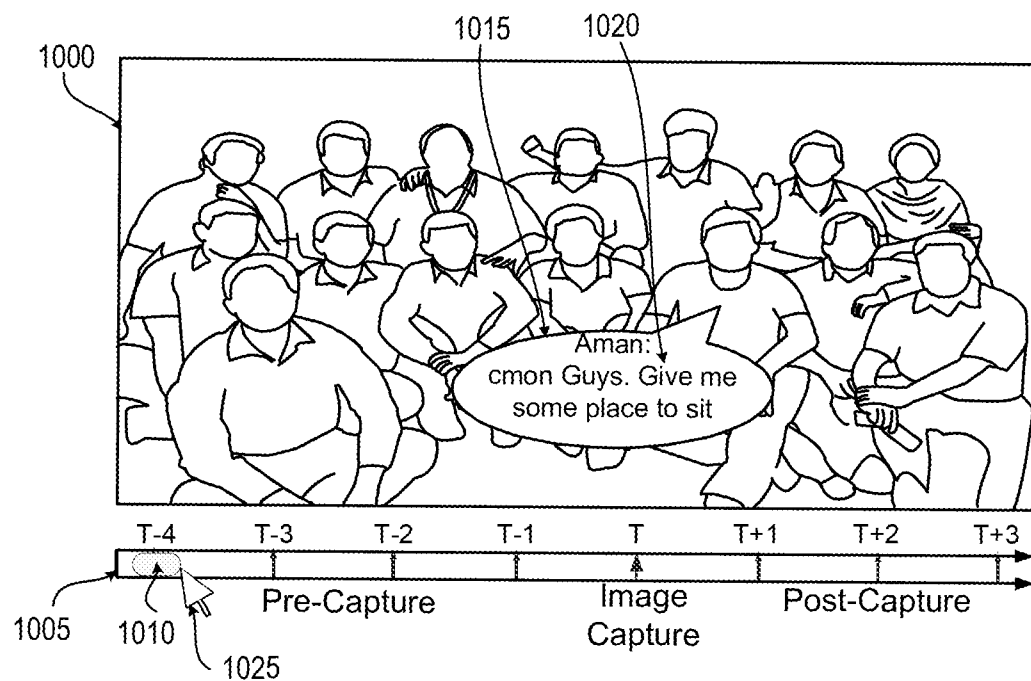


FIG. 10

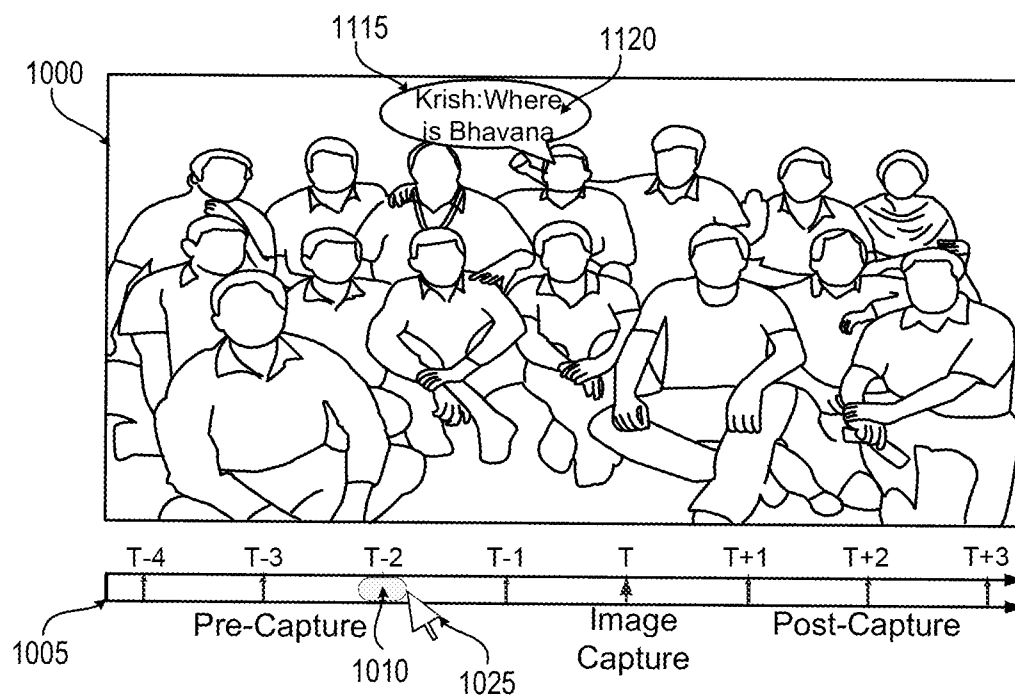


FIG. 11

9/15

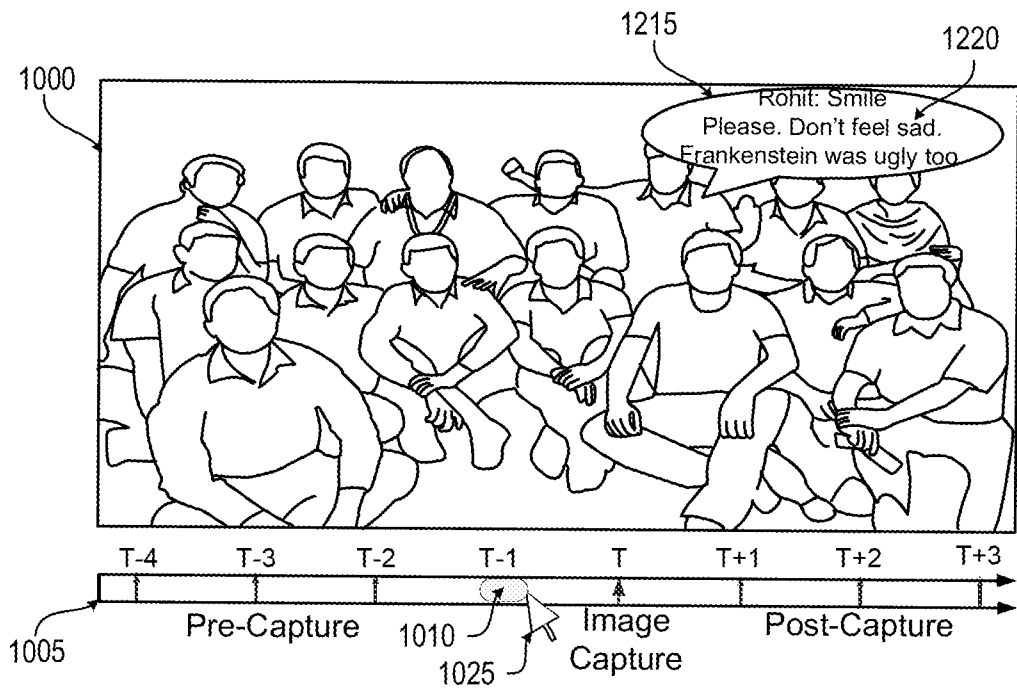


FIG. 12

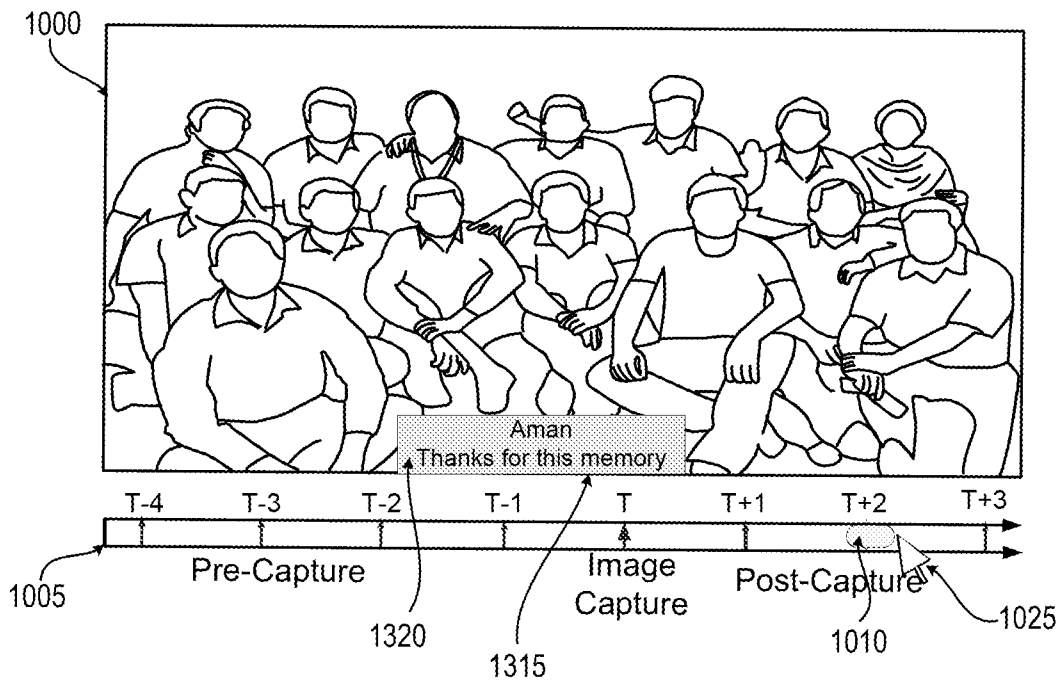


FIG. 13

10/15

1400

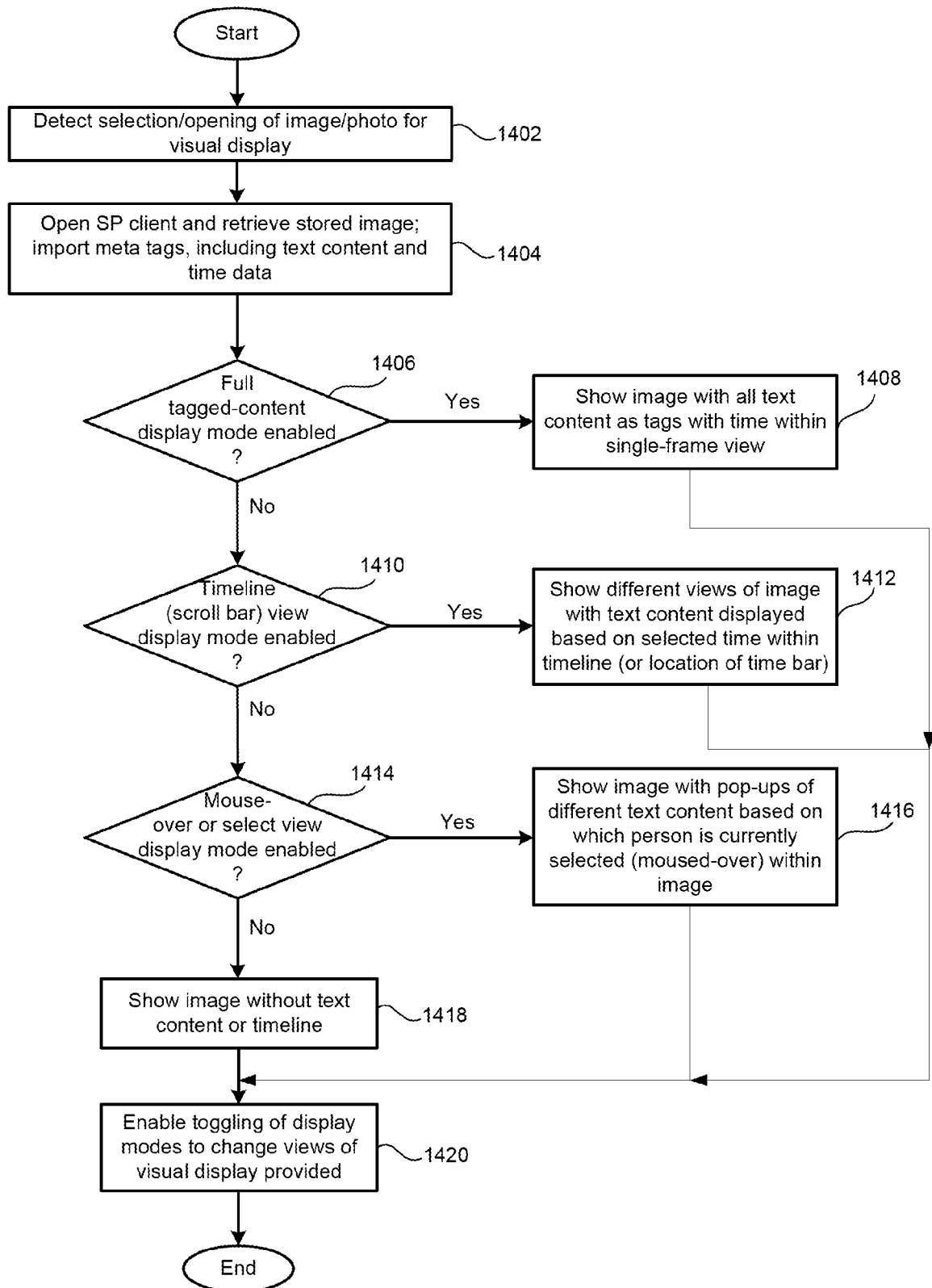


FIG. 14

11/15

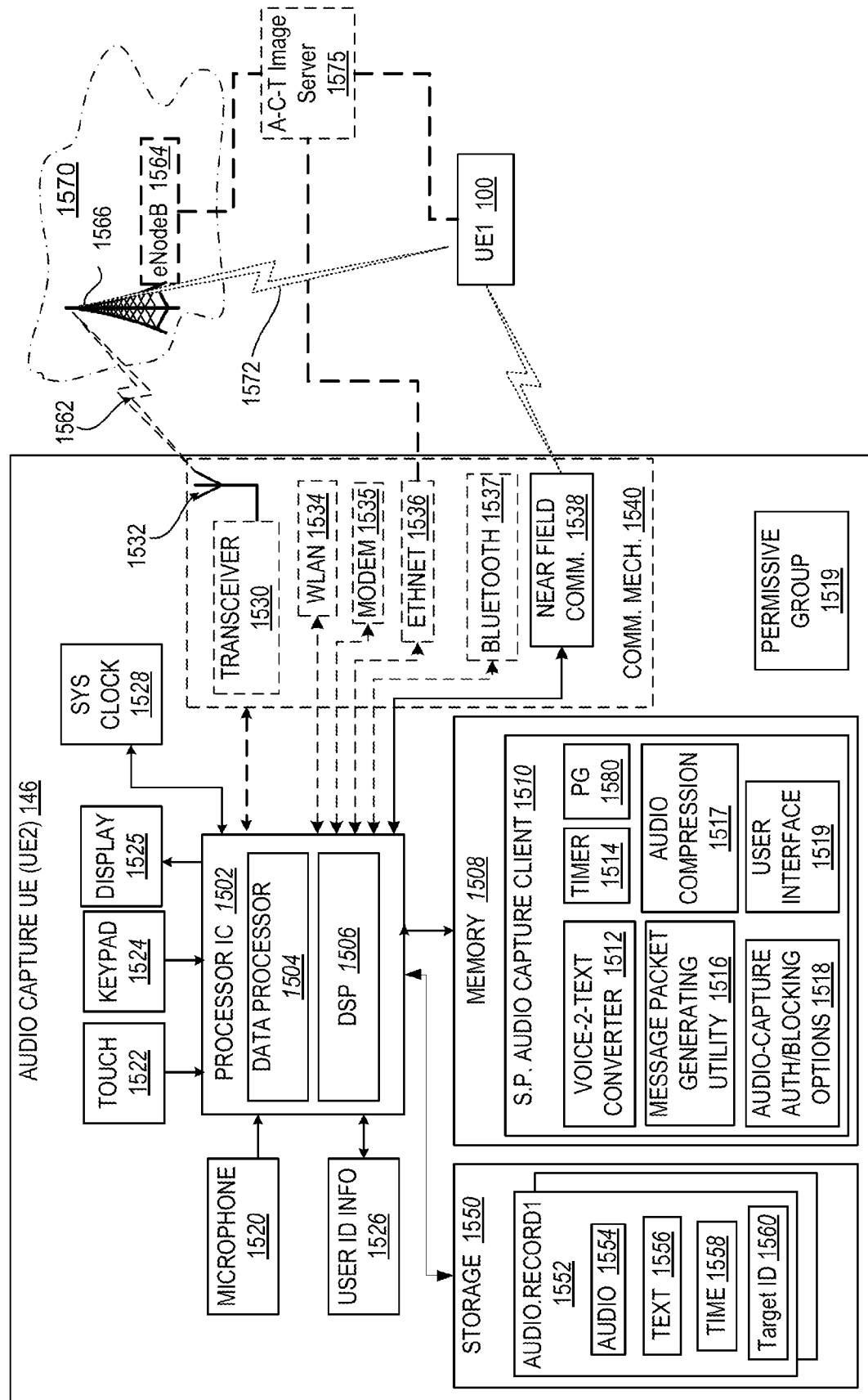


FIG. 15

12/15

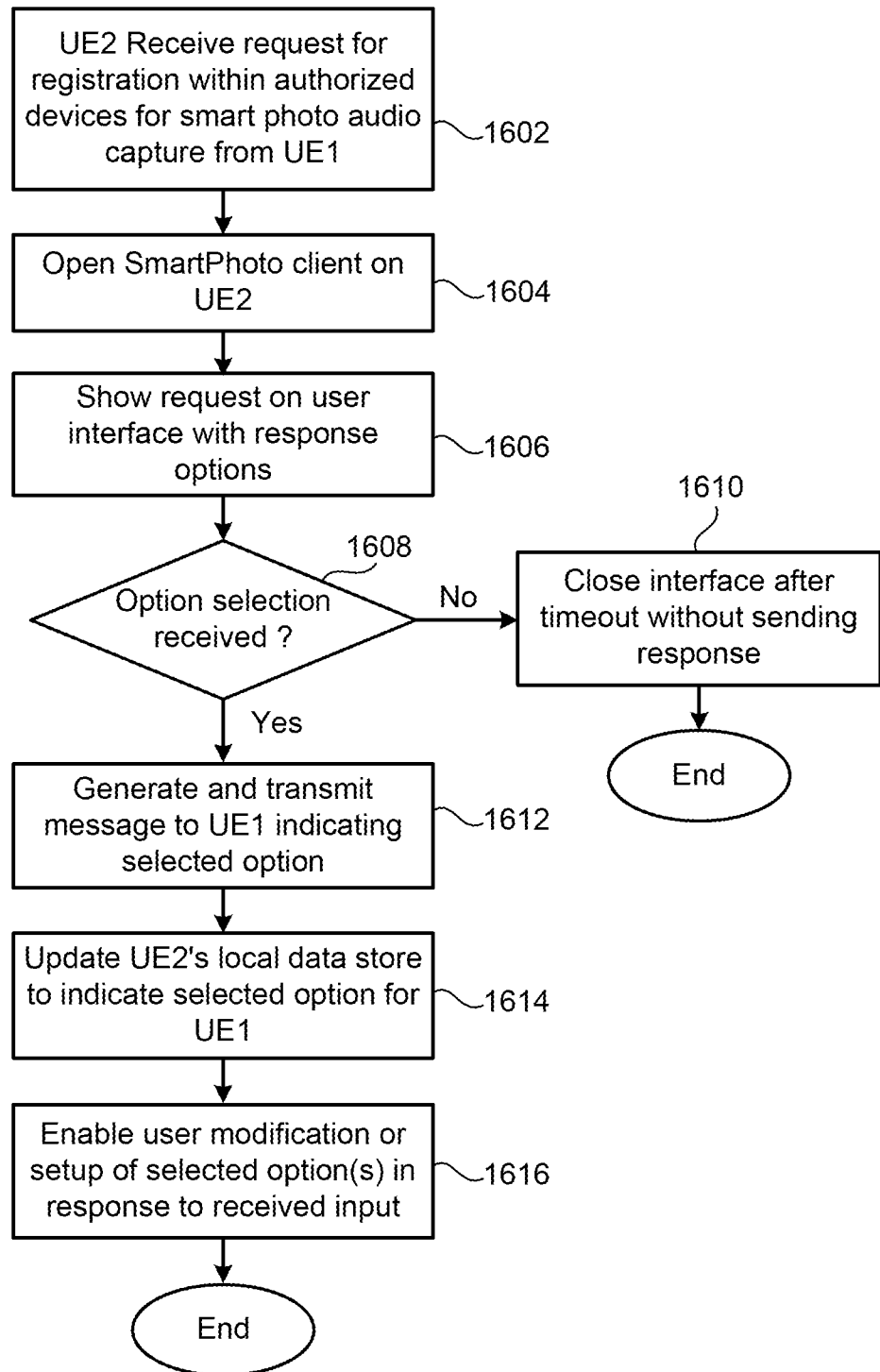
1600

FIG. 16

13/15

1700

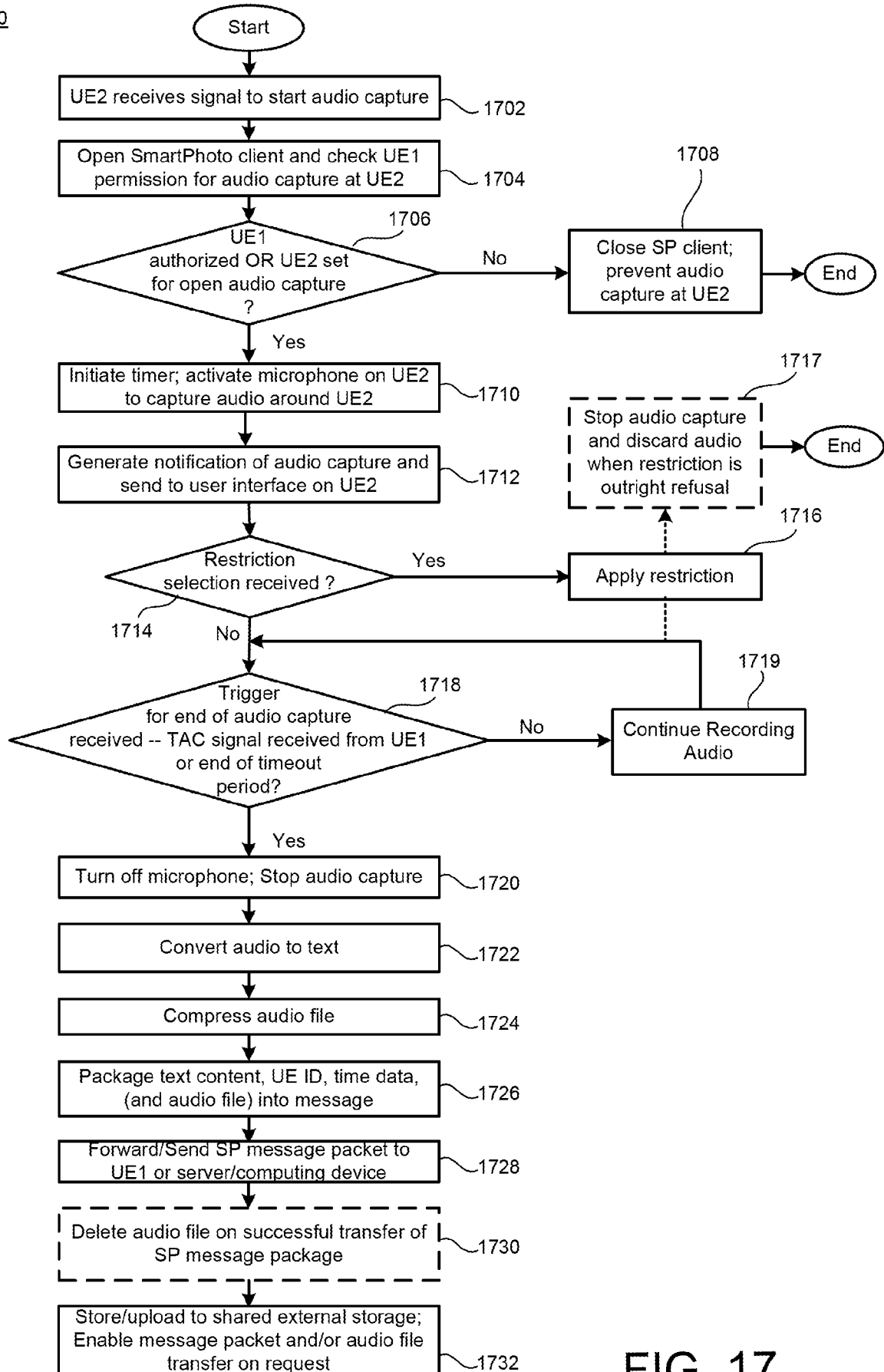


FIG. 17

14/15

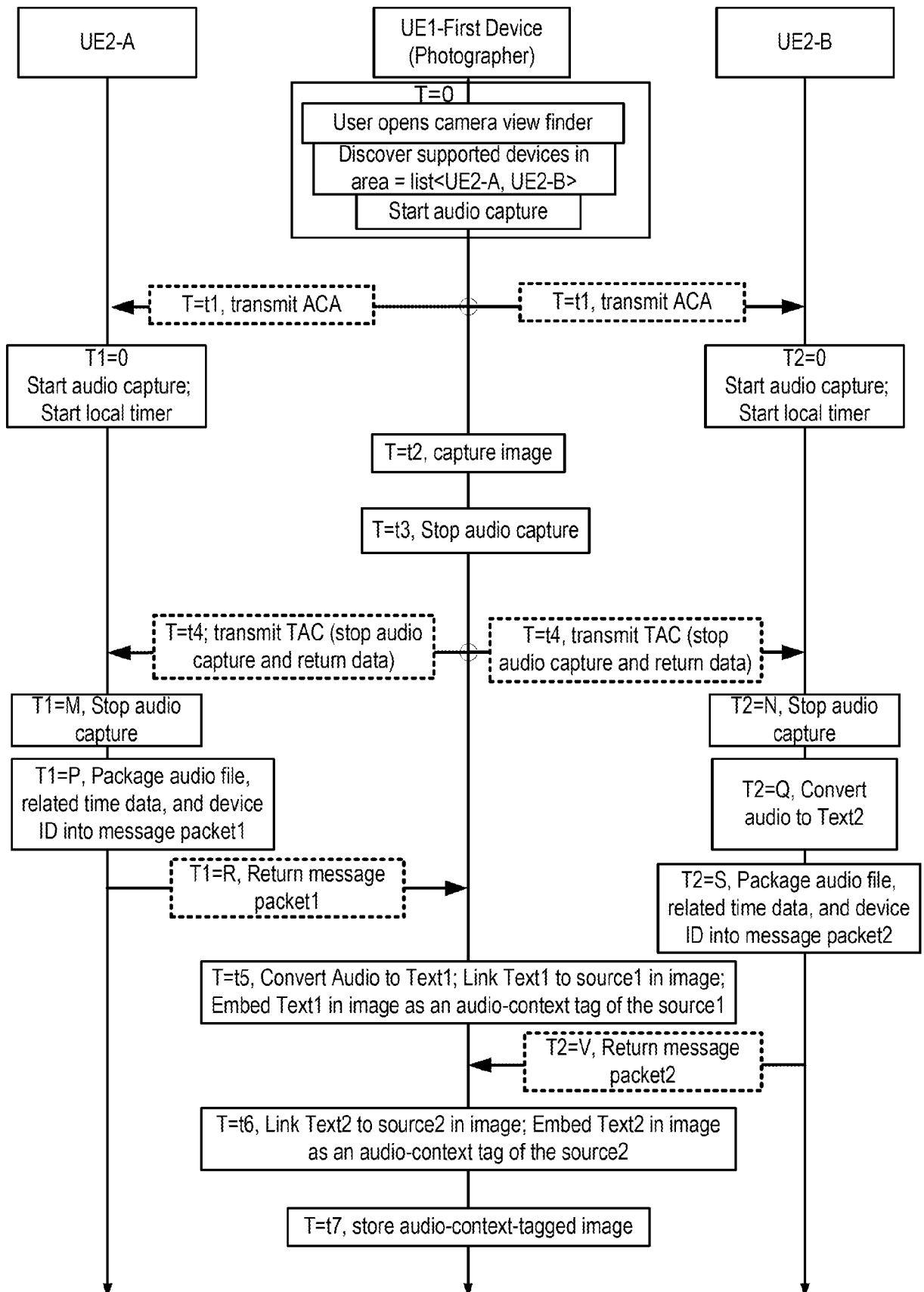


FIG. 18

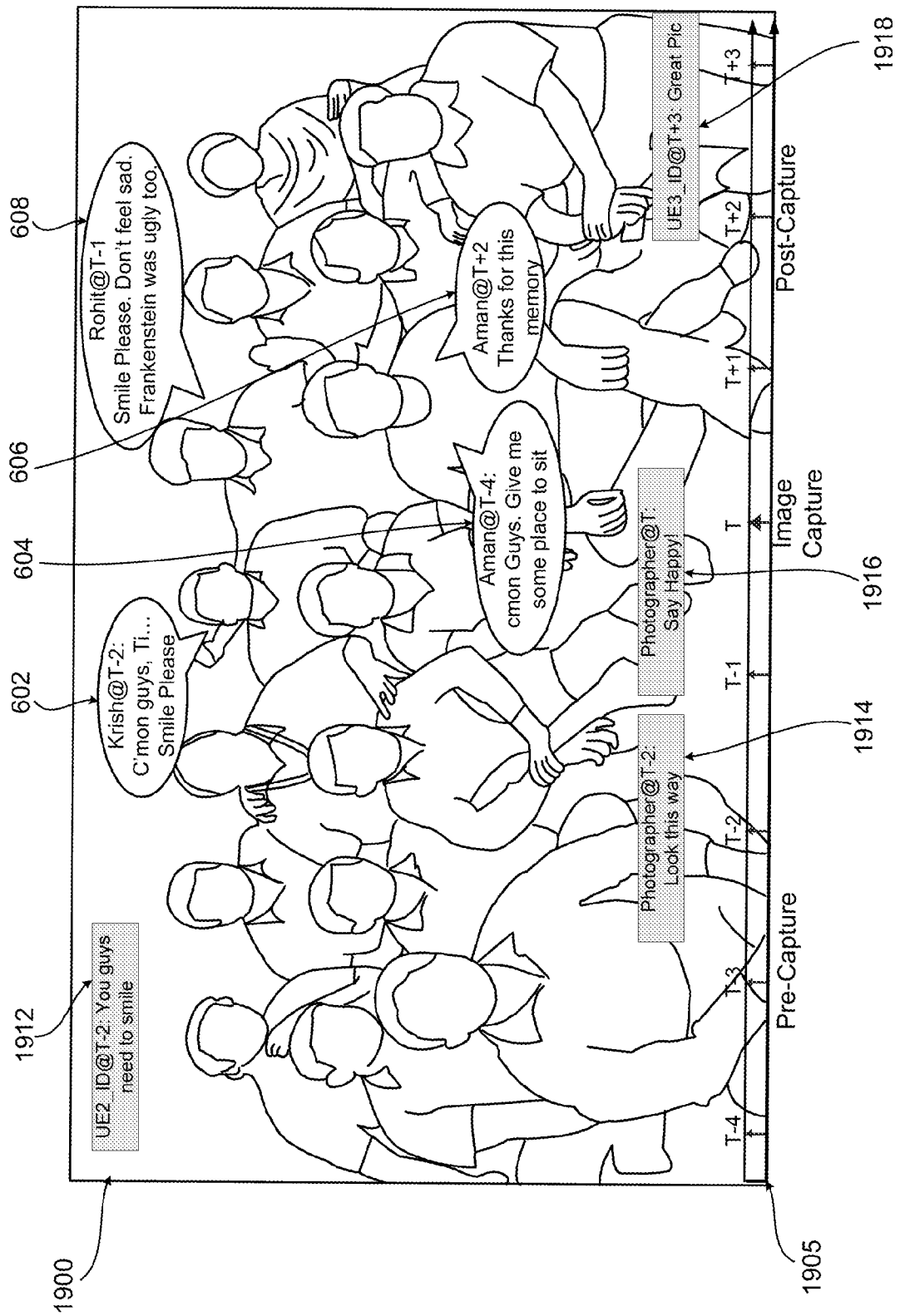


FIG. 19

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/US2014/017549

A. CLASSIFICATION OF SUBJECT MATTER  
INV. H04N5/262 G11B27/34 H04N5/265  
ADD.

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

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
H04N G11B

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

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

EPO-Internal, WPI Data

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category*        | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                                                                                                                                     | Relevant to claim No.                                                              |
|------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|
| X<br>A<br>A<br>A | <p>US 2010/238323 A1 (ENGLUND HAKAN [SE])<br/>23 September 2010 (2010-09-23)<br/>paragraphs [0002] - [0021], [0061] -<br/>[0064], [0117] - [0127]; figures 1-4, 8,<br/>16, 17</p> <p>-----</p> <p>US 2010/123797 A1 (YAMAMOTO YASUHIRO [JP])<br/>20 May 2010 (2010-05-20)<br/>abstract; figures 1, 3, 5</p> <p>-----</p> <p>US 2012/315013 A1 (WING TSE HONG [HK] ET<br/>AL) 13 December 2012 (2012-12-13)<br/>abstract; figures 4, 7</p> <p>-----</p> | <p>1-8, 11,<br/>15-20<br/>9, 10,<br/>12-14,<br/>21, 22</p> <p>1-22</p> <p>1-22</p> |



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

27 May 2014

Date of mailing of the international search report

06/06/2014

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

Trimeche, Mejdi

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2014/017549

| Patent document<br>cited in search report | Publication<br>date | Patent family<br>member(s) | Publication<br>date |
|-------------------------------------------|---------------------|----------------------------|---------------------|
| US 2010238323 A1                          | 23-09-2010          | EP 2411980 A1              | 01-02-2012          |
|                                           |                     | JP 5331936 B2              | 30-10-2013          |
|                                           |                     | JP 2012521705 A            | 13-09-2012          |
|                                           |                     | US 2010238323 A1           | 23-09-2010          |
|                                           |                     | WO 2010109274 A1           | 30-09-2010          |
| -----                                     |                     |                            |                     |
| US 2010123797 A1                          | 20-05-2010          | JP 2010124039 A            | 03-06-2010          |
|                                           |                     | US 2010123797 A1           | 20-05-2010          |
| -----                                     |                     |                            |                     |
| US 2012315013 A1                          | 13-12-2012          | CN 102932623 A             | 13-02-2013          |
|                                           |                     | US 2012315013 A1           | 13-12-2012          |
| -----                                     |                     |                            |                     |