



(51) International Patent Classification:

H04N 21/8549 (2011.01) *H04N 5/265* (2006.01)
H04N 5/262 (2006.01) *G11B 27/031* (2006.01)

(21) International Application Number:

PCT/IB2012/054794

(22) International Filing Date:

14 September 2012 (14.09.2012)

(25) Filing Language:

English

(26) Publication Language:

English

(71) Applicant (for all designated States except US): **Nokia Corporation** [FI/FI]; Keilalahdentie 4, FI-02150 Espoo (FI).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **ARRASVUORI, Juha Henrik** [FI/FI]; Kruutinkuja 2 B 13, FI-33710 Tampere (FI). **ERONEN, Antti Johannes** [FI/FI]; Rintamäenkatu 13 A 6, FI-33820 Tampere (FI). **LEHTINIEMI, Arto Juhani** [FI/FI]; Lomarantatie 17, FI-33880 Lempäälä (FI).

(74) Agents: **ANDERSON, Oliver** et al.; 200 Aldersgate, London EC1A 4HD (GB).

(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, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

[Continued on next page]

(54) Title: HANDLING USER-GENERATED CONTENT

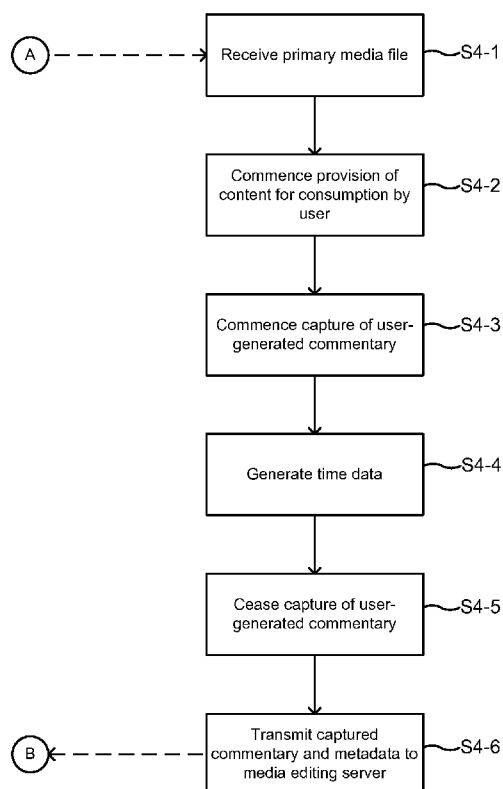


Figure 4

(57) Abstract: A method comprises receiving user-generated auxiliary content and associated time data, the user-generated auxiliary content being associated with primary content from a primary media file, the time data indicating a least one portion of the primary content to which the user-generated auxiliary content relates; identifying, using the time data, at least a portion of the user-generated auxiliary content that relates to a portion of the primary content; storing, in a first storage file, the identified at least a portion of the user-generated auxiliary content; and storing, in association with the first storage file information identifying the portion of the primary content to which the at least a portion of the user-generated auxiliary content is related. A further method comprises receiving user-generated auxiliary content relating to at least one portion of the primary media content; and causing transmission of the received user-generated auxiliary content along with time data, the time data indicating at least one portion of the primary media content to which the user-generated auxiliary content relates.

WO 2014/041399 A1



TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, **Published:**
ML, MR, NE, SN, TD, TG).

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

Handling User-Generated Content

Field

The invention relates to the handling of user-generated content.

5

Background

Media remixing is an application where multiple video media recordings are combined in order to obtain a video media mix that contains some segments selected from the plurality of video media recordings obtained, for example, through crowdsourcing principles. Video remixing, as such, is one of the basic manual video editing applications, for which various software products and services are already available. Furthermore, there exist automatic video remixing or editing systems, which use multiple instances of user-generated or professional recordings to automatically generate a remix that combines content from the available source content.

15

Summary

In a first aspect, this specification describes a method comprising: receiving user-generated auxiliary content and associated time data, the user-generated auxiliary content being associated with primary content from a primary media file, the time data indicating a least one portion of the primary content to which the user-generated auxiliary content relates; identifying, using the time data, at least a portion of the user-generated auxiliary content that relates to a portion of the primary content; storing, in a first storage file, the identified at least a portion of the user-generated auxiliary content; and storing, in association with the first storage file information identifying the portion of the primary content to which the at least a portion of the user-generated auxiliary content is related.

20

25

The time data may indicate a time, relative to playback or capture of the primary content, at which the user-generated auxiliary content was captured.

30

The primary content from the primary media file may comprise first and second sections, wherein a first portion of the user-generated auxiliary content relates to the first section and a second portion relates to the second section, and the method may comprise: identifying, using the time data included in the metadata, the first portion of

35

the user-generated auxiliary content and the second portion of the user-generated auxiliary content; storing, in the first storage file, the first portion of the user-generated auxiliary content, wherein the information stored in association with first storage file identifies the first section of the primary content; storing, in a second storage file, the
5 second portion of the user-generated auxiliary content; and storing, in association with the second storage file, information indentifying the second section of the primary content.

The method may comprise: in response to determining that the portion of the primary
10 content is to be included in a composite media file, retrieving the first storage file from storage; and modifying the portion of the primary content to include the auxiliary content and including the modified portion in the composite media file; or including the portion of the primary content in the composite media file and preparing the auxiliary content from the first storage file for consumption by a user along with the
15 composite media file.

The primary content may comprise a video content part and a corresponding audio content part. The user-generated auxiliary content may comprise video content and the method may comprise modifying the portion of the primary content by modifying the
20 video content part of the portion of the primary content to include the video content of the at least a portion of the user-generated auxiliary content. The user-generated auxiliary content may alternatively or additionally comprise audio content and the method may comprise modifying the portion of the primary content by modifying the audio content part of the portion of the primary content to include the audio content of
25 the at least a portion of the user-generated auxiliary content.

The method may comprise prior to receiving the user-generated auxiliary content and associated metadata, providing the primary media file, for consumption by the user. The method may comprise outputting, for consumption by the user, the primary
30 content from the first media file; and concurrently, capturing the user-generated auxiliary content. The method may comprise generating the time data, the time data identifying a time, relative to the beginning of the primary content, at which the capturing of the user-generated auxiliary content was initiated. The method may comprise generating, for transmission along with the time data, information identifying
35 the first media file.

The received user-generated auxiliary content may be received at server apparatus from a user terminal and the auxiliary content may have been captured by the user terminal.

In a second aspect, this specification describes a method comprising: receiving user-generated auxiliary content relating to at least one portion of the primary media
5 content; and causing transmission of the received user-generated auxiliary content along with time data, the time data indicating at least one portion of the primary media content to which the user-generated auxiliary content relates. The method may comprise causing transmission, along with the time data, of information identifying a
10 primary media file containing the primary content to which the captured auxiliary content relates.

The method may comprise causing provision, for consumption by a user, of the primary media content; and receiving the user-generated auxiliary content during provision of
15 the primary media content. The method may comprise: receiving, during provision of the primary media content, an instruction to initiate capture of the user-generated auxiliary content; and responding to the instruction, by causing capture of the user-generated auxiliary content to be initiated and by generating the time data, the time data indicating a time, relative to the provision of the primary media content, at which
20 auxiliary content capture was initiated.

Alternatively, the method may comprise: causing capture of the primary media content; and receiving the user-generated auxiliary content during capture of the primary media content. The method may comprise: receiving, during capture of the primary media
25 content, an instruction to initiate capture of the user-generated auxiliary content; and responding to the instruction, by causing capture of the user-generated auxiliary content to be initiated and by generating the time data, the time data indicating a time, relative to the capture of the primary media content, at which auxiliary content capture was initiated. The method may comprise causing transmission of the captured primary
30 media content along with the auxiliary content.

In a third aspect, this specification describes apparatus comprising at least one processor and at least one memory, the at least one memory having stored thereon computer-readable code which, when executed by the at least one processor, causes the
35 apparatus: to receive user-generated auxiliary content and associated time data, the user-generated auxiliary content being associated with primary content from a primary

media file, the time data indicating a least one portion of the primary content to which the user-generated auxiliary content relates; to identify, using the time data, at least a portion of the user-generated auxiliary content that relates to a portion of the primary content; to store, in a first storage file, the identified at least a portion of the user-generated auxiliary content; and to store, in association with the first storage file information identifying the portion of the primary content to which the at least a portion of the user-generated auxiliary content is related.

The time data may indicate a time, relative to playback or capture of the primary content, at which the user-generated auxiliary content was captured.

The primary content from the primary media file may comprise first and second sections, wherein a first portion of the user-generated auxiliary content relates to the first section and a second portion relates to the second section, and the computer-readable code may, when executed by the at least one processor, cause the apparatus: to identify, using the time data included in the metadata, the first portion of the user-generated auxiliary content and the second portion of the user-generated auxiliary content; to store, in the first storage file, the first portion of the user-generated auxiliary content, wherein the information stored in association with first storage file identifies the first section of the primary content; to store, in a second storage file, the second portion of the user-generated auxiliary content; and to store, in association with the second storage file, information indentifying the second section of the primary content.

The computer-readable code may, when executed by the at least one processor, cause the apparatus: in response to determining that the portion of the primary content is to be included in a composite media file, to retrieve the first storage file from storage; and to modify the portion of the primary content to include the auxiliary content and to include the modified portion in the composite media file; or to include the portion of the primary content in the composite media file and to prepare the auxiliary content from the first storage file for consumption by a user along with the composite media file.

The primary content may comprise a video content part and a corresponding audio content part. The user-generated auxiliary content may comprise video content, and the computer-readable code may, when executed by the at least one processor, cause

the apparatus to modify the portion of the primary content by modifying the video content part of the portion of the primary content to include the video content of the at least a portion of the user-generated auxiliary content. Alternatively or additionally, the user-generated auxiliary content may comprise audio content and the computer-readable code may, when executed by the at least one processor cause the apparatus to
5 modify the portion of the primary content by modifying the audio content part of the portion of the primary content to include the audio content of the at least a portion of the user-generated auxiliary content.

10 The computer-readable code may, when executed by the at least one processor, cause the apparatus: prior to receiving the user-generated auxiliary content and associated metadata, to cause provision of the primary media file, for consumption by the user. The computer-readable code may, when executed by the at least one processor, cause the apparatus: to cause to be outputted, for consumption by the user, the primary
15 content from the first media file; and concurrently, to cause the user-generated auxiliary content to be captured.

The computer-readable code may, when executed by the at least one processor, cause the apparatus: to generate the time data, the time data identifying a time, relative to the
20 beginning of the primary content, at which the capturing of the user-generated auxiliary content was initiated. The computer-readable code may, when executed by the at least one processor, cause the apparatus: to generate, for transmission along with the time data, information identifying the first media file.

25 The received user-generated auxiliary content may be received at server apparatus from a user terminal and the auxiliary content may have been captured by the user terminal.

In a fourth aspect, this specification describes apparatus comprising at least one processor and at least one memory, the at least one memory having stored thereon
30 computer-readable code which, when executed by the at least one processor, causes the apparatus: to receive user-generated auxiliary content relating to at least one portion of the primary media content; and to cause transmission of the received user-generated auxiliary content along with time data, the time data indicating at least one portion of the primary media content to which the user-generated auxiliary content relates. The
35 computer-readable code may, when executed by the at least one processor, cause the apparatus to cause transmission, along with the time data, of information identifying a

primary media file which contains the primary media content to which the auxiliary content relates.

5 The computer-readable code may, when executed by the at least one processor, cause the apparatus: to cause provision, for consumption by a user, of the primary media content; and to receive the user-generated auxiliary content during provision of the primary media content. The computer-readable code may, when executed by the at least one processor, cause the apparatus: to receive, during provision of the primary media content, an instruction to initiate capture of the user-generated auxiliary
10 content; and to respond to the instruction, by causing capture of the user-generated auxiliary content to be initiated and by generating the time data, the time data indicating a time, relative to the provision of the primary media content, at which auxiliary content capture was initiated.

15 The computer-readable code may, when executed by the at least one processor, cause the apparatus: to cause capture of the primary media content; and to receive the user-generated auxiliary content during capture of the primary media content. The computer-readable code may, when executed by the at least one processor, cause the apparatus: to receive, during capture of the primary media content, an instruction to
20 initiate capture of the user-generated auxiliary content; and to respond to the instruction, by causing capture of the user-generated auxiliary content to be initiated and by generating the time data, the time data indicating a time, relative to the capture of the primary media content, at which auxiliary content capture was initiated. The computer-readable code may, when executed by the at least one processor, cause the
25 apparatus to cause transmission of the captured primary media content along with the auxiliary content.

In a fifth aspect, this specification describes at least one non-transitory computer-readable medium having computer-readable instructions stored thereon, the computer-readable instructions, when executed by at least one processor, causing the at least one
30 processor: to receive user-generated auxiliary content and associated time data, the user-generated auxiliary content being associated with primary content from a primary media file, the time data indicating a least one portion of the primary content to which the user-generated auxiliary content relates; to identify, using the time data, at least a portion of the user-generated auxiliary content that relates to a portion of the primary
35 content; to store, in a first storage file, the identified at least a portion of the user-generated auxiliary content; and to store, in association with the first storage file

information identifying the portion of the primary content to which the at least a portion of the user-generated auxiliary content is related.

5 In a sixth aspect, this specification describes at least one non-transitory computer-readable medium having computer-readable instructions stored thereon, the computer-readable instructions, when executed by at least one processor, causing the at least one processor: to receive user-generated auxiliary content relating to at least one portion of the primary media content; and to cause transmission of the received user-generated auxiliary content along with time data, the time data indicating at least one portion of
10 the primary media content to which the user-generated auxiliary content relates.

In a seventh aspect, this specification describes computer-readable code which, when executed by computing apparatus, causes the computing apparatus to perform the method of either of the first and second aspects.

15

In an eighth aspect, this specification describes apparatus comprising: means for receiving user-generated auxiliary content and associated time data, the user-generated auxiliary content being associated with primary content from a primary media file, the time data indicating a least one portion of the primary content to which the user-generated auxiliary content relates; means for identifying, using the time data, at least a
20 portion of the user-generated auxiliary content that relates to a portion of the primary content; means for storing, in a first storage file, the identified at least a portion of the user-generated auxiliary content; and means for storing, in association with the first storage file information identifying the portion of the primary content to which the at least a portion of the user-generated auxiliary content is related. The apparatus may
25 further comprise means for carrying out any of the operations described with reference to the first aspect.

In a ninth aspect, this specification describes apparatus comprising: means for
30 receiving user-generated auxiliary content relating to at least one portion of the primary media content; and means for causing transmission of the received user-generated auxiliary content along with time data, the time data indicating at least one portion of the primary media content to which the user-generated auxiliary content relates.

35

Brief Description of the Figures

For a more complete understanding of example embodiments, reference is now made to the following description taken in connection with the accompanying drawings in which:

- 5 Figure 1 is a schematic diagram of a system including a media editing server and a plurality of terminals;
- Figure 2 is a schematic diagram of components a terminal, such as one of those shown in Figure 1;
- Figure 3 is a schematic diagram of components of the media editing server shown in
- 10 Figure 1;
- Figure 4 is a flow chart showing a method that may performed by the terminal of Figure 2;
- Figure 5 is a method that may be performed by the media editing server of Figure 3; and
- 15 Figures 6A and 6B are examples of user interfaces that may be displayed on the display of the terminal.

Detailed Description of Example Embodiments

In the description and drawings, like reference numerals refer to like elements

20 throughout.

Figure 1 is a schematic drawing of a system 1 for providing “crowdsourced” media services. Specifically, the system 1 enables users of user terminals 100, 102, 104 to capture media content and to upload the captured content, via a network 300, to a

25 media editing server 500. The network 300 may be any data network such as a Local Area Network (LAN), Wide Area Network (WAN) or the Internet. In examples described herein, the media content may comprise video content, audio content or audio-visual (AV) content (i.e. video content and corresponding audio content). The media editing server 500 is operable to edit the uploaded media content and to create

30 composite media files for provision back to the same or other users. The composite media files may include media content from many different users.

In one example, the users of the user terminals 100, 102, 104 may all be at the same event (such as a music concert). Each of the users captures AV content, using a video

35 camera and microphone of their user terminal 100, 102, 104. Because each of the user terminals 100, 102, 104 cannot be in exactly the same location, the respective video

portions of the captured AV content will be different but the audio portions may be the same provided that all the users are capturing over a common time period.

5 The users of the terminals 100, 102, 104 subsequently upload the captured AV content to the media editing server 500, either using an application stored on the user terminal 100, 102, 104 or from another terminal, such as a computer with which the user terminal 100, 102, 104 synchronises. At the same time, users are prompted to identify the event, either by entering a description of the event, or by selecting an already-registered event from a pull-down menu. Alternative identification methods may be
10 envisaged, for example by using associated Global Positioning System (GPS) data from the terminals 100, 102, 104 to identify the capture location.

At the media editing server 500, received video clips from the terminals 100, 102, 104 are identified as being associated with a common event. Subsequent analysis of the AV
15 content from each user video clip can then be performed. This allows composite media files, comprising portions of AV content from one or more of the users, to be created. The media editing server 500 may determine the moments at which to switch between content from different users using, for example, beat analysis on the audio part of the AV content. A mechanism for carrying out the beat analysis is described in co-pending
20 patent application PCT/IB2012/052157, which is hereby incorporated by reference in its entirety. The resulting composite media file thus includes video content which switches between content from different users at appropriate moments. The resulting composite media file is split into plural sections or scenes. A scene may be comprised of content from a single source or content from plural different sources.

25 It may be desirable for users to be able to personalize a composite media file by creating audio or audio-visual auxiliary content to a video remix. The auxiliary content may be, for instance, a commentary track. Furthermore, it may be desirable for other users to be able to consume such auxiliary content with the video remix and for these users to
30 be able to control the presentation of the auxiliary content.

It may be desirable for users to provide audio or AV commentary in relation to these composite media files for consumption by other users. However, there may be many different versions of composite media files relating to a particular event. Each version
35 may include some scenes in common with another version, but also some scenes which are not included in the other version. The commentary may hereafter be referred to as

user-generated auxiliary content. The video or AV content to which the user-generated auxiliary content relates may hereafter be referred to as primary content. Described herein are methods and apparatuses for synchronising user-generated auxiliary content with primary content to which the user-generated auxiliary content relates in reconfigurable manner. In other words, the methods and apparatuses allow received user-generated auxiliary content to be utilised with different composite media files to that which was being viewed by the user at the time of creation of the user-generated auxiliary content.

Figure 2 is a schematic illustration of a user terminal 200 according to example embodiments. The user terminal 200 may be one of those shown in of Figure 1. The user terminal 200 is operable, in use, to communicate with the media editing server 500 via the network 300, in order to receive media files (including audio and video content) from the media editing server 500. The user terminal 200 is also operable to capture user-generated auxiliary content and to upload this to the media editing server 500, via the network 300. In some examples, the user terminal 200 may be a portable device such as, but not limited to, a mobile telephone, a personal digital assistant (PDA), a tablet computer, a portable media player, an e-reader or a laptop. In other examples, the user terminal 200 may be stationary user terminal such as a desktop computer.

The user terminal 200 comprises apparatus 2 comprising a controller 106 and at least one non-transitory memory medium 108. The controller 106 is operable to execute computer-readable code 108A stored on the at least on memory 108 and to control the other components of the user terminal 200. The at least one memory 108 may be comprised of any suitable type, or any combination of suitable types of non-transitory memory medium. Suitable types of non-transitory memory include, but are not limited to, ROM, RAM, flash memory and solid state memory. The controller 106 comprises at least one processor 106A, which is operable to execute the computer-readable code 108A stored on the memory 108. The controller 106 may also comprise one or more application specific integrated circuits (not shown).

In addition to the apparatus 2, the user terminal 200 comprises a display 110, a user input interface 112, a communication interface 114, a speaker 116 and a microphone 118. In some examples, in which the user terminal 200 is operable to capture video

content, the user terminal also includes a camera 120. The controller 106 is connected to each of the other components in order to control operation thereof.

5 The communication interface 114 is configured to allow two-way communication with the media editing server 500, via the network 300. The communication interface may be configured to communicate wirelessly via one or more of several protocols such as Global System for Mobile Communications (GSM), Code Division Multiple Access (CDMA), Universal Mobile Telecommunications System (UMTS), Bluetooth and IEEE 802.11 (Wi-Fi). Alternatively or additionally, the communication interface 114 may be
10 configured for wired communication with the network 300.

The display 110 is operable under the control of the controller 106 to output video content for consumption by the user. The video content may have been received from the media editing server 500 or may be provided to the terminal in some other way,
15 such as a wired or wireless connection to a video capture device (not shown), or via the camera 120.

The user input interface 112 allows the user to provide commands to the user terminal 200. These commands are interpreted by the controller 106 which responds
20 accordingly. The user input interface 112 may be a touch sensitive transducer for receiving touch inputs. In such examples, the user input interface 112 and the display 110 may form a touchscreen display. In other examples, the user input interface 112 may be of another type. For example, the user input interface 112 may comprise one or more of mechanical keys, a mouse and cursor, a scroll wheel, a trackball and a voice
25 input interface.

The speaker 116 is operable to output audio content for consumption by the user. As with the video content output by the display 110, the audio content may have been received in any suitable way. For example, the audio content may have been received
30 from the media editing server 500 as part of a received media file, or may have been received via the microphone 118 or via wired or wireless connection to an audio content capture device.

The microphone 118 is configured to capture audio content generated by the user. The
35 camera 120 is configured to capture video content. When the camera 120 and the

microphone operate simultaneously, the resulting captured content may be collectively referred to as AV content.

It will of course be appreciated that the user terminal 200 may include other
5 components, which are not shown in Figure 2. For example, the user terminal 200 may include a power source or an interface configured to receive power from a power source.

Figure 3 is a schematic illustration of a media editing server 500 according to example
10 embodiments.

The media editing server 500 comprises apparatus 5 comprising a controller 506 and at least one non-transitory memory medium 508. The controller 506 is operable to execute computer-readable code 508A stored on the at least on memory 508 and to
15 control the other components of server 500. The at least one memory 508 may be comprised of any suitable type, or any combination of suitable types of memory medium. Suitable types of memory include, but are not limited to, ROM, RAM, flash memory and solid state memory. The controller 506 comprises at least one processor 506A, which is operable to execute the computer-readable code 108A stored on the
20 memory 508. The controller 506 may also comprise one or more application specific integrated circuits (not shown).

The media editing server 500 also comprises a content storage module 512, which comprises at lease one discrete non-transitory memory medium (not shown). The
25 controller 106 is operable to store files containing content in the content storage module 512 and also to retrieve the files therefrom.

The media editing server 500 also comprises an input/output interface 510 configured to allow two-way communication with the network 300. The input/output interface
30 510 may use any suitable type, or combination of suitable types, of communication protocol.

Although referred to as a single server, the media editing server 500 may be constituted by plural server computers, which may or may not be distributed over multiple
35 locations.

Referring now to Figure 2, the apparatus 2 is operable to cause the terminal 200 to receive a primary media file, containing primary media content (e.g. video or AV content), from the media editing server 500. The primary media file may be a composite media file, which is constituted by content from many different sources or users. Receipt of the primary media file may be in response to a request (or command) from the user of the terminal 200. The apparatus 2 subsequently causes the primary media content from the primary media file to be provided by the display 110, and the speaker 112 if required, for consumption by the user. Provision of the primary content for consumption by the user may be referred to as "playback". During playback of the primary media content, the user may input a command which indicates that they would like to record some auxiliary content (e.g. commentary) relating to the primary media content. In response to receipt of this command, the controller 106 causes the microphone 118, and the camera 120 if required, to capture the auxiliary content provided by the user. In addition, the controller 106 also generates time data (e.g. a timestamp) indicating a time, relative to a reference time in the primary media content, at which the capture of auxiliary content was initiated. The reference time in the primary media content may be, for example, the beginning of the primary media content. As such, if the user initiated capture of the auxiliary content sixty-five seconds after the start of the primary media content, the time data would indicate sixty-five seconds

In some examples, the apparatus 2 may be configured such that, in response to receipt of the command, the playback of the primary media content is paused until the relevant hardware and software is ready to begin capture of the auxiliary content. Once the relevant hardware and software is ready to begin auxiliary content capture, playback may continue automatically or in response to a further user input.

Capture of the auxiliary content may continue until the user provides a command to finish content capture or until the primary media content finishes. In response to the termination of the auxiliary content capture, the controller 106 may generate additional time data. This additional time data may indicate a time at which the auxiliary content capture ends. The additional time data may be, for example, a duration indicating the duration for which the auxiliary content capture was performed. Alternatively, the additional time data may be a time relative to the reference point in the primary media content. So, if the auxiliary content capture was commenced at 65s and finished at

180s, the additional time data may specify 180s (i.e. the end point) or may indicate 115s (i.e. the duration).

When the auxiliary content capture is finished, the user may indicate via the user input
5 interface 112 that they want the captured auxiliary content to be uploaded to the media editing server 500. In response to this, the apparatus 2 causes the user-generated auxiliary content to be uploaded to the media editing server 500. The auxiliary content is uploaded to the media editing server 500 along with associated metadata. The metadata includes the generated time data and, in some examples, an identifier for
10 identifying the primary media content to which the user-generated auxiliary content relates. The metadata may be included in the same file as the auxiliary content or may be included in a separate file.

Referring back to Figure 3, the apparatus 5 of the media editing server 500 is
15 configured to cause the media editing server 500 to receive the user-generated auxiliary content and the associated metadata. The media editing server 500 is configured, subsequently, to use the time data included in the metadata to identify the portion of the primary content to which the auxiliary content relates. This may be carried out in two steps. Firstly, the media editing server 500 identifies the primary content file to
20 which the auxiliary content relates. This may be determined based on an identifier which identifies the primary media file and which is present in the metadata. Alternatively, if the server 500 maintained a communication session with the user terminal 200, the primary media file to which the auxiliary content relates may be identified using an identifier stored at the media editing server 500, which identifies
25 the last media file consumed by the user terminal 200. Next, the media editing server 500 utilises the time data extracted from the metadata to identify a portion of the primary media content from the primary media file to which auxiliary content relates. As such, if the time data indicated that the auxiliary content capture was started 65s
30 after the beginning of the primary content and that the auxiliary content is 115s in length, the media editing server 500 would identify the portion of the primary content to which the auxiliary content relates as the portion of the content that is located between 65 and 180 seconds from the beginning of the primary content.

The media editing server is configured, subsequently, to store in the content storage 512
35 the auxiliary content along with an identifier identifying the portion of the primary content to which the auxiliary content relates. The auxiliary content is stored in at least

one storage file. The identifier may be stored as metadata of the storage file. Alternatively, the identifier may be associated with the storage file in some other way.

As mentioned above, in some examples, the primary content conveyed to the user terminal includes content from different scenes. In such cases, the media editing server 500 is configured not only to use the time data extracted from the received metadata to identify the portion of the primary content to which the auxiliary content relates, but also to identify portions of the auxiliary content which relate to primary content from different scenes. This is determined using knowledge of the scenes which constitute the primary media file.

The media editing server 500 is configured to separate the auxiliary content into segments, with each segment relating to a particular scene of the primary content. Each segment of the auxiliary content is then stored in a separate storage file in the content storage module 512. Metadata, which identifies the portion of the primary content to which the segment of auxiliary content relates, is stored in association with each storage file. The metadata may also identify the creator of the auxiliary content.

Let us consider the earlier example in which the auxiliary content relates to a portion of the primary content which runs from 65 seconds to 180 seconds. Let us now assume that the identified portion of the primary content is constituted by two scenes, the first running from 65s to 120s and the second running from 120s to 180s. In this case, the stored data may look as follows:

```
auxiliary_content_segment_1.mp4
primary_media_content_scene_2.mp4_01:05_02:00
user_x

auxiliary_content_segment_2.mp4
primary_media_content_scene_3.mp4_2:00_03:00
user_x
```

The first line in each case is the name of the storage file. The second line is metadata which is associated with the storage file and which identifies the portion (or scene or section) of the primary content to which the segment of the auxiliary content relates. This portion of the metadata may also be referred to as a link. The third line, which

may also be part of the metadata identifies the creator of the auxiliary content, in this case "user x". This allows the server to link the auxiliary content to its creator.

To use general terms, in some example embodiments in which the primary content in the primary media file comprises a first scene and a second scene, the apparatus 5 of the media editing server 500 is configured to identify a first portion of the user-generated auxiliary content which relates to the first scene and a second portion of the user-generated auxiliary content which relates to the second scene. This is carried out using the time data included in the metadata in combination with time data indicating the locations within the primary media file of the first and second scenes. The apparatus 5 is configured, subsequently, to store, in a first storage file, the first portion of the user-generated auxiliary content. Information which identifies the first scene is stored in association with the first storage file. The apparatus 5 also causes the second portion of the user-generated auxiliary content to be stored in a second storage file and information identifying the second scene to be stored in association with the second storage file.

By dividing the auxiliary content up into plural segments, each relating to primary content from a different scene and each stored in a separate file, the media editing server 500 is able subsequently to use the auxiliary content (e.g. user commentary) when creating new composite media files which are made-up of different combinations of scenes.

The media editing 500 server is configured to create a new composite media file by modifying or supplementing one or more portions (or scenes) of primary content with related auxiliary content that is stored in the content storage 512. In order to do this, when the editing media server selects a scene of primary content for inclusion in the new composite media file, the controller 506 examines the metadata associated with the auxiliary content storage files to determine if a segment of auxiliary content which is related to that scene is present in the content storage 512. If such a segment is found, the media editing server 500 modifies or supplements the primary content of that scene to include that segment of auxiliary content.

The media editing server 500 may modify the primary content, for example by replacing the audio part of the primary content with the audio part of the user-generated auxiliary content. In examples in which the auxiliary content alternatively or

additionally includes video content, the media editing server may modify the primary content by including the video content of the auxiliary content in a window within the video of the primary content (i.e. “picture-in-picture”).

5 The media editing server 500 may supplement the primary content by transmitting the segment of auxiliary content to the user terminal when it transmits the new composite media file to the user terminal 200. In this way, the user may then be able to select at the user terminal 200 whether or not to listen to and/or watch auxiliary content. In examples in which two auxiliary content segments provided by different users but
10 which relate to the same portion of primary content are stored in the content storage, the media editing server 500 may supplement the primary content with both segments of auxiliary content. The user may subsequently be able to choose between the auxiliary content from the different sources.

15 Example embodiments have been described above in general terms. More specific example embodiments will now be described below with reference to the flow charts of Figures 4 and 5. Figure 4 is a flow chart illustrating a method according to example embodiments that may be performed by a user terminal (e.g. user terminal 200). Figure 5 is a flow chart illustrating a method according to example embodiments which
20 may be performed by the media editing server 500. Dotted arrows are included in Figures 4 and 5 to indicate the flow of data between the user terminal 200 and the media editing server 500. This provides an indication as to the relative orders of the operations illustrated in the two flow charts.

25 In step 4-1, the user terminal 200 receives a primary media file including AV content. The AV content of the primary media file is separated in to plural different scenes and may be derived from plural different sources. This may be received from the media editing server in response to a request provided by the user, via the user input interface 112 of the terminal.

30 Next, in step S4-2, the controller 106 causes the user terminal 200 to initiate provision, via the display 110 and speakers 116, of the AV content in the received media file for consumption by the user.

35 Subsequent to initiating the provision of the AV content, the controller 106 is responsive to a user input to begin capture of commentary relating to the AV content.

Initiation of the commentary capture is performed in step S4-3. Commentary may be captured using the camera 120 and/or the microphone 118.

5 In step S4-4, the controller 106 generates time data. The time data indicates a time, relative to the AV content, at which capture of the commentary was initiated. The time data is temporarily stored in the memory 108.

Subsequently, in step S4-5, the controller 106 responds to another user input or a determination that the end of AV content has been reached by ceasing capture of the
10 commentary.

Following step S4-5, the method proceeds to step S4-6. In step S4-6, the terminal controller 106 packages and transmits the captured commentary to the media editing server 500 along with associated metadata. The metadata includes the time data
15 generated in step S4-4. The metadata may also include a parameter indicating the length of the captured commentary. In addition, in some example embodiments, the metadata includes an identifier which identifies the primary media file to which the captured commentary relates. The metadata may be packaged in the same file as the captured commentary or may alternatively be packaged in a separate file.

20

Referring now to Figure 5, in step S5-1, the media editing server provides the primary media file to the user terminal 200 for consumption by the user. The primary media file is received by the user terminal 200 in step S4-1 of Figure 4. In some examples, the media editing server 500 may keep a record, in the at least one memory 108, of the
25 identity of the primary media file.

Next, in step S5-2, the media editing server 500 receives, from the user terminal 200, a file containing the captured commentary. The media editing server 500 also receives metadata including time data and, in some examples, an identifier identifying the
30 primary media file to which the commentary relates.

Subsequently, in step S5-3, the media editing server 500 uses the time data extracted from the metadata, as well as knowledge of the composition of the AV content in the primary media file, to identify segments of the commentary which relate to different
35 scenes of content in the primary media file.

Next, in step S5-4, the controller 506 of the media server 500 causes the identified segments of commentary to be stored in separate files in the content storage 512. Next in step S5-5, the controller 506 generates and stores, in association with each separate file, information which identifies the scene of primary content to which the
5 commentary in the file relates.

Next, in step S5-6, which may occur some time after step S5-5, the media editing server 500 creates a composite media file for consumption by the user of a user terminal (which may or may not be the same user terminal 200 that uploaded the commentary)
10 along with one or more of the stored commentary segments. The composite media file may be created in response to a request from a user of the system. In other examples, the media editing server 500 may create the composite media file automatically or in response to another trigger event.

15 In step S5-7, the controller 506 modifies or supplements the AV content in the composite media file using one or more stored commentary segments. This step may include determining which scenes of AV content are to be included in the composite media file. Subsequently, the controller 506 identifies one or more commentary segments which are stored in the content storage 512 and which relate to one or more
20 of the scenes which are to be included in the composite media file. Next, the controller 506 modifies one or more scenes of the composite media file to include a stored commentary segment which relates to that scene. Where the commentary segment includes only audio content, modifying the scene may include replacing or overlaying the audio portion of the scene with the audio content of the stored commentary
25 segment. In examples in which the commentary segment includes video content, modifying the scene may include modifying the video portion of the scene to include (for example, as "picture-in-picture") the video content of the commentary segment.

In alternative embodiments, the AV content of the composite media file may be
30 unmodified, but the controller 506 may prepare the commentary segments which relate to the scenes in the composite media file for transmission to a user terminal 200 and subsequent consumption by a user along with the composite media file. In some examples, the controller 106 may identify plural commentary segments (e.g. from different users) which relate to the same scene. In such embodiments, the controller
35 506 may select some or all of these commentary segments for provision to the user terminal 200 along with the composite media file. The selection as to which

commentary segments are provided to the user may be based on a request received from the user terminal.

Figure 6A shows an example of a graphical user-interface 50 which may be caused, by
5 the controller 106 of the user terminal 200, to be displayed on the display 110 of the
user terminal 200 to allow the user to select which commentary segments are provided.
As can be seen, the graphical user interface (GUI) 50 indicates plural different creators
52 who have created commentary segments which are related to at least one scene of
the composite media file. In addition, the GUI 50 provides a mechanism 54 for
10 allowing the user of the user terminal to select those creators from whom they wish to
receive commentary. In this example, the selection mechanism 54 comprises check
boxes although it will of course be appreciated that another selection mechanism may
instead be used. The GUI 50 also includes a field 56 which indicates the type of
commentary (e.g. audio, video, text, AV etc) created by each different creator. In
15 addition the GUI 50 also indicates a rating field 58 for indicating a rating associated
with the commentary from one or more of the creators. In other examples, the GUI 50
may include a field for indicating the number of scenes of the composite file for which a
particular creator has provided commentary. After the user of the terminal 200 has
selected the creators from which they wish to receive commentary, a request which
20 indicates the selected creators is sent from the user terminal 200 to the media editing
server 500. The GUI of Figure 6A may be displayed to the user of a user terminal after
step S5-6 of Figure 5 in which the media editing server 500 creates the composite
media file. The modification or supplementation of step S5-7 of Figure 5 may be
performed in response to the request from the user terminal 200 which indicates the
25 commentary segments that the user wishes to receive.

Returning now to Figure 5, in step S5-8, the controller 506 causes the created
composite media file to be transmitted to a requesting user terminal 200. In some
examples, one more commentary files including the identified (and optionally selected)
30 commentary segments may also be transmitted to the requesting terminal along with
the composite media file. The one or more commentary files may include, or have
associated therewith, information for allowing the commentary segments to be output
to the user of the terminal at the same time as the scene of the primary media file to
which it relates. This may include time data, which identifies a time, relative to the
35 content of the composite media file, which indicates when the commentary segment

should be output by the user terminal. This information may be generated and included in, or associated with, the one or more commentary files in step S5-7.

5 Following receipt of the composite media file and related commentary segments, the recipient user terminal 200 is operable to output the content in the composite media file for consumption by the user. In some examples, the related commentary segments are automatically output when the composite media file is output. In other examples, such as those in which commentary segment from more than one creator are received. The commentary segments may be provided to the user following in response to a user
10 command. Figure 6B shows a GUI 60 which may be caused to be displayed on the display 110 of the user terminal 200 and which allows the user to select which commentary they wish to view/hear in conjunction with the composite media file.

In this example, the GUI 60 comprises first second and third regions 61, 62, 63. The
15 first region 63 is a content display region in which the video content of the composite media file is displayed the user. The first region 61 also comprises one or more commentary sub-regions 61A. Video commentaries are displayed in the commentary sub-windows. In this example, there is only one sub-window 61A. However, in other examples, plural sub-windows 61A one for each different video commentary may be
20 provided. The size of the sub-windows 61A may be dynamically altered by the controller 106 depending on the number of sub-windows 61A that are currently displayed in the first region 61.

The second region 62 is a commentary control region. The control region includes a
25 mechanism 64 (in this example, check boxes) for allowing the user to select which commentary is output. The second region 62 also includes fields 65 for identifying each of the different commentaries. The second region 62 also includes a control mechanism for allowing the user to control one or more parameters associated with the commentaries. In this example, the control mechanism 66 comprises a slider. The
30 second region 62 also includes a field 67 for indicating the type of each commentary. The commentary control mechanism 66 may allow the user to alter, for example, the volume of the commentary, the size or location of the commentary sub-window or the location in audio space of each commentary.

35 The third region 63 is a content control region. The content control region 63 includes standard controls 68 for allowing the user to control the playback of the content of the

composite media file and the commentary content. The content control region 63 also includes a progress bar 69. The progress bar 69 is configured to indicate to the user the progress of the content by virtue of the position along the progress bar of a slider 69A. The GUI 60 may be configured such that it is responsive to a user input (such as swipe
5 input) to move the slider 69A along the progress bar 69 and thereby to cause a different part of the content to be displayed. Provided adjacent the progress bar is a set of commentary markers 70 for indicating one or more portions (or scenes) of the content for which commentary by a particular creator is available. The locations and lengths of the commentary markers 70 relative to the progress bar indicate the portions or scenes
10 of the content for which commentary is available. In this example, there are two sets of commentary markers 70, a first relating to commentary from a first creator ("USER 1") and the second relating to commentary from a second creator ("USER 2"). The two sets of commentary markers are visually distinguishable from one another. In this example, they include different patterns. However, it will be appreciated that they may be
15 distinguished from one another in any suitable way, such as by using different colours. In this example, the commentary control region 62 includes an indication as to which commentary markers relate to which commentary. This may be provided in any suitable way. In this case, an example of the commentary marker for the commentary from each creator is provided next to the creator's identifier.

20

The GUIs 50, 60 may be provided to the user via a web browser application stored at the user terminal 200 which interacts with an application stored at the server 500. Alternatively, the GUIs may be part of a dedicated application that is stored at the user terminal.

25

In the above described examples, the segments of auxiliary content are provided to the user concurrently with the scenes of primary content with they relate. In other examples, the segments of auxiliary content may be provided to the user immediately before or after the scene of primary content to which they relate. In other words, the
30 scenes of the primary content may be interspersed with segments of auxiliary content. In some examples, one or more segments of auxiliary content may be appended to the primary content.

As mentioned briefly above, in some examples, the auxiliary content may be textual
35 content in addition to or instead of the audio and/or video content. In examples in which the auxiliary content is textual content only, the time data that is transmitted to

the media editing server 500 as metadata may be user entered. In other words, the user terminal 200 may receive the textual content, for example via a keyboard (either a mechanical or graphical, touch-sensitive keyboard) or via a voice-to-text interface.

Subsequently, the user may provide the time data, for example, by defining a portion, in
5 time, of the composite media file to which the auxiliary content relates. For example, the user may define a start time and an end time for the portion. Alternatively, the user terminal 500 may automatically generate the time data by recording the time, relative to playback of the composite media file, at which the textual content was provided by the user. It will also be appreciated that the auxiliary content may alternatively
10 comprise one or more still images (e.g. photographs taken by the user at the event to which the primary media relates). In such examples, the user may specify a portion of the primary content to which the image relates. This may be done, for example, by specifying two times during the playback of the related primary content time between which the still image should be displayed (e.g. as picture-in-picture).

15 Although, in the above examples, the user provides auxiliary content (e.g. commentary) in respect of a composite media file received from the media editing server 500, it will be appreciated that this may not always be the case. In some examples, the user may provide commentary in respect of a non-composite primary media file (e.g. a
20 continuous portion of primary content which was captured by a single user). The content of the primary media file may have been captured by the user who is providing the auxiliary content or may have been captured by a different user to that providing the auxiliary content. In some of these examples, the primary content may have already been uploaded to the media editing server 500 when the auxiliary content is
25 provided. In other examples, the user may provide commentary in respect of primary content which they have captured and which has not yet been uploaded to the media editing server 500. In such examples, the captured primary content may be uploaded to the server 500 along with the auxiliary content.

30 In some examples, the auxiliary content (e.g. commentary) may be captured at the same time as the capturing the primary content. In such examples, the user may be capturing the primary media content using their user terminal 200 and may then indicate to the terminal 200 that they wish also to provide auxiliary content. In response to this indication, the user terminal 200 records data indicative of a time,
35 relative to the capture of the primary content, at which the auxiliary content capture is initiated. When the user wishes to cease capture of the auxiliary content, another

indication is provided to the user terminal 200. The user terminal 200 may respond to this by recording additional time data which indicates the end point of the auxiliary content relative to the primary content. Once the user has finished capturing the primary content and the auxiliary content, the primary content, the auxiliary content
5 and the time data may be uploaded to the media editing server 500.

In such examples, the user terminal 200 may comprise a first camera 120 (e.g. on the back of the terminal) to capture the primary content and a second camera (not shown) (e.g. on the front of the terminal) to capture the video part of the auxiliary content. In
10 addition, there may be a second microphone (not shown) in addition to the main microphone 118. The second microphone may be configured along with an audio processing module (not shown) to utilise audio beam-forming methods to capture primarily sounds coming from the direction of the user rather than sounds which come from elsewhere, which may be captured by the main microphone 118 of the terminal
15 200. In this way, the audio received from the main microphone may form part of the primary content, whereas the audio captured by the additional microphone may form part of the auxiliary content. The additional microphone may be provided at a location of the user terminal 200 that is proximate to the second camera. The additional microphone may be provided on the same side of the user terminal 200 as the second
20 camera. The audio processing module may comprise a combination of hardware and software. In some examples, the additional microphone may be a peripheral device that is temporarily connected to the terminal 200 for the purpose of capturing the auxiliary content. A peripheral microphone such as this may be positioned proximate to the user's mouth, thereby negating or reducing the need to for the terminal to be
25 capable of beam-forming.

It should be realized that the foregoing embodiments should not be construed as limiting. Other variations and modifications will be apparent to persons skilled in the art upon reading the present application. Moreover, the disclosure of the present
30 application should be understood to include any novel features or any novel combination of features either explicitly or implicitly disclosed herein or any generalization thereof and during the prosecution of the present application or of any application derived therefrom, new claims may be formulated to cover any such features and/or combination of such features.

Claims

1. A method comprising:
 - receiving user-generated auxiliary content and associated time data, the user-
 - 5 generated auxiliary content being associated with primary content from a primary media file, the time data indicating at least one portion of the primary content to which the user-generated auxiliary content relates;
 - identifying, using the time data, at least a portion of the user-generated auxiliary content that relates to a portion of the primary content;
 - 10 storing, in a first storage file, the identified at least a portion of the user-generated auxiliary content; and
 - storing, in association with the first storage file information identifying the portion of the primary content to which the at least a portion of the user-generated auxiliary content is related.
 - 15
2. The method of claim 1, wherein the time data indicates a time, relative to playback or capture of the primary content, at which the user-generated auxiliary content was captured.
- 20 3. The method of claim 1 or claim 2, wherein the primary content from the primary media file comprises first and second sections, wherein a first portion of the user-generated auxiliary content relates to the first section and a second portion relates to the second section, the method comprising:
 - identifying, using the time data included in the metadata, the first portion of the
 - 25 user-generated auxiliary content and the second portion of the user-generated auxiliary content;
 - storing, in the first storage file, the first portion of the user-generated auxiliary content, wherein the information stored in association with first storage file identifies the first section of the primary content;
 - 30 storing, in a second storage file, the second portion of the user-generated auxiliary content; and
 - storing, in association with the second storage file, information indentifying the second section of the primary content.
- 35 4. The method of any of claims 1 to 3, comprising:

in response to determining that the portion of the primary content is to be included in a composite media file, retrieving the first storage file from storage; and
modifying the portion of the primary content to include the auxiliary content and including the modified portion in the composite media file; or
5 including the portion of the primary content in the composite media file and preparing the auxiliary content from the first storage file for consumption by a user along with the composite media file.

5. The method of claim 4, wherein the primary content comprises a video content
10 part and a corresponding audio content part.

6. The method of claim 5, wherein the user-generated auxiliary content comprises video content, the method comprising modifying the portion of the primary content by modifying the video content part of the portion of the primary content to include the
15 video content of the at least a portion of the user-generated auxiliary content.

7. The method of claim 5 or claim 6, wherein the user-generated auxiliary content comprises audio content, the method comprising modifying the portion of the primary content by modifying the audio content part of the portion of the primary content to
20 include the audio content of the at least a portion of the user-generated auxiliary content.

8. The method of any preceding claim, comprising:
prior to receiving the user-generated auxiliary content and associated metadata,
25 providing the primary media file, for consumption by the user.

9. The method of claim 8, comprising:
outputting, for consumption by the user, the primary content from the first
media file; and
30 concurrently, capturing the user-generated auxiliary content.

10. The method of claim 9, comprising:
generating the time data, the time data identifying a time, relative to the beginning of the primary content, at which the capturing of the user-generated
35 auxiliary content was initiated.

11. The method of claim 10, comprising:
generating, for transmission along with the time data, information identifying the first media file.
- 5 12. The method of any preceding claim, wherein the received user-generated auxiliary content is received at server apparatus from a user terminal, the auxiliary content having been captured by the user terminal.
13. A method comprising:
10 receiving user-generated auxiliary content relating to at least one portion of primary media content; and
causing transmission of the received user-generated auxiliary content along with time data, the time data indicating at least one portion of the primary media content to which the user-generated auxiliary content relates.
- 15 14. The method of claim 13, comprising:
causing provision, for consumption by a user, of the primary media content; and
receiving the user-generated auxiliary content during provision of the primary media content.
- 20 15. The method of claim 13 or claim 14, comprising:
receiving, during provision of the primary media content, an instruction to initiate capture of the user-generated auxiliary content; and
responding to the instruction, by causing capture of the user-generated auxiliary
25 content to be initiated and by generating the time data, the time data indicating a time, relative to the provision of the primary media content, at which auxiliary content capture was initiated.
- 30 16. The method of claim 13, comprising:
causing capture of the primary media content; and
receiving the user-generated auxiliary content during capture of the primary media content.
- 35 17. The method of any of claims 13 to 16, comprising causing transmission, along with the time data, of information identifying a primary media file containing the primary content to auxiliary content relates.

18. Apparatus comprising at least one processor and at least one memory, the at least one memory having stored thereon computer-readable code which, when executed by the at least one processor, causes the apparatus:

- 5 to receive user-generated auxiliary content and associated time data, the user-generated auxiliary content being associated with primary content from a primary media file, the time data indicating a least one portion of the primary content to which the user-generated auxiliary content relates;
- to identify, using the time data, at least a portion of the user-generated auxiliary
10 content that relates to a portion of the primary content;
- to store, in a first storage file, the identified at least a portion of the user-generated auxiliary content; and
- to store, in association with the first storage file information identifying the portion of the primary content to which the at least a portion of the user-generated
15 auxiliary content is related.

19. The apparatus of claim 16, wherein the time data indicates a time, relative to playback or capture of the primary content, at which the user-generated auxiliary content was captured.

20

20. The apparatus of claim 16 or claim 19, wherein the primary content from the primary media file comprises first and second sections, wherein a first portion of the user-generated auxiliary content relates to the first section and a second portion relates to the second section, and wherein the computer-readable code, when executed by the
25 at least one processor, causes the apparatus:
- to identify, using the time data included in the metadata, the first portion of the user-generated auxiliary content and the second portion of the user-generated auxiliary content;
- to store, in the first storage file, the first portion of the user-generated auxiliary
30 content, wherein the information stored in association with first storage file identifies the first section of the primary content;
- to store, in a second storage file, the second portion of the user-generated auxiliary content; and
- to store, in association with the second storage file, information indentifying the
35 second section of the primary content.

21. The apparatus of any of claims 18 to 20, wherein the computer-readable code, when executed by the at least one processor, causes the apparatus:

in response to determining that the portion of the primary content is to be included in a composite media file, to retrieve the first storage file from storage; and

5 to modify the portion of the primary content to include the auxiliary content and to include the modified portion in the composite media file; or

to include the portion of the primary content in the composite media file and to prepare the auxiliary content from the first storage file for consumption by a user along with the composite media file.

10

22. The apparatus of claim 21, wherein the primary content comprises a video content part and a corresponding audio content part.

23. The apparatus of claim 22, wherein the user-generated auxiliary content
15 comprises video content, and wherein the computer-readable code, when executed by the at least one processor, causes the apparatus to modify the portion of the primary content by modifying the video content part of the portion of the primary content to include the video content of the at least a portion of the user-generated auxiliary content.

20

24. The apparatus of claim 22 or claim 23, wherein the user-generated auxiliary content comprises audio content, and wherein the computer-readable code, when executed by the at least one processor causes the apparatus to modify the portion of the primary content by modifying the audio content part of the portion of the primary
25 content to include the audio content of the at least a portion of the user-generated auxiliary content.

25. The apparatus of any of claims 18 to 24, wherein the computer-readable code, when executed by the at least one processor, causes the apparatus:

30 prior to receiving the user-generated auxiliary content and associated metadata, to cause provision of the primary media file, for consumption by the user.

26. The apparatus of any of claims 18 to 25, wherein the received user-generated auxiliary content is received at a server from a user terminal, the auxiliary content
35 having been captured by the user terminal.

27. Apparatus comprising at least one processor and at least one memory, the at least one memory having stored thereon computer-readable code which, when executed by the at least one processor, causes the apparatus:

5 to receive user-generated auxiliary content relating to at least one portion of the primary media content; and

to cause transmission of the received user-generated auxiliary content along with time data, the time data indicating at least one portion of the primary media content to which the user-generated auxiliary content relates.

10 28. The apparatus of claim 27, wherein the computer-readable code, when executed by the at least one processor, causes the apparatus:

to cause provision, for consumption by a user, of the primary media content; and

15 to receive the user-generated auxiliary content during provision of the primary media content.

29. The apparatus of claim 27 or claim 28, wherein the computer-readable code, when executed by the at least one processor, causes the apparatus:

20 to receive, during provision of the primary media content, an instruction to initiate capture of the user-generated auxiliary content; and

to respond to the instruction, by causing capture of the user-generated auxiliary content to be initiated and by generating the time data, the time data indicating a time, relative to the provision of the primary media content, at which auxiliary content capture was initiated.

25

30. The apparatus of claim 29, wherein the computer-readable code, when executed by the at least one processor, causes the apparatus:

30 to cause capture of the primary media content; and to receive the user-generated auxiliary content during capture of the primary media content.

31. The apparatus of any of claims 27 to 30, wherein the computer-readable code, when executed by the at least one processor, causes the apparatus to cause transmission, along with the time data, of information identifying the primary media file.

35

32. At least one non-transitory computer-readable medium having computer-readable instructions stored thereon, the computer-readable instructions, when executed by at least one processor, causing the at least one processor:

5 to receive user-generated auxiliary content and associated time data, the user-generated auxiliary content being associated with primary content from a primary media file, the time data indicating a least one portion of the primary content to which the user-generated auxiliary content relates;

to identify, using the time data, at least a portion of the user-generated auxiliary content that relates to a portion of the primary content;

10 to store, in a first storage file, the identified at least a portion of the user-generated auxiliary content; and

to store, in association with the first storage file information identifying the portion of the primary content to which the at least a portion of the user-generated auxiliary content is related.

15

33. At least one non-transitory computer-readable medium having computer-readable instructions stored thereon, the computer-readable instructions, when executed by at least one processor, causing the at least one processor:

20 to receive user-generated auxiliary content relating to at least one portion of the primary media content; and

to cause transmission of the received user-generated auxiliary content along with time data, the time data indicating at least one portion of the primary media content to which the user-generated auxiliary content relates.

25 34. Computer-readable code which, when executed by computing apparatus, causes the computing apparatus to perform the method of any of claims 1 to 17.

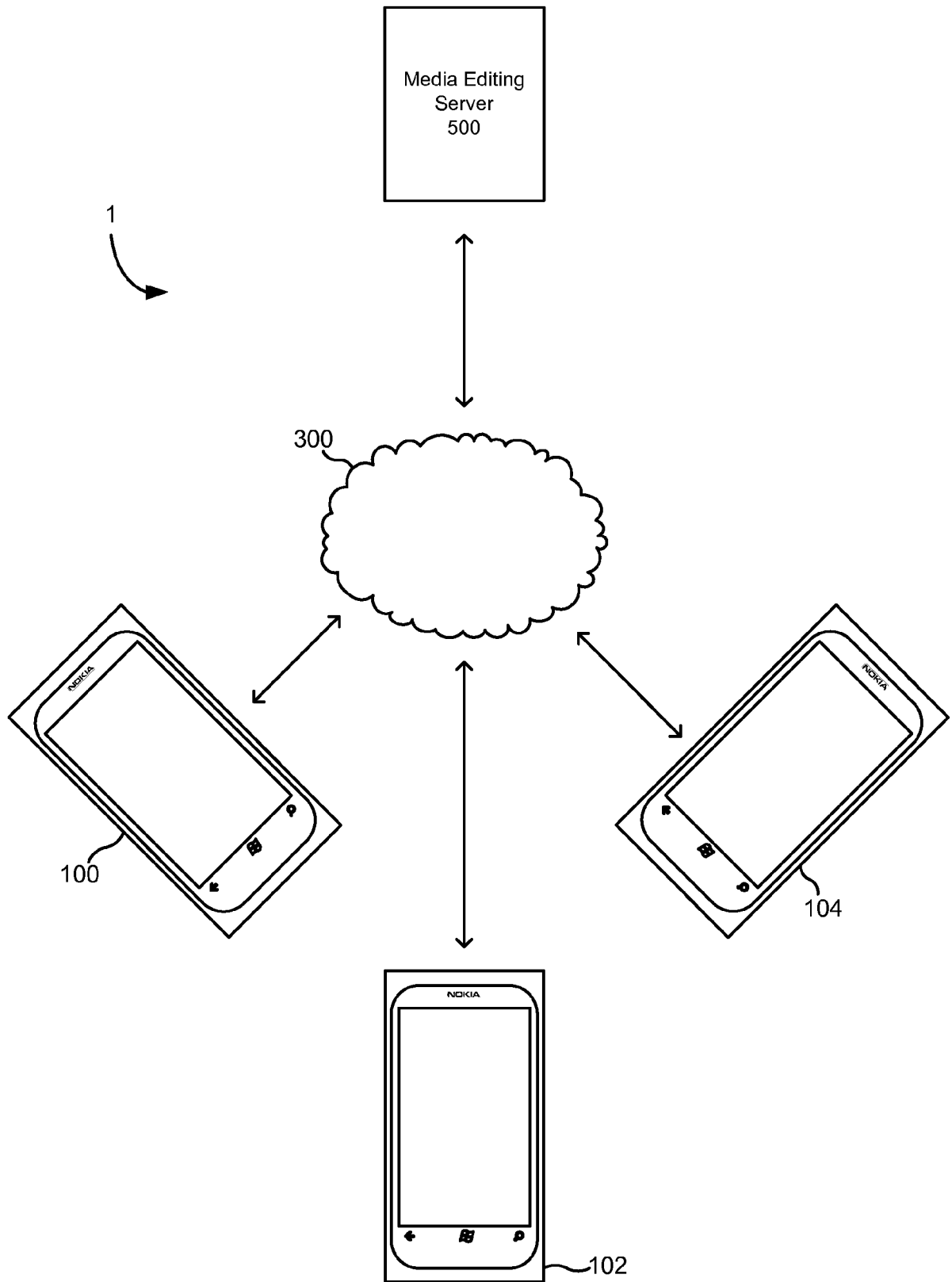


Figure 1

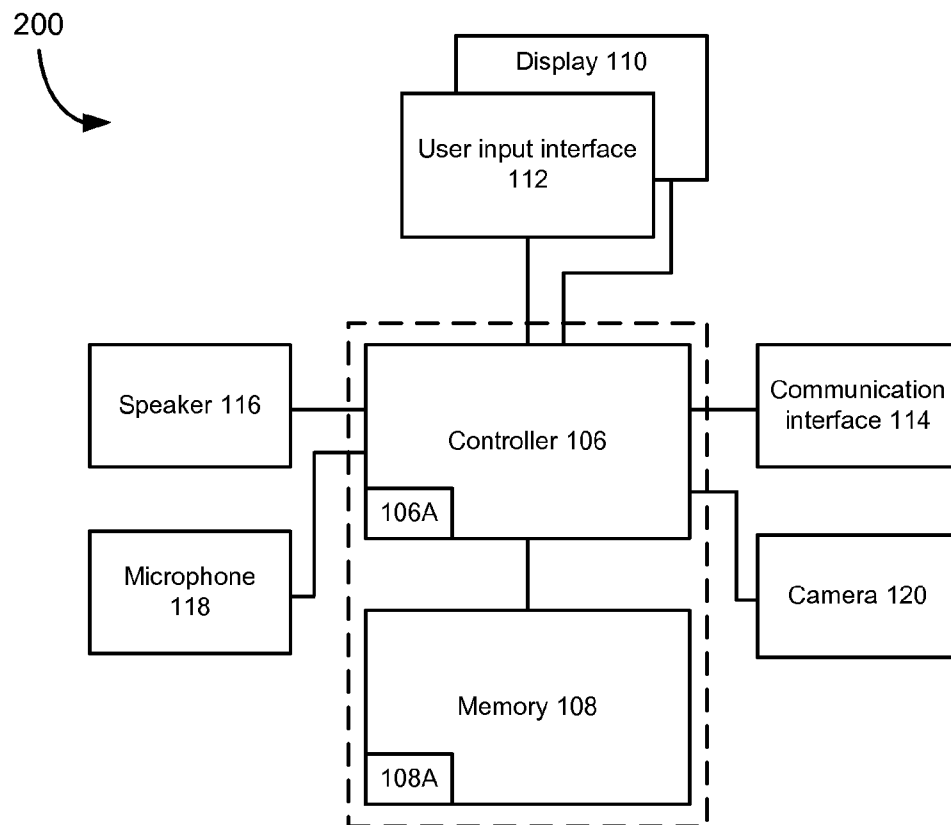
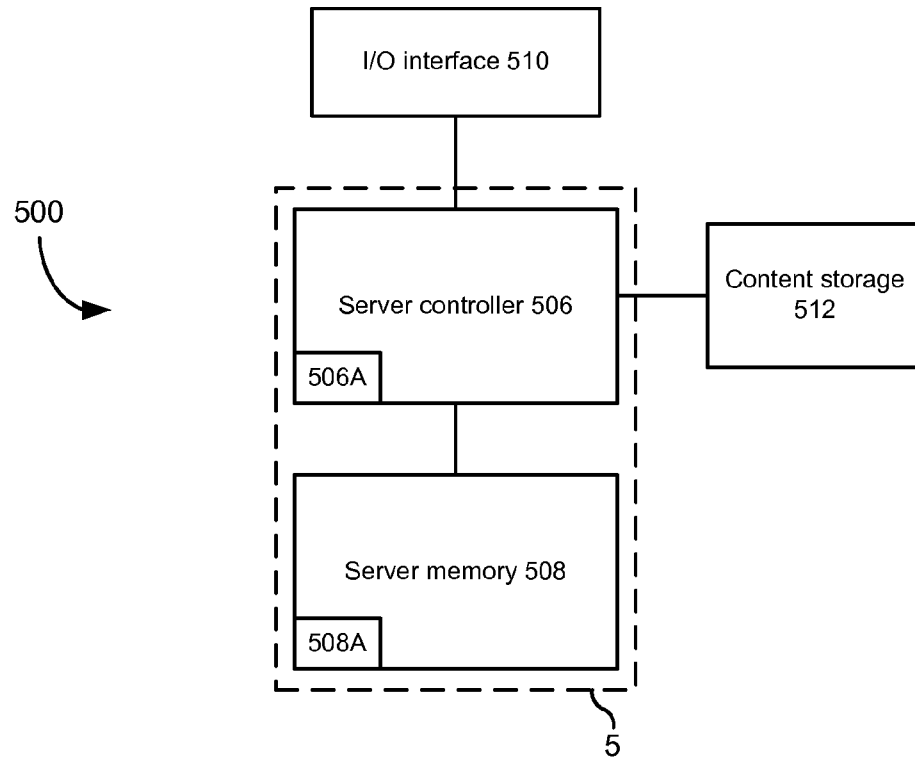
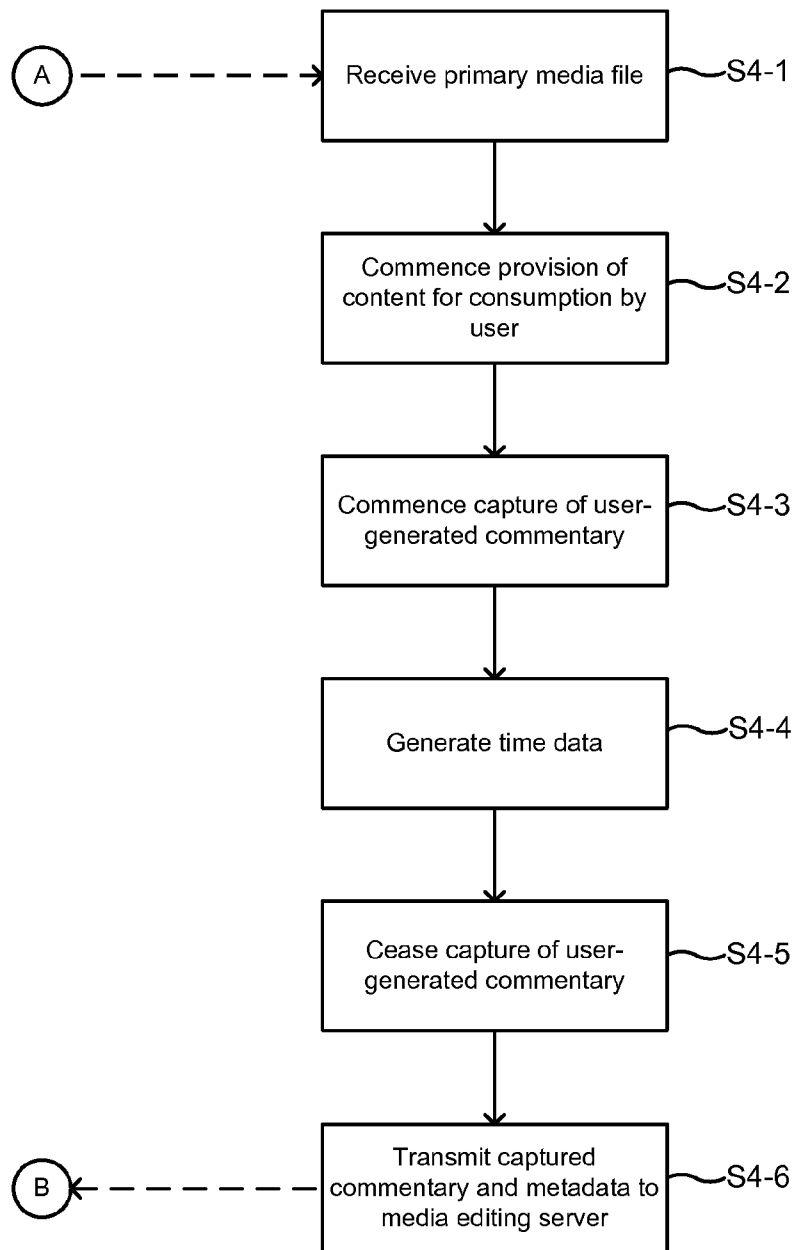
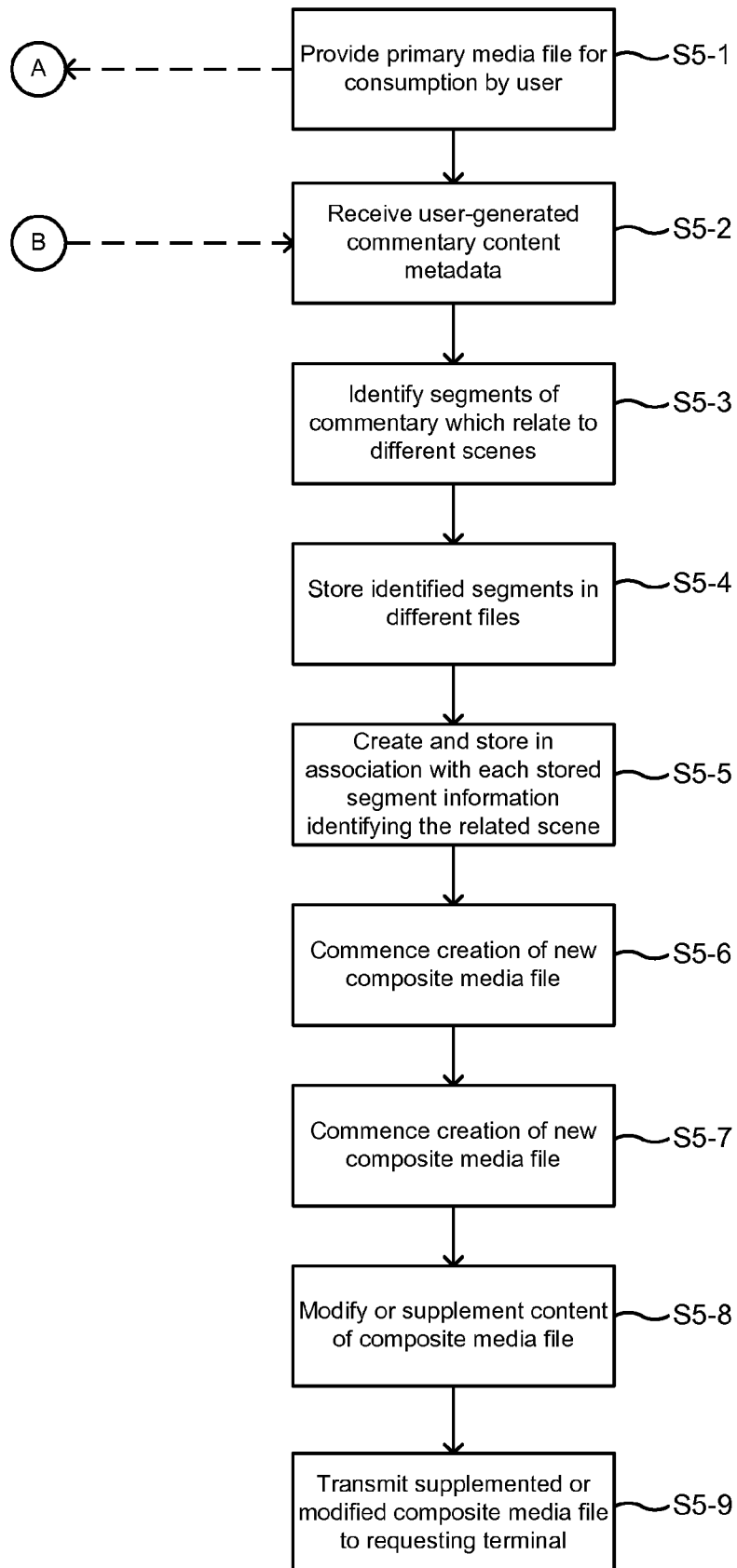


Figure 2

**Figure 3**

**Figure 4**

5/6

**Figure 5**

50

54	✓	52 USER 1	56 Audio	***
	✓	USER 2	Video	*
		USER 3	Video	
		USER 4	Text	**

58

Figure 6A

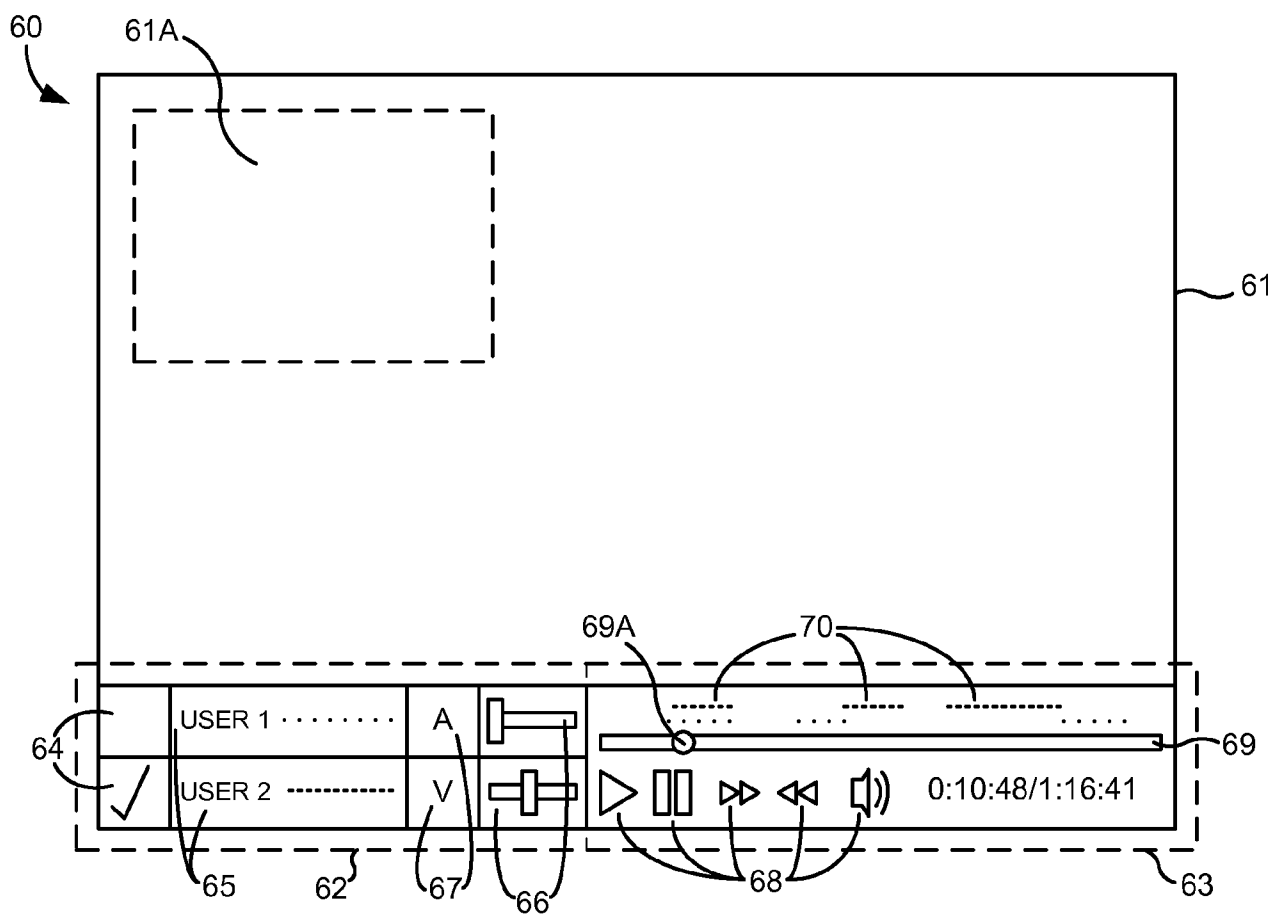


Figure 6B

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IB2012/054794

A. CLASSIFICATION OF SUBJECT MATTER

See extra sheet

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)

IPC: H04N, G11B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
FI, SE, NO, DKElectronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI, IEEE Xplore, Google Scholar

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2007300260 A1 (HOLM JUKKA A [FI] et al.) 27 December 2007 (27.12.2007) abstract; Fig. 1; paragraphs [0002]-[0007], [0016], [0034]-[0049]; claims 1, 2	1-34
X	US 2010125571 A1 (WHITE WILLIAM MACDONALD [US]) 20 May 2010 (20.05.2010) abstract; Figs. 1A, 3, 4; par. [0001]-[0008], [0024]-[0026], [0031]-[0033]; claims 1-3	1-34
X	US 2010130226 A1 (ARRASVUORI JUHA [FI] et al.) 27 May 2010 (27.05.2010) abstract; Figs. 3, 4; par. [0037]-[0060]; claim 1	1-34
A	EP 2151770 A1 (CYBERLINK CORP [TW]) 10 February 2010 (10.02.2010) abstract; Figs. 2, 8; par. [0002], [0026]-[0028], [0044]-[0046]; claims 1, 2	1-34

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

* Special categories of cited documents:

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search
14 June 2013 (14.06.2013)

Date of mailing of the international search report
18 June 2013 (18.06.2013)

Name and mailing address of the ISA/FI
National Board of Patents and Registration of Finland
P.O. Box 1160, FI-00101 HELSINKI, Finland
Facsimile No. +358 9 6939 5328

Authorized officer
Timo Laakso
Telephone No. +358 9 6939 500

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IB2012/054794

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 8051446 B1 (QIAN RICHARD [US] et al.) 01 November 2011 (01.11.2011) the whole document	1-34

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/IB2012/054794

Patent document cited in search report	Publication date	Patent family members(s)	Publication date
US 2007300260 A1	27/12/2007	None	
US 2010125571 A1	20/05/2010	None	
US 2010130226 A1	27/05/2010	WO 2010058064 A1	27/05/2010
EP 2151770 A1	10/02/2010	US 2010036808 A1	11/02/2010
US 8051446 B1	01/11/2011	None	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IB2012/054794

CLASSIFICATION OF SUBJECT MATTER

Int.Cl.

H04N 21/8549 (2011.01)

H04N 5/262 (2006.01)

H04N 5/265 (2006.01)

G11B 27/031 (2006.01)