



(51) International Patent Classification:
G06F 17/30 (2006.01)

(21) International Application Number:
PCT/US2012/042456

(22) International Filing Date:
14 June 2012 (14.06.2012)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US): **EMPIRE TECHNOLOGY DEVELOPMENT LLC** [US/US];
2711 Centerville Road, Suite 400, Wilmington, DE 19808 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **KIM, Seungil** [KR/KR]; 108-605, Dohwa Hyundai 1st Apt, Dohwa-Dong, Mapo-Gu, Seoul, 121-77 (KR). **KO, Youngil** [KR/KR]; 106-1601 Wolkok-Samsung Remian Apt, Hawolkok-Dong, Sungbuk-gu, Seoul, 136-748 (KR).

(74) Agent: **LEE, David, S.**; Brundidge & Stanger, P.C., 2318 Mill Road, Suite 1020, Alexandria, VA 22314 (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, 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, 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.

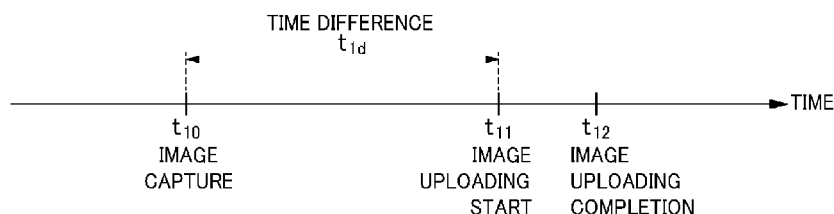
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE,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

Published:

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

(54) Title: DATA DEDUPLICATION MANAGEMENT

FIG. 1



(57) Abstract: Technologies are generally described for a data deduplication management scheme for media files uploaded or to be uploaded to a server. In some examples, a method may include identifying, by a server, a creation time of a media file based at least in part on metadata of the media file; identifying, by the server, an uploading time of the media file; calculating, by the server, a difference between the creation time and the uploading time; and performing, by the server, a data deduplication process when the difference is greater than a predetermined value.



DATA DEDUPLICATION MANAGEMENT

BACKGROUND

[0001] Data deduplication is a data compression technique for eliminating redundant data. In a typical deduplication process, first data is compared to stored data to detect duplicates, that is, to identify or determine whether the first data is unique or not. Then, when the first data is identified as not being unique, the redundant first data is eliminated and replaced with a small reference that points to the stored data.

SUMMARY

[0002] In an example, a method may include identifying, by a server, a creation time of a media file based at least in part on metadata of the media file; identifying, by the server, an uploading time of the media file; calculating, by the server, a difference between the creation time and the uploading time; and performing, by the server, a data deduplication process when the difference is greater than a predetermined value.

[0003] In another example, a data deduplication management system may include a creation time identifying unit configured to identify a creation time of a media file, an uploading time identifying unit configured to identify an uploading time of the media file, and a determination unit configured to determine whether it is permissible to perform a data deduplication process for the media file based at least in part on the creation time identified by the creation time identifying unit and the uploading time identified by the uploading time identifying unit.

[0004] In yet another example, a computer-readable storage medium may store thereon computer-executable instructions that, in response to execution, cause a data deduplication management system to perform operations, including identifying a creation time of a media file, identifying an uploading time of the media file, and determining whether to perform a data deduplication process for the media file based

at least in part on the creation time of the media file and the uploading time of the media file.

[0005] The foregoing summary is illustrative only and is not intended to be in any way limiting. In addition to the illustrative aspects, embodiments, and features described above, further aspects, embodiments, and features will become apparent by reference to the drawings and the following detailed description.

BRIEF DESCRIPTION OF THE FIGURES

[0006] The foregoing and other features of this disclosure will become more apparent from the following description and appended claims, taken in conjunction with the accompanying drawings. Understanding that these drawings depict only several embodiments in accordance with the disclosure and are, therefore, not to be considered limiting of its scope, the disclosure will be described with additional specificity and detail through use of the accompanying drawings, in which:

[0007] **Fig. 1** schematically shows an illustrative example of a timeline related to a process of capturing an image file and uploading the image file to a server, arranged in accordance with at least some embodiments described herein;

[0008] **Fig. 2** schematically shows an illustrative example of a timeline related to a process of capturing a video file and uploading the video file to a server, arranged in accordance with at least some embodiments described herein;

[0009] **Fig. 3** shows a schematic block diagram illustrating an example architecture of a data deduplication management system for providing a data deduplication management scheme for a server, arranged in accordance with at least some embodiments described herein;

[0010] **Fig. 4** shows an example flow diagram of a process for providing a data deduplication management scheme for a server, arranged in accordance with at least some embodiments described herein;

[0011] **Fig. 5** illustrates example computer program products that may be utilized to provide a data deduplication management scheme, arranged in accordance with at least some embodiments described herein; and

[0012] Fig. 6 is a block diagram illustrating an example computing device that may be utilized to provide a data deduplication management scheme, arranged in accordance with at least some embodiments described herein.

DETAILED DESCRIPTION

[0013] In the following detailed description, reference is made to the accompanying drawings, which form a part hereof. In the drawings, similar symbols typically identify similar components, unless context dictates otherwise. The illustrative embodiments described in the detailed description, drawings, and claims are not meant to be limiting. Other embodiments may be utilized, and other changes may be made, without departing from the spirit or scope of the subject matter presented herein. It will be readily understood that the aspects of the present disclosure, as generally described herein, and illustrated in the drawings, can be arranged, substituted, combined, separated, and designed in a wide variety of different configurations, all of which are explicitly contemplated herein.

[0014] This disclosure is generally drawn, *inter alia*, to methods, apparatuses, systems, devices, and computer program products related to a data deduplication management scheme for media files uploaded or to be uploaded to a server.

[0015] Technologies are generally described for determining whether to perform a data deduplication process for a media file based at least in part on a creation time of the media file and an uploading time of the media file, when a user uploads or tries to upload the media file to a server. The aforementioned creation time of the media file may be provided by, *e.g.*, metadata associated with the media file. By way of example, but not limitation, the metadata of the media file may include data regarding means of creation of the media file, purpose of the media file, time and date of creation of the media file, creator or author of the media file, location where the media file was created, standards used for the media file, etc.

[0016] By way of example, but not limitation, the server may include a single server, a collection of server, a server farm, a server cluster, etc. of a photo sharing service provider and/or a video sharing service provider.

[0017] In some examples, the server may identify the creation time of the media file and the uploading time of the media file, and calculate a time difference between the creation time and the uploading time. Then, in some examples, the server may perform the data deduplication process if the time difference is greater than a predetermined value, while otherwise skip the data deduplication process, on the presumption that if there is no or very little time difference between the creation time and the uploading time, the media file may be considered as unique data not requiring the data deduplication process.

[0018] In some examples, the server may identify the creation time of the media file based at least in part on the metadata of the media file. In some examples, when the media file is an image file, the creation time of the media file may be a time when the image file was initially captured. In some examples, when the media file is a video file, the creation time of the media file may be a time when capture of the video file was completed, which may be identified based at least in part on when capture of the video file started and a running time of the video file.

[0019] In some examples, the server may identify the uploading time of the media file as a time when the server starts to receive the media file. In some examples, the server may identify the uploading time of the media file as a time when the server receives from a client device a request for uploading the media file.

[0020] In some examples, the server may keep or store the identified creation time and/or the identified uploading time.

[0021] In some examples, the server may set the predetermined value based at least in part on a processing power and/or a storage capacity thereof.

[0022] Fig. 1 schematically shows an illustrative example of a timeline related to a process of capturing an image file and uploading the image file to a server, arranged in accordance with at least some embodiments described herein.

[0023] As depicted in Fig. 1, it may be assumed that a device captures the image file at t_{10} and starts to upload the image file to the server at t_{11} , and the uploading of the image file is completed at t_{12} . In some embodiments, when the device creates the image file, the device may store t_{10} as metadata of the image file, which may be a portion of the image file. Examples of the device may include, but

are not limited to, a smartphone, a mobile phone, a personal digital assistant (PDA), a tablet, a mobile game console, and any other devices that may access to the server via any network. Examples of the server may include, but are not limited to, a single server, a collection of server, a server farm, a server cluster, etc. of a photo sharing service provider such as, for example, InstagramTM, FlickrTM, PicasaTM, etc.

[0024] In some embodiments, the server may identify a creation time of the image file as t_{10} . In some embodiments, the server may identify the creation time based at least in part on the metadata of the image file.

[0025] In some embodiments, the server may identify an uploading time of the image file as t_{11} . In some alternative embodiments, the server may identify the uploading time of the image file as a time when the server receives from the device a request for uploading the image file (not shown in Fig. 1).

[0026] Then, in some embodiments, the server may calculate a time difference t_{1d} as $t_{11} - t_{10}$, and compare time difference t_{1d} with a predetermined value, to determine whether it is permissible to perform a data deduplication process for the image file.

[0027] Fig. 2 schematically shows an illustrative example of a timeline related to a process of capturing a video file and uploading the video file to a server, arranged in accordance with at least some embodiments described herein.

[0028] As depicted in Fig. 2, it may be assumed that a device starts to capture the video file at t_{20} , stops capturing of the video file at t_{21} , and starts to upload the video file to the server at t_{22} , and the uploading of the video file is completed at t_{23} . In such cases, a running time t_{2r} of the video file may be $t_{21} - t_{20}$. In some embodiments, when the device creates the video file, the device may store t_{20} , t_{21} , and/or t_{2r} as metadata of the video file, which may be a portion of the video file. As with the aforementioned image capture, examples of the device may include, but are not limited to, a smartphone, a mobile phone, a personal digital assistant (PDA), a tablet, a mobile game console, and any other devices that may access to the server via any network. Examples of the server may include, but are not limited to, a single server, a collection of server, a server farm, a server cluster, etc. of a video sharing service provider such as, for example, YouTubeTM, FlickrTM, GoogleTM Video, etc.

[0029] In some embodiments, the server may identify a creation time of the video file as t_{21} . In some embodiments, the server may identify the creation time by identifying t_{20} and t_{2r} based at least in part on the metadata of the video file, and calculating a sum of t_{20} and t_{2r} .

[0030] In some embodiments, the server may identify an uploading time of the video file as t_{22} . In some alternative embodiments, the server may identify the uploading time of the video file as a time when the server receives from the device a request for uploading the video file (not shown in Fig. 2).

[0031] Then, in some embodiments, the server may calculate a time difference t_{2d} as $t_{22} - t_{21}$, and compare time difference t_{2d} with a predetermined value, to determine whether it is permissible to perform a data deduplication process for the video file.

[0032] Fig. 3 shows a schematic block diagram illustrating an example architecture of a data deduplication management system for providing a data deduplication management scheme for a server, arranged in accordance with at least some embodiments described herein.

[0033] As depicted, a data deduplication management system 300 may include a creation time identifying unit 310, an uploading time identifying unit 320, a determination unit 330 and a data deduplication unit 340. Although illustrated as discrete components, various components may be divided into additional components, combined into fewer components, or eliminated altogether while being contemplated within the scope of the disclosed subject matter. Those skilled in the art will recognize that data deduplication management system 300 may be a component of the server or be separate from but still controlled by the server. Those skilled in the art will also recognize that each component of data deduplication management system 300 may be controlled by one or more segments of a server farm or a server cluster. In such cases, the respective components may be controlled by the same segments or different segments, depending on the desired implementation.

[0034] Creation time identifying unit 310 may be configured to identify a creation time of a media file. In some embodiments, creation time identifying unit 310 may be configured to identify the creation time of the media file based at least in

part on metadata of the media file, which may include, but are not limited to, data regarding means of creation of the media file, purpose of the media file, time and date of creation of the media file, creator or author of the media file, location where the media file was created, standards used for the media file, etc. That is, creation time identifying unit 310 may be configured to extract the metadata from the media file, analyze the metadata, and identify the creation time of the media file based on the metadata.

[0035] In some embodiments, when the media file is an image file, creation time identifying unit 310 may identify the creation time as a time when the image file was initially captured. In some embodiments, when the media file is a video file, creation time identifying unit 310 may identify the creation time as a time when capture of the video file was completed. By way of example, but not limitation, creation time identifying unit 310 may identify the time when capture of the video file was completed based at least in part on when capture of the video file started and a running time of the video file.

[0036] Uploading time identifying unit 320 may be configured to identify an uploading time of the media file. In some embodiments, uploading time identifying unit 320 may be configured to identify the uploading time of the media file as a time when the server starts to receive the media file. In some alternative embodiments, uploading time identifying unit 320 may identify the uploading time of the media file as a time when the server receives a request for uploading the media file.

[0037] Determination unit 330 may be configured to determine whether it is permissible to perform a data deduplication process for the media file based at least in part on the creation time identified by creation time identifying unit 310 and the uploading time identified by uploading time identifying unit 320. In some embodiments, determination unit 330 may be configured to determine that it is permissible to perform the data deduplication process for the media file when a difference between the creation time and the uploading time is greater than a predetermined value.

[0038] In some embodiments, the predetermined value may vary based at least in part on a processing power of the server and/or a storage capacity of the

server. By way of example, but not limitation, when the storage capacity of the server is relatively large, and/or the processing power of the server is relatively insufficient, the predetermined value may set relatively high. Meanwhile, when the storage capacity of the server is relatively insufficient, and/or the processing power of the server is relatively sufficient, the predetermined value may set relatively low.

[0039] Data deduplication unit 340 may be configured to perform the data deduplication process for the media file when determination unit 330 determines that it is permissible to perform the data deduplication process for the media file.

[0040] As such, data deduplication management system 300 may reduce the server's analysis overhead for deduplication by skipping the data deduplication process for some media files determined as being unique based at least in part on their creation times and uploading times.

[0041] Fig. 4 shows an example flow diagram of a process for providing a data deduplication management scheme for a server, arranged in accordance with at least some embodiments described herein.

[0042] The process in Fig. 4 may be implemented in a server or a data deduplication management system such as data deduplication management system 300 described above, which may be a component of the server or separate from but controlled by the server. An example process 400 may include one or more operations, actions, or functions as illustrated by one or more blocks 410, 420, 430, 440 and/or 450. Although illustrated as discrete blocks, various blocks may be divided into additional blocks, combined into fewer blocks, or eliminated, depending on the desired implementation. Processing may begin at block 410.

[0043] At block 410 (Identify Creation Time and Uploading Time of Media File), the server or the data deduplication management system may identify a creation time and an uploading time of a media file. In some embodiments, the server or the data deduplication management system may identify the creation time of the media file based at least in part on metadata of the media file.

[0044] In some embodiments, when the media file is an image file, the server or the data deduplication management system may identify the creation time as a time when the image file was initially captured. In some embodiments, when the media

file is a video file, the server or the data deduplication management system may identify the creation time as a time when capture of the video file was completed. By way of example, but not limitation, the server or the data deduplication management system may identify the time when capture of the video file was completed based at least in part on when capture of the video file started and a running time of the video file.

[0045] In some embodiments, the server or the data deduplication management system may identify the uploading time of the media file as a time when the server starts to receive the media file. In some alternative embodiments, the server or the data deduplication management system may identify the uploading time of the media file as a time when the server receives a request for uploading the media file. Processing may continue from block 410 to block 420.

[0046] At block 420 (Calculate Time Difference between Creation Time and Uploading Time), the server or the data deduplication management system may calculate a time difference between the creation time and the uploading time. Processing may continue from block 420 to block 430.

[0047] At block 430 (Time Difference > Predetermined Value?), the server or the data deduplication management system may determine whether the time difference is greater than a predetermined value. In some embodiments, the server or the data deduplication management system may set the predetermined value based at least in part on at least one of a processing power of the server and a storage capacity of the server. When the server or the data deduplication management system determines that the time difference is greater than a predetermined value, processing may continue from block 430 to block 440. Otherwise, processing may continue from block 430 to block 450.

[0048] At block 440 (Perform Data Deduplication Process), the server or the data deduplication management system may perform a data deduplication process for the media file. By way of example, but not limitation, the server or the data deduplication management system may compare the media file with other stored data to detect duplicates, and when the media file is identified as not being unique, the

server or the data deduplication management system may eliminate the redundant media file.

[0049] At block 450 (Skip Data Deduplication Process), the server or the data deduplication management system may skip the data deduplication process for the media file, on the presumption that if the time difference between the creation time and the uploading time is less than the predetermined value, the media file may be considered as unique data not requiring the data deduplication process. The server or the data deduplication management system may store the media file without performing the data deduplication process.

[0050] As such, the server's analysis overhead for deduplication may be reduced, by skipping the data deduplication process for some media files determined as being unique based at least in part on their creation times and uploading times.

[0051] One skilled in the art will appreciate that, for this and other processes and methods disclosed herein, the functions performed in the processes and methods may be implemented in differing order. Furthermore, the outlined steps and operations are only provided as examples, and some of the steps and operations may be optional, combined into fewer steps and operations, or expanded into additional steps and operations without detracting from the essence of the disclosed embodiments.

[0052] Fig. 5 illustrates example computer program products that may be utilized to provide a data deduplication management scheme, arranged in accordance with at least some embodiments described herein.

[0053] Program product 500 may include a signal bearing medium 502. Signal bearing medium 502 may include one or more instructions 504 that, when executed by, for example, a processor, may provide the functionality described above with respect to Figs. 1-4. By way of example, instructions 504 may include: one or more instructions for identifying a creation time of a media file; one or more instructions for identifying an uploading time of the media file; or one or more instructions for determining whether to perform a data deduplication process for the media file based at least in part on the creation time of the media file and the uploading time of the media file. Thus, for example, referring to Fig. 3, data

deduplication management system 300 may undertake one or more of the blocks shown in Fig. 4 in response to instructions 504.

[0054] In some implementations, signal bearing medium 502 may encompass a computer-readable medium 506, such as, but not limited to, a hard disk drive, a CD, a DVD, a digital tape, memory, etc. In some implementations, signal bearing medium 502 may encompass a recordable medium 508, such as, but not limited to, memory, read/write (R/W) CDs, R/W DVDs, etc. In some implementations, signal bearing medium 502 may encompass a communications medium 510, such as, but not limited to, a digital and/or an analog communication medium (e.g., a fiber optic cable, a waveguide, a wired communications link, a wireless communication link, etc.). Thus, for example, program product 500 may be conveyed to one or more modules of data deduplication management system 300 by an RF signal bearing medium 502, where the signal bearing medium 502 is conveyed by a wireless communications medium 510 (e.g., a wireless communications medium conforming with the IEEE 802.11 standard).

[0055] Fig. 6 is a block diagram illustrating an example computing device that may be utilized to provide a data deduplication management scheme, arranged in accordance with at least some embodiments described herein.

[0056] In these examples, elements of computing device 600 may be arranged or configured for a server. In a very basic configuration 602, computing device 600 typically includes one or more processors 604 and a system memory 606. A memory bus 608 may be used for communicating between processor 604 and system memory 606.

[0057] Depending on the desired configuration, processor 604 may be of any type including but not limited to a microprocessor (μ P), a microcontroller (μ C), a digital signal processor (DSP), or any combination thereof. Processor 604 may include one or more levels of caching, such as a level one cache 610 and a level two cache 612, a processor core 614, and registers 616. An example processor core 614 may include an arithmetic logic unit (ALU), a floating point unit (FPU), a digital signal processing core (DSP Core), or any combination thereof. An example memory

controller 618 may also be used with processor 604, or in some implementations memory controller 618 may be an internal part of processor 604.

[0058] Depending on the desired configuration, system memory 606 may be of any type including but not limited to volatile memory (such as RAM), non-volatile memory (such as ROM, flash memory, etc.) or any combination thereof. System memory 606 may include an operating system 620, one or more applications 622, and program data 624. Application 622 may include instructions 626 that may be arranged to perform the functions as described herein including the actions described with respect to the data deduplication management system 300 architecture as shown in Fig. 3 or including the actions described with respect to the flow charts shown in Fig. 4. In some examples, application 622 may be arranged to operate with program data 624 on an operating system 620 such that implementations for instructions for an electronic device as described herein.

[0059] Computing device 600 may have additional features or functionality, and additional interfaces to facilitate communications between basic configuration 602 and any required devices and interfaces. For example, a bus/interface controller 630 may be used to facilitate communications between basic configuration 602 and one or more data storage devices 632 via a storage interface bus 634. Data storage devices 632 may be removable storage devices 636, non-removable storage devices 638, or a combination thereof. Examples of removable storage and non-removable storage devices include magnetic disk devices such as flexible disk drives and hard-disk drives (HDD), optical disk drives such as compact disk (CD) drives or digital versatile disk (DVD) drives, solid state drives (SSD), and tape drives to name a few. Example computer storage media may include volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information, such as computer readable instructions, data structures, program modules, or other data.

[0060] System memory 606, removable storage devices 636 and non-removable storage devices 638 are examples of computer storage media. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical

storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which may be used to store the desired information and which may be accessed by computing device 600. Any such computer storage media may be part of computing device 600.

[0061] Computing device 600 may also include an interface bus 640 for facilitating communication from various interface devices (e.g., output devices 642, peripheral interfaces 644, and communication devices 646) to basic configuration 602 via bus/interface controller 630. Example output devices 642 include a graphics processing unit 648 and an audio processing unit 650, which may be configured to communicate to various external devices such as a display or speakers via one or more A/V ports 652. Example peripheral interfaces 644 include a serial interface controller 654 or a parallel interface controller 656, which may be configured to communicate with external devices such as input devices (e.g., keyboard, mouse, pen, voice input device, touch input device, etc.) or other peripheral devices (e.g., printer, scanner, etc.) via one or more I/O ports 658. An example communication device 646 includes a network controller 660, which may be arranged to facilitate communications with one or more other computing devices 662 over a network communication link via one or more communication ports 664.

[0062] The network communication link may be one example of a communication media. Communication media may typically be embodied by computer readable instructions, data structures, program modules, or other data in a modulated data signal, such as a carrier wave or other transport mechanism, and may include any information delivery media. A "modulated data signal" may be a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal. By way of example, and not limitation, communication media may include wired media such as a wired network or direct-wired connection, and wireless media such as acoustic, radio frequency (RF), microwave, infrared (IR) and other wireless media. The term computer readable media as used herein may include both storage media and communication media.

[0063] Computing device 600 may be implemented as a portion of a small-form factor portable (or mobile) electronic device such as a cell phone, a personal

data assistant (PDA), a personal media player device, a wireless web-watch device, a personal headset device, an application specific device, or a hybrid device that include any of the above functions. Computing device 600 may also be implemented as a personal computer including both laptop computer and non-laptop computer configurations.

[0064] The present disclosure is not to be limited in terms of the particular embodiments described in this application, which are intended as illustrations of various aspects. Many modifications and variations can be made without departing from its spirit and scope, as will be apparent to those skilled in the art. Functionally equivalent methods and apparatuses within the scope of the disclosure, in addition to those enumerated herein, will be apparent to those skilled in the art from the foregoing descriptions. Such modifications and variations are intended to fall within the scope of the appended claims. The present disclosure is to be limited only by the terms of the appended claims, along with the full scope of equivalents to which such claims are entitled. It is to be understood that this disclosure is not limited to particular methods, reagents, compounds, compositions or biological systems, which can, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting.

[0065] With respect to the use of substantially any plural and/or singular terms herein, those having skill in the art can translate from the plural to the singular and/or from the singular to the plural as is appropriate to the context and/or application. The various singular/plural permutations may be expressly set forth herein for sake of clarity.

[0066] It will be understood by those within the art that, in general, terms used herein, and especially in the appended claims (*e.g.*, bodies of the appended claims) are generally intended as "open" terms (*e.g.*, the term "including" should be interpreted as "including but not limited to," the term "having" should be interpreted as "having at least," the term "includes" should be interpreted as "includes but is not limited to," etc.). It will be further understood by those within the art that if a specific number of an introduced claim recitation is intended, such an intent will be explicitly

recited in the claim, and in the absence of such recitation no such intent is present. For example, as an aid to understanding, the following appended claims may contain usage of the introductory phrases "at least one" and "one or more" to introduce claim recitations. However, the use of such phrases should not be construed to imply that the introduction of a claim recitation by the indefinite articles "a" or "an" limits any particular claim containing such introduced claim recitation to embodiments containing only one such recitation, even when the same claim includes the introductory phrases "one or more" or "at least one" and indefinite articles such as "a" or "an" (*e.g.*, "a" and/or "an" should be interpreted to mean "at least one" or "one or more"); the same holds true for the use of definite articles used to introduce claim recitations. In addition, even if a specific number of an introduced claim recitation is explicitly recited, those skilled in the art will recognize that such recitation should be interpreted to mean at least the recited number (*e.g.*, the bare recitation of "two recitations," without other modifiers, means at least two recitations, or two or more recitations). Furthermore, in those instances where a convention analogous to "at least one of A, B, and C, etc." is used, in general such a construction is intended in the sense one having skill in the art would understand the convention (*e.g.*, "a system having at least one of A, B, and C" would include but not be limited to systems that have A alone, B alone, C alone, A and B together, A and C together, B and C together, and/or A, B, and C together, etc.). In those instances where a convention analogous to "at least one of A, B, or C, etc." is used, in general such a construction is intended in the sense one having skill in the art would understand the convention (*e.g.*, "a system having at least one of A, B, or C" would include but not be limited to systems that have A alone, B alone, C alone, A and B together, A and C together, B and C together, and/or A, B, and C together, etc.). It will be further understood by those within the art that virtually any disjunctive word and/or phrase presenting two or more alternative terms, whether in the description, claims, or drawings, should be understood to contemplate the possibilities of including one of the terms, either of the terms, or both terms. For example, the phrase "A or B" will be understood to include the possibilities of "A" or "B" or "A and B."

[0067] In addition, where features or aspects of the disclosure are described in terms of Markush groups, those skilled in the art will recognize that the disclosure is also thereby described in terms of any individual member or subgroup of members of the Markush group.

[0068] As will be understood by one skilled in the art, for any and all purposes, such as in terms of providing a written description, all ranges disclosed herein also encompass any and all possible subranges and combinations of subranges thereof. Any listed range can be easily recognized as sufficiently describing and enabling the same range being broken down into at least equal halves, thirds, quarters, fifths, tenths, etc. As a non-limiting example, each range discussed herein can be readily broken down into a lower third, middle third and upper third, etc. As will also be understood by one skilled in the art all language such as "up to," "at least," and the like include the number recited and refer to ranges which can be subsequently broken down into subranges as discussed above. Finally, as will be understood by one skilled in the art, a range includes each individual member. Thus, for example, a group having 1-3 cells refers to groups having 1, 2, or 3 cells. Similarly, a group having 1-5 cells refers to groups having 1, 2, 3, 4, or 5 cells, and so forth.

[0069] From the foregoing, it will be appreciated that various embodiments of the present disclosure have been described herein for purposes of illustration, and that various modifications may be made without departing from the scope and spirit of the present disclosure. Accordingly, the various embodiments disclosed herein are not intended to be limiting, with the true scope and spirit being indicated by the following claims.

CLAIMS

What is claimed is:

1. A method comprising:
identifying, by a server, a creation time of a media file based at least in part on metadata of the media file;
identifying, by the server, an uploading time of the media file;
calculating, by the server, a difference between the creation time and the uploading time; and
performing, by the server, a data deduplication process when the difference is greater than a predetermined value.
2. The method of Claim 1, wherein the media file is an image file, and the creation time of the media file is a time when the image file was initially made.
3. The method of Claim 1, wherein the media file is a video file, and the creation time of the media file is a time when capture of the video file was completed.
4. The method of Claim 1, wherein the media file is a video file, and wherein the identifying of the creation time of the media file is based at least in part on when capture of the video file started and a running time of the video file.
5. The method of Claim 1, wherein the uploading time of the media file is a time when the server starts to receive the media file.
6. The method of Claim 1, wherein the uploading time of the media file is a time when the server receives from a client device a request for uploading the media file.

7. The method of Claim 1, wherein the predetermined value is set by the server based at least in part on at least one of a processing power of the server and a storage capacity of the server.

8. The method of Claim 1, wherein the server stores at least one of the creation time and the uploading time.

9. The method of Claim 1, further comprising:
skipping the data deduplication process when the difference is less than the predetermined value.

10. A data deduplication management system, comprising:
a creation time identifying unit configured to identify a creation time of a media file;

an uploading time identifying unit configured to identify an uploading time of the media file; and

a determination unit configured to determine whether it is permissible to perform a data deduplication process for the media file based at least in part on the creation time identified by the creation time identifying unit and the uploading time identified by the uploading time identifying unit.

11. The data deduplication management system of Claim 10, further comprising:

a data deduplication unit configured to perform the data deduplication process for the media file when the determination unit determines that it is permissible to perform the data deduplication process for the media file.

12. The data deduplication management system of Claim 10, wherein the creation time identifying unit is further configured to identify the creation time of the media file based at least in part on metadata of the media file.

13. The data deduplication management system of Claim 10, wherein the determination unit is further configured to determine that it is permissible to perform the data deduplication process for the media file when a difference between the creation time and the uploading time is greater than a predetermined value.

14. The data deduplication management system of Claim 13, wherein the predetermined value is set by a server based at least in part on at least one of a processing power of the server and a storage capacity of the server.

15. The data deduplication management system of Claim 10, wherein the media file is an image file and the creation time of the media file is a time when the image file was initially captured.

16. The data deduplication management system of Claim 10, wherein the media file is a video file and the creation time of the media file is a time when capture of the video file was completed.

17. The data deduplication management system of Claim 10, wherein the media file is a video file, and

wherein the creation time identifying unit is further configured to identify the capture time of the media file based at least in part on when capture of the video file started and a running time of the video file.

18. The data deduplication management system of Claim 10, wherein the uploading time of the media file is a time when a server starts to receive the media file.

19. The data deduplication management system of Claim 10, wherein the uploading time of the media file is a time when a server receives from a client device a request for uploading the media file.

20. A computer-readable storage medium having stored thereon computer-executable instructions that, in response to execution, cause a data deduplication management system to perform operations, comprising:

identifying a creation time of a media file;

identifying an uploading time of the media file; and

determining whether to perform a data deduplication process for the media file based at least in part on the creation time of the media file and the uploading time of the media file.

21. The computer-readable storage medium of Claim 20, wherein the operations further comprise:

performing the data deduplication process for the media file when a difference between the creation time and the uploading time is greater than a predetermined value.

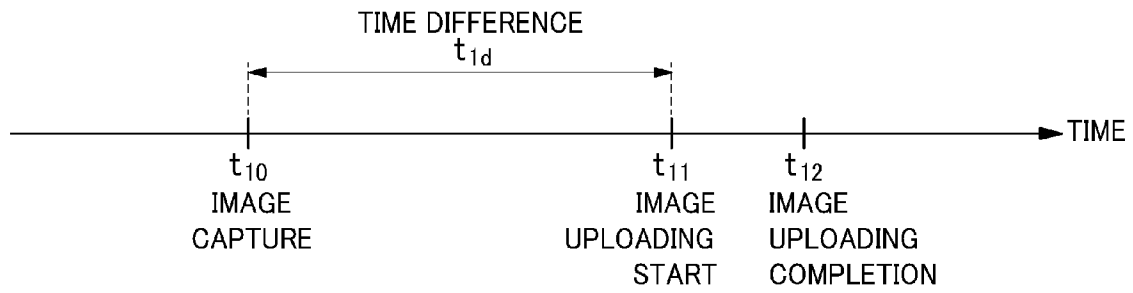
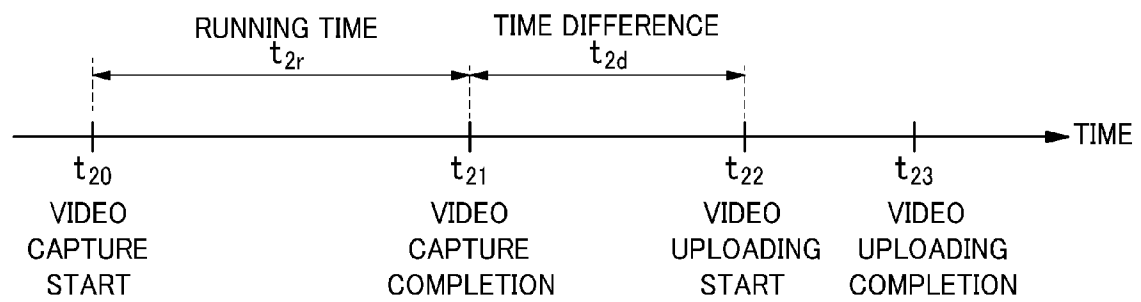
FIG. 1*FIG. 2*

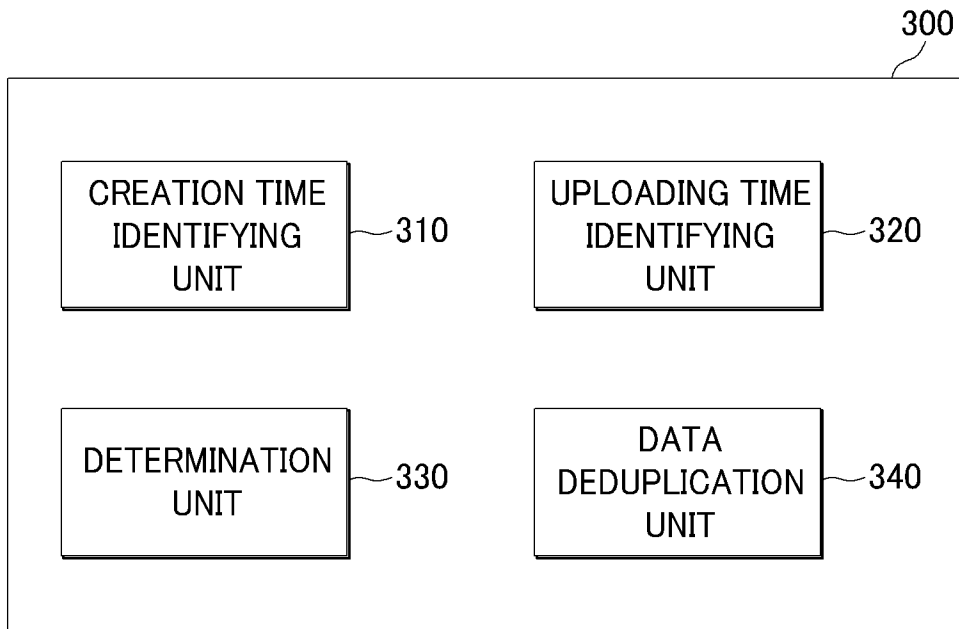
FIG. 3

FIG. 4

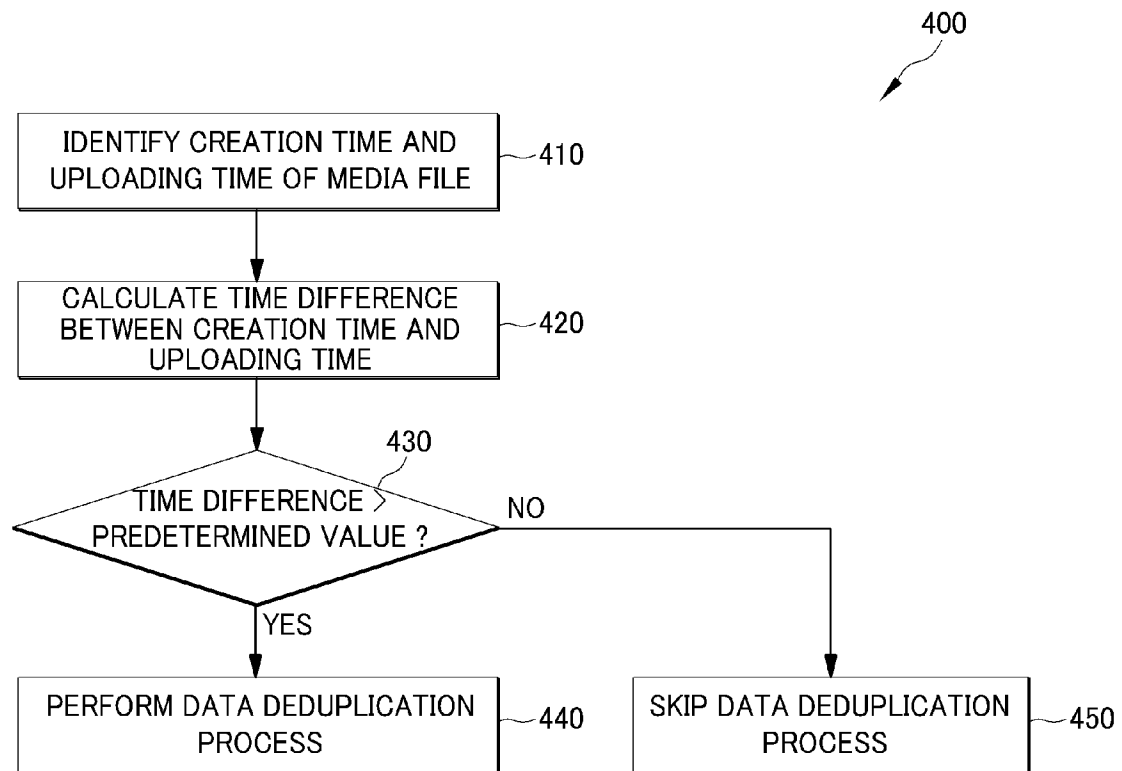


FIG. 5

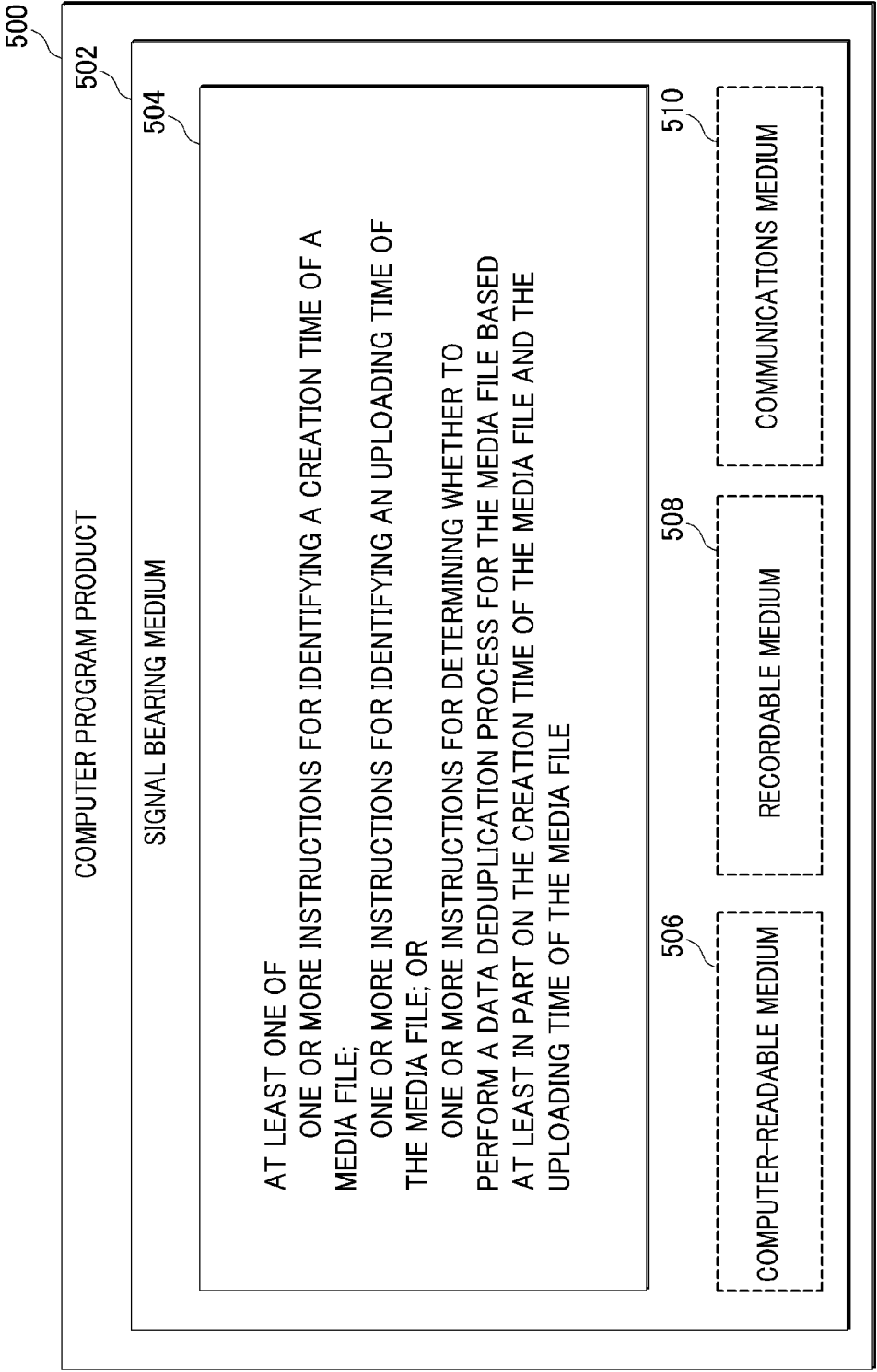


FIG. 6

