US 20090317061A1

(19) **United States**

(12) **Patent Application Publication**

(10) **Pub. No.: US 2009/0317061 A1**

(43) **Pub. Date: Dec. 24, 2009**

JUNG et al.

(54) **IMAGE GENERATING METHOD AND APPARATUS AND IMAGE PROCESSING METHOD AND APPARATUS**

(75) Inventors: **Kil-soo JUNG**, Osan-si (KR); **Hyun-kwon CHUNG**, Seoul (KR); **Dae-jong LEE**, Suwon-si (KR)

Correspondence Address:
STEIN MCEWEN, LLP
1400 EYE STREET, NW, SUITE 300
WASHINGTON, DC 20005 (US)

(73) Assignee: **Samsung Electronics Co., Ltd.**, Suwon-si (KR)

(21) Appl. No.: **12/489,758**

(22) Filed: **Jun. 23, 2009**

Related U.S. Application Data

(60) Provisional application No. 61/075,184, filed on Jun. 24, 2008.

(30) **Foreign Application Priority Data**

Sep. 17, 2008 (KR) 10-2008-0091269

Publication Classification

(51) **Int. Cl.**
H04N 5/91 (2006.01)

(52) **U.S. Cl.** **386/95**; 386/E05.001

(57) **ABSTRACT**

An image processing method and apparatus and an image generating method and apparatus, the image processing method to output a video data being a two-dimensional (2D) image as the 2D image or a three-dimensional (3D) image including: extracting information about the video data from metadata associated with the video data; and outputting the video data as the 2D image or the 3D image by using the extracted information about the video data.

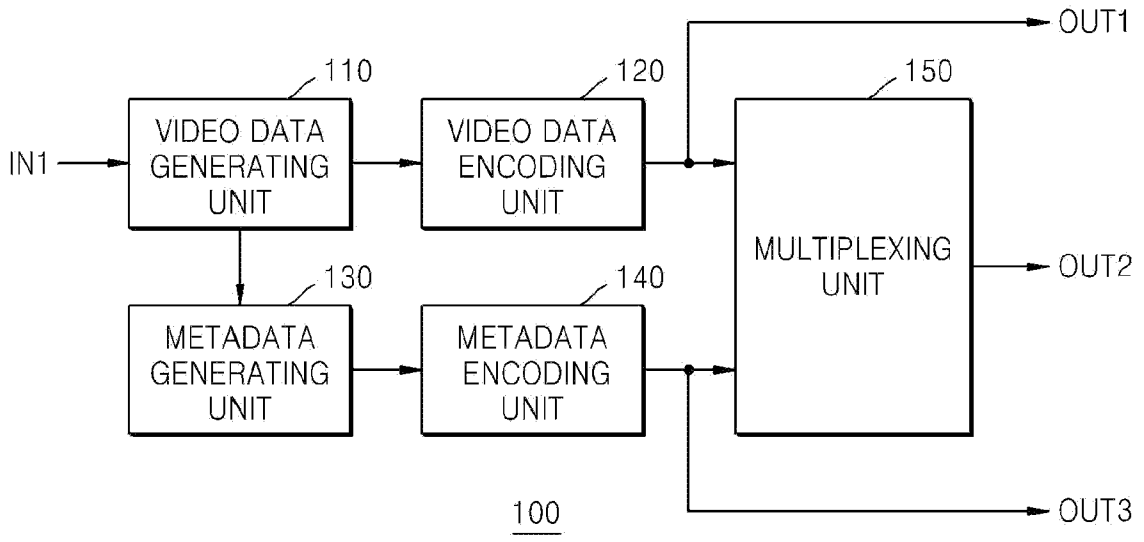


FIG. 1

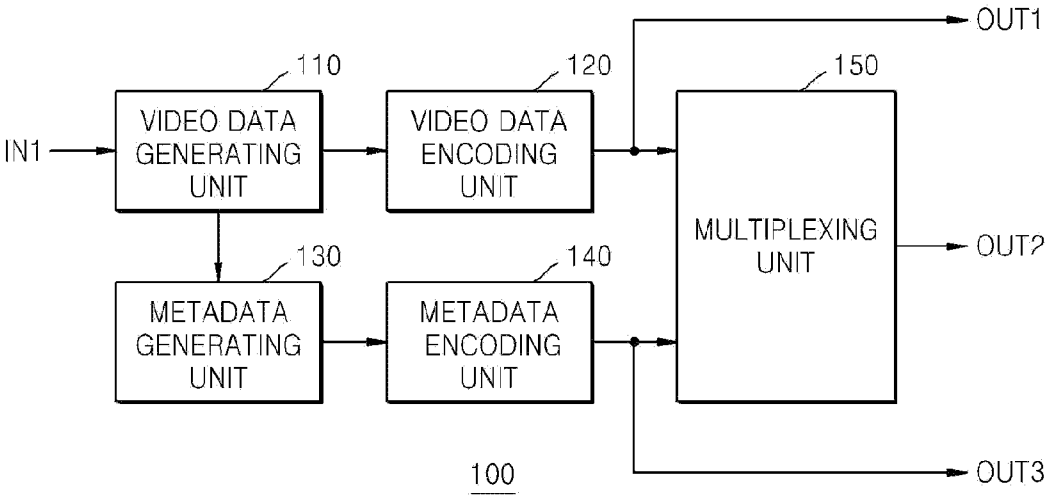
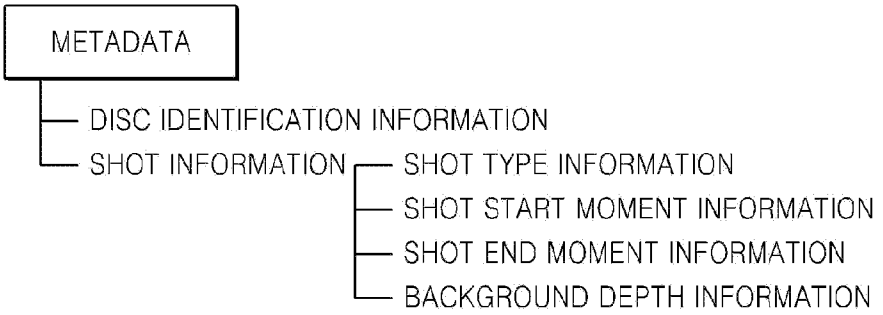

IN1
110
VIDEO DATA GENERATING UNIT
120
VIDEO DATA ENCODING UNIT
150
MULTIPLEXING UNIT
OUT1
OUT2
OUT3
130
METADATA GENERATING UNIT
140
METADATA ENCODING UNIT
100

FIG. 2


METADATA
DISC IDENTIFICATION INFORMATION
SHOT INFORMATION
SHOT TYPE INFORMATION
SHOT START MOMENT INFORMATION
SHOT END MOMENT INFORMATION
BACKGROUND DEPTH INFORMATION

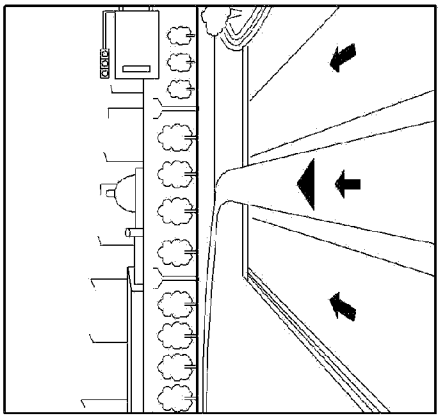


FIG. 3A

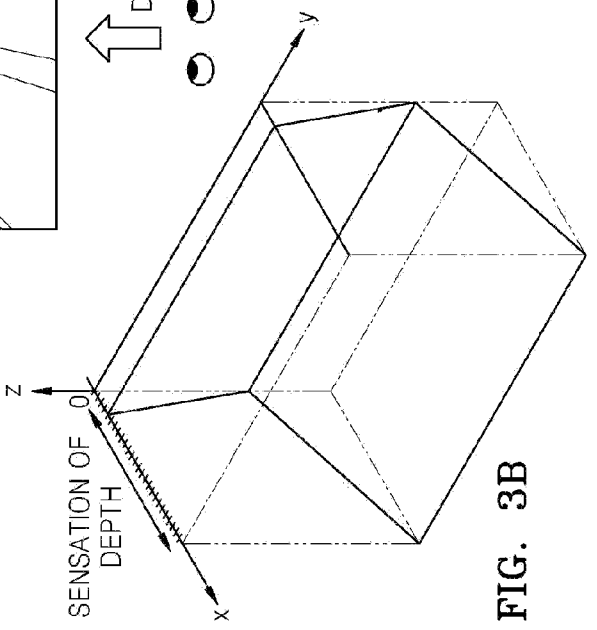


FIG. 3B

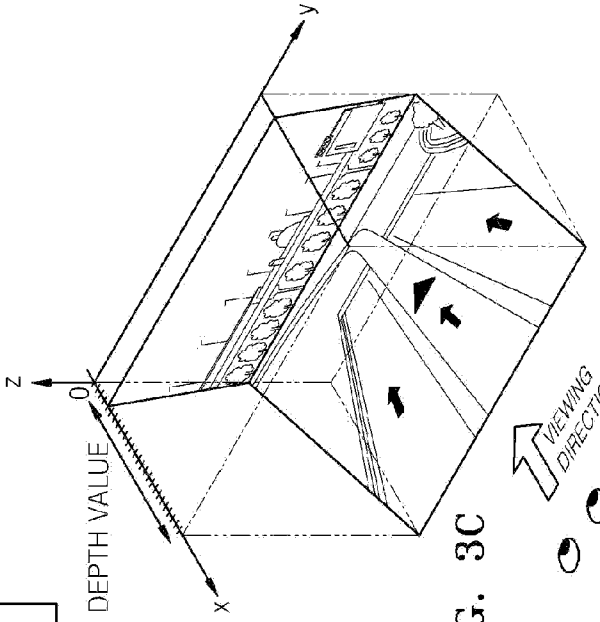


FIG. 3C

FIG. 4

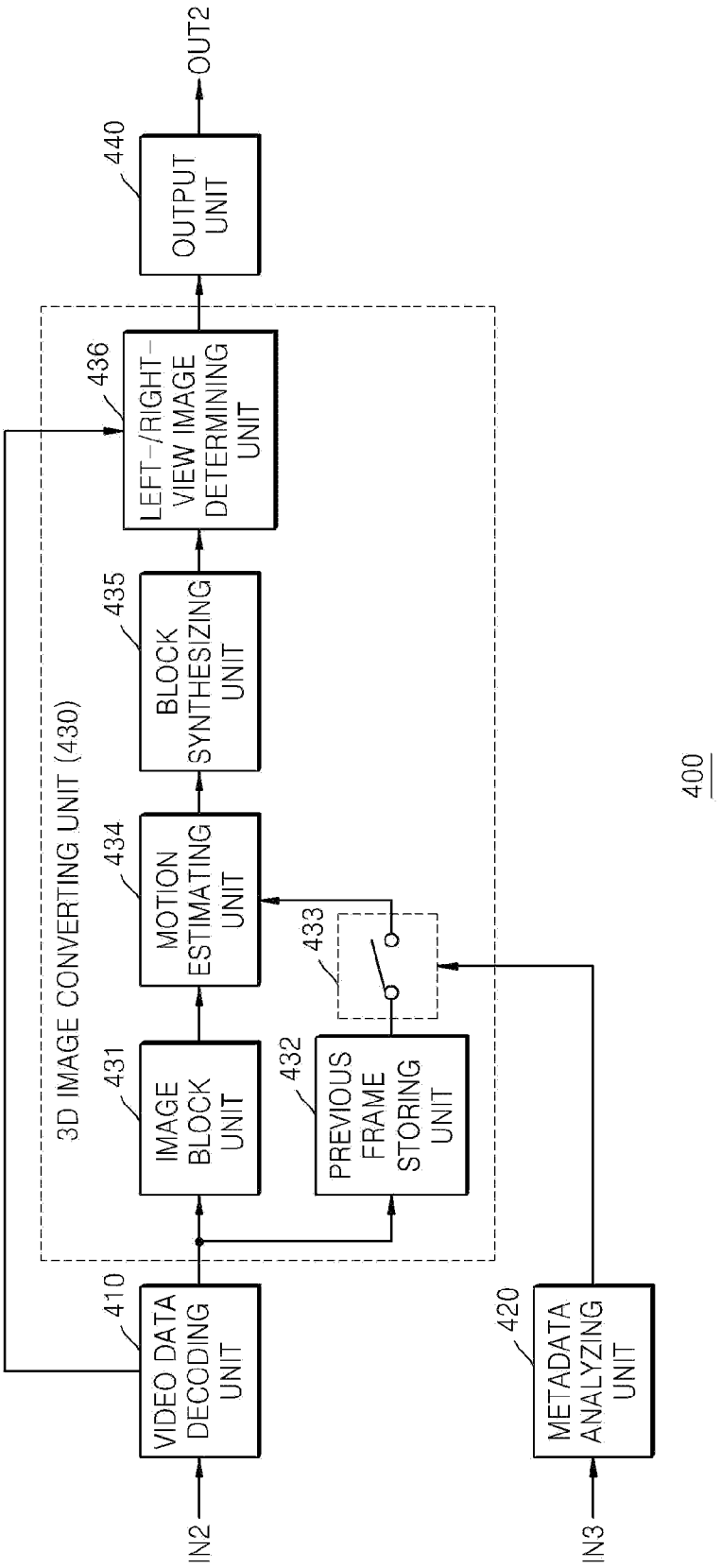

400
IN2
IN3
410
VIDEO DATA DECODING UNIT
420
METADATA ANALYZING UNIT
3D IMAGE CONVERTING UNIT (430)
431
IMAGE BLOCK UNIT
432
PREVIOUS FRAME STORING UNIT
433
434
MOTION ESTIMATING UNIT
435
BLOCK SYNTHESIZING UNIT
436
LEFT-/RIGHT-VIEW IMAGE DETERMINING UNIT
440
OUTPUT UNIT
OUT2

FIG. 5

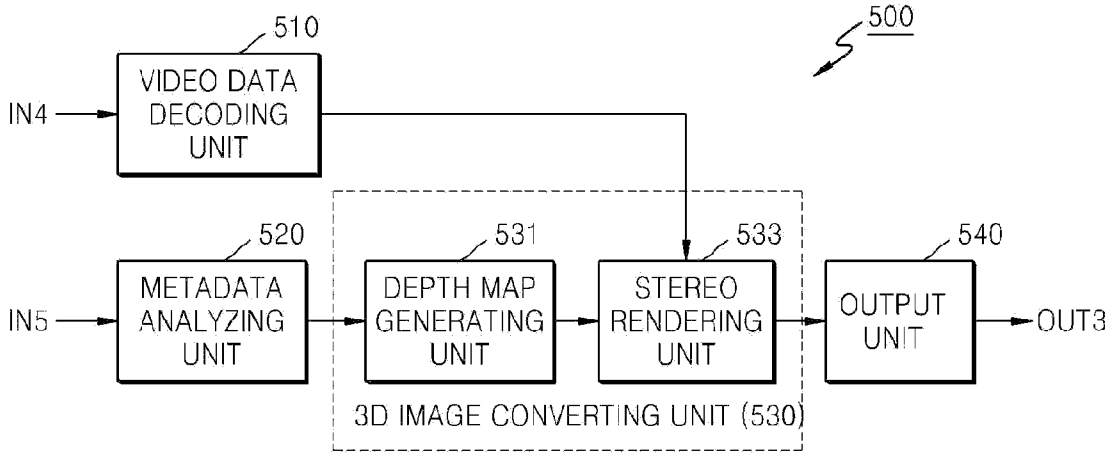

500
510
VIDEO DATA DECODING UNIT
IN4
520
METADATA ANALYZING UNIT
IN5
531
DEPTH MAP GENERATING UNIT
533
STEREO RENDERING UNIT
540
OUTPUT UNIT
OUT3
3D IMAGE CONVERTING UNIT (530)

FIG. 6

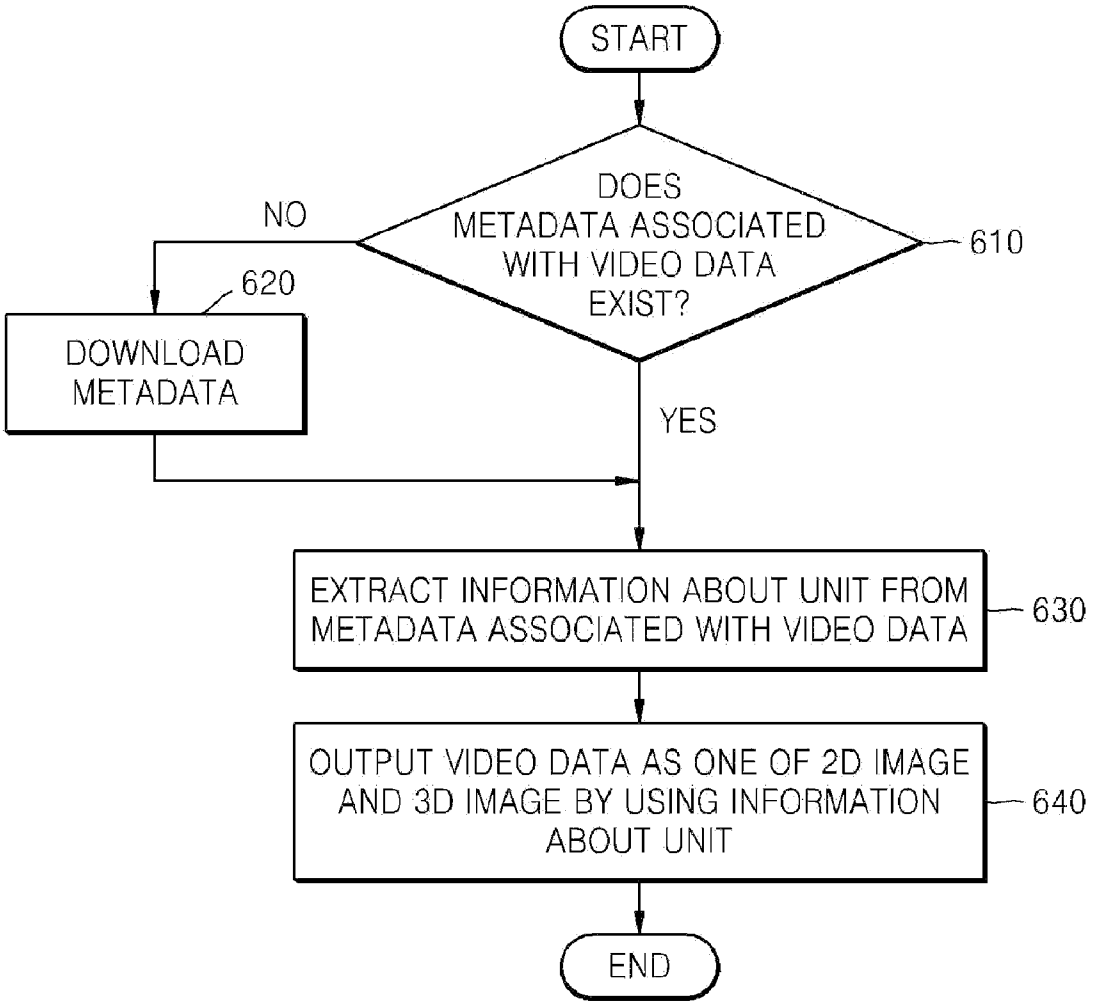

START
DOES METADATA ASSOCIATED WITH VIDEO DATA EXIST?
610
NO
620
DOWNLOAD METADATA
YES
EXTRACT INFORMATION ABOUT UNIT FROM METADATA ASSOCIATED WITH VIDEO DATA
630
OUTPUT VIDEO DATA AS ONE OF 2D IMAGE AND 3D IMAGE BY USING INFORMATION ABOUT UNIT
640
END

FIG. 7

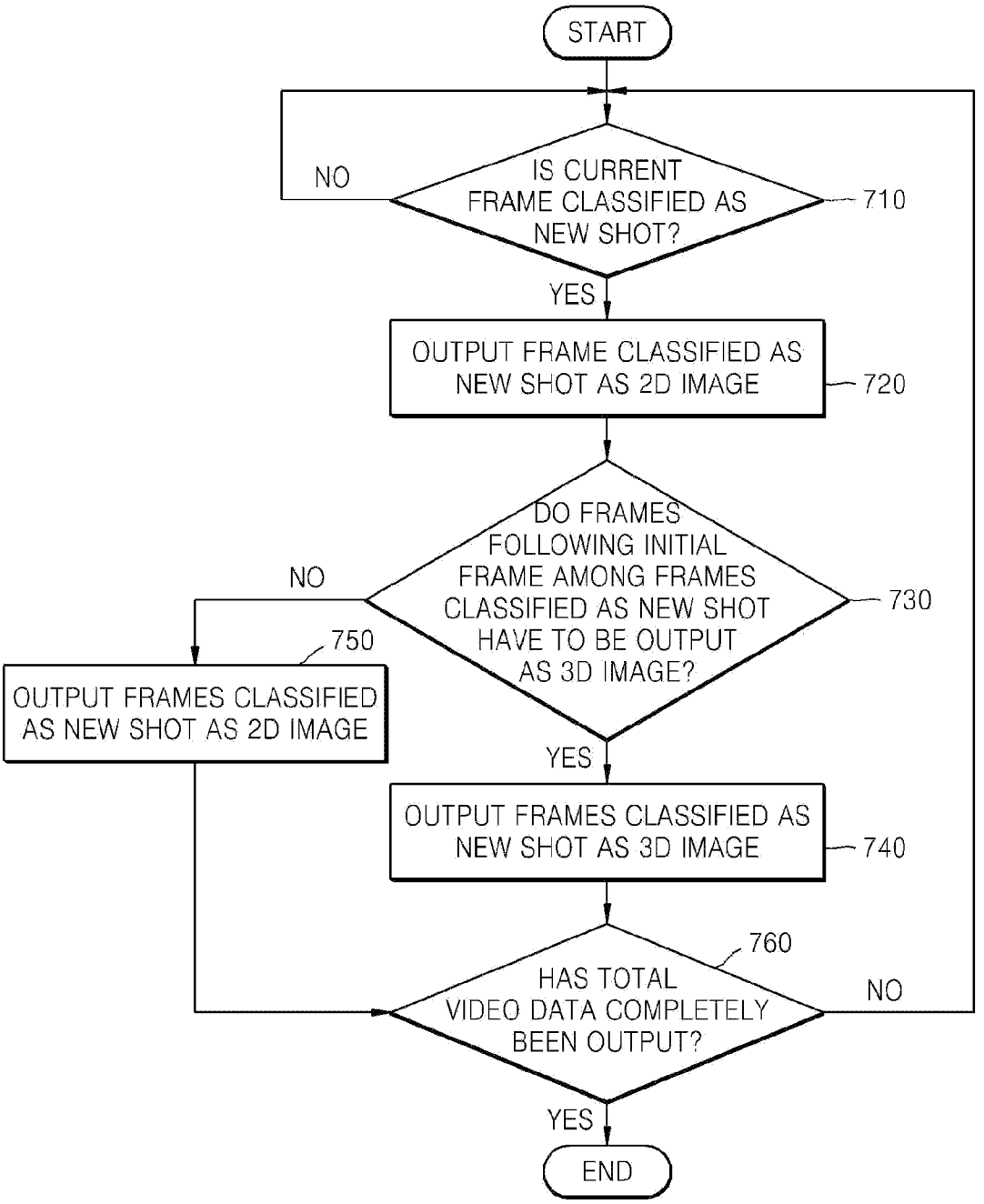

START
IS CURRENT FRAME CLASSIFIED AS NEW SHOT?
710
NO
YES
OUTPUT FRAME CLASSIFIED AS NEW SHOT AS 2D IMAGE
720
DO FRAMES FOLLOWING INITIAL FRAME AMONG FRAMES CLASSIFIED AS NEW SHOT HAVE TO BE OUTPUT AS 3D IMAGE?
730
NO
YES
750
OUTPUT FRAMES CLASSIFIED AS NEW SHOT AS 2D IMAGE
OUTPUT FRAMES CLASSIFIED AS NEW SHOT AS 3D IMAGE
740
760
HAS TOTAL VIDEO DATA COMPLETELY BEEN OUTPUT?
NO
YES
END

IMAGE GENERATING METHOD AND APPARATUS AND IMAGE PROCESSING METHOD AND APPARATUS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 61/075,184, filed on Jun. 24, 2008 in the U.S. Patent and Trademark Office, and the benefit of Korean Patent Application No. 10-2008-0091269, filed on Sep. 17, 2008 in the Korean Intellectual Property Office, the disclosures of which are incorporated herein in their entirety by reference.

BACKGROUND OF THE INVENTION

[0002] 1. Field of the Invention

[0003] Aspects of the present invention generally relate to an image generating method and apparatus and an image processing method and apparatus, and more particularly, to an image generating method and apparatus and an image processing method and apparatus in which video data is output as a two-dimensional (2D) image or a three-dimensional (3D) image by using metadata associated with the video data.

[0004] 2. Description of the Related Art

[0005] With the development of digital technology, three-dimensional (3D) image technology has widely spread. The 3D image technology expresses a more realistic image by adding depth information to a two-dimensional (2D) image. The 3D image technology can be classified into technology to generate video data as a 3D image and technology to convert video data generated as a 2D image into a 3D image. Both technologies have been studied together.

SUMMARY OF THE INVENTION

[0006] Aspects of the present invention provide an image processing method and apparatus to output video data as a two-dimensional image or a three-dimensional image by using metadata associated with the video data can be provided.

[0007] According to an aspect of the present invention, there is provided an image processing method to output video data being a two-dimensional (2D) image as the 2D image or a three-dimensional (3D) image, the image processing method including: extracting information about the video data from metadata associated with the video data; and outputting the video data as the 2D image or the 3D image by using the extracted information about the video data, wherein the information about the video data includes information to classify frames of the video data into predetermined units.

[0008] According to an aspect of the present invention, the information to classify the frames of the video data as the predetermined units may be shot information to classify a group of frames in which a background composition of a current frame is predictable by using a previous frame preceding the current frame as a shot.

[0009] According to an aspect of the present invention, the shot information may include output moment information of a frame being output first and output moment information of a frame being output last from among the group of frames classified as the shot.

[0010] According to an aspect of the present invention, the metadata may include shot type information indicating whether the frames classified as the shot are to be output as the 2D image or the 3D image, and the outputting of the video data may include outputting the frames classified as the shot as the 2D image or the 3D image by using the shot type information.

[0011] According to an aspect of the present invention, the outputting of the video data may include determining, by using the metadata, whether a background composition of a current frame is not predictable by using a previous frame preceding the current frame and thus the current frame is classified as a new shot, outputting the current frame as the 2D image when the current frame is classified as the new shot, and converting the remaining frames of the frames classified as the new shot into the 3D image and outputting the converted 3D image.

[0012] According to an aspect of the present invention, the outputting of the video data may include determining, by using the metadata, whether a background composition of a current frame is not predictable by using a previous frame preceding the current frame and thus the current frame is classified as a new shot, extracting background depth information to be applied to the current frame classified as the new shot from the metadata when the current frame is classified as the new shot, and generating a depth map for the current frame by using the background depth information.

[0013] According to an aspect of the present invention, the generating of the depth map for the current frame may include generating the depth map for a background of the current frame by using coordinate point values of the background of the current frame, depth values corresponding to the coordinate point values, and a panel position value, in which the coordinate point values, the depth value, and the panel position value are included in the background depth information.

[0014] According to an aspect of the present invention, the image processing method may further include reading the metadata from a disc recorded with the video data or downloading the metadata from a server through a communication network.

[0015] According to an aspect of the present invention, the metadata may include identification information to identify the video data, and the identification information may include a disc identifier (ID) to identify a disc recorded with the video data and a title ID to indicate a title including the video data among a plurality of titles recorded in the disc identified by the disc ID.

[0016] According to another aspect of the present invention, there is provided an image generating method including: receiving video data being a two-dimensional (2D) image; and generating metadata associated with the video data, the metadata including information to classify frames of the video data as predetermined units, wherein the information to classify the frames of the video data as the predetermined units is shot information to classify a group of frames in which a background composition of a current frame is predictable by using a previous frame preceding the current frame as a shot.

[0017] According to an aspect of the present invention, the shot information may include output moment information of a frame being output first and output moment information of a frame being output last from among the frames classified as the shot, and/or may include shot type information indicating whether the frames classified as the shot are to be output as the 2D image or a three-dimensional (3D) image.

[0018] According to an aspect of the present invention, the metadata may include background depth information for

frames classified as a predetermined shot and the background depth information may include coordinate point values of a background of the frame classified as the predetermined shot, depth values corresponding to the coordinate point values, and a panel position value.

[0019] According to another aspect of the present invention, there is provided an image processing apparatus to output video data being a two-dimensional (2D) image as the 2D image or a three-dimensional (3D) image, the image processing apparatus including: a metadata analyzing unit to determine whether the video data is to be output as the 2D image or the 3D image by using metadata associated with the video data; a 3D image converting unit to convert the video data into the 3D image when the video data is to be output as the 3D image; and an output unit to output the video data as the 2D image or the 3D image, wherein the metadata includes information to classify frames of the video data into predetermined units.

[0020] According to an aspect of the present invention, the information to classify the frames of the video data into the predetermined units may be shot information to classify a group of frames in which a background composition of a current frame is predictable by using a previous frame preceding the current frame as a shot.

[0021] According to an aspect of the present invention, the shot information may include output moment information of a frame being output first and output moment information of a frame being output last from among the frames classified as the shot.

[0022] According to an aspect of the present invention, the metadata may include shot type information indicating whether the frames classified as the shot are to be output as the 2D image or the 3D image.

[0023] According to an aspect of the present invention, the metadata may include background depth information for a frame classified as a predetermined shot, and the background depth information may include coordinate point values of a background of the frame classified as the predetermined shot, depth values corresponding to the coordinate point values, and a panel position value.

[0024] According to another aspect of the present invention, there is provided an image generating apparatus including: a video data encoding unit to encode video data being a two-dimensional (2D) image; a metadata generating unit to generate metadata associated with the video data, the metadata including information to classify frames of the video data into predetermined units; and a metadata encoding unit to encode the metadata, in which the information to classify the frames of the video data into the predetermined units is shot information to classify a group of frames in which a background composition of a current frame is predictable by using a previous frame preceding the current frame as a shot.

[0025] According to yet another aspect of the present invention, there is provided a computer-readable information storage medium including video data being a two-dimensional (2D) image and metadata associated with the video data, the metadata including information to classify frames of the video data into predetermined units, wherein the information to classify the frames of the video data into the predetermined units is shot information to classify a group of frames in which a background composition of a current frame is predictable by using a previous frame preceding the current frame as a shot.

[0026] According to still another aspect of the present invention, there is provided a computer-readable information storage medium having recorded thereon a program to execute an image processing method to output video data being a two-dimensional (2D) image as the 2D image or a three-dimensional (3D) image, the image processing method including: extracting information about the video data from metadata associated with the video data; and outputting the video data as the 2D image or the 3D image by using the extracted information about the video data, wherein the information about the video data includes information to classify frames of the video data into predetermined units.

[0027] According to an aspect of the present invention, there is provided a system to output video data as a two-dimensional (2D) image or a three-dimensional (3D) image, the system including: an image generating apparatus including: a video data encoding unit to encode the video data being the 2D image, a metadata generating unit to generate metadata associated with the video data, the metadata comprising information to classify frames of the video data as predetermined units and used to determine whether each of the classified frames is to be converted to the 3D image; and an image processing apparatus to receive the encoded video data and the generated metadata, and to output the video data as the 2D image or the 3D image, the image processing apparatus including: a metadata analyzing unit to determine whether the video data is to be output as the 2D image or the 3D image by using the information to classify the frames of the video data comprised in the received metadata associated with the video data, a 3D image converting unit to convert the video data into the 3D image when the metadata analyzing unit determines that the video data is to be output as the 3D image, and an output unit to output the video data as the 2D image or the 3D image according to the determination of the metadata analyzing unit, wherein the information to classify the frames of the video data as the predetermined units is shot information to classify a group of frames in which a background composition of a current frame is predictable by using a previous frame preceding the current frame in the group of frames as a shot.

[0028] According to another aspect of the present invention, there is provided a computer-readable information storage medium including: metadata associated with video data comprising two-dimensional (2D) frames, the metadata comprising information used by an image processing apparatus to classify the frames of the video data as predetermined units and used by the image processing apparatus to determine whether each of the classified frames is to be converted by the image processing apparatus to a three-dimensional (3D) image, wherein the information to classify the frames of the video data as the predetermined units comprises shot information to classify, as a shot, a group of frames in which a background composition of a current frame is predictable by using a previous frame preceding the current frame in the group of frames.

[0029] According to another aspect of the present invention, there is provided an image processing method to output video data having two-dimensional (2D) images as the 2D images or three-dimensional (3D) images, the image processing method including: determining, by an image processing apparatus, whether metadata associated with the video data exists on a disc comprising the video data; reading, by the image processing apparatus, the metadata from the disc if the metadata is determined to exist on the disc; retrieving, by the

image processing apparatus, the metadata from a server if the metadata is determined to not exist on the disc; and outputting, by the image processing apparatus, the video data as selectable between the 2D image and the 3D image according to the metadata.

[0030] Additional aspects and/or advantages of the invention will be set forth in part in the description which follows and, in part, will be obvious from the description, or may be learned by practice of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0031] These and/or other aspects and advantages of the invention will become apparent and more readily appreciated from the following description of the embodiments, taken in conjunction with the accompanying drawings of which:

[0032] FIG. **1** is a block diagram of an image generating apparatus according to an embodiment of the present invention;

[0033] FIG. **2** illustrates metadata generated by the image generating apparatus illustrated in FIG. **1**;

[0034] FIGS. **3**A through **3**C are views to explain a depth map generated by using background depth information;

[0035] FIG. **4** is a block diagram of an image processing apparatus according to an embodiment of the present invention;

[0036] FIG. **5** is a block diagram of an image processing apparatus according to another embodiment of the present invention;

[0037] FIG. **6** is a flowchart illustrating an image processing method according to an embodiment of the present invention; and

[0038] FIG. **7** is a flowchart illustrating in detail an operation illustrated in FIG. **6** where video data is output as a two-dimensional (2D) image or a three-dimensional (3D) image.

DETAILED DESCRIPTION OF THE EMBODIMENTS

[0039] Reference will now be made in detail to the present embodiments of the present invention, examples of which are illustrated in the accompanying drawings, wherein like reference numerals refer to the like elements throughout. The embodiments are described below in order to explain the present invention by referring to the figures.

[0040] FIG. **1** is a block diagram of an image generating apparatus **100** according to an embodiment of the present invention. Referring to FIG. **1**, the image generating apparatus **100** includes a video data generating unit **110**, a video data encoding unit **120**, a metadata generating unit **130**, a metadata encoding unit **140**, and a multiplexing unit **150**. The video data generating unit **110** generates video data and outputs the generated video data to the video data encoding unit **120**. The video data encoding unit **120** encodes the input video data and outputs the encoded video data (OUT**1**) to the multiplexing unit **150**, and/or to an image processing apparatus (not shown) through a communication network, though it is understood that the video data encoding unit **120** may output the encoded video data to the image processing apparatus through any wired and/or wireless connection (such as IEEE 1394, universal serial bus, a Bluetooth, an infrared, etc.). The image generating apparatus **100** may be a computer, a workstation, a camera device, a mobile device, a stand-alone device, etc. Moreover, while not required, each of the units **110**, **120**, **130**, **140**, **150** can be one or more processors or processing elements on one or more chips or integrated circuits.

[0041] The metadata generating unit **130** analyzes the video data generated by the video data generating unit **110** to generate metadata including information about frames of the video data. The metadata includes information to convert the generated video data from a two-dimensional (2D) image into a three-dimensional (3D) image. The metadata also includes information to classify the frames of the video data as predetermined units. The metadata generated by the metadata generating unit **130** will be described in more detail with reference to FIG. **2**. The metadata generating unit **130** outputs the generated metadata to the metadata encoding unit **140**.

[0042] The metadata encoding unit **140** encodes the input metadata and outputs the encoded metadata (OUT**3**) to the multiplexing unit **150** and/or to the image processing apparatus. The multiplexing unit **150** multiplexes the encoded video data (OUT**1**) and the encoded metadata (OUT**3**) and transmits the multiplexing result (OUT**2**) to the image processing apparatus through a wired and/or wireless communication network, or any wired and/or wireless connection, as described above. The metadata encoding unit **140** may transmit the encoded metadata (OUT**3**), separately from the encoded video data (OUT**1**), to the image processing apparatus, instead of to or in addition to the multiplexing unit **150**. In this way, the image generating apparatus **100** generates metadata associated with video data, the metadata including information to convert the video data from a 2D image into a 3D image.

[0043] FIG. **2** illustrates metadata generated by the image generating apparatus **100** illustrated in FIG. **1**. The metadata includes information about video data. In order to indicate with which video data the information included in the metadata is associated, disc identification information to identify a disc in which the video data is recorded is included in the metadata, though it is understood that the metadata does not include the disc identification information in other embodiments. The disc identification information may include a disc identifier (ID) to identify the disc recorded with the video data and a title ID to identify a title including the video data among a plurality of titles recorded in the disc identified by the disc ID.

[0044] Since the video data has a series of frames, the metadata includes information about the frames. The information about the frames may include information to classify the frames according to a predetermined criterion. Assuming that a group of similar frames is a unit, total frames of the video data can be classified as a plurality of units. In the present embodiment, information to classify the frames of the video data as predetermined units is included in the metadata. Specifically, a group of frames having similar background compositions in which a background composition of a current frame can be predicted by using a previous frame preceding the current frame is classified as a shot. The metadata generating unit **130** classifies the frames of the video data as a predetermined shot and incorporates information about the shot (i.e., shot information) into the metadata. When the background composition of the current frame is different from that of the previous frame due to a significant change in the frame background composition, the current frame and the previous frame are classified as different shots.

[0045] The shot information includes information about output moments of frames classified within the shot. For

example, such information includes output moment information of a frame being output first (shot start moment information in FIG. **2**) and output moment information of a frame being output last (shot end moment information in FIG. **2**) among the frames classified as each shot, though aspects of the present invention are not limited thereto. For example, according to other aspects, the shot information includes the shot start moment information and information on a number of frames included in the shot. The metadata further includes shot type information about frames classified as a shot. The shot type information indicates for each shot whether frames classified as a shot are to be output as a 2D image or a 3D image. The metadata also includes background depth information, which will be described in detail with reference to FIGS. **3**A through **3**C.

[0046] FIGS. **3**A through **3**C are views to explain a depth map generated by using the background depth information. FIG. **3**A illustrates a 2D image, FIG. **3**B illustrates a depth map to be applied to the 2D image illustrated in FIG. **3**A, and FIG. **3**C illustrates a result of applying the depth map to the 2D image. In order to add a cubic effect to a 2D image, a sense of depth is given to the 2D image. When a user sees a screen, an image projected on the screen is formed in each of the user's two eyes. A distance between two points of the images formed in the eyes is called parallax, and the parallax can be classified into positive parallax, zero parallax, and negative parallax. The positive parallax refers to parallax corresponding to a case when the image appears to be formed inside the screen, and the positive parallax is less than or equal to a distance between the eyes. As the positive parallax increases, more cubic effect by which the image appears to lie behind the screen is given. When the image appears to be two-dimensionally formed on the screen plane, a parallax is 0 (i.e., zero parallax). In the case of the zero parallax, the user cannot feel a cubic effect because the image is formed on the screen plane. The negative parallax refers to parallax corresponding to a case when the image appears to lie in front of the screen. This parallax is generated when lines of sight to the user's eyes intersect. The negative parallax gives a cubic effect by which the image appears to protrude forward.

[0047] In order to generate a 3D image by adding the sense of depth to a 2D image, a motion of a current frame may be predicted by using a previous frame and the sense of depth may be added to an image of the current frame by using the predicted motion. For the same purpose, a depth map for a frame may be generated by using a composition of the frame and the sense of depth may be added to the frame by using the depth map. The former will be described in detail with reference to FIG. **4**, and the latter will be described in detail with reference to FIG. **5**.

[0048] As stated previously, metadata includes information to classify frames of video data as predetermined shots. When a composition of a current frame cannot be predicted by using a previous frame due to no similarity in composition between the current frame and the previous frame, the current frame and the previous frame are classified as different shots. The metadata includes information about compositions to be applied to frames classified as a shot due to their similarity in composition, and/or includes information about a composition to be applied to each shot.

[0049] Background compositions of frames may vary. The metadata includes background depth information to indicate a composition of a corresponding frame. The background depth information may include type information of a background included in a frame, coordinate point information of the background, and a depth value of the background corresponding to a coordinate point. The type information of the background may be an ID indicating a composition of the background from among a plurality of compositions.

[0050] Referring to FIG. **3**A, a frame includes a background including the ground and the sky. In this frame, the horizon where the ground and the sky meet is the farthest point from the perspective of a viewer, and an image corresponding to the bottom portion of the ground is the nearest point from the perspective of the viewer. The image generating apparatus **100** determines that a composition of a type illustrated in FIG. **3**B is to be applied to the frame illustrated in FIG. **3**A, and generates metadata including type information indicative of the composition illustrated in FIG. **3**B for the frame illustrated in FIG. **3**A.

[0051] Coordinate point values refer to values of a coordinate point of a predetermined position in 2D images. A depth value refers to the degree of depth of an image. In aspects of the present invention, the depth value may be one of 256 values ranging from 0 to 255. As the depth value decreases, the depth becomes greater and thus an image appears to be farther from a viewer. Conversely, as the depth value increases, an image appears nearer to a viewer. Referring to FIGS. **3**B and **3**C, it can be seen that a portion where the ground and the sky meets (i.e., the horizon portion) has a smallest depth value and the bottom portion of the ground has a largest depth value in the frame. The image processing apparatus (not shown) extracts the background depth information included in the metadata, generates the depth map as illustrated in FIG. **3**C by using the extracted depth information, and outputs a 2D image as a 3D image by using the depth map.

[0052] FIG. **4** is a block diagram of an image processing apparatus **400** according to an embodiment of the present invention. Referring to FIG. **4**, the image processing apparatus **400** includes a video data decoding unit **410**, a metadata analyzing unit **420**, and a 3D image converting unit **430**, and an output unit **440** to output a 3D image to a screen. However, it is understood that the image processing apparatus **400** need not include the output unit **440** in all embodiments, and/or the output unit **440** may be provided separately from the image processing apparatus **400**. Moreover, the image processing apparatus **400** may be a computer, a mobile device, a set-top box, a workstation, etc. The output unit **440** may be a cathode ray tube device, a liquid crystal display device, a plasma display device, an organic light emitting diode display device, etc. and/or be connected to the same and or connected to goggles through wired and/or wireless protocols.

[0053] The video data decoding unit **410** reads video data (IN**2**) from a disc (such as a DVD, Blu-ray, etc.), a local storage, transmitted the image generating device **100** of FIG. **1**, or any external storage device (such as a hard disk drive, a flash memory, etc.) and decodes the read video data. The metadata analyzing unit **420** decodes metadata (IN**3**) to extract information about frames of the read video data from the metadata, and analyzes the extracted information. By using the metadata, the metadata analyzing unit **420** controls a switching unit **433** included in the 3D image converting unit **430** in order to output a frame as a 2D image or a 3D image. The metadata analyzing unit **420** receives the metadata IN**3** from a disc, a local storage, transmitted from the image generating device **100** of FIG. **1**, or any external storage device

(such as a hard disk drive, a flash memory, etc.). The metadata need not be stored with the video data in all aspects of the invention.

[0054] The 3D image converting unit **430** converts the video data from a 2D image received from the video data decoding unit **410** into a 3D image. In FIG. **4**, the 3D image converting unit **430** estimates a motion of a current frame by using a previous frame in order to generate a 3D image for the current frame.

[0055] The metadata analyzing unit **420** extracts, from the metadata, output moment information of a frame being output first and/or output moment information of a frame being output last among frames classified as a shot, and determines whether a current frame being currently decoded by the video data decoding unit **410** is classified as a new shot, based on the extracted output moment information. When the metadata analyzing unit **420** determines that the current frame is classified as a new shot, the metadata analyzing unit **420** controls the switching unit **433** in order to not convert the current frame into a 3D image such that a motion estimating unit **434** does not estimate the motion of the current frame by using a previous frame stored in a previous frame storing unit **432**. This is because motion information of a current frame is extracted by referring to a previous frame in order to convert video data from a 2D image into a 3D image. However, if the current frame and the previous frame are classified as different shots, the current frame and the previous frame do not have sufficient similarity therebetween, and thus a composition of the current frame cannot be predicted by using the previous frame. As shown, the switch unit **433** disconnects the storing unit **432** to prevent use of the previous frame, but aspects of the invention are not limited thereto.

[0056] When the video data is not to be converted into a 3D image (for example, when the video data is a warning sentence, a menu screen, an ending credit, etc.), the metadata includes the shot type information indicating that frames of the video data are to be output as a 2D image. The metadata analyzing unit **420** determines whether the video data is to be output as a 2D image or a 3D image for each shot using the shot type information and controls the switching unit **433** depending on a result of the determination. Specifically, when the metadata analyzing unit **420** determines, based on the shot type information, that video data classified as a predetermined shot does is not to be converted into a 3D image, the metadata analyzing unit **420** controls the switching unit **433** such that the 3D image converting unit **430** does not estimate the motion of the current frame by using the previous frame by disconnected the storing unit **432** from the motion estimating unit **434**. When the metadata analyzing unit **420** determines, based on the shot type information, that video data classified as a predetermined shot is to be converted into a 3D image, the metadata analyzing unit **420** controls the switching unit **433** such that the image converting unit **430** converts the current frame into a 3D image by using the previous frame by connecting the storing unit **432** and the motion estimating unit **434**.

[0057] When the video data is classified as a predetermined shot and is to be output as a 3D image, the 3D image converting unit **430** converts the video data being a 2D image received from the video data decoding unit **410** into the 3D image. The 3D image converting unit **430** includes an image block unit **431**, the previous frame storing unit **432**, the motion estimating unit **434**, a block synthesizing unit **435**, a left-/right-view image determining unit **436**, and the switching unit **433**. The image block unit **431** divides a frame of video data, which is a 2D image, into blocks of a predetermined size. The previous frame storing unit **432** stores a predetermined number of previous frames preceding a current frame. Under the control of the metadata analyzing unit **420**, the switching unit **433** enables or disables outputting of previous frames stored in the previous frame storing unit **432** to the motion estimating unit **434**.

[0058] The motion estimating unit **434** obtains a per-block motion vector regarding the amount and direction of motion using a block of a current frame and a block of a previous frame. The block synthesizing unit **435** synthesizes blocks selected by using the motion vectors obtained by the motion estimating unit **434** from among predetermined blocks of previous frames in order to generate a new frame. When the motion estimating unit **434** does not use a previous frame due to the control of the switching unit **433** by the metadata analyzing unit **420**, the motion estimating unit **434** outputs the current frame received from the image block unit **431** to the block synthesizing unit **435**.

[0059] The generated new frame or the current frame is input to the left-/right-view image determining unit **436**. The left-/right-view image determining unit **436** determines a left-view image and a right-view image by using the frame received from the block synthesizing unit **435** and a frame received from the video data decoding unit **410**. When the metadata analyzing unit **420** controls the switching unit **433** to not convert video data into a 3D image, the left-/right-view image determining unit **436** generates the left-view image and the right-view image that are the same as each other by using the frame with a 2D image received from the block synthesizing unit **435** and the frame with a 2D image received from the video data decoding unit **410**. The left-/right-view image determining unit **436** outputs the left-view image and the right-view image to the output unit **440**, an external output device, and/or an external terminal (such as a computer, an external display device, a server, etc.).

[0060] The image processing apparatus **400** further includes the output unit **440** to output the left-view image and the right-view image (OUT**2**) determined by the left-/right-view image determining unit **436** to the screen alternately at lest every $1/120$ second. As such, by using the shot information included in the metadata, the image processing apparatus **400** according to an embodiment of the present invention does not convert video data corresponding to a shot change point or video data for which 3D image conversion is not required according to the determination based on the shot information provided in metadata, thereby reducing unnecessary computation and complexity of the apparatus **400**. While not required, the output image OUT**2** can be received at a receiving unit through which a user sees the screen, such as goggles, through wired and/or wireless protocols.

[0061] FIG. **5** is a block diagram of an image processing apparatus **500** according to another embodiment of the present invention. Referring to FIG. **5**, the image processing apparatus **500** includes a video data decoding unit **510**, a metadata analyzing unit **520**, a 3D image converting unit **530**, and an output unit **540**. However, it is understood that the image processing apparatus **500** need not include the output unit **540** in all embodiments, and/or the output unit **540** may be provided separately from the image processing apparatus **500**. Moreover, the image processing apparatus **500** may be a computer, a mobile device, a set-top box, a workstation, etc. The output unit **540** may be a cathode ray tube device, a liquid

crystal display device, a plasma display device, an organic light emitting diode display device, etc. and/or connected to the same or connected to goggles through wired and/or wireless protocols. Moreover, while not required, each of the units **510**, **520**, **530** can be one or more processors or processing elements on one or more chips or integrated circuits.

[0062] When video data that is a 2D image and metadata associated with the video data are recorded in a disc (not shown) in a multiplexed state or separately from each other, upon loading of the disc recorded with the video data and the metadata into the image processing apparatus **500**, the video data decoding unit **510** and the metadata analyzing unit **520** read the video data (IN**4**) and the metadata (IN**5**) from the loaded disc. The metadata may be recorded in a lead-in region, a user data region, and/or a lead-out region of the disc. However, it is understood that aspects of the present invention are not limited to receiving the video data and the metadata from a disc. For example, according to other aspects, the image processing apparatus **500** may further include a communicating unit (not shown) to communicate with an external server or an external terminal (for example, through a communication network and/or any wired/wireless connection). The image processing apparatus **500** may download video data and/or metadata associated therewith from the external server or the external terminal and store the downloaded data in a local storage (not shown). Furthermore, the image processing apparatus **500** may receive the video data and/or metadata from any external storage device different from the disc (for example, a flash memory).

[0063] The video data decoding unit **510** reads the video data from the disc, the external storage device, the external terminal, or the local storage and decodes the read video data. The metadata analyzing unit **520** reads the metadata associated with the video data from the disc, the external storage device, the external terminal, or the local storage and analyzes the read metadata. When the video data is recorded in the disc, the metadata analyzing unit **520** extracts, from the metadata, a disc ID to identify the disc recorded with the video data and a title ID indicating titles including the video data among a plurality of titles in the disc, and determines which video data the metadata is associated with by using the extracted disc ID and title ID.

[0064] The metadata analyzing unit **520** analyzes the metadata to extract information about frames of the video data classified as a predetermined shot. The metadata analyzing unit **520** determines whether a current frame is video data corresponding to a shot change point (i.e., is classified as a new shot), in order to control a depth map generating unit **531**. The metadata analyzing unit **520** determines whether the frames classified as the predetermined shot are to be output as a 2D image or a 3D image by using shot type information, and controls the depth map generating unit **531** according to a result of the determination. Furthermore, the metadata analyzing unit **520** extracts depth information from the metadata and outputs the depth information to the depth map generating unit **531**.

[0065] The 3D image converting unit **530** generates a 3D image for video data. The 3D image converting unit **530** includes the depth map generating unit **531** and a stereo rendering unit **533**. The depth map generating unit **531** generates a depth map for a frame by using the background depth information received from the metadata analyzing unit **520**. The background depth information includes coordinate point values of a background included in a current frame, a depth value corresponding to the coordinate point values, and a panel position value that represents a depth value of the screen on which an image is output. The depth map generating unit **531** generates a depth map for the background of the current frame by using the background depth information and outputs the generated depth map to the stereo rendering unit **533**. However, when the current frame is to be output as a 2D image, the depth map generating unit **531** outputs the current frame to the stereo rendering unit **533** without generating the depth map for the current frame.

[0066] The stereo rendering unit **533** generates a left-view image and a right-view image by using the video data received from the video data decoding unit **510** and the depth map received from the depth map generating unit **531**. Accordingly, the stereo rendering unit **533** generates a 3D-format image including both the generated left-view image and the generated right-view image. When the current frame is to be output as a 2D image, a frame received from the depth map generating unit **531** and a frame received from the video data decoding unit **510** are the same as each other, and thus the left-view image and the right-view image generated by the stereo rendering unit **533** are also the same as each other. The 3D format may be a top-and-down format, a side-by-side format, or an interlaced format. The stereo rendering unit **533** outputs the left-view image and the right-view image to the output unit **540**, an external output device, and/or an external terminal (such as a computer, an external display device, a server, etc.).

[0067] In the present embodiment, the image processing apparatus **500** further includes the output unit **540** that operates as an output device. In this case, the output unit **540** sequentially outputs the left-view image and the right-view image received from the stereo rendering unit **533** to the screen. A viewer perceives that an image is sequentially and seamlessly reproduced when the image is output at a frame rate of at least 60 Hz as viewed from a single eye. Therefore, the output unit **540** outputs the screen at a frame rate of at least 120 Hz so that the viewer can perceive that a 3D image is seamlessly reproduced. Accordingly, the output unit **540** sequentially outputs the left-view image and the right-view image (OUT**3**) included in a frame to the screen at least every $1/120$ second. The viewer can have his/her view selectively blocked using goggles to alternate which eye receives the image and/or using polarized light.

[0068] FIG. **6** is a flowchart illustrating an image processing method according to an embodiment of the present invention. Referring to FIG. **6**, the image processing apparatus **400** or **500** determines whether metadata associated with read video data exists in operation **610**. For example, when the video data and metadata are provided on a disc and the disc is loaded and the image processing apparatus **400** or **500** is instructed to output a predetermined title of the loaded disc, the image processing apparatus **400** or **500** determines whether metadata associated with the title exists therein by using a disc ID and a title ID in operation **610**. If the image processing apparatus **400** or **500** determines that the disc does not have the metadata therein, the image forming apparatus **400** or **500** may download the metadata from an external server or the like through a communication network in operation **620**. In this manner, existing video (such as movies on DVD and Blu-ray discs or computer games) can become 3D by merely downloading the corresponding metadata. Alter-

natively, the disc could only contain the metadata, and when the metadata for a particular video is selected, the video is downloaded from the server.

[0069] The image processing apparatus **400** or **500** extracts information about a unit in which the video data is classified from the metadata associated with the video data in operation **630**. As previously described, the information about a unit may be information about a shot (i.e., shot information) in some aspects of the present invention. The shot information indicates whether a current frame is classified as the same shot as a previous frame, and may include shot type information indicating whether the current frame is to be output as a 2D image or a 3D image. The image processing apparatus **400** or **500** determines whether to output frames as a 2D image or a 3D image by using the shot information, and outputs frames classified as a predetermined shot as a 2D image or a 3D image according to a result of the determination in operation **640**.

[0070] FIG. **7** is a flowchart illustrating in detail operation **640** of FIG. **6**. Referring to FIG. **7**, the image processing apparatus **400** or **500**, when outputting video data, determines whether a current frame has a different composition from a previous frame and is, thus, classified as a new shot in operation **710**. When the image processing apparatus **400** or **500** determines that the current frame is classified as the new shot, the image processing apparatus **400** or **500** outputs an initial frame included in the new shot as a 2D image without converting the initial frame into a 3D image in operation **720**.

[0071] The image processing apparatus **400** or **500** determines whether to output the remaining frames following the initial frame among total frames classified as the new shot as a 2D image or a 3D image by using shot type information regarding the new shot, provided in metadata, in operation **730**. When the shot type information regarding the new shot indicates that video data classified as the new shot is to be output as a 3D image, the image processing apparatus **400** or **500** converts the video data classified as the new shot into a 3D image in operation **740**. Specifically, the image processing apparatus **400** or **500** determines a left-view image and a right-view image from the video data converted into the 3D image and the video data being a 2D image and outputs the video data classified as the new shot as a 3D image in operation **740**. When the image processing apparatus **500** generates a 3D image by using composition information as in FIG. **5**, the image processing apparatus **500** extracts background depth information to be applied to a current frame classified as a new shot from metadata and generates a depth map for the current frame by using the background depth information.

[0072] When the shot type information regarding the new shot indicates that the video data classified as the new shot is to be output as a 2D image (operation **730**), the image processing apparatus **400** and **500** outputs the video data as a 2D image without converting the video data into a 3D image in operation **750**. The image processing apparatus **400** or **500** determines whether the entire video data has been completely output in operation **760**. If not, the image processing apparatus **400** or **500** repeats operation **710**.

[0073] In this way, according to aspects of the present invention, by using shot information included in metadata, video data can be output as a 2D image at a shot change point. Moreover, according to an embodiment of the present invention, it is determined for each shot whether to output video data as a 2D image or a 3D image and the video data is output according to a result of the determination, thereby reducing the amount of computation that may increase due to conversion of total video data into a 3D image.

[0074] While not restricted thereto, aspects of the present invention can also be embodied as computer-readable code on a computer-readable recording medium. The computer-readable recording medium is any data storage device that can store data that can be thereafter read by a computer system. Examples of the computer-readable recording medium include read-only memory (ROM), random-access memory (RAM), CD-ROMs, magnetic tapes, floppy disks, and optical data storage devices. The computer-readable recording medium can also be distributed over network-coupled computer systems so that the computer-readable code is stored and executed in a distributed fashion. Aspects of the present invention may also be realized as a data signal embodied in a carrier wave and comprising a program readable by a computer and transmittable over the Internet. Moreover, while not required in all aspects, one or more units of the image processing apparatus **400** and **500** can include a processor or microprocessor executing a computer program stored in a computer-readable medium, such as a local storage (not shown). Furthermore, it is understood that the image generating apparatus **100** and the image processing apparatus **400** or **500** may be provided in a single apparatus in some embodiments.

[0075] Although a few embodiments of the present invention have been shown and described, it would be appreciated by those skilled in the art that changes may be made in this embodiment without departing from the principles and spirit of the invention, the scope of which is defined in the claims and their equivalents.

What is claimed is:

1. An image processing method to output video data having two-dimensional (2D) images as the 2D images or three-dimensional (3D) images, the image processing method comprises:

extracting, by an image processing apparatus, information about the video data from metadata associated with the video data; and

outputting, by the image processing apparatus, the video data as selectable between the 2D image and the 3D image according to the extracted information about the video data,

wherein the information about the video data includes information to classify frames of the video data into predetermined units.

2. The image processing method as claimed in claim **1**, wherein the information to classify the frames of the video data into the predetermined units is shot information to classify, as a shot, a group of frames in which a background composition of a current frame is predictable by using a previous frame preceding the current frame in the group of frames.

3. The image processing method as claimed in claim **2**, wherein the shot information comprises output moment information of a frame being output first from among the group of frames classified as the shot and/or output moment information of a frame being output last from among the group of frames classified as the shot.

4. The image processing method as claimed in claim **2**, wherein:

the metadata comprises shot type information indicating whether the group of frames classified as the shot are to be output as the 2D image or the 3D image; and

the outputting of the video data comprises outputting the group of frames classified as the shot as the 2D image or the 3D image according to the shot type information.

5. The image processing method as claimed in claim **2**, wherein the outputting of the video data comprises:

according to the metadata, determining that a current frame is classified as a new shot as compared to a previous frame preceding the current frame when a background composition of the current frame is not predictable by using the previous frame;

when the current frame is classified as the new shot, outputting the current frame as the 2D image; and

converting other frames of a group of frames classified as the new shot into the 3D image and outputting the converted 3D image.

6. The image processing method as claimed in claim **2**, wherein the outputting of the video data comprises:

according to the metadata, determining that a current frame is classified as a new shot as compared to a previous frame preceding the current frame when a background composition of the current frame is not predictable by using the previous frame;

when the current frame is classified as the new shot, extracting background depth information to be applied to the current frame classified as the new shot from the metadata; and

when the current frame is classified as the new shot, generating a depth map for the current frame by using the background depth information.

7. The image processing method as claimed in claim **6**, wherein:

the background depth information comprises coordinate point values of a background of the current frame, depth values respectively corresponding to the coordinate point values, and a panel position value; and

the generating of the depth map for the current frame comprises generating the depth map for the background of the current frame by using the coordinate point values, the depth values, and the panel position value that represents a depth value of an output screen.

8. The image processing method as claimed in claim **1**, further comprising reading the metadata from a disc recorded with the video data or downloading the metadata from a server through a communication network.

9. The image processing method as claimed in claim **1**, wherein the metadata comprises identification information to identify the video data, and the identification information comprises a disc identifier (ID) to identify a disc recorded with the video data and a title ID to indicate a title including the video data from among a plurality of titles recorded in the disc identified by the disc ID.

10. An image generating method comprising:

receiving, by an image generating apparatus, video data as two-dimensional (2D) images; and

generating, by the image generating apparatus, metadata associated with the video data, the metadata comprising information to classify frames of the video data as predetermined units and used to determine whether each of the classified frames is to be converted to a three-dimensional (3D) image,

wherein the information to classify the frames of the video data as the predetermined units comprises shot information to classify a group of frames, as a shot, in which a background composition of a current frame is predictable by using a previous frame preceding the current frame in the group of frames.

11. The image generating method as claimed in claim **10**, wherein the shot information comprises output moment information of a frame being output first from among the group of frames classified as the shot, output moment information of a frame being output last from among the group of frames classified as the shot, and/or shot type information indicating whether the group of frames classified as the shot are to be output as the 2D image or the 3D image.

12. The image generating method as claimed in claim **10**, wherein:

the metadata further comprises background depth information for the group of frames classified as the predetermined shot; and

the background depth information comprises coordinate point values of a background of the group of frames classified as the predetermined shot, depth values corresponding to the coordinate point values, and a panel position value that represents a depth value of an output screen.

13. An image processing apparatus to output video data having two-dimensional (2D) images as the 2D images or three-dimensional (3D) images, the image processing apparatus comprising:

a metadata analyzing unit to determine whether the video data is to be output as the 2D image or the 3D image by using metadata associated with the video data;

a 3D image converting unit to convert the video data into the 3D image when the metadata analyzing unit determines that the video data is to be output as the 3D image; and

an output unit to output the video data as the 2D image or the 3D image according to the determination of the metadata analyzing unit,

wherein the metadata includes information to classify frames of the video data into predetermined units.

14. The image processing apparatus as claimed in claim **13**, wherein the information to classify the frames of the video data into the predetermined units comprises shot information to classify, as a shot, a group of frames in which a background composition of a current frame is predictable by using a previous frame preceding the current frame in the group of frames.

15. The image processing apparatus as claimed in claim **14**, wherein the shot information comprises output moment information of a frame being output first from among the group of frames classified as the shot and/or output moment information of a frame being output last from among the group of frames classified as the shot.

16. The image processing apparatus as claimed in claim **14**, wherein:

the metadata comprises shot type information indicating whether the group of frames classified as the shot are to be output as the 2D image or the 3D image; and

the metadata analyzing unit determines whether the group of frames classified as the shot are to be output as the 2D image or the 3D image according to the shot type information.

17. The image processing apparatus as claimed in claim **14**, wherein the metadata analyzing unit determines, according to the metadata, that a current frame is classified as a new shot as compared to a previous frame preceding the current frame when a background composition of the current frame is not

predictable by using the previous frame, determines that the current frame is to be output as the 2D image when the current frame is classified as the new shot, and determines that the current frame is to be output as the 3D image when the current frame is not classified as the new shot.

18. The image processing apparatus as claimed in claim **14**, wherein:

the metadata analyzing unit determines, according to the metadata, that a current frame is classified as a new shot as compared to a previous frame preceding the current frame when a background composition of the current frame is not predictable by using the previous frame; and

when the current frame is classified as the new shot, the 3D image converting unit extracts background depth information to be applied to the current frame classified as the new shot from the metadata and generates a depth map for the current frame by using the background depth information.

19. The image processing apparatus as claimed in claim **18**, wherein:

the background depth information comprises coordinate point values of a background of the current frame, depth values respectively corresponding to the coordinate point values, and a panel position value that represents a depth value of an output screen; and

the 3D image converting unit generates the depth map for a background of the current frame by using the coordinate point values of the background of the current frame, the depth values respectively corresponding to the coordinate point values, and the panel position value.

20. The image processing apparatus as claimed in claim **13**, wherein the metadata is read from a disc recorded with the video data or downloaded from a server through a communication network.

21. The image processing apparatus as claimed in claim **13**, wherein the metadata comprises identification information to identify the video data, and the identification information comprises a disc identifier (ID) to identify a disc recorded with the video data and a title ID to indicate a title including the video data from among a plurality of titles recorded in the disc identified by the disc ID.

22. An image generating apparatus comprising:

a video data encoding unit to encode video data as two-dimensional (2D) images;

a metadata generating unit to generate metadata associated with the video data, the metadata comprising information to classify frames of the video data as predetermined units and used to determine whether each of the classified frames is to be converted to a three-dimensional (3D) image; and

a metadata encoding unit to encode the metadata,

wherein the information to classify the frames of the video data as the predetermined units comprises shot information to classify, as a shot, a group of frames in which a background composition of a current frame is predictable by using a previous frame preceding the current frame in the group of frames.

23. The image generating apparatus as claimed in claim **22**, wherein the shot information comprises output moment information of a frame being output first from among the group of frames classified as the shot, output moment information of a frame being output last from among the group of frames classified as the shot, and/or includes shot type information indicating whether the group of frames classified as the shot are to be output as the 2D image or the 3D image.

24. The image generating apparatus as claimed in claim **22**, wherein:

the metadata further comprises background depth information for the group of frames classified as the predetermined shot; and

the background depth information comprises coordinate point values of a background of the group of frames classified as the predetermined shot, depth values corresponding to the coordinate point values, and a panel position value that represents a depth value of an output screen.

25. A computer-readable information storage medium comprising:

video data recorded as two-dimensional (2D) images; and

metadata associated with the video data, the metadata comprising information used by an image processing apparatus to classify frames of the video data as predetermined units and used by the image processing apparatus to determine whether each of the classified frames is to be converted by the image processing apparatus to a three-dimensional (3D) image,

wherein the information to classify the frames of the video data as the predetermined units comprises shot information used by the used by the image processing apparatus to classify, as a shot, a group of frames in which a background composition of a current frame is predictable by using a previous frame preceding the current frame in the group of frames.

26. The computer-readable information storage medium as claimed in claim **25**, wherein the shot information comprises output moment information of a frame being output first from among the group of frames classified as the shot, output moment information of a frame being output last from among the group of frames classified as the shot, and/or shot type information indicating whether the group of frames classified as the shot are to be output as the 2D image or the 3D image.

27. The computer-readable information storage medium as claimed in claim **25**, wherein:

the metadata further comprises background depth information for the group of frames classified as the predetermined shot; and

the background depth information comprises coordinate point values of a background of the group of frames classified as the predetermined shot, depth values corresponding to the coordinate point values, and a panel position value that represents to the image processing apparatus a depth value of an output screen.

28. A computer-readable information storage medium having recorded thereon a program to execute the image processing method of claim **1** and implemented by the image processing apparatus.

29. A computer-readable information storage medium having recorded thereon a program to execute the image generating method of claim **10** and implemented by the image generating apparatus.

* * * * *