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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 November 2011 (10.11.2011)

PCT

(10) International Publication Number
WO 2011/139448 A2

(51) International Patent Classification:
G06Q 50/00 (2006.01) *H04N 21/236* (2011.01)

(21) International Application Number:
PCT/US2011/031046

(22) International Filing Date:
4 April 2011 (04.04.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
12/767,114 26 April 2010 (26.04.2010) US

(71) Applicant *(for all designated States except US)*: **MICROSOFT CORPORATION** [US/US]; One Microsoft Way, Redmond, Washington 98052-6399 (US).

(72) Inventors: **MEI, Tao**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, Washington 98052-6399 (US). **HUA, Xian-Sheng**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, Washington 98052-6399 (US). **LI, Shipeng**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, Washington 98052-6399 (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, RO, RS, RU, 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, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

[Continued on next page]

(54) Title: ENRICHING ONLINE VIDEOS BY CONTENT DETECTION, SEARCHING, AND INFORMATION AGGREGATION

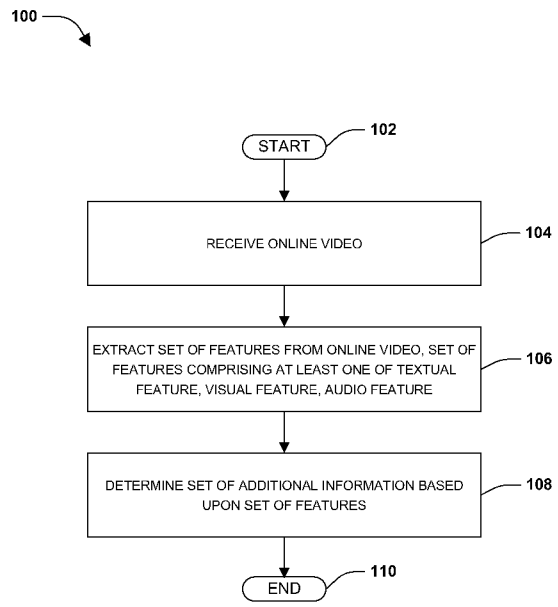

FIG. 1

(57) Abstract: Many internet users consume content through online videos. For example, users may view movies, television shows, music videos, and/or homemade videos. It may be advantageous to provide additional information to users consuming the online videos. Unfortunately, many current techniques may be unable to provide additional information relevant to the online videos from outside sources. Accordingly, one or more systems and/or techniques for determining a set of additional information relevant to an online video are disclosed herein. In particular, visual, textual, audio, and/or other features may be extracted from an online video (e.g., original content of the online video and/or embedded advertisements). Using the extracted features, additional information (e.g., images, advertisements, etc.) may be determined based upon matching the extracted features with content of a database. The additional information may be presented to a user consuming the online video.

WO 2011/139448 A2

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

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

ENRICHING ONLINE VIDEOS BY CONTENT DETECTION, SEARCHING, AND INFORMATION AGGREGATION

BACKGROUND

[0001] When browsing the internet, users interact with a variety of content through online videos. For example, users may view full length movies, advertisements, home videos captured by cell phones, television shows, and other content. Not only may users view online videos through a host website, but may also upload and share online videos with other users. Because online videos play such as integral role in a user internet experience, additional information, such as advertisements and related services (e.g., news, weather, nearby traffic, etc.), may be incorporated with the online video to enhance the user's experience. Unfortunately, current techniques may provide additional information that is irrelevant to the original or embedded content of the online video. Thus, a user may often ignore irrelevant additional information because it may not appeal to the interest of the user.

SUMMARY

[0002] This Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key factors or essential features of the claimed subject matter, nor is it intended to be used to limit the scope of the claimed subject matter.

[0003] Among other things, one or more systems and/or techniques for determining a set of additional information relevant to an online video are disclosed herein. It may be appreciated that a set of additional information may comprise any digital content or combination thereof, such as images, text, audio, video, etc. In one example, a set of additional information may be a video advertisement. In another example, a set of additional information may be a set of images coupled with a URL linking to a webpage.

[0004] In one embodiment, an online video may be received. It may be appreciated that the online video may comprise the original content and/or embedded content, such as overlaid advertisements. A set of features from the online video may be extracted. The set of features may comprise textual features (e.g., embedded text), visual features (e.g., textures, SIFT descriptors, etc.), audio features, and/or other descriptive data of the online video. A set of additional information may be determined based upon the set of features. For example, a multimodal relevance matching algorithm may be executed against a database (e.g., an advertisement database) using the set of features to determine the set of

additional information. In one example, the set of additional information may be presented to a user consuming the online video. In another example, the set of additional information may be saved for later utilization.

[0005] In another embodiment, a parsing component may be configured to parse an online video into one or more segments, a segment comprising a sequence of frames of the online video. For respective segments, the parsing component may extract one or more keyframes representative of a sequence of frames of a segment. Because an entire segment may comprise unnecessary amounts of information, one or more keyframes are extracted to represent a segment. A feature extraction component may be configured to extract a set of feature from the one or more keyframes of respective segments. The set of features may comprise textual, visual, audio, and/or other features relating to the online video. An information extraction component may be configured to execute a multimodal relevance matching algorithm against a database using the set of features to determine the set of additional information. A presentation component may be configured to present the set of additional information. In one example, the set of additional information may be presented within a browser hosting the online video. In another example, the set of additional information may be presented overlaid the online video. In another example, the set of additional information may be presented within a skin of a media player hosting the online video. A variety of different presentation modes are contemplated as falling within the scope of the claimed subject matter.

[0006] To the accomplishment of the foregoing and related ends, the following description and annexed drawings set forth certain illustrative aspects and implementations. These are indicative of but a few of the various ways in which one or more aspects may be employed. Other aspects, advantages, and novel features of the disclosure will become apparent from the following detailed description when considered in conjunction with the annexed drawings.

DESCRIPTION OF THE DRAWINGS

[0007] Fig. 1 is a flow chart illustrating an exemplary method of determining a set of additional information relevant to an online video.

[0008] Fig. 2 is a component block diagram illustrating an exemplary system for determining a set of additional information relevant to an online video.

[0009] Fig. 3 is a component block diagram illustrating an exemplary system for determining a set of additional information relevant to an online video.

[0010] Fig. 4 is an illustration of an example of detecting visual features of an online video hosted within a web browsing environment.

[0011] Fig. 5 is an illustration of an example of detecting visual features and/or textual features of an online video.

[0012] Fig. 6 is an illustration of an example of an information extraction component determining a set of additional information based upon a set of features extracted from an online video.

[0013] Fig. 7 is an illustration of an example of presenting additional information to a user consuming an online video.

[0014] Fig. 8 is an illustration of an example of presenting additional information to a user consuming an online video.

[0015] Fig. 9 is an illustration of an example of presenting additional information to a user consuming an online video.

[0016] Fig. 10 is an illustration of an exemplary computer-readable medium wherein processor-executable instructions configured to embody one or more of the provisions set forth herein may be comprised.

[0017] Fig. 11 illustrates an exemplary computing environment wherein one or more of the provisions set forth herein may be implemented.

DETAILED DESCRIPTION

[0018] The claimed subject matter is now described with reference to the drawings, wherein like reference numerals are used to refer to like elements throughout. In the following description, for purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the claimed subject matter. It may be evident, however, that the claimed subject matter may be practiced without these specific details. In other instances, structures and devices are illustrated in block diagram form in order to facilitate describing the claimed subject matter.

[0019] Today, a substantial portion of internet content is in the form online videos. For example, television programs may be streamed as online videos for users to consume. It may be advantageous to provide additional relevant information along with the online videos. For example, if a user is watching an online video comprising a promotional car video, then it may be useful to determine that the content of the online video relates to a car. Using the detected car feature, additional information (e.g., URL to car websites) relating to cars may be provided to the user to enhance their experience.

[0020] Unfortunately, current techniques may be unable to detect features of the online video and/or features of embedded content within the online video. Furthermore, current techniques may not use detected features of embedded content to search outside sources for additional information relevant to the embedded content. That is, current techniques may not search outside databases, such as an advertisement database, for additional information generated by sources other than the content producer of the online video and/or embedded content within the online video.

[0021] Accordingly, one or more systems and/or techniques for determining a set of additional information relevant to an online video are provided herein. In particular, features of an online video may be extracted from original content and/or embedded content (e.g., advertisements) of the online video. The features may be used as search criteria against a database, such as an advertisement database, to determine additional information relevant to the online video. It may be appreciated that the user may have an interest in the content of the online video because the user may choose to spend their browsing time watching the online video. In this way, a user's experience may be enhanced by providing the user with additional information relevant to the online video and/or the user's interest.

[0022] One embodiment of determining a set of additional information relevant to an online video is illustrated by an exemplary method 100 in Fig. 1. At 102, the method begins. At 104, an online video may be received. The online video may comprise original content and/or additional embedded content, such as overlaid advertisements. It may be appreciated that receiving the online video may be interoperated as gaining access to the content of the online video (e.g., accessing a real-time data stream of the online video while a user consumes the online video using a web browser or media player, accessing a stored copy of the online video, requesting the online video from a source hosting the online video, accessing the online video through a browser displaying the online video, etc.).

[0023] At 106, a set of features may be extracted from the online video. The set of features may comprise at least one of a textual feature, a visual feature, and/or an audio feature. In one example, a textual feature (e.g., superimposed text) may be extracted using one or more textual feature extraction techniques, such as an OCR text recognition technique. In another example, a visual feature (e.g., a color histogram, textures, SIFT descriptors) may be extracted by one or more visual feature extraction techniques, such as a scale-invariant feature transformation. It may be appreciated that in one example, more

than one feature may be extracted as the set of features (e.g., two textual features as a set of features, one textual feature and two audio features as a set of features, one hundred visual textual and thirty textual features, etc.).

[0024] At 108, a set of additional information may be determined based upon the set of features. In one example, a multimodal relevance matching algorithm may be executed against a database using the set of features. In particular, the multimodal relevance matching algorithm may match features within the set of features with features of additional information, such as advertisements, images, videos, etc., within the database. For example, the set of features may comprise a visual feature of a car and a textual feature of the word "tires." The multimodal relevance matching algorithm may attempt to match the visual car feature and/or the textual tire feature with images, videos, web page URLs, and/or other content relating to cars and/or tires. The set of additional information relevant to the online video may be determined based upon the matched content. The set of additional information may be presented on a display. For example, the set of additional information may be presented as a video overlaid on or embedded within the online video. At 110, the method ends.

[0025] One embodiment of determining a set of additional information relevant to an online video is illustrated by an exemplary method 200 in Fig. 2. At 202, the method begins. At 204, an online video may be received. At 206, the online video may be parsed into one or more segments, a segment comprising a sequence of frames. The online video may be segmented to parse down the online video into manageable segments from which features may be identified and extracted. To aid in segmentation, boundaries may be determined based upon large visual differences between two sequential frames. Furthermore, segments may be designated as promotional content (e.g., embedded promotional content) or program content (e.g., original content of the online video), for example.

[0026] To further narrow the amount of information utilized from the online video, one or more keyframes may be extracted for respective segments, at 208. A keyframe may be a frame within a segment that is representative of a sequence of frames of the segment. For example, a middle frame or a frame having a video quality above a predetermined threshold may be selected as the keyframe. Selecting keyframes reduces the amount of unnecessary and/or redundant data utilized from the online video.

[0027] At 210, a set of features may be extracted from one or more keyframes of respective segments. In one example, a set of features may comprise one or more textual

features and/or one or more visual features. At 212, a text based search algorithm may be performed upon a database using the one or more textual features to determine a first list of candidate additional information. For example, textual features may be used as keywords to search the database for content comprising matching keywords. At 214, a visual based search algorithm may be performed upon the database using one or more visual features to determine a second list of candidate additional information. For example, the visual features may be matched against content within the database comprising matching visual features.

[0028] At 216, a linear combination may be performed on the first list and second list to generate a candidate set of additional information. For example, the candidate set of additional information may comprise images, URLs, videos, advertisements, and/or other content matching features of the online video. It may be appreciated that the candidate set of additional information or portion thereof may be used as a set of additional information. In one example, the candidate set of additional information may be used as the set of additional information. In another example, a predetermined number of candidates may be selected from within the candidate set of additional information to generate the set of additional information. A predetermined number of candidate content may be selected because the candidate set of additional information may comprise an enormous amount of content that may be used as additional information. That is, the candidate set of additional information may comprise a combination of advertisement videos spanning a time frame longer than the online video. In one example, one or more candidates having high relevancy to the online video may be selected. The set of additional information may be aggregated into a video, which may be presented to a user engaged with the online video. At 218, the method ends.

[0029] Fig. 3 illustrates an example of a system 300 configured for determining a set of additional information 314 relevant to an online video 302. The system 300 may comprise a parsing component 304, a feature extraction component 308, an information extraction component 312, and/or a presentation component 316. The system 300 may utilize a database 320. The database 320 may comprise digital content (potential candidates for additional information) in a format such as text, URLs, images, videos, audio, and/or other digital formats (e.g., advertisement videos, promotional images or text, URLs linking to web pages, coupon data, customer opinions, non-promotional content, etc.).

[0030] The parsing component 304 may be configured to parse the online video 302 into one or more segments, a segment comprising a sequence of frames of the online video

302. The parsing component 304 may be configured to extract keyframes 306 from the one or more segments. In particular, the parsing component 304 may extract one or more keyframes representative of a sequence of frames of a segment.

[0031] The feature extraction component 308 may be configured to extract a set of features 310 from the keyframes 306. The set of features 310 may comprise at least one of a textual feature, a visual feature, and/or an audio feature. In one example, the feature extraction component 308 may comprise a text recognition component, a visual recognition component, and/or an audio recognition component. The text recognition component may be configured to extract one or more textual features using an OCR text recognition technique and/or other textual recognition techniques. The visual recognition component may be configured to extract one or more visual features using a scale-invariant feature transformation and/or other visual recognition techniques. The audio recognition component may be configured to extract one or more audio features using one or more audio recognition techniques. A variety of different visual, textual, and audio recognitions techniques are contemplated as falling within the scope of the claimed subject matter.

[0032] The information extraction component 312 may be configured to execute a multimodal relevance matching algorithm against the database 320 using the set of features 310 to determine the set of additional information 314. In one example, the multimodal relevance matching algorithm may be configured to perform a text based search algorithm upon the database 320 using textual features to determine a first list of candidate additional information. The multimodal relevance matching algorithm may be configured to perform a visual based search algorithm upon the database 320 using visual features to determine a second list of candidate additional information. The multimodal relevance matching algorithm may perform a linear combination of the first list and second list to generate the set of additional information 314. The information extraction component 312 may be configured to aggregate the set of additional information 314 into a video.

[0033] In one example, the presentation component 316 may be configured to present the set of additional information 314. In another example, the presentation component 316 may be configured to present the video to a user consuming the online video 302 within a web browsing environment 318.

[0034] Fig. 4 illustrates an example 400 of detecting visual features of an online video 404 hosted within a web browsing environment. It may be appreciated that the online video

404 may be hosted within a web page displayed by a web browser, a media player plug-in of a web browser, a media player separate from a web browser, and/or other online video playing devices. Visual features may be detected within the online video 404 and/or extracted as part of a set of features. For example, clouds 412, sun 410, trees 402, car 406, tires 408, and/or other visual features may be detected and/or extracted. Additional information relating to the extracted visual features may be determined and/or presented to a user consuming the online video 404. In one example, additional information regarding vacations may be presented based upon the sun 410, the clouds 412, and/or trees 402. In another example, additional information regarding car reviews may be presented based upon the car 406 and/or the tires 408. In this way, the user's experience may be enhanced because additional relevant information, which the user may have an interest in, may be presented to the user.

[0035] Fig. 5 illustrates an example 500 of detecting visual features and/or textual features of an online video 504. Visual features may be detected within the online video 504 and/or extracted as part of a set of features. For example, car 508, tires 510, and/or road 502 may be detected and/or extracted. Textual features may be detected within the online video 504 and/or extracted as part of the set of features. For example, the embedded text 506 "Test drive a new car today by visiting car maker website…" or a portion thereof may be detected and/or extracted. It may be appreciated that letters, single words, phrases, sentences, and/or other logical groupings of words may be detected as a textual feature. That is, the embedded text 506 may be parsed into one or more features (e.g., test drive, Car Maker, etc.). Using the visual features and/or textual features, additional information may be presented to a user consuming the online video 504. For example, additional information of a URL to a website hosted by the Car Maker may be presented based upon the car 508 visual feature and/or the Car Maker textual feature.

[0036] Fig. 6 illustrates an example 600 of an information extraction component 602 determining a set of additional information 608 based upon a set of features 604 extracted from an online video. In one example, the information extraction component 602 may have extracted visual features (car, tree, sun, clouds, road, etc.), textual features (car, drive, Car Maker, website, etc.), audio features (car engine noise, spoken words, etc.), and/or other features from the online video. The information extraction component 602 may determine the set of additional information 608 from within a database 606 using the set of features 604. It may be appreciated that the database 606 may comprise a vast array

of content (e.g., images, URLs, advertisements, videos, audio, etc.) that may be determined as additional information.

[0037] It may be appreciated that some content within the database 606 (e.g., the set of additional information 608) may match one or more of the features within the set of features 604, while other content within the database 606 (e.g., irrelevant content 610) may not match features within the set of features 604. In this way, additional information relevant to the online video (e.g., content within the database 606 matching features within the set of features 604) may be determined within the database 606. For example, the set of additional information 608 may comprise a car image, a car maker website URL, an advertisement for vacation, a video of a car race, and/or other content matching the set of features 604. Irrelevant content 610 may be excluded from the set of additional information 608 because it does not relate to the set of features 604. For example, a textual description about how to install gutters may not match visual features of a car, tree, sun, clouds, road, etc., and/or textual features of a car, drive, Car Maker, website, etc.

[0038] Fig. 7 illustrates an example 700 of presenting additional information 704 to a user consuming an online video 702. In one example, a set of features relating to cars may have been extracted from the online video 702. Using the set of features relating to cars, the additional information 704 of a car race video may have been determined. In one example, the additional information 704 may be presented to a user, for example, as a video within a web browsing environment. In another example, the additional information 704 may be presented as a video within a media player separate from the web browsing environment. A variety of different presentation modes are contemplated as falling within the scope of the claimed subject matter.

[0039] Fig. 8 illustrates an example 800 of presenting additional information 802 to a user consuming an online video. In one example, a set of features relating to cars may have been extracted from the online video. Using the set of features relating to cars, the additional information 802 of a car race video may have been determined. In one example, the additional information 802 of the car race video may be presented overlaid the online video.

[0040] Fig. 9 illustrates an example 900 of presenting additional information 904 to a user consuming an online video 902. In one example, a set of features relating to vacationing may have been extracted from the online video 902. Using the set of features relating to vacationing, the additional information 904 of an advertisement for a vacation (e.g., an image of a vacation coupled with a URL to a website offering deals on vacations) may

have been determined. In one example, the additional information 904 of the advertisement for a vacation may be presented to a user consuming the online video 902. In this way, the user's experience is enhanced with rich additional information, which may be gathered from sources outside of the online video 902 host, for example.

[0041] Still another embodiment involves a computer-readable medium comprising processor-executable instructions configured to implement one or more of the techniques presented herein. An exemplary computer-readable medium that may be devised in these ways is illustrated in Fig. 10, wherein the implementation 1000 comprises a computer-readable medium 1016 (e.g., a CD-R, DVD-R, or a platter of a hard disk drive), on which is encoded computer-readable data 1014. This computer-readable data 1014 in turn comprises a set of computer instructions 1012 configured to operate according to one or more of the principles set forth herein. In one such embodiment 1000, the processor-executable computer instructions 1012 may be configured to perform a method 1010, such as the exemplary method 100 of Fig. 1 and exemplary method 200 of Fig. 2, for example. In another such embodiment, the processor-executable instructions 1012 may be configured to implement a system, such as the exemplary system 300 of Fig. 3, for example. Many such computer-readable media may be devised by those of ordinary skill in the art that are configured to operate in accordance with the techniques presented herein.

[0042] Although the subject matter has been described in language specific to structural features and/or methodological acts, it is to be understood that the subject matter defined in the appended claims is not necessarily limited to the specific features or acts described above. Rather, the specific features and acts described above are disclosed as example forms of implementing the claims.

[0043] As used in this application, the terms "component," "module," "system," "interface," and the like are generally intended to refer to a computer-related entity, either hardware, a combination of hardware and software, software, or software in execution. For example, a component may be, but is not limited to being, a process running on a processor, a processor, an object, an executable, a thread of execution, a program, and/or a computer. By way of illustration, both an application running on a controller and the controller can be a component. One or more components may reside within a process and/or thread of execution and a component may be localized on one computer and/or distributed between two or more computers.

[0044] Furthermore, the claimed subject matter may be implemented as a method, apparatus, or article of manufacture using standard programming and/or engineering techniques to produce software, firmware, hardware, or any combination thereof to control a computer to implement the disclosed subject matter. The term "article of manufacture" as used herein is intended to encompass a computer program accessible from any computer-readable device, carrier, or media. Of course, those skilled in the art will recognize many modifications may be made to this configuration without departing from the scope or spirit of the claimed subject matter.

[0045] Fig. 11 and the following discussion provide a brief, general description of a suitable computing environment to implement embodiments of one or more of the provisions set forth herein. The operating environment of Fig. 11 is only one example of a suitable operating environment and is not intended to suggest any limitation as to the scope of use or functionality of the operating environment. Example computing devices include, but are not limited to, personal computers, server computers, hand-held or laptop devices, mobile devices (such as mobile phones, Personal Digital Assistants (PDAs), media players, and the like), multiprocessor systems, consumer electronics, mini computers, mainframe computers, distributed computing environments that include any of the above systems or devices, and the like.

[0046] Although not required, embodiments are described in the general context of "computer readable instructions" being executed by one or more computing devices. Computer readable instructions may be distributed via computer readable media (discussed below). Computer readable instructions may be implemented as program modules, such as functions, objects, Application Programming Interfaces (APIs), data structures, and the like, that perform particular tasks or implement particular abstract data types. Typically, the functionality of the computer readable instructions may be combined or distributed as desired in various environments.

[0047] Fig. 11 illustrates an example of a system 1110 comprising a computing device 1112 configured to implement one or more embodiments provided herein. In one configuration, computing device 1112 includes at least one processing unit 1116 and memory 1118. Depending on the exact configuration and type of computing device, memory 1118 may be volatile (such as RAM, for example), non-volatile (such as ROM, flash memory, etc., for example) or some combination of the two. This configuration is illustrated in Fig. 11 by dashed line 1114.

[0048] In other embodiments, device 1112 may include additional features and/or functionality. For example, device 1112 may also include additional storage (e.g., removable and/or non-removable) including, but not limited to, magnetic storage, optical storage, and the like. Such additional storage is illustrated in Fig. 11 by storage 1120. In one embodiment, computer readable instructions to implement one or more embodiments provided herein may be in storage 1120. Storage 1120 may also store other computer readable instructions to implement an operating system, an application program, and the like. Computer readable instructions may be loaded in memory 1118 for execution by processing unit 1116, for example.

[0049] The term "computer readable media" as used herein includes computer storage media. Computer storage media includes volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information such as computer readable instructions or other data. Memory 1118 and storage 1120 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 (DVDs) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store the desired information and which can be accessed by device 1112. Any such computer storage media may be part of device 1112.

[0050] Device 1112 may also include communication connection(s) 1126 that allows device 1112 to communicate with other devices. Communication connection(s) 1126 may include, but is not limited to, a modem, a Network Interface Card (NIC), an integrated network interface, a radio frequency transmitter/receiver, an infrared port, a USB connection, or other interfaces for connecting computing device 1112 to other computing devices. Communication connection(s) 1126 may include a wired connection or a wireless connection. Communication connection(s) 1126 may transmit and/or receive communication media.

[0051] The term "computer readable media" may include communication media. Communication media typically embodies computer readable instructions or other data in a "modulated data signal" such as a carrier wave or other transport mechanism and includes any information delivery media. The term "modulated data signal" may include a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal.

[0052] Device 1112 may include input device(s) 1124 such as keyboard, mouse, pen, voice input device, touch input device, infrared cameras, video input devices, and/or any other input device. Output device(s) 1122 such as one or more displays, speakers, printers, and/or any other output device may also be included in device 1112. Input device(s) 1124 and output device(s) 1122 may be connected to device 1112 via a wired connection, wireless connection, or any combination thereof. In one embodiment, an input device or an output device from another computing device may be used as input device(s) 1124 or output device(s) 1122 for computing device 1112.

[0053] Components of computing device 1112 may be connected by various interconnects, such as a bus. Such interconnects may include a Peripheral Component Interconnect (PCI), such as PCI Express, a Universal Serial Bus (USB), firewire (IEEE 1394), an optical bus structure, and the like. In another embodiment, components of computing device 1112 may be interconnected by a network. For example, memory 1118 may be comprised of multiple physical memory units located in different physical locations interconnected by a network.

[0054] Those skilled in the art will realize that storage devices utilized to store computer readable instructions may be distributed across a network. For example, a computing device 1130 accessible via a network 1128 may store computer readable instructions to implement one or more embodiments provided herein. Computing device 1112 may access computing device 1130 and download a part or all of the computer readable instructions for execution. Alternatively, computing device 1112 may download pieces of the computer readable instructions, as needed, or some instructions may be executed at computing device 1112 and some at computing device 1130.

[0055] Various operations of embodiments are provided herein. In one embodiment, one or more of the operations described may constitute computer readable instructions stored on one or more computer readable media, which if executed by a computing device, will cause the computing device to perform the operations described. The order in which some or all of the operations are described should not be construed as to imply that these operations are necessarily order dependent. Alternative ordering will be appreciated by one skilled in the art having the benefit of this description. Further, it will be understood that not all operations are necessarily present in each embodiment provided herein.

[0056] Moreover, the word "exemplary" is used herein to mean serving as an example, instance, or illustration. Any aspect or design described herein as "exemplary" is not necessarily to be construed as advantageous over other aspects or designs. Rather, use of

the word exemplary is intended to present concepts in a concrete fashion. As used in this application, the term "or" is intended to mean an inclusive "or" rather than an exclusive "or." That is, unless specified otherwise, or clear from context, "X employs A or B" is intended to mean any of the natural inclusive permutations. That is, if X employs A; X employs B; or X employs both A and B, then "X employs A or B" is satisfied under any of the foregoing instances. In addition, the articles "a" and "an" as used in this application and the appended claims may generally be construed to mean "one or more" unless specified otherwise or clear from context to be directed to a singular form.

[0057] Also, although the disclosure has been shown and described with respect to one or more implementations, equivalent alterations and modifications will occur to others skilled in the art based upon a reading and understanding of this specification and the annexed drawings. The disclosure includes all such modifications and alterations and is limited only by the scope of the following claims. In particular regard to the various functions performed by the above described components (e.g., elements, resources, etc.), the terms used to describe such components are intended to correspond, unless otherwise indicated, to any component which performs the specified function of the described component (e.g., that is functionally equivalent), even though not structurally equivalent to the disclosed structure which performs the function in the herein illustrated exemplary implementations of the disclosure. In addition, while a particular feature of the disclosure may have been disclosed with respect to only one of several implementations, such feature may be combined with one or more other features of the other implementations as may be desired and advantageous for any given or particular application. Furthermore, to the extent that the terms "includes," "having," "has," "with," or variants thereof are used in either the detailed description or the claims, such terms are intended to be inclusive in a manner similar to the term "comprising."

CLAIMS

1. A method for determining a set of additional information relevant to an online video, comprising:

receiving an online video;

extracting a set of features from the online video, the set of features comprising at least one of a textual feature, a visual feature, and an audio feature; and

determining a set of additional information based upon the set of features.

2. The method of claim 1, comprising:

presenting the set of additional information on a display.

3. The method of claim 2, the presenting comprising at least one of:

presenting the set of additional information as a video;

presenting the set of additional information as an animation;

presenting the set of additional information as an image; and

presenting the set of additional information as text.

4. The method of claim 1, the extracting comprising:

extracting a textual feature using an OCR text recognition technique.

5. The method of claim 1, the extracting comprising:

extracting a visual feature using a scale-invariant feature transformation.

6. The method of claim 1, the visual feature comprising at least one of:

a color histogram;

textures; and

SIFT descriptors.

7. The method of claim 1, the determining comprising:

executing a multimodal relevance matching algorithm against a database using the set of features to determine the set of additional information.

8. A system for determining a set of additional information relevant to an online video, comprising:

a parsing component configured to:

parse an online video into one or more segments, a segment comprising a sequence of frames; and

for respective segments, extract one or more keyframes representative of a sequence of frames of a segment;

a feature extraction component configured to:

extract a set of features from the one or more keyframes of respective segments, the set of features comprising at least one of a textual feature, a visual feature, and an audio feature; and

an information extraction component configured to:

execute a multimodal relevance matching algorithm against a database using the set of features to determine the set of additional information.

9. The system of claim 8, the information extraction component configured to:

aggregate the set of additional information into a video.

10. The system of claim 9, comprising:

a presentation component configured to:

present the video to a user consuming the online video.

11. The system of claim 8, comprising:

a presentation component configured to:

present the set of additional information.

12. The system of claim 8, the feature extraction component comprising:

a text recognition component configured to extract one or more textual features using an OCR text recognition technique;

a visual recognition component configured to extract one or more visual features using a scale-invariant feature transformation; and

an audio recognition component configured to extract one or more audio features.

13. The system of claim 8, the multimodal relevance matching algorithm configured to:

perform a text based search algorithm upon the database using the one or more textual features to determine a first list of candidate additional information;

perform a visual feature matching algorithm upon the database using the one or more visual features to determine a second list of candidate additional information; and

perform a linear combination of the first list and second list to generate the set of additional information.

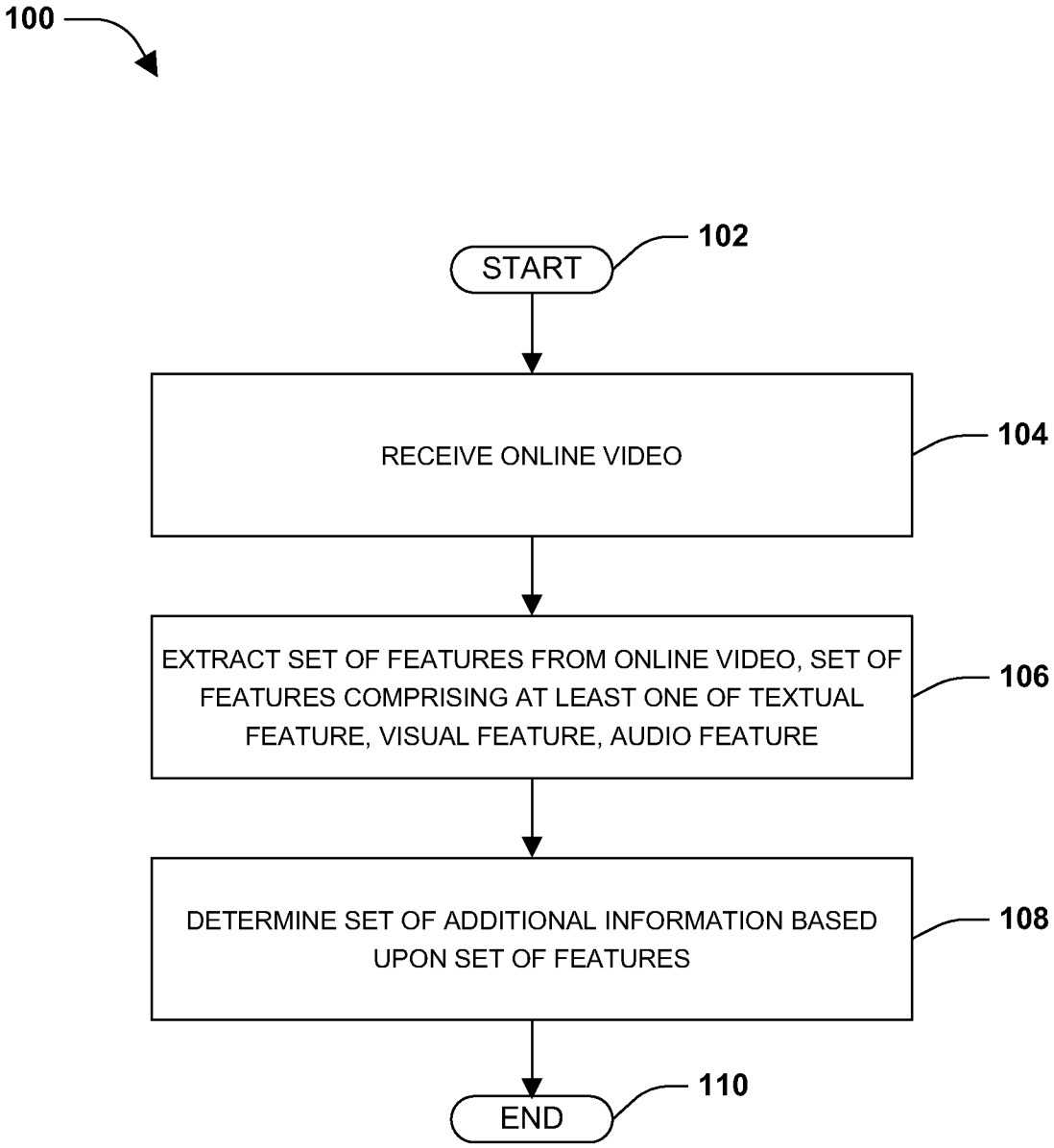
100
START
102
RECEIVE ONLINE VIDEO
104
EXTRACT SET OF FEATURES FROM ONLINE VIDEO, SET OF FEATURES COMPRISING AT LEAST ONE OF TEXTUAL FEATURE, VISUAL FEATURE, AUDIO FEATURE
106
DETERMINE SET OF ADDITIONAL INFORMATION BASED UPON SET OF FEATURES
108
END
110

FIG. 1

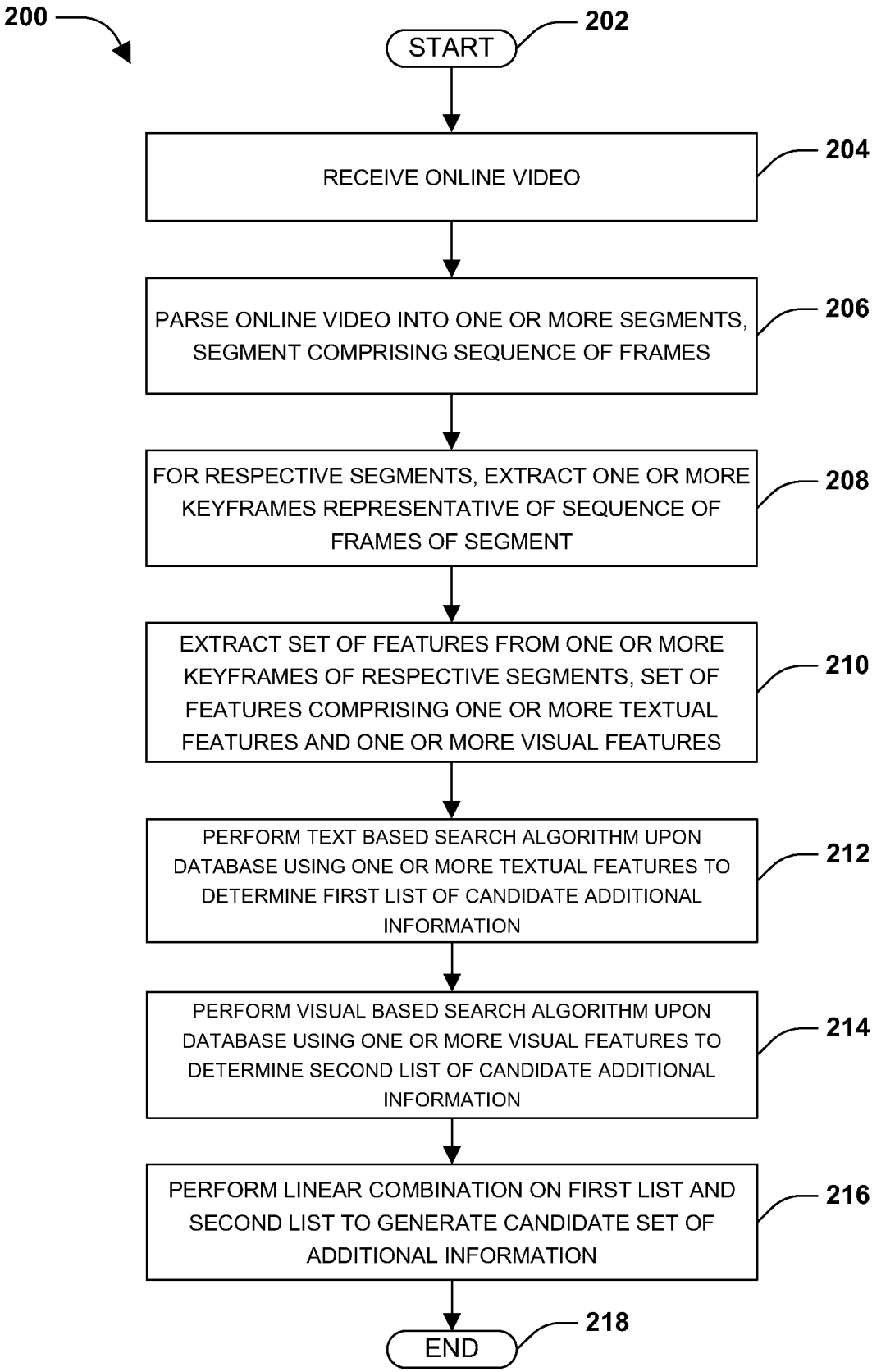
200
START
202
RECEIVE ONLINE VIDEO
204
PARSE ONLINE VIDEO INTO ONE OR MORE SEGMENTS, SEGMENT COMPRISING SEQUENCE OF FRAMES
206
FOR RESPECTIVE SEGMENTS, EXTRACT ONE OR MORE KEYFRAMES REPRESENTATIVE OF SEQUENCE OF FRAMES OF SEGMENT
208
EXTRACT SET OF FEATURES FROM ONE OR MORE KEYFRAMES OF RESPECTIVE SEGMENTS, SET OF FEATURES COMPRISING ONE OR MORE TEXTUAL FEATURES AND ONE OR MORE VISUAL FEATURES
210
PERFORM TEXT BASED SEARCH ALGORITHM UPON DATABASE USING ONE OR MORE TEXTUAL FEATURES TO DETERMINE FIRST LIST OF CANDIDATE ADDITIONAL INFORMATION
212
PERFORM VISUAL BASED SEARCH ALGORITHM UPON DATABASE USING ONE OR MORE VISUAL FEATURES TO DETERMINE SECOND LIST OF CANDIDATE ADDITIONAL INFORMATION
214
PERFORM LINEAR COMBINATION ON FIRST LIST AND SECOND LIST TO GENERATE CANDIDATE SET OF ADDITIONAL INFORMATION
216
END
218

FIG. 2

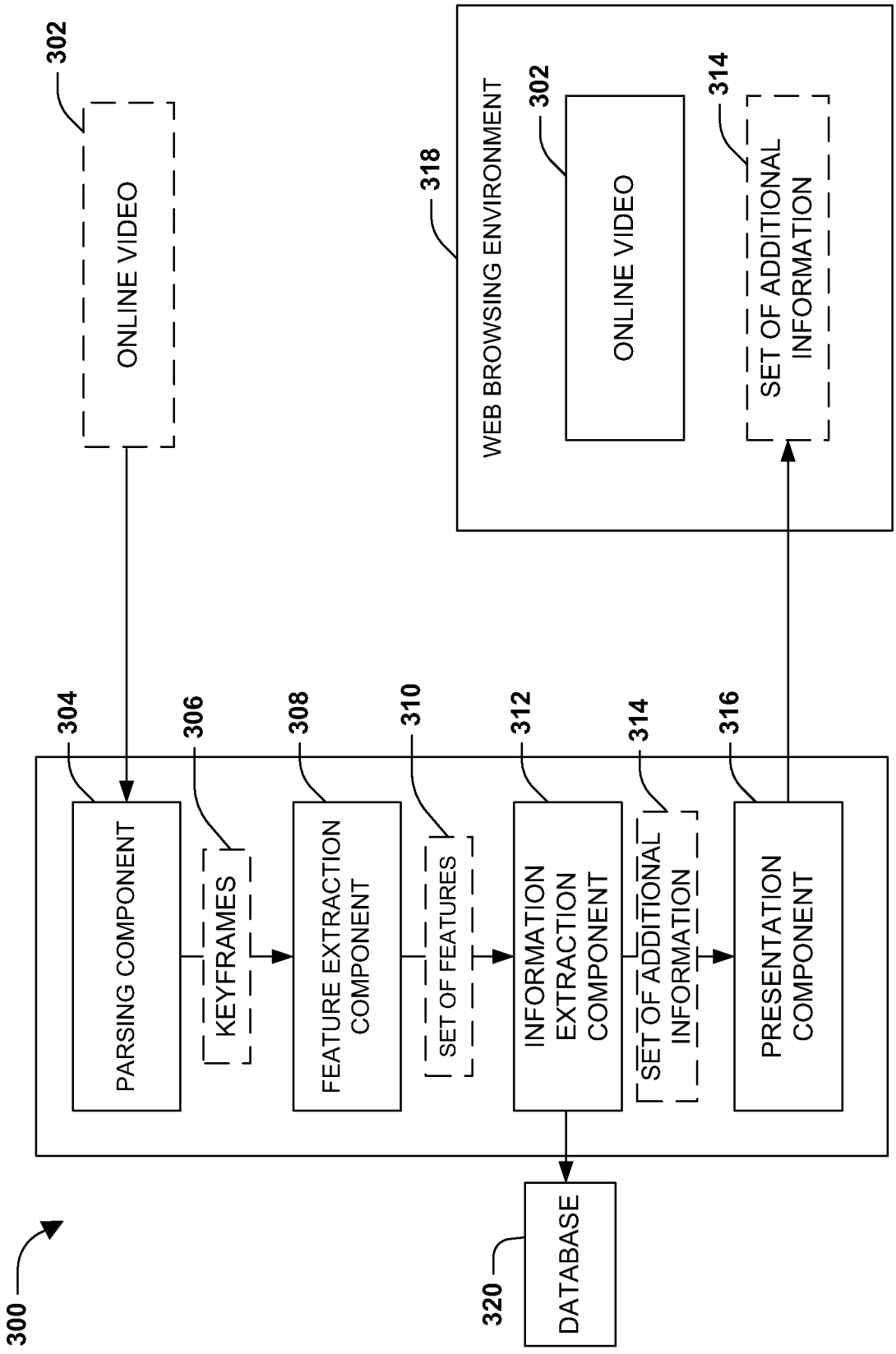
300
302
ONLINE VIDEO
318
WEB BROWSING ENVIRONMENT
302
ONLINE VIDEO
314
SET OF ADDITIONAL INFORMATION
304
PARSING COMPONENT
306
KEYFRAMES
308
FEATURE EXTRACTION COMPONENT
310
SET OF FEATURES
312
INFORMATION EXTRACTION COMPONENT
314
SET OF ADDITIONAL INFORMATION
316
PRESENTATION COMPONENT
320
DATABASE

FIG. 3

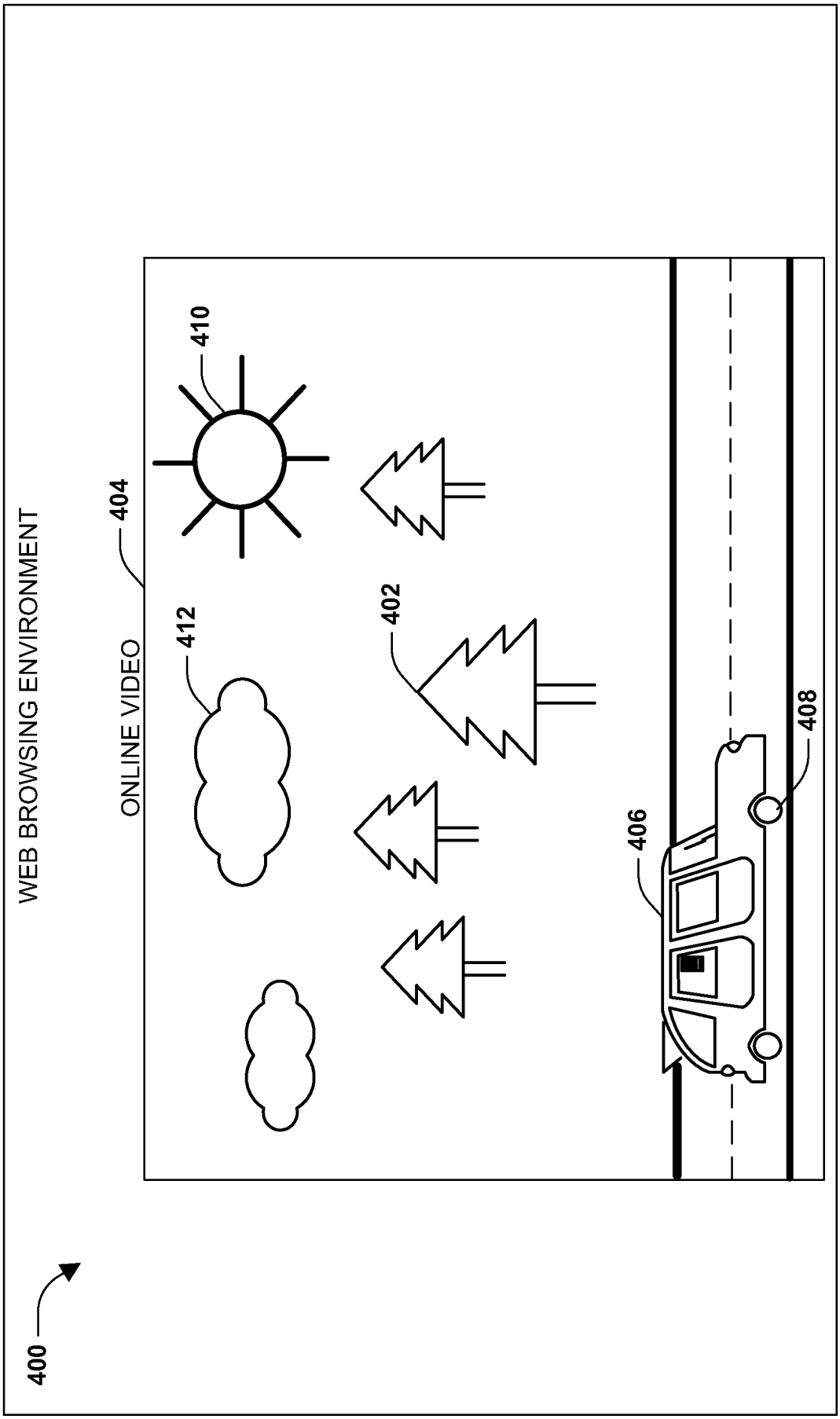
WEB BROWSING ENVIRONMENT
ONLINE VIDEO
400
404
410
412
402
406
408

FIG. 4

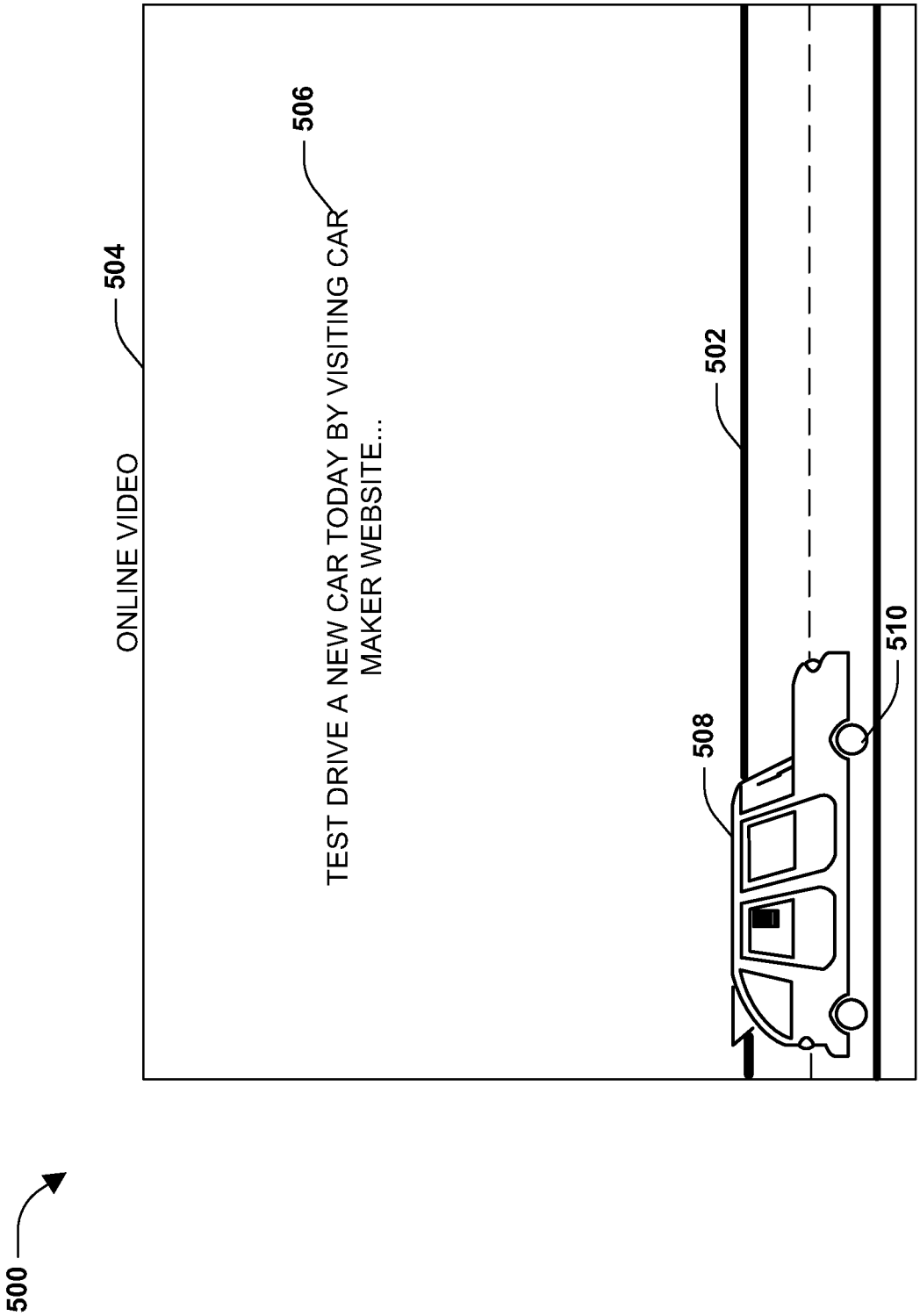
500
504
ONLINE VIDEO
506
TEST DRIVE A NEW CAR TODAY BY VISITING CAR MAKER WEBSITE...
502
508
510

FIG. 5

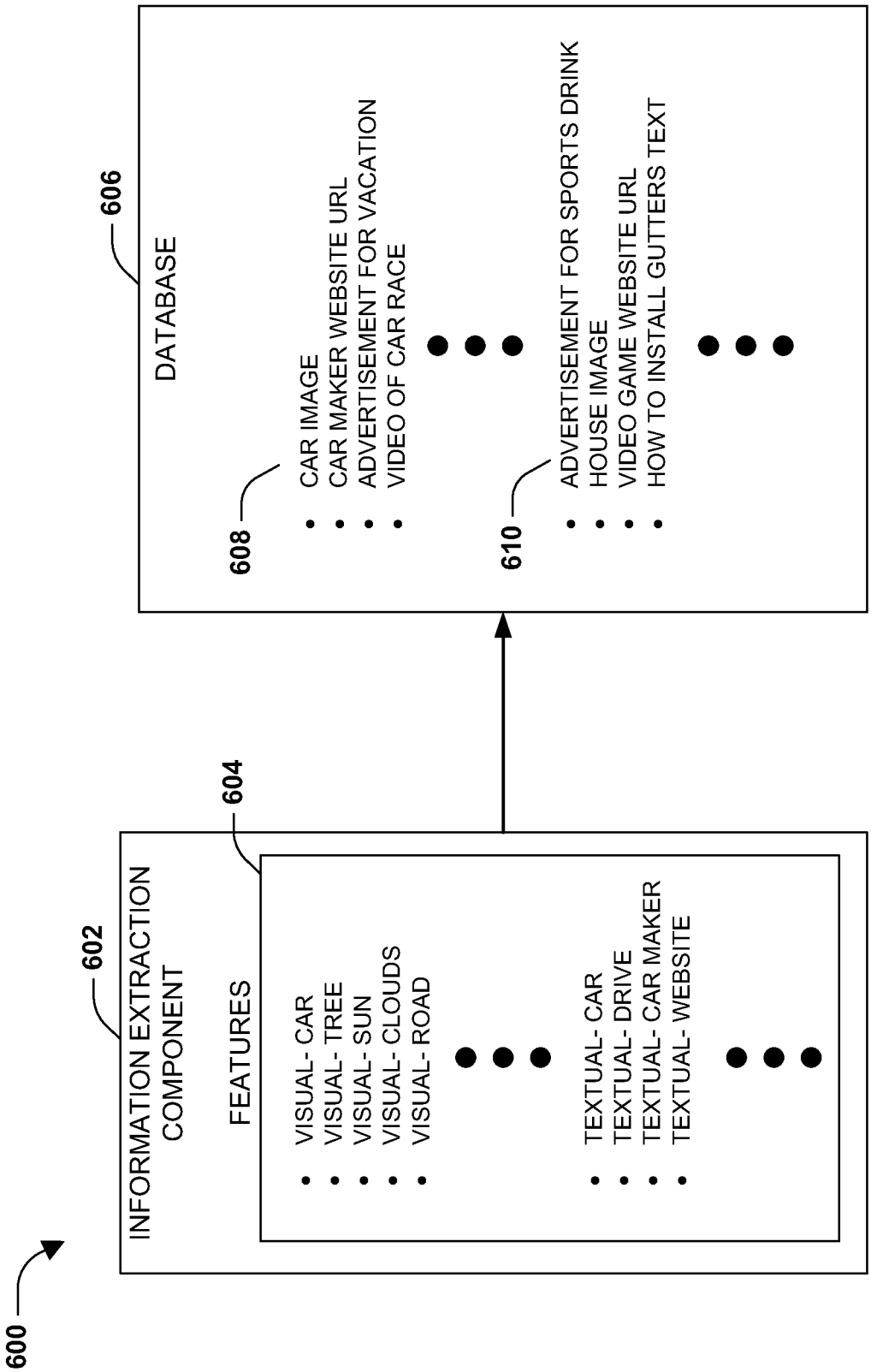
600
602
INFORMATION EXTRACTION COMPONENT
604
FEATURES
VISUAL- CAR
VISUAL- TREE
VISUAL- SUN
VISUAL- CLOUDS
VISUAL- ROAD
TEXTUAL- CAR
TEXTUAL- DRIVE
TEXTUAL- CAR MAKER
TEXTUAL- WEBSITE
606
DATABASE
608
CAR IMAGE
CAR MAKER WEBSITE URL
ADVERTISEMENT FOR VACATION
VIDEO OF CAR RACE
610
ADVERTISEMENT FOR SPORTS DRINK
HOUSE IMAGE
VIDEO GAME WEBSITE URL
HOW TO INSTALL GUTTERS TEXT

FIG. 6

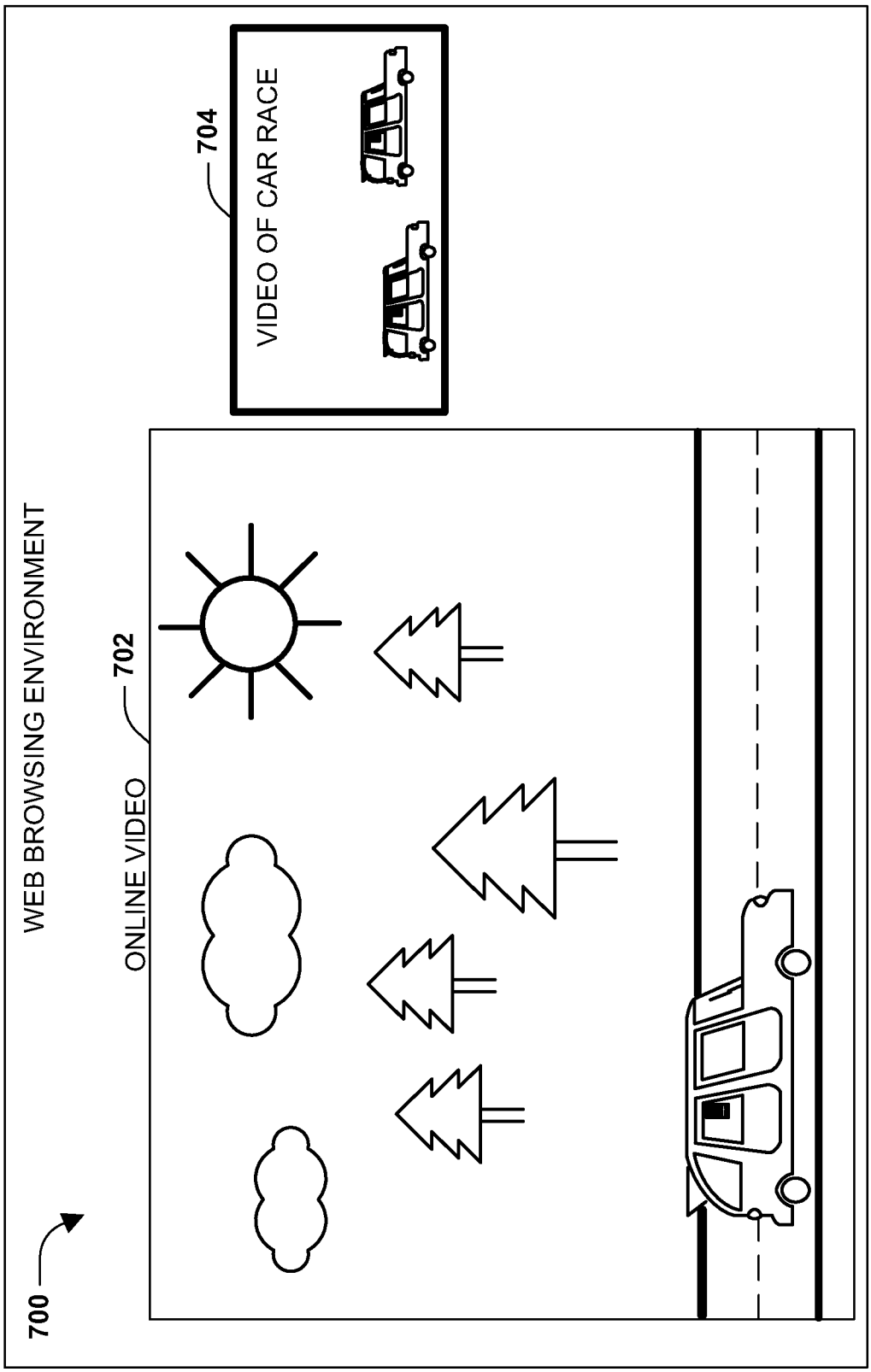
WEB BROWSING ENVIRONMENT
700
ONLINE VIDEO
702
704
VIDEO OF CAR RACE

FIG. 7

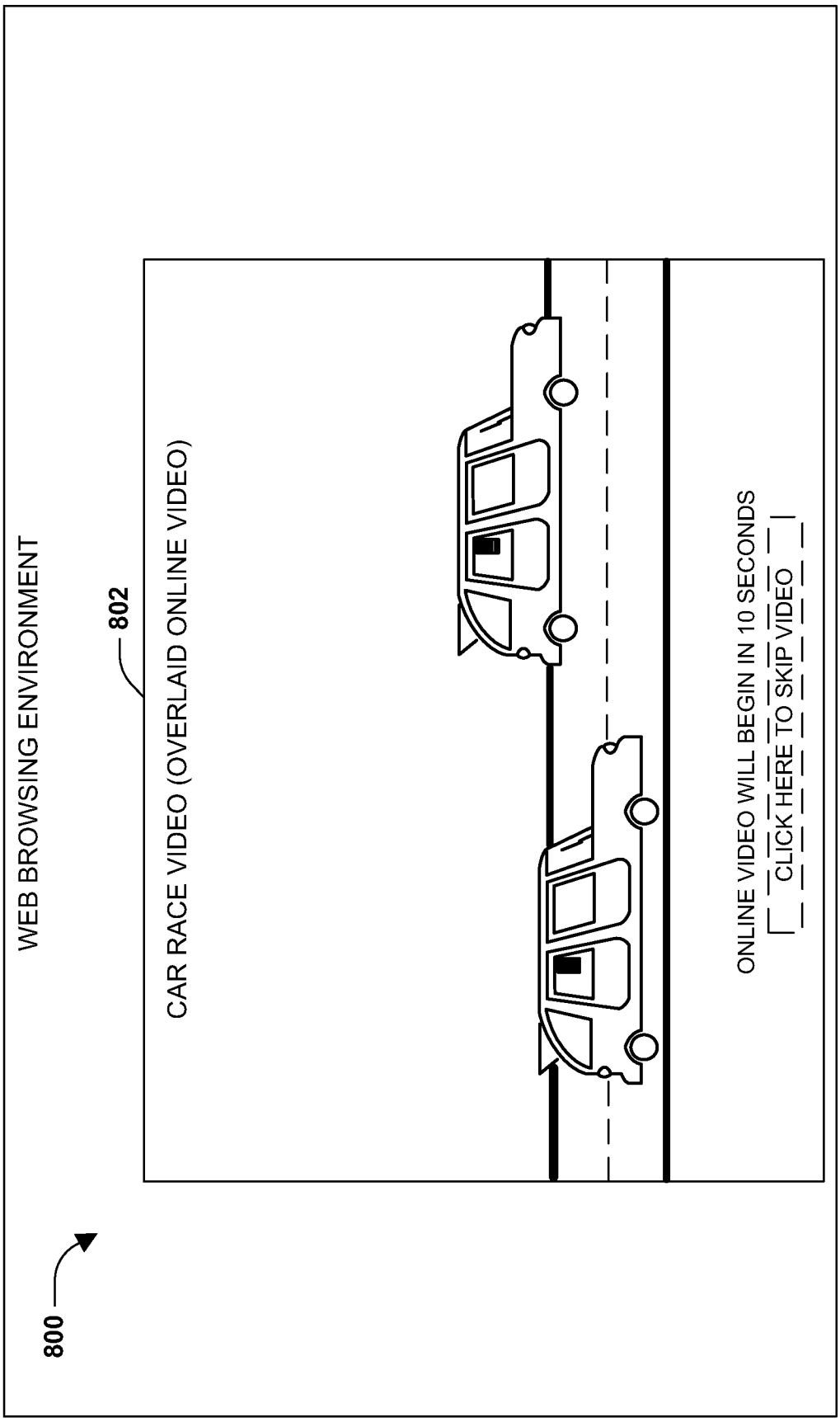
WEB BROWSING ENVIRONMENT
800
802
CAR RACE VIDEO (OVERLAID ONLINE VIDEO)
ONLINE VIDEO WILL BEGIN IN 10 SECONDS
CLICK HERE TO SKIP VIDEO

FIG. 8

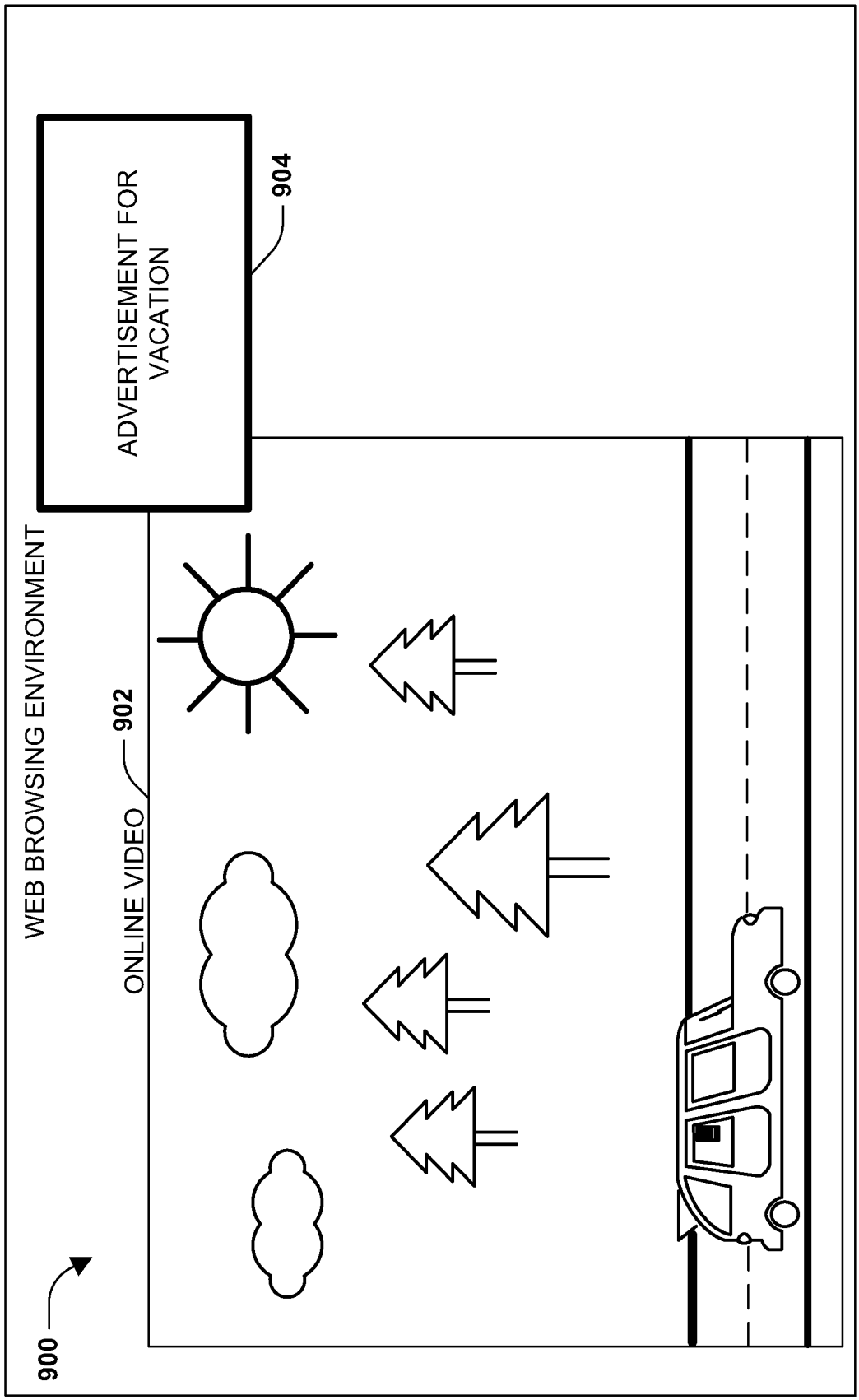
WEB BROWSING ENVIRONMENT
900
ONLINE VIDEO
902
ADVERTISEMENT FOR VACATION
904

FIG. 9

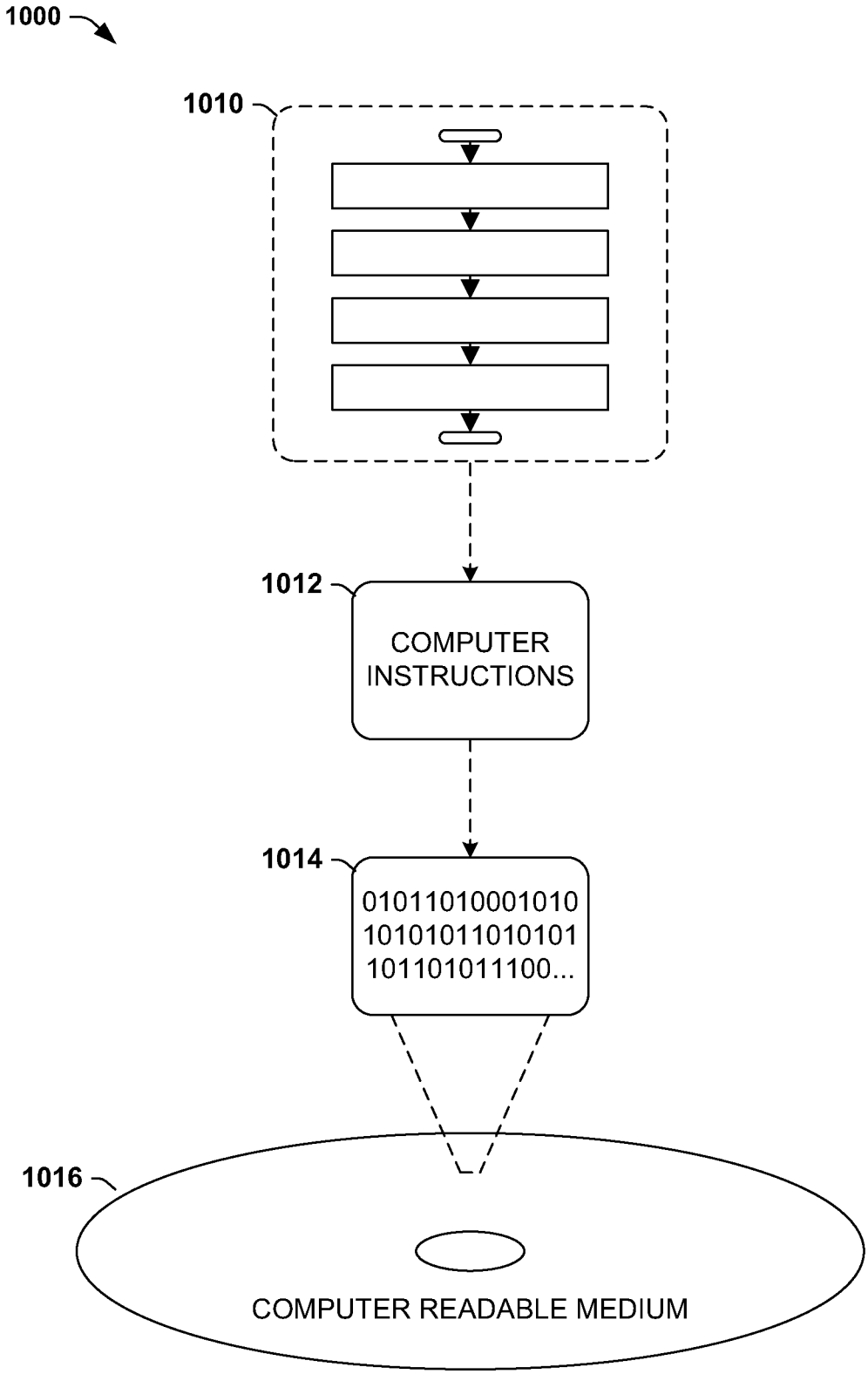
1000
1010
1012
COMPUTER INSTRUCTIONS
1014
01011010001010
10101011010101
101101011100...
1016
COMPUTER READABLE MEDIUM

FIG. 10

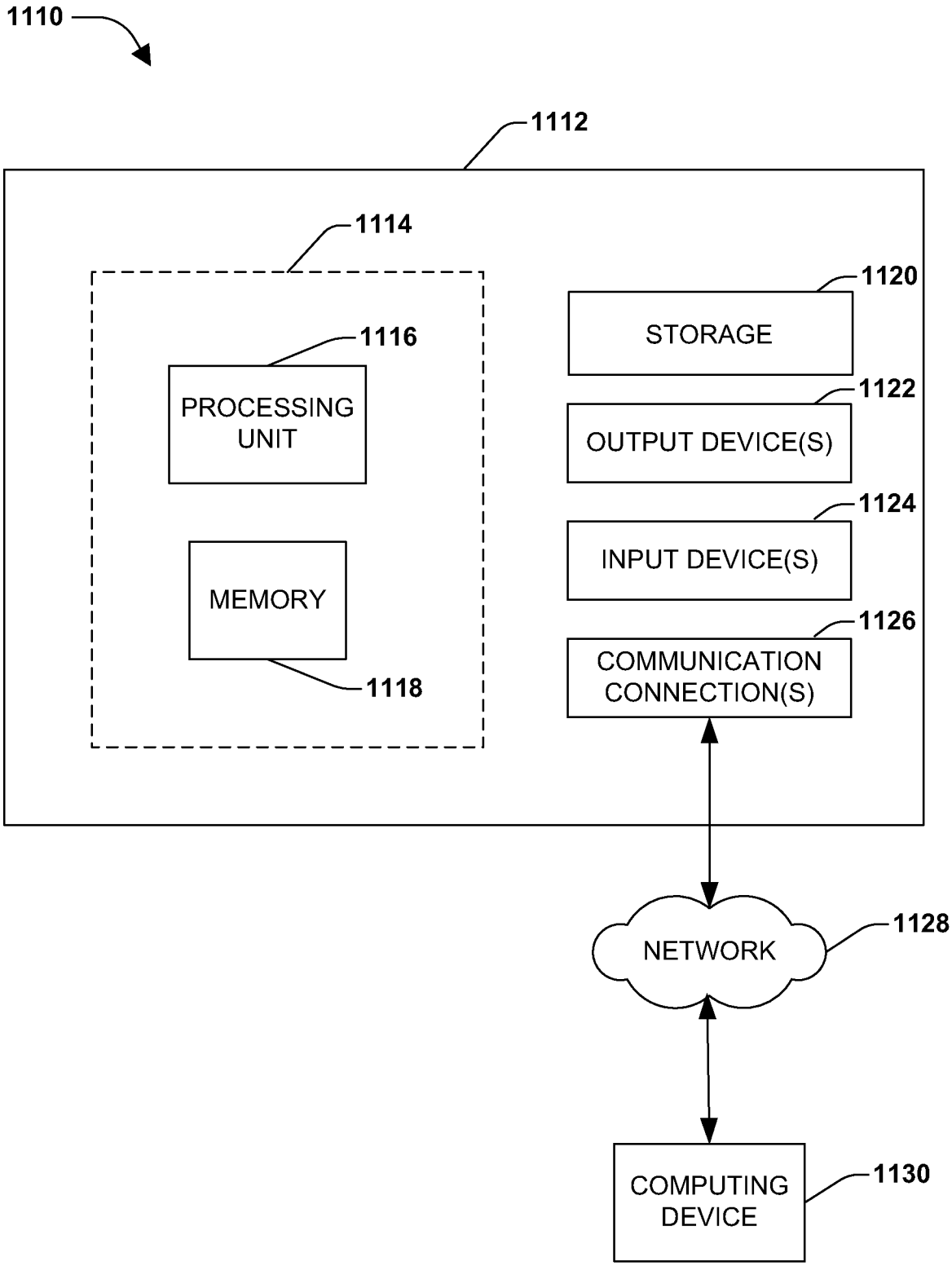
1110
1112
1114
1116
PROCESSING UNIT
MEMORY
1118
1120
STORAGE
1122
OUTPUT DEVICE(S)
1124
INPUT DEVICE(S)
1126
COMMUNICATION CONNECTION(S)
NETWORK
1128
COMPUTING DEVICE
1130

FIG. 11