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

(51) International Patent Classification⁷: G06K 9/00

(21) International Application Number: PCT/US01/03379

(22) International Filing Date: 1 February 2001 (01.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/180,364 4 February 2000 (04.02.2000) US
09/525,865 15 March 2000 (15.03.2000) US

(71) Applicant (for all designated States except US): DIGI-MARC CORPORATION [US/US]; 19801 SW 72nd Avenue, Suite 250, Tualatin, OR 97062 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): DAVIS, Bruce, L. [US/US]; 15599 Village Drive, Lake Oswego, OR 97034 (US). RODRIGUEZ, Tony, F. [US/US]; 3104 NE 31st Avenue, Portland, OR 97212 (US). RHOADS, Geoffrey, B. [US/US]; 2961 SW Turner Road, West Linn, OR 97068 (US).

(74) Agent: MEYER, Joel, R.; Digimarc Corporation, 19801 S.W. 72nd Avenue, Suite 250, Tualatin, OR 97062 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, 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, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— with international search 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: INTEGRATING DIGITAL WATERMARKS IN MULTIMEDIA CONTENT
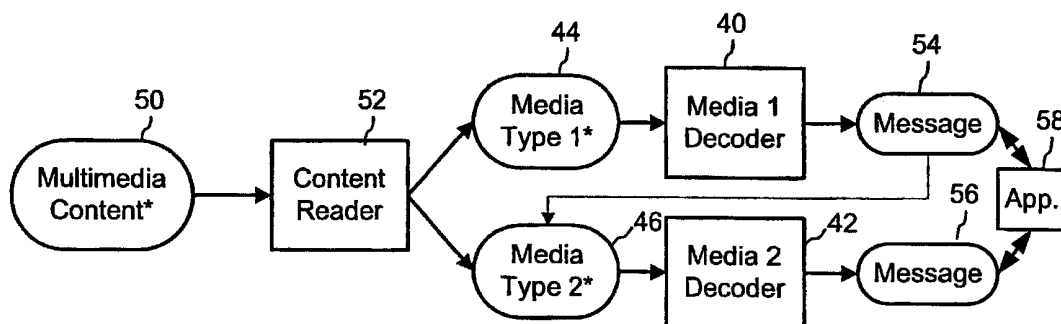
(57) Abstract: A method for decoding auxiliary data from media signals in multimedia content decodes watermarks (40 and 42) from different media signals (44 and 46) and uses the watermarks to control processing of the multimedia content (58). A copy control method decodes a watermark from one of the media signals in multimedia content, and uses the watermark to control processing of the multimedia content. Another method uses a watermark decoded from a media signal to decode metadata associated with the media signal. Finaly, another method forms a key for decoding data from at least first and second watermarks extracted from first and second media signals.

-1-

Integrating Digital Watermarks in Multimedia Content

Technical Field

The invention relates to digital watermarking, and more specifically relates to
applications of digital watermarks in multimedia data.

Background and Summary

Digital watermarking is a process for modifying media content to embed a machine-
readable code into the data content. The data may be modified such that the embedded code is
imperceptible or nearly imperceptible to the user, yet may be detected through an automated
detection process. Most commonly, digital watermarking is applied to media such as images,
audio signals, and video signals. However, it may also be applied to other types of data,
including documents (e.g., through line, word or character shifting), software, multi-
dimensional graphics models, and surface textures of objects.

Digital watermarking systems have two primary components: an embedding
component that embeds the watermark in the media content, and a reading component that
detects and reads the embedded watermark. The embedding component embeds a watermark
pattern by altering data samples of the media content in the spatial or frequency domains. The
reading component analyzes target content to detect whether a watermark pattern is present. In
applications where the watermark encodes information, the reader extracts this information
from the detected watermark.

Recently, digital watermarks have been used in applications for encoding auxiliary data
in video, audio and still images. Despite the pervasiveness of multimedia content, such
applications generally focus on ways to embed and detect watermarks in a single media type.

One aspect of the invention is a method for decoding auxiliary data in multimedia
content with two or more media signals of different media types. This method decodes
watermarks in the media signals, uses the watermarks from the different media signals to
control processing of the multimedia content. There are many applications of this method.
One application is to use the watermark in one media signal to locate the watermark in another
media signal. This is applicable to movies where a watermark in one media signal, such as the
audio or video track, is used to locate the watermark in another media signal.

The watermark messages from different media signals may be combined for a variety
of applications. One such application is to control processing of the multimedia signal. For
example, the combined message can be used to control playback, copying or recording of the
multimedia content.

-2-

Another aspect of the invention is a method for copy control of multimedia content where a watermark from one media signal is used to control processing of the multimedia content. An audio watermark may be used to control processing of the video signal in a movie, or a video watermark may be used to control processing of the audio signal in the movie.

5   Another aspect of the invention is a method for watermark decoding where a watermark decoded from a first media signal of a first media type is used to decoding a second media signal. The first and second media signals may be of the same or different types. Also, they may be part of the same composite media signal, such as an audio or video sequence. The term, "composite," refers to a collection of media signals, which may be temporal portions

10   (e.g., time frames in audio or video), or spatial portions (e.g., blocks of pixels in an image or video frame) of a visual, audio, or audio visual work. As an example, the first media signal may be an audio or video frame (or frames) in an audio or video sequence and the second media signal may be subsequent frames in the same sequence.

This method may be used in a variety of applications. The watermark in the first media

15   signal may be used to de-scramble, decrypt, or decompress the second media signal. In addition, the watermark in the first media signal may be used to decode a different watermark from the second signal.

Another aspect of the invention is a method that uses.a watermark decoded from a first media signal of a first media type to decode metadata associated with the first media signal.

20   The watermark may be used to locate the metadata, which may be hidden for security purposes. The metadata located from the watermark may be located on the same storage medium that includes the first media signal. For example, the metadata may be located on portable storage device, such as flash memory, a magnetic memory device (e.g.., tape or disk), or an optical memory device (e.g., CD, DVD, minidisk, etc.). The metadata may be located in a file header

25   or some other place (e.g., encoded in the disk wobble).

There are a variety of applications of the watermark in this context. It may carry a key to decrypt, decompress, descramble, or locate the metadata. The metadata, in turn, may be used to control processing of the media signal in a computer or consumer electronic device. For example, it may be used to control usage rights, playback, recording, copying, transfer, etc.

30   Yet another aspect of the invention is a method that decodes first and second watermarks and forms a key for decoding data from the first and second watermarks. The watermarks may be decoded from the same or different media signals. For example, the watermarks may be decoded from media signals from the same composite signal. They may be derived from different types of media signals, such as the audio and video tracks of a movie.

35   Alternatively, they may be derived from different parts of the same type of media signal, such

-3-

as an audio sequence, video sequence, or image. The watermarks may be extracted from a signal or signals stored in a storage device, such as a portable storage device (e.g., optical or magnetic disk or tape, flash memory, etc.).

The key formed from the watermarks may be used for a variety of applications. It may
5 be used as a watermark key to decode a watermark from a media signal. It may be used as a decryption or de-scrambling key. Also, it may be used as a decompression key (e.g., a parameter used to decompress a media signal).

Further features of the invention will become apparent with reference to the following detailed description and accompanying drawings.
10

Brief Description of the Drawings

Fig. 1 is a diagram of a watermark encoder system for encoding watermarks in multimedia content.

Fig. 2 is a diagram of a watermark decoder system for multimedia data.
15   Fig. 3 is a diagram of a watermark decoder system where watermark detectors for different media types collaborate.

Fig. 4 is a diagram of a watermark decoder system where watermark readers for different media types collaborate.

Fig. 5 illustrates an operating environment for implementations of the invention.
20   Detailed Description

1.0 Introduction

The following sections describe applications for integrating watermarks in multimedia data. In general, these applications exploit some level of interaction between watermarks and/or metadata associated with two or more different media types. The types of media
25 supported in a given implementation vary with the application, and may include, for example, audio (e.g., speech, music, etc.), video, images, graphical models, etc.

The initial sections describe ways to integrate watermark embedder and detector systems in multimedia data. These techniques may be applied to many different applications, including, for example, copy protection, content authentication, binding media content with
30 external data or machine instructions, etc.

Later sections discuss specific application scenarios.

-4-

2.0 Integration of Watermarks and Metadata of Different Data Types

2.1 Defining Multimedia

The term, multimedia, as used in this document, refers to any data that has a collection of two or more different media types. One example is a movie, which has an audio and video track. Other examples include multimedia collections that are packaged together on a storage device, such as optical or magnetic storage device. For example, media signals such as still images, music, graphical models and videos may be packaged on a portable storage device such as CD, DVD, tape, or flash memory card. Different media signals may be played back concurrently, such as the video and audio tracks of a movie, or may be played independently.

2.2 Levels of Integration of Watermark Systems

The extent of integration of watermark systems for different media types ranges from a low level of integration, where watermark decoders operate independently on different media types, to a high level of integration, where the decoders functionally interact. At a low level of integration, the watermark systems for different media types operate on their respective media types independently, yet there is some relationship between the auxiliary data embedded in each type. At a high level of integration, components of the watermark detectors and readers share information and assist each other to perform their respective functions.

Fig. 1 illustrates an encoder system for embedding messages into a multimedia content with two or more media types. One example of multimedia content is a movie with video and audio tracks. For the purpose of illustrating the system, the following sections use a movie as an example of multimedia content. Similar methods may be implemented for other forms of multimedia content, such as combinations of three-dimensional/two-dimensional graphics and animation, audio, video, and still images.

In the encoder system shown in Fig. 1, there is a watermark encoder 20, 22 for each media type. Each encoder may embed a message 24, 26 into the corresponding media type 28, 30 in the native domain of the signal (e.g., a spatial or temporal domain) or in some transform domain (e.g., frequency coefficients). The result is multimedia content 32 having watermarks in different media types. The multimedia content 32 may be packaged and distributed on a portable storage device, such as a CD, DVD, flash memory, or delivered electronically from one machine or device to another in a file or streaming format.

There are a variety of ways to integrate the encoder functions. One way is to use a unified key that controls how a given message or set of messages are encoded and located within the respective media types. Another way is to insert a common message component in

-5-

two or more different media types. Yet another way is to make a message inserted in one media type dependent on the content of one or more other media types. For example, attributes of an image may be extracted from the image and encoded into an audio track, and similarly, attributes of an audio track may be extracted and encoded in an image. Finally, the message in one media type may be used to control the processing of another media type. For example, copy control flags in a movie's audio track may be used to control copying of the movie's video track or the movie; and, copy control flags in the video track may be used to control copying of the audio track or the movie.

The following sub-sections describe various scenarios for integrating watermarks in different media types from the perspective of the decoder.

2.2.1 Auxiliary Data Embedded in Different Media Types

Fig. 2 depicts a framework for low level integration, where watermark decoders 40, 42 for different media types 44, 46 operate independently, yet an application 58 uses the auxiliary data associated with each of the media types. The auxiliary data may be encoded in a watermark message within a media signal or may be located in metadata accompanying the media signal (e.g., on the storage device and/or within a header of a file or data packet encapsulating the media). The multimedia content 50 is annotated with a "*" to reflect that it may not be identical to the original version of the content (e.g., the content shown at item 32, Fig. 1) at the time of encoding due to intentional or unintentional corruption (e.g., filtering, compression, geometric or temporal transforms, analog to digital, and digital to analog conversion). A content reader 52 receives the multimedia data and identifies the distinct media types within it. The functionality of the content reader may be built into a watermark decoder or provided by a separate computer program or device. In the example of a movie, the content reader identifies the audio and video tracks.

Watermark decoders for each media type operate on their respective media data. In extracting the watermark from the signal domain in which the embedder inserted it, the decoder functions compliment the embedder functions. In many applications, the media types may be coded in a standard or proprietary format. In the example of a movie, both the audio and video tracks are typically compressed (e.g., using some lossy transform domain compression codec like MPEG). The watermark decoders may operate on compressed, partially compressed or uncompressed data. For example, the decoders may operate on frequency coefficients in the compressed image, video or audio data. As shown in Fig. 2, the decoders 40, 42 operate independently on corresponding media types to extract messages 54, 56 from watermarks in each media type.

-6-

In the low level integration scenario of Fig. 2, an application 58 uses the messages from different media types to process the multimedia content. The application is a device, software process, or combination of a device and software. The specific nature of this processing depends on the requirements of a particular application. In some cases, the message embedded

5   in one media type references content of another type (e.g., link 60 from message 54 to media type 2). For example, text sub-titles in a movie may be embedded in the audio track, and may be linked to specific frames of video in the video track via frame identifiers, such as frame numbers or addresses. The application, in this scenario, controls the playback by superimposing the text sub-titles on the linked frames.

10   In many applications, it may be useful to insert a link in one media type to content of another media type within the multimedia data. For example, one might want to link a still image or a video texture to a graphical model. Then, a graphics rendering application may use the link to determine which image (or video) to map to the surface of a graphical model. As another example, one might link an audio clip to an image, graphical model or other media

15   object. When instructed to render the image, model or other media object, the rendering application then uses the link to also initiate playback of the linked audio clip, and optionally, to synchronize playback of the linking media signal with the signal linked by the watermark. For example, the video watermark could specify which audio clip to play and when to initiate playback of parts of the audio clip. Stated more generally, the embedded link from one media

20   type to another may be used by the rendering application to control the relationship between the linked media objects during playback and to control the playback process.

The media signals within multimedia content can be linked together through watermarks and embedded with control information and metadata that is used to control playback. The entire script for controlling playback of a multimedia file or collection may be

25   embedded in watermarks in the media signals. For example, a user could initiate playback by clicking on an image from the multimedia content. In response, the rendering application extracts control instructions, links, and/or metadata to determine how to playback video, audio, animation and other media signals in the multimedia content. The rendering application can execute a script embedded in a watermark or linked via a reference in the watermark (e.g., a

30   watermark message includes a pointer to, or an index or address of a script program stored elsewhere). The watermark message may also specify the order of playback, either by including a script, or linking to a script that contains this ordering. Several media signals may be tied together in a playback sequence via a linked list structure where watermarks embedded in the media signals reference the next media signal to be played back (as well as media signals

35   to be played back concurrently). Each media signal may link to another one by providing a

-7-

media signal identifier in the watermark message, such as an address, pointer, index, name of media title, etc.

As the rendering application plays back multimedia content, it can also display metadata about the media signals (e.g., the content owner, a description of the content, time and location of creation, etc.). The watermark messages embedded in the media signals can either include this metadata or link to it. In addition, the watermark messages may include instructions (or a link to instructions) for indicating how and when to display metadata. The metadata need not be in text form. For example, metadata may be in the form of speech output (via a text to speech synthesis system), a pre-recorded audio clip, video clip, or animation.

To embed a variety of different information, instructions and links into the media signals within multimedia content, the embedder can locate watermark messages in different temporal portions (e.g., time multiplex different messages) of a time varying signal like audio or video. Similarly, the embedder can locate different watermark messages in different spatial portions of images, graphical models, or video frames. Finally, the embedder can locate different watermark messages in different transform domains (e.g., Discrete Fourier Transform, Discrete Cosine Transform, Wavelet transform, etc.) of image or audio signals.

The following sub-sections describe additional application scenarios.

2.2.1.1 Copy Protection

In a copy protection application, the messages embedded in each media type convey information to the application specifying how it may use the content. For example, each message may provide copy control flags specifying "copy once", "copy no more", "copy freely", and "copy never." These flags indicate whether the application may copy the media type or the multimedia content as a whole, and if so, how many times it may copy the pertinent content.

The application collects the copy control flags from the different media types and determines the extent to which it may copy the content or selected media types within it.

2.2.1.2 Ownership Management

In multimedia content, each media type may be owned by different entities. The messages embedded in the content may contain an owner identifier or link to an owner. An ownership management application can then collect the ownership information, either from each of the messages in each media type, or by requesting this information by following the link to the owner. For example, the link may be associated with an external database that provides this information. The application may use the link to query a local database for the

-8-

information. Alternatively, the application may use the link to query a remote database via a wire, wireless, or combination of wire and wireless connections to a remote database on a communication network (e.g., the Internet). One or more intermediate processing stages may be invoked to convert the link into a query to the remote database. For example, the link may be a unique number, index or address that cross-references the URL of a database server on the Internet.

2.2.1.3 Media Authentication

An authentication application may use watermark messages and/or metadata to authenticate media signals within the multimedia content. One or more of the media signals in multimedia content may be tampered with. Multimedia content poses an additional problem because media signals may be swapped into the content in place of the original signals. For example, in a video used as evidence, one might swap in a fake audio clip or remove a portion of the audio track. One way to authenticate the media signals is to extract features from them, hash the features, and insert the hashed features into the watermark messages of one or more of the media signals at encoding time.

To verify authenticity, the application at the decoder side repeats the process of extracting the features from the received media types (e.g., 44, 46), hashing these features, and then comparing the new hash with the hash extracted from the watermark message or messages. The objective of the hash is to create a content dependent parameter that may be inserted into a watermark message, or in some cases, in metadata associated with a media signal. The hash is not necessary if the size of the extracted features is such that they fit within a message.

Examples of features in images include the location of identifiable objects (such as the location of eyes and noses of human subjects), the shape of objects (e.g., a binary mask or chain code of an object in an image), the inertia of an image, a low pass filtering of an image, the Most Significant Bit of every pixel in a selected color plane (luminance, chrominance, Red, Green, Blue, etc.).

Examples of features in audio include the temporal location of certain aural attributes (e.g., a transition from quiet to high intensity, sharp transitions in spectral energy, etc.), a low pass filter of an audio clip, etc.

Features from one media type may be inserted into a watermark or the metadata of another media type. Alternatively, they may be combined and inserted in one or more of the media types in a watermark embedded in a watermark of the media signal or its metadata.

An additional level of security may be added using public key encryption techniques to create a digital signature that identifies the source of the multimedia content. Some examples

-9-

of public key cryptography include RSA, DES, IDEA (International Data Encryption

Algorithm), skipjack, discrete log systems (e.g., El Gamal Cipher), elliptic curve systems,

cellular automata, etc. Public key cryptography systems employ a private and public key. The

private key is kept secret, and the public key is distributed to users. To digitally sign a

5   message, the originator of the message encrypts the message with his private key. The private

key is uniquely associated with the originator. Those users having a public key verify that the

message has originated from the holder of the private key by using the public key to decrypt the

message.

10

2.2.2 Integrating Watermark Detection Processes

          Another way to integrate processing of media types is to integrate watermark detectors

for different media types. One function of some watermark detectors is to determine the

orientation and strength of a watermark within a host media signal. The orientation may

15   provide the watermark location, and possibly other orientation parameters like warp (e.g., an

affine or non-linear warp, temporal and/or spatial), scale, rotation, shear, etc. As the media

content is subjected to various transformations, the watermark orientation and strength may

change. Watermark detectors use attributes of the watermark signal to identify its location and

orientation within a host signal. In multimedia content where different media signals are

20   watermarked, detectors for the respective media signals can assist each other by sharing

information about the orientation and/or strength of a watermark in the media signals. While

the watermarks in different media types may be transformed in different ways, the orientation

information found in one media signal might help locate a watermark in a different media

signal.

25   Fig. 3 depicts a watermark decoder framework in which the watermark detectors for

different media types collaborate. Each detector 70, 72 operates on its respective media type

74, 76, yet the detectors share information. The detectors determine the presence, and in some

cases, the strength and/or orientation of a watermark in a host media signal. In some

applications, such as authentication, the detector identifies portions of the media signal that

30   have a valid watermark signal, and portions where the watermark has been degraded (e.g., the

watermark is no longer detectable, or its strength is reduced). Depending on the nature of the

host signal, these portions may be temporal portions (e.g., a time segment within an audio

signal where the watermark is missing or degraded) or spatial portions (e.g., groups of pixels in

an image where the watermark is missing or degraded). The absence of a watermark signal, or

35   a degraded watermark signal, may evidence that the host signal has been tampered with.

-10-

In applications where the watermark carries a message, each detector may invoke a watermark reader 78, 80 to extract a message from the watermark. In some cases, the reader uses the orientation to locate and read the watermark. The strength of the watermark signal may also be used to give signal samples more or less weight in message decoding. Preferably,

5   each reader should be able to read a watermark message 82, 84 from a media signal without requiring the original, un-watermarked media signal.

One example of integrated detection is a scheme where watermark detectors operate on respective media types concurrently and share orientation parameters. To illustrate the scheme, consider the example of a movie that has a watermarked audio and video track. While video

10   and audio are distinct media signals in the content delivery and storage formats, the video and audio tracks are carefully synchronized so that the audio closely tracks the movement of actors' mouths and other motion depicted in the video. The embedding scheme places audio watermarks within a specified temporal range of the video watermarks. Because the video and audio tracks need to be temporally synchronized to avoid noticeable artifacts during playback,

15   the temporal locations of the audio and video watermarks are likely to remain within a predictable temporal distance in their respective host signals. As such, the watermark detectors can take advantage of the temporal relationship of the watermarks in different media types to facilitate detection.

The location of a watermark detected in one media signal can provide information

20   about the location of a watermark yet to be detected in another media signal. For example, when the video watermark detector finds a watermark in a video frame (e.g., an I frame in MPEG video), it signals the other detector, passing information about the temporal location of the video watermark. Leveraging the temporal relationship between the video and audio watermarks, the audio watermark detector confines its search for an audio watermark to a

25   specified temporal range in the audio signal relative to the location of the corresponding video watermark in the video signal.

In this scenario, the audio watermark detector may provide similar information to the video watermark detector to help it identify the frame or sequence of frames to be analyzed for a video watermark.

30   Another example is a scheme where one watermark detector operates on a media type, and then passes orientation parameters to a detector of another media type. This scheme reduces the complexity of the second detector because it uses the orientation parameters extracted from a first media type to assist computation of the orientation in another media type. Applying this scheme to the previous example of a movie, the watermark decoder method

35   reduces the complexity of the audio detector by confining its search to a specified range defined

-11-

relative to the location of a video watermark. This is a simpler case than the previous example in the sense that the orientation information flows solely from a first detector to a second one. The second detector searches in a confined space around the location specified by the other detector, and does not have to pass orientation information to the other detector.

2.2.3.1 Applications of Integrated Watermark Detectors

As in the previous sections, there are a variety of applications for watermark systems with integrated detectors. The watermarks may be used to encode data or links to external data or other media signals within the multimedia content.

The watermarks may also be used to encode authentication information. In the movie example, the watermarks in one media type can reference one or more watermarks in another media type. For example, if an audio detector does not find an audio watermark designated by the video watermark to be in a specified range within the audio signal, then it can mark that specified range as being corrupted. Similarly, the video detector can authenticate video frames based on presence or absence of video watermarks designated by audio watermarks.

In copy control applications for mixed media like movies, integrated detectors can be used to locate audio and video watermarks carrying copy control flags. If the audio or the video tracks have been tampered with or transformed in a way that removes or degrades the watermarks, then a copy control application can take the appropriate action in response to detecting the absence of a watermark or a degraded watermark. The actions triggered in response may include, for example, preventing copying, recording, playback, etc.

2.2.4 Integrating Watermark Message Reading of Different Media Types

Fig. 4 illustrates yet another scenario for integrating watermark decoders where the watermark readers for different media types collaborate. In this scheme, watermark detectors 100, 102 for different media types 104, 106 operate independently (or collaborate as described above) to detect the presence, and optionally the orientation, of watermarks in their respective media types. Watermark readers 108, 110 then extract messages from the detected watermarks. The watermark readers pool the message data 112 that they extract from the different media types.

Then, a message decoder 114 attempts to decode the pooled message data. The message decoder may perform various error correction decoding operations, such as Reed Solomon, BCH, Turbo, Convolution operations. In cases where the watermark embedder uses spread spectrum modulation to spread raw message bits in the host media signal into chips, the

-12-

message decoder may perform the inverse of a spread spectrum modulation function to convert spread spectrum chip values back to raw message values.

The result of the decoding operations provides information about the media signals. Depending on the application and implementation, the decoded message 116 can be interpreted in different ways. For example, in some cases, to generate a valid decoded message (as indicated by an error detection process such as a CRC or parity check), watermark message data from each media signal must be valid. In other cases, the decoded message may specify which media signals have valid messages, and which do not.

2.2.4.1 Applications

Like the other scenarios described above, the scheme for integrating watermark readers of different media types can be applied to many applications, including data embedding and linking, content authentication, broadcast monitoring, copy control, etc. This scheme is particularly suited for content authentication and copy control because it can be used to indicate content tampering and to disable various operations, such as copying, playback, recording, etc. For example, it can be used in a copy control scheme for content with audio and video tracks. Each track contains watermark messages that must be detected and converted to the raw message data 112 before the decoder 114 can decode a valid message. Thus, valid copy control information in both the video and audio tracks must be present before a valid copy control message 116 will be produced. A player can then process the multimedia content based on the control information in the valid copy control message. Alternatively, the content can be prevented from being passed into a player or other application or device if a valid control message is not generated.

2.2.5 Using Watermark Messages to Store Keys to Other Watermarks or Metadata

The watermark message in one media signal may be used to specify a key of a watermark in another media signal. In this scenario, the watermark reader for one media type supplies the watermark decoder for another media type with the key. This key may specify the location of the watermark as well as information about how to extract the watermark from another media signal, and information to decode or decrypt the watermark message.

The watermark message in a media signal may also specify a key to access other metadata on the storage device of the media signal. For example, the message may specify a key to decrypt or decode metadata on the storage device, such as metadata in a header file or encoded within tracks of a CD or DVD (e.g., encoded within the disk wobble). The key may also specify the location of the associated metadata.

-13-

2.2.5.1 Applications

The scheme described in the previous section may be used in many applications, including those discussed previously. This scheme is particularly suited for content

5   authentication and copy protection. In order to authenticate the content, each of the media signals in multimedia content need to have valid watermarks. The watermark in one media signal cannot be located without extracting a key from a watermark in another media signal.

In copy protection applications, the decoding system would need to find the watermarks in each of the media signals before enabling certain actions (e.g., playback,

10   recording, copying, etc.).

2.3 Using Watermark Data in One Media Type to Control Playback of Another Media Type

For some applications, it is not necessary that each media signal in multimedia content have a watermark. For example, a watermark in one media signal could provide the desired

15   functionality for the entire content, or for selected portions of the content. For example, in copy protection applications for movies, a watermark in the audio track could be used to encode copy control flags to control copying, playback, or recording of audio and/or video tracks.

20   2.4 Using Watermark Data in Conjunction with Other Data or Applications

The watermark message data can be used in conjunction with other data or applications to control processing of the multimedia or single media content. Using any of the scenarios above, for example, a decoder can extract a message that is used to control further media processing.

25   One example is where the watermark message is used as a necessary key for decoding or decrypting the media content. For example, the watermark message may contain necessary bits for decompressing (e.g., MPEG decoding) of the media signal or signals within the content (audio, video or both). Examples of necessary bits are CRC bits that are required to reconstruct coded video or audio data. This technique is particularly useful when the message is derived

30   from watermark messages embedded in different media signals. In a movie copy control application, for instance, the decoder would have to generate a valid message based on decoding the raw message information from audio and video watermark messages before allowing playback, recording, etc. In this case, the embedder would spread the necessary control information into watermark messages inserted in the audio and video tracks. For

-14-

example, watermark messages in audio or video frames include decompression parameters or descrambling keys to decompress or descramble subsequent audio or video frames.

The same approach can be implemented by embedding other forms of control data in one or more watermark messages in different media signals. Another example is a decryption key that is necessary to decrypt other media signals within the content, or other portions of the same media signal. Watermark messages in audio or video frames may include decryption keys to decrypt subsequent frames. One watermark message may include a key, or a portion of a key, needed to decrypt or unscramble other signal portions or other watermark messages. In the case where the watermark message includes only a portion of a key (e.g., one parameter in a key comprising two or more parameters), the other portion may be constructed by extracting another component of the key from another watermark message (in the same or different media signals) or from other metadata (e.g., in the disk wobble, the header file of MPEG content, etc.).

Another form of control data is region data that indicates that a particular media signal may only be played when the region data of the media signal and the player match. A similar region data scheme is understood to be implemented in the Content Scrambling System currently used for DVDs. The region data can be embedded in one or more watermarks in the same or different media signals. By placing this information in different media signals, the decoder must be able to extract consistent region data from watermarks in each of the media signals as a pre-requisite to further use of the content. Then, assuming all of the region data creates a valid region data message, then the copy control application would control playback based on whether the region data decoded from the watermarks (and/or metadata of the different media signals) matches the region data of the player.

3.0 Implementation of Watermark Encoders and Decoders

The state of watermark encoders and decoders for audio, video and still images is quite advanced. Some examples of watermark systems for multimedia data include US Patent Nos. 5,862,260, 5,930,369, and US patent application no. 09/503,881. Examples of watermark systems targeted to audio signals include 5,945,932, 5,940,135, 6,005,501, and 5,828,325. Other watermark systems are described in 5,940,429, 5,613,004, 5,889,868, WO 99/45707, WO 99/45706, WO 99/45705, and WO 98/54897.

Examples of watermark systems used in copy control are: WO 00/04688, WO 00/04712, WO 00/04727, and WO 99/65240. These documents include examples where a copy protection scheme uses watermark data and metadata to control processing of a media signal.

-15-

Watermark systems that operate on compressed content include: 5,687,191; and WO 00/04722.

These watermark systems may be used to implement the scenarios described above.

5   3.1 Location of the Watermark Decoder

The watermark decoder may be implemented in one or more components. The location of these components varies depending on the application. For multimedia content on portable memory devices like DVDs or CDs, the decoder may be implemented in the drive hardware or in an interface to the drive hardware. Alternatively, the decoder may be located in an

10   application program or device. One example is a media codec, like an MPEG decoder. If the media signals are compressed, the detector may have to implement at least portions of the codec. For example, if the watermark is coded in frequency coefficients in MPEG video and audio, the decoder system may include an MPEG parser and dequantizer to identify the media signals (audio and video signals) and extract the coefficients from each of the media signals.

15   Placing the watermark decoder in the media codec, such as the MPEG codec, saves resources because many of the resources used for decoding the media signals may also be used for detecting and reading the watermarks.

3.2 Operating Environment

20   Figure 5 illustrates an example of a computer system that may serve as an operating environment for software implementations of the watermarking systems described above. The encoder and decoder implementations as well as related media codecs and applications may be implemented in C/C++ and are portable to many different computer systems. Components may also be implemented in hardware devices or in a combination of hardware and software

25   components. These components may be installed in a computing device such as a Personal Digital Assistant, Personal Computer, Hand-held media player, media players (DVD players, CD players, etc.) or implemented in a hardware module such as an integrated circuit module, ASIC, etc. Fig. 5 generally depicts one example of an operating environment for encoder and decoder systems.

30   The computer system shown in Fig. 5 includes a computer 1220, including a processing unit 1221, a system memory 1222, and a system bus 1223 that interconnects various system components including the system memory to the processing unit 1221.

The system bus may comprise any of several types of bus structures including a memory bus or memory controller, a peripheral bus, and a local bus using a bus architecture

35   such as PCI, VESA, Microchannel (MCA), ISA and EISA, to name a few.

-16-

The system memory includes read only memory (ROM) 1224 and random access memory (RAM) 1225. A basic input/output system 1226 (BIOS), containing the basic routines that help to transfer information between elements within the computer 1220, such as during start-up, is stored in ROM 1224.

5   The computer 1220 further includes a hard disk drive 1227, a magnetic disk drive 1228, e.g., to read from or write to a removable disk 1229, and an optical disk drive 1230, e.g., for reading a CD-ROM or DVD disk 1231 or to read from or write to other optical media. The hard disk drive 1227, magnetic disk drive 1228, and optical disk drive 1230 are connected to the system bus 1223 by a hard disk drive interface 1232, a magnetic disk drive interface 1233,

10   and an optical drive interface 1234, respectively. The drives and their associated computer-readable media provide nonvolatile storage of data, data structures, computer-executable instructions (program code such as dynamic link libraries, and executable files), etc. for the computer 1220.

       Although the description of computer-readable media above refers to a hard disk, a

15   removable magnetic disk and an optical disk, it can also include other types of media that are readable by a computer, such as magnetic cassettes, flash memory cards, digital video disks, and the like.

       A number of program modules may be stored in the drives and RAM 1225, including an operating system 1235, one or more application programs 1236, other program modules

20   1237, and program data 1238.

       A user may enter commands and information into the personal computer 1220 through a keyboard 1240 and pointing device, such as a mouse 1242. Other input devices may include a microphone, sound card, radio or television tuner, joystick, game pad, satellite dish, digital camera, scanner, or the like. A digital camera or scanner 43 may be used to capture the target

25   image for the detection process described above. The camera and scanner are each connected to the computer via a standard interface 44. Currently, there are digital cameras designed to interface with a Universal Serial Bus (USB), Peripheral Component Interconnect (PCI), and parallel port interface. Two emerging standard peripheral interfaces for cameras include USB2 and 1394 (also known as firewire and iLink).

30   In addition to a camera or scanner, watermarked images or video may be provided from other sources, such as a packaged media devices (e.g., CD, DVD, flash memory, etc), streaming media from a network connection, television tuner, etc. Similarly, watermarked audio may be provided from packaged devices, streaming media, radio tuner, etc.

-17-

These and other input devices are often connected to the processing unit 1221 through a port interface 1246 that is coupled to the system bus, either directly or indirectly. Examples of such interfaces include a serial port, parallel port, game port or universal serial bus (USB).

A monitor 1247 or other type of display device is also connected to the system bus 1223 via an interface, such as a video adapter 1248. In addition to the monitor, personal computers typically include other peripheral output devices (not shown), such as speakers and printers.

The computer 1220 operates in a networked environment using logical connections to one or more remote computers, such as a remote computer 1249. The remote computer 1249 may be a server, a router, a peer device or other common network node, and typically includes many or all of the elements described relative to the computer 1220, although only a memory storage device 1250 has been illustrated in Figure 5. The logical connections depicted in Figure 5 include a local area network (LAN) 1251 and a wide area network (WAN) 1252. Such networking environments are commonplace in offices, enterprise-wide computer networks, intranets and the Internet.

When used in a LAN networking environment, the computer 1220 is connected to the local network 1251 through a network interface or adapter 1253. When used in a WAN networking environment, the personal computer 1220 typically includes a modem 1254 or other means for establishing communications over the wide area network 1252, such as the Internet. The modem 1254, which may be internal or external, is connected to the system bus 1223 via the serial port interface 1246.

In a networked environment, program modules depicted relative to the personal computer 1220, or portions of them, may be stored in the remote memory storage device. The processes detailed above can be implemented in a distributed fashion, and as parallel processes. It will be appreciated that the network connections shown are exemplary and that other means of establishing a communications link between the computers may be used.

4.0 Relationship with Other Applications of Metadata

Watermarks can facilitate and cooperate with other applications that employ metadata of multimedia objects. As demonstrated above, this is particularly true in copy protection/control applications where the copy control information in the watermark and the metadata are used to control playback. The watermark message and metadata (in the MPEG file header or encoded in the disk wobble) can form components in a unified key that is a necessary prerequisite to playback or some other use of the content.

-18-

The watermarks in the media signals can each act as persistent links to metadata stored elsewhere, such as a metadata database server on the Internet or some other wire or wireless network. Applications for viewing and playing content can display metadata by extracting the link and querying a metadata database server to return the metadata (e.g., owner name, content description, sound or video annotation, etc.). The watermark decoder or an application program in communication with it can issue the query over the Internet using standard communication protocols like TCP/IP, database standards like ODBC, and metadata standards like XML. The query may be sent to a metadata router that maps the link to a metadata database server, which in turn, returns the metadata to the viewing application for display or playback to the user.

5.0 Concluding Remarks

The watermarking technology detailed herein can be employed in numerous diverse applications. See, e.g., the applications for watermarking detailed in commonly-owned patent 5,862,260, and copending applications 09/292,569, 60/134,782, 09/343,104, 09/473,396, 09/476,686, and 60/141,763.

Having described and illustrated the principles of the invention with reference to several specific embodiments, it will be recognized that the principles thereof can be implemented in other, different, forms.

To provide a comprehensive disclosure without unduly lengthening the specification, applicant incorporates by reference any patents and patent applications referenced above.

The particular combinations of elements and features in the above-detailed embodiments are exemplary only; the interchanging and substitution of these teachings with other teachings in this and the incorporated-by-reference patents/applications are also contemplated.

In view of the wide variety of embodiments to which the principles of the invention can be applied, it should be recognized that the detailed embodiment is illustrative only and should not be taken as limiting the scope of the invention. Rather, we claim as our invention all such embodiments as may come within the scope and spirit of the following claims, and equivalents thereto.

-19-

1.   A method for decoding auxiliary data in multimedia content with two or more media signals of different media types, comprising:

    decoding watermarks in the media signals; and

    using the watermarks from the different media signals to control processing of the multimedia content.

2.   The method of claim 1 wherein the watermark in one media signal is used to locate the watermark in another media signal.

3.   The method of claim 2 wherein the one media signal is a video track and the other is an audio track of a movie, and there is a temporal relationship between the video and audio track such that a watermark in one track is within a predefined temporal distance of a watermark in another track.

4.   The method of claim 1 wherein watermark messages from different signal types are combined.

5.   The method of claim 4 wherein the combined message is used to control playback, copying or recording of the multimedia content.

6.   A method for copy control of multimedia content with two or more media signals of different media types, comprising:

    decoding a watermark from one of the media signals; and

    using the watermark to control processing of the multimedia content.

7.   The method of claim 6 wherein a watermark in an audio signal is used to control processing of a video signal in the multimedia content.

8.   The method of claim 6 wherein a watermark in a video signal is used to control processing of an audio signal in the multimedia content.

9.   A method for watermark decoding comprising:

    decoding a watermark from a first media signal of a first media type;

    and decoding a second media signal using the watermark.

-20-

10. The method of claim 9 wherein the second media signal is a part of a composite media signal including both the first and second media signals.

11. The method of claim 9 wherein the first and second media signals are the same media type.

12. The method of claim 11 wherein the first and second media signals are parts of a composite media signal.

13. The method of claim 9 wherein the decoding comprises decompressing the second media signal.

14. The method of claim 9 wherein the decoding comprises de-scrambling or decrypting the second media signal.

15. The method of claim 9 wherein the decoding comprises using the watermark as a watermark key for decoding a different watermark from the second media signal.

16. The method of claim 15 wherein the watermark key specifies a location of the different watermark in the second media signal.

17. A method for watermark decoding comprising:
decoding a watermark from a first media signal of a first media type;
and decoding metadata associated with the first media signal using the watermark.

18. The method of claim 17 wherein the watermark specifies a location of metadata and decoding comprises using the watermark to locate the metadata.

19. The method of claim 18 wherein the metadata is located on a storage medium that also includes the first media signal.

20. The method of claim 19 wherein the storage medium is a portable storage device.

21. The method of claim 17 wherein the watermark includes a decryption key and the decoding comprises decrypting metadata with the decryption key.

22. A method of watermark decoding comprising:

decoding at least a first watermark from a first media signal;

decoding at least a second watermark from a second media signal; and

5   forming a key for decoding data from at least the first and second watermarks.

23. The method of claim 22 wherein the first and second media signals are of the same media type.

10   24. The method of claim 22 wherein the first and second media signals are of different media types.

25. The method of claim 22 wherein the first and second media signals are stored in the same storage medium.

15

26. The method of claim 25 wherein the storage media is a portable storage medium.

27. The method of claim 22 wherein the key comprises a watermark key used to decode a watermark from a media signal.

20

28. The method of claim 22 wherein the key comprises a decryption key used to decrypt a media signal.

29. The method of claim 22 wherein the key comprises a parameter used to
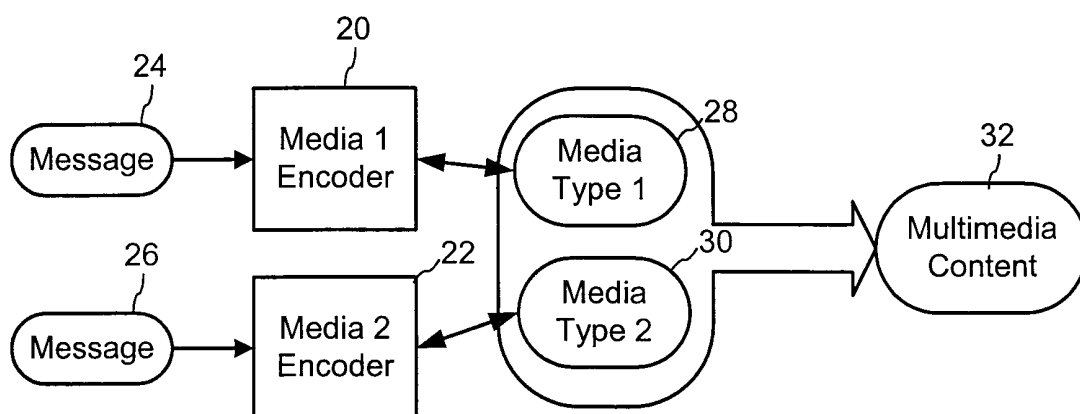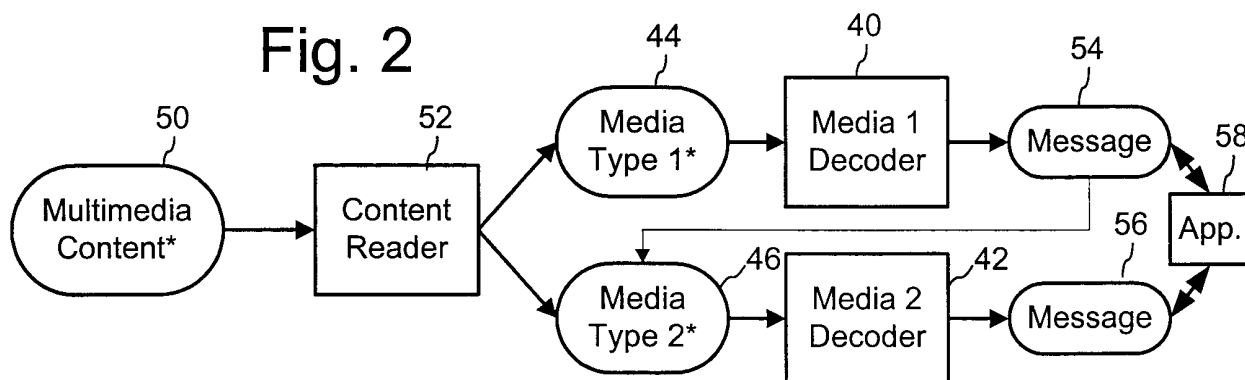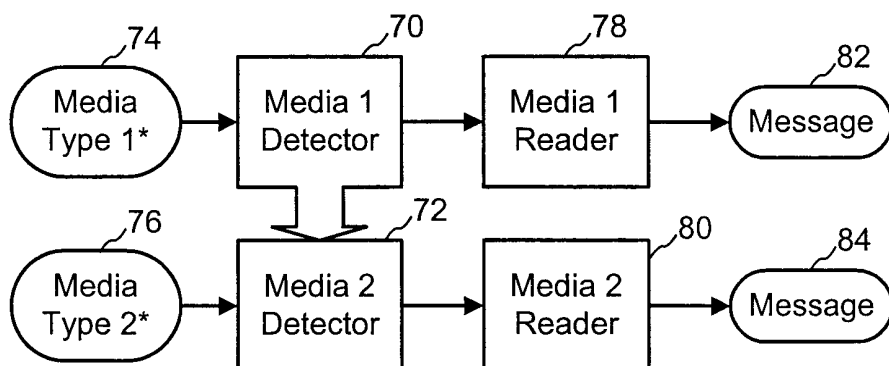
25   decompress a media signal.

Fig. 1

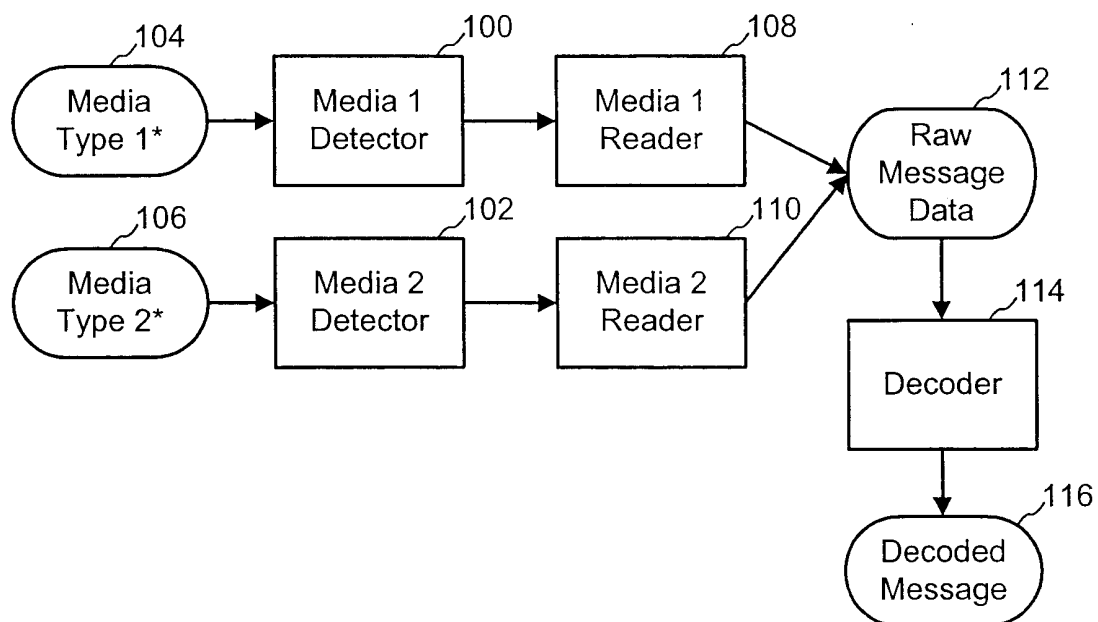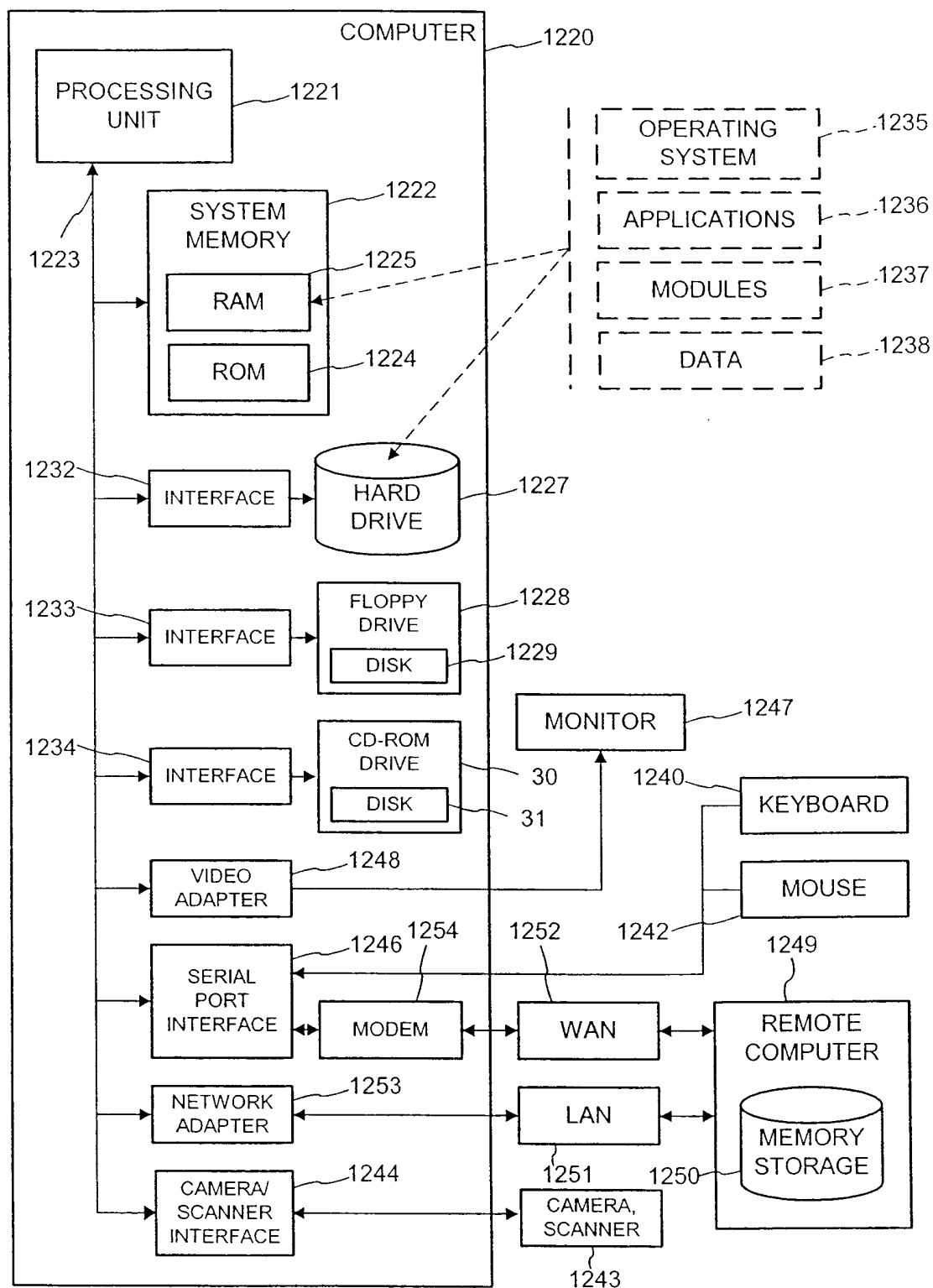Message (24) → Media 1 Encoder (20) → Media Type 1 (28)

Message (26) → Media 2 Encoder (22) → Media Type 2 (30)

Media Type 1 (28) + Media Type 2 (30) → Multimedia Content (32)

Fig. 2

Multimedia Content* (50) → Content Reader (52)

Media Type 1* (44) → Media 1 Decoder (40) → Message (54) ↔ App. (58)

Media Type 2* (46) → Media 2 Decoder (42) → Message (56) ↔ App. (58)

Fig. 3

Media Type 1* (74) → Media 1 Detector (70) → Media 1 Reader (78) → Message (82)

Media Type 2* (76) → Media 2 Detector (72) → Media 2 Reader (80) → Message (84)

Fig. 4

FIG. 5

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7)   :   G06K 9/00
US CL   :   382/100, 232

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : Please See Continuation Sheet

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	MASURI et al., "Video-Steganography: How to Secretly Embed a Signature in a Picture", PROCEEDINGS: Technological Strategies for Protecting Intellectual Property in the Networked Multimedia Environment, Vol. 1, Issue 1, January 1994, pp. 1-264.	1-2, 4-6, 9-16 and 22-29 ---------------------------- 3, 7-8 and 17-21
Y	COX et al., "Secure Spread Spectrum Watermarking for Images, Audio and Video", Proc. IEEE Southcon 96, June 1996, pp. 243-246.	3 and 7-8
Y	US 5,721,788 A (POWELL et al.) 24 February 1998 (24.02.1998), figures 1-5, and column 1, line 1 through column7, line 35.	1-29
Y	US 5,493,677 A (BALOGH et al.) 20 February 1996 (20.02.1996), figures 1-16, column 1, line 1 through column 18, line 30.	1-29

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

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance		
"E"	earlier application or patent published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
04 April 2001 (04.04.2001)	**27 APR 2001**
Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Jose L. Cousø
Facsimile No. (703)305-3230	Telephone No. (703) 305-8576

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

Continuation of B. FIELDS SEARCHED Item 1: 382/100, 112, 115, 232, 233, 236, 244, 248, 254, 280, 287; 358/85-86, 400, 402, 403, 407, 442; 379/94-98, 100, 105, 112-114, 118, 125; 380/3-5, 9, 16, 22, 23, 25, 49, 50, 54.