



(51) International Patent Classification:
H04N 9/80 (2006.01)

(21) International Application Number:
PCT/US2015/041624

(22) International Filing Date:
22 July 2015 (22.07.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/028,254 23 July 2014 (23.07.2014) US
62/039,849 20 August 2014 (20.08.2014) US
14/513,149 13 October 2014 (13.10.2014) US
14/513,150 13 October 2014 (13.10.2014) US
14/513,151 13 October 2014 (13.10.2014) US
14/513,153 13 October 2014 (13.10.2014) US
14/705,864 6 May 2015 (06.05.2015) US

(71) Applicant: **GOPRO, INC.** [US/US]; 3000 Clearview Way, Bldg. E, San Mateo, CA 94402 (US).

(72) Inventors: **NEWMAN, David, A.**; C/o Gopro, Inc., 3000 Clearview Way, Bldg. E, San Mateo, CA 94402 (US). **SIL-**

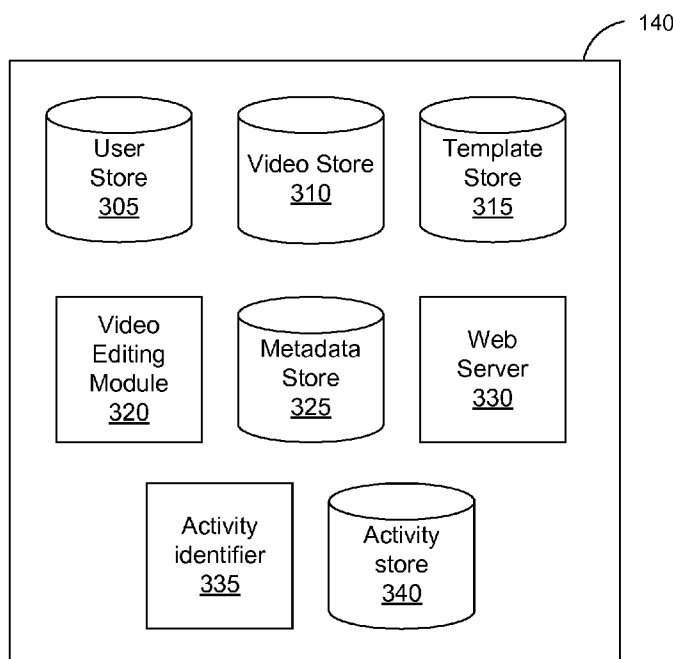
VER, Adam; C/o Gopro, Inc., 3000 Clearview Way, Bldg. E, San Mateo, CA 94402 (US). **HODULIK, Nick**; C/o Gopro, Inc., 3000 Clearview Way, Bldg. E, San Mateo, CA 94402 (US). **TAYLOR, Jonathan**; C/o Gopro, Inc., 3000 Clearview Way, Bldg. E, San Mateo, CA 94402 (US). **SCHUNCK, Brian, G.**; C/o Gopro, Inc., 3000 Clearview Way, Bldg. E, San Mateo, CA 94402 (US). **ADUMILLI, Balineedu, Chowdary**; C/o Gopro, Inc., 3000 Clearview Way, Bldg. E, San Mateo, CA 94402 (US). **LUSTIG, Ryan**; C/o Gopro, Inc., 3000 Clearview Way, Bldg. E, San Mateo, CA 94402 (US).

(74) Agents: **JACOBSON, Anthony** et al.; Fenwick & West LLP, Silicon Valley Center, 801 California Street, Mountain View, CA 94041 (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, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM,

[Continued on next page]

(54) Title: SCENE AND ACTIVITY IDENTIFICATION IN VIDEO SUMMARY GENERATION



(57) Abstract: Video and corresponding metadata is accessed. Events of interest within the video are identified based on the corresponding metadata, and best scenes are identified based on the identified events of interest. A video summary can be generated including one or more of the identified best scenes. The video summary can be generated using a video summary template with slots corresponding to video clips selected from among sets of candidate video clips. Best scenes can also be identified by receiving an indication of an event of interest within video from a user during the capture of the video. Metadata patterns representing activities identified within video clips can be identified within other videos, which can subsequently be associated with the identified activities.

FIG. 3



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.

SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE,

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

SCENE AND ACTIVITY IDENTIFICATION IN VIDEO SUMMARY GENERATION

CROSS REFERENCE TO RELATED APPLICATIONS

[0026] This application claims priority to, and the benefit of, U.S. Provisional Application No. 62/028,254, filed July 23, 2014, U.S. Provisional Application 62/039,849, filed August 20, 2014, U.S. Application No. 14/513,149, filed October 13, 2014, U.S. Application No. 14/513,150, filed October 13, 2014, U.S. Application No. 14/513,151, filed October 13, 2014, U.S. Application No. 14/513,153, filed October 13, 2014, and U.S. Application No. 14/705,864, filed May 6, 2015, which are incorporated by reference herein in their entirety.

BACKGROUND

TECHNICAL FIELD

[0001] This disclosure relates to a camera system, and more specifically, to processing video data captured using a camera system.

DESCRIPTION OF THE RELATED ART

[0002] Digital cameras are increasingly used to capture videos in a variety of settings, for instance outdoors or in a sports environment. However, as users capture increasingly more and longer videos, video management becomes increasingly difficult. Manually searching through raw videos (“scrubbing”) to identify the best scenes is extremely time consuming. Automated video processing to identify the best scenes can be very resource-intensive, particularly with high-resolution raw-format video data. Accordingly, an improved method of automatically identifying the best scenes in captured videos and generating video summaries including the identified best scenes can beneficially improve a user’s video editing experience.

BRIEF DESCRIPTIONS OF THE DRAWINGS

[0003] The disclosed embodiments have other advantages and features which will be more readily apparent from the following detailed description of the invention and the appended claims, when taken in conjunction with the accompanying drawings, in which:

[0004] Figure (or “Fig.”) 1 is a block diagram of a camera system environment according to one embodiment.

[0005] Fig. 2 is a block diagram illustrating a camera system, according to one embodiment.

[0006] Fig. 3 is a block diagram of a video server, according to one embodiment.

[0007] Fig. 4 is a flowchart illustrating a method for selecting video portions to include in a video summary, according to one embodiment.

[0008] Fig. 5 is a flowchart illustrating a method for generating video summaries using video templates, according to one embodiment.

[0009] Fig. 6 is a flowchart illustrating a method for generating video summaries of videos associated with user-tagged events, according to one embodiment.

[0010] Fig. 7 is a flowchart illustrating a method of identifying an activity associated with a video, according to one embodiment.

[0011] Fig. 8 is a flowchart illustrating a method of sharing a video based on an identified activity within the video, according to one embodiment.

DETAILED DESCRIPTION

[0012] The figures and the following description relate to preferred embodiments by way of illustration only. It should be noted that from the following discussion, alternative embodiments of the structures and methods disclosed herein will be readily recognized as viable alternatives that may be employed without departing from the principles of what is claimed.

[0013] Reference will now be made in detail to several embodiments, examples of which are illustrated in the accompanying figures. It is noted that wherever practicable similar or like reference numbers may be used in the figures and may indicate similar or like functionality. The figures depict embodiments of the disclosed system (or method) for purposes of illustration only. One skilled in the art will readily recognize from the following description that alternative embodiments of the structures and methods illustrated herein may be employed without departing from the principles described herein.

EXAMPLE CAMERA SYSTEM CONFIGURATION

[0014] Fig. 1 is a block diagram of a camera system environment, according to one embodiment. The camera system environment 100 includes one or more metadata sources 110, a network 120, a camera 130, a client device 135 and a video server 140. In alternative configurations, different and/or additional components may be included in the camera system environment 100. Examples of metadata sources 110 include sensors (such as accelerometers, speedometers, rotation sensors, GPS sensors, altimeters, and the like), camera inputs (such as

an image sensor, microphones, buttons, and the like), and data sources (such as external servers, web pages, local memory, and the like). Although not shown in Fig. 1, it should be noted that in some embodiments, one or more of the metadata sources 110 can be included within the camera 130.

[0015] The camera 130 can include a camera body having a camera lens structured on a front surface of the camera body, various indicators on the front of the surface of the camera body (such as LEDs, displays, and the like), various input mechanisms (such as buttons, switches, and touch-screen mechanisms), and electronics (e.g., imaging electronics, power electronics, metadata sensors, etc.) internal to the camera body for capturing images via the camera lens and/or performing other functions. As described in greater detail in conjunction with Fig. 2 below, the camera 130 can include sensors to capture metadata associated with video data, such as motion data, speed data, acceleration data, altitude data, GPS data, and the like. A user uses the camera 130 to record or capture videos in conjunction with associated metadata which the user can edit at a later time.

[0016] The video server 140 receives and stores videos captured by the camera 130 allowing a user to access the videos at a later time. In one embodiment, the video server 140 provides the user with an interface, such as a web page or native application installed on the client device 135, to interact with and/or edit the videos captured by the user. In one embodiment, the video server 140 generates video summaries of various videos stored at the video server, as described in greater detail in conjunction with Fig. 3 and Fig. 4 below. As used herein, “video summary” refers to a generated video including portions of one or more other videos. A video summary often includes highlights (or “best scenes”) of a video captured by a user. In some embodiments, best scenes include events of interest within the captured video, scenes associated with certain metadata (such as an above threshold altitude or speed), scenes associated with certain camera or environment characteristics, and the like. For example, in a video captured during a snowboarding trip, the best scenes in the video can include jumps performed by the user or crashes in which the user was involved. In addition to including one or more highlights of the video, a video summary can also capture the experience, theme, or story associated with the video without requiring significant manual editing by the user. In one embodiment, the video server 140 identifies the best scenes in raw video based on the metadata associated with the video. The video server 140 may then generate a video summary using the identified best scenes of the video. The metadata can either be captured by the camera 130 during the capture of the video or can be retrieved from one or more metadata sources 110 after the capture of the video.

[0017] Metadata includes information about the video itself, the camera used to capture the video, the environment or setting in which a video is captured or any other information associated with the capture of the video. For example, metadata can include acceleration data representative of the acceleration of a camera 130 attached to a user as the user captures a video while snowboarding down a mountain. Such acceleration metadata helps identify events representing a sudden change in acceleration during the capture of the video, such as a crash the user may encounter or a jump the user performs. Thus, metadata associated with captured video can be used to identify best scenes in a video recorded by a user without relying on image processing techniques or manual curation by a user.

[0018] Examples of metadata include: telemetry data (such as motion data, velocity data, and acceleration data) captured by sensors on the camera 130; location information captured by a GPS receiver of the camera 130; compass heading information; altitude information of the camera 130; biometric data such as the heart rate of the user, breathing of the user, eye movement of the user, body movement of the user, and the like; vehicle data such as the velocity or acceleration of the vehicle, the brake pressure of the vehicle, or the rotations per minute (RPM) of the vehicle engine; or environment data such as the weather information associated with the capture of the video. The video server 140 may receive metadata directly from the camera 130 (for instance, in association with receiving video from the camera), from a client device 135 (such as a mobile phone, computer, or vehicle system associated with the capture of video), or from external metadata sources 110 such as web pages, blogs, databases, social networking sites, or servers or devices storing information associated with the user (e.g., a user may use a fitness device recording fitness data).

[0019] A user can interact with interfaces provided by the video server 140 via the client device 135. The client device 135 is any computing device capable of receiving user inputs as well as transmitting and/or receiving data via the network 120. In one embodiment, the client device 135 is a conventional computer system, such as a desktop or a laptop computer. Alternatively, the client device 135 may be a device having computer functionality, such as a personal digital assistant (PDA), a mobile telephone, a smartphone or another suitable device. The user can use the client device to view and interact with or edit videos stored on the video server 140. For example, the user can view web pages including video summaries for a set of videos captured by the camera 130 via a web browser on the client device 135.

[0020] One or more input devices associated with the client device 135 receive input from the user. For example, the client device 135 can include a touch-sensitive display, a keyboard, a trackpad, a mouse, a voice recognition system, and the like. In some embodiments, the

client device 135 can access video data and/or metadata from the camera 130 or one or more metadata sources 110, and can transfer the accessed metadata to the video server 140. For example, the client device may retrieve videos and metadata associated with the videos from the camera via a universal serial bus (USB) cable coupling the camera 130 and the client device 135. The client device can then upload the retrieved videos and metadata to the video server 140.

[0021] In one embodiment, the client device 135 executes an application allowing a user of the client device 135 to interact with the video server 140. For example, a user can identify metadata properties using an application executing on the client device 135, and the application can communicate the identified metadata properties selected by a user to the video server 140 to generate and/or customize a video summary. As another example, the client device 135 can execute a web browser configured to allow a user to select video summary properties, which in turn can communicate the selected video summary properties to the video server 140 for use in generating a video summary. In one embodiment, the client device 135 interacts with the video server 140 through an application programming interface (API) running on a native operating system of the client device 135, such as IOS® or ANDROID™. While FIG. 1 shows a single client device 135, in various embodiments, any number of client devices 135 may communicate with the video server 140.

[0022] The video server 140 communicates with the client device 135, the metadata sources 110, and the camera 130 via the network 120, which may include any combination of local area and/or wide area networks, using both wired and/or wireless communication systems. In one embodiment, the network 120 uses standard communications technologies and/or protocols. In some embodiments, all or some of the communication links of the network 120 may be encrypted using any suitable technique or techniques. It should be noted that in some embodiments, the video server 140 is located within the camera 130 itself.

EXAMPLE CAMERA CONFIGURATION

[0023] Fig. 2 is a block diagram illustrating a camera system, according to one embodiment. The camera 130 includes one or more microcontrollers 202 (such as microprocessors) that control the operation and functionality of the camera 130. A lens and focus controller 206 is configured to control the operation and configuration of the camera lens. A system memory 204 is configured to store executable computer instructions that, when executed by the microcontroller 202, perform the camera functionalities described herein. A synchronization interface 208 is configured to synchronize the camera 130 with other cameras or with other

external devices, such as a remote control, a second camera 130, a smartphone, a client device 135, or a video server 140.

[0024] A controller hub 230 transmits and receives information from various I/O components. In one embodiment, the controller hub 230 interfaces with LED lights 236, a display 232, buttons 234, microphones such as microphones 222, speakers, and the like.

[0025] A sensor controller 220 receives image or video input from an image sensor 212. The sensor controller 220 receives audio inputs from one or more microphones, such as microphone 212a and microphone 212b. Metadata sensors 224, such as an accelerometer, a gyroscope, a magnetometer, a global positioning system (GPS) sensor, or an altimeter may be coupled to the sensor controller 220. The metadata sensors 224 each collect data measuring the environment and aspect in which the video is captured. For example, the accelerometer 220 collects motion data, comprising velocity and/or acceleration vectors representative of motion of the camera 130, the gyroscope provides orientation data describing the orientation of the camera 130, the GPS sensor provides GPS coordinates identifying the location of the camera 130, and the altimeter measures the altitude of the camera 130. The metadata sensors 224 are rigidly coupled to the camera 130 such that any motion, orientation or change in location experienced by the camera 130 is also experienced by the metadata sensors 224. The sensor controller 220 synchronizes the various types of data received from the various sensors connected to the sensor controller 220. For example, the sensor controller 220 associates a time stamp representing when the data was captured by each sensor. Thus, using the time stamp, the measurements received from the metadata sensors 224 are correlated with the corresponding video frames captured by the image sensor 212. In one embodiment, the sensor controller begins collecting metadata from the metadata sources when the camera 130 begins recording a video. In one embodiment, the sensor controller 220 or the microcontroller 202 performs operations on the received metadata to generate additional metadata information. For example, the microcontroller may integrate the received acceleration data to determine the velocity profile of the camera 130 during the recording of a video.

[0026] Additional components connected to the microcontroller 202 include an I/O port interface 238 and an expansion pack interface 240. The I/O port interface 238 may facilitate the receiving or transmitting video or audio information through an I/O port. Examples of I/O ports or interfaces include USB ports, HDMI ports, Ethernet ports, audioports, and the like. Furthermore, embodiments of the I/O port interface 238 may include wireless ports that can accommodate wireless connections. Examples of wireless ports include Bluetooth, Wireless USB, Near Field Communication (NFC), and the like. The expansion pack interface 240 is

configured to interface with camera add-ons and removable expansion packs, such as a display module, an extra battery module, a wireless module, and the like.

EXAMPLE VIDEO SERVER ARCHITECTURE

[0027] FIG. 3 is a block diagram of an architecture of the video server. The video server 140 in the embodiment of FIG. 3 includes a user storage module 305 (“user store” hereinafter), a video storage module 310 (“video store” hereinafter), a template storage module 315 (“template store” hereinafter), a video editing module 320, a metadata storage module 325 (“metadata store” hereinafter), a web server 330, an activity identifier 335, and an activity storage module 340 (“activity store” hereinafter). In other embodiments, the video server 140 may include additional, fewer, or different components for performing the functionalities described herein. Conventional components such as network interfaces, security functions, load balancers, failover servers, management and network operations consoles, and the like are not shown so as to not obscure the details of the system architecture.

[0028] Each user of the video server 140 creates a user account, and user account information is stored in the user store 305. A user account includes information provided by the user (such as biographic information, geographic information, and the like) and may also include additional information inferred by the video server 140 (such as information associated with a user’s previous use of a camera). Examples of user information include a username, a first and last name, contact information, a user’s hometown or geographic region, other location information associated with the user, and the like. The user store 305 may include data describing interactions between a user and videos captured by the user. For example, a user account can include a unique identifier associating videos uploaded by the user with the user’s user account.

[0029] The video store 310 stores videos captured and uploaded by users of the video server 140. The video server 140 may access videos captured using the camera 130 and store the videos in the video store 310. In one example, the video server 140 may provide the user with an interface executing on the client device 135 that the user may use to upload videos to the video store 315. In one embodiment, the video server 140 indexes videos retrieved from the camera 130 or the client device 135, and stores information associated with the indexed videos in the video store. For example, the video server 140 provides the user with an interface to select one or more index filters used to index videos. Examples of index filters include but are not limited to: the type of equipment used by the user (e.g., ski equipment, mountain bike equipment, etc.), the type of activity being performed by the user while the

video was captured (e.g., snowboarding, mountain biking, etc.), the time and data at which the video was captured, or the type of camera 130 used by the user.

[0030] In some embodiments, the video server 140 generates a unique identifier for each video stored in the video store 310. In some embodiments, the generated identifier for a particular video is unique to a particular user. For example, each user can be associated with a first unique identifier (such as a 10-digit alphanumeric string), and each video captured by a user is associated with a second unique identifier made up of the first unique identifier associated with the user concatenated with a video identifier (such as an 8-digit alphanumeric string unique to the user). Thus, each video identifier is unique among all videos stored at the video store 310, and can be used to identify the user that captured the video.

[0031] The metadata store 325 stores metadata associated with videos stored by the video store 310. For instance, the video server 140 can retrieve metadata from the camera 130, the client device 135, or one or more metadata sources 110, can associate the metadata with the corresponding video (for instance by associating the metadata with the unique video identifier), and can store the metadata in the metadata store 325. The metadata store 325 can store any type of metadata, including but not limited to the types of metadata described herein. It should be noted that in some embodiments, metadata corresponding to a video is stored within a video file itself, and not in a separate storage module.

[0032] The web server 330 provides a communicative interface between the video server 140 and other entities of the environment of Fig. 1. For example, the web server 330 can access videos and associated metadata from the camera 130 or the client device 135 to store in the video store 310 and the metadata store 325, respectively. The web server 330 can also receive user input provided to the client device 135, can request video summary templates or other information from a client device 135 for use in generating a video summary, and can provide a generated video summary to the client device or another external entity.

EVENT OF INTEREST/ACTIVITY IDENTIFICATION

[0033] The video editing module 320 analyzes metadata associated with a video to identify best scenes of the video based on identified events of interest or activities, and generates a video summary including one or more of the identified best scenes of the video. The video editing module 320 first accesses one or more videos from the video store 310, and accesses metadata associated with the accessed videos from the metadata store 325. The video editing module 320 then analyzes the metadata to identify events of interest in the metadata.

Examples of events of interest can include abrupt changes or anomalies in the metadata, such

as a peak or valley in metadata maximum or minimum values within the metadata,, metadata exceeding or falling below particular thresholds, metadata within a threshold of predetermine values (for instance, within 20 meters of a particular location or within), and the like. The video editing module 320 can identify events of interest in videos based on any other type of metadata, such as a heart rate of a user, orientation information, and the like.

[0034] For example, the video editing module 320 can identify any of the following as an event of interest within the metadata: a greater than threshold change in acceleration or velocity within a pre-determined period of time, a maximum or above-threshold velocity or acceleration, a maximum or local maximum altitude, a maximum or above-threshold heart rate or breathing rate of a user, a maximum or above-threshold audio magnitude, a user location within a pre-determined threshold distance from a pre-determined location, a threshold change in or pre-determined orientation of the camera or user, a proximity to another user or location, a time within a threshold of a pre-determined time, a pre-determined environmental condition (such as a particular weather event, a particular temperature, a sporting event, a human gathering, or any other suitable event), or any other event associated with particular metadata.

[0035] In some embodiments, a user can manually indicate an event of interest during capture of the video. For example, a user can press a button on the camera or a camera remote or otherwise interact with the camera during the capture of video to tag the video as including an event of interest. The manually tagged event of interest can be indicated within metadata associated with the captured video. For example, if a user is capturing video while snowboarding and presses a camera button associated with manually tagging an event of interest, the camera creates metadata associated with the captured video indicating that the video includes an event of interest, and indicating a time or portion within the captured video at which the tagged event of interest occurs. In some embodiments, the manual tagging of an event of interest by a user while capturing video is stored as a flag within a resulting video file. The location of the flag within the video file corresponds to a time within the video at which the user manually tags the event of interest.

[0036] In some embodiments, a user can manually indicate an event of interest during capture of the video using a spoken command or audio signal. For instance, a user can say “Tag” or “Tag my moment” during the capture of video to tag the video as including an event of interest. The audio-tagged event of interest can be indicated within metadata associated with the captured video. The spoken command can be pre-programmed, for instance by a manufacturer, programmer, or seller of the camera system, or can be customized by a user of

the camera system. For instance, a user can speak a command or other audio signal into a camera during a training period (for instance, in response to configuring the camera into a training mode, or in response to the selection of a button or interface option associated with training a camera to receive a spoken command). The spoken command or audio signal can be repeated during the training mode a threshold number of times (such as once, twice, or any number of times necessary for the purposes of identifying audio patterns as described herein), and the camera system can identify an audio pattern associated with the spoken commands or audio signals received during the training period. The audio pattern is then stored at the camera, and, during a video capture configuration, the camera can identify the audio pattern in a spoken command or audio signal received from a user of the camera, and can manually tag an event of interest during the capture of video in response to detecting the stored audio pattern within the received spoken command or audio signal. In some embodiments, the audio pattern is specific to spoken commands or audio signals received from a particular user and can be detected only in spoken commands or audio signals received from the particular user. In other embodiments, the audio pattern can be identified within spoken commands or audio signals received from any user. It should be noted that manually identified events of interest can be associated with captured video by the camera itself, and can be identified by a system to which the captured video is uploaded from the camera without significant additional post-processing.

[0037] As noted above, the video editing module 320 can identify events of interest based on activities performed by users when the videos are captured. For example, a jump while snowboarding or a crash while skateboarding can be identified as events of interest. Activities can be identified by the activity identifier module 335 based on metadata associated with the video captured while performing the activities. Continuing with the previous example, metadata associated with a particular altitude and a parabolic upward and then downward velocity can be identified as a “snowboarding jump”, and a sudden slowdown in velocity and accompanying negative acceleration can be identified as a “skateboarding crash”.

[0038] The video editing module 320 can identify events of interest based on audio captured in conjunction with the video. In some embodiments, the video editing module identifies events of interest based on one or more spoken words or phrases in captured audio. For example, if audio of a user saying “Holy Smokes!” is captured, the video editing module can determine that an event of interest just took place (e.g., within the previous 5 seconds or other threshold of time), and if audio of a user saying “Oh no! Watch out!” is captured, the video editing module can determine that an event of interest is about to occur (e.g., within the next

5 seconds or other threshold of time). In addition to identifying events of interest based on captured dialogue, the video editing module can identify an event of identify based on captured sound effects, captured audio exceeding a magnitude or pitch threshold, or captured audio satisfying any other suitable criteria.

[0039] In some embodiments, the video editing module 320 can identify video that does not include events of interest. For instance, the video editing module 320 can identify video that is associated with metadata patterns determined to not be of interest to a user. Such patterns can include metadata associated with a below-threshold movement, a below-threshold luminosity, a lack of faces or other recognizable objects within the video, audio data that does not include dialogue or other notable sound effects, and the like. In some embodiments, video determined to not include events of interest can be disqualified from consideration for inclusion in a generated video summary, or can be hidden from a user viewing captured video (in order to increase the chance that the remaining video presented to the user does include events of interest).

[0040] The activity identifier module 335 can receive a manual identification of an activity within videos from one or more users. In some embodiments, activities can be tagged during the capture of video. For instance, if a user is about to capture video while performing a snowboarding jump, the user can manually tag the video being captured or about to be captured as “snowboarding jump”. In some embodiments, activities can be tagged after the video is captured, for instance during playback of the video. For instance, a user can tag an activity in a video as a skateboarding crash upon playback of the video.

[0041] Activity tags in videos can be stored within metadata associated with the videos. For videos stored in the video store 310, the metadata including activity tags associated with the videos is stored in the metadata store 325. In some embodiments, the activity identifier module 335 identifies metadata patterns associated with particular activities and/or activity tags. For instance, metadata associated with several videos tagged with the activity “skydiving” can be analyzed to identify similarities within the metadata, such as a steep increase in acceleration at a high altitude followed by a high velocity at decreasing altitudes. Metadata patterns associated with particular activities are stored in the activity store 340.

[0042] In some embodiments, metadata patterns associated with particular activities can include audio data patterns. For instance, particular sound effects, words or phrases of dialogue, or the like can be associated with particular activities. For example, the spoken phrase “nice wave” can be associated with surfing, and the sound of a revving car engine can be associated with driving or racing a vehicle. In some embodiments, metadata patterns used

to identify activities can include the use of particular camera mounts associated with the activities in capturing video. For example, a camera can detect that it is coupled to a snowboard mount, and video captured while coupled to the snowboard mount can be associated with the activity of snowboarding.

[0043] Once metadata patterns associated with particular activities are identified, the activity identifier module 335 can identify metadata patterns in metadata associated with other videos, and can tag or associate other videos associated with metadata including the identified metadata patterns with the activities associated with the identified metadata patterns. The activity identifier module 335 can identify and store a plurality of metadata patterns associated with a plurality of activities within the activity store 340. Metadata patterns stored in the activity store 340 can be identified within videos captured by one user, and can be used by the activity identifier module 335 to identify activities within videos captured by the user. Alternatively, metadata patterns can be identified within videos captured by a first plurality of users, and can be used by the activity identifier module 335 to identify activities within videos captured by a second plurality of users including at least one user not in the first plurality of users. In some embodiments, the activity identifier module 335 aggregates metadata for a plurality of videos associated with an activity and identifies metadata patterns based on the aggregated metadata. As used herein, “tagging” a video with an activity refers to the association of the video with the activity. Activities tagged in videos can be used as a basis to identify best scenes in videos (as described above), and to select video clips for inclusion in video summary templates (as described below).

[0044] Videos tagged with activities can be automatically uploaded to or shared with an external system. For instance, if a user captures video, the activity identifier module 335 can identify a metadata pattern associated with an activity in metadata of the captured video, in real-time (as the video is being captured), or after the video is captured (for instance, after the video is uploaded to the video server 140). The video editing module 320 can select a portion of the captured video based on the identified activity, for instance a threshold amount of time or frames around a video clip or frame associated with the identified activity. The selected video portion can be uploaded or shared to an external system, for instance via the web server 330. The uploading or sharing of video portions can be based on one or more user settings and/or the activity identified. For instance, a user can select one or more activities in advance of capturing video, and captured video portions identified as including the selected activities can be uploaded automatically to an external system, and can be automatically shared via one or more social media outlets.

BEST SCENE IDENTIFICATION AND VIDEO SUMMARY GENERATION

[0045] The video editing module 320 identifies best scenes associated with the identified events of interest for inclusion in a video summary. Each best scene is a video clip, portion, or scene (“video clips” hereinafter), and can be an entire video or a portion of a video. For instance, the video editing module 320 can identify video clips occurring within a threshold amount of time of an identified event of interest (such as 3 seconds before and after the event of interest), within a threshold number of frames of an identified event of interest (such as 24 frames before and after the event of interest), and the like. The amount of length of a best scene can be pre-determined, and/or can be selected by a user.

[0046] The amount or length of video clip making up a best scene can vary based on an activity associated with captured video, based on a type or value of metadata associated with captured video, based on characteristics of the captured video, based on a camera mode used to capture the video, or any other suitable characteristic. For example, if an identified event of interest is associated with an above-threshold velocity, the video editing module 320 can identify all or part of the video corresponding to above-threshold velocity metadata as the best scene. In another example, the length of a video clip identified as a best scene can be greater for events of interest associated with maximum altitude values than for events of interest associated with proximity to a pre-determined location.

[0047] For events of interest manually tagged by a user, the length of a video clip identified as a best scene can be pre-defined by the user, can be manually selected by the user upon tagging the event of interest, can be longer than automatically-identified events of interest, can be based on a user-selected tagging or video capture mode, and the like. The amount or length of video clips making up best scenes can vary based on the underlying activity represented in captured video. For instance, best scenes associated with events of interest in videos captured while boating can be longer than best scenes associated with events of interest in videos captured while skydiving.

[0048] The identified video portions make up the best scenes as described herein. The video editing module 320 generates a video summary by combining or concatenating some or all of the identified best scenes into a single video. The video summary thus includes video portions of events of interest, beneficially resulting in a playable video including scenes likely to be of greatest interest to a user. The video editing module 320 can receive one or more video summary configuration selections from a user, each specifying one or more properties of the video summary (such as a length of a video summary, a number of best scenes for inclusion in the video summary, and the like), and can generate the video summary

according to the one or more video summary configuration selections. In some embodiments, the video summary is a renderable or playable video file configured for playback on a viewing device (such as a monitor, a computer, a mobile device, a television, and the like). The video summary can be stored in the video store 310, or can be provided by the video server 140 to an external entity for subsequent playback. Alternatively, the video editing module 320 can serve the video summary from the video server 140 by serving each best scene directly from a corresponding best scene video file stored in the video store 310 without compiling a singular video summary file prior to serving the video summary. It should be noted that the video editing module 320 can apply one or more edits, effects, filters, and the like to one or more best scenes within the video summary, or to the entire video summary during the generation of the video summary.

[0049] In some embodiments, the video editing module 320 ranks identified best scenes. For instance, best scenes can be ranked based on activities with which they are associated, based on metadata associated with the best scenes, based on length of the best scenes, based on a user-selected preference for characteristics associated with the best scenes, or based on any other suitable criteria. For example, longer best scenes can be ranked higher than shorter best scenes. Likewise, a user can specify that best scenes associated with above-threshold velocities can be ranked higher than best scenes associated with above-threshold heart rates. In another example, best scenes associated with jumps or crashes can be ranked higher than best scenes associated with sitting down or walking. Generating a video summary can include identifying and including the highest ranked best scenes in the video summary.

[0050] In some embodiments, the video editing module 320 classifies scenes by generating a score associated with each of one or more video classes based on metadata patterns associated with the scenes. Classes can include but are not limited to: content-related classes (“snow videos”, “surfing videos”, etc.), video characteristic classes (“high motion videos”, “low light videos”, etc.), video quality classes, mode of capture classes (based on capture mode, mount used, etc.), sensor data classes (“high velocity videos”, “high acceleration videos”, etc.), audio data classes (“human dialogue videos”, “loud videos”, etc.), number of cameras used (“single-camera videos”, “multi-camera videos”, etc.), activity identified within the video, and the like. Scenes can be scored for one or more video classes, the scores can be weighted based on a pre-determined or user-defined class importance scale, and the scenes can be ranked based on the scores generated for the scenes.

[0051] In one example, the video editing module 320 analyzes metadata associated with accessed videos chronologically to identify an order of events of interest presented within the

video. For example, the video editing module 320 can analyze acceleration data to identify an ordered set of video clips associated with acceleration data exceeding a particular threshold. In some embodiments, the video editing module 320 can identify an ordered set of events occurring within a pre-determined period of time. Each event in the identified set of events can be associated with a best scene; if the identified set of events is chronologically ordered, the video editing module 320 can generate a video summary by a combining video clips associated with each identified event in the order of the ordered set of events.

[0052] In some embodiments, the video editing module 320 can generate a video summary for a user using only videos associated with (or captured by) the user. To identify such videos, the video editing module 320 can query the video store 310 to identify videos associated with the user. In some embodiments, each video captured by all users of the video server 140 includes a unique identifier identifying the user that captured the video and identifying the video (as described above). In such embodiments, the video editing module 320 queries the video store 310 with an identifier associated with a user to identify videos associated with the user. For example, if all videos associated with User A include a unique identifier that starts with the sequence “X1Y2Z3” (an identifier unique to User A), the video editing module 320 can query the video store 310 using the identifier “X1Y2Z3” to identify all videos associated with User A. The video editing module 320 can then identify best scenes within such videos associated with a user, and can generate a video summary including such best scenes as described herein.

[0053] In addition to identifying best scenes, the video editing module 320 can identify one or more video frames that satisfy a set of pre-determined criteria for inclusion in a video summary, or for flagging to a user as candidates for saving as images/photograph stills. The pre-determined criteria can include metadata criteria, including but not limited to: frames with high motion (or blur) in a first portion of a frame and low motion (or blur) in another portion of a frame, frames associated with particular audio data (such as audio data above a particular magnitude threshold or audio data associated with voices or screaming), frames associated with above-threshold acceleration data, or frames associated with metadata that satisfies any other metadata criteria as described herein. In some embodiments, users can specify metadata criteria for use in flagging one or more video frames that satisfy pre-determined criteria. Similarly, in some embodiments, the video editing module 320 can identify metadata patterns or similarities in frames selected by a user to save as images/photograph stills, and can identify subsequent video frames that include the identified

metadata patterns or similarities for flagging as candidates to save as images/photograph stills.

VIDEO SUMMARY TEMPLATES

[0054] In one embodiment, the video editing module 320 retrieves video summary templates from the template store 315 to generate a video summary. The template store 315 includes video summary templates each describing a sequence of video slots for including in a video summary. In one example, each video summary template may be associated with a type of activity performed by the user while capturing video or the equipment used by the user while capturing video. For example, a video summary template for generating video summaries of a ski tip can differ from the video summary template for generating video summaries of a mountain biking trip.

[0055] Each slot in a video summary template is a placeholder to be replaced by a video clip or scene when generating a video summary. Each slot in a video summary template can be associated with a pre-defined length, and the slots collectively can vary in length. The slots can be ordered within a template such that once the slots are replaced with video clips, playback of the video summary results in the playback of the video clips in the order of the ordered slots replaced by the video clips. For example, a video summary template may include an introductory slot, an action slot, and a low-activity slot. When generating the video summary using such a template, a video clip can be selected to replace the introductory slot, a video clip of a high-action event can replace the action slot, and a video clip of a low-action event can replace the low-activity slot. It should be noted that different video summary templates can be used to generate video summaries of different lengths or different kinds.

[0056] In some embodiments, video summary templates include a sequence of slots associated with a theme or story. For example, a video summary template for a ski trip may include a sequence of slots selected to present the ski trip narratively or thematically. In some embodiments, video summary templates include a sequence of slots selected based on an activity type. For example, a video summary template associated with surfing can include a sequence of slots selected to highlight the activity of surfing.

[0057] Each slot in a video summary template can identify characteristics of a video clip to replace the slot within the video summary template, and a video clip can be selected to replace the slot based on the identified characteristics. For example, a slot can identify one or more of the following video clip characteristics: motion data associated with the video clip, altitude information associated with the video clip, location information associated with the

video clip, weather information associated with the clip, or any other suitable video characteristic or metadata value or values associated with a video clip. In these embodiments, a video clip having one or more of the characteristics identified by a slot can be selected to replace the slot.

[0058] In some embodiments, a video clip can be selected based on a length associated with a slot. For instance, if a video slot specifies a four-second length, a four-second (give or take a pre-determined time range, such as 0.5 seconds) video clip can be selected. In some embodiments, a video clip shorter than the length associated with a slot can be selected, and the selected video clip can replace the slot, reducing the length of time taken by the slot to be equal to the length of the selected video clip. Similarly, a video clip longer than the length associated with a slot can be selected, and either 1) the selected video clip can replace the slot, expanding the length of time associated with the slot to be equal to the length of the selected video clip, or 2) a portion of the selected video clip equal to the length associated with the slot can be selected and used to replace the slot. In some embodiments, the length of time of a video clip can be increased or decreased to match the length associated with a slot by adjusting the frame rate of the video clip to slow down or speed up the video clip, respectively. For example, to increase the amount of time taken by a video clip by 30%, 30% of the frames within the video clip can be duplicated. Likewise, to decrease the amount of time taken by a video clip by 60%, 60% of the frames within the video clip can be removed.

[0059] To generate a video summary using a video summary template, the video editing module 320 accesses a video summary template from the template store 315. The accessed video summary template can be selected by a user, can be automatically selected (for instance, based on an activity type or based on characteristics of metadata or video for use in generating the video summary), or can be selected based on any other suitable criteria. The video editing module 320 then selects a video clip for each slot in the video summary template, and inserts the selected video clips into the video summary in the order of the slots within the video summary template.

[0060] To select a video clip for each slot, the video editing module 320 can identify a set of candidate video clips for each slot, and can select from the set of candidate video clips (for instance, by selecting the determined best video from the set of candidate video clips according to the principles described above). In some embodiments, selecting a video clip for a video summary template slot identifying a set of video characteristics includes selecting a video clip from a set of candidate video clips that include the identified video characteristics. For example, if a slot identifies a video characteristic of “velocity over 15mph”, the video

editing module 320 can select a video clip associated with metadata indicating that the camera or a user of the camera was traveling at a speed of over 15 miles per hour when the video was captured, and can replace the slot within the video summary template with the selected video clip.

[0061] In some embodiments, video summary template slots are replaced by video clips identified as best scenes (as described above). For instance, if a set of candidate video clips are identified for each slot in a video summary template, if one of the candidate video clips identified for a slot is determined to be a best scene, the best scene is selected to replace the slot. In some embodiments, multiple best scenes are identified for a particular slot; in such embodiments, one of the best scenes can be selected for inclusion into the video summary based on characteristics of the best scenes, characteristics of the metadata associated with the best scenes, a ranking of the best scenes, and the like. It should be noted that in some embodiments, if a best scene or other video clip cannot be identified as an above-threshold match for clip requirements associated with a slot, the slot can be removed from the template without replacing the slot with a video clip.

[0062] In some embodiments, instead of replacing a video summary template slot with a video clip, an image or frame can be selected and can replace the slot. In some embodiments, an image or frame can be selected that satisfies one or more pre-determined criteria for inclusion in a video summary as described above. In some embodiments, an image or frame can be selected based on one or more criteria specified by the video summary template slot. For example, if a slot specifies one or more characteristics, an image or frame having one or more of the specified characteristics can be selected. In some embodiments, the video summary template slot can specify that an image or frame is to be selected to replace the slot. When an image or frame is selected and used to replace a slot, the image or frame can be displayed for the length of time associated with the slot. For instance, if a slot is associated with a four-second period of display time, an image or frame selected and used to replace the slot can be displayed for the four-second duration.

[0063] In some embodiments, when generating a video summary using a video summary template, the video editing module 320 can present a user with a set of candidate video clips for inclusion into one or more video summary template slots, for instance using a video summary generation interface. In such embodiments, the user can be presented with a pre-determined number of candidate video clips for a particular slot, and, in response to a selection of a candidate scene by the user, the video editing module 320 can replace the slot with the selected candidate video clip. In some embodiments, the candidate video clips

presented to the user for each video summary template slot are the video clips identified as best scenes (as described above). Once a user has selected a video clip for each slot in a video summary template, the video editing module 320 generates a video summary using the user-selected video clips based on the order of slots within the video summary template.

[0064] In one embodiment, the video editing module 320 generates video summary templates automatically, and stores the video summary templates in the template store 315. The video summary templates can be generated manually by experts in the field of video creation and video editing. The video editing module 320 may provide a user with a user interface allowing the user to generate video summary templates. Video summary templates can be received from an external source, such as an external template store. Video summary templates can be generated based on video summaries manually created by users, or based on an analysis of popular videos or movies (for instance by including a slot for each scene in a video).

SYSTEM OPERATION

[0055] Fig. 4 is a flowchart illustrating a method for selecting video portions to include in a video summary, according to one embodiment. A request to generate a video summary is received 410. The request can identify one or more videos for which a video summary is to be generated. In some embodiments, the request can be received from a user (for instance, via a video summary generation interface on a computing device), or can be received from a non-user entity (such as the video server 140 of Fig. 1). In response to the request, video and associated metadata is accessed 420. The metadata includes data describing characteristics of the video, the context or environment in which the video was captured, characteristics of the user or camera that captured the video, or any other information associated with the capture of the video. As described above, examples of such metadata include telemetry data describing the acceleration or velocity of the camera during the capture of the video, location or altitude data describing the location of the camera, environment data at the time of video capture, biometric data of a user at the time of video capture, and the like.

[0056] Events of interest within the accessed video are identified 430 based on the accessed metadata associated with the video. Events of interest can be identified based on changes in telemetry or location data within the metadata (such as changes in acceleration or velocity data), based on above-threshold values within the metadata (such as a velocity threshold or altitude threshold), based on local maximum or minimum values within the data (such as a maximum heart rate of a user), based on the proximity between metadata values and other

values, or based on any other suitable criteria. Best scenes are identified 440 based on the identified events of interest. For instance, for each event of interest identified within a video, a portion of the video corresponding to the event of interest (such as a threshold amount of time or a threshold number of frames before and after the time in the video associated with the event of interest) is identified as a best scene. A video summary is then generated 450 based on the identified best scenes, for instance by concatenating some or all of the best scenes into a single video.

[0057] Fig. 5 is a flowchart illustrating a method for generating video summaries using video templates, according to one embodiment. A request to generate a video summary is received 510. A video summary template is selected 520 in response to receiving the request. The selected video summary template can be a default template, can be selected by a user, can be selected based on an activity type associated with captured video, and the like. The selected video summary template includes a plurality of slots, each associated with a portion of the video summary. The video slots can specify video or associated metadata criteria (for instance, a slot can specify a high-acceleration video clip).

[0058] A set of candidate video clips is identified 530 for each slot, for instance based on the criteria specified by each slot, based on video clips identified as “best scenes” as described above, or based on any other suitable criteria. For each slot, a candidate video clip is selected 540 from among the set of candidate video clips identified for the slot. In some embodiments, the candidate video clips in each set of candidate video clips are ranked, and the most highly ranked candidate video clip is selected. The selected candidate video clips are combined 550 to generate a video summary. For instance, the selected candidate video clips can be concatenated in the order of the slots of the video summary template with which the selected candidate video clips correspond.

[0059] Fig. 6 is a flowchart illustrating a method for generating video summaries of videos associated with user-tagged events, according to one embodiment. Video is captured 610 by a user of a camera. During video capture, an input is received 620 from the user indicating an event of interest within the captured video. The input can be received, for instance, through the selection of a camera button, a camera interface, or the like. An indication of the user-tagged event of interest is stored in metadata associated with the captured video. A video portion associated with the tagged event of interest is selected 630, and a video summary including the selected video portion is generated 640. For instance, the selected video portion can be a threshold number of video frames before and after a frame associated with the user-

tagged event, and the selected video portion can be included in the generated video summary with one or more other video portions.

[0060] Fig. 7 is a flowchart illustrating a method 700 of identifying an activity associated with a video, according to one embodiment. A first video and associated metadata is accessed 710. An identification of an activity associated with the first video is received 720. For instance, a user can identify an activity in the first video during post-processing of the first video, or during the capture of the first video. A metadata pattern associated with the identified activity is identified 730 within the accessed metadata. The metadata pattern can include, for example, a defined change in acceleration metadata and altitude metadata.

[0061] A second video and associated metadata is accessed 740. The metadata pattern is identified 750 within the metadata associated with the second video. Continuing with the previous example, the metadata associated with the second video is analyzed and the defined change in acceleration metadata and altitude metadata is identified within the examined metadata. In response to identifying the metadata pattern within the metadata associated with the second video, the second video is associated 750 with the identified activity.

[0062] Fig. 8 is a flowchart illustrating a method 800 of sharing a video based on an identified activity within the video, according to one embodiment. Metadata patterns associated with one or more pre-determined activities are stored 810. Video and associated metadata are subsequently captured 820, and a stored metadata pattern associated with an activity is identified 830 within the captured metadata. A portion of the captured video associated with the metadata pattern is selected 840, and is outputted 850 based on the activity associated with the identified metadata pattern and/or one or more user settings. For instance, a user can select “snowboarding jump” and “3 seconds before and after” as an activity and video portion length, respectively. In such an example, when a user captures video, a metadata pattern associated with a snowboarding jump can be identified, and a video portion consisting of 3 seconds before and 3 seconds after the video associated with the snowboarding jump can automatically be uploaded to a social media outlet.

ADDITIONAL CONFIGURATION CONSIDERATIONS

[0063] Throughout this specification, some embodiments have used the expression “coupled” along with its derivatives. The term “coupled” as used herein is not necessarily limited to two or more elements being in direct physical or electrical contact. Rather, the term “coupled” may also encompass two or more elements are not in direct contact with each other, but yet

still co-operate or interact with each other, or are structured to provide a thermal conduction path between the elements.

[0064] Likewise, as used herein, the terms “comprises,” “comprising,” “includes,” “including,” “has,” “having” or any other variation thereof, are intended to cover a non-exclusive inclusion. For example, a process, method, article, or apparatus that comprises a list of elements is not necessarily limited to only those elements but may include other elements not expressly listed or inherent to such process, method, article, or apparatus.

[0065] In addition, use of the “a” or “an” are employed to describe elements and components of the embodiments herein. This is done merely for convenience and to give a general sense of the invention. This description should be read to include one or at least one and the singular also includes the plural unless it is obvious that it is meant otherwise.

[0066] Finally, as used herein any reference to “one embodiment” or “an embodiment” means that a particular element, feature, structure, or characteristic described in connection with the embodiment is included in at least one embodiment. The appearances of the phrase “in one embodiment” in various places in the specification are not necessarily all referring to the same embodiment.

[0067] Upon reading this disclosure, those of skill in the art will appreciate still additional alternative structural and functional designs for a camera expansion module as disclosed from the principles herein. Thus, while particular embodiments and applications have been illustrated and described, it is to be understood that the disclosed embodiments are not limited to the precise construction and components disclosed herein. Various modifications, changes and variations, which will be apparent to those skilled in the art, may be made in the arrangement, operation and details of the method and apparatus disclosed herein without departing from the spirit and scope defined in the appended claims.

CLAIMS

What is claimed is:

1. A method for identifying scenes in captured video for inclusion in a video summary, the method comprising:
 - accessing metadata associated with a video, the accessed metadata representative of capture of the video;
 - identifying a plurality of events of interest within the video based on the accessed metadata;
 - for each identified event of interest, identifying a best scene in the video associated with the identified event of interest, the identified best scene comprising a threshold amount of video occurring before and after a video frame or portion corresponding to the identified event of interest; and
 - selecting one or more identified best scenes for inclusion in a video summary.
2. The method of claim 1, further comprising:
 - receiving a request for a video summary from the user; and
 - generating a video summary including a plurality of selected best scenes.
3. The method of claim 2, wherein generating the video summary comprises concatenating the plurality of selected best scenes.
4. The method of claim 1, wherein the accessed metadata is generated by a camera during the capture of the video.
5. The method of claim 4, wherein the accessed metadata comprises telemetry data describing a motion of the camera during the capture of the video.
6. The method of claim 4, wherein the accessed metadata comprises location data describing a location of the camera during the capture of the video.
7. The method of claim 4, wherein the accessed metadata comprises biometric data describing characteristics of a user of the camera during the capture of the video.
8. The method of claim 1, wherein the accessed metadata is accessed from an external entity after the capture of the video.

9. The method of claim 8, wherein the accessed metadata comprises environment data describing characteristics of an environment in which the video was captured.
10. The method of claim 1, further comprising:
 ranking the identified best scenes based on a likelihood that the identified best scene will be of interest to a user;
 wherein the selected one or more best scenes are selected based on the ranking.
11. The method of claim 1, wherein selecting one or more identified best scenes comprises:
 presenting the identified best scenes to the user; and
 receiving, from the user, a selection of one or more identified best scenes.
12. A system for identifying scenes in captured video for inclusion in a video summary, the system comprising;
 a non-transitory computer-readable storage medium storing instructions configured to, when executed:
 access metadata associated with a video, the accessed metadata representative of capture of the video;
 identify a plurality of events of interest within the video based on the accessed metadata;
 for each identified event of interest, identify a best scene in the video associated with the identified event of interest, the identified best scene comprising a threshold amount of video occurring before and after a video frame or portion corresponding to the identified event of interest;
 and
 select one or more identified best scenes for inclusion in a video summary;
 and
 a processor configured to execute the instructions.
13. The system of claim 12, wherein the instructions are further configured to:
 receive a request for a video summary from the user; and
 generate a video summary including a plurality of selected best scenes.
14. The system of claim 12, wherein the accessed metadata is generated by a camera during the capture of the video and comprises one or more of: telemetry data describing a

motion of the camera during the capture of the video, location data describing a location of the camera during the capture of the video, and biometric data describing characteristics of a user of the camera during the capture of the video.

15. The system of claim 12, wherein the accessed metadata is accessed from an external entity after the capture of the video and comprises environment data describing characteristics of an environment in which the video was captured.

16. The system of claim 12, wherein the instructions are further configured to:
rank the identified best scenes based on a likelihood that the identified best scene will be of interest to a user;
wherein the selected one or more best scenes are selected based on the ranking.

17. A non-transitory computer-readable storage medium storing instructions for identifying scenes in captured video for inclusion in a video summary, the instructions configured to, when executed:
access metadata associated with a video, the accessed metadata representative of capture of the video;
identify a plurality of events of interest within the video based on the accessed metadata;
for each identified event of interest, identify a best scene in the video associated with the identified event of interest, the identified best scene comprising a threshold amount of video occurring before and after a video frame or portion corresponding to the identified event of interest; and
select one or more identified best scenes for inclusion in a video summary.

18. The computer-readable storage medium of claim 17, wherein the instructions are further configured to:
receive a request for a video summary from the user; and
generate a video summary including a plurality of selected best scenes.

19. The computer-readable storage medium of claim 17, wherein the accessed metadata is generated by a camera during the capture of the video and comprises one or more of: telemetry data describing a motion of the camera during the capture of the video, location data describing a location of the camera during the capture of the video, and biometric data describing characteristics of a user of the camera during the capture of the video.

20. The computer-readable storage medium of claim 17, wherein the accessed metadata is accessed from an external entity after the capture of the video and comprises environment data describing characteristics of an environment in which the video was captured.

21. The computer-readable storage medium of claim 17, wherein the instructions are further configured to:

rank the identified best scenes based on a likelihood that the identified best scene will be of interest to a user;

wherein the selected one or more best scenes are selected based on the ranking.

22. A method for identifying scenes in captured video for inclusion in a video summary, the method comprising;

accessing metadata associated with a video, the accessed metadata representative of capture of the video;

identifying a plurality of events of interest within the video based on the accessed metadata;

for each identified event of interest, identifying a best scene in the video associated with the identified event of interest, the identified best scene comprising a threshold amount of video occurring before and after a video frame or portion corresponding to the identified event of interest; and

selecting one or more identified best scenes for inclusion in a video summary.

23. The method of claim 22, further comprising:

receiving a request for a video summary from the user; and

generating a video summary including a plurality of selected best scenes.

24. The method of claim 23, wherein generating the video summary comprises concatenating the plurality of selected best scenes.

25. The method of claim 22, wherein the accessed metadata is generated by a camera during the capture of the video.

26. The method of claim 25, wherein the accessed metadata comprises telemetry data describing a motion of the camera during the capture of the video.

27. The method of claim 25, wherein the accessed metadata comprises location data describing a location of the camera during the capture of the video.

28. The method of claim 25, wherein the accessed metadata comprises biometric data describing characteristics of a user of the camera during the capture of the video.
29. The method of claim 22, wherein the accessed metadata is accessed from an external entity after the capture of the video.
30. The method of claim 29, wherein the accessed metadata comprises environment data describing characteristics of an environment in which the video was captured.
31. The method of claim 22, further comprising:
 ranking the identified best scenes based on a likelihood that the identified best scene will be of interest to a user;
 wherein the selected one or more best scenes are selected based on the ranking.
32. The method of claim 22, wherein selecting one or more identified best scenes comprises:
 presenting the identified best scenes to the user; and
 receiving, from the user, a selection of one or more identified best scenes.
33. A system for identifying scenes in captured video for inclusion in a video summary, the system comprising;
 a non-transitory computer-readable storage medium storing instructions configured to, when executed:
 access metadata associated with a video, the accessed metadata representative of capture of the video;
 identify a plurality of events of interest within the video based on the accessed metadata;
 for each identified event of interest, identify a best scene in the video associated with the identified event of interest, the identified best scene comprising a threshold amount of video occurring before and after a video frame or portion corresponding to the identified event of interest;
 and
 select one or more identified best scenes for inclusion in a video summary;
 and
 a processor configured to execute the instructions.

34. The system of claim 33, wherein the instructions are further configured to:
- receive a request for a video summary from the user; and
 - generate a video summary including a plurality of selected best scenes.
35. The system of claim 33, wherein the accessed metadata is generated by a camera during the capture of the video and comprises one or more of: telemetry data describing a motion of the camera during the capture of the video, location data describing a location of the camera during the capture of the video, and biometric data describing characteristics of a user of the camera during the capture of the video.
36. The system of claim 33, wherein the accessed metadata is accessed from an external entity after the capture of the video and comprises environment data describing characteristics of an environment in which the video was captured.
37. The system of claim 33, wherein the instructions are further configured to:
- rank the identified best scenes based on a likelihood that the identified best scene will be of interest to a user;
 - wherein the selected one or more best scenes are selected based on the ranking.
38. A non-transitory computer-readable storage medium storing instructions for identifying scenes in captured video for inclusion in a video summary, the instructions configured to, when executed:
- access metadata associated with a video, the accessed metadata representative of capture of the video;
 - identify a plurality of events of interest within the video based on the accessed metadata;
 - for each identified event of interest, identify a best scene in the video associated with the identified event of interest, the identified best scene comprising a threshold amount of video occurring before and after a video frame or portion corresponding to the identified event of interest; and
 - select one or more identified best scenes for inclusion in a video summary.
39. The computer-readable storage medium of claim 38, wherein the instructions are further configured to:
- receive a request for a video summary from the user; and
 - generate a video summary including a plurality of selected best scenes.

40. The computer-readable storage medium of claim 38, wherein the accessed metadata is generated by a camera during the capture of the video and comprises one or more of: telemetry data describing a motion of the camera during the capture of the video, location data describing a location of the camera during the capture of the video, and biometric data describing characteristics of a user of the camera during the capture of the video.

41. The computer-readable storage medium of claim 38, wherein the accessed metadata is accessed from an external entity after the capture of the video and comprises environment data describing characteristics of an environment in which the video was captured.

42. The computer-readable storage medium of claim 38, wherein the instructions are further configured to:

rank the identified best scenes based on a likelihood that the identified best scene will be of interest to a user;
wherein the selected one or more best scenes are selected based on the ranking.

43. A method for generating a video summary using a video summary template, the method comprising:

receiving a request for a video summary of one or more videos;
selecting a video summary template for use in generating the video summary, the video summary template comprising a plurality of template slots;
identifying, for each template slot, a set of candidate video clips associated with the template slot;
selecting, for each template slot, an associated candidate video clip from the set of candidate video clips associated with the template slot; and
generating the video summary by replacing each template slot in the video summary template with the associated selected candidate video clip.

44. The method of claim 43, wherein a template slot specifies one or more video clip requirements, and wherein each video clip in the set of candidate video clips associated with the template slot satisfy the specified video clip requirements.

45. The method of claim 43, wherein a template slot specifies a video clip length, and wherein each video clip in the set of candidate video clips associated with the template slot is the specified video clip length.

46. The method of claim 43, wherein identifying a set of candidate video clips comprises:
- identifying a plurality of events of interest in the one or more videos based on metadata associated with the one or more videos;
 - identifying a video clip associated with each identified event of interest;
 - ranking the identified video clips based on a likelihood that the identified video clips will be of interest to a user; and
 - selecting a set of candidate video clips based on the ranking.
47. The method of claim 43, wherein selecting a candidate video clip from the set of candidate video clips comprises selecting a highest ranked video clip.
48. The method of claim 43, wherein selecting a candidate video clip from the set of candidate video clips comprises:
- presenting the set of candidate video clips to a user; and
 - receiving, from the user, a selection of one of the set of candidate video clips.
49. A system for generating a video summary using a video summary template, the system comprising:
- a non-transitory computer-readable storage medium storing instructions configured to, when executed:
 - receive a request for a video summary of one or more videos;
 - select a video summary template for use in generating the video summary, the video summary template comprising a plurality of template slots;
 - identify, for each template slot, a set of candidate video clips associated with the template slot;
 - select, for each template slot, an associated candidate video clip from the set of candidate video clips associated with the template slot; and
 - generate the video summary by replacing each template slot in the video summary template with the associated selected candidate video clip;
 - and
 - a processor configured to execute the instructions.
50. The system of claim 49, wherein a template slot specifies one or more video clip requirements, and wherein each video clip in the set of candidate video clips associated with the template slot satisfy the specified video clip requirements.

51. The system of claim 49, wherein a template slot specifies a video clip length, and wherein each video clip in the set of candidate video clips associated with the template slot is the specified video clip length.

52. The system of claim 49, wherein identifying a set of candidate video clips comprises:
identifying a plurality of events of interest in the one or more videos based on metadata associated with the one or more videos;
identifying a video clip associated with each identified event of interest;
ranking the identified video clips based on a likelihood that the identified video clips will be of interest to a user; and
selecting a set of candidate video clips based on the ranking.

53. The system of claim 49, wherein selecting a candidate video clip from the set of candidate video clips comprises selecting a highest ranked video clip.

54. The system of claim 49, wherein selecting a candidate video clip from the set of candidate video clips comprises:
presenting the set of candidate video clips to a user; and
receiving, from the user, a selection of one of the set of candidate video clips.

55. A non-transitory computer-readable storage medium storing instructions for generating a video summary using a video summary template, the instructions configured to, when executed:

receive a request for a video summary of one or more videos;
select a video summary template for use in generating the video summary, the video summary template comprising a plurality of template slots;
identify, for each template slot, a set of candidate video clips associated with the template slot;
select, for each template slot, an associated candidate video clip from the set of candidate video clips associated with the template slot; and
generate the video summary by replacing each template slot in the video summary template with the associated selected candidate video clip.

56. The computer-readable storage medium of claim 55, wherein a template slot specifies one or more video clip requirements, and wherein each video clip in the set of candidate video clips associated with the template slot satisfy the specified video clip requirements.

57. The computer-readable storage medium of claim 55, wherein a template slot specifies a video clip length, and wherein each video clip in the set of candidate video clips associated with the template slot is the specified video clip length.

58. The computer-readable storage medium of claim 55, wherein identifying a set of candidate video clips comprises:

- identifying a plurality of events of interest in the one or more videos based on metadata associated with the one or more videos;
- identifying a video clip associated with each identified event of interest;
- ranking the identified video clips based on a likelihood that the identified video clips will be of interest to a user; and
- selecting a set of candidate video clips based on the ranking.

59. The computer-readable storage medium of claim 55, wherein selecting a candidate video clip from the set of candidate video clips comprises selecting a highest ranked video clip.

60. The computer-readable storage medium of claim 55, wherein selecting a candidate video clip from the set of candidate video clips comprises:

- presenting the set of candidate video clips to a user; and
- receiving, from the user, a selection of one of the set of candidate video clips.

61. A method for identifying scenes in captured video, the method comprising:

- capturing video with a camera;
- receiving, from a user of the camera during the capture of the video, an input identifying an event of interest within the captured video;
- storing an indication of the event of interest in metadata associated with the captured video; and
- identifying a video clip associated with the event of interest based on the metadata, the video clip comprising an amount of video occurring before and after a video portion corresponding to the event of interest.

62. The method of claim 61, wherein receiving an input identifying an event of interest comprises receiving a selection of a camera button associated with tagging events of interest.

63. The method of claim 61, wherein the event of interest is an activity type, wherein the received input identifies the activity type, and wherein storing an indication of the event of interest in metadata comprises storing an indication of the activity type in the metadata.
64. The method of claim 61, further comprising:
receiving a request to generate a video summary; and
generating a video summary comprising a plurality of video clips including the identified video clip associated with the event of interest.
65. The method of claim 61, further comprising:
receiving, from the user of the camera during the capture of the video, one or more additional inputs identifying one or more additional events of interest;
storing an indication of each additional event of interest in metadata associated with the captured video; and
identifying an additional video clip associated with each additional event of interest.
66. The method of claim 65, further comprising:
displaying each identified video clip;
receiving a selection of one or more displayed video clips; and
generating a video summary comprising the selected one or more displayed video clips.
67. A system for identifying scenes in captured video, the system comprising:
a non-transitory computer-readable storage medium storing instructions configured to, when executed:
capture video;
receive, from a user during the capture of the video, an input identifying an event of interest within the captured video;
store an indication of the event of interest in metadata associated with the captured video; and
identify a video clip associated with the event of interest based on the metadata, the video clip comprising an amount of video occurring before and after a video portion corresponding to the event of interest;
and
a processor configured to execute the instructions.

68. The system of claim 67, wherein receiving an input identifying an event of interest comprises receiving a selection of a button associated with tagging events of interest.
69. The system of claim 67, wherein the event of interest is an activity type, wherein the received input identifies the activity type, and wherein storing an indication of the event of interest in metadata comprises storing an indication of the activity type in the metadata.
70. The system of claim 67, wherein the instructions are further configured to:
receive a request to generate a video summary; and
generate a video summary comprising a plurality of video clips including the
identified video clip associated with the event of interest.
71. The system of claim 67, wherein the instructions are further configured to:
receive, from the user during the capture of the video, one or more additional inputs
identifying one or more additional events of interest;
store an indication of each additional event of interest in metadata associated with the
captured video; and
identify an additional video clip associated with each additional event of interest.
72. The system of claim 71, wherein the instructions are further configured to:
display each identified video clip;
receive a selection of one or more displayed video clips; and
generate a video summary comprising the selected one or more displayed video clips.
73. A non-transitory computer-readable storage medium storing instructions for
identifying scenes in captured video, the instructions configured to, when executed:
capture video;
receive, from a user during the capture of the video, an input identifying an event of
interest within the captured video;
store an indication of the event of interest in metadata associated with the captured
video; and
identify a video clip associated with the event of interest based on the metadata, the
video clip comprising an amount of video occurring before and after a video
portion corresponding to the event of interest.

74. The computer-readable storage medium of claim 73, wherein receiving an input identifying an event of interest comprises receiving a selection of a button associated with tagging events of interest.

75. The computer-readable storage medium of claim 73, wherein the event of interest is an activity type, wherein the received input identifies the activity type, and wherein storing an indication of the event of interest in metadata comprises storing an indication of the activity type in the metadata.

76. The computer-readable storage medium of claim 73, wherein the instructions are further configured to:

- receive a request to generate a video summary; and
- generate a video summary comprising a plurality of video clips including the identified video clip associated with the event of interest.

77. The computer-readable storage medium of claim 73, wherein the instructions are further configured to:

- receive, from the user during the capture of the video, one or more additional inputs identifying one or more additional events of interest;
- store an indication of each additional event of interest in metadata associated with the captured video; and
- identify an additional video clip associated with each additional event of interest.

78. The computer-readable storage medium of claim 77 wherein the instructions are further configured to:

- display each identified video clip;
- receive a selection of one or more displayed video clips; and
- generate a video summary comprising the selected one or more displayed video clips.

79. A method for identifying activities in videos, the method comprising:

- accessing a first video and first metadata associated with the first video;
- receiving an identification of an activity associated with the first video;
- identifying a metadata pattern associated with the identified activity within the first metadata;
- accessing a second video and second metadata associated with the second video;

identifying the metadata pattern within the second metadata; and
associating the second video with the identified activity.

80. The method of claim 79, wherein receiving an identification of an activity associated with the first video comprises receiving a tag from a user tagging the activity within the video.

81. The method of claim 79, wherein the received identification of an activity identifies a portion of the first video associated with the activity, and wherein identifying a metadata pattern associated with the identified activity comprises identifying a pattern within metadata associated with the portion of the first video.

82. The method of claim 79, wherein identifying a metadata pattern comprises identifying changes to one or more values within the first metadata.

83. The method of claim 82, wherein identifying the metadata pattern within the second metadata comprises identifying corresponding changes to the one or more values within the second metadata.

84. The method of claim 79, wherein the metadata pattern comprises a pattern within one or more of: telemetric data describing a motion of a camera, location data describing a location of a camera, biometric data describing characteristics of a user of a camera, and environment data describing characteristics of an environment of a camera.

85. The method of claim 79, wherein associating the second video with the identified activity comprises tagging the second video with the identified activity.

86. A system for identifying activities in videos, the system comprising:
a non-transitory computer-readable storage medium storing instructions configured to,
when executed:
access a first video and first metadata associated with the first video;
receive an identification of an activity associated with the first video;
identify a metadata pattern associated with the identified activity within the
first metadata;
access a second video and second metadata associated with the second video;
identify the metadata pattern within the second metadata; and

associate the second video with the identified activity; and
a processor configured to execute the instructions.

87. The system of claim 86, wherein receiving an identification of an activity associated with the first video comprises receiving a tag from a user tagging the activity within the video.

88. The system of claim 86, wherein the received identification of an activity identifies a portion of the first video associated with the activity, and wherein identifying a metadata pattern associated with the identified activity comprises identifying a pattern within metadata associated with the portion of the first video.

89. The system of claim 86, wherein identifying a metadata pattern comprises identifying changes to one or more values within the first metadata.

90. The system of claim 89, wherein identifying the metadata pattern within the second metadata comprises identifying corresponding changes to the one or more values within the second metadata.

91. The system of claim 86, wherein the metadata pattern comprises a pattern within one or more of: telemetric data describing a motion of a camera, location data describing a location of a camera, biometric data describing characteristics of a user of a camera, and environment data describing characteristics of an environment of a camera.

92. The system of claim 86, wherein associating the second video with the identified activity comprises tagging the second video with the identified activity.

93. A non-transitory computer-readable storage medium storing instructions for identifying activities in videos, the instructions configured to, when executed:
access a first video and first metadata associated with the first video;
receive an identification of an activity associated with the first video;
identify a metadata pattern associated with the identified activity within the first metadata;
access a second video and second metadata associated with the second video;
identify the metadata pattern within the second metadata; and
associate the second video with the identified activity.

94. The computer-readable storage medium of claim 93, wherein receiving an identification of an activity associated with the first video comprises receiving a tag from a user tagging the activity within the video.

95. The computer-readable storage medium of claim 93, wherein the received identification of an activity identifies a portion of the first video associated with the activity, and wherein identifying a metadata pattern associated with the identified activity comprises identifying a pattern within metadata associated with the portion of the first video.

96. The computer-readable storage medium of claim 93, wherein identifying a metadata pattern comprises identifying changes to one or more values within the first metadata.

97. The computer-readable storage medium of claim 96, wherein identifying the metadata pattern within the second metadata comprises identifying corresponding changes to the one or more values within the second metadata.

98. The computer-readable storage medium of claim 93, wherein the metadata pattern comprises a pattern within one or more of: telemetric data describing a motion of a camera, location data describing a location of a camera, biometric data describing characteristics of a user of a camera, and environment data describing characteristics of an environment of a camera.

99. The computer-readable storage medium of claim 93, wherein associating the second video with the identified activity comprises tagging the second video with the identified activity.

100. A method for identifying scenes in captured video for inclusion in a video summary, the method comprising;

- accessing a video, the video including a plurality of frames;

- identifying a motion value associated with each frame of the plurality of frames;

- identifying a differential motion value associated with each frame of the plurality of frames based on the motion value associated with each of one or more of the frames;

- determining a score for each frame of the plurality of frames based on the motion value and differential motion value associated with each frame;

selecting one or more frames in the video based on the determined scores;
identifying, for each selected frame, a corresponding video scene, the video
comprising a threshold amount of video occurring before and after the selected
frame; and
selecting one or more identified video scenes for inclusion in a video summary.

101. The method of claim 100, further comprising:

receiving a request for a video summary from the user; and
generating a video summary including a plurality of selected video scenes.

102. The method of claim 101, wherein generating the video summary comprises
concatenating the plurality of selected video scenes.

103. The method of claim 100, wherein identifying a motion value associated with a
frame of the plurality of frames comprises:

identifying a magnitude of each of a plurality of motion vectors in the frame; and
determining the motion value associated with the frame based on the magnitude of
each of the plurality of motion vectors.

104. The method of claim 100, wherein identifying a differential motion value associated
with a frame of the plurality of frames comprises:

identifying a motion value associated with the frame and an adjacent previous frame;
and
determining the differential motion value of the frame based on the difference
between the motion value of the frame and the motion value of the previous
frame.

105. The method of claim 104, further comprising:

identifying a threshold number of differential motion values associated with a
sequence of frames; and
aggregating the threshold number of differential motion values to determine a single
differential motion value for each of the frames in the sequence of frames.

106. The method of claim 100, further comprising:

ranking the identified video scenes based on a likelihood that the identified video
scenes will be of interest to a user;

wherein the selected one or more video scenes are selected based on the ranking.

107. A system for identifying scenes in captured video for inclusion in a video summary, the system comprising;

a non-transitory computer-readable storage medium comprising executable computer instructions configured to, when executed, perform steps comprising:

- identifying a motion value associated with each frame of the plurality of frames;
- identifying a differential motion value associated with each frame of the plurality of frames based on the motion value associated with each of one or more of the frames;
- determining a score for each frame of the plurality of frames based on the motion value and differential motion value associated with each frame;
- selecting one or more frames in the video based on the determined scores;
- identifying, for each selected frame, a corresponding video scene, the video comprising a threshold amount of video occurring before and after the selected frame; and
- selecting one or more identified video scenes for inclusion in a video summary; and

a hardware processor.

108. The system of claim 107, the instructions configured to, when executed, perform further steps comprising:

- receiving a request for a video summary from the user; and
- generating a video summary including a plurality of selected video scenes.

109. The system of claim 108, wherein generating the video summary comprises concatenating the plurality of selected video scenes.

110. The system of claim 107, wherein identifying a motion value associated with a frame of the plurality of frames comprises:

- identifying a magnitude of each of a plurality of motion vectors in the frame; and
- determining the motion value associated with the frame based on the magnitude of each of the plurality of motion vectors.

111. The system of claim 107, wherein identifying a differential motion value associated with a frame of the plurality of frames comprises:

- identifying a motion value associated with the frame and an adjacent previous frame;
- and
- determining the differential motion value of the frame based on the difference between the motion value of the frame and the motion value of the previous frame.

112. The system of claim 111, the instructions configured to, when executed, perform further steps comprising:

- identifying a threshold number of differential motion values associated with a sequence of frames; and
- aggregating the threshold number of differential motion values to determine a single differential motion value for each of the frames in the sequence of frames.

113. The system of claim 107, the instructions configured to, when executed, perform further steps comprising:

- ranking the identified video scenes based on a likelihood that the identified video scenes will be of interest to a user;
- wherein the selected one or more video scenes are selected based on the ranking.

114. A non-transitory computer-readable storage medium storing executable computer instructions configured to, when executed by a processor, perform steps comprising:

- accessing a video, the video including a plurality of frames;
- identifying a motion value associated with each frame of the plurality of frames;
- identifying a differential motion value associated with each frame of the plurality of frames based on the motion value associated with each of one or more of the frames;
- determining a score for each frame of the plurality of frames based on the motion value and differential motion value associated with each frame;
- selecting one or more frames in the video based on the determined scores;
- identifying, for each selected frame, a corresponding video scene, the video comprising a threshold amount of video occurring before and after the selected frame; and
- selecting one or more identified video scenes for inclusion in a video summary.

115. The computer-readable storage medium of claim 114, comprising instructions configured to, when executed, perform further steps comprising:

receiving a request for a video summary from the user; and
generating a video summary including a plurality of selected video scenes.

116. The computer-readable storage medium of claim 115, wherein generating the video summary comprises concatenating the plurality of selected video scenes.

117. The computer-readable storage medium of claim 114, wherein identifying a motion value associated with a frame of the plurality of frames comprises:

identifying a magnitude of each of a plurality of motion vectors in the frame; and
determining the motion value associated with the frame based on the magnitude of each of the plurality of motion vectors.

118. The computer-readable storage medium of claim 114, wherein identifying a differential motion value associated with a frame of the plurality of frames comprises:

identifying a motion value associated with the frame and an adjacent previous frame;
and
determining the differential motion value of the frame based on the difference
between the motion value of the frame and the motion value of the previous
frame.

119. The computer-readable storage medium of claim 118, comprising instructions configured to, when executed, perform further steps comprising:

identifying a threshold number of differential motion values associated with a
sequence of frames; and
aggregating the threshold number of differential motion values to determine a single
differential motion value for each of the frames in the sequence of frames.

120. The computer-readable storage medium of claim 114, comprising instructions configured to, when executed, perform further steps comprising:

ranking the identified video scenes based on a likelihood that the identified video
scenes will be of interest to a user;
wherein the selected one or more video scenes are selected based on the ranking.

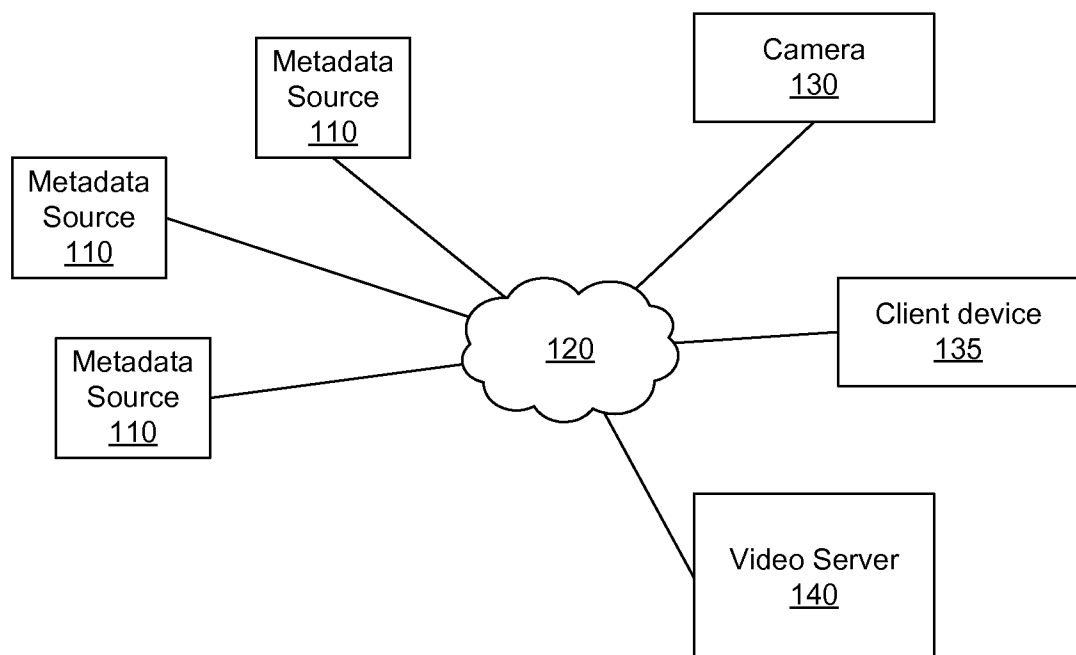
100

FIG. 1

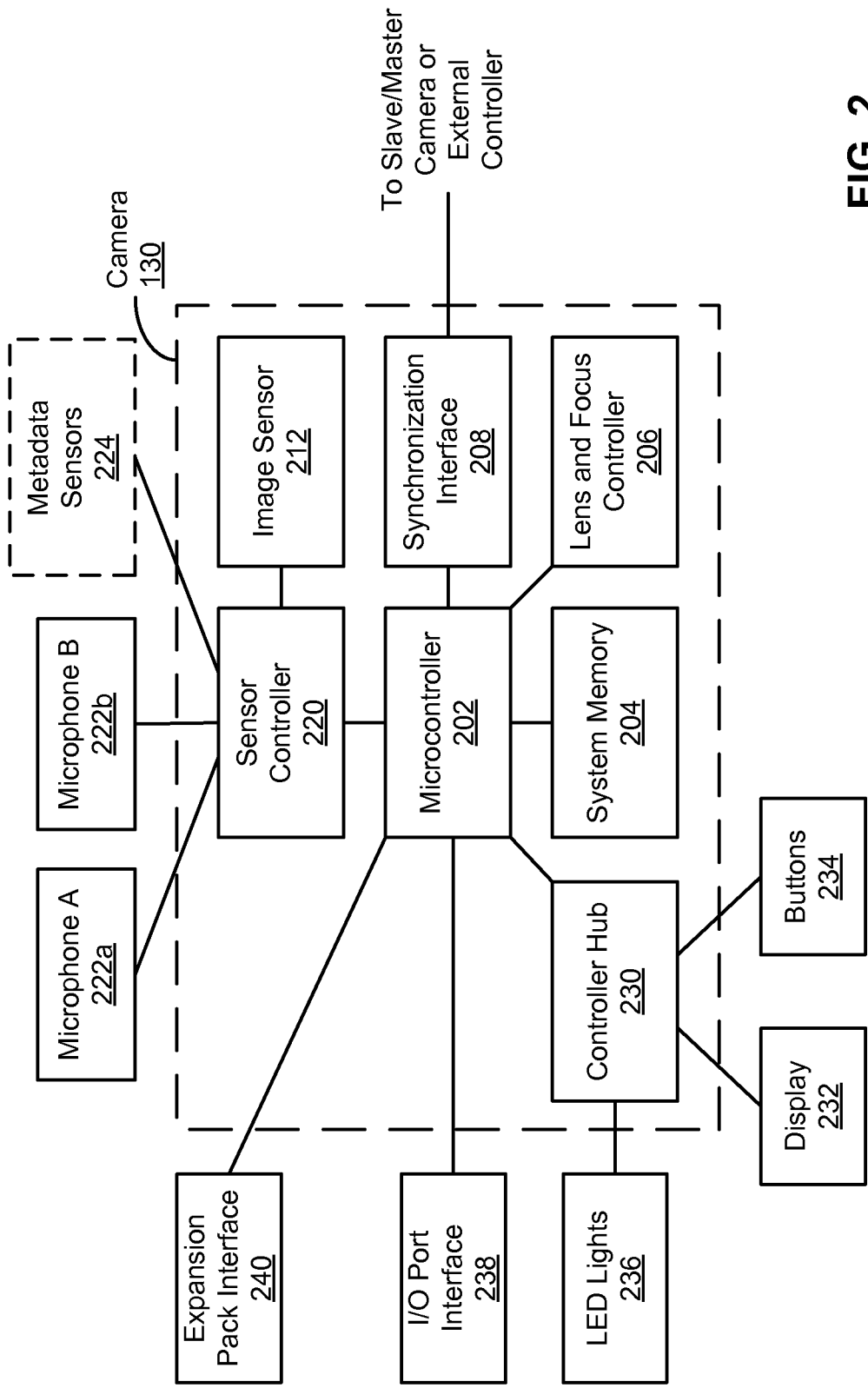


FIG. 2

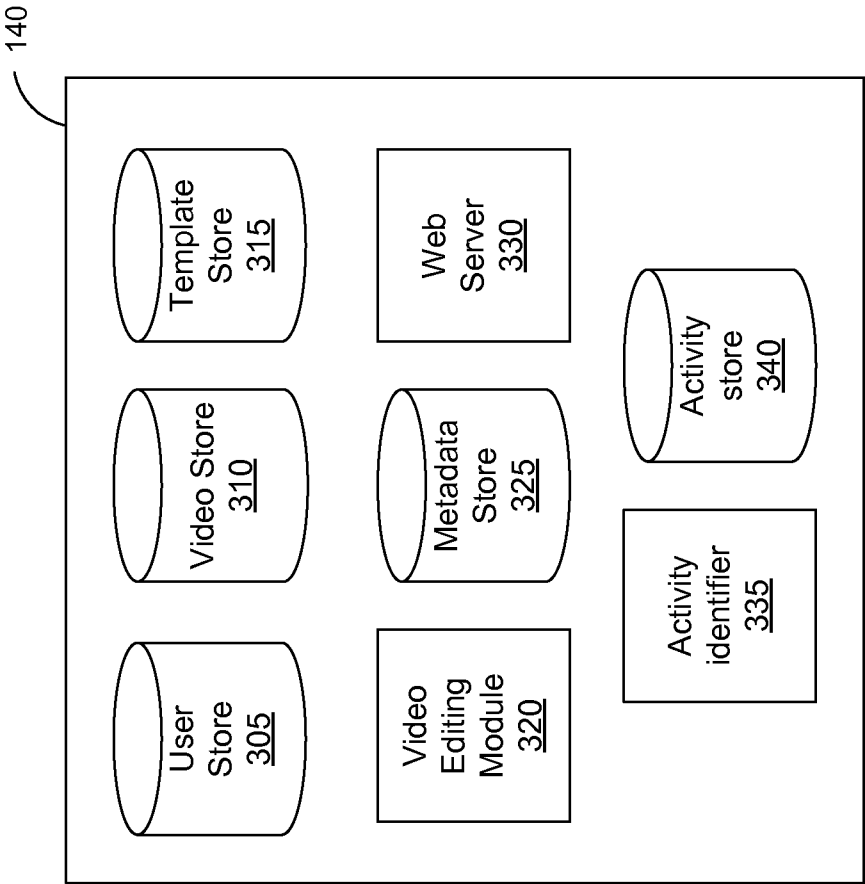
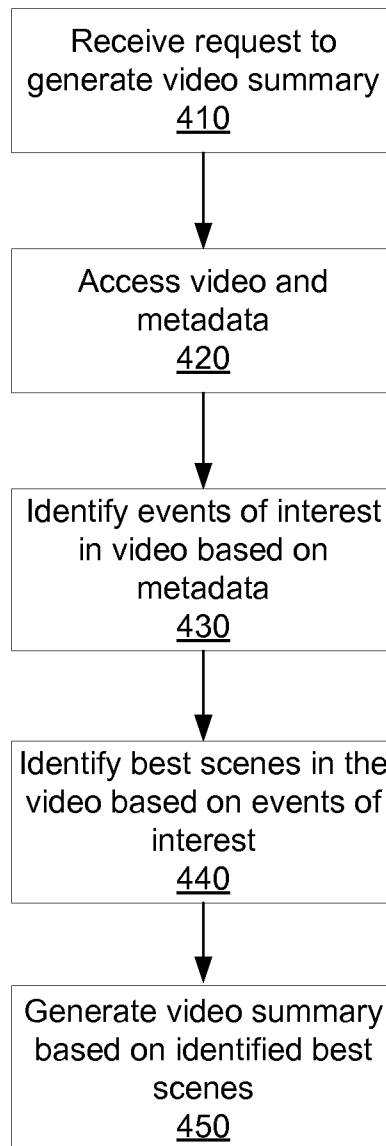
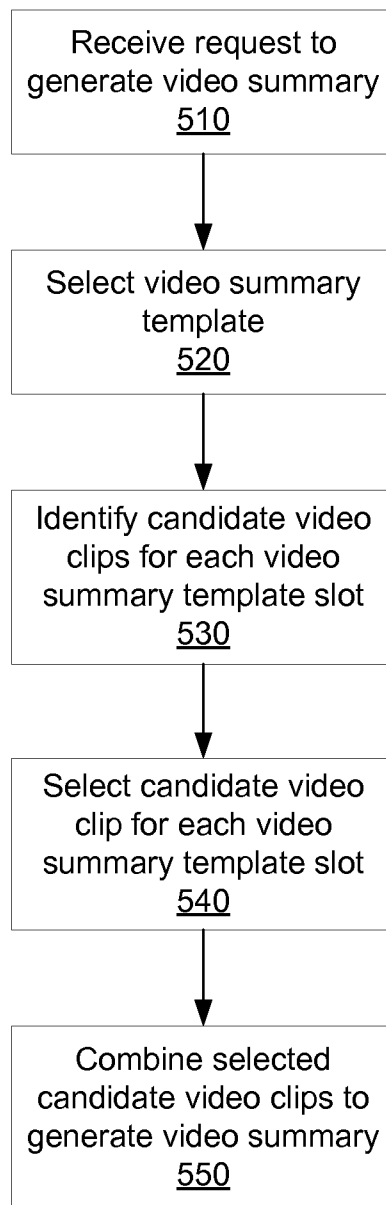
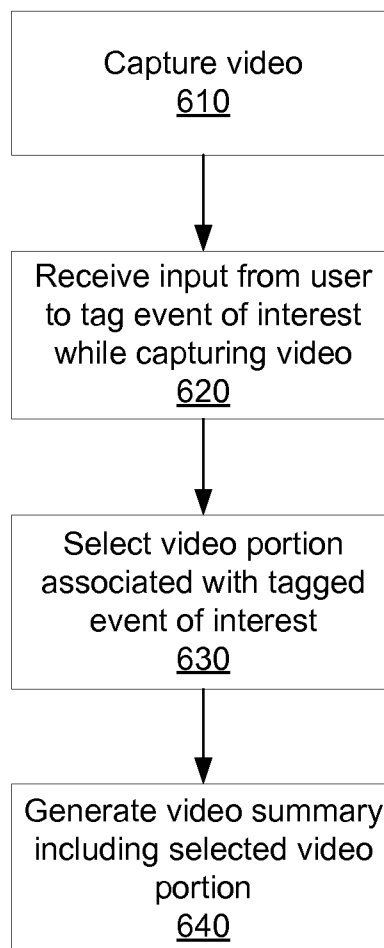
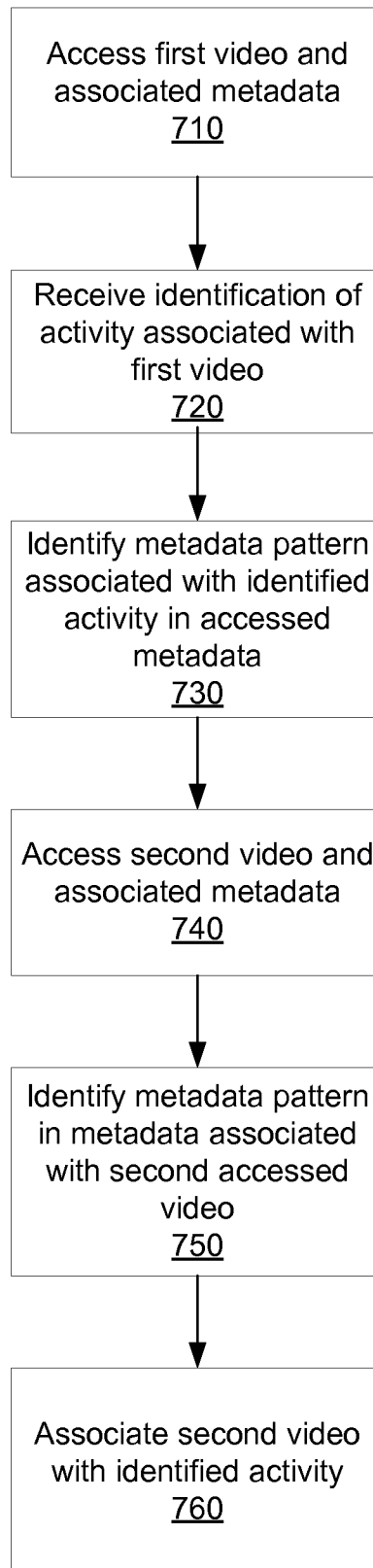


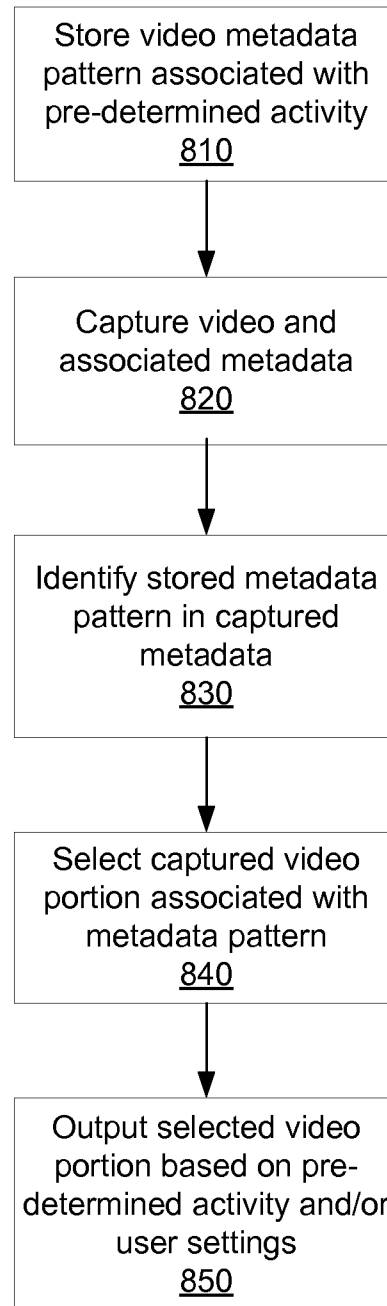
FIG. 3

**FIG. 4**

**FIG. 5**

**FIG. 6**

**FIG. 7**

**FIG. 8**

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 15/41624

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - H04N 9/80 (2015.01)

CPC - H04N 21/8549

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)
CPC: H04N21/8549; IPC(8): H04N 9/80 (2015.01); USPC: 386/239Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
CPC: G06K9/00711, G06F17/30843, G06F17/30811, G06F17/30787, H04N21/8453, H04N21/25891, H04N21/8549, G06F17/30802;
USPC: 725/19, 382/190, 707/E17.028, 725/9, 382/181, 382/205, 382/195, 386/241, 386/239 (keyword limited; terms below)

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

PatBase; Google Scholar

Search Terms: scenes, video summary, metadata, interest, best scene, capture, telemetry, location, position, geo, biometric, clip, segment

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X. --- Y	US 2010/0104261 A1 (LIU et al.) 29 April 2010 (29.04.2010), entire document, especially abstract and para [0002]-[0003], [0018]-[0021], [0024], [0038], [0040]-[0042], [0045], [0050]-[0051].	1-5, 10, 12-14, 16-19, 21-26, 31, 33-35, 37-40, 42 6-9, 11, 15, 20, 27-30, 32, 36, 41
Y	US 2012/0057852 A1 (DEVLEESCHOUWER et al.) 08 March 2012 (08.03.2012), entire document, especially abstract and para [0057]-[0058], [0091], [0152], [0154], [0187]-[0193], [0201]-[0204], [0208], Fig. 4, Fig. 6.	6, 8-9, 15, 20, 27, 29-30, 36, 41
Y	US 2010/0245626 A1 (WOYCECHOWSKY) 30 September 2010 (30.09.2010), entire document, especially abstract and para [0036].	7, 28
Y	US 2013/0318443 A1 (BACHMAN et al.) 28 November 2013 (28.11.2013), entire document, especially abstract and para [0003], [0020], [0023].	11, 32
A	US 2012/0123780 A1 (GAO et al.) 17 May 2012 (17.05.2012), entire document.	1-42
A	US 2014/0072285 A1 (SHYNAR et al.) 13 March 2014 (13.03.2014), entire document.	1-42

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

* 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

04 November 2015 (04.11.2015)

Date of mailing of the international search report

17 DEC 2015

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 15/41624

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

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

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

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

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

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

This International Searching Authority found multiple inventions in this international application, as follows:
This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1.

Group I: Claims 1-42, directed to identifying scenes in captured video for inclusion in a video summary by using metadata associated with a video, identifying events of interest within the video, identifying a best scene in the video, and selecting one or more best scenes for inclusion in a video summary.

Group II: Claims 43-60, directed to generating a video summary using a video summary template by receiving a request for a video summary, selecting a video summary template comprising a plurality of template slots for use in generating the video summary, selecting, for each template slot, an associated candidate video clip, and generating the video summary by replacing each template slot in the video summary template with the associated selected candidate video clip.

(See extra sheet)

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-42

Remark on Protest

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

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 15/41624

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

Group III: Claims 61-78, directed to identifying scenes in captured video by capturing video, receiving, from a user during the capture of the video, input identifying an event of interest within the captured video, storing an indication of the event of interest in metadata, and identifying a video clip associated with the event of interest based on the metadata.

Group IV: Claims 79-99, directed to identifying activities in videos by accessing videos and video metadata, receiving an identification of an activity associated with a video, identifying metadata patterns associated with the identified activity, and associating a video with the identified activity.

Group V: Claims 100-120, directed to identifying scenes in captured video for inclusion in a video summary by accessing a video including a plurality of frames, identifying a motion value associated with each frame, identifying a differential motion value associated with each frame based on the motion value, determining a score for each frame based on the motion value and differential motion value, selecting one or more frames based on the scores, identifying a corresponding video scene, and selecting one or more scenes for inclusion in a video summary.

The groups of inventions listed above do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The special technical feature of the Group I invention is using metadata associated with a video, identifying events of interest within the video, identifying a best scene in the video, and selecting one or more best scenes for inclusion in a video summary, not required by Group II-V. (See Fig. 4 for example).

The special technical feature of the Group II invention is receiving a request for a video summary, selecting a video summary template comprising a plurality of template slots for use in generating the video summary, selecting, for each template slot, an associated candidate video clip, and generating the video summary by replacing each template slot in the video summary template with the associated selected candidate video clip, not required by Group I or III-V. (See Fig. 5 for example).

The special technical feature of the Group III invention is receiving, from a user during the capture of the video, input identifying an event of interest within the captured video, storing an indication of the event of interest in metadata, and identifying a video clip associated with the event of interest based on the metadata, not required by Group I-II or IV-V. (See Fig. 6 for example).

The special technical feature of the Group IV invention is accessing videos and video metadata, receiving an identification of an activity associated with a video, identifying metadata patterns associated with the identified activity, and associating a video with the identified activity, not required by Group I-III or V. (See Fig. 7 for example).

The special technical feature of the Group V invention is accessing a video including a plurality of frames, identifying a motion value associated with each frame, identifying a differential motion value associated with each frame based on the motion value, determining a score for each frame based on the motion value and differential motion value, selecting one or more frames based on the scores, identifying a corresponding video scene, and selecting one or more scenes for inclusion in a video summary, not required by Group I-IV.

Groups I, II and V share the technical features of, in various combination, utilizing a plurality of video segments / clips / frames, and generating a video summary.

However, these shared technical features fail to represent a contribution over the prior art of US 2014/0161351 A1 (Yagnik), which discloses utilizing a plurality of video segments / clips / frames (para [0025]-[0028] - "receiving a video 102, which is comprised of a number of frames"), and generating a video summary (para [0009], claim 1 - "combining selected scenes to produce the summary for the video").

Groups I, III and IV share the technical features of, in various combination, utilizing metadata associated with video, and identifying activities / scenes in captured video associated with events of interest / activities within captured video.

However, these shared technical features fail to represent a contribution over the prior art of US 2014/0093164 A1 to Noorkami et al. (hereinafter 'Noorkami'), which discloses utilizing metadata associated with video (para [0062]-[0064] - "the method 700 may then insert metadata into the video content"), and identifying activities / scenes in captured video associated with events of interest / activities (para [0062]-[0064], claim 1 - "the method 700 may examine all of the shots in the video content 34 and efficiently determine one or more scene boundaries at one or more target shots, where a new scene begins at each scene boundary"; "detecting a scene in video content") within captured video (para [0019] - "each of the frames (i-1), (i), (i+1), (i+2), etc. may represent an image captured by an image capture device, such as a video camera").

(See next extra sheet)

(Continuation of previous extra sheet)

Groups I, III and V share the technical features of, in various combination, identifying activities / scenes in captured video, and an amount of video occurring before and after a video frame or portion corresponding to and identified event of interest.

However, these shared technical features fail to represent a contribution over the prior art of US 2013/0300939 A1 to Chou et al. (hereinafter 'Chou'), which discloses identifying activities / scenes (para [0013], [0050] - "scene recognition in a video/audio processing environment") in captured video (para [0033] - "embodiments of system 10 can operate on videos and audios captured from any capture system"), and an amount of video occurring before and after a video frame or portion corresponding to and identified event of interest (para [0019], [0022]-[0028], [0035], [0057] - "User 32 may search the example video for various scenes (e.g., Welcome Scene, Farewell Scene, etc.); "order scenes into a scene sequence"; "user 32 may be provided with options to click through to particular scenes of interest, or speakers of interest, as the case may be. Because each scene sequence 52 and speaker sequence 54 may include scenes tagged with scene identifiers, and speakers tagged with speaker identifiers, respectively, searching for particular scenes and/or speakers in report 24 may be effected easily").

Groups I, III, IV and V share the technical features of, in various combination, identifying activities / scenes in captured video.

However, these shared technical features fail to represent a contribution over the prior art of US 2014/0105573 A1 to Hanckmann et al. (hereinafter 'Hanckmann'), which discloses identifying activities / scenes (para [0019], [0021], [0072] - "Action type detector 14 is a device that detects which of a predetermined set of action types is seen to occur in a time range in a video stream"; "identify the subset of likely actions, and to subsequently describe them with a rule-based method that relates the actions to entities in the scene") in captured video (para [0018] - "Camera 10 is configured to generate a stream of data that represents a temporal sequence of video images, which is stored by video data storage device").

Thus, the inventions listed as Groups I-V lack unity of invention because they do not share a same or corresponding special technical feature providing a contribution over the prior art.