(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 July 2008 (03.07.2008)

PCT

(10) International Publication Number
WO 2008/079850 A2

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

(21) International Application Number:
PCT/US2007/088067

(22) International Filing Date:
19 December 2007 (19.12.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/615,771 22 December 2006 (22.12.2006) US

(71) Applicant *(for all designated States except US)*: **GOOGLE INC.** [US/US]; 1600 Amphitheatre Parkway, Building 41, Mountain View, CA 94043 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **DATAR, Mayur** [IN/US]; C/o Google Inc., 1600 Amphiteather Parkway, Building 41, Mountain View, CA 94043 (US). **GARG, Ashutosh** [IN/US]; C/o Google Inc., 1600 Amphitheatre Parkway, Building 41, Mountain View, CA 94043 (US). **MITTAL, Vibhu** [US/US]; C/o Google Inc., 1600 Amphitheatre Parkway, Building 41, Mountain View, CA 94043 (US).

(74) Agents: **SACHS, Robert, R.** et al.; Fenwick & West LLP, Silicon Valley Center, 801 California Street, Mountain View, CA 94041 (US).

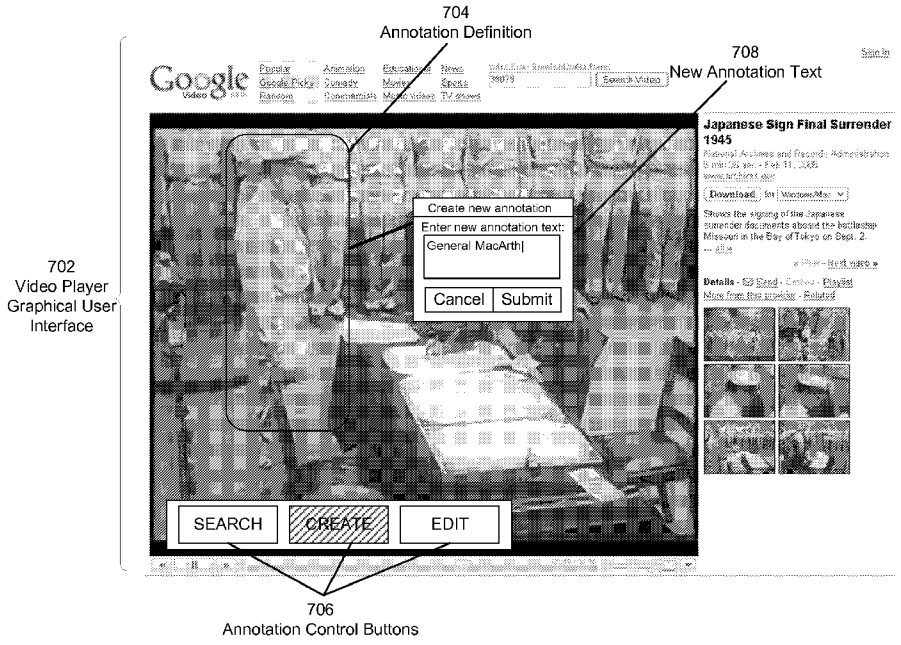
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, 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*

(54) **Title:** ANNOTATION FRAMEWORK FOR VIDEO



(57) **Abstract:** A system and method for transferring annotations associated with a media file. An annotation associated with a media file is indexed to a first instance of that media file. By comparing features of the two instances, a mapping is created between the first instance of the media file and a second instance of the media file. The annotation can be indexed to the second instance using the mapping between the first and second instances. The annotation can be processed (displayed, stored, or modified) based on the index to the second instance.

WO 2008/079850 A2

ANNOTATION FRAMEWORK FOR VIDEO

TECHNICAL FIELD

[0001] The disclosed embodiments relate generally to the authoring and display of annotations for video, and to the collaborative sharing and editing of annotations over a network.

BACKGROUND

[0002] Annotations provide a mechanism for supplementing video with useful information. Annotations can contain, for example, metadata describing the content of the video, subtitles, or additional audio tracks. Annotations can be of various data types, including text, audio, graphics, or other forms. To make their content meaningful, annotations are typically associated with a particular video, or with a particular portion of a video.

[0003] One method by which the useful information contained in annotations can be exchanged is by transferring annotated video over a network. However, transferring video content over a network introduces several obstacles. First, video files are generally quite large, and transferring video requires substantial amounts of bandwidth, as well as host and recipient computers that can support the required bandwidth and storage needs. Second, many video files are likely to be copyrighted, or to be otherwise prohibited from distribution without payment of a fee. Compliance with copyright restrictions requires additional software and hardware investments to prevent unauthorized copying. Third, as the recipient of an annotated video may already have an unannotated copy of the video, from a data efficiency perspective the transfer of an annotated copy of the video to such a recipient unnecessarily consumes both bandwidth and storage.

[0004] Thus, exchanging annotated video by transferring a complete copy of the video is an inadequate solution.

SUMMARY

[0005] Annotations associated with a media file are transferred between devices independently of the associated media file, while maintaining the appropriate temporal or spatial relationship of the annotation with any segment of the media file. An annotation associated with a media file is indexed to a first instance of that media file. A mapping is created between the first instance of the media file and a second instance of the media file by comparing features of the two instances. The annotation can be indexed to the second instance using the mapping between the first and second instances. The annotation can be displayed, stored, or modified based on the index to the second instance.

[0006] Comparing features of instances allows the annotations to be consistently indexed to a plurality of independently acquired instances of a media file. Consistent indexing of annotations supports sharing of annotations and allows for a collaborative community of annotation authors, editors, and consumers. Annotations can include advertisements or premium for-pay content. Privileges for submitting, editing or viewing annotations can be offered for sale on a subscription basis, free of charge, or can be bundled with purchase of media files.

[0007] According to one embodiment, a first user submits to an annotation server annotations that are indexed to his instance of a media file. The annotation server maps the first user's instance of the media file to a canonical instance of the media file and stores the submitted annotation indexed to the canonical instance of the media file. A second user requests annotations, and the annotation server maps the second user's instance of the media file to the canonical instance of the media file. The annotation server sends the annotation to the second user indexed to the second user's instance of the media file.

[0008] The features and advantages described in this summary and the following detailed description are not all-inclusive. Many additional features and advantages will be

apparent to one of ordinary skill in the art in view of the drawings, specification, and claims hereof.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] FIG. 1 shows a network connecting a community of video providers and consumers.

[0010] FIG. 2 illustrates frames of a video, and the indexing of annotations to one or more frames.

[0011] FIG. 3 illustrates frames of two instances of a video.

[0012] FIG. 4(a) illustrates annotations indexed to a canonical instance of video.

[0013] FIG. 4(b) illustrates mapping a client instance of video to a canonical instance of video.

[0014] FIG. 5 illustrates one embodiment for storing video and annotations.

[0015] FIG. 6 is an event trace of the display and modification of annotations associated with a video.

[0016] FIG. 7(a) illustrates a user interface for viewing, creating, and editing annotations.

[0017] FIG. 7(b) illustrates a user interface for creating a new annotation.

[0018] FIG. 8 illustrates a method for determining which annotations to display.

[0019] The figures depict various embodiments of the present invention for purposes of illustration only. One skilled in the art will readily recognize from the following discussion that alternative embodiments of the structures and methods illustrated herein may be employed without departing from the principles of the invention described herein.

DESCRIPTION OF EMBODIMENTS

[0020] FIG. 1 shows a network connecting a community of video providers and consumers. FIG. 1 illustrates one embodiment by which a plurality of users can exchange videos and annotations. Video is used herein as an example of a media file with which annotation can be associated. This example is chosen for the purposes of illustration and is not limiting. Other types of media files with which annotations can be associated include, but are not limited to, audio programs, Flash, movies (in any encoding format), slide presentations, photo collections, animated programs, and other documents. Other examples will be apparent to one of skill in the art without departing from the scope of the present invention.

[0021] A user views, authors, and edits annotations using a client 104. An annotation is any data which can usefully supplement a media file. For example, an annotation can be an audio or textual commentary, translation, advertisement or summary, rating on a predetermined scale (1-5 stars), metadata, or a command for how the media file should be displayed. An annotation can also include video content. The clients 104 include software and hardware for displaying video. For example, a client 104 can be implemented as a television, a personal computer, a digital video recorder (DVR), a personal digital assistant (PDA), a cellular telephone, or another device having or connected to a display device; software includes any video player adapted to decode video files, such as MPEG-2, MPEG-4, QuickTime, VCD, or any other current or future video format. Other examples of clients will be apparent to one of skill in the art without departing from the scope of the present invention. A graphical user interface used by the client 104 according to one embodiment is described herein with references to FIGS. 7(a) and 7(b).

[0022] The clients 104 are connected to a network 105. The network 105 can be implemented as any electronic medium by which annotation content can be transferred.

Through the network 105, the clients 104 can send and receive data from other clients 104. The network 105 can be a global (e.g., the Internet), regional, wide-area, or local area network.

[0023] A video server 106 stores a collection of videos on an electronic medium. Responsive to a request by a client 104 for a particular video (or a set of videos matching certain criteria), the video server 106 transfers a video over the network 105 to the client 104. The video server 106 may be configured to charge a fee for the service of providing the video to the client, or it may provide the video free of charge. The video server 106 can be implemented, for example, as an on-demand content service, an online store, or a streaming video server. Other examples of video servers will be apparent to one of skill in the art without departing from the scope of the present invention.

[0024] Some of the clients 104 are also connected to video sources 102. A video source 102 is a device providing video to the client. For example, a video source 102 could be a cable box, a television antenna, a digital video recorder, a video cassette player, a camera, a game console, a digital video disk (DVD) unit, or any other device capable of producing a video output in a format readable by the client 104. Other examples of video sources 102 will be apparent to one of skill in the art without departing from the scope of the present invention.

[0025] According to one embodiment of the present invention, clients 104 can send video over the network 105. For example, the client 104B can receive video from the video source 102B and transfer it through the network to another client, such as the client 104D. Clients 104 can also send video through the network 105 to the video server 106. Video sent from a client 104 to the video server 106 is stored on an electronic medium and is available to other clients 104.

[0026] Annotation server 110 is connected to the network 105. The annotation server 110 stores annotations on an electronic medium. Responsive to a request from a client 104 for an annotation associated with a particular media file, the annotation server 110 sends one or more annotations associated with the media file to the client 104 through the network 105. Responsive to a submission by the client 104 of one or more annotations associated with a media file, the annotation server 110 stores the one or more annotations in association with the media file. The annotation server 110 stores annotations indexed to instances of one or more media files or portions thereof. A method used by the annotation server 110, according to various embodiments of the present invention, is described herein with reference to FIGS. 4 - 6.

[0027] Optionally, a video server 108 is communicatively connected to the annotation server 110, either locally or over the network 105. The video server 108 can have many of the same capabilities as described herein with reference to the video server 106. The video server 108 can transfer video to the clients 104 over the network 105. In one embodiment, the annotation server 110 and video server 108 in combination transfer annotated video to a client 104. In another embodiment, the video server 108 stores a canonical instance of a video, as described herein with reference to FIG. 5.

[0028] As shown in the figure, any given client may have access to video from a variety of sources. For example, the client 104A can receive video directly from the video source 102A or from the video server 106 via the network 105. Different clients sometimes have access to different video sources. For example, like the client 104A, the client 104B can receive video from the video server 106 via the network 105, but, in contrast to the client 104A, has direct access to the video source 102B instead of the video source 102A.

[0029] Although a client can obtain video from a potentially wide range of video sources, the present invention allows annotations sent from the annotation server 110 to the

client to be consistently associated with a particular media file and portion thereof, regardless of the source from which the client's copy of the video was obtained. The consistent association of annotations with media files facilitates the exchange of annotations between users having different instances (or copies) of a given media file. The present invention enables the sharing and exchange of annotations among a plurality of clients by reindexing annotations for various instances of client media files. For example, the annotation server 110 sends annotations indexed to the client 104A's instance of a video and sends annotations indexed to the client 104B's instance of the video, despite the fact that the two clients may have acquired their copies of the video from different sources. The annotation server 110 beneficially provides annotations that are not only appropriate for the video displayed by the client 104, but for the particular instance of the video which the client 104 is displaying, as described herein with reference to FIG. 4.

[0030] Referring now to FIG. 2, there is shown a conceptual diagram illustrating how annotations are associated temporally and/or spatially with a video file and one or more frames of thereof. FIG. 2 shows a series of video frames, running from frame 200 to frame 251. The client 104 displays these frames, and can also pause, rewind, fast-forward, skip, or otherwise adjust the order or speed with which the frames are displayed.

[0031] For the purposes of illustration, the following discussion refers to a video as being composed of frames. Video is sometimes stored or transmitted as blocks of frames, fields, macroblocks, or in sections of incomplete frames. When reference is made herein to video being composed of frames, it should be understood that during intermediate steps video may in fact be stored as any one of various other forms. The term "frame" is used herein for the sake of clarity, and is not limiting to any particular format or convention for the storage or display of video.

[0032] Some of the frames have annotations associated with them as provided by a particular user. In the example illustrated, frame 201 is drawn in greater to detail to illustrate some of its associated annotations. As shown in the figure, annotations can be associated with a particular spatial location of a frame, or they can be associated with an entire frame. For example, annotation 1 is associated with a rectangular box in the upper-left corner of frame 201. In contrast, annotation 4 is associated with the entire frame.

[0033] Annotations can also be associated with overlapping spatial locations. For example, annotation 1 is associated with a rectangular box overlapping a different rectangular box associated with annotation 2. In one embodiment, annotations can be associated with a spatial location defined by any closed form shape. For example, as shown in FIG. 2, annotation 3 is associated with spatial locations defined by an elliptical shape.

[0034] Annotation list 280 maintains associations between the spatial definition of annotations and the content of annotations. Annotation 1, associated with a rectangular box in frame 201, includes the text "Vice President." Annotation 1 is an example of an annotation useful for highlighting or adding supplemental information to particular portions of a frame. Annotation 4 is associated with the entire frame 201 and contains the text "State of the Union." Annotation 4 is an example of an annotation used to summarize the content of a frame. Annotation 5 is associated with the entire frame 201 and contains some audio, which, in this case, is a French audio translation. Annotation 5 is an example of an annotation used to provide supplemental audio content.

[0035] Annotations can also have temporal associations with a media file or any portion thereof. For example, an annotation can be associated with a specific frame, or a specific range of frames. In FIG. 2, for example, annotation 2 could be associated with frame 200 to frame 251, while annotation 5 is associated only with frame 201. The spatial definition associated with an annotation can also change over time. For example, annotation

1 can be associated with a first region in frame 201, and with a second region in frame 202. Time and spatially-dependent annotation associations are particularly useful for providing supplemental information regarding objects in motion, and can accommodate, as in the example shown in the figure, the movement of the Vice-President of the United States. The temporal associations can be defined in terms of frame numbers, timecodes, or any other indexing basis. The illustration of the annotation list 280 as a table is not meant to limit the underlying storage format used; any format or organization of the annotation information may be employed including optimized formats that reduce storage requirements and/or increase retrieval speed.

[0036] During playback of a media file, the client 104 is adapted to display the annotations associated with the frames of the file. Annotations can be displayed, for example, as text superimposed on the video frame, as graphics shown alongside the frame, or as audio reproduced simultaneously with video; annotations may also appear in a separate window or frame proximate to the video. Annotations can also include commands for how the media file with which they are associated is to be displayed. Displaying command annotations can include displaying video as instructed by the annotation. For example, responsive to an annotation, the client 104 might skip to a different place in a video, display a portion of the video in slow motion, or jump to a different video altogether.

[0037] The client 104 is capable of displaying a subset of the available annotations. For example, a user watching the video of FIG. 2 can select which annotations should be displayed by the client 104 by designation of various criteria. The user can choose to receive only certain types of annotations (e.g. commentary, text, graphic, audio), or only annotations that are defined by a particular region of the display. The user can choose to receive only annotations in a particular language, matching a certain search criteria (such as keywords), or authored by a particular user. As another example, when annotations are written and edited

in a collaborative community of users, a user can choose to receive only annotations authored by users with reputations above a certain threshold, or to receive only annotations with ratings above a certain threshold. Users can also search for annotations, and retrieve associated video based on the results of the annotation search.

[0038] Certain annotations can be given a priority that does not allow a user to prevent them from being displayed. For example, annotations can include advertisements, which may be configured so that no other annotations are displayed unless the advertisement annotations are also displayed. Such a configuration would prevent users from viewing certain annotations while avoiding paid advertisement annotations. A method for determining which annotations to display is described herein with reference to FIG. 8.

[0039] Users can also edit annotations using the client 104. For example, a user viewing the annotations shown in FIG. 2 may be dissatisfied with annotation 1. The user changes the annotation text "Vice President" to "Vice President of the United States" using an input device connected to the client 104. Future display of the annotation (to this user or possibly other users) would include the modified text "Vice President of the United States." As another option, a user can change the temporal or spatial definition with which annotations or associated. For example, the astute user may recognize that the documents shown on the right side of the frame are actually excerpts from 15 USC §§78dd-1, and that the Constitution (despite being almost completely obscured by the position of the President) is just barely visible on the left side of the frame. The user can change the temporal definition with which Annotation 3 is associated accordingly, for example, by dragging (for example, in a direct manipulation user interface illustrating frames of the video) the spatial definition to a different location using an input device connected to the client 104.

[0040] The annotation list 280 is shown in FIG. 2 for the purposes of illustration as one example of how a client can organize annotations and their associated frames. The

annotation list 280 is useful for managing and displaying annotations associated with a frame or range of frames, but various clients can organize annotations differently without departing from the scope of the present invention.

[0041] As shown in FIG. 1, a client sometimes has access to multiple instances of the same video, and different clients frequently have access to various different instances. FIG. 3 illustrates sequences of the frames making up two instances of the same video. For example, video instance 302 could be a copy of a video received from a cable channel, while video instance 304 is a copy of the same video received from an online video store. As another example, video instance 302 could be a copy of a video recorded by a first user's digital video recorder receiving a signal from a first broadcast station, while video instance 304 is a copy of the same video recorded by a second user's digital video recorder receiving a signal from a second broadcast station.

[0042] As video instance 302 is acquired independently of video instance 304, it is likely that the two copies are not time-synchronized, and/or are of different lengths. For example, video instance 302 might have been recorded from The Zurich Channel, a television affiliate known for its punctuality and good taste. Video instance 304, on the other hand, might have been recorded from TV Tulsa, a television affiliate known for its slipshod programming and haphazard timing. Thus, as shown in FIG. 3, the frames of the first instance might not necessarily correspond to the frames of the second instance. In addition, there are numerous other types of differences that can arise between different instances of a given program or broadcast. These include, and are not limited to, differences in encoding parameters (e.g., resolution, frame rate) and differences in file formats.

[0043] In the example illustrated, the frames 306 of video instance 302 are time-shifted with respect to the frames 308 of the video instance 304. The first frame of the frames 308 contains the same content as the third frame of the frames 306. When annotations

are associated with specific frames of a video by one user, it is desirable that they be displayed with those frames when shown to another user in spite of the possibility of time shifting between various instances of the video. Notice as well that video instance 302 has 6 frames, whereas video instance 304 has 4 frames.

[0044] The annotation server 110 accounts for this time shifting of frames so that annotations can be properly displayed with various instances of the video. For example, suppose an annotation describes the driver who enters the third frame of the frames 306. If this annotation is indexed with respect to the frames 306, the annotation server 110 translates this index to an index with respect to the frames 308 so that the annotation can be properly displayed with the video instance 304. The annotation server 110 translates the annotation indexes by mapping one video instance to another.

[0045] Referring now to FIG. 4(a), annotations 404 are indexed to a canonical instance of video 406. For the purposes of illustration, the instance of video having annotations indexed to it is referred to as the canonical instance, and the instance of video that will be displayed at the client is referred to as the client instance. According to one embodiment, annotations can be shared in multiple directions between two or more client peers. As such, it is possible that there is no definitively canonical instance of video. It should be understood that the term "canonical instance" refers to a role that an instance of video plays in one case of annotation exchange, and not necessarily to the status of that copy of the video in the video distribution system or in the annotation framework as a whole.

[0046] The video server 108 may store video content in chunks. One system and method for storing video in chunks is disclosed in U.S. Patent Application Serial No. 11/428,319, titled "Dynamic Media Serving Infrastructure" to Manish Gupta, et al., Attorney Docket No. 24207-11584, filed June 30, 2006, and U.S. Provisional Patent Application Serial No. 60/756,787, titled "Discontinuous Download of Media Articles" to Michael Yu, et al.,

Attorney Docket No. 24207-11081, filed January 6, 2006, both of which are incorporated herein by reference in their entirety. FIG. 4(a) shows a canonical instance of video 406 stored as chunk 402A and chunk 402B. A chunk is a data element for storing video. Storing video in chunks is beneficial for the efficient indexing and transfer of video, and allows for the manipulation as video data of more manageable size.

[0047] As described herein with reference to FIG. 2, an annotation can be associated with a specific frame in a video. The association between the annotation and the specific frame is stored by indexing the annotation to a frame in a particular instance of the video. Annotation 404A, for example, is indexed to a frame of the canonical instance of video 406, in this case to a frame in the chunk 402A.

[0048] As also described herein with reference to FIG. 2, an annotation can be associated with a range of frames in a video. A set of one or more frames of video is sometimes referred to as a segment of video. Annotation 404D, for example, is indexed to a segment of video of the canonical instance of video 406, in this case the segment including one or more frames of the chunk 402B.

[0049] The client receives a video from a video source or server (such as one of those described herein with reference to FIG. 1) and stores a copy as the client instance of video 408. As the client displays the video, the client periodically requests, from the annotation server, annotations associated with frames of video about to be displayed. To ensure that annotations are requested, retrieved, transmitted and received in sufficient time for display with their associated frames, the client requests annotations associated with a frame some time before that frame is to be displayed.

[0050] For increased efficiency, the client can combine requests for annotations associated with particular frames into a request for annotations associated with a segment of video. A request could, for example, seek to retrieve all of the annotations associated with a

given video. In the example shown, the client requests annotations associated with the segment of video 409. The request for annotations will return annotations associated with individual frames of the segment, or annotations associated with a superset or subset of the frames of the segment. For example, the client can request annotations associated with exactly the segment of video 409, associated with individual frames of the segment of video 409, or associated with the entire video.

[0051] Referring now to FIG. 4(b), the annotation server 110 maps the client instance of video 408 to a canonical instance of video 406. The mapping 412 describes the correspondence between frames of the client instance of video 408 and frames in the canonical instance of video 406. The annotation server 110 can map the client instance of the video 408 to the canonical instance of video 406 using a variety of techniques. According to one embodiment of the present invention, the client's request for annotations includes a feature of the client instance of video 408. A feature is a succinct representation of the content of one or more frames of video that are similar. For example, the annotation server 110 may group the frames into logical units, such as scenes or shots. The annotation server 110 may use scene detection algorithms to group the frames automatically. One scene detection algorithm is described in Naphade, M.R., et al., "A High-Performance Shot Boundary Detection Algorithm Using Multiple Cues", 1998 International Conference on Image Processing (Oct 4-7 1998), vol.1, pp. 884-887, which is incorporated by reference herein.

[0052] Thus, the annotation server 110 can compute one feature set for all frames that belong to the same scene. The feature can be, for example, a description of a characteristic in the time, spatial, or frequency domains. For example, a client can request annotations associated with a specific frame, and can describe that frame by its time, position, and frequency domain characteristics. The client can use any technique for determining features

of video, such as those described in Zabih, R., Miller, J., and Mai, K., "Feature-Based Algorithms for Detecting and Classifying Scene Breaks", Proc. ACM Multimedia 95, San Francisco, CA (Nov. 1993), pp. 189-200; Arman, F., Hsu, A., and Chiu, M-Y., "Image Processing on Encoded Video Sequences", Multimedia Systems (1994), vol. 1, no. 5, pp. 211-219; Ford, R.M., et al., "Metrics for Shot Boundary Detection in Digital Video Sequences", Multimedia Systems (2000), vol. 8, pp. 37-46, all of the foregoing being incorporated by reference herein. One of ordinary skill in the art would recognize various techniques for determining features of video.

[0053] Generally, a distance function is defined over the universe of features that captures the closeness of the underlying sets of frames. When the annotation server 110 receives a request for annotation for a frame, along with its feature set, the server first attempts to map the frame in the request to the closest frame in the canonical instance of video 406. The annotation server 110 uses the temporal position of the frame in the client instance of video 408 (one of the features in the feature set) to narrow down the set of frames in the canonical video 406 that may potentially map to this frame, e.g., by limiting the candidate set to frames within a fixed amount of time or frames before and after the selected frame. For all of the frames in the candidate set, the annotation server 110 computes the distance between the feature set of the frame from the client 408 and feature set of the frame from canonical video 406. The frame from the canonical video 406 with the shortest distance is termed as the matching frame. The client frame is then mapped to the matching frame. If the distance to the closest frame is greater than a certain threshold, indicating absence of a good match, no annotations are returned. The components described by a feature used to create the mapping can reside in the segment of video for which annotations are being requested, but need not be. Similarly, the components described by a feature may or may not reside in the segment of video to which an annotation is indexed.

[0054] Features may be represented as strings, allowing the annotation server 110 to search for features using an inverted index from feature strings to frames, for example. The annotation server 110 may also search for features by defining a distance metric over the feature set and selecting the candidate frame with the smallest distance. Such mapping could take place at the time the server 110 receives the client request, or the annotation server 110 can pre-compute and maintain the distances in an offline process.

[0055] Using the mapping 412, the annotation server 110 determines a corresponding segment of video 414 in the canonical instance of video. The corresponding segment of video 414 has content that closely matches the content of the segment of video 409, as described above. Under ideal conditions, the corresponding segment of video 414 contains instances of the same frames as the segment of video 409. The annotation server 110 associates each frame in the client video 408 that maps to a frame in the canonical instance of video with a frame number and maintains a list of frame numbers for each frame mapping. In one example, the length of the list of frame numbers is equal to the number of frames in the client instance of video 408, where each entry maps the corresponding frame to the frame in the canonical instance of video 406.

[0056] The annotation server determines the annotations that are indexed to the corresponding segment of video 414 (or to a superset or subset of the corresponding segment of video 414). As the example of FIG. 4(b) illustrates, the annotation 404D is indexed to a segment of video that falls in the corresponding segment of video 414. In response to the request for annotations for the segment 409, the annotation server 110 transmits the annotation 404D to the client.

[0057] Optionally, the annotation server can also transmit information describing the segment of the video that the annotation is associated with. For example, using a feature as a

reference point, the annotation server can describe a frame (or range of frames) with respect to that reference point.

[0058] FIG. 5 illustrates the organization of video and annotations. FIG. 5 shows how annotations can be indexed to a canonical instance of video in an annotation server.

[0059] According to one embodiment, annotations are stored in an annotation repository. Canonical instances of video are stored in a video repository. The annotation and repositories can be included in the same server, or they can be included in different servers. For example, the annotations can be stored in the annotation server 110 and video can be stored in the video server 108.

[0060] An annotation includes a reference to a segment of video. For example, the annotation 404D includes a temporal definition 501D. A temporal definition specifies one or more frames of a canonical instance of video. In the example illustrated, the temporal definition 501D refers to one of the frames 504 of the canonical instance of video 406. As another example, the annotation 404F includes temporal definition 510F. Temporal definition 510F refers to a range of the frames of the canonical instance of video 406. A temporal definition can described using a variety of metrics including, but not limited to, document identifiers, frame identifiers, timecodes, length in frames, length in milliseconds, and various other combinations.

[0061] The temporal definition is one example of how annotations can be associated with segments of video. Other methods for associating annotations with segments of video will be apparent to one of skill in the art without departing from the scope of the present invention.

[0062] An annotation also includes annotation content 511. Annotation content can include, for example, audio, text, metadata, commands, or any other data useful to be associated with a media file. An annotation can optionally include a spatial definition 509,

which specifies the area of the frame (or frames) with which that annotation is associated. Use of a spatial definition 509 is an example of one method for associating an annotation with a specific spatial location on a frame.

[0063] As an example, suppose the corresponding segment of video 414 includes the frames 504. The corresponding segment of video 414 can be defined as a range of timecodes. The annotation server retrieves annotations by searching for annotations with references to timecodes that are within or overlapping with the range of timecodes defining the corresponding segment of video 414. The annotation server retrieves annotation 404D, including the annotation content 511D. The annotation server transmits the annotation content 511D (or the annotation 404D, which includes the annotation content 511D) to the client, which displays the annotation content 511D.

[0064] FIG. 6 is an event trace of the display and modification of annotations associated with a video, according to one embodiment of the present invention. The client 104 receives a segment of video from a video server 106 or a video source 102, and stores a copy as the client instance of video. The client processes the segment using a feature detection algorithm and determines 602 a feature based on a first segment of video. The client sends a request for annotations associated with a second segment of video, the request including the feature, to the annotation server 110.

[0065] The first segment of video may contain some frames in common with the second segment of video, but need not. The feature included in the request for annotations associated with the second segment of video may additionally include features from adjacent segments to the second segment of video.

[0066] The request can also include metadata describing the content or title of the video so that the annotation server can retrieve the appropriate annotations. For example, video purchased from an online store may have a video title that can be used to filter the set

of available annotations. As another example, the metadata sent to the annotation server for video acquired from broadcast television or cable can include a description of the time and channel at which the video was acquired. The annotation server can use this time and channel information to determine the appropriate video and retrieve annotations associated with that video.

[0067] The annotation server 110 receives the request for annotations. The annotation server 110 searches 604 for the feature included in the request in a canonical instance of the video and creates a mapping between the client instance of the video and the canonical instance of the video. In one embodiment, the request for annotations includes metadata indicating a particular video for which to retrieve annotations, and the annotation server 110 searches 604 in a canonical instance in the video indicated by this metadata for the feature.

[0068] The annotation server 110 searches 608 an annotation repository for annotations associated with the video and returns an annotation. For example, the annotation server 110 can search for annotations indexed to the canonical instance of the video. Using the mapping between the two instances, the annotation server 110 can translate the index to the canonical instance of the video to an index to the client instance of the video

[0069] The annotation server 110 transmits an annotation associated with the video to the client. According to one embodiment, the annotation also includes index information defining the set of one or more frames associated with the annotation. The annotation server 110 can define frames associated with the annotation, for example, by indexing the association with respect to the feature.

[0070] The client 104 receives and displays 610 the annotation. The client 104 can also process index information for the annotation so that the annotation is displayed appropriately along with the client instance of the video.

[0071] Optionally, the client receives 612 changes to the annotation from the user. For example, a user can edit text, re-record audio, modify metadata included in the annotation content, or change an annotation command. The client 104 transmits the modified annotation to the annotation server 110, or, alternatively, transmits a description of the modifications the annotation server 110.

[0072] The annotation server 110 receives the modified annotation. The annotation server 110 stores 614 the modified annotation and indexes the modified annotation to the canonical instance of the video. The annotation server 110 can index the modified annotation with the canonical instance of the video using a variety of methods. For example, the annotation server 110 can translate an index to the client instance of the video using a previously established mapping. As another example, the client 104 can include a feature with the modified annotation, and the annotation server 110 can establish a new mapping between the client instance of the video and the canonical instance of the video.

[0073] For the purposes of illustration, features have been shown as flowing from the client 104 to the annotation server 110. However, for the purpose of establishing a mapping between the client instance of the video and the canonical instance of the video, features can flow in either direction. The example of the annotation server 110 maintaining this mapping on the basis of features sent by the client 104 is given for the purposes of illustration and is not limiting. In another embodiment, the client maintains the mapping between the client instance of the video and the canonical instance of the video, for example, on the basis of features of the canonical instance of the video sent by the annotation server 110 to the client 104. In yet another embodiment, a third party maintains the mapping between the client instance of the video and the canonical instance of the video by receiving features from both the annotation server 110 and the client 104.

[0074] The client 104 can also be used to submit a new annotation. For example, a user can create annotation content and associate it with a video. The user can also specify a spatial definition for the new annotation and choose a range of frames of the client instance of the video to which the annotation will be indexed. The client 104 transmits the new annotation to the annotation server 110 for storage.

[0075] Referring now to FIG. 7(a), a user can search, create, or edit annotations using a graphical user interface. In the example illustrated, the graphical user interface for annotations is integrated into a video player graphical user interface 702. The video player graphical user interface 702 is an example of an interface that might be shown on the display device of a client 104. The video player graphical user interface 702 includes a display area for presenting the media file (in the example illustrated, a video), as well as control buttons for selecting, playing, pausing, fast forwarding, and rewinding the media file. The video player graphical user interface 702 can also include advertisements, such as the advertisement for the National Archives and Records Administration shown in FIG. 7(a).

[0076] The video player graphical user interface 702 presents a frame of video. Shown along with the frame of video is an annotation definition 704. The annotation definition 704 graphically illustrates the spatial definition and/or the temporal definition of an annotation. For example, the annotation definition 704 shown in FIG. 7(a) delineates a subset of the frame with which an annotation is associated. As another example, an annotation definition 704 can delineate a range of frames with which an annotation is associated. While a single annotation definition 704 is shown in FIG. 7(a), the video player graphical user interface 702 can include a plurality of annotation definitions 704 without departing from the scope of the invention.

[0077] The annotation definition 704 can be displayed in response to a user selection, or as part of the display of an existing annotation. For example, the user can use an input

device to select a region of the frame with which a new annotation will be associated, and in response to that selection the video player graphical user interface 702 displays the annotation definition 704 created by the user. As another example, the video player graphical user interface 702 can display video and associated annotations, and can display the annotation definition 704 in conjunction with displaying an associated annotation.

[0078] The video player graphical user interface 702 also includes annotation control buttons 706, which allow the user to control the content and display of annotations. For example, the video player graphical user interface 702 can include a button for searching annotations. In response to the selection of the search annotations button, the client searches for annotations associated with the annotation definition 704 (or a similar definition), or for annotations associated with a keyword. The results of the search can then be displayed on the video player graphical user interface 702. As another example, the video player graphical user interface 702 can include a button for editing annotations. In response to the selection of the edit annotations button, the video player graphical user interface 702 displays one or more annotations associated with the annotation definition 704 and allows the user to modify the one or more annotations. As yet another example, the video player graphical user interface 702 can include a button for creating a new annotation. In response to the selection of the create new annotation button, the video player graphical user interface 702 displays options such as those shown in FIG. 7(b).

[0079] Referring now to FIG. 7(b), the annotation control buttons 706 indicate that the create new annotation button has been selected. The video player graphical user interface 702 includes a display area for receiving user input of the new annotation content. In the example illustrated, the new annotation content includes some new annotation text 708. As shown in FIG. 7(b), as the user enters the description "General MacArthur", the new annotation text 708 is displayed. In response to a further user selection indicating the

authoring of annotation content is complete, the new annotation is submitted, for example, to the annotation server 110, and displayed in the video player graphical user interface 702.

[0080] The entering of new annotation text 708 has been shown as an example of the authoring of annotation content. The video player graphical user interface 702 can be adapted to receive other types of annotation content as well. For example, annotation content can include audio, and the video player graphical user interface 702 can include a button for starting recording of audio through a microphone, or for selecting an audio file from a location on a storage medium. Other types of annotations and similar methods for receiving their submission by a user will be apparent to one of skill in the art without departing from the scope of the invention.

[0081] FIG. 8 illustrates a method for determining which annotations to display. In one embodiment, the client 104 displays only some of the received annotations. The client 104 performs a method such as the one illustrated in FIG. 8 to determine which annotations should be displayed and which should not.

[0082] The client 104 receives 802 an annotation. The client determines 804 if the annotation is high-priority. A high-priority annotation is displayed regardless of user settings for the display of annotations. High-priority annotations can include, for example, advertisements, emergency broadcast messages, or other communications whose importance that should supersede local user settings.

[0083] If the client 104 determines 804 that the annotation is high-priority, the client displays 812 the annotation. If the client 104 determines 804 that the annotation is not high-priority, the client determines 806 if annotations are enabled. Annotations can be enabled or disabled, for example, by a user selection of an annotation display mode. If the user has selected to disable annotations, the client 104 does not display 810 the annotation. If the user

has selected to enable annotations, the client 104 determines 808 if the annotation matches user-defined criteria.

[0084] As described herein, the client 104 allows the user to select annotations for display based on various criteria. In one embodiment, the user-defined criteria can be described in the request for annotation, limiting the annotations sent by the annotation server 110. In another embodiment, the user-defined criteria can be used to limit which annotations to display once annotations have been received at the client 104. User defined-criteria can specify which annotations to display, for example, on the basis of language, annotation content, particular authors or groups of authors, or other annotation properties.

[0085] If the client 104 determines 808 that the annotation satisfies the user-defined criteria, the client 104 displays 812 the annotation. If the client 104 determines 808 that the annotation does not satisfy the user-defined criteria, the client 104 does not display 810 the annotation.

[0086] FIG. 8 illustrates one example of how the client 104 may determine which annotations to display. Other methods for arbitrating annotation priorities established by the annotation provider and the annotation consumer will be apparent to one of skill in the art without departing from the scope of the present invention.

[0087] Turning now to the canonical instance of video disclosed herein, the canonical instance of video can be implemented in a variety of ways according to various embodiments. In some cases, the annotation server 110 has selected a canonical instance of the video prior to the submission of the new annotation. The client 104 can send a feature to facilitate the indexing of the new annotation to the canonical instance of the video. In other cases, for example, when the annotation is the first to be associated with a particular video, the annotation server 110 may not have yet identified a canonical instance of the video. The annotation server 110 stores the annotation indexed to the client instance of the video, and

establishes the client instance of the video as the canonical instance of the video for future annotation transactions.

[0088] According to one embodiment of the present invention, annotations are stored indexed to features of the instance of video used by the client that submitted that annotation. Annotations can be stored and retrieved without any underlying canonical instance of video. For example, each annotation can be indexed to its own "canonical instance of video", which refers to the instance of video of the submitter. Such an approach is particularly beneficial for situations in which the annotation server 110 does not maintain or have access to copies of the video itself. Essentially, the annotation server 110 can serve as a blind broker of annotations, passing annotations from authors to consumers without its own copy of the video with which those annotations are associated.

[0089] A content-blind annotation server can be beneficial, for example, when the video content is copyrighted, private, or otherwise confidential. For example, a proud mother may want to annotate a film of her son's first bath, but might be reticent to submit even a reference instance of the video to a central annotation server. The content-blind annotation server stores annotations indexed to the mother's instance of the video, without access to an instance of its own. When an aunt, uncle, or other trusted user with a instance of the video requests annotations, his instance is mapped to the mother's instance by comparison of features of his instance to features of the mother's instance received with the submission of the annotation. Features can be determined in such a way that cannot be easily reversed to find the content of a frame, thus preserving the privacy of the video.

[0090] The case of an annotation server and a client is but one example in which the present invention may be usefully employed for the sharing and distribution of annotations for video. It will be apparent to one of skill in the art that the methods described herein for transmitting annotations without the need to transmit associated video will have a variety of

other uses without departing from the scope of the present invention. For example, the features described herein could be used in an online community in which users can author, edit, review, publish, and view annotations collaboratively, without the burdens of transferring or hosting video directly. Such a community would allow for open-source style production of annotations without infringing the copyright protections of the video with which those annotations are associated.

[0091] As an added feature, a user in such a community could also accumulate a reputation, for example based on other users' review of the quality of that user's previous authoring or editing. A user who wants to view annotations could have the option of ignoring annotations from users with reputations below a certain threshold, or to search for annotations by users with reputations of an exceedingly high caliber. As another example, a user could select to view annotations only from a specific user, or from a specific group of users.

[0092] As described herein, annotations can also include commands describing how video should be displayed, for example, commands that instruct a display device to skip forward in that video, or to jump to another video entirely. A user could author a string of jump-to command annotations, effectively providing a suggestion for the combination of video segments into a larger piece. As an example, command annotations can be used to create a new movie from component parts of one or more other movies. The annotation server provides the annotations to the client, which acquires the various segments specified by the annotations and assembles the pieces for display to the user.

[0093] The present invention has applicability to any of a variety of hosting models, including but not limited to peer-to-peer, distributed hosting, wiki-style hosting, centralized serving, or other known methods for sharing data over a network.

[0094] The annotation framework described herein presents the opportunity for a plurality of revenue models. As an example, the owner of the annotation server can charge of

fee for including advertisements in annotations. The annotation server can target advertisement annotations to the user based on a variety of factors. For example, the annotation server could select advertisements for transmission to the client based on the title or category of the video that the client is displaying, known facts about the user, recent annotation search requests (such as keyword searches), other annotations previously submitted for the video, the geographic location of the client, or other criteria useful for effectively targeting advertising.

[0095] Access to annotations could be provided on a subscription basis, or annotations could be sold in a package with the video content itself. For example, a user who purchases a video from an online video store might be given permission for viewing, editing, or authoring annotations, either associated with that video or with other videos. An online video store might have a promotion, for example, in which the purchase of a certain number of videos in a month gives the user privileges on an annotation server for that month.

[0096] Alternatively, the purchase of a video from an online video store might be coupled to privileges to author, edit, or view annotations associated with that video. If a particular annotation server becomes particularly popular with users, controlled access to the annotation server could assist with the protection of the copyrights of the video. For example, a user might have to prove that he has a certified legitimately acquired copy of a video before being allowed to view, edit, or author annotations. Such a requirement could reduce the usefulness or desirability of illegally acquired copies of video.

[0097] These examples of revenue models have been given for the purposes of illustration and are not limiting. Other applications and potentially profitable uses will be apparent to one of skill in the art without departing from the scope of the present invention.

[0098] Reference in the specification to "one embodiment" or to "an embodiment" means that a particular feature, structure, or characteristic described in connection with the

embodiments is included in at least one embodiment of the invention. The appearances of the phrase "in one embodiment" in various places in the specification are not necessarily all referring to the same embodiment.

[0099] It should be noted that the process steps and instructions of the present invention can be embodied in software, firmware or hardware, and when embodied in software, can be downloaded to reside on and be operated from different platforms used by a variety of operating systems.

[00100] The present invention also relates to an apparatus for performing the operations herein. This apparatus may be specially constructed for the required purposes, or it may comprise a general-purpose computer selectively activated or reconfigured by a computer program stored in the computer. Such a computer program may be stored in a computer readable storage medium, such as, but is not limited to, any type of disk including floppy disks, optical disks, CD-ROMs, magnetic-optical disks, read-only memories (ROMs), random access memories (RAMs), EPROMs, EEPROMs, magnetic or optical cards, application specific integrated circuits (ASICs), or any type of media suitable for storing electronic instructions, and each coupled to a computer system bus. Furthermore, the computers referred to in the specification may include a single processor or may be architectures employing multiple processor designs for increased computing capability.

[00101] The algorithms and displays presented herein are not inherently related to any particular computer or other apparatus. Various general-purpose systems may also be used with programs in accordance with the teachings herein, or it may prove convenient to construct more specialized apparatus to perform the required method steps. The required structure for a variety of these systems will appear from the description below. In addition, the present invention is not described with reference to any particular programming language. It will be appreciated that a variety of programming languages may be used to implement the

teachings of the present invention as described herein, and any references below to specific languages are provided for disclosure of enablement and best mode of the present invention.

[00102] While the invention has been particularly shown and described with reference to a preferred embodiment and several alternate embodiments, it will be understood by persons skilled in the relevant art that various changes in form and details can be made therein without departing from the spirit and scope of the invention.

[00103] Finally, it should be noted that the language used in the specification has been principally selected for readability and instructional purposes, and may not have been selected to delineate or circumscribe the inventive subject matter. Accordingly, the disclosure of the present invention is intended to be illustrative, but not limiting, of the scope of the invention, which is set forth in the following claims.

CLAIMS

What is claimed is:

1. A method for retrieving annotations, the method comprising:

receiving from a client device a request for annotations associated with a segment of a first instance of a media file, the first instance of the media file being displayed at the client device;

mapping the segment of the first instance of the media file to a corresponding segment of a second instance of the media file, the second instance of the media file stored at a host device remotely located from the client device;

retrieving an annotation associated with the corresponding segment of the second instance of the media file; and

transmitting the annotation to the client device for processing thereon.

2. The method of claim 1, wherein the request for annotations includes a feature of the first instance of the media file.

3. The method of claim 2, wherein mapping the segment of the first instance of the media file to a corresponding segment of a second instance of the media file comprises searching for the feature in the second instance of the media file.

4. The method of claim 1, wherein the media file comprises a video.

5. The method of claim 1, wherein the segment of the first instance of the media file comprises a first frame and wherein the corresponding segment of the second instance of the media file comprises a second frame.

6. The method of claim 1, wherein the annotation comprises a segment of audio.

7. The method of claim 1, wherein the annotation comprises a description of the media file.

8. The method of claim 1, wherein the annotation comprises an advertisement.

9. The method of claim 1, wherein the annotation comprises a command.

10. A method for processing annotations associated with a media file, the method comprising:

determining a feature of a first segment of the media file;

requesting from a server an annotation associated with the media file, the request including the feature of the first segment of the media file;

receiving from the server a response to the request, said response comprising an annotation associated with a second segment of the media file; and

processing the annotation.

11. The method of claim 10, wherein the feature identifies content in the second segment of the media file.

12. The method of claim 10, wherein the media file comprises a video.

13. The method of claim 10, wherein the first segment comprises a first frame and wherein the second segment comprises a second frame.

14. The method of claim 10, wherein the annotation comprises a segment of audio.

15. The method of claim 10, wherein the annotation comprises a description of the media file.

16. The method of claim 10, wherein the annotation comprises an advertisement.

17. The method of claim 10, wherein the annotation comprises a command.

18. A method for storing annotations comprising:

receiving from a first client device a first annotation, wherein the first annotation is associated with a segment of a first instance of a media file, the first instance of the media file being displayed at the first client device;

mapping the segment of the first instance of the media file to a first corresponding segment in a second instance of the media file, the second instance of the media file stored at a host device remotely located from the first client device; and

storing the first annotation, wherein the first annotation is indexed to the first corresponding segment of the second instance of the media file.

19. The method of claim 18, further comprising:

receiving a feature of the first instance of the media file.

20. The method of claim 19, wherein mapping the segment of the first instance of the media file to the first corresponding segment of the second instance of the media file comprises searching for the feature in the second instance of the media file.

21. The method of claim 18, wherein the media file comprises a video.

22. The method of claim 18, wherein the segment of the first instance of the media file comprises a first frame and wherein the corresponding segment of the second instance of the media file comprises a second frame.

23. The method of claim 18, wherein the first annotation comprises a segment of audio.

24. The method of claim 18, wherein the first annotation comprises a description of the media file.

25. The method of claim 18, wherein the first annotation comprises an advertisement.

26. The method of claim 18, wherein the first annotation comprises a command.

27. The method of claim 18, further comprising:

receiving from a second client device a second annotation; and

storing the second annotation, wherein the second annotation is indexed to the first

corresponding segment of the second instance of the media file.

28. A system for indexing annotations comprising:

a feature detector, configured to create a mapping of a first instance of a media file to a second instance of the media file;

an annotation retriever, configured to retrieve an annotation indexed to the first instance of the media file; and

an annotation indexer, configured to index the annotation to the second instance of the media file using the mapping.

29. The system of claim 28, further comprising:

an annotation displayer, configured to display the annotation together with the second instance of the media file.

30. A system for retrieving annotations, the system comprising:

means for receiving from a client device a request for annotations associated with a segment of a first instance of a video, the first instance of the video being displayed at the client device, the request for annotations includes a feature of the first instance of the video;

means for mapping the segment of the first instance of the video to a corresponding segment of a second instance of the video, the second instance of the video stored at a host device remotely located from the client device, said means further comprising means for searching for the feature in the second instance of the video;

means for retrieving an annotation associated with the corresponding segment of the second instance of the video; and

means for transmitting the annotation to the client device for display thereon.

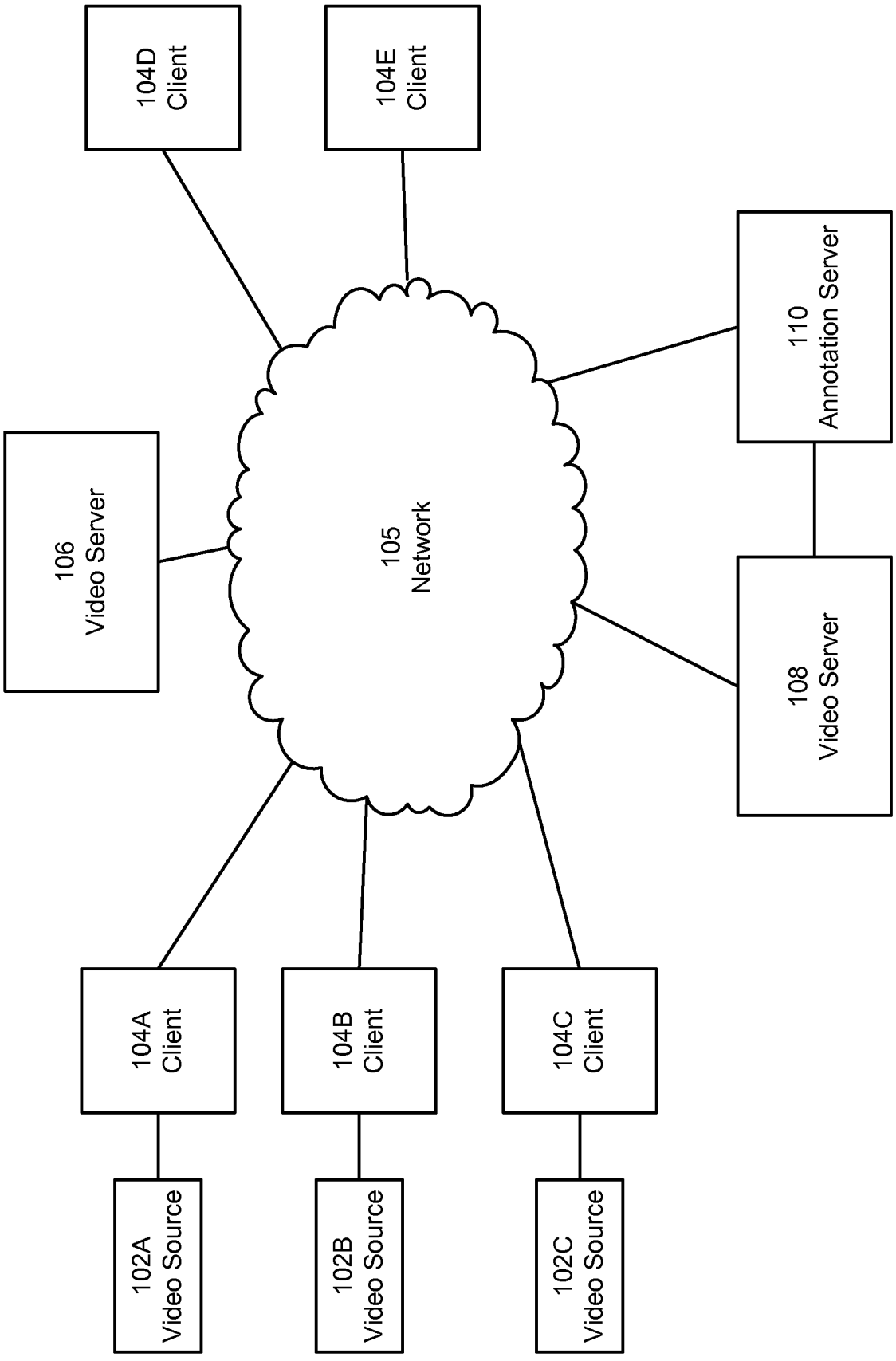

104D Client
104E Client
110 Annotation Server
106 Video Server
105 Network
108 Video Server
104A Client
104B Client
104C Client
102A Video Source
102B Video Source
102C Video Source

FIG. 1

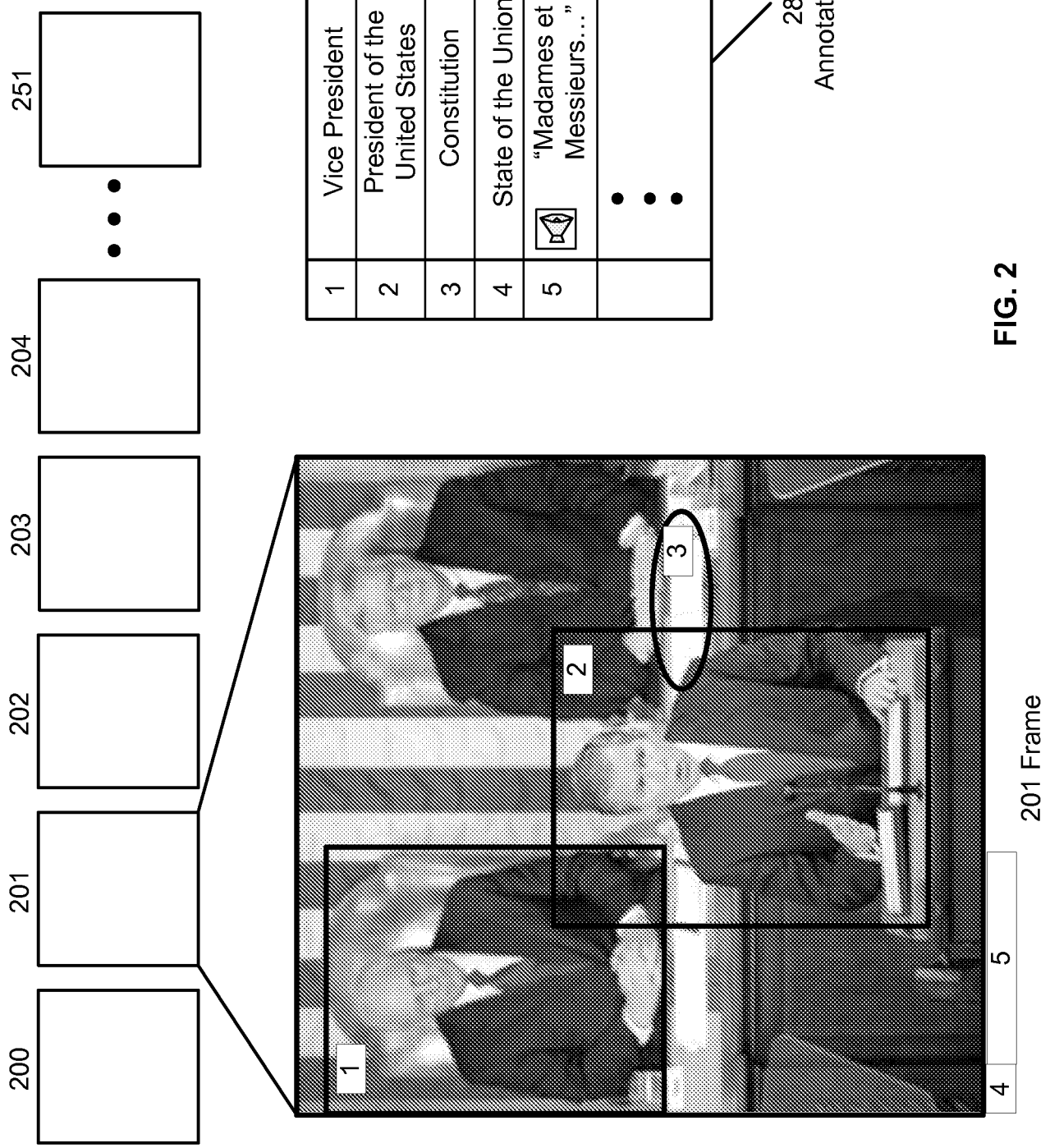

200
201
202
203
204
251
1
2
3
4
5
201 Frame
280
Annotation List
1 Vice President
2 President of the United States
3 Constitution
4 State of the Union
5 "Madames et Messieurs..."

FIG. 2

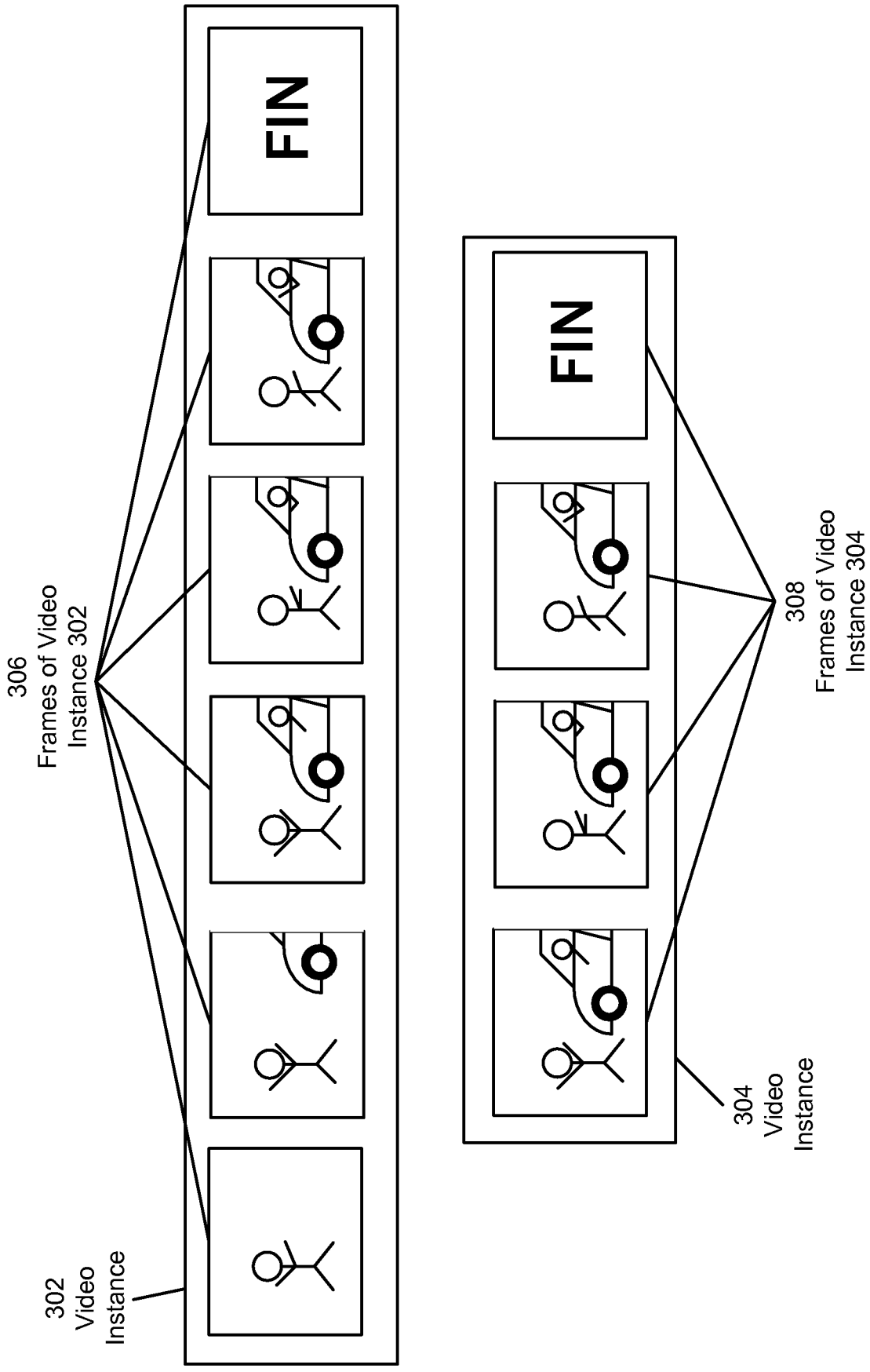

FIN
FIN
306
Frames of Video
Instance 302
308
Frames of Video
Instance 304
302
Video
Instance
304
Video
Instance

FIG. 3

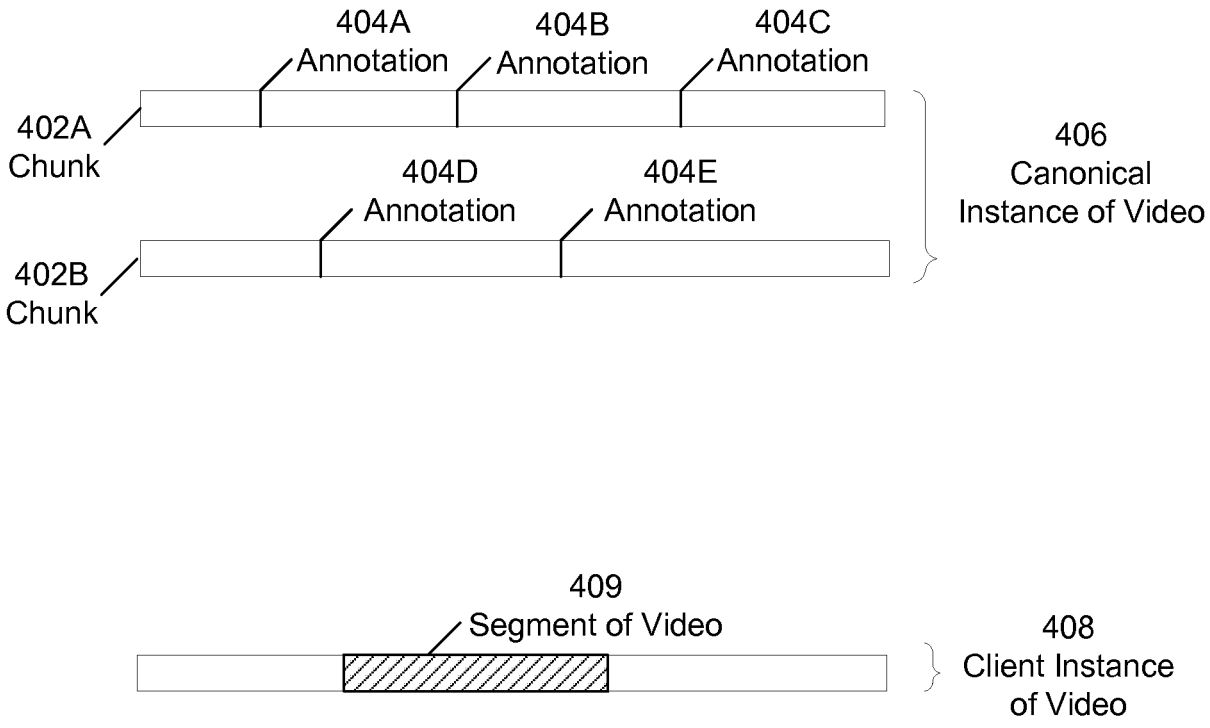

404A
Annotation
404B
Annotation
404C
Annotation
402A
Chunk
404D
Annotation
404E
Annotation
402B
Chunk
406
Canonical
Instance of Video
409
Segment of Video
408
Client Instance
of Video

FIG. 4(a)

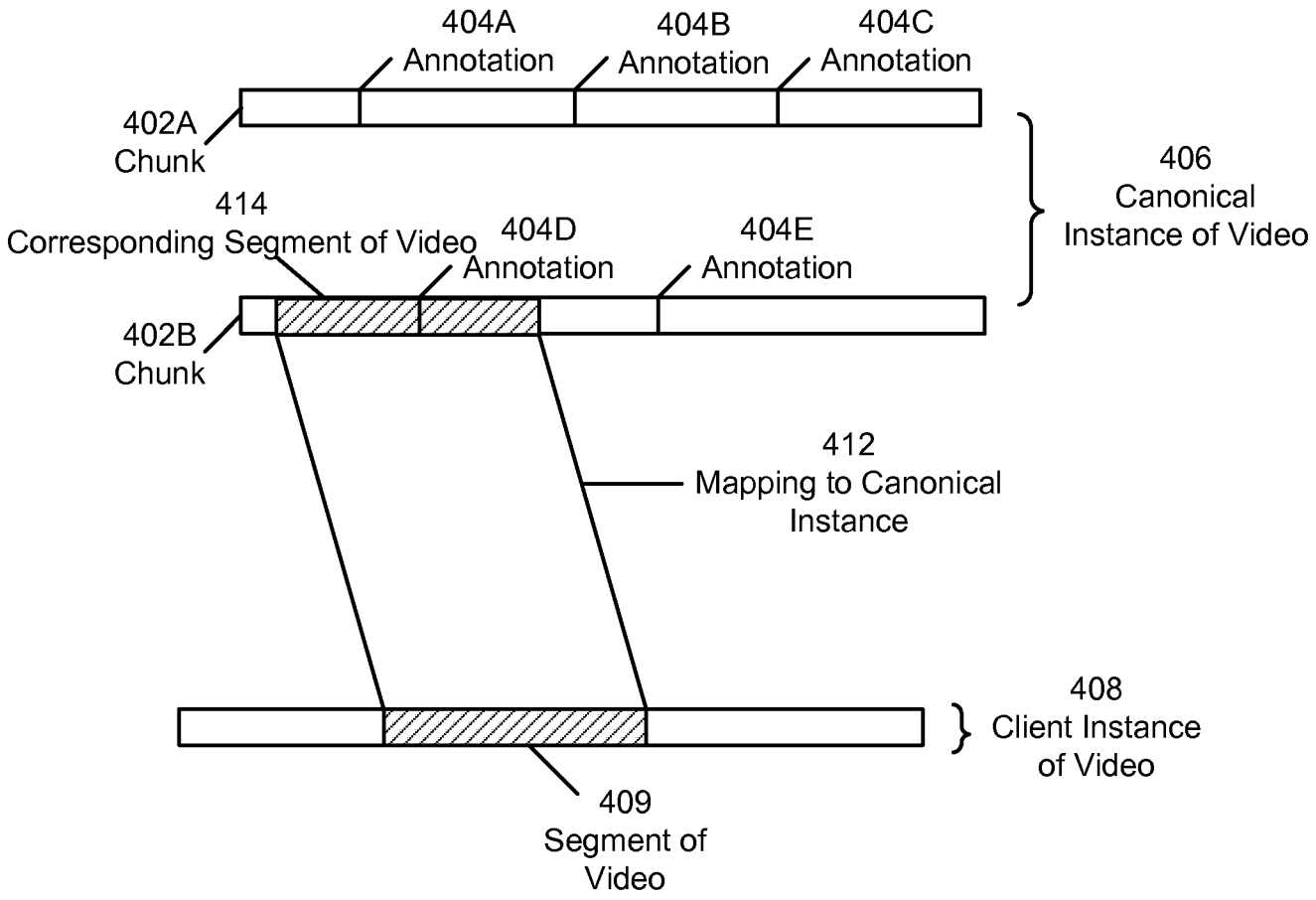

404A
Annotation
404B
Annotation
404C
Annotation
402A
Chunk
406
Canonical
Instance of Video
414
Corresponding Segment of Video
404D
Annotation
404E
Annotation
402B
Chunk
412
Mapping to Canonical
Instance
408
Client Instance
of Video
409
Segment of
Video

FIG. 4(b)

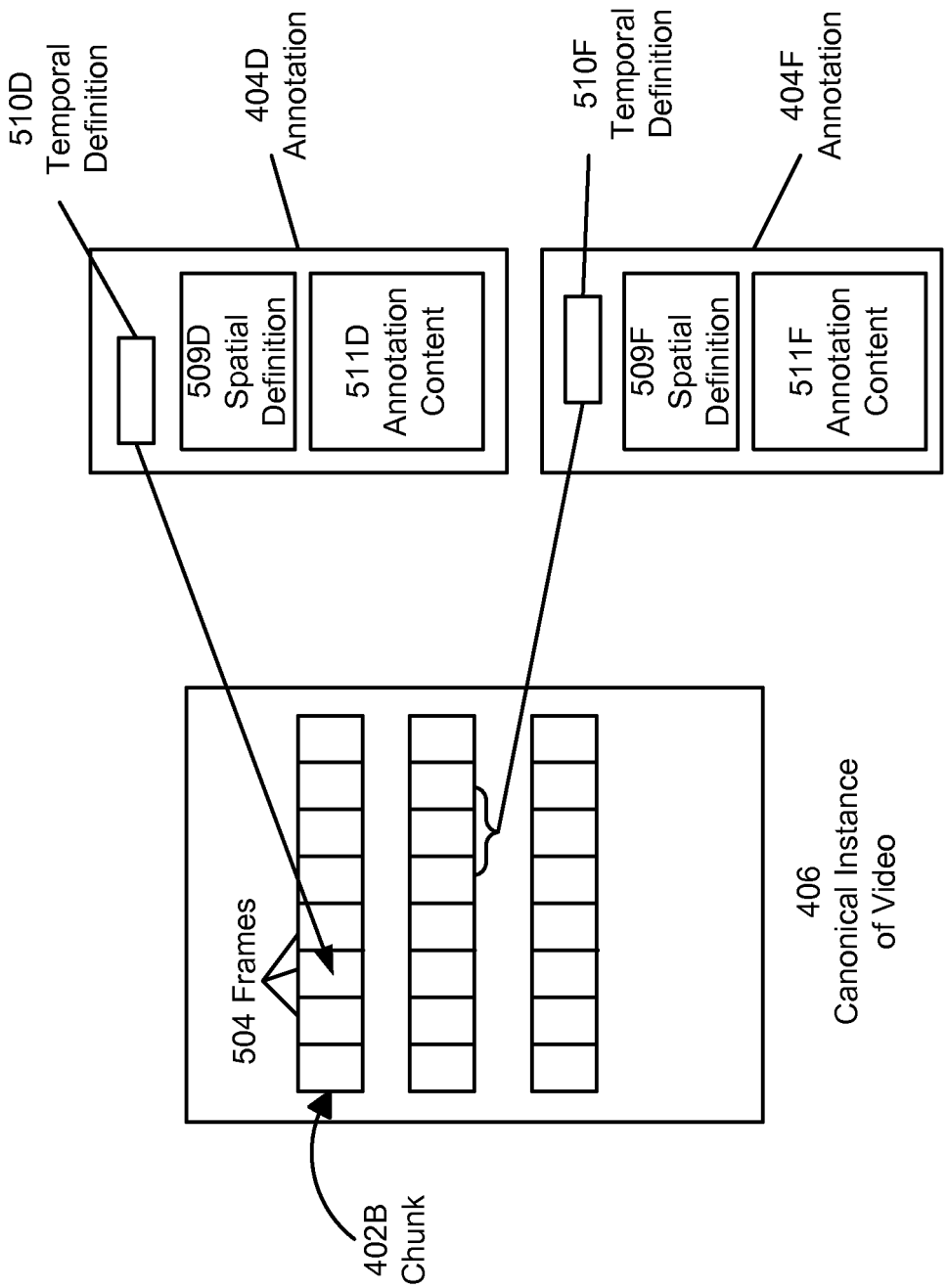

510D Temporal Definition
404D Annotation
509D Spatial Definition
511D Annotation Content
510F Temporal Definition
404F Annotation
509F Spatial Definition
511F Annotation Content
504 Frames
406 Canonical Instance of Video
402B Chunk

FIG. 5

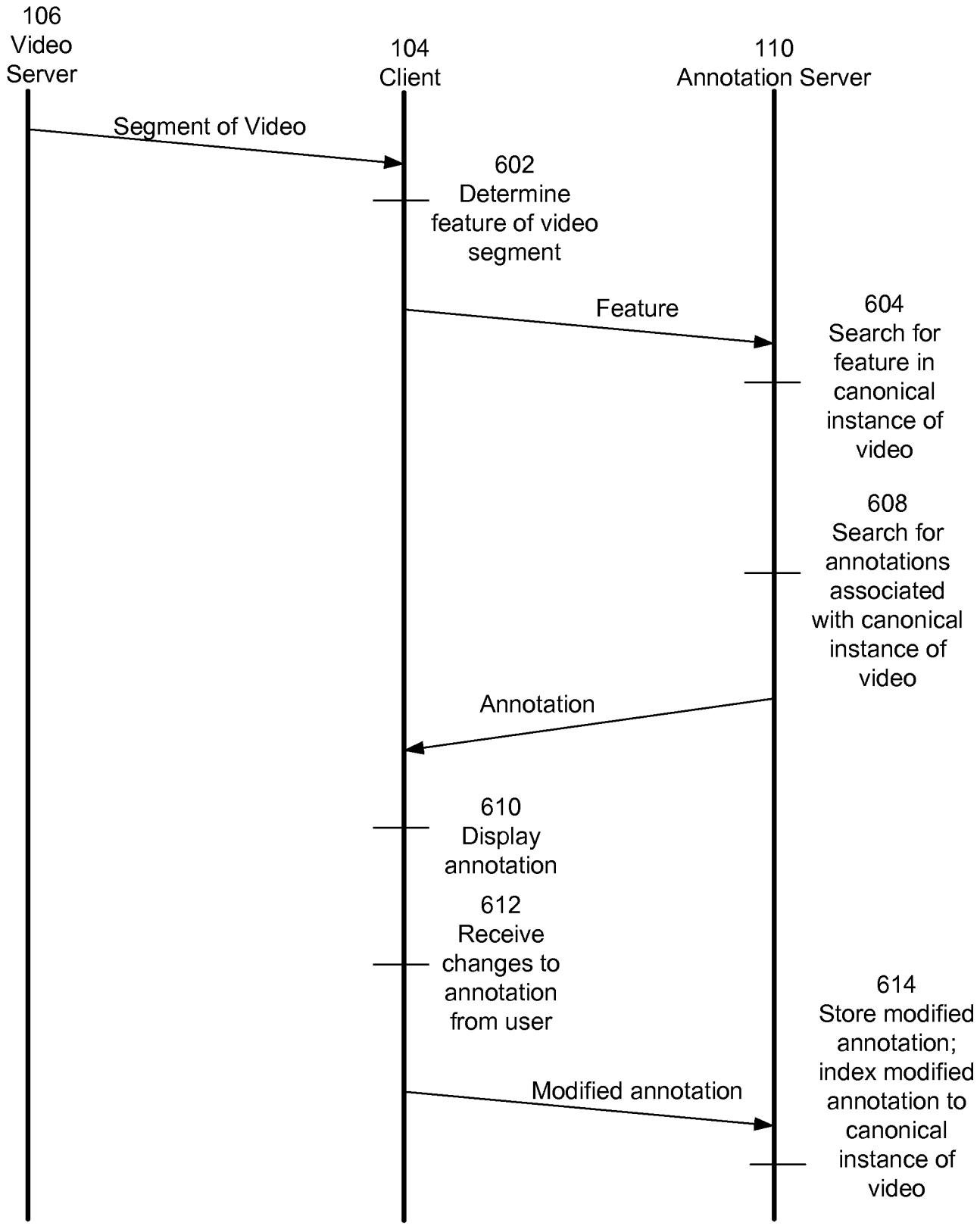

106
Video
Server
104
Client
110
Annotation Server
Segment of Video
602
Determine
feature of video
segment
Feature
604
Search for
feature in
canonical
instance of
video
608
Search for
annotations
associated
with canonical
instance of
video
Annotation
610
Display
annotation
612
Receive
changes to
annotation
from user
614
Store modified
annotation;
index modified
annotation to
canonical
instance of
video
Modified annotation

FIG. 6

704
Annotation Definition
Google Video
Search Video
Japanese Sign Final Surrender 1945
SEARCH
CREATE
EDIT
702
Video Player
Graphical User
Interface
706
Annotation Control Buttons

FIG. 7(a)

704
Annotation Definition
708
New Annotation Text
702
Video Player
Graphical User
Interface
Google
Video
Search Video
Create new annotation
Enter new annotation text:
General MacArth
Cancel
Submit
Japanese Sign Final Surrender
1945
SEARCH
CREATE
EDIT
706
Annotation Control Buttons

FIG. 7(b)

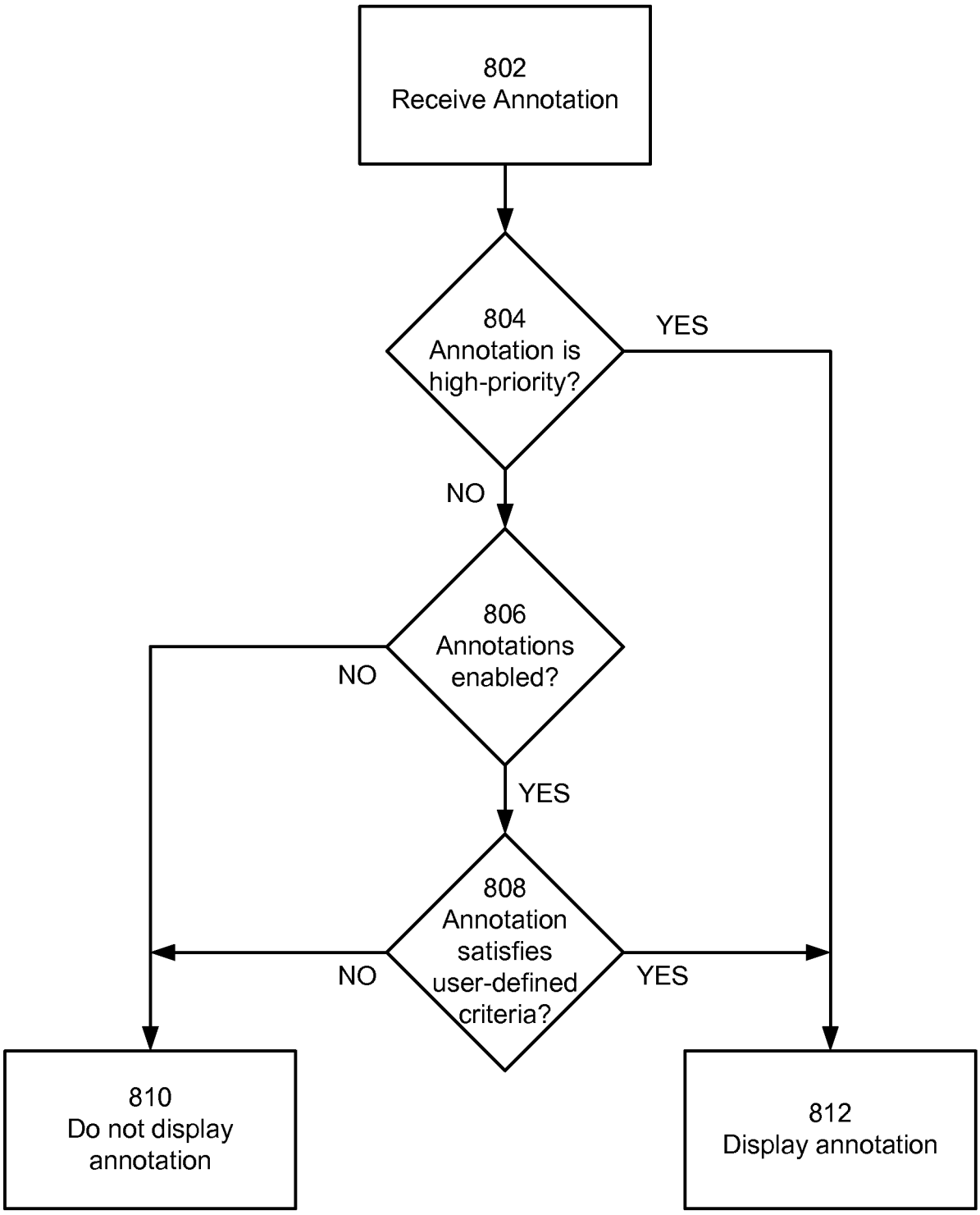

802
Receive Annotation
804
Annotation is high-priority?
YES
NO
806
Annotations enabled?
NO
YES
808
Annotation satisfies user-defined criteria?
NO
YES
810
Do not display annotation
812
Display annotation

FIG. 8