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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 February 2004 (05.02.2004)

PCT

(10) International Publication Number
WO 2004/012105 A2

(51) International Patent Classification[7]: G06F 17/30

(21) International Application Number:
PCT/US2003/023809

(22) International Filing Date: 29 July 2003 (29.07.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/207,279 29 July 2002 (29.07.2002) US

(71) Applicant: **HEWLETT-PACKARD DEVELOPMENT COMPANY, L.P.** [US/US]; 20555 S.H. 249, Houston, TX 77070 (US).

(72) Inventor: **OBRADOR, Pere**; 2309 Rock Street, #21, Mountain View, CA 94043 (US).

(74) Agent: **HEMINGER, Susan, E.**; Hewlett-Packard Company, IP Administration, P.O. Box 272400, Ft. Collins, CO 80527-2400 (US).

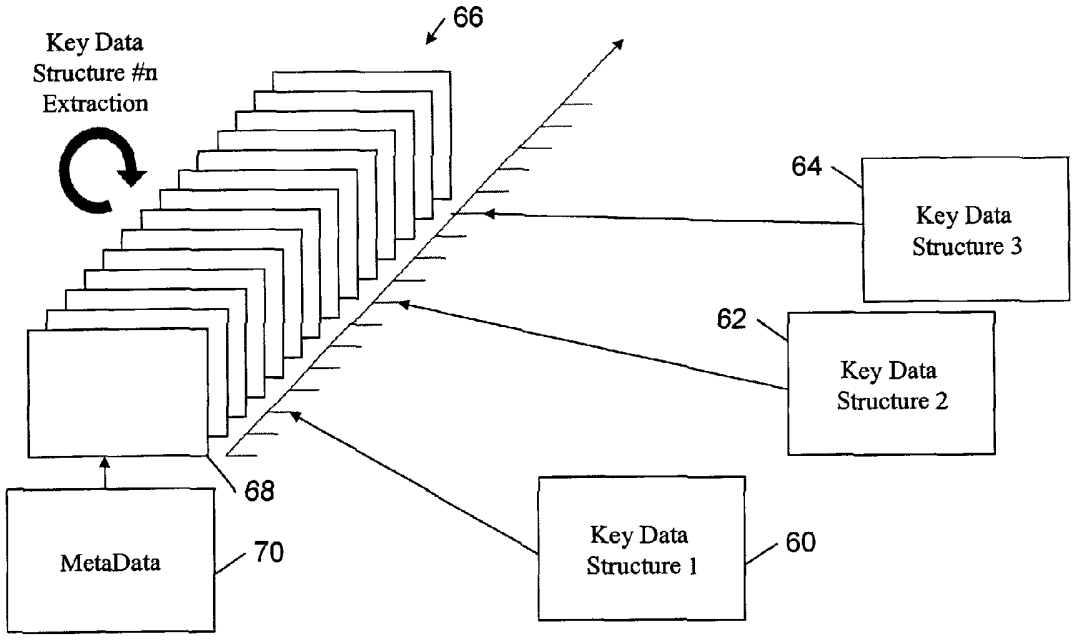
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (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*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: PRESENTING A COLLECTION OF MEDIA OBJECTS



(57) Abstract: Systems and methods of presenting media objects are described. In one aspect, a group of media objects is selected from the collection based upon media object relevance to one or more data structures of a selected media file of indexed, temporally-ordered data structures. The selected media file and the media objects of the selected group are presented contemporaneously. In another aspect, media objects in the collection are grouped into multiple clusters based upon one or more media object relevance criteria. The media object clusters are arranged into a hierarchy of two or more levels. A selected media object cluster is presented contemporaneously. Systems for implementing these media object presentation methods also are described.

WO 2004/012105 A2

PRESENTING A COLLECTION OF MEDIA OBJECTS

TECHNICAL FIELD

This invention relates to systems and methods of presenting media objects.

BACKGROUND

Individuals and organizations are rapidly accumulating large collections of digital content, including text, audio, graphics, animated graphics and full-motion video. This content may be presented individually or combined in a wide variety of different forms, including documents, presentations, music, still photographs, commercial videos, home movies, and meta data describing one or more associated digital content files. As these collections grow in number and diversity, individuals and organizations increasingly will require systems and methods for organizing and browsing the digital content in their collections. To meet this need, a variety of different systems and methods for browsing selected kinds of digital content have been proposed.

For example, storyboard browsing has been developed for browsing full-motion video content. In accordance with this technique, video information is condensed into meaningful representative snapshots and corresponding audio content. One known video browser of this type divides a video sequence into equal length segments and denotes the first frame of each segment as its key frame. Another known video browser of this type stacks every frame of the sequence and provides the user with information regarding the camera and object motions.

Content-based video browsing techniques also have been proposed. In these techniques, a long video sequence typically is classified into story units based on video content. In some approaches, scene change detection (also called temporal segmentation of video) is used to give an indication of when a new shot starts and ends. Scene change detection algorithms, such as scene transition detection algorithms based on DCT (Discrete Cosine Transform) coefficients of an encoded image, and algorithms that are configured to identify both abrupt and gradual scene transitions using the DCT coefficients of an encoded video sequence are known in the art.

In one video browsing approach, Rframes (representative frames) are used to organize the visual contents of video clips. Rframes may be grouped according to various criteria to aid the user in identifying the desired material. In this approach, the user may

select a key frame, and the system then uses various criteria to search for similar key frames and present them to the user as a group. The user may search representative frames from the groups, rather than the complete set of key frames, to identify scenes of interest. Language-based models have been used to match incoming video sequences with the expected grammatical elements of a news broadcast. In addition, *a priori* models of the expected content of a video clip have been used to parse the clip.

In another approach, U.S. Patent No. 5,821,945 has proposed a technique for extracting a hierarchical decomposition of a complex video selection for video browsing purposes. This technique combines visual and temporal information to capture the important relations within a scene and between scenes in a video, thus allowing the analysis of the underlying story structure with no *a priori* knowledge of the content. A general model of hierarchical scene transition graph is applied to an implementation for browsing. Video shots are first identified and a collection of key frames is used to represent each video segment. These collections are then classified according to gross visual information. A platform is built on which the video is presented as directed graphs to the user, with each category of video shots represented by a node and each edge denoting a temporal relationship between categories. The analysis and processing of video is carried out directly on the compressed videos.

A variety of different techniques that allow media files to be searched through associated annotations also have been proposed. For example, U.S. Patent No. 6,332,144 has proposed a technique in accordance with which audio/video media is processed to generate annotations that are stored in an index server. A user may browse through a collection of audio/video media by submitting queries to the index server. In response to such queries, the index server transmits to a librarian client each matching annotation and a media identification number associated with each matching annotation. The librarian client transmits to the user the URL (uniform resource locator) of the digital representation from which each matching annotation was generated and an object identification number associated with each matching annotation. The URL may specify the location of all or a portion of a media file.

SUMMARY

In one aspect, the invention features a method of presenting a collection of media objects. In accordance with this inventive method, a group of media objects is selected

from the collection based upon media object relevance to one or more data structures of a selected media file of indexed, temporally-ordered data structures. The selected media file and the media objects of the selected group are presented contemporaneously.

In another aspect of the invention, media objects in the collection are grouped into multiple clusters based upon one or more media object relevance criteria. The media object clusters are arranged into a hierarchy of two or more levels. A selected media object cluster is presented contemporaneously.

The invention also features systems for implementing the above-described media object presentation methods.

Other features and advantages of the invention will become apparent from the following description, including the drawings and the claims.

DESCRIPTION OF DRAWINGS

FIG. 1 is a diagrammatic view of a media management node coupled directly to a set of local media files and coupled indirectly to multiple sets of remote media files over a local area network and a global network infrastructure.

FIG. 2 is a diagrammatic view of a computer system that is programmable to implement a method of managing media objects.

FIG. 3 is a diagrammatic perspective view of a media file of indexed, temporally-ordered data structures and an automatically-generated selection of key data structures.

FIG. 4 is a diagrammatic perspective view of the media file of FIG. 3 after the selection of key data structures has been modified by a user.

FIG. 5 is a diagrammatic perspective view of an indexed media file containing a sequence of full-motion video frames, a selection of keyframes, and a high resolution still photograph.

FIG. 6 is a diagrammatic perspective view of the indexed media file, keyframe selection and high resolution still photograph of FIG. 5, along with multiple user-selected media objects that are linked to respective video frames of the indexed media file.

FIG. 7A is a diagrammatic perspective view of the links connecting the keyframes, the high resolution still photograph, and the media objects to the indexed media file of FIG. 6.

FIG. 7B is a diagrammatic perspective view of a database storing the indexed media file, keyframes, high resolution still photograph, media objects and connecting links of FIG. 7A.

FIG. 8A is a diagrammatic perspective view of a video file mapped into a set of video sequences.

FIG. 8B is a diagrammatic perspective view of a set of video sequences mapped into a common video file.

FIG. 8C is a diagrammatic perspective view of a set of consecutive video sequences mapped into two video files.

FIG. 8D is a diagrammatic perspective view of a set of non-consecutive video sequences mapped into two video files.

FIG. 9 is a flow diagram of a method of presenting a collection of media objects.

FIGS. 10A-10C are diagrammatic views of a multimedia album page during presentation of a video file and a selected group of media objects that are linked to the video frames of the video file.

FIG. 11 is a diagrammatic view of a graphical user interface simultaneously presenting a selected video file and a group of media objects selected based upon a metadata similarity criterion.

FIG. 12 is a diagrammatic view of a graphical user interface presenting the media objects of FIG. 11 as a video card stack adjacent to the selected video file.

FIG. 13 is a hierarchical arrangement of media objects that are grouped into clusters based upon one or more media object relevance criteria.

DETAILED DESCRIPTION

In the following description, like reference numbers are used to identify like elements. Furthermore, the drawings are intended to illustrate major features of exemplary embodiments in a diagrammatic manner. The drawings are not intended to depict every feature of actual embodiments nor relative dimensions of the depicted elements, and are not drawn to scale.

MEDIA MANAGEMENT SYSTEM OVERVIEW

Referring to FIG. 1, in one embodiment, a media management node 10 includes a media manager 12 that is configured to enable all forms of digital content in a selected collection of media objects to be organized into a browsable context-sensitive,

temporally-referenced media database. As used herein, the term "media object" refers broadly to any form of digital content, including text, audio, graphics, animated graphics and full-motion video. This content may be packaged and presented individually or in some combination in a wide variety of different forms, including documents, annotations, presentations, music, still photographs, commercial videos, home movies, and meta data describing one or more associated digital content files. The media objects may be stored physically in a local database 14 of media management node 10 or in one or more remote databases 16, 18 that may be accessed over a local area network 20 and a global communication network 22, respectively. Some media objects also may be stored in a remote database 24 that is accessible over a peer-to-peer network connection. In some embodiments, digital content may be compressed using a compression format that is selected based upon digital content type (e.g., an MP3 or a WMA compression format for audio works, and an MPEG or a motion JPEG compression format for audio/video works). The requested digital content may be formatted in accordance with a user-specified transmission format. For example, the requested digital content may be transmitted to the user in a format that is suitable for rendering by a computer, a wireless device, or a voice device. In addition, the requested digital content may be transmitted to the user as a complete file or in a streaming file format.

A user may interact with media manager 12 locally, at media management node 10, or remotely, over local area network 20 or global communication network 22. Transmissions between media manager 12, the user, and the content providers may be conducted in accordance with one or more conventional secure transmission protocols. For example, each digital work transmission may involve packaging the digital work and any associated meta-data into an encrypted transfer file that may be transmitted securely from one entity to another.

Global communication network 22 may include a number of different computing platforms and transport facilities, including a voice network, a wireless network, and a computer network. Media object requests may be transmitted, and media object replies may be presented in a number of different media formats, such as voice, Internet, e-mail and wireless formats. In this way, users may access the services provided by media management node 10 and the remote media objects 16 provided by service provider 26 and peer-to-peer node 24 using any one of a wide variety of different communication devices. For example, in one illustrative implementation, a wireless device (e.g., a

wireless personal digital assistant (PDA)) may connect to media management node 10, service provider 26, and peer-to-peer node 24 over a wireless network. Communications from the wireless device may be in accordance with the Wireless Application Protocol (WAP). A wireless gateway converts the WAP communications into HTTP messages that may be processed by service provider 10. In another illustrative implementation, a voice device (e.g., a conventional telephone) may connect to media management node 10, service provider 26 and peer-to-peer node 24 over a voice network. Communications from the voice device may be in the form of conventional analog or digital audio signals, or they may be formatted as VoxML messages. A voice gateway may use speech-to-text technology to convert the audio signals into HTTP messages; VoxML messages may be converted to HTTP messages based upon an extensible style language (XSL) style specification. The voice gateway also may be configured to receive real time audio messages that may be passed directly to the voice device. Alternatively, the voice gateway may be configured to convert formatted messages (e.g., VoxML, XML, WML, e-mail) into a real time audio format (e.g., using text-to-speech technology) before the messages are passed to the voice device. In a third illustrative implementation, a software program operating at a client personal computer (PC) may access the services of media management node 10 and the media objects provided by service provider 26 and peer-to-peer node 24 over the Internet.

As explained in detail below, media manager 12 enables a user to organize and browse through a selected collection of media objects by means of a set of links between media objects. In general, all media objects may be indexed by any other media object in the selected collection. Each link may be browsed from one media object to a linked media object, and vice versa. The set of links between media objects may be generated by a user, a third party, or automatically by media manager 12. These links are stored separately from the media objects in one or more media object linkage data structures that are accessible by the media manager 12.

Content manager 12 may provide access to a selected digital content collection in a variety of different ways. In one embodiment, a user may organize and browse through a personal collection of a diverse variety of interlinked media objects. In another embodiment, content manager 12 may operate an Internet web site that may be accessed by a conventional web browser application program executing on a user's computer system. The web site may present a collection of personal digital content, commercial

digital content and/or publicly available digital content. The web site also may provide additional information in the form of media objects that are linked to the available digital content. Users may specify links to be generated and browse through the collection of digital content using media objects as links into and out of specific digital content files.

In an alternative embodiment, a traditional brick-and-mortar retail establishment (e.g., a bookstore or a music store) may contain one or more kiosks (or content preview stations). The kiosks may be configured to communicate with media manager 12 (e.g., over a network communication channel) to provide user access to digital content that may be rendered at the kiosk or transferred to a user's portable media device for later playback. A kiosk may include a computer system with a graphical user interface that enables users to establish links and navigate through a collection of digital content that is stored locally at the retail establishment or that is stored remotely and is retrievable over a network communication channel. A kiosk also may include a cable port that a user may connect to a portable media device for downloading selected digital content.

In embodiments in which a user interacts remotely with media manager 12, the user may store the media object linkage data structures that are generated during a session in a portable storage device or on a selected network storage location that is accessible over a network connection.

Referring to FIG. 2, in one embodiment, content manager 12 may be implemented as one or more respective software modules operating on a computer 30. Computer 30 includes a processing unit 32, a system memory 34, and a system bus 36 that couples processing unit 32 to the various components of computer 30. Processing unit 32 may include one or more processors, each of which may be in the form of any one of various commercially available processors. System memory 34 may include a read only memory (ROM) that stores a basic input/output system (BIOS) containing start-up routines for computer 30 and a random access memory (RAM). System bus 36 may be a memory bus, a peripheral bus or a local bus, and may be compatible with any of a variety of bus protocols, including PCI, VESA, Microchannel, ISA, and EISA. Computer 30 also includes a persistent storage memory 38 (e.g., a hard drive, a floppy drive 126, a CD ROM drive, magnetic tape drives, flash memory devices, and digital video disks) that is connected to system bus 36 and contains one or more computer-readable media disks that provide non-volatile or persistent storage for data, data structures and computer-executable instructions. A user may interact (e.g., enter commands or data) with

computer 30 using one or more input devices 40 (e.g., a keyboard, a computer mouse, a microphone, joystick, and touch pad). Information may be presented through a graphical user interface (GUI) that is displayed to the user on a display monitor 42, which is controlled by a display controller 44. Computer 30 also may include peripheral output devices, such as speakers and a printer. One or more remote computers may be connected to computer 30 through a network interface card (NIC) 46.

As shown in FIG. 2, system memory 34 also stores media manager 12, a GUI driver 48, and one or more media object linkage structures 50. Media manager 12 interfaces with the GUI driver 48 and the user input 40 to control the creation of the media object linkage data structures 50. Media manager 12 also interfaces with the GUI driver 48 and the media object linkage data structures to control the media object browsing experience presented to the user on display monitor 42. The media objects in the collection to be linked and browsed may be stored locally in persistent storage memory 38 or stored remotely and accessed through NIC 46, or both.

LINKING MEDIA OBJECTS

Referring to FIG. 3, in one embodiment, media manager 12 may be configured to automatically generate a selection of key data structures 60, 62, 64 from a media file 66 of indexed, temporally-ordered data structures 68. Media file 66 may correspond to any kind of digital content that is indexed and temporally-ordered (i.e., ordered for playback in a specific time sequence), including frames of a full-motion video, animated graphics, slides (e.g., PowerPoint® slides, text slides, and image slides) organized into a slideshow presentation, and segments of digital audio. Key data structures 60-64 may be extracted in accordance with any one of a variety of conventional automatic key data structure extraction techniques (e.g., automatic keyframe extraction techniques used for full-motion video). Media manager 12 also may be configured to link meta data 70 with the first data structure 68 of media file 66. In this embodiment, each of the media file data structures 68 is associated with an index value (e.g., a frame number or time-stamp number for full-motion video). Each of the links between media objects 60-64, 70 and media file data structures 68 is a pointer between the index value associated with the media file data structure 68 and the address of one of the linked media objects 60-64, 70. Each link is browsable from a given data structure 68 of media file 66 to a media object 60-64, 70, and vice versa. The links may be stored in one or more media object data structures in, for example, an XML (Extensible Markup Language) format.

As shown in FIG.4, in one embodiment, media manager 12 is configured to modify the initial selection of key data structures in response to user input. For example, in the illustrated embodiment, a user may remove key data structure 64 and add a new key data structure 72. In addition, a user may change the data structure 68 of media file 66 to which key data structure 62 is linked. In this embodiment, the data structures 68 of media file 68 preferably are presented to the user in the graphical user interface as a card stack. In this presentation, the user may select one of the data structures 68 with a pointing device (e.g., a computer mouse) and media manager 12 will present the contents of the selected data structure to the user for review. In other embodiments, the data structures 68 of media file 66 may be presented to the user in an array or one-by-one in sequence.

Referring to FIGS. 5 and 6, in one illustrative embodiment, media file 66 corresponds to a video file sequence 73 of full-motion video frames 74. After automatic keyframe extraction and user-modification, two keyframes 76, 78 and a high resolution still photograph 80 are linked to video file 73. As shown in FIG. 6, in addition to modifying the selection of keyframes 76-80, a user may link other media objects to the video frames 74. For example, the user may link a text file annotation 82 to video file 73. The user also may link an XHTML (Extensible HyperText Markup Language) document 84 to the video frame corresponding to keyframe 78. XHTML document 84 may include a hypertext link 86 that contains the URL (Uniform Resource Locator) for another media object (e.g., a web page). The user also may link an audio file 88 to the video frame corresponding to keyframe 80. In the illustrated embodiment, for example, the linked audio file 88 may correspond to the song being played by a person appearing in the associated video keyframe 80. The user also may link a full-motion video file 90 to a frame 92 of video file 73. In the illustrated embodiment, for example, the linked video file 90 may correspond to a video of a person appearing in the associated video frame 92. The user also may link to the video frame corresponding to keyframe 80 a text file 94 containing meta data relating to the associated video frame 80. For example, in the illustrated embodiment, video frame 80 may correspond to a high-resolution still image and meta data file 94 may correspond to the meta data that was automatically generated by the video camera that captured the high-resolution still image.

Referring to FIGS. 7A and 7B, in one embodiment, after video file 73 has been enriched with links to other media objects, the resulting collection of media objects and

media object linkage data structures (shown in FIG. 7A) may be stored as a context-sensitive, temporally-referenced media database 96 (shown in FIG. 7B). This database 96 preserves temporal relationships and associations between media objects. The database 96 may be browsed in a rich and meaningful way that allows target contents to be found rapidly and efficiently from associational links that may evolve over time. All media objects linked to the video file 73 may share annotations and links with other media objects. In this way, new or forgotten associations may be discovered while browsing through the collection of media objects.

Referring to FIGS. 8A-8D, in some embodiments, all media files in a selected collection are stored only once in data base 96 (FIG. 7B). Each media file (e.g., video file 73) of indexed, temporally-ordered data structures may be split logically into a set of data structure sequences that are indexed with logical links into the corresponding media file. Media objects 98 may be indexed with logical links into the set of data structure sequences, as shown in FIG. 8A. Each data structure sequence link into a media file may identify a starting point in the media file and the length of the corresponding sequence. The data structure sequences may be consecutive, as shown in FIG. 8B, or non-consecutive. In addition, the set of data structure sequences may map consecutively into multiple media files, as shown in FIG. 8C. Alternatively, the set of data structure sequences may be mapped non-consecutively into multiple media files.

PRESENTING AND BROWSING A COLLECTION A MEDIA OBJECTS

Referring to FIG. 9, in some embodiments, media manager 12 may be configured to present a collection of media objects to a user as follows. A group of media objects is selected from the collection based upon their relevance to one or more data structures of a selected media file of indexed, temporally-ordered data structures (step 100). As explained in detail below, the relevance criterion for selecting media objects may relate to browsable links between media objects and the selected media file, or the relevance criteria may relate to a selected metadata similarity between media objects and the selected media file. Next, the media file and the media objects of the selected group are presented contemporaneously to a user for browsing (step 102). The media file and the media objects preferably are presented to the user through multimedia album page, which is a windows-based GUI that is displayed on display monitor 42 (FIG. 2).

BROWSING MEDIA OBJECT LINKS

In some embodiments, the inter-media-object linking architecture described above may be used by media manager 12 to present the media objects of a collection to a user in a context-sensitive, temporally-referenced way through a multimedia album page 104, which is anchored to a selected media file (e.g., video file 73). For example, media manager 12 may present the data structures of the media file in sequence and, during this presentation, media manager 12 may present media objects in the selected group at times when they are relevant to the data structure (or data structure sequence) currently being presented to the user. In these embodiments, the relevance criterion for selecting the group of objects contemporaneously with the media file relates to the existence of browsable links between the selected media objects and data structures of the media file. The browsable links may be established automatically by media manager 12 and modified by a user, as described above. The browsable links also may be established by media manager 12 just before (e.g., during a preprocessing period) or on the fly during the presentation of the media file by instantiating one or more context matching processes (e.g., a face recognition process, a voice recognition process, or other pattern matching process) between the data structures of the media file being presented and the other media objects in the collection.

Referring to FIG. 10A, in one illustrative embodiment, the media file of indexed, temporally-ordered data structures being presented by media manager 12 corresponds to video file sequence 73. Initially, media manager 12 may present the video file 73 to a user by displaying the first keyframe 76 of the video file sequence. In addition, media manager 12 may simultaneously present the text file annotation 82, which is linked directly to keyframe 76. Text file annotation 82 may be presented only during the presentation of keyframe 76, or it may be presented during the presentation of keyframe 76 and for some prescribed time after keyframe 76 has been presented. The text file annotation presentation period preferably is shorter than the presentation period of the entire video file sequence 73. In the illustrated embodiment, text file annotation file 82 is presented only when keyframe 76 is presented.

Referring to FIG. 10B, media manager 12 presents the frames of video file 73 to a user in sequence. When one or more media object links to a video frame are detected, media manager 12 presents to the user the linked media objects along with the corresponding linked video frame. In the illustrated embodiment, when video frame 78 is

displayed, media manager 12 also displays the text file annotation 84, which is linked directly to video frame 88. A user may jump from multimedia album page 104 to another page (e.g., another multimedia album page, a web page, or some other media object) by selecting hypertext link 86 that is associated with text file annotation 84.

Media manager 12 also may determine that another media object 80 in the collection is relevant to the context of video frame 78. For example, media manager 12 may instantiate a face recognition process to find a match between the person displayed in video frame 78 and the person displayed in media object 80. In the illustrated embodiment, the relevant media object identified by media manager 12 corresponds to high resolution still photograph 80; in other embodiments, however, the identified media object may correspond to a different media object. Media manager 12 preferably displays the identified media object 80 only as long as it is relevant to the context currently being presented (e.g., as long as the same person appears in the presented video frames). In addition to displaying media object 80, media manager 12 displays graphical representations of links to media objects that are linked to media object 80. In the illustrated embodiment, media manager 12 displays a link 106 to a another page for displaying a larger version of the high resolution still photograph, a link 108 to an Internet web page corresponding to the home page of the person displayed in the high resolution photograph, and a link 110 to an audio file relating to the person displayed in the high resolution photograph (e.g., a voice greeting recorded by the displayed person). A user may select one or more of the hyperlinks 106-110 to jump from multimedia album page 104 to the corresponding linked media objects.

As shown in FIG. 10C, when the context changes, media manager 12 discontinues the presentation of media objects 80, 84. When a new media object link is detected or when a new relevant media object is identified, media manager 12 presents a new media object 90 in the multimedia album page 104. In the illustrated embodiment, media object 90 corresponds to a video file that is represented by a keyframe 112 showing a person with a face that matches the person being displayed in a frame 92 of video file 73. Media manager 12 also displays graphical representations of links to media objects that are linked to media object 90. In the illustrated embodiment, media manager 12 displays a link 114 to a multimedia album page corresponding to media object 90, a link 116 to a window for presenting the video file 92, and a link 118 to an Internet web page corresponding to the home page of the person displayed in keyframe 112. A user may

select one or more of the hyperlinks 114-118 to jump from multimedia album page 104 to the corresponding linked media objects.

Thus, in these embodiments, during presentation of a selected media file of indexed, temporally-ordered data structures a user may navigate through a collection of associated media objects by activating the links that are displayed to the user in a way that is context-sensitive and temporally-referenced to the playback of the selected media file of indexed, temporally-ordered data structures. In this way, new or forgotten associations may be discovered while browsing through the collection of media objects.

BROWSING A MEDIA OBJECT CLUSTER HIERARCHY

In some embodiments, the relevance criteria used to select the media objects that will be presented contemporaneously with the selected media file may relate to a selected metadata similarity between media objects and the selected media file. The metadata similarity may correspond to low-level features (e.g., motion activity, texture or color content, and audio content) or high-level features (e.g., meta data, such as keywords and names; objects, such as persons, places and structures; and time-related information, such as playback length and media object creation date). One or more known media object processing techniques (e.g., pattern recognition techniques, voice recognition techniques, color histogram-based techniques, and automatic pan/zoom motion characterization processing techniques) may be used to compare media objects to the selected media file in accordance with the selected metadata similarity criterion. In response to a user's indication to organize media objects in accordance with a selected metadata similarity criterion, media manager 12 selects a group of media objects having the highest metadata similarity to a selected media file. Media manager 12 then contemporaneously presents the selected media file and the media objects of the selected group.

As shown in FIG. 11, in these embodiments, the media objects of the selected group may be presented simultaneously around a selected media file 120. Alternatively, as shown in FIG. 12, the media objects of the selected group may be presented in the form of a media object stack 122. Individual media objects within the media object stack may be selected for comparison with the selected media file 120 by, for example, a pointer 124. In the illustrated embodiment, each of the selected media objects corresponds to a video file; in other embodiments, however, one or more of the selected media objects may be non-video media objects. A user may select a metadata similarity criterion in accordance with which media manager 12 will organize the presentation of

media objects by selecting one of the metadata similarity buttons contained in an "Organize by" window 126.

Referring to FIG. 13, in some embodiments, the media objects in the collection may be organized into a cluster hierarchy for each metadata similarity hierarchy. In this way, a user may efficiently browse through a large collection of media objects that are sorted in accordance with a context that is relevant to the user. In these embodiments, the media objects are grouped into clusters, each of which preferably contains a fixed number of media objects (e.g., the number of media objects that may be conveniently presented to a user at the same time). The media objects in each cluster preferably are all sufficiently similar to logically be placed in the same cluster. In general, the media objects of a cluster at one level in the hierarchy are represented by a single media object in a cluster at the next higher level. Thus, in the illustrated embodiment, media object 128 represents the cluster of media objects 130, 132, 134, 136, and media object 130 represents the cluster of media objects 138, 140, 142, 144. The representative media object may be determined statistically. For example, the media objects may be ordered in accordance with a selected context criterion, and the representative media object may correspond to the centroid or some other statistically-weighted average of a selected cluster of the ordered media objects.

In operation, a user may select (or anchor) a media object (e.g., video file 120) and organize the collection of media objects in accordance with a selected metadata similarity criteria (e.g., "Date" in FIGS. 11 and 12). In response, media manager 12 selects a group of media objects having the highest metadata similarity to the anchored media object (e.g., media objects that were created nearest in time to the date the anchored media file was created). The selected group of media objects are presented to the user either simultaneously, as shown in FIG. 11, or as a video card stack, as shown in FIG. 12. The user may browse through the presented media objects (e.g., playback one or more of the video files) or the user may jump to another presentation view. For example, the user may navigate through the media objects linked to video file 73, or the user may drill down into a media object cluster represented by one of the high level media objects in the selected group that is currently being presented to the user.

In some embodiments, a user may navigate through the collection of media objects without anchoring a particular media object. For example, in response to a user's selection of a desired metadata similarity criterion, media manager 12 may automatically

organize the collection of media objects into a browsable metadata similarity hierarchy. In one illustrative example, a user may select to organize the media object collection in terms of color content. In response, media manager 12 performs a color histogram analysis (e.g., RGB or YUV color histogram analysis) to generate a cluster hierarchy in which each lower level cluster is represented by a media object in a higher level cluster. Media manager 12 presents the highest level of the media object cluster hierarchy for browsing by the user. In this example, the highest level media objects may represent respective media object clusters having different respective color profiles (e.g., predominantly green, predominantly red, etc.). The user may then may direct media manager 12 to perform one or more of the following actions with respect to a selected one of the presented media objects: show the current scene; show the next scene; show the previous scene; show a slide show summary of the media object; show a photo album presentation of the media object; show a slide show presentation of the media object; navigate into a cluster represented by the selected media object; and navigate out of a cluster containing the selected media object. The user may drill down through the entire cluster hierarchy, or the user may select another metadata similarity criterion on which to organize the media object collection. At any stage during this process, the user may anchor a particular media object and organize the media object collection around the anchored media object in accordance with one or more selected metadata similarity criteria.

The systems and methods described herein are not limited to any particular hardware or software configuration, but rather they may be implemented in any computing or processing environment, including in digital electronic circuitry or in computer hardware, firmware or software. These systems and methods may be implemented, in part, in a computer program product tangibly embodied in a machine-readable storage device for execution by a computer processor. In some embodiments, these systems and methods preferably are implemented in a high level procedural or object oriented programming language; however, the algorithms may be implemented in assembly or machine language, if desired. In any case, the programming language may be a compiled or interpreted language. The media object management methods described herein may be performed by a computer processor executing instructions organized, e.g., into program modules to carry out these methods by operating on input data and generating output. Suitable processors include, e.g., both general and special purpose

microprocessors. Generally, a processor receives instructions and data from a read-only memory and/or a random access memory. Storage devices suitable for tangibly embodying computer program instructions include all forms of non-volatile memory, including, e.g., semiconductor memory devices, such as EPROM, EEPROM, and flash memory devices; magnetic disks such as internal hard disks and removable disks; magneto-optical disks; and CD-ROM. Any of the foregoing technologies may be supplemented by or incorporated in specially-designed ASICs (application-specific integrated circuits).

Other embodiments are within the scope of the claims.

WHAT IS CLAIMED IS:

1. A method of presenting a collection of media objects, comprising:

selecting a group of media objects from the collection based upon media object relevance to one or more data structures of a selected media file (66) of indexed, temporally-ordered data structures (68); and

contemporaneously presenting the selected media file (66) and the media objects of the selected group.

2. The method of claim 1, wherein media object relevance corresponds to browsable links between media objects of the selected group.

3. The method of claim 2, wherein the selected group consists of media objects linked directly to one or more data structures (68) of the selected media file (66).

4. The method of claim 3, wherein contemporaneously presenting comprises presenting the data structures (68) of the selected media file (66) in sequence and presenting each media object of the selected group in timed relationship with a respective linked data structure (68) of the selected media file (66).

5. The method of claim 4, wherein presentation of a media object of the selected group is initiated when a respective linked data structure (68) of the selected media file (66) is presented.

6. The method of claim 2, wherein presenting a given media object further comprises displaying one or more links from the given media object to respective linked media objects in the collection.

7. The method of claim 1, wherein media object relevance corresponds to a selected metadata similarity between media objects and the one or more data structures (68) of the selected media file (66).

8. A system for presenting a collection of media objects, comprising a media manager (12) operable to:

select a group of media objects from the collection based upon media object relevance to one or more data structures of a selected media file (66) of indexed, temporally-ordered data structures (68); and

contemporaneously present the selected media file (66) and the media objects of the selected group.

9. A method of presenting a collection of media objects, comprising:

grouping media objects (130-144) in the collection into multiple clusters based upon one or more media object relevance criteria;

arranging the media object clusters into a hierarchy of two or more levels; and

contemporaneously presenting a selected media object cluster.

10. The method of claim 9, wherein grouping media objects (130-144) into clusters comprises assigning media objects (130-144) to clusters based upon degree of mutual metadata similarity.

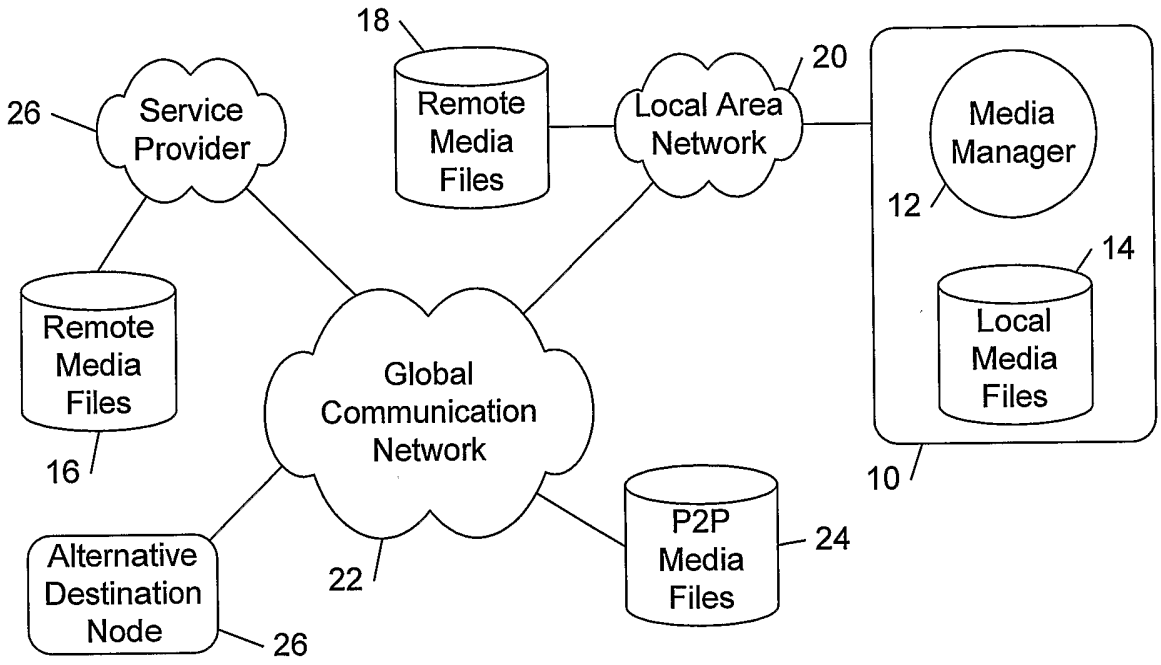

18
Remote Media Files
20
Local Area Network
Media Manager
12
14
Local Media Files
10
26
Service Provider
Remote Media Files
16
Global Communication Network
22
P2P Media Files
24
Alternative Destination Node
26

FIG. 1

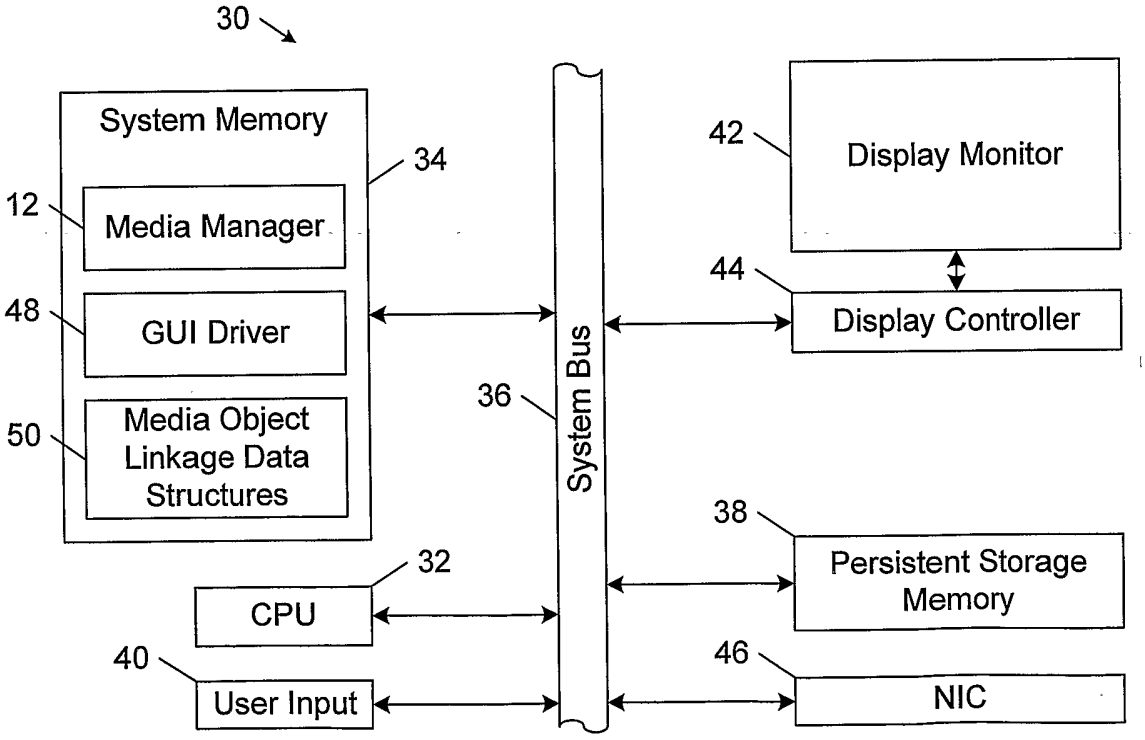

30
System Memory
34
12
Media Manager
48
GUI Driver
50
Media Object Linkage Data Structures
36
System Bus
42
Display Monitor
44
Display Controller
38
Persistent Storage Memory
32
CPU
40
User Input
46
NIC

FIG. 2

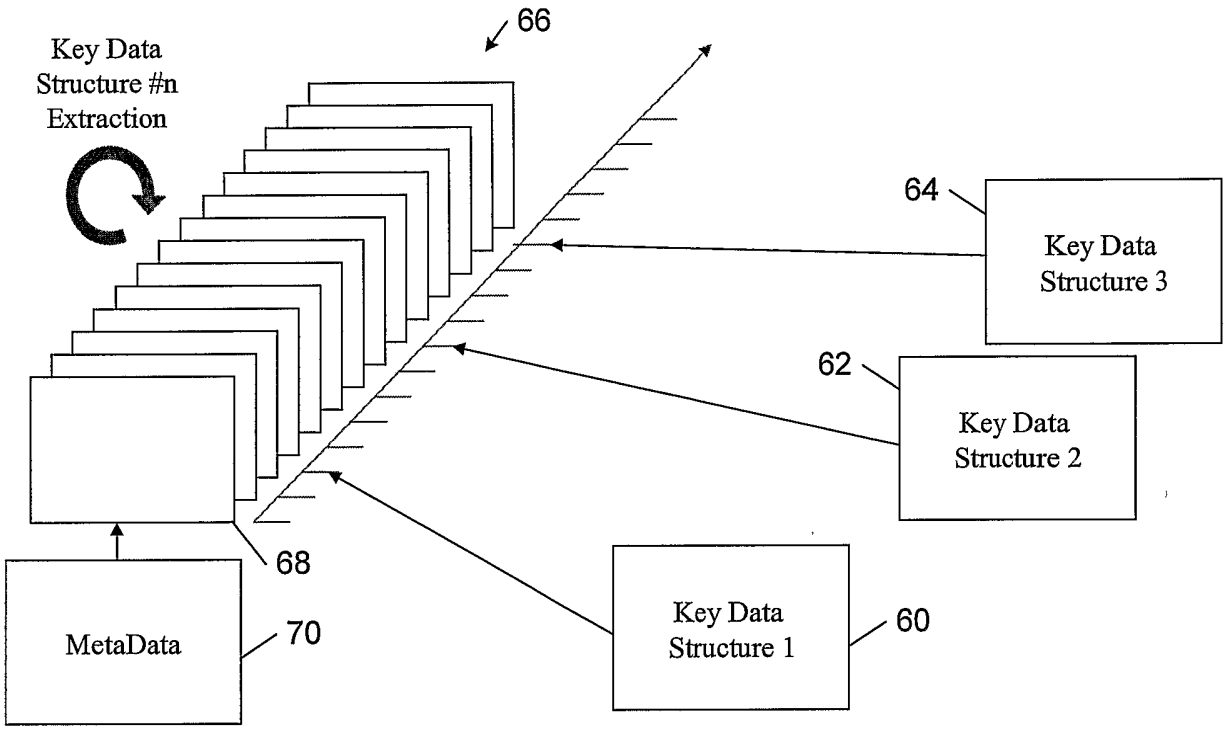

66
Key Data
Structure #n
Extraction
64
Key Data
Structure 3
62
Key Data
Structure 2
68
MetaData
70
Key Data
Structure 1
60

FIG. 3

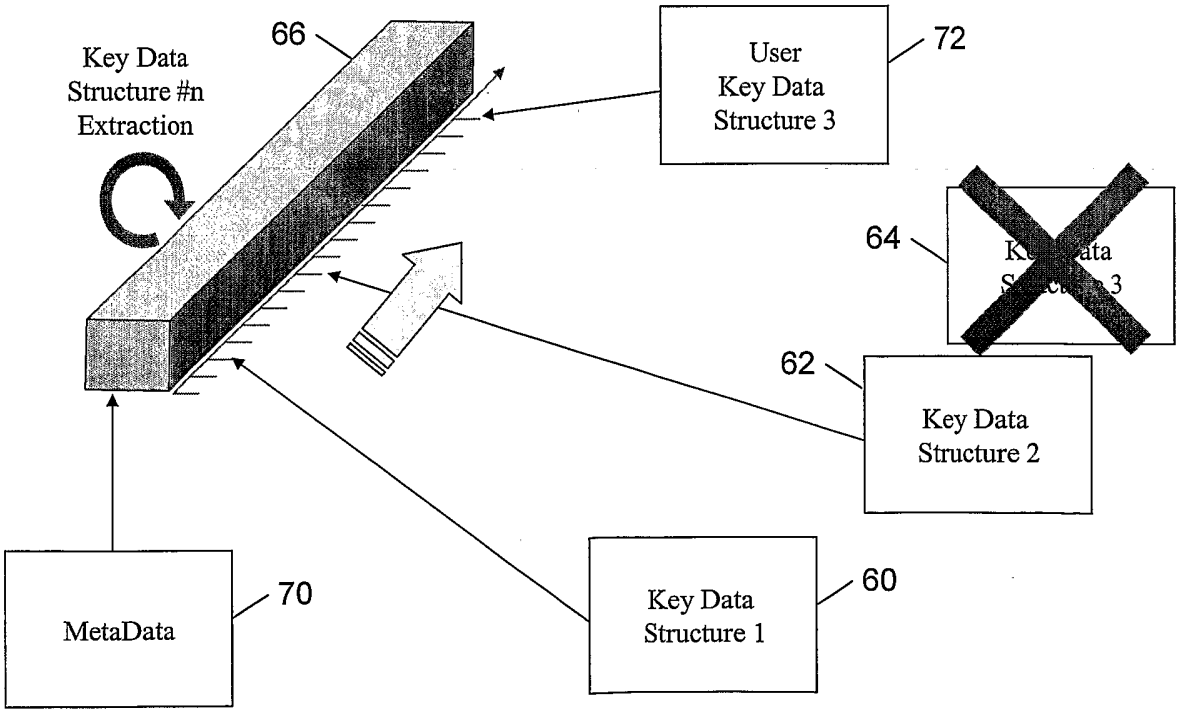

66
Key Data
Structure #n
Extraction
User
Key Data
Structure 3
72
64
62
Key Data
Structure 2
MetaData
70
Key Data
Structure 1
60

FIG. 4

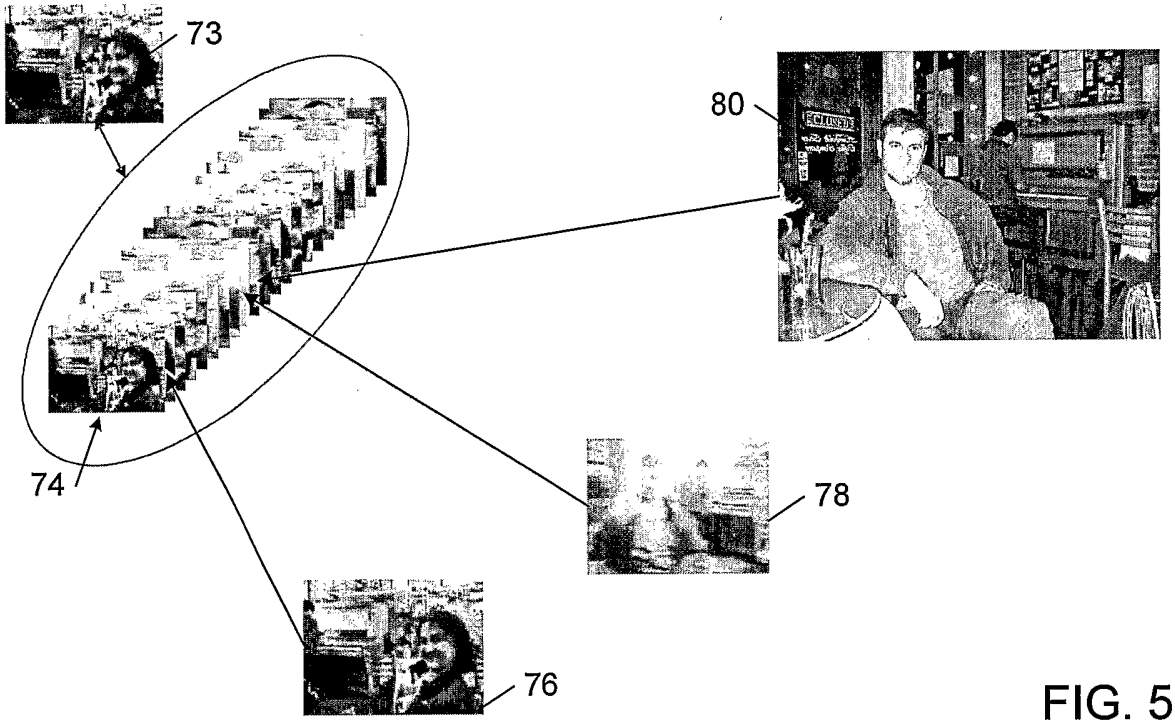

73
80
74
78
76

FIG. 5

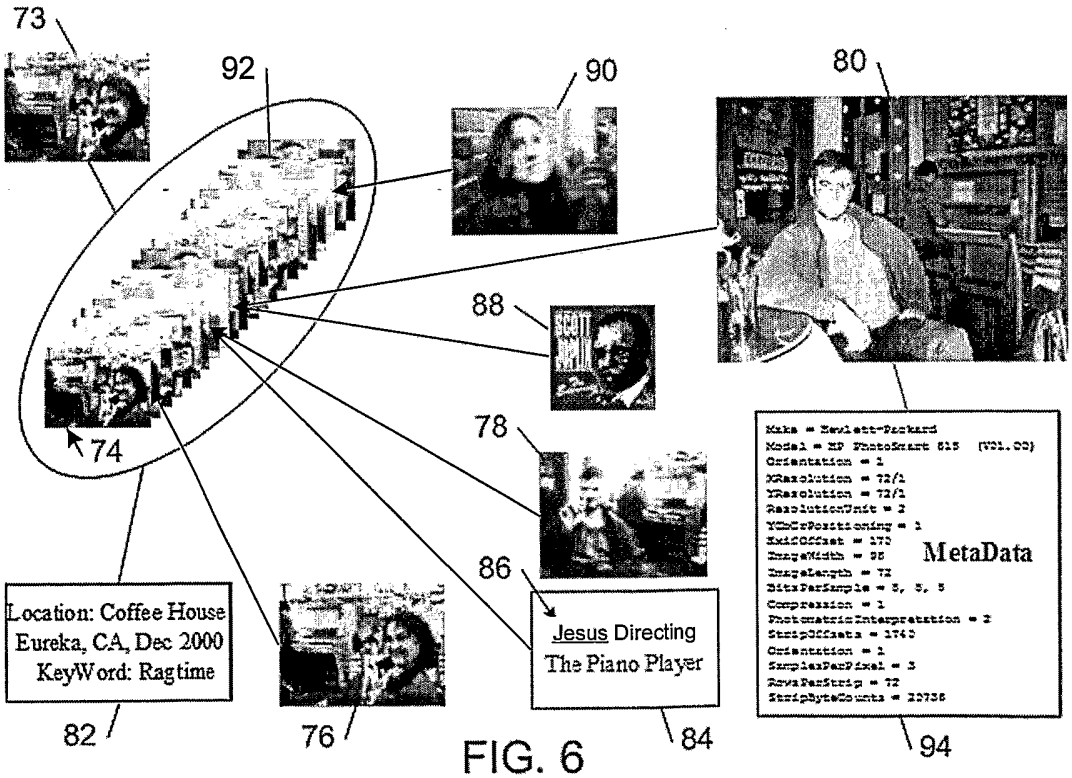

73
92
90
80
88
78
74
86
Location: Coffee House
Eureka, CA, Dec 2000
KeyWord: Ragtime
Jesus Directing
The Piano Player
MetaData
82
76
84
94

FIG. 6

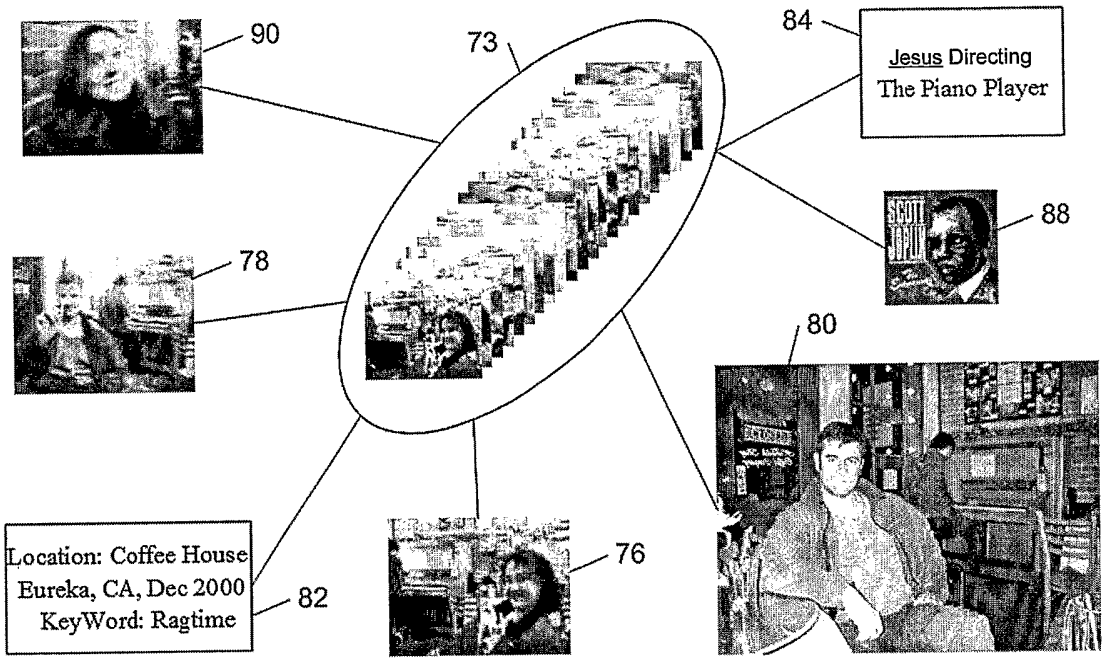

90
73
84
Jesus Directing
The Piano Player
88
78
80
Location: Coffee House
Eureka, CA, Dec 2000
KeyWord: Ragtime
82
76

FIG. 7A

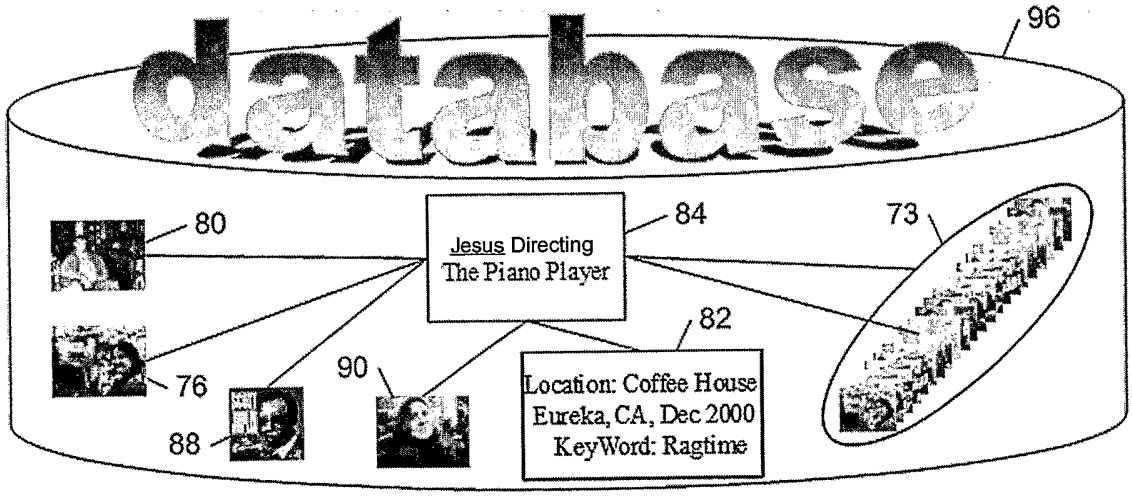

database
96
80
84
Jesus Directing
The Piano Player
73
82
76
90
88
Location: Coffee House
Eureka, CA, Dec 2000
KeyWord: Ragtime

FIG. 7B

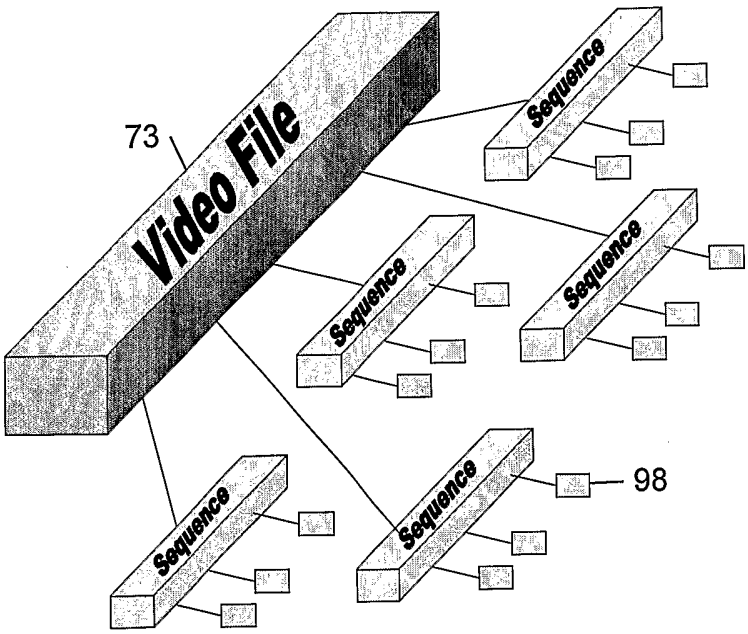

73
Video File
Sequence
Sequence
Sequence
Sequence
Sequence
Sequence
98

FIG. 8A

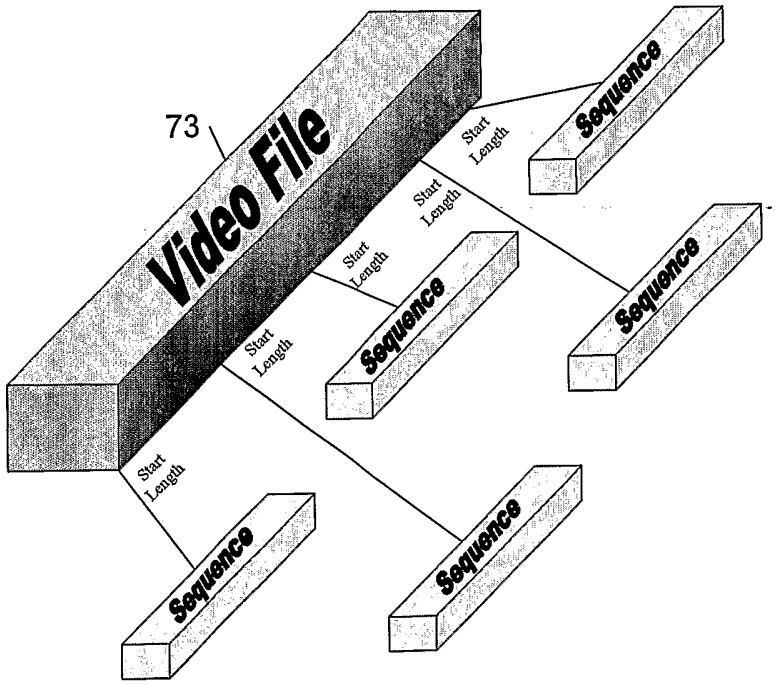

73
Video File
Start
Length
Sequence
Start
Length
Start
Length
Sequence
Sequence
Start
Length
Start
Length
Sequence
Sequence

FIG. 8B

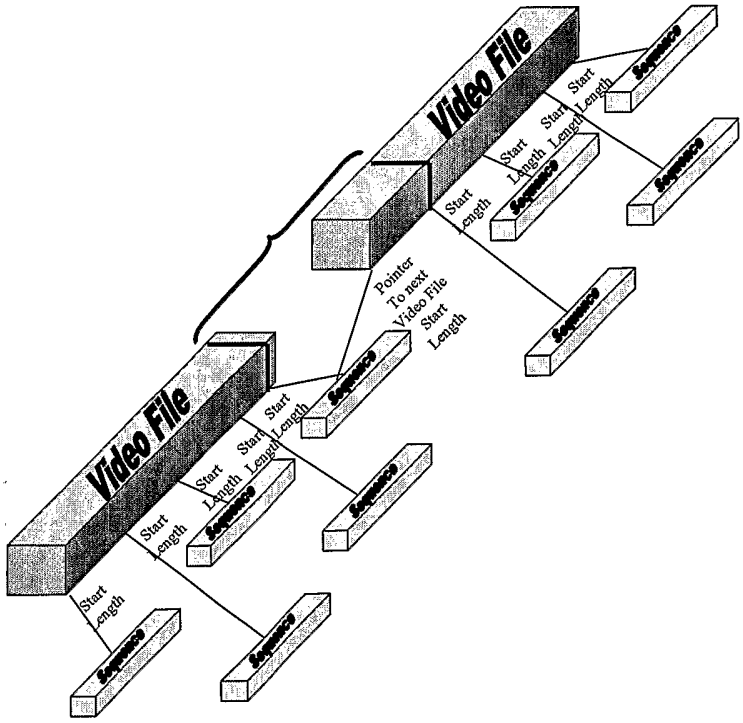

Video File
Video File
Pointer
To next
Video File
Start
Length
Start
Length
Sequence

FIG. 8C

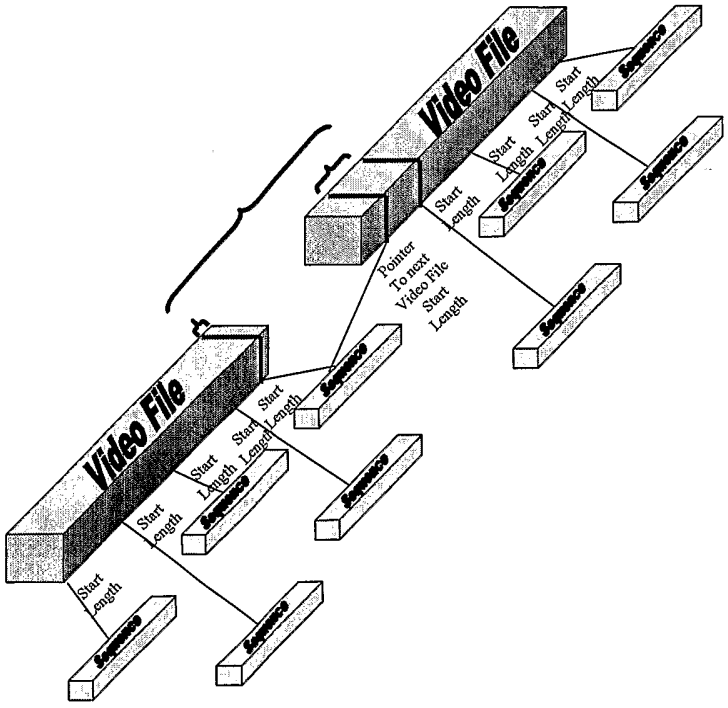

Video File
Video File
Pointer
To next
Video File
Start
Length
Start
Length
Sequence

FIG. 8D

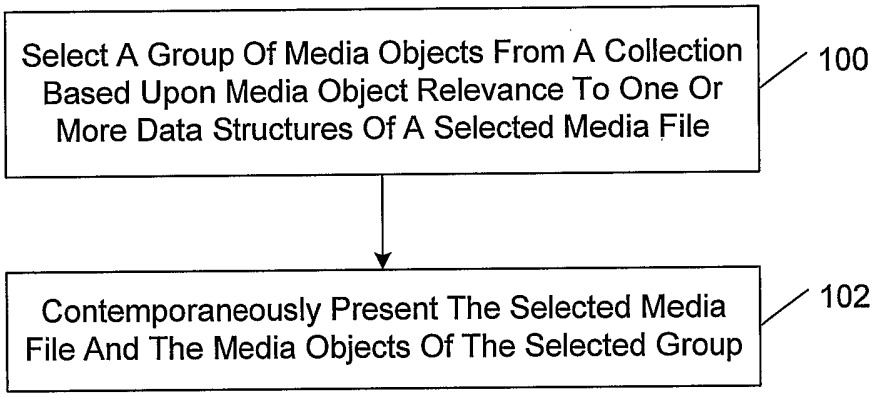

Select A Group Of Media Objects From A Collection Based Upon Media Object Relevance To One Or More Data Structures Of A Selected Media File
100
Contemporaneously Present The Selected Media File And The Media Objects Of The Selected Group
102

FIG. 9

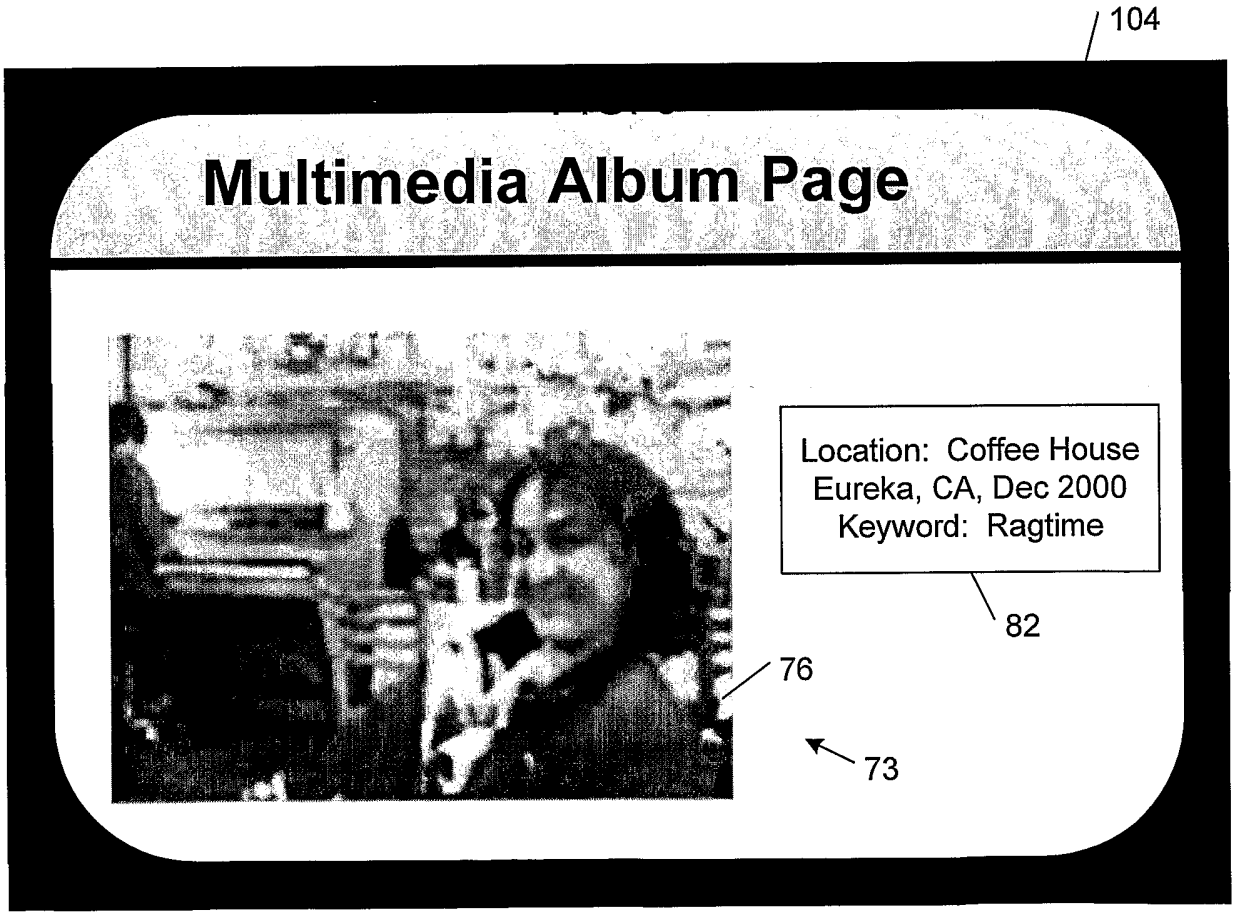

104
Multimedia Album Page
Location: Coffee House
Eureka, CA, Dec 2000
Keyword: Ragtime
82
76
73

FIG. 10A

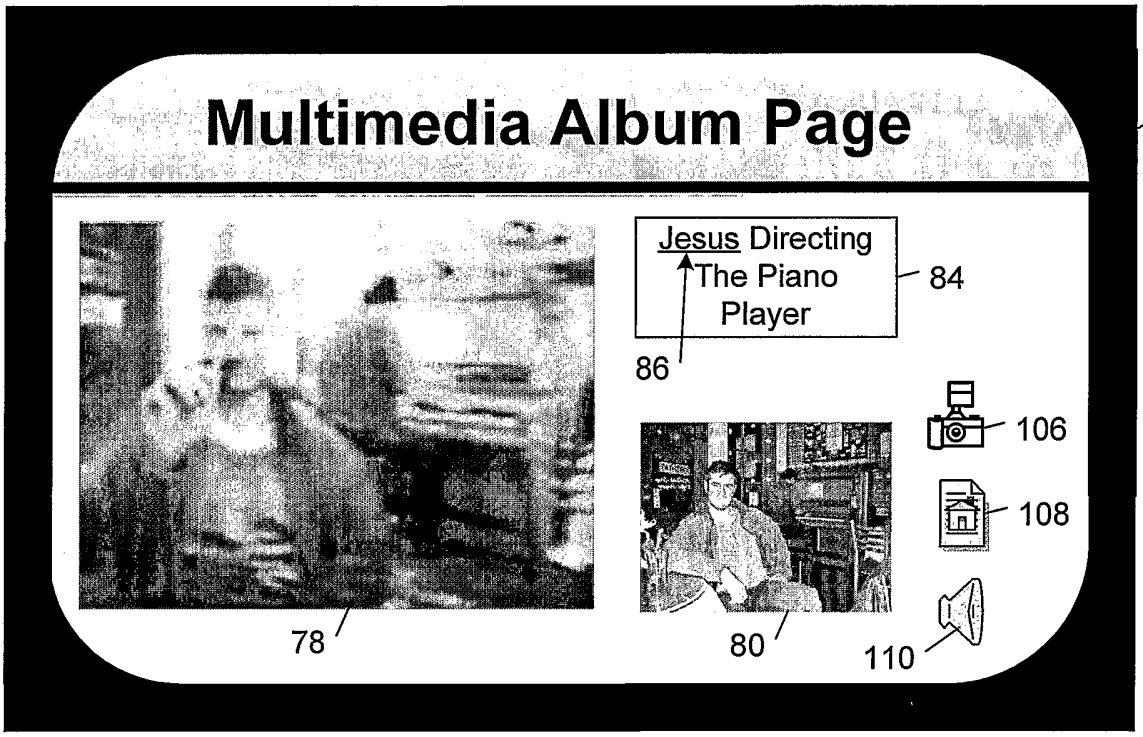


FIG. 10B

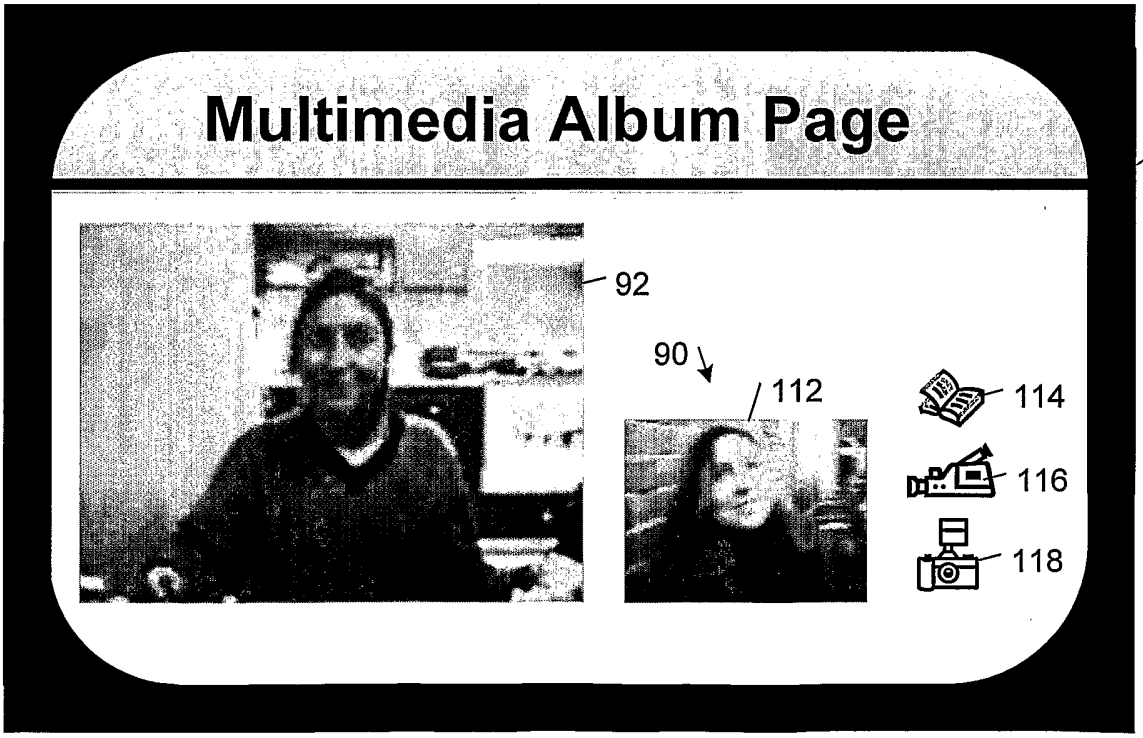


FIG. 10C

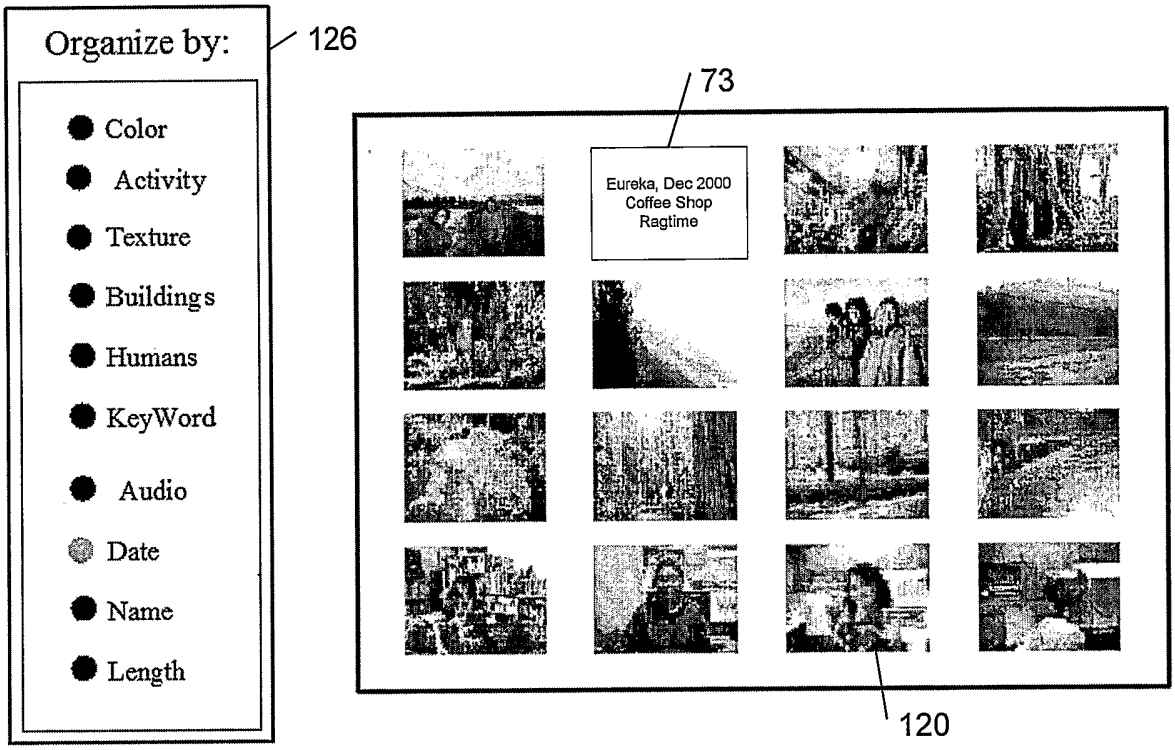

Organize by:
126
Color
Activity
Texture
Buildings
Humans
KeyWord
Audio
Date
Name
Length
73
Eureka, Dec 2000
Coffee Shop
Ragtime
120

FIG. 11

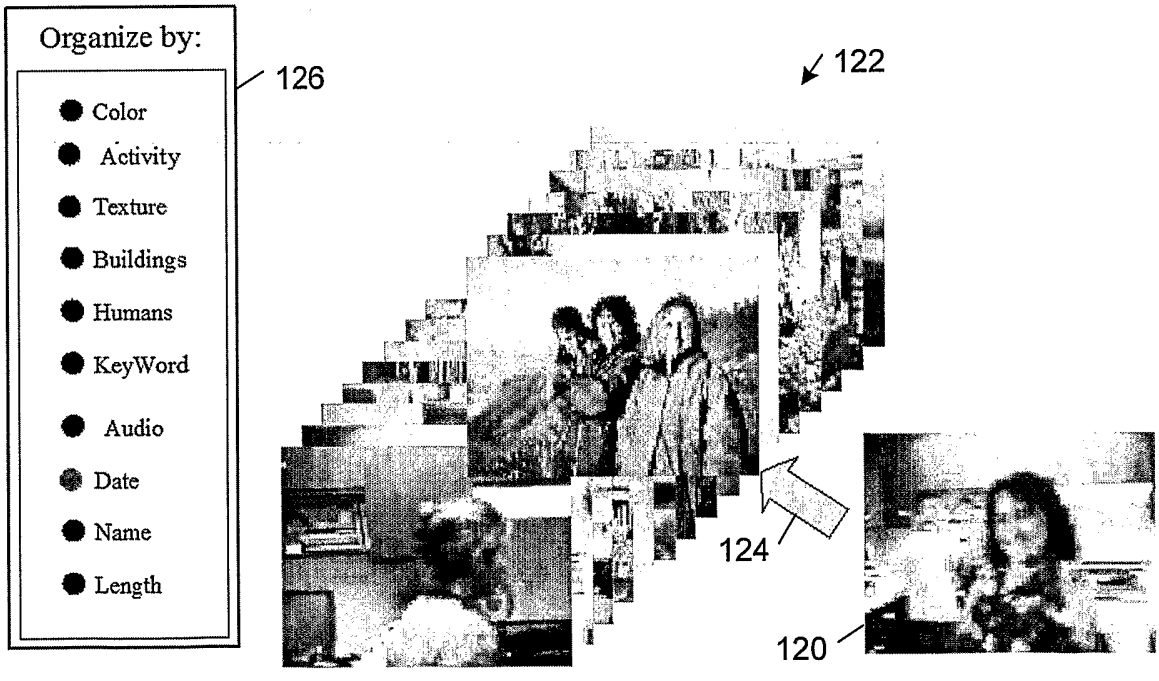

Organize by:
126
Color
Activity
Texture
Buildings
Humans
KeyWord
Audio
Date
Name
Length
122
124
120

FIG. 12

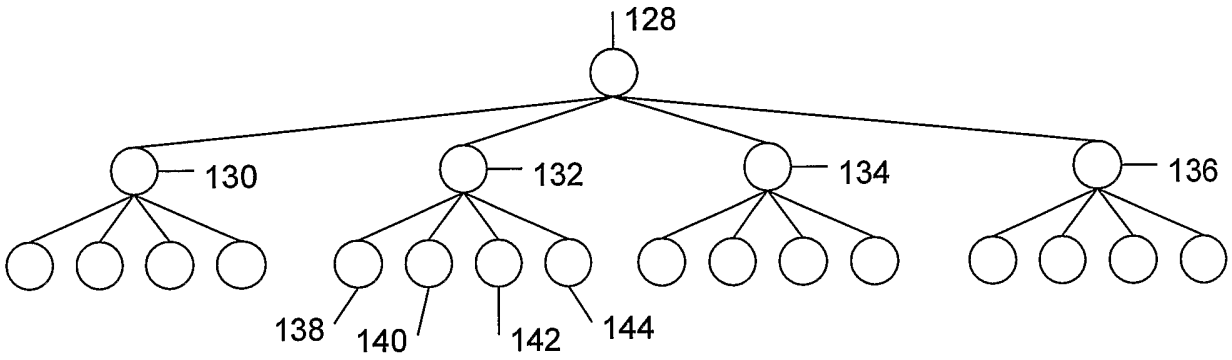


FIG. 13