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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
06 April 2023 (06.04.2023)

WIPO | PCT

(10) International Publication Number
WO 2023/056111 A1

(51) International Patent Classification:
H04N 19/597 (2014.01) *H04N 19/44* (2014.01)
H04N 19/30 (2014.01) *H04N 19/70* (2014.01)

(21) International Application Number: PCT/US2022/072712

(22) International Filing Date: 02 June 2022 (02.06.2022)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/250,182 29 September 2021 (29.09.2021) US
17/824,639 25 May 2022 (25.05.2022) US

(71) Applicant: **TENCENT AMERICA LLC** [US/US]; 2747 Park Boulevard, Palo Alto, California 94306 (US).

(72) Inventors: **CHOI, Byeongdoo**; c/o Tencent America LLC, 2747 Park Boulevard, Palo Alto, California 94306 (US). **LIU, Shan**; c/o Tencent America LLC, 2747 Park Boulevard, Palo Alto, California 94306 (US). **WENGER, Stephan**; c/o Tencent America LLC, 2747 Park Boulevard, Palo Alto, California 94306 (US). **HINDS, Arianne**; c/o Tencent America LLC, 2747 Park Boulevard, Palo Alto, California 94306 (US).

(74) Agent: **MA, Johnny**; ArentFox Schiff LLP, 1717 K Street NW, Washington, District of Columbia 20006 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI,

(54) Title: TECHNIQUES FOR SIGNALING MULTIVIEW VIEW POSITIONS IN SEI MESSAGE

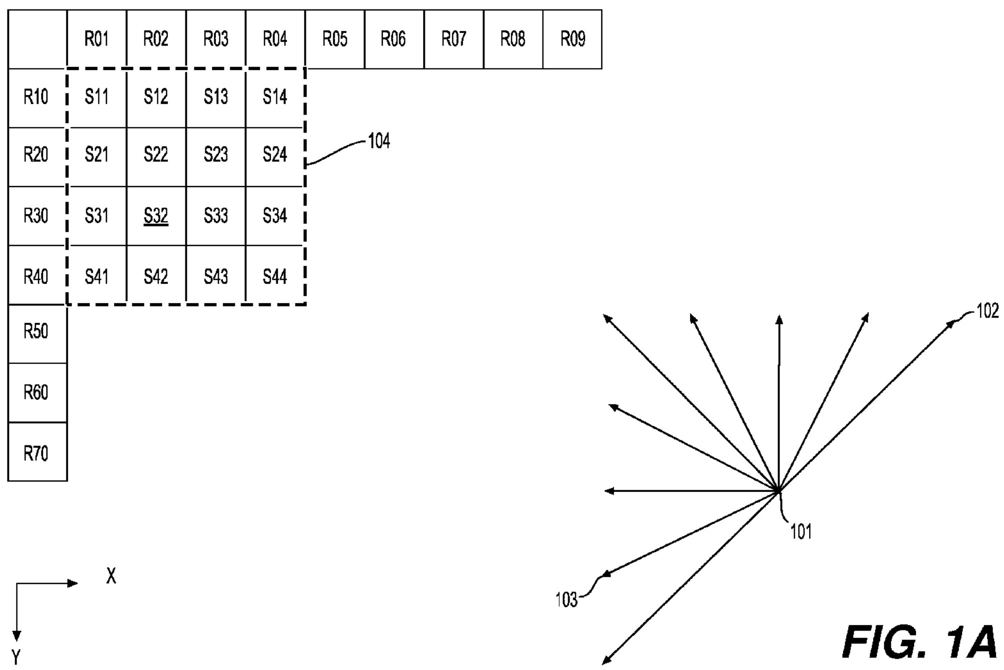

FIG. 1A

(57) Abstract: Aspects of the disclosure provide methods and apparatuses for video encoding/decoding. In some examples, an apparatus for video decoding includes processing circuitry. The processing circuitry decodes pictures associated with views in a bitstream and determines, from a supplemental enhancement information (SEI) message, positions of multidimensional coordinates in a multidimensional space respectively for the views. In an example, a position of a view is defined in the SEI message by at least a vertical view position and a horizontal view position in the multidimensional space. Further, the processing circuitry determines a rendering picture from the pictures based on a rendering view that is defined by at least the vertical view position and the horizontal view position in the multidimensional space.

WO 2023/056111 A1

[Continued on next page]

NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

TECHNIQUES FOR SIGNALING MULTIVIEW VIEW POSITIONS IN SEI MESSAGE

INCORPORATION BY REFERENCE

[0001] This present application claims the benefit of priority to U.S. Patent Application No. 17/824,639, “TECHNIQUES FOR SIGNALING MULTIVIEW VIEW POSITIONS IN SEI MESSAGE” filed on May 25, 2022, which claims the benefit of priority to U.S. Provisional Application No. 63/250,182, "TECHNIQUES FOR TWO DIMENSIONAL MULTIVIEW VIEW POSITION SEI MESSAGE FOR CODED VIDEO STREAM" filed on September 29, 2021. The disclosures of the prior applications are hereby incorporated by reference in their entirety.

TECHNICAL FIELD

[0002] The present disclosure describes embodiments generally related to video coding.

BACKGROUND

[0003] The background description provided herein is for the purpose of generally presenting the context of the disclosure. Work of the presently named inventors, to the extent the work is described in this background section, as well as aspects of the description that may not otherwise qualify as prior art at the time of filing, are neither expressly nor impliedly admitted as prior art against the present disclosure.

[0004] Uncompressed digital video can include a series of pictures, each picture having a spatial dimension of, for example, 1920 x 1080 luminance samples and associated chrominance samples. The series of pictures can have a fixed or variable picture rate (informally also known as frame rate), of, for example 60 pictures per second or 60 Hz. Uncompressed video has specific bitrate requirements. For example, 1080p60 4:2:0 video at 8 bit per sample (1920x1080 luminance sample resolution at 60 Hz frame rate) requires close to 1.5 Gbit/s bandwidth. An hour of such video requires more than 600 GBytes of storage space.

[0005] One purpose of video coding and decoding can be the reduction of redundancy in the input video signal, through compression. Compression can help reduce the aforementioned bandwidth and/or storage space requirements, in some cases by two orders of magnitude or more. Both lossless compression and lossy compression, as well as a combination thereof can be employed. Lossless compression refers to techniques where an exact copy of the original signal can be reconstructed from the compressed original signal. When using lossy compression, the reconstructed signal may not be identical to the original signal, but the distortion between original and reconstructed signals is small enough to make the reconstructed signal useful for the intended application. In the case of video, lossy compression is widely employed. The amount of distortion tolerated depends on the application; for example, users of certain consumer streaming applications may tolerate higher distortion than users of television distribution applications. The compression ratio achievable can reflect that: higher allowable/tolerable distortion can yield higher compression ratios.

[0006] A video encoder and decoder can utilize techniques from several broad categories, including, for example, motion compensation, transform, quantization, and entropy coding.

[0007] Video codec technologies can include techniques known as intra coding. In intra coding, sample values are represented without reference to samples or other data from previously reconstructed reference pictures. In some video codecs, the picture is spatially subdivided into blocks of samples. When all blocks of samples are coded in intra mode, that picture can be an intra picture. Intra pictures and their derivations such as independent decoder refresh pictures, can be used to reset the decoder state and can, therefore, be used as the first picture in a coded video bitstream and a video session, or as a still image. The samples of an intra block can be exposed to a transform, and the transform coefficients can be quantized before entropy coding. Intra prediction can be a technique that minimizes sample values in the pre-transform domain. In some cases, the smaller the DC value after a transform is, and the smaller the AC coefficients are, the fewer the bits that are required at a given quantization step size to represent the block after entropy coding.

[0008] Traditional intra coding such as known from, for example MPEG-2 generation coding technologies, does not use intra prediction. However, some newer video compression technologies include techniques that attempt, from, for example, surrounding sample data and/or metadata obtained during the encoding and/or decoding of spatially neighboring, and preceding in decoding order, blocks of data. Such techniques are henceforth called “intra prediction” techniques. Note that in at least some cases, intra prediction is using reference data only from the current picture under reconstruction and not from reference pictures.

[0009] There can be many different forms of intra prediction. When more than one of such techniques can be used in a given video coding technology, the technique in use can be coded in an intra prediction mode. In certain cases, modes can have submodes and/or parameters, and those can be coded individually or included in the mode codeword. Which codeword to use for a given mode, submode, and/or parameter combination can have an impact in the coding efficiency gain through intra prediction, and so can the entropy coding technology used to translate the codewords into a bitstream.

[0010] A certain mode of intra prediction was introduced with H.264, refined in H.265, and further refined in newer coding technologies such as joint exploration model (JEM), versatile video coding (VVC), and benchmark set (BMS). A predictor block can be formed using neighboring sample values belonging to already available samples. Sample values of neighboring samples are copied into the predictor block according to a direction. A reference to the direction in use can be coded in the bitstream or may itself be predicted.

[0011] Referring to FIG. 1A, depicted in the lower right is a subset of nine predictor directions known from H.265's 33 possible predictor directions (corresponding to the 33 angular modes of the 35 intra modes). The point where the arrows converge (101) represents the sample being predicted. The arrows represent the direction from which the sample is being predicted. For example, arrow (102) indicates that sample (101) is predicted from a sample or samples to the upper right, at a 45 degree angle from the horizontal. Similarly, arrow (103) indicates that sample (101) is predicted from a sample or samples to the lower left of sample (101), in a 22.5 degree angle from the horizontal.

[0012] Still referring to FIG. 1A, on the top left there is depicted a square block (104) of 4 x 4 samples (indicated by a dashed, boldface line). The square block (104) includes 16 samples, each labelled with an "S", its position in the Y dimension (e.g., row index) and its position in the X dimension (e.g., column index). For example, sample S21 is the second sample in the Y dimension (from the top) and the first (from the left) sample in the X dimension. Similarly, sample S44 is the fourth sample in block (104) in both the Y and X dimensions. As the block is 4 x 4 samples in size, S44 is at the bottom right. Further shown are reference samples that follow a similar numbering scheme. A reference sample is labelled with an R, its Y position (e.g., row index) and X position (column index) relative to block (104). In both H.264 and H.265, prediction samples neighbor the block under reconstruction; therefore no negative values need to be used.

[0013] Intra picture prediction can work by copying reference sample values from the neighboring samples as appropriated by the signaled prediction direction. For example, assume the coded video bitstream includes signaling that, for this block, indicates a prediction direction consistent with arrow (102)—that is, samples are predicted from a prediction sample or samples to the upper right, at a 45 degree angle from the horizontal. In that case, samples S41, S32, S23, and S14 are predicted from the same reference sample R05. Sample S44 is then predicted from reference sample R08.

[0014] In certain cases, the values of multiple reference samples may be combined, for example through interpolation, in order to calculate a reference sample; especially when the directions are not evenly divisible by 45 degrees.

[0015] The number of possible directions has increased as video coding technology has developed. In H.264 (year 2003), nine different direction could be represented. That increased to 33 in H.265 (year 2013), and JEM/VVC/BMS, at the time of disclosure, can support up to 65 directions. Experiments have been conducted to identify the most likely directions, and certain techniques in the entropy coding are used to represent those likely directions in a small number of bits, accepting a certain penalty for less likely directions. Further, the directions themselves can sometimes be predicted from neighboring directions used in neighboring, already decoded, blocks.

[0016] FIG. 1B shows a schematic (110) that depicts 65 intra prediction directions according to JEM to illustrate the increasing number of prediction directions over time.

[0017] The mapping of intra prediction directions bits in the coded video bitstream that represent the direction can be different from video coding technology to video coding technology; and can range, for example, from simple direct mappings of prediction direction to intra prediction mode, to codewords, to complex adaptive schemes involving most probable modes, and similar techniques. In all cases, however, there can be certain directions that are statistically less likely to occur in video content than certain other directions. As the goal of video compression is the reduction of redundancy, those less likely directions will, in a well working video coding technology, be represented by a larger number of bits than more likely directions.

[0018] Video coding and decoding can be performed using inter-picture prediction with motion compensation. Motion compensation can be a lossy compression technique and can relate to techniques where a block of sample data from a previously reconstructed picture or part thereof (reference picture), after being spatially shifted in a direction indicated by a motion vector (MV henceforth), is used for the prediction of a newly reconstructed picture or picture part. In some cases, the reference picture can be the same as the picture currently under reconstruction. MVs can have two dimensions X and Y, or three dimensions, the third being an indication of the reference picture in use (the latter, indirectly, can be a time dimension).

[0019] In some video compression techniques, an MV applicable to a certain area of sample data can be predicted from other MVs, for example from those related to another area of sample data spatially adjacent to the area under reconstruction, and preceding that MV in decoding order. Doing so can substantially reduce the amount of data required for coding the MV, thereby removing redundancy and increasing compression. MV prediction can work effectively, for example, because when coding an input video signal derived from a camera (known as natural video) there is a statistical likelihood that areas larger than the area to which a single MV is applicable move in a similar direction and, therefore, can in some cases be predicted using a similar motion vector derived from MVs of neighboring area. That results in the MV found for a given area to be similar or the same as the MV predicted from the surrounding MVs, and that in turn can be represented, after entropy coding, in a smaller number of bits than what would be used if coding the MV directly. In some cases, MV prediction can be an example of lossless compression of a signal (namely: the MVs) derived from the original signal (namely: the sample stream). In other cases, MV prediction itself can be lossy, for example because of rounding errors when calculating a predictor from several surrounding MVs.

[0020] Various MV prediction mechanisms are described in H.265/HEVC (ITU-T Rec. H.265, "High Efficiency Video Coding", December 2016). Out of the many MV prediction mechanisms that H.265 offers, described here is a technique henceforth referred to as "spatial merge".

[0021] Referring to FIG. 2, a current block (201) comprises samples that have been found by the encoder during the motion search process to be predictable from a previous block of the same size that has been spatially shifted. Instead of coding that MV directly, the MV can be derived from metadata associated with one or more reference pictures, for example from the most recent (in decoding order) reference picture, using the MV associated with either one of five surrounding samples, denoted A0, A1, and B0, B1, B2 (202 through 206, respectively). In H.265, the MV prediction can use predictors from the same reference picture that the neighboring block is using.

SUMMARY

[0022] Aspects of the disclosure provide methods and apparatuses for video encoding/decoding. In some examples, an apparatus for video decoding includes processing circuitry. The processing circuitry receives, from a bitstream, pictures associated with a plurality of views. The processing circuitry decodes the pictures associated with the plurality of views, and determines, from a supplemental enhancement information (SEI) message in the bitstream, positions of multidimensional coordinates in a multidimensional space respectively for the plurality of views. The positions includes at least both a vertical view position and a horizontal view position. The processing circuitry renders the pictures based

on a rendering view corresponding to the vertical view position and the horizontal view position. The processing circuitry reorders the pictures based on the rendered pictures and the positions of multidimensional coordinates in the multidimensional space.

[0023] In some embodiments, the processing circuitry determines, from the SEI message, positions of two dimensional coordinates in a two dimensional space for the views. In some examples, the processing circuitry obtains, from the SEI message, a first value indicative of a number of views from the SEI message; a second value indicative of a first number of positions in a vertical dimension of the two dimensional space; and a third value indicative of a second number of positions in a horizontal dimension of the two dimensional space. Further, the processing circuitry obtains, for each view of the number of views, a first coordinate value of a first dimension and a second coordinate value of a second dimension as a view position in the two dimensional space.

[0024] In some examples, the SEI message is associated with an intra random access point (IRAP) access unit of a coded video sequence.

[0025] In some examples, the SEI message is not in another SEI message.

[0026] In some examples, the SEI message is in a coded video sequence carried by the bitstream, and the positions of the multidimensional coordinates for the views are applied to access units in the coded video sequence.

[0027] In some examples, the SEI message is denoted as a multiview view position (MVP) SEI message, and the coded video sequence includes a scalability dimension information (SDI) SEI message. The processing circuitry can derive a first value indicative of a number of the views from the SDI SEI message, and obtain a second value associated with the number from the MVP SEI message. Further, the processing circuitry can compare one plus the second value with the first value in a conformance check.

[0028] Aspects of the disclosure also provide a non-transitory computer-readable medium storing instructions which when executed by a computer for video decoding cause the computer to perform the method for video decoding.

BRIEF DESCRIPTION OF THE DRAWINGS

[0029] Further features, the nature, and various advantages of the disclosed subject matter will be more apparent from the following detailed description and the accompanying drawings in which:

[0030] FIG. 1A is a schematic illustration of an exemplary subset of intra prediction modes.

[0031] FIG. 1B is an illustration of exemplary intra prediction directions.

[0032] FIG. 2 is a schematic illustration of a current block and its surrounding spatial merge candidates in one example.

[0033] FIG. 3 is a schematic illustration of a simplified block diagram of a communication system (300) in accordance with an embodiment.

[0034] FIG. 4 is a schematic illustration of a simplified block diagram of a communication system (400) in accordance with an embodiment.

[0035] FIG. 5 is a schematic illustration of a simplified block diagram of a decoder in accordance with an embodiment.

[0036] FIG. 6 is a schematic illustration of a simplified block diagram of an encoder in accordance with an embodiment.

[0037] FIG. 7 shows a block diagram of an encoder in accordance with another embodiment.

[0038] FIG. 8 shows a block diagram of a decoder in accordance with another embodiment.

[0039] FIG. 9 shows a diagram of auto-stereoscopic display in some examples.

[0040] FIGs. 10A-B show an example of reordering pictures according to multiview view positions in an example.

[0041] FIG. 11 shows a syntax example in a supplemental enhancement information (SEI) message for indicating view positions for a multiview video.

[0042] FIG. 12 shows a diagram of auto-stereoscopic display in some examples.

[0043] FIG. 13 shows an example of associating pictures according to view positions in an example.

[0044] FIG. 14 shows a syntax example in an SEI message for indicating view positions in two dimensional matrix for a multiview video.

[0045] FIG. 15 shows a flow chart outlining a process according to some embodiment of the disclosure.

[0046] FIG. 16 shows a flow chart outlining another process according to some embodiment of the disclosure.

[0047] FIG. 17 is a schematic illustration of a computer system in accordance with an embodiment.

DETAILED DESCRIPTION OF EMBODIMENTS

[0048] FIG. 3 illustrates a simplified block diagram of a communication system (300) according to an embodiment of the present disclosure. The communication system (300) includes a plurality of terminal devices that can communicate with each other, via, for example, a network (350). For example, the communication system (300) includes a first pair of terminal devices (310) and (320) interconnected via the network (350). In the FIG. 3 example, the first pair of terminal devices (310) and (320) performs unidirectional transmission of data. For example, the terminal device (310) may code video data (e.g., a stream of video pictures that are captured by the terminal device (310)) for transmission to the other terminal device (320) via the network (350). The encoded video data can be transmitted in the form of one or more coded video bitstreams. The terminal device (320) may receive the coded video data from the network (350), decode the coded video data to recover the video pictures and display video pictures according to the recovered video data. Unidirectional data transmission may be common in media serving applications and the like.

[0049] In another example, the communication system (300) includes a second pair of terminal devices (330) and (340) that performs bidirectional transmission of coded video data that may occur, for example, during videoconferencing. For bidirectional transmission of data, in an example, each terminal device of the terminal devices (330) and (340) may code video data (e.g., a stream of video pictures that are captured by the terminal device) for transmission to the other terminal device of the terminal devices (330) and (340) via the network (350). Each terminal device of the terminal devices (330) and (340) also may receive the coded video data transmitted by the other terminal device of the terminal devices (330) and (340), and may decode the coded video data to recover the video pictures and may display video pictures at an accessible display device according to the recovered video data.

[0050] In the FIG. 3 example, the terminal devices (310), (320), (330) and (340) may be illustrated as servers, personal computers and smart phones but the principles of the present disclosure may be not so limited. Embodiments of the present disclosure find application with laptop computers, tablet computers, media players and/or dedicated video conferencing equipment. The network (350) represents any number of networks that convey coded video data among the terminal devices (310), (320), (330) and (340), including for example wireline (wired) and/or wireless communication networks. The communication network (350) may exchange data in circuit-switched and/or packet-switched channels. Representative networks include telecommunications networks, local area networks, wide area networks and/or the Internet. For the purposes of the present discussion, the architecture and topology of the network (350) may be immaterial to the operation of the present disclosure unless explained herein below.

[0051] FIG. 4 illustrates, as an example for an application for the disclosed subject matter, the placement of a video encoder and a video decoder in a streaming environment.

The disclosed subject matter can be equally applicable to other video enabled applications, including, for example, video conferencing, digital TV, storing of compressed video on digital media including CD, DVD, memory stick and the like, and so on.

[0052] A streaming system may include a capture subsystem (413), that can include a video source (401), for example a digital camera, creating for example a stream of video pictures (402) that are uncompressed. In an example, the stream of video pictures (402) includes samples that are taken by the digital camera. The stream of video pictures (402), depicted as a bold line to emphasize a high data volume when compared to encoded video data (404) (or coded video bitstreams), can be processed by an electronic device (420) that includes a video encoder (403) coupled to the video source (401). The video encoder (403) can include hardware, software, or a combination thereof to enable or implement aspects of the disclosed subject matter as described in more detail below. The encoded video data (404) (or encoded video bitstream (404)), depicted as a thin line to emphasize the lower data volume when compared to the stream of video pictures (402), can be stored on a streaming server (405) for future use. One or more streaming client subsystems, such as client subsystems (406) and (408) in FIG. 4 can access the streaming server (405) to retrieve copies (407) and (409) of the encoded video data (404). A client subsystem (406) can include a video decoder (410), for example, in an electronic device (430). The video decoder (410) decodes the incoming copy (407) of the encoded video data and creates an outgoing stream of video pictures (411) that can be rendered on a display (412) (e.g., display screen) or other rendering device (not depicted). In some streaming systems, the encoded video data (404), (407), and (409) (e.g., video bitstreams) can be encoded according to certain video coding/compression standards. Examples of those standards include ITU-T Recommendation H.265. In an example, a video coding standard under development is informally known as Versatile Video Coding (VVC). The disclosed subject matter may be used in the context of VVC.

[0053] It is noted that the electronic devices (420) and (430) can include other components (not shown). For example, the electronic device (420) can include a video decoder (not shown) and the electronic device (430) can include a video encoder (not shown) as well.

[0054] FIG. 5 shows a block diagram of a video decoder (510) according to an embodiment of the present disclosure. The video decoder (510) can be included in an electronic device (530). The electronic device (530) can include a receiver (531) (e.g., receiving circuitry). The video decoder (510) can be used in the place of the video decoder (410) in the FIG. 4 example.

[0055] The receiver (531) may receive one or more coded video sequences to be decoded by the video decoder (510); in the same or another embodiment, one coded video sequence at a time, where the decoding of each coded video sequence is independent from other coded video sequences. The coded video sequence may be received from a channel (501), which may be a hardware/software link to a storage device which stores the encoded video data. The receiver (531) may receive the encoded video data with other data, for example, coded audio data and/or ancillary data streams, that may be forwarded to their respective using entities (not depicted). The receiver (531) may separate the coded video sequence from the other data. To combat network jitter, a buffer memory (515) may be coupled in between the receiver (531) and an entropy decoder / parser (520) ("parser (520)" henceforth). In certain applications, the buffer memory (515) is part of the video decoder (510). In others, it can be outside of the video decoder (510) (not depicted). In still others, there can be a buffer memory (not depicted) outside of the video decoder (510), for example to combat network jitter, and in addition another buffer memory (515) inside the video decoder (510), for example to handle playout timing. When the receiver (531) is receiving data from a store/forward device of sufficient bandwidth and controllability, or from an isosynchronous network, the buffer memory (515) may not be needed, or can be small. For use on best effort packet networks such as the Internet, the buffer memory (515) may be required, can be comparatively large and can be advantageously of adaptive size, and may at least partially be implemented in an operating system or similar elements (not depicted) outside of the video decoder (510).

[0056] The video decoder (510) may include the parser (520) to reconstruct symbols (521) from the coded video sequence. Categories of those symbols include information used to manage operation of the video decoder (510), and potentially information to control a rendering device such as a render device (512) (e.g., a display screen) that is not an integral part of the electronic device (530) but can be coupled to the electronic device (530), as was shown in FIG. 5. The control information for the rendering device(s) may be in the form of supplemental enhancement information (SEI messages) or Video Usability Information (VUI) parameter set fragments (not depicted). The parser (520) may parse / entropy-decode the coded video sequence that is received. The coding of the coded video sequence can be in accordance with a video coding technology or standard, and can follow various principles, including variable length coding, Huffman coding, arithmetic coding with or without context sensitivity, and so forth. The parser (520) may extract from the coded video sequence, a set of subgroup parameters for at least one of the subgroups of pixels in the video decoder, based upon at least one parameter corresponding to the group. Subgroups can include Groups of Pictures (GOPs), pictures, tiles, slices, macroblocks, Coding Units (CUs), blocks, Transform Units (TUs), Prediction Units (PUs) and so forth. The parser (520) may also extract from the coded video sequence information such as transform coefficients, quantizer parameter values, motion vectors, and so forth.

[0057] The parser (520) may perform an entropy decoding / parsing operation on the video sequence received from the buffer memory (515), so as to create symbols (521).

[0058] Reconstruction of the symbols (521) can involve multiple different units depending on the type of the coded video picture or parts thereof (such as: inter and intra picture, inter and intra block), and other factors. Which units are involved, and how, can be controlled by the subgroup control information that was parsed from the coded video sequence by the parser (520). The flow of such subgroup control information between the parser (520) and the multiple units below is not depicted for clarity.

[0059] Beyond the functional blocks already mentioned, the video decoder (510) can be conceptually subdivided into a number of functional units as described below. In a practical implementation operating under commercial constraints, many of these units interact closely with each other and can, at least partly, be integrated into each other. However, for the purpose of describing the disclosed subject matter, the conceptual subdivision into the functional units below is appropriate.

[0060] A first unit is the scaler / inverse transform unit (551). The scaler / inverse transform unit (551) receives a quantized transform coefficient as well as control information, including which transform to use, block size, quantization factor, quantization scaling matrices, etc. as symbol(s) (521) from the parser (520). The scaler / inverse transform unit (551) can output blocks comprising sample values, that can be input into aggregator (555).

[0061] In some cases, the output samples of the scaler / inverse transform (551) can pertain to an intra coded block; that is: a block that is not using predictive information from previously reconstructed pictures, but can use predictive information from previously reconstructed parts of the current picture. Such predictive information can be provided by an intra picture prediction unit (552). In some cases, the intra picture prediction unit (552) generates a block of the same size and shape of the block under reconstruction, using surrounding already reconstructed information fetched from the current picture buffer (558). The current picture buffer (558) buffers, for example, partly reconstructed current picture and/or fully reconstructed current picture. The aggregator (555), in some cases, adds, on a per sample basis, the prediction information the intra prediction unit (552) has generated to the output sample information as provided by the scaler / inverse transform unit (551).

[0062] In other cases, the output samples of the scaler / inverse transform unit (551) can pertain to an inter coded, and potentially motion compensated block. In such a case, a motion compensation prediction unit (553) can access reference picture memory (557) to fetch samples used for prediction. After motion compensating the fetched samples in accordance with the symbols (521) pertaining to the block, these samples can be added by the aggregator (555) to the output of the scaler / inverse transform unit (551) (in this case called the residual samples or residual signal) so as to generate output sample information. The addresses within the reference picture memory (557) from where the motion compensation prediction unit (553) fetches prediction samples can be controlled by motion vectors, available to the motion compensation prediction unit (553) in the form of symbols (521) that can have, for example X, Y, and reference picture components. Motion compensation also can include interpolation of sample values as fetched from the reference picture memory (557) when sub-sample exact motion vectors are in use, motion vector prediction mechanisms, and so forth.

[0063] The output samples of the aggregator (555) can be subject to various loop filtering techniques in the loop filter unit (556). Video compression technologies can include in-loop filter technologies that are controlled by parameters included in the coded video sequence (also referred to as coded video bitstream) and made available to the loop filter unit (556) as symbols (521) from the parser (520), but can also be responsive to meta-information obtained during the decoding of previous (in decoding order) parts of the coded picture or coded video sequence, as well as responsive to previously reconstructed and loop-filtered sample values.

[0064] The output of the loop filter unit (556) can be a sample stream that can be output to the render device (512) as well as stored in the reference picture memory (557) for use in future inter-picture prediction.

[0065] Certain coded pictures, once fully reconstructed, can be used as reference pictures for future prediction. For example, once a coded picture corresponding to a current picture is fully reconstructed and the coded picture has been identified as a reference picture (by, for example, the parser (520)), the current picture buffer (558) can become a part of the reference picture memory (557), and a fresh current picture buffer can be reallocated before commencing the reconstruction of the following coded picture.

[0066] The video decoder (510) may perform decoding operations according to a predetermined video compression technology in a standard, such as ITU-T Rec. H.265. The coded video sequence may conform to a syntax specified by the video compression technology or standard being used, in the sense that the coded video sequence adheres to both the syntax of the video compression technology or standard and the profiles as documented in the video compression technology or standard. Specifically, a profile can select certain tools as the only tools available for use under that profile from all the tools available in the video compression technology or standard. Also necessary for compliance can be that the complexity of the coded video sequence is within bounds as defined by the level of the video compression technology or standard. In some cases, levels restrict the maximum picture size, maximum frame rate, maximum reconstruction sample rate (measured in, for example megasamples per second), maximum reference picture size, and so on. Limits set by levels can, in some cases, be further restricted through Hypothetical Reference Decoder (HRD) specifications and metadata for HRD buffer management signaled in the coded video sequence.

[0067] In an embodiment, the receiver (531) may receive additional (redundant) data with the encoded video. The additional data may be included as part of the coded video sequence(s). The additional data may be used by the video decoder (510) to properly decode the data and/or to more accurately reconstruct the original video data. Additional data can be in the form of, for example, temporal, spatial, or signal noise ratio (SNR) enhancement layers, redundant slices, redundant pictures, forward error correction codes, and so on.

[0068] FIG. 6 shows a block diagram of a video encoder (603) according to an embodiment of the present disclosure. The video encoder (603) is included in an electronic device (620). The electronic device (620) includes a transmitter (640) (e.g., transmitting circuitry). The video encoder (603) can be used in the place of the video encoder (403) in the FIG. 4 example.

[0069] The video encoder (603) may receive video samples from a video source (601) (that is not part of the electronic device (620) in the FIG. 6 example) that may capture video image(s) to be coded by the video encoder (603). In another example, the video source (601) is a part of the electronic device (620).

[0070] The video source (601) may provide the source video sequence to be coded by the video encoder (603) in the form of a digital video sample stream that can be of any suitable bit depth (for example: 8 bit, 10 bit, 12 bit, ...), any colorspace (for example, BT.601 Y CrCB, RGB, ...), and any suitable sampling structure (for example Y CrCb 4:2:0, Y CrCb

4:4:4). In a media serving system, the video source (601) may be a storage device storing previously prepared video. In a videoconferencing system, the video source (601) may be a camera that captures local image information as a video sequence. Video data may be provided as a plurality of individual pictures that impart motion when viewed in sequence. The pictures themselves may be organized as a spatial array of pixels, wherein each pixel can comprise one or more samples depending on the sampling structure, color space, etc. in use. A person skilled in the art can readily understand the relationship between pixels and samples. The description below focuses on samples.

[0071] According to an embodiment, the video encoder (603) may code and compress the pictures of the source video sequence into a coded video sequence (643) in real time or under any other time constraints as required by the application. Enforcing appropriate coding speed is one function of a controller (650). In some embodiments, the controller (650) controls other functional units as described below and is functionally coupled to the other functional units. The coupling is not depicted for clarity. Parameters set by the controller (650) can include rate control related parameters (picture skip, quantizer, lambda value of rate-distortion optimization techniques, ...), picture size, group of pictures (GOP) layout, maximum motion vector search range, and so forth. The controller (650) can be configured to have other suitable functions that pertain to the video encoder (603) optimized for a certain system design.

[0072] In some embodiments, the video encoder (603) is configured to operate in a coding loop. As an oversimplified description, in an example, the coding loop can include a source coder (630) (e.g., responsible for creating symbols, such as a symbol stream, based on an input picture to be coded, and a reference picture(s)), and a (local) decoder (633) embedded in the video encoder (603). The decoder (633) reconstructs the symbols to create the sample data in a similar manner as a (remote) decoder also would create (as any compression between symbols and coded video bitstream is lossless in the video compression technologies considered in the disclosed subject matter). The reconstructed sample stream (sample data) is input to the reference picture memory (634). As the decoding of a symbol stream leads to bit-exact results independent of decoder location (local or remote), the content in the reference picture memory (634) is also bit exact between the local encoder and remote encoder. In other words, the prediction part of an encoder "sees" as reference picture samples exactly the same sample values as a decoder would "see" when using prediction during decoding. This fundamental principle of reference picture synchronicity (and resulting drift, if synchronicity cannot be maintained, for example because of channel errors) is used in some related arts as well.

[0073] The operation of the "local" decoder (633) can be the same as of a "remote" decoder, such as the video decoder (510), which has already been described in detail above in conjunction with FIG. 5. Briefly referring also to FIG. 5, however, as symbols are available and encoding/decoding of symbols to a coded video sequence by an entropy coder (645) and the parser (520) can be lossless, the entropy decoding parts of the video decoder (510), including the buffer memory (515), and parser (520) may not be fully implemented in the local decoder (633).

[0074] An observation that can be made at this point is that any decoder technology except the parsing/entropy decoding that is present in a decoder also necessarily needs to be present, in substantially identical functional form, in a corresponding encoder. For this reason, the disclosed subject matter focuses on decoder operation. The description of encoder technologies can be abbreviated as they are the inverse of the comprehensively described decoder technologies. Only in certain areas a more detail description is required and provided below.

[0075] During operation, in some examples, the source coder (630) may perform motion compensated predictive coding, which codes an input picture predictively with reference to one or more previously coded picture from the video sequence that were designated as "reference pictures." In this manner, the coding engine (632) codes differences between pixel blocks of an input picture and pixel blocks of reference picture(s) that may be selected as prediction reference(s) to the input picture.

[0076] The local video decoder (633) may decode coded video data of pictures that may be designated as reference pictures, based on symbols created by the source coder (630). Operations of the coding engine (632) may advantageously be lossy processes. When the coded video data may be decoded at a video decoder (not shown in FIG. 6), the reconstructed video sequence typically may be a replica of the source video sequence with some errors. The local video decoder (633) replicates decoding processes that may be performed by the video decoder on reference pictures and may cause reconstructed reference pictures to be stored in the reference picture cache (634). In this manner, the video encoder (603) may store copies of reconstructed reference pictures locally that have common content as the reconstructed reference pictures that will be obtained by a far-end video decoder (absent transmission errors).

[0077] The predictor (635) may perform prediction searches for the coding engine (632). That is, for a new picture to be coded, the predictor (635) may search the reference picture memory (634) for sample data (as candidate reference pixel blocks) or certain metadata such as reference picture motion vectors, block shapes, and so on, that may serve as an appropriate prediction reference for the new pictures. The predictor (635) may operate on a sample block-by-pixel block basis to find appropriate prediction references. In some cases, as determined by search results obtained by the predictor (635), an input picture may have prediction references drawn from multiple reference pictures stored in the reference picture memory (634).

[0078] The controller (650) may manage coding operations of the source coder (630), including, for example, setting of parameters and subgroup parameters used for encoding the video data.

[0079] Output of all aforementioned functional units may be subjected to entropy coding in the entropy coder (645). The entropy coder (645) translates the symbols as generated by the various functional units into a coded video sequence, by lossless compressing the symbols according to technologies such as Huffman coding, variable length coding, arithmetic coding, and so forth.

[0080] The transmitter (640) may buffer the coded video sequence(s) as created by the entropy coder (645) to prepare for transmission via a communication channel (660), which may be a hardware/software link to a storage device which would store the encoded video data. The transmitter (640) may merge coded video data from the video coder (603) with other data to be transmitted, for example, coded audio data and/or ancillary data streams (sources not shown).

[0081] The controller (650) may manage operation of the video encoder (603). During coding, the controller (650) may assign to each coded picture a certain coded picture type, which may affect the coding techniques that may be applied to the respective picture. For example, pictures often may be assigned as one of the following picture types:

[0082] An Intra Picture (I picture) may be one that may be coded and decoded without using any other picture in the sequence as a source of prediction. Some video codecs allow for different types of intra pictures, including, for example Independent Decoder Refresh ("IDR") Pictures. A person skilled in the art is aware of those variants of I pictures and their respective applications and features.

[0083] A predictive picture (P picture) may be one that may be coded and decoded using intra prediction or inter prediction using at most one motion vector and reference index to predict the sample values of each block.

[0084] A bi-directionally predictive picture (B Picture) may be one that may be coded and decoded using intra prediction or inter prediction using at most two motion vectors and reference indices to predict the sample values of each block. Similarly, multiple-predictive pictures can use more than two reference pictures and associated metadata for the reconstruction of a single block.

[0085] Source pictures commonly may be subdivided spatially into a plurality of sample blocks (for example, blocks of 4x4, 8x8, 4x8, or 16x16 samples each) and coded on a block-by-block basis. Blocks may be coded predictively with reference to other (already coded) blocks as determined by the coding assignment applied to the blocks' respective pictures. For example, blocks of I pictures may be coded non-predictively or they may be coded predictively with reference to already coded blocks of the same picture (spatial prediction or intra prediction). Pixel blocks of P pictures may be coded predictively, via spatial prediction or via temporal prediction with reference to one previously coded reference picture. Blocks of B pictures may be coded predictively, via spatial prediction or via temporal prediction with reference to one or two previously coded reference pictures.

[0086] The video encoder (603) may perform coding operations according to a predetermined video coding technology or standard, such as ITU-T Rec. H.265. In its operation, the video encoder (603) may perform various compression operations, including predictive coding operations that exploit temporal and spatial redundancies in the input video sequence. The coded video data, therefore, may conform to a syntax specified by the video coding technology or standard being used.

[0087] In an embodiment, the transmitter (640) may transmit additional data with the encoded video. The source coder (630) may include such data as part of the coded video sequence. Additional data may comprise temporal/spatial/SNR enhancement layers, other forms of redundant data such as redundant pictures and slices, SEI messages, VUI parameter set fragments, and so on.

[0088] A video may be captured as a plurality of source pictures (video pictures) in a temporal sequence. Intra-picture prediction (often abbreviated to intra prediction) makes use of spatial correlation in a given picture, and inter-picture prediction makes uses of the (temporal or other) correlation between the pictures. In an example, a specific picture under encoding/decoding, which is referred to as a current picture, is partitioned into blocks. When a block in the current picture is similar to a reference block in a previously coded and still buffered reference picture in the video, the block in the current picture can be coded by a vector that is referred to as a motion vector. The motion vector points to the reference block in the reference picture, and can have a third dimension identifying the reference picture, in case multiple reference pictures are in use.

[0089] In some embodiments, a bi-prediction technique can be used in the inter-picture prediction. According to the bi-prediction technique, two reference pictures, such as a first reference picture and a second reference picture that are both prior in decoding order to the current picture in the video (but may be in the past and future, respectively, in display order) are used. A block in the current picture can be coded by a first motion vector that points to a first reference block in the first reference picture, and a second motion vector that points to a second reference block in the second reference picture. The block can be predicted by a combination of the first reference block and the second reference block.

[0090] Further, a merge mode technique can be used in the inter-picture prediction to improve coding efficiency.

[0091] According to some embodiments of the disclosure, predictions, such as inter-picture predictions and intra-picture predictions are performed in the unit of blocks. For example, according to the HEVC standard, a picture in a sequence of video pictures is partitioned into coding tree units (CTU) for compression, the CTUs in a picture have the same size, such as 64x64 pixels, 32x32 pixels, or 16x16 pixels. In general, a CTU includes three coding tree blocks (CTBs), which are one luma CTB and two chroma CTBs. Each CTU can be recursively quadtree split into one or multiple coding units (CUs). For example, a CTU of 64x64 pixels can be split into one CU of 64x64 pixels, or 4 CUs of 32x32 pixels, or 16 CUs of 16x16 pixels. In an example, each CU is analyzed to determine a prediction type for the CU, such as an inter prediction type or an intra prediction type. The CU is split into one or more prediction units (PUs) depending on the temporal and/or spatial predictability. Generally, each PU includes a luma prediction block (PB), and two chroma PBs. In an embodiment, a prediction operation in coding (encoding/decoding) is performed in the unit of a prediction block. Using a luma prediction block as an example of a prediction block, the prediction block includes a matrix of values (e.g., luma values) for pixels, such as 8x8 pixels, 16x16 pixels, 8x16 pixels, 16x8 pixels, and the like.

[0092] FIG. 7 shows a diagram of a video encoder (703) according to another embodiment of the disclosure. The video encoder (703) is configured to receive a processing block (e.g., a prediction block) of sample values within a current video picture in a sequence of video pictures, and encode the processing block into a coded picture that is part of a coded video sequence. In an example, the video encoder (703) is used in the place of the video encoder (403) in the FIG. 4 example.

[0093] In an HEVC example, the video encoder (703) receives a matrix of sample values for a processing block, such as a prediction block of 8x8 samples, and the like. The video encoder (703) determines whether the processing block is best coded using intra mode, inter mode, or bi-prediction mode using, for example, rate-distortion optimization. When the processing block is to be coded in intra mode, the video encoder (703) may use an intra prediction technique to encode the processing block into the coded picture; and when the processing block is to be coded in inter mode or bi-prediction mode, the video encoder (703) may use an inter prediction or bi-prediction technique, respectively, to encode the processing block into the coded picture. In certain video coding technologies, merge mode can be an inter picture prediction submode where the motion vector is derived from one or more motion vector predictors without the benefit of a coded motion vector component outside the predictors. In certain other video coding technologies, a motion vector component applicable to the subject block may be present. In an example, the video encoder (703) includes other components, such as a mode decision module (not shown) to determine the mode of the processing blocks.

[0094] In the FIG. 7 example, the video encoder (703) includes the inter encoder (730), an intra encoder (722), a residue calculator (723), a switch (726), a residue encoder (724), a general controller (721), and an entropy encoder (725) coupled together as shown in FIG. 7.

[0095] The inter encoder (730) is configured to receive the samples of the current block (e.g., a processing block), compare the block to one or more reference blocks in reference pictures (e.g., blocks in previous pictures and later pictures), generate inter prediction information (e.g., description of redundant information according to inter encoding technique, motion vectors, merge mode information), and calculate inter prediction results (e.g., predicted block) based on the inter prediction information using any suitable technique. In some examples, the reference pictures are decoded reference pictures that are decoded based on the encoded video information.

[0096] The intra encoder (722) is configured to receive the samples of the current block (e.g., a processing block), in some cases compare the block to blocks already coded in the same picture, generate quantized coefficients after transform, and in some cases also intra prediction information (e.g., an intra prediction direction information according to one or more intra encoding techniques). In an example, the intra encoder (722) also calculates intra prediction results (e.g., predicted block) based on the intra prediction information and reference blocks in the same picture.

[0097] The general controller (721) is configured to determine general control data and control other components of the video encoder (703) based on the general control data. In an example, the general controller (721) determines the mode of the block, and provides a control signal to the switch (726) based on the mode. For example, when the mode is the intra mode, the general controller (721) controls the switch (726) to select the intra mode result for use by the residue calculator (723), and controls the entropy encoder (725) to select the intra prediction information and include the intra prediction information in the bitstream; and when the mode is the inter mode, the general controller (721) controls the switch (726) to select the inter prediction result for use by the residue calculator (723), and controls the entropy encoder (725) to select the inter prediction information and include the inter prediction information in the bitstream.

[0098] The residue calculator (723) is configured to calculate a difference (residue data) between the received block and prediction results selected from the intra encoder (722) or the inter encoder (730). The residue encoder (724) is configured to operate based on the residue data to encode the residue data to generate the transform coefficients. In an example, the residue encoder (724) is configured to convert the residue data from a spatial domain to a frequency domain, and generate the transform coefficients. The transform coefficients are then subject to quantization processing to obtain quantized transform coefficients. In various embodiments, the video encoder (703) also includes a residue decoder (728). The residue decoder (728) is configured to perform inverse-transform, and generate the decoded residue data. The decoded residue data can be suitably used by the intra encoder (722) and the inter encoder (730). For example, the inter encoder (730) can generate decoded blocks based on the decoded residue data and inter prediction information, and the intra encoder (722) can generate decoded blocks based on the decoded residue data and the intra prediction information. The decoded blocks are suitably processed to generate decoded pictures and the decoded pictures can be buffered in a memory circuit (not shown) and used as reference pictures in some examples.

[0099] The entropy encoder (725) is configured to format the bitstream to include the encoded block. The entropy encoder (725) is configured to include various information according to a suitable standard, such as the HEVC standard. In an example, the entropy encoder (725) is configured to include the general control data, the selected prediction information (e.g., intra prediction information or inter prediction information), the residue information, and other suitable information in the bitstream. Note that, according to the disclosed subject matter, when coding a block in the merge submode of either inter mode or bi-prediction mode, there is no residue information.

[0100] FIG. 8 shows a diagram of a video decoder (810) according to another embodiment of the disclosure. The video decoder (810) is configured to receive coded pictures that are part of a coded video sequence, and decode the coded pictures to generate reconstructed pictures. In an example, the video decoder (810) is used in the place of the video decoder (410) in the FIG. 4 example.

[0101] In the FIG. 8 example, the video decoder (810) includes an entropy decoder (871), an inter decoder (880), a residue decoder (873), a reconstruction module (874), and an intra decoder (872) coupled together as shown in FIG. 8.

[0102] The entropy decoder (871) can be configured to reconstruct, from the coded picture, certain symbols that represent the syntax elements of which the coded picture is made up. Such symbols can include, for example, the mode in which a block is coded (such as, for example, intra mode, inter mode, bi-predicted mode, the latter two in merge submode or another submode), prediction information (such as, for example, intra prediction information or inter prediction information) that can identify certain sample or metadata that is used for prediction by the intra decoder (872) or the inter decoder (880), respectively, residual information in the form of, for example, quantized transform coefficients, and the like. In an example, when the prediction mode is inter or bi-predicted mode, the inter prediction information is provided to the inter decoder (880); and when the prediction type is the intra prediction type, the intra prediction information is provided to the intra decoder (872). The residual information can be subject to inverse quantization and is provided to the residue decoder (873).

[0103] The inter decoder (880) is configured to receive the inter prediction information, and generate inter prediction results based on the inter prediction information.

[0104] The intra decoder (872) is configured to receive the intra prediction information, and generate prediction results based on the intra prediction information.

[0105] The residue decoder (873) is configured to perform inverse quantization to extract de-quantized transform coefficients, and process the de-quantized transform coefficients to convert the residual from the frequency domain to the spatial domain. The residue decoder (873) may also require certain control information (to include the Quantizer

Parameter (QP)), and that information may be provided by the entropy decoder (871) (data path not depicted as this may be low volume control information only).

[0106] The reconstruction module (874) is configured to combine, in the spatial domain, the residual as output by the residue decoder (873) and the prediction results (as output by the inter or intra prediction modules as the case may be) to form a reconstructed block, that may be part of the reconstructed picture, which in turn may be part of the reconstructed video. It is noted that other suitable operations, such as a deblocking operation and the like, can be performed to improve the visual quality.

[0107] It is noted that the video encoders (403), (603), and (703), and the video decoders (410), (510), and (810) can be implemented using any suitable technique. In an embodiment, the video encoders (403), (603), and (703), and the video decoders (410), (510), and (810) can be implemented using one or more integrated circuits. In another embodiment, the video encoders (403), (603), and (603), and the video decoders (410), (510), and (810) can be implemented using one or more processors that execute software instructions.

[0108] Aspects of the disclosure provide techniques for signal multidimensional view positions for multiview video using supplemental enhancement information (SEI) message for coded video stream.

[0109] According to some aspects of the disclosure, video can be categorized into single view videos and multiview videos. For example, a single view video (also referred to as monoscopic video) is a two-dimensional medium that provides viewers a single view of a scene; and a multiview video can provide multiple viewpoints of a scene, and can provide viewers the sensation of realism. In an example, a 3D video can provide two views, such as a left view and a right view corresponding to a human viewer. The two views may be displayed (presented) simultaneously or near simultaneously using different polarizations of light, and a viewer may wear polarized glasses such that each of the viewer's eyes receives a respective one of the views.

[0110] In another example, some display devices, such as some auto-stereoscopic display devices can emit different pictures depending on positions of a viewer's eyes, and do not require glasses for viewing. These display devices are referred to as glassless 3D displays.

[0111] A multiview video is typically created by capturing a scene using multiple cameras simultaneously, where the multiple cameras are properly located so that each camera captures the scene from one viewpoint. Accordingly, the multiple cameras will capture multiple video sequences corresponding to multiple viewpoints. To provide more views, more cameras can be used to generate a multiview video with a large number of video sequences associated with the views. Accordingly, the multi-view video may require a large storage space to store and/or a high bandwidth to transmit. Therefore, multiview video coding techniques have been developed in the field to reduce the required storage space or the transmission bandwidth.

[0112] In order to improve efficiency of multiview video coding, the similarities between views are exploited. In some embodiments, one of the views that is referred to as a base view, is encoded like a monoscopic video. For example, during the encoding of the base view, the intra(frame) and/or temporal inter(frame) predictions are used. The base view may be decoded using a (monoscopic) decoder that performs intra(frame) prediction and the inter(frame) prediction. Other views beside the base view in the multiview video can be referred to as dependent views. To code the dependent views, in addition to intra(frame) and inter(frame) predictions, the inter-view prediction with disparity compensation may be used. In an example, in an inter-view prediction, a current block in a dependent view is predicted using a reference block of samples from a frame of another view in the same time instant. The location of the reference block is pointed out by a disparity vector. This inter-view prediction is similar to the inter(frame) prediction, but the motion vectors are replaced by the disparity vectors, and the temporal reference frames are replaced by the reference frames from other views.

[0113] According to some aspects of the disclosure, multiview coding can employ a multi-layer approach. The multi-layer approach can multiplex different HEVC-coded representations of video sequences, called layers, into one bitstream. The layers can depend on each other. Dependencies are used by inter-layer prediction to achieve increased compression performance by exploiting similarities among different layers. A layer can represent texture, depth or other auxiliary information of a scene related to a particular camera perspective. In some examples, all layers belonging to the same camera perspective are denoted as a view; and layers carrying the same type of information (e.g., texture or depth) are usually called components in the scope of multiview video.

[0114] According to an aspect of the disclosure, the multiview video coding can include high level syntax (HLS) (e.g., higher than slice level) additions with existing single layer decoding cores. Thus, in some examples, the multiview view coding does not change the syntax or decoding process required for (e.g., HEVC) single-layer coding below the slice level. Thus, re-use of existing implementations without major changes for building multiview video decoders is allowed. For example, a multiview video decoder can be implemented based on the video decoder (510) or the video decoder (810).

[0115] In some examples, all pictures associated with the same capturing or display time instant are contained in one access unit (AU) and have the same picture order count (POC). The multiview video coding allows inter-view prediction that performs prediction from pictures in the same AU. For example, the decoded pictures from other views can be inserted into one or both of the reference picture lists of a current picture. Further, in some examples, the motion vectors may be actual temporal motion vectors when related to temporal reference pictures of the same view, or may be disparity vectors when related to inter-view reference pictures. Thus, block-level motion compensation modules (e.g., block level encoding software or hardware, block level decoding software or hardware) can be used which operate the same way regardless of whether a motion vector is a temporal motion vector or a disparity vector.

[0116] According to an aspect of the disclosure, the multiview video coding codes the pictures of different views at a display time instant in an order that is not necessary in association with the display position order.

[0117] FIG. 9 shows a diagram of auto-stereoscopic display (900) in some examples. The auto-stereoscopic display (900) can display pictures of different views in response to detected eye positions of an observer. In the FIG. 9 example, the eye positions of the observer can be detected in one dimension, such as between a leftmost position and a rightmost position. For example, when the observer's eye position is at E0, the auto-stereoscopic display (900) displays pictures of a view with a view identifier ViewId[0]; when the observer's eye position is at E1, the auto-stereoscopic display (900) displays pictures of a view with a view identifier ViewId[1]; when the observer's eye position is at E2, the auto-stereoscopic display (900) displays pictures of a view with a view identifier ViewId[2]; and when the observer's eye position is at E3, the auto-stereoscopic display (900) displays pictures of a view with a view identifier ViewId[3]. The order of observer's eye position from left to right is E2, E0, E1 and E3.

[0118] In some examples, an AU includes coded information of pictures of different views associated with the same capturing or display time instant. For example, an AU includes coded information of a picture P0 of the view ViewId[0], a picture P1 of the view ViewId[1], a picture P2 of the view ViewId[2], and a picture P3 of the view ViewId[3]. The pictures P0-P3 can be coded (encoded or decoded) in a coding order that may be determined based on certain coding requirements, such as rate distortion optimization, and the like. The coding order does not necessarily follow the order of the observer's eye position.

[0119] FIGs. 10A-B show an example of reordering pictures according to multiview view positions in an example.

[0120] FIG. 10A shows pictures in an AU that is decoded according to a decoding order. In the FIG. 10A example, the pictures P0-P3 in an AU are decoded in an order of P0, P1, P2 and P3. Because the order of observer's eye positions from left to right is E2, E0, E1 and E3. The decoded picture order does not correspond to the order of the observer's eye positions from left to right.

[0121] In some examples, the decoded pictures can be re-ordered according to a display order that is signaled for example for the intended user experience. In an example, the display order is associated with the order of the observer's eye positions, such as from left to right.

[0122] FIG. 10B shows decoded pictures in an AU that are re-ordered according to a display order in some examples. For example, the display order is associated with the order of the observer's eye positions from left to right.

[0123] According to an aspect of the disclosure, the display order can be signaled using a supplemental enhancement information (SEI) message.

[0124] According to some aspects of the disclosure, supplemental enhancement information (SEI) messages can be included in an encoded bitstream to assist in the decoding and/or display of the encoded bitstream, or for another purpose. In some examples, SEI messages is not required for reconstructing luma or chroma samples during the decoding process. Additionally, decoders that conform to a video coding standard that supports SEI messages are not required to process SEI messages in order to be conforming. For some coding standards, some SEI message information may be required to check bitstream conformance or for outputting timing decoder conformance.

[0125] SEI messages can contain various types of data that indicate the timing of the video pictures or describe various properties of the coded video or how the various properties can be used or enhanced. In some examples, SEI messages do not affect the core decoding process, but can indicate how the video is recommended to be post-processed or displayed.

[0126] SEI messages can be used to provide additional information about an encoded bitstream, which can be used to change the presentation of the bitstream once the bitstream is decoded, or to provide information to a decoder. For example, SEI messages have been used to provide frame packing information (e.g., describing the manner in which video data is arranged in a video frame), content descriptions (e.g., to indicate that the encoded bitstream is, for example, 360-degree video), and color information (e.g., color gamut and/or color range), among other things.

[0127] In some examples, an SEI message can be used to signal to a decoder that the encoded bitstream includes 360-degree video. The decoder can use this information to render the video data for a 360-degree presentation. Alternatively, if the decoder is not capable of rendering 360-degree video, the decoder can use this information to not render the video data.

[0128] In some related examples, the SEI message can include information indicating view positions in one dimension. For example, the SEI message can include the view positions in one dimension correspond to observer eye positions.

[0129] FIG. 11 shows a syntax example (1100) in an SEI message for indicating view positions for a multiview video. The SEI message for indicating the view positions can be referred to as multiview view position (MVP) SEI message. In some examples, the MVP SEI message can signal the number of views, and then signal positions of the views respectively.

[0130] In the FIG. 11 example, a parameter denoted by num_views_minus1 can be signaled by the MVP SEI message as shown by (1110). The parameter num_views_minus1 is indicative of the number of views, for example in an access unit. For example, the number of views is equal to a sum of 1 and the value of the parameter num_views_minus1. The views in a decoding order have view identifiers ViewId[0] to ViewId[3]. Then, the positions of the views in the display order can be respectively signaled in the MVP SEI message, such as view_positions[i] where i is an integer from 0 to num_views_minus1, as shown by (1120) in FIG. 11.

[0131] In an example, to signal the view positions for the example in FIG. 9, “3” can be signaled as the parameter num_views_minus1, and then “1” is signaled as the view_position[0], “2” is signaled as the view_position[1], “0” is signaled as the view_position[2], and “3” is signaled as the view_position[3], where “0” is the leftmost, and “3” is the rightmost in position from left to right. Then, when views ViewId[0]-ViewId[3] are decoded from an access unit, the view ViewId[0] has the view_position[0], the view ViewId[1] has the view_position[1], the view ViewId[2] has the view_position[2], and the view ViewId[3] has the view_position[3]. The views ViewId[0]-ViewId[3] can be re-ordered according to corresponding view positions view_position[0] to view_position[3] to obtain the display order in FIG. 10B.

[0132] The present disclosure provides techniques for use in MVP SEI message to represent view positions in multidimensional coordinates in a multidimensional space, such as two-dimensional array along both vertical and horizontal axes. In some examples, the techniques for two-dimensional representation of multiview view positions can be used to support proper display in 3D display devices, such as light-field display device, holography display device, that can support 2D representation of multiview video.

[0133] In some embodiments, the techniques can employ a list of vertical view position (e.g., denoted by view_position_y[i]) as well as a list of horizontal view position (e.g., denoted by view_position_x[i]), and employ a parameter denoted by num_vertical_view_positions_minus1 to indicate the number of vertical view positions, and a parameter denoted by num_horizontal_view_positions_minus1 to indicate the number of horizontal view positions.

[0134] According to some aspects of the disclosure, the MVP SEI message may need to subject to similar constraints for other multiview related SEI messages, such as scalability dimension information (SDI) SEI message, multiview acquisition information (MAI) SEI message, and the like. In some examples, the MVP SEI message is constrained not to be contained in a scalable nesting SEI message. A scalable nesting SEI message refers to an SEI message that contains one or more additional SEI messages. The SEI messages contained in the scalable nesting SEI messages can be referred to as scalable nested SEI messages.

[0135] An SDI SEI message refers to a SEI message that signals scalability dimensions information of a multiview video. For example, the SDI SEI message may include number and types of scalability dimensions, such as information indicating the number of views and the like for multiview video. In some examples (e.g., a version of versatile SEI, such as JVET-W2006), SDI SEI messages may subject to a constraint that an SDI SEI message shall not be contained in a scalable nesting SEI message.

[0136] According to an aspect of the disclosure, the MVP SEI message includes one or more parameters that have semantic dependency on information in the SDI SEI message. For example, a syntax element num_views_minus1 has a semantic dependency on the value of a parameter NumViews that is derived from the SDI SEI message. Thus, the MVP SEI message may subject to constraints in association with the SDI SEI message. In an example, when a coded video sequence (CVS) does not contain an SDI SEI message, the CVS shall not contain an MVP SEI message. In some examples, an MVP SEI message shall not be present when the associated SDI SEI message is not present. In another example, an MVP SEI message shall not be contained in a scalable nesting SEI message because SDI SEI message shall not be contained in a scalale nesting SEI message.

[0137] In some examples, a bitstream can include one or more coded video sequences (CVSs). A CVS is independently coded from other CVSs. Each CVS can include one or more layers, each layer is a representation of the video with a specific quality or spatial resolution, or a representation of some component interpretation property, e.g., as depth or transparency maps or perspective views. In a temporal dimension, each CVS includes one or more access units (AUs). Each AU includes one or more pictures of different layers that belong to a same time instant. A coded layer video sequence (CLVS) is a layer-wise CVS that includes a sequence of picture units in the same layer. If a bitstream has multiple layers, a CVS in the bitstream has one or more CLVSs for each layer.

[0138] FIG. 12 shows a diagram of auto-stereoscopic display (1200) in some examples. The auto-stereoscopic display (1200) can display pictures of different views in response to, for example, detected eye positions of an observer, detected head postures of an observer, and the like. In the FIG. 12 example, the eye positions of the observer can be detected in a two-dimensional plane. For example, when the observer's eye position is at E0, the auto-stereoscopic display (1200) displays pictures of a view with a view identifier ViewId[0]; when the observer's eye position is at E1, the auto-stereoscopic display (1200) displays pictures of a view with a view identifier ViewId[1]; when the observer's eye position is at E2, the auto-stereoscopic display (1200) displays pictures of a view with a view identifier ViewId[2]; when the observer's eye position is at E3, the auto-stereoscopic display (1200) displays pictures of a view with a view identifier ViewId[3]; when the observer's eye position is at E4, the auto-stereoscopic display (1200) displays pictures of a view with a view identifier ViewId[4]; when the observer's eye position is at E5, the auto-stereoscopic display (1200) displays pictures of a view with a view identifier ViewId[5]; when the observer's eye position is at E6, the auto-stereoscopic display (1200) displays pictures of a view with a view identifier ViewId[6]; and when the observer's eye position is at E7, the auto-stereoscopic display (1200) displays pictures of a view with a view identifier ViewId[7].

[0139] In some examples, an AU includes coded information of pictures of different views associated with the same capturing or display time instant. For example, pictures in the AU can be decoded as a picture P0 of view ViewId[0], a picture P1 of view ViewId[1], a picture P2 of view ViewId[2], a picture P3 of view ViewId[3], a picture P4 of view ViewId[4], a picture P5 of view ViewId[5], a picture P6 of view ViewId[6], and a picture P7 of view ViewId[7]. The pictures P0-P7 can be coded (encoded or decoded) in a coding order that may be determined based on certain coding requirements, such as rate distortion optimization, and the like. The coding order does not necessarily follow the order of the observer's eye position.

[0140] FIG. 13 shows an example of associating pictures according to multiview view positions in an example.

[0141] In the FIG. 13 example, the pictures in an AU are decoded in an order of P0, P1, P2, P3, P4, P5, P6 and P7. In some embodiments, an MVP SEI message can provide view positions in a 2-D matrix, then the pictures can be associated with view positions in the 2D matrix, thus the pictures are ordered according to the view positions. The ordered pictures in 2D matrix can be provided for display in response to detected eye positions of the observer in some examples.

[0142] In the FIG. 13 example, view positions are provided, for example in a form of two-dimensional coordinates in a 2D matrix. In the FIG. 13 example, the 2D matrix has a horizontal axis that is represented by an arrow from left to right, and has a vertical axis that is represented by an arrow from top to bottom. The pictures P0-P7 are associated with the view positions.

[0143] It is noted that while FIG. 13 shows a cartesian coordinate system, other suitable coordinate system, such as polar coordinate system, spherical coordinate system, and the like can be used in some examples.

[0144] The view positions can be provided by an SEI message that is also referred to as MVP SEI message. The MVP SEI message specifies the relative view positions along a horizontal axis and a vertical axis of views within a coded video sequence (CVS). In some examples, when an MVP SEI message is presented in a CVS, the MVP SEI message shall be associated with an intra random access picture (IRAP) access unit. A random access point picture refers to a picture where a decoder may start decoding a coded video sequence. When the random access point picture is intra coded, the random access point picture is referred to as intra random access pictures (IRAP). It is noted that, in some examples, the information signaled in the MVP SEI message in a CVS applies to the entire CVS.

[0145] According to an aspect of the disclosure, the MVP SEI message includes one or more parameters that have semantic dependency on information in the SDI SEI message. For example, a syntax element num_views_minus1 has a semantic dependency on the value of a parameter NumViews that is derived from the SDI SEI message. Thus, the MVP SEI message may subject to constraints in association with the SDI SEI message. In an example, when a coded video sequence (CVS) does not contain an SDI SEI message, the CVS shall not contain an MVP SEI message. In some examples, an MVP SEI message shall not be present when the associated SDI SEI message is not present. In another example, an MVP SEI message shall not be contained in a scalable nesting SEI message, because SDI SEI message shall not be contained in a scalale nesting SEI message.

[0146] FIG. 14 shows a syntax example (1400) in an MVP SEI message for indicating view positions in 2D matrix for a multiview video. In some examples, the MVP SEI message can signal a first value indicating a number of views, a second value indicating a number of view positions in horizontal direction, a third value indicating a number of view positions in vertical direction, and then signal relative view positions in the horizontal direction and vertical direction respectively.

[0147] In the FIG. 14 example, a parameter denoted by num_views_minus1 can be signaled by the MVP SEI message as shown by (1410). The parameter num_views_minus1 is indicative of the number of views, for example in an access unit. For example, the number of views is equal to a sum of 1 and the value of the parameter num_views_minus1. In some examples, the number of views shall be equal to NumViews derived from the SDI SEI message for the CVS.

[0148] In the FIG. 14 example, a parameter denoted by num_vertical_view_positions_minus1 can be signaled by the MVP SEI message as shown by (1420). The parameter num_vertical_view_positions_minus1 is indicative of the number of vertical view positions, for example in the 2D matrix. For example, the number of vertical view positions is equal to a sum of 1 and the value of the parameter num_vertical_view_positions_minus1. In some examples, the value of num_vertical_view_positions_minus1 shall be in the range of 0 to 62, inclusive.

[0149] In the FIG. 14 example, a parameter denoted by num_horizontal_view_positions_minus1 can be signaled by the MVP SEI message as shown by (1430). The parameter num_horizontal_view_positions_minus1 is indicative of the number of horizontal view positions, for example in the 2D matrix. For example, the number of horizontal view positions is equal to a sum of 1 and the value of the parameter num_horizontal_view_positions_minus1. In some examples, the value of num_horizontal_view_positions_minus1 shall be in the range of 0 to 62, inclusive.

[0150] Further, in the FIG. 14 example, vertical view positions and horizontal view positions are respective signaled by the MVP SEI message as shown by (1460).

[0151] In some examples, the views can be ordered in a decoding order. Then, the vertical positions and horizontal of the views in the decoding order can be respectively signaled in the MVP SEI message, such as view_position_y[i] and view_position_x[i] where i is an integer from 0 to num_views_minus1, as shown by (1440) and (1450) in FIG. 14.

[0152] The parameter view_position_y[i] indicates the vertical order of the view with view identifier equal to ViewId[i] among all the views from top to bottom for the purpose of display, with the order for the topmost view being equal to 0 and the value of the order increasing by 1 for next view from top to bottom. The value of view_position_y[i] shall be in the range of 0 to num_vertical_view_positions_minus1, inclusive.

[0153] The parameter view_position_x[i] indicates the horizontal order of the view with view identifier equal to ViewId[i] among all the views from left to right for the purpose of display, with the order for the leftmost view being equal to 0 and the value of the order increasing by 1 for next view from left to right. The value of view_position_x[i] shall be in the range of 0 to num_horizontal_view_positions_minus1, inclusive.

[0154] Using the view positions in FIG. 13 as an example, 0 is signaled as view_position_y[0] and 0 is signaled as view_position_x[0] for ViewId[0]; 0 is signaled as view_position_y[1] and 1 is signaled as view_position_x[1] for ViewId[1]; 0 is signaled as view_position_y[2] and 2 is signaled as view_position_x[2] for ViewId[2]; 0 is signaled as view_position_y[3] and 3 is signaled as view_position_x[3] for ViewId[3]; 1 is signaled as view_position_y[4] and 0 is signaled as view_position_x[4] for ViewId[4]; 1 is signaled as view_position_y[5] and 1 is signaled as view_position_x[5] for ViewId[5]; 1 is signaled as view_position_y[6] and 2 is signaled as view_position_x[6] for ViewId[6]; 1 is signaled as view_position_y[7] and 3 is signaled as view_position_x[7] for ViewId[7].

[0155] FIG. 15 shows a flow chart outlining a process (1500) according to an embodiment of the disclosure. The process (1500) can be used in a video encoder. In various embodiments, the process (1500) is executed by processing circuitry, such as the processing circuitry in the terminal devices (310), (320), (330) and (340), the processing circuitry that performs functions of the video encoder (403), the processing circuitry that performs functions of the video encoder (603), the processing circuitry that performs functions of the video encoder (703), and the like. In some embodiments, the process (1500) is implemented in software instructions, thus when the processing circuitry executes the software instructions, the processing circuitry performs the process (1500). The process starts at (S15301) and proceeds to (S1510).

[0156] At (S1510), pictures associated with views are encoded in a bitstream. In some examples, the pictures are taken by multiple cameras of a same scene at a same time instant. In some examples, the pictures are generated for displaying a same scene at a same time instant from different views. In an example, the pictures are encoded in an access unit in a coded video sequence carried by the bitstream.

[0157] In some examples, the pictures are videos taken by the multiple cameras. The pictures are encoded into access units of a coded video sequence.

[0158] At (S1520), a SEI message is formed to include positions of multidimensional coordinates in a multidimensional space respectively for the views.

[0159] In some embodiments, the multidimensional space is a two dimensional space, thus positions of two dimensional coordinates in the two dimensional space for the views are included in the SEI message.

[0160] In some examples, a first value (e.g., the value of num_views_minus1) indicative of a number of views is included in the SEI message; a second value (e.g., the value of num_vertical_view_positions_minus1) indicative of a first number of positions in a vertical dimension of the two dimensional space is included in the SEI message; and a third value (e.g., the value of num_horizontal_view_positions_minus1) indicative of a second number of positions in a horizontal dimension of the two dimensional space is included in the SEI message.

[0161] In some examples, for each view of the number of views, a first coordinate value of a first dimension and a second coordinate value of a second dimension are included in the SEI message as a view position in the two dimensional space.

[0162] At (S1530), the SEI message is included in the bitstream.

[0163] In some examples, the SEI message is associated with a coded video sequence, and all access units in the entire coded video sequence have the same views as the positions defined in the SEI message. In some examples, the SEI message is associated with an intra random access point (IRAP) access unit of the coded video sequence.

[0164] It is noted that, in some examples, the SEI message is a non nested SEI message, and the non nested SEI message is not in another SEI message.

[0165] In some examples, the SEI message is denoted as a multiview view position (MVP) SEI message. When the coded video sequence includes the MVP SEI message, the coded video sequence also includes a scalability dimension information (SDI) SEI message. In an example, a first value in the SDI SEI message is indicative of a number of the views from the SDI SEI message, and a second value (e.g., the value of num_views_minus1) in the MVP SEI message is associated with the first value in the SDI SEI message, and a conformance check can be applied to check a relationship of the first value and the second value.

[0166] Then, the process proceeds to (S1599) and terminates.

[0167] The process (1500) can be suitably adapted. Step(s) in the process (1500) can be modified and/or omitted. Additional step(s) can be added. Any suitable order of implementation can be used.

[0168] FIG. 16 shows a flow chart outlining a process (1600) according to an embodiment of the disclosure. The process (1600) can be used in a video decoder. In various embodiments, the process (1600) is executed by processing circuitry, such as the processing circuitry in the terminal devices (310), (320), (330) and (340), the processing circuitry that performs functions of the video decoder (410), the processing circuitry that performs functions of the video decoder (510), and the like. In some embodiments, the process (1600) is implemented in software instructions, thus when the processing circuitry executes the software instructions, the processing circuitry performs the process (1600). The process starts at (S1601) and proceeds to (S1610).

[0169] At (S1610), pictures associated with views are decoded from a bitstream. In an example, the pictures are decoded from an access unit in a coded video sequence carried by the bitstream. In another example, the pictures are decoded from access units of the coded video sequence.

[0170] At (S1620), from an SEI message, positions of multidimensional coordinates in a multidimensional space are determined respectively for the views.

[0171] In some embodiments, the multidimensional space is a two dimensional space, thus positions of two dimensional coordinates in the two dimensional space for the views are determined from the SEI message.

[0172] In some examples, a first value obtained from the SEI message is indicative of a number of views; a second value obtained from the SEI message is indicative of a first number of positions in a vertical dimension of the two dimensional space; and a third value obtained from the SEI message is indicative of a second number of positions in a horizontal dimension of the two dimensional space.

[0173] In some examples, for each view of the number of views, a first coordinate value of a first dimension and a second coordinate value of a second dimension can be obtained from the SEI message as a view position in the two dimensional space.

[0174] In some examples, the SEI message is associated with an intra random access point (IRAP) access unit of a coded video sequence.

[0175] It is noted that, in some examples, the SEI message is a non nested SEI message, and the non nested SEI message is not in another SEI message.

[0176] In some examples, the SEI message is in a coded video sequence carried by the bitstream, and the positions of the multidimensional coordinates for the views are applied to all access units in the entire coded video sequence.

[0177] In some examples, the SEI message is denoted as a multiview view position (MVP) SEI message. When the coded video sequence includes the MVP SEI message, the coded video sequence also includes a scalability dimension information (SDI) SEI message. In an example, a first value indicative of a number of the views is derived from the SDI SEI message, and a second value associated with the number is obtained from the MVP SEI message. Then, one plus the second value is compared with the first value in a conformance check.

[0178] At (S1630), a rendering picture is determined from the pictures based on a rendering view in the multidimensional space. In some examples, the rendering view is one of the views, and the rendering picture is one of the pictures selected based on the rendering view. In some examples, the rendering picture can be derived from the pictures when the rendering view is not any of the views. In some examples, the rendering view is determined based on an observer's viewport information, such as eye position of the observer, head posture of the observer, the observer's location, and the like. Then, the process proceeds to (S1699) and terminates.

[0179] The process (1600) can be suitably adapted. Step(s) in the process (1600) can be modified and/or omitted. Additional step(s) can be added. Any suitable order of implementation can be used. In some examples, processing circuitry can receive, from a bitstream, pictures associated with a plurality of views. The processing circuitry decodes the pictures associated with the plurality of views, and determines, from a supplemental enhancement information (SEI) message in the bitstream, positions of multidimensional coordinates in a multidimensional space respectively for the plurality of views. The positions includes at least both a vertical view position and a horizontal view position. The processing circuitry renders the pictures based on a rendering view corresponding to the vertical view position and the horizontal view position. The processing circuitry reorders the pictures based on the rendered pictures and the positions of multidimensional coordinates in the multidimensional space.

[0180] The techniques described above, can be implemented as computer software using computer-readable instructions and physically stored in one or more computer-readable media. For example, FIG. 17 shows a computer system (1700) suitable for implementing certain embodiments of the disclosed subject matter.

[0181] The computer software can be coded using any suitable machine code or computer language, that may be subject to assembly, compilation, linking, or like mechanisms to create code comprising instructions that can be executed directly, or through interpretation, micro-code execution, and the like, by one or more computer central processing units (CPUs), Graphics Processing Units (GPUs), and the like.

[0182] The instructions can be executed on various types of computers or components thereof, including, for example, personal computers, tablet computers, servers, smartphones, gaming devices, internet of things devices, and the like.

[0183] The components shown in FIG. 17 for computer system (1700) are exemplary in nature and are not intended to suggest any limitation as to the scope of use or functionality of the computer software implementing embodiments of the present disclosure. Neither should the configuration of components be interpreted as having any dependency or requirement relating to any one or combination of components illustrated in the exemplary embodiment of a computer system (1700).

[0184] Computer system (1700) may include certain human interface input devices. Such a human interface input device may be responsive to input by one or more human users through, for example, tactile input (such as: keystrokes, swipes, data glove movements), audio input (such as: voice, clapping), visual input (such as: gestures), olfactory input (not depicted). The human interface devices can also be used to capture certain media not necessarily directly related to conscious input by a human, such as audio (such as: speech, music, ambient sound), images (such as: scanned images, photographic images obtain from a still image camera), video (such as two-dimensional video, three-dimensional video including stereoscopic video).

[0185] Input human interface devices may include one or more of (only one of each depicted): keyboard (1701), mouse (1702), trackpad (1703), touch screen (1710), data-glove (not shown), joystick (1705), microphone (1706), scanner (1707), camera (1708).

[0186] Computer system (1700) may also include certain human interface output devices. Such human interface output devices may be stimulating the senses of one or more human users through, for example, tactile output, sound, light, and smell/taste. Such human interface output devices may include tactile output devices (for example tactile feedback by the touch-screen (1710), data-glove (not shown), or joystick (1705), but there can also be tactile feedback devices that do not serve as input devices), audio output devices (such as: speakers (1709), headphones (not depicted)), visual output devices (such as screens (1710) to include CRT screens, LCD screens, plasma screens, OLED screens, each with or without touch-screen input capability, each with or without tactile feedback capability—some of which may be capable to output two dimensional visual output or more than three dimensional output through means such as stereographic output; virtual-reality glasses (not depicted), holographic displays and smoke tanks (not depicted)), and printers (not depicted).

[0187] Computer system (1700) can also include human accessible storage devices and their associated media such as optical media including CD/DVD ROM/RW (1720) with CD/DVD or the like media (1721), thumb-drive (1722), removable hard drive or solid state drive (1723), legacy magnetic media such as tape and floppy disc (not depicted), specialized ROM/ASIC/PLD based devices such as security dongles (not depicted), and the like.

[0188] Those skilled in the art should also understand that term "computer readable media" as used in connection with the presently disclosed subject matter does not encompass transmission media, carrier waves, or other transitory signals.

[0189] Computer system (1700) can also include an interface (1754) to one or more communication networks (1755). Networks can for example be wireless, wireline, optical. Networks can further be local, wide-area, metropolitan, vehicular and industrial, real-time, delay-tolerant, and so on. Examples of networks include local area networks such as Ethernet, wireless LANs, cellular networks to include GSM, 3G, 4G, 5G, LTE and the like, TV wireline or wireless wide area digital networks to include cable TV, satellite TV, and terrestrial broadcast TV, vehicular and industrial to include CANBus, and so forth. Certain networks commonly require external network interface adapters that attached to certain general purpose data ports or peripheral buses (1749) (such as, for example USB ports of the computer system (1700)); others are commonly integrated into the core of the computer system (1700) by attachment to a system bus as described below (for example Ethernet interface into a PC computer system or cellular network interface into a smartphone computer system). Using any of these networks, computer system (1700) can communicate with other entities. Such communication can be uni-directional, receive only (for example, broadcast TV), uni-directional send-only (for example CANbus to certain CANbus devices), or bi-directional, for example to other computer systems using local or wide area digital networks. Certain protocols and protocol stacks can be used on each of those networks and network interfaces as described above.

[0190] Aforementioned human interface devices, human-accessible storage devices, and network interfaces can be attached to a core (1740) of the computer system (1700).

[0191] The core (1740) can include one or more Central Processing Units (CPU) (1741), Graphics Processing Units (GPU) (1742), specialized programmable processing units in the form of Field Programmable Gate Areas (FPGA) (1743), hardware accelerators for certain tasks (1744), graphics adapters (1750), and so forth. These devices, along with Read-only memory (ROM) (1745), Random-access memory (1746), internal mass storage such as internal non-user accessible hard drives, SSDs, and the like (1747), may be connected through a system bus (1748). In some computer systems, the system bus (1748) can be accessible in the form of one or more physical plugs to enable extensions by additional CPUs, GPU, and the like. The peripheral devices can be attached either directly to the core's system bus (1748), or through a peripheral bus (1749). In an example, the screen (1710) can be connected to the graphics adapter (1750). Architectures for a peripheral bus include PCI, USB, and the like.

[0192] CPUs (1741), GPUs (1742), FPGAs (1743), and accelerators (1744) can execute certain instructions that, in combination, can make up the aforementioned computer code. That computer code can be stored in ROM (1745) or RAM (1746). Transitional data can be also be stored in RAM (1746), whereas permanent data can be stored for example, in the internal mass storage (1747). Fast storage and retrieve to any of the memory devices can be enabled through the use of cache memory, that can be closely associated with one or more CPU (1741), GPU (1742), mass storage (1747), ROM (1745), RAM (1746), and the like.

[0193] The computer readable media can have computer code thereon for performing various computer-implemented operations. The media and computer code can be those specially designed and constructed for the purposes of the present disclosure, or they can be of the kind well known and available to those having skill in the computer software arts.

[0194] As an example and not by way of limitation, the computer system having architecture (1700), and specifically the core (1740) can provide functionality as a result of processor(s) (including CPUs, GPUs, FPGA, accelerators, and the like) executing software embodied in one or more tangible, computer-readable media. Such computer-readable media can be media associated with user-accessible mass storage as introduced above, as well as certain storage of the core (1740) that are of non-transitory nature, such as core-internal mass storage (1747) or ROM (1745). The software implementing various embodiments of the present disclosure can be stored in such devices and executed by core (1740). A computer-readable medium can include one or more memory devices or chips, according to particular needs. The software can cause the core (1740) and specifically the processors therein (including CPU, GPU, FPGA, and the like) to execute particular processes or particular parts of particular processes described herein, including defining data structures stored in RAM (1746) and modifying such data structures according to the processes defined by the software.

In addition or as an alternative, the computer system can provide functionality as a result of logic hardwired or otherwise embodied in a circuit (for example: accelerator (1744)), which can operate in place of or together with software to execute particular processes or particular parts of particular processes described herein. Reference to software can encompass logic, and vice versa, where appropriate. Reference to a computer-readable media can encompass a circuit (such as an integrated circuit (IC)) storing software for execution, a circuit embodying logic for execution, or both, where appropriate. The present disclosure encompasses any suitable combination of hardware and software.

Appendix A: Acronyms

JEM: joint exploration model

VVC: versatile video coding

BMS: benchmark set

MV: Motion Vector

HEVC: High Efficiency Video Coding

SEI: Supplementary Enhancement Information

VUI: Video Usability Information

GOPs: Groups of Pictures

TUs: Transform Units,

PUs: Prediction Units

CTUs: Coding Tree Units

CTBs: Coding Tree Blocks

PBs: Prediction Blocks

HRD: Hypothetical Reference Decoder

SNR: Signal Noise Ratio

CPUs: Central Processing Units

GPUs: Graphics Processing Units

CRT: Cathode Ray Tube

LCD: Liquid-Crystal Display

OLED: Organic Light-Emitting Diode

CD: Compact Disc

DVD: Digital Video Disc

ROM: Read-Only Memory

RAM: Random Access Memory

ASIC: Application-Specific Integrated Circuit

PLD: Programmable Logic Device

LAN: Local Area Network

GSM: Global System for Mobile communications

LTE: Long-Term Evolution

CANBus: Controller Area Network Bus

USB: Universal Serial Bus

PCI: Peripheral Component Interconnect

FPGA: Field Programmable Gate Areas

SSD: solid-state drive

IC: Integrated Circuit

CU: Coding Unit

[0195] While this disclosure has described several exemplary embodiments, there are alterations, permutations, and various substitute equivalents, which fall within the scope of the disclosure. It will thus be appreciated that those skilled in the art will be able to devise numerous systems and methods which, although not explicitly shown or described herein, embody the principles of the disclosure and are thus within the spirit and scope thereof.

WHAT IS CLAIMED IS:

1. A method of video processing, comprising:

receiving, from a bitstream, pictures associated with a plurality of views;

decoding the pictures associated with the plurality of views;

determining, from a supplemental enhancement information (SEI) message in the bitstream, positions of multidimensional coordinates in a multidimensional space respectively for the plurality of views, the positions comprising at least both a vertical view position and a horizontal view position;

rendering the pictures based on a rendering view corresponding to the vertical view position and the horizontal view position; and

reordering the pictures based on the rendered pictures and the positions of multidimensional coordinates in the multidimensional space.

2. The method of claim 1, wherein the determining, from the SEI message, the positions of the multidimensional coordinates in the multidimensional space for the plurality of views further comprises:

determining, from the SEI message, positions of two dimensional coordinates in a two dimensional space for the plurality of views.

3. The method of claim 2, wherein the determining, from the SEI message, the positions of the two dimensional coordinates in the two dimensional space for the plurality of views further comprises:

obtaining a first value from the SEI message, the first value indicative of a number of the plurality of views;

obtaining a second value from the SEI message, the second value indicative of a first number of positions in a vertical dimension of the two dimensional space; and

obtaining a third value from the SEI message, the third value indicative of a second number of positions in a horizontal dimension of the two dimensional space.

4. The method of claim 3, wherein the determining, from the SEI message, the positions of the two dimensional coordinates in the two dimensional space for the views further comprises:

obtaining, for each view of the plurality of views, a first coordinate value of a vertical dimension and a second coordinate value of a horizontal dimension as a view position in the two dimensional space.

5. The method of claim 1, wherein the SEI message is associated with an intra random access point (IRAP) access unit of a coded video sequence.

6. The method of claim 1, wherein the SEI message is not in another SEI message.

7. The method of claim 1, wherein the SEI message is in a coded video sequence carried by the bitstream, and the positions of the multidimensional coordinates for the views are applied to access units in the coded video sequence.

8. The method of claim 7, wherein the SEI message is denoted as a multiview view position (MVP) SEI message, and the coded video sequence comprises a scalability dimension information (SDI) SEI message.

9. The method of claim 8, further comprising:

deriving a first value indicative of a number of the views from the SDI SEI message;

obtaining a second value associated with the number from the MVP SEI message; and

comparing one plus the second value with the first value in a conformance check.

10. An apparatus for video processing, comprising processing circuitry configured to:

receive, from a bitstream, pictures associated with a plurality of views;

decode the pictures associated with the plurality of views;

determine, from a supplemental enhancement information (SEI) message in the bitstream, positions of multidimensional coordinates in a multidimensional space respectively for the plurality of views, the positions comprising at least both a vertical view position and a horizontal view position;

render the pictures based on a rendering view corresponding to the vertical view position and the horizontal view position; and reorder the pictures based on the rendered pictures and the positions of multidimensional coordinates in the multidimensional space.

11. The apparatus of claim 10, wherein the processing circuitry is configured to:

determine, from the SEI message, positions of two dimensional coordinates in a two dimensional space for the plurality of views.

12. The apparatus of claim 11, wherein the processing circuitry is configured to:

obtain a first value from the SEI message, the first value indicative of a number of views;

obtain a second value from the SEI message, the second value indicative of a first number of positions in a vertical dimension of the two dimensional space; and

obtain a third value from the SEI message, the third value indicative of a second number of positions in a horizontal dimension of the two dimensional space.

13. The apparatus of claim 12, wherein the processing circuitry is configured to:

obtain, for each view of the plurality of views, a first coordinate value of a first dimension and a second coordinate value of a second dimension as a view position in the two dimensional space.

14. The apparatus of claim 10, wherein the SEI message is associated with an intra random access point (IRAP) access unit of a coded video sequence.

15. The apparatus of claim 10, wherein the SEI message is not in another SEI message.

16. The apparatus of claim 10, wherein the SEI message is in a coded video sequence carried by the bitstream, and the positions of the multidimensional coordinates for the views are applied to access units in the coded video sequence.

17. The apparatus of claim 16, wherein the SEI message is denoted as a multiview view position (MVP) SEI message, and the coded video sequence comprises a scalability dimension information (SDI) SEI message.

18. The apparatus of claim 17, wherein the processing circuitry is configured to:

derive a first value indicative of a number of the views from the SDI SEI message;

obtain a second value associated with the number from the MVP SEI message; and

compare one plus the second value with the first value in a conformance check.

19. A non-transitory computer-readable storage medium storing instructions which when executed by at least one processor cause the at least one processor to perform:

receiving, from a bitstream, pictures associated with a plurality of views;

decoding the pictures associated with the plurality of views;

determining, from a supplemental enhancement information (SEI) message in the bitstream, positions of multidimensional coordinates in a multidimensional space respectively for the plurality of views, the positions comprising at least both a vertical view position and a horizontal view position;

rendering the pictures based on a rendering view corresponding to the vertical view position and the horizontal view position; and

reordering the pictures based on the rendered pictures and the positions of multidimensional coordinates in the multidimensional space.

20. The non-transitory computer-readable storage medium of claim 19, wherein the instructions cause the at least one processor to perform:

obtaining a first value from the SEI message, the first value indicative of a number of views;

obtaining a second value from the SEI message, the second value indicative of a first number of positions in a vertical dimension of a two dimensional space;

obtaining a third value from the SEI message, the third value indicative of a second number of positions in a horizontal dimension of the two dimensional space; and

obtaining, for each view of the plurality of views, a first coordinate value of a first dimension and a second coordinate value of a second dimension as a view position in the two dimensional space.

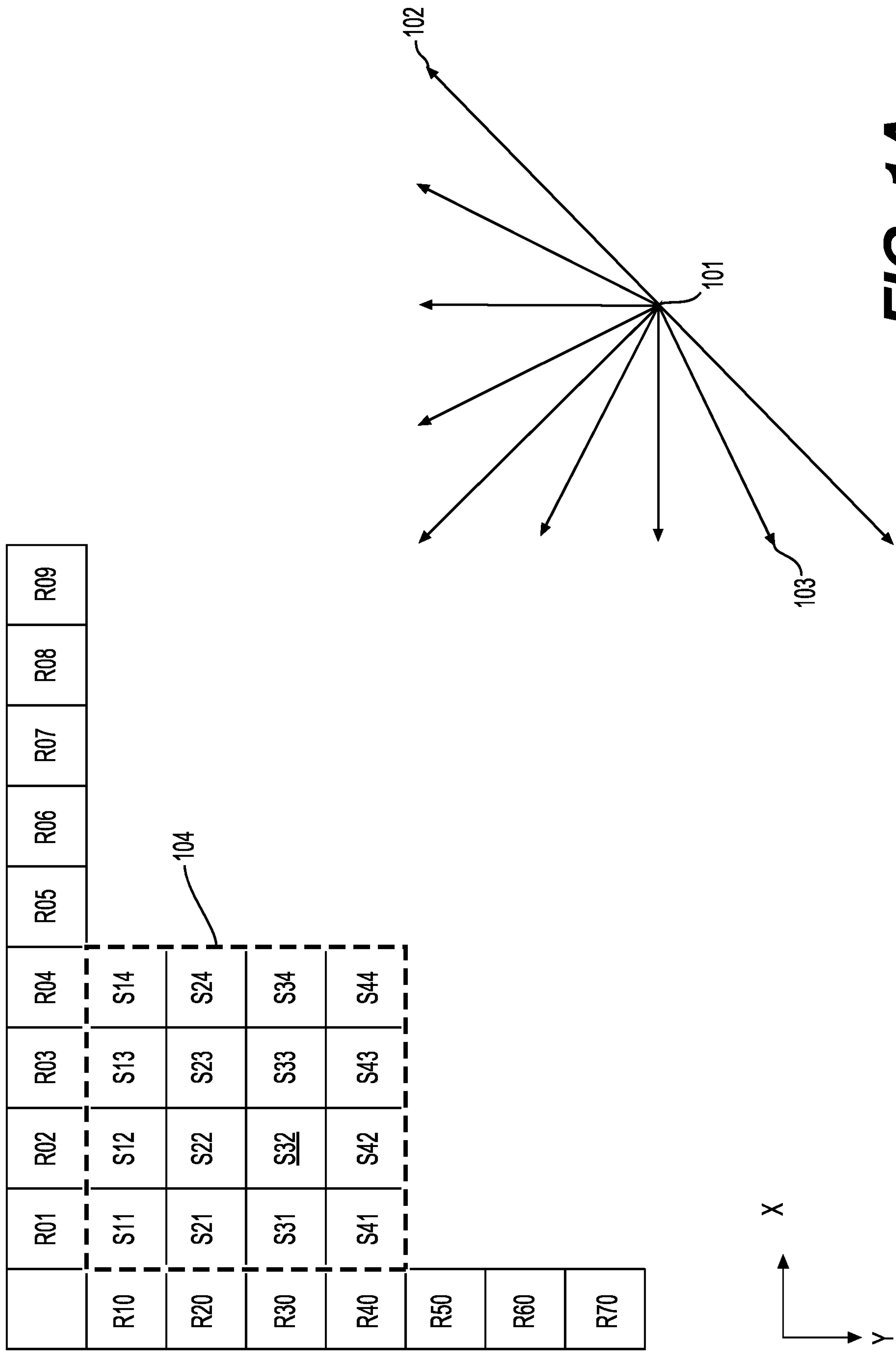
102
101
103
104
R01
R02
R03
R04
R05
R06
R07
R08
R09
R10
R20
R30
R40
R50
R60
R70
S11
S12
S13
S14
S21
S22
S23
S24
S31
S32
S33
S34
S41
S42
S43
S44
X
Y

FIG. 1A

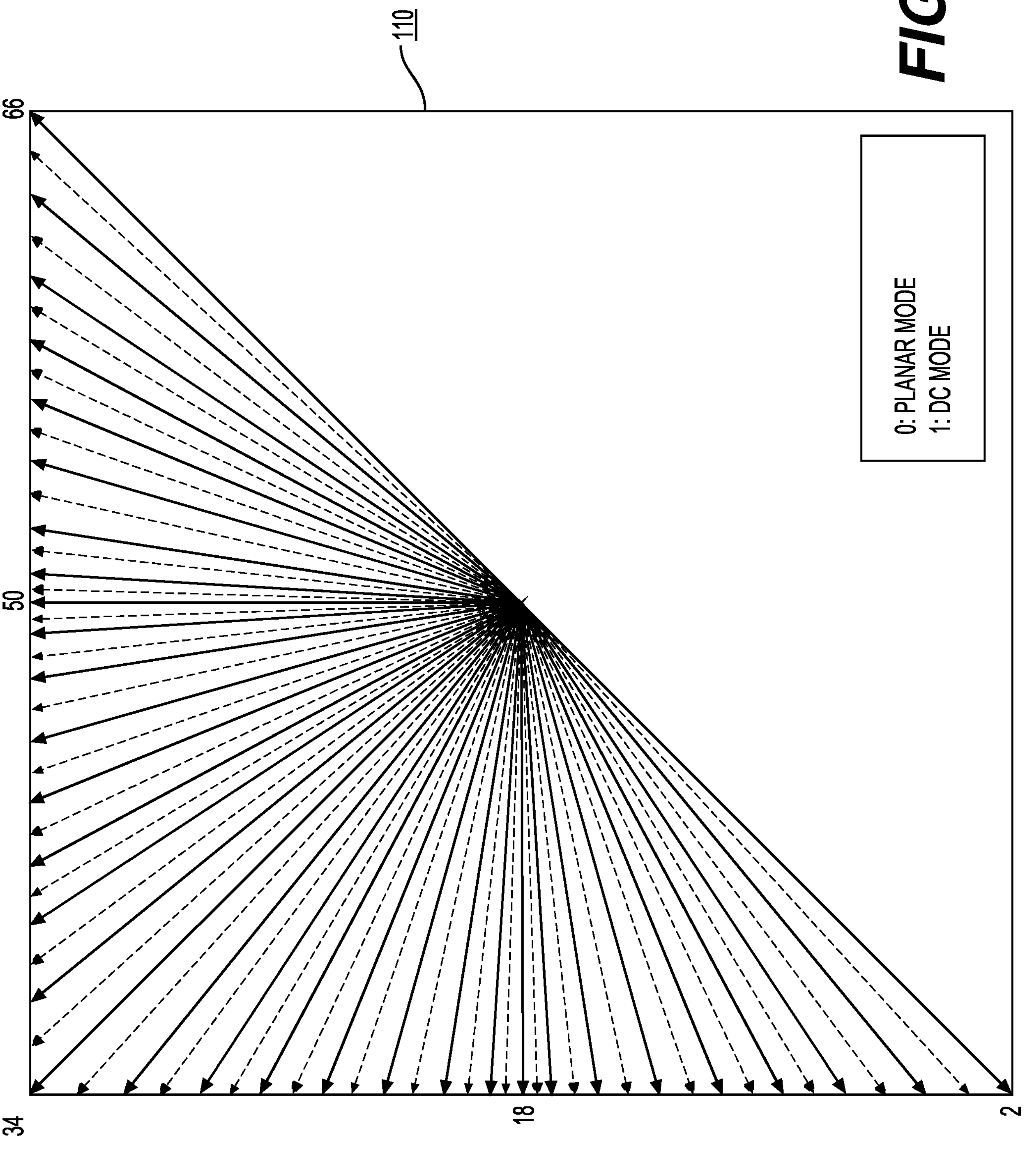

FIG. 1B

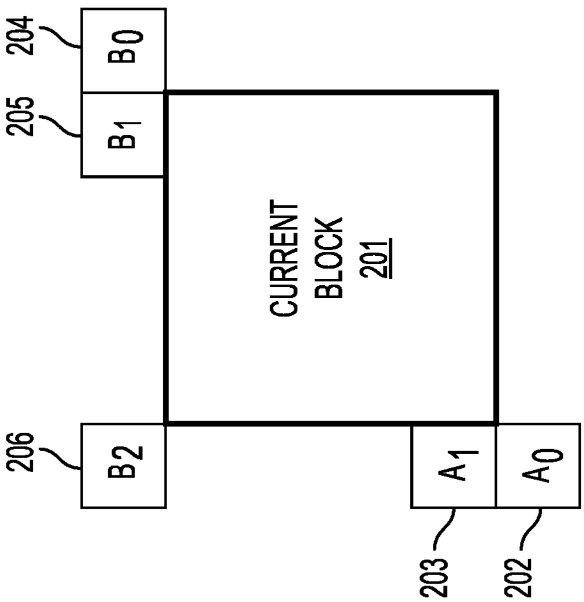

FIG. 2
(Related Art)

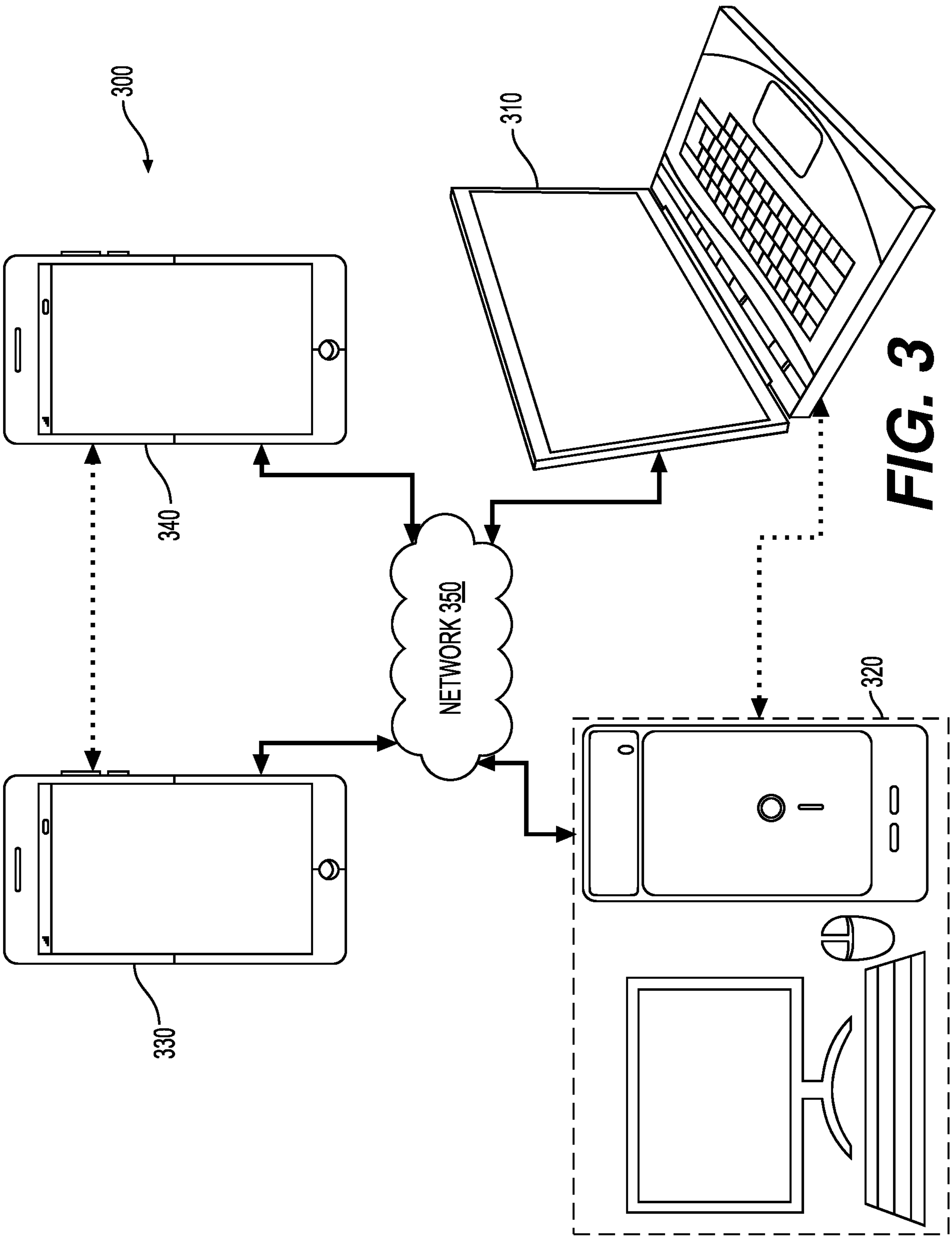

FIG. 3

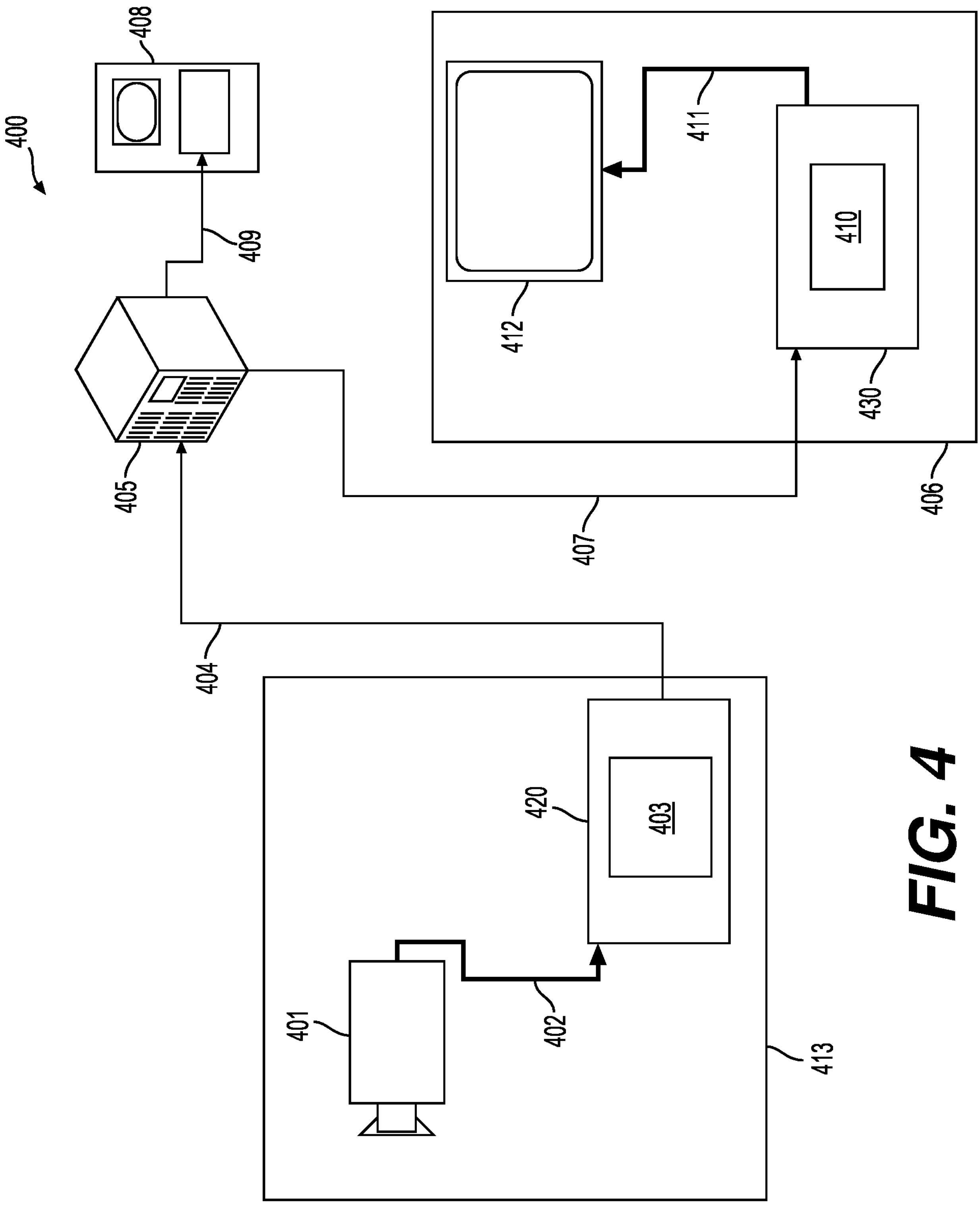

FIG. 4

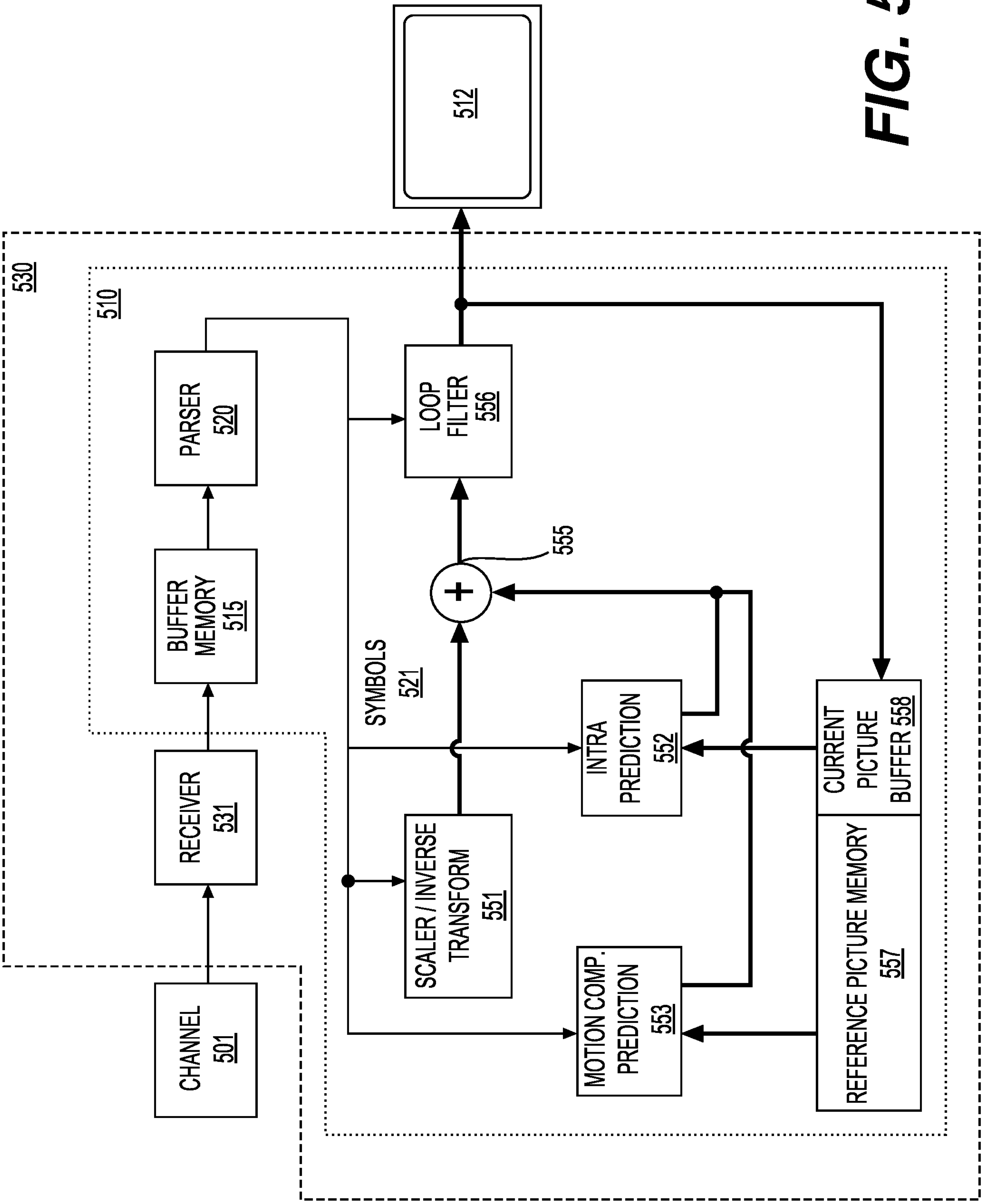
530
510
CHANNEL 501
RECEIVER 531
BUFFER MEMORY 515
PARSER 520
SYMBOLS 521
SCALER / INVERSE TRANSFORM 551
MOTION COMP. PREDICTION 553
INTRA PREDICTION 552
555
LOOP FILTER 556
REFERENCE PICTURE MEMORY 557
CURRENT PICTURE BUFFER 558
512

FIG. 5

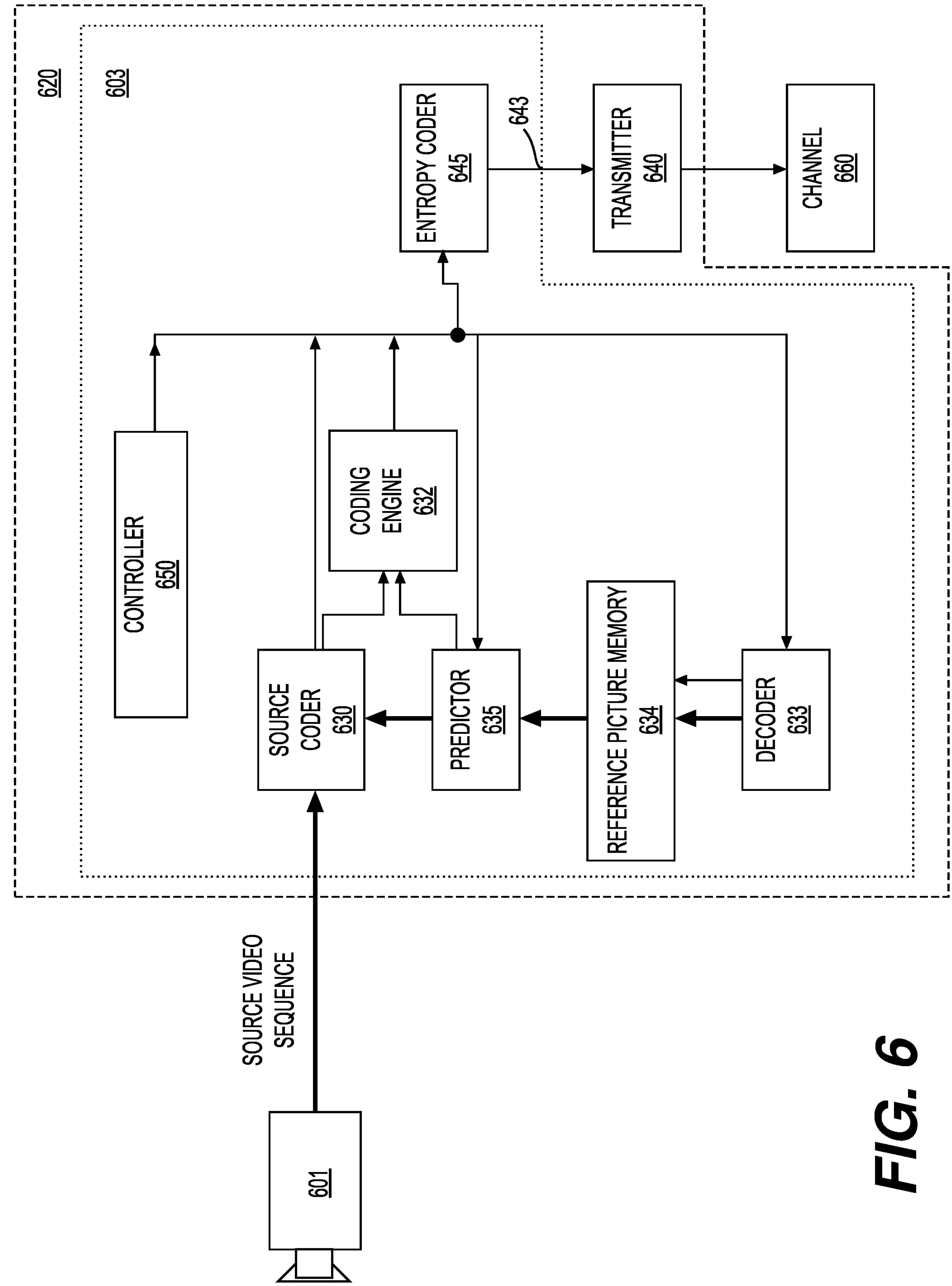
620
603
ENTROPY CODER 645
643
TRANSMITTER 640
CHANNEL 660
CONTROLLER 650
CODING ENGINE 632
SOURCE CODER 630
PREDICTOR 635
REFERENCE PICTURE MEMORY 634
DECODER 633
SOURCE VIDEO SEQUENCE
601

FIG. 6

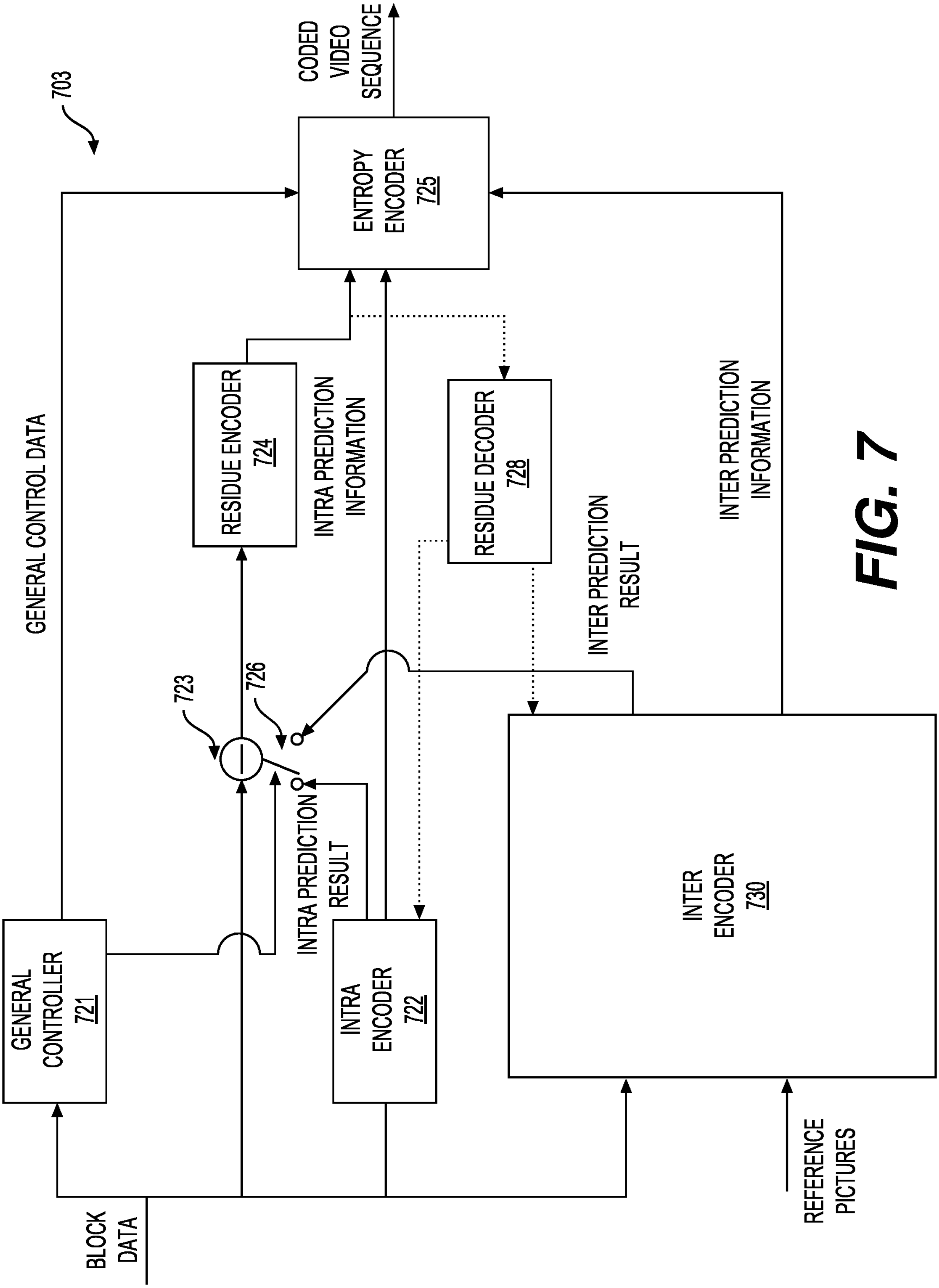
703
CODED VIDEO SEQUENCE
ENTROPY ENCODER 725
GENERAL CONTROL DATA
RESIDUE ENCODER 724
INTRA PREDICTION INFORMATION
RESIDUE DECODER 728
INTER PREDICTION RESULT
INTER PREDICTION INFORMATION
723
726
INTRA PREDICTION RESULT
GENERAL CONTROLLER 721
INTRA ENCODER 722
INTER ENCODER 730
BLOCK DATA
REFERENCE PICTURES

FIG. 7

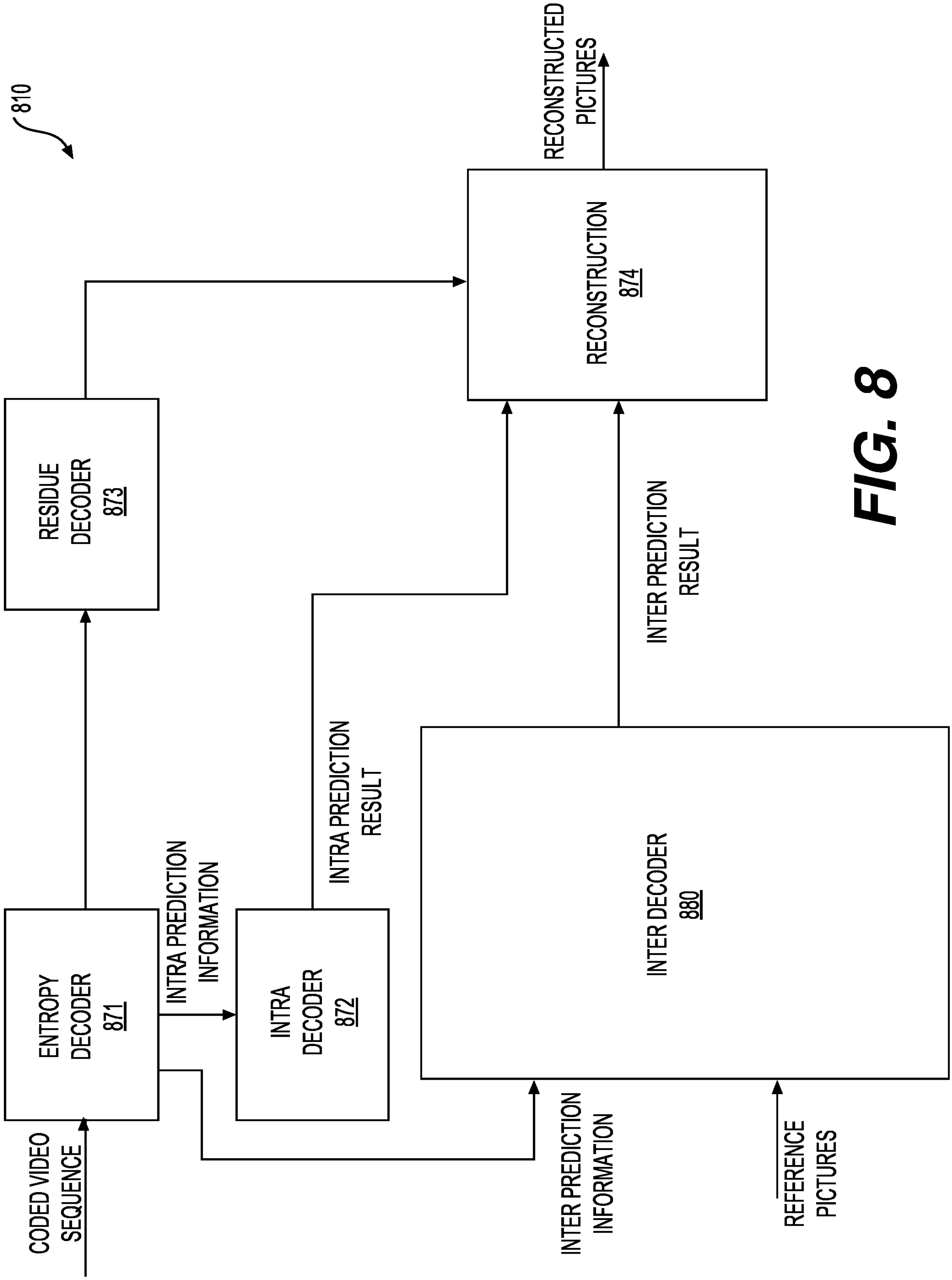
810
CODED VIDEO SEQUENCE
ENTROPY DECODER 871
INTRA PREDICTION INFORMATION
INTRA DECODER 872
INTRA PREDICTION RESULT
RESIDUE DECODER 873
RECONSTRUCTION 874
RECONSTRUCTED PICTURES
INTER PREDICTION INFORMATION
INTER DECODER 880
INTER PREDICTION RESULT
REFERENCE PICTURES

FIG. 8

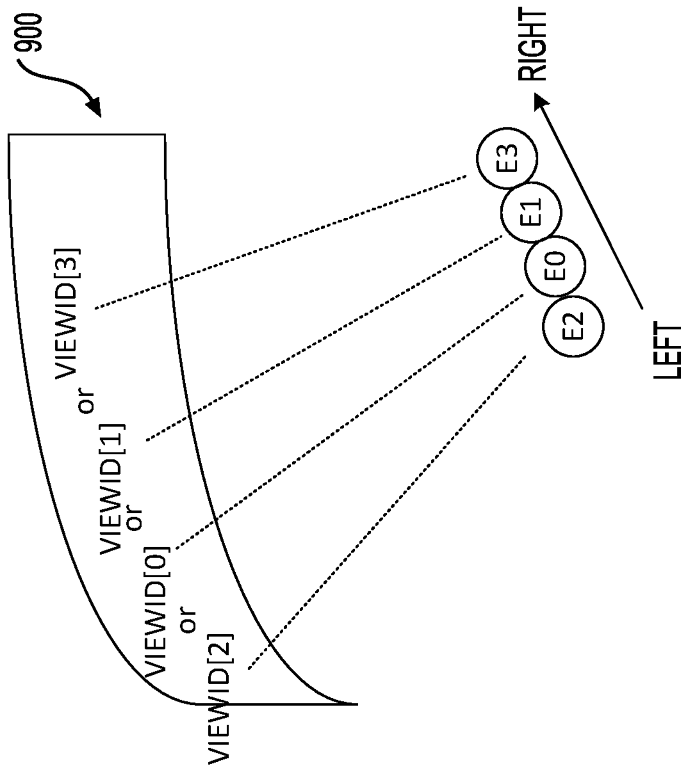
900
VIEWID[3]
or
VIEWID[1]
or
VIEWID[0]
or
VIEWID[2]
E3
E1
E0
E2
RIGHT
LEFT

FIG. 9

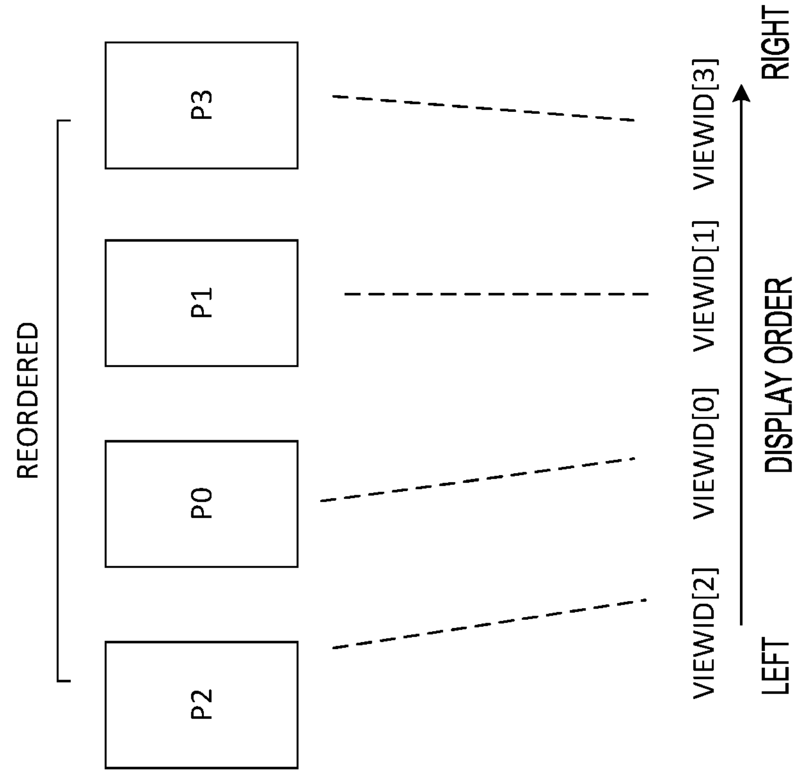

FIG. 10B

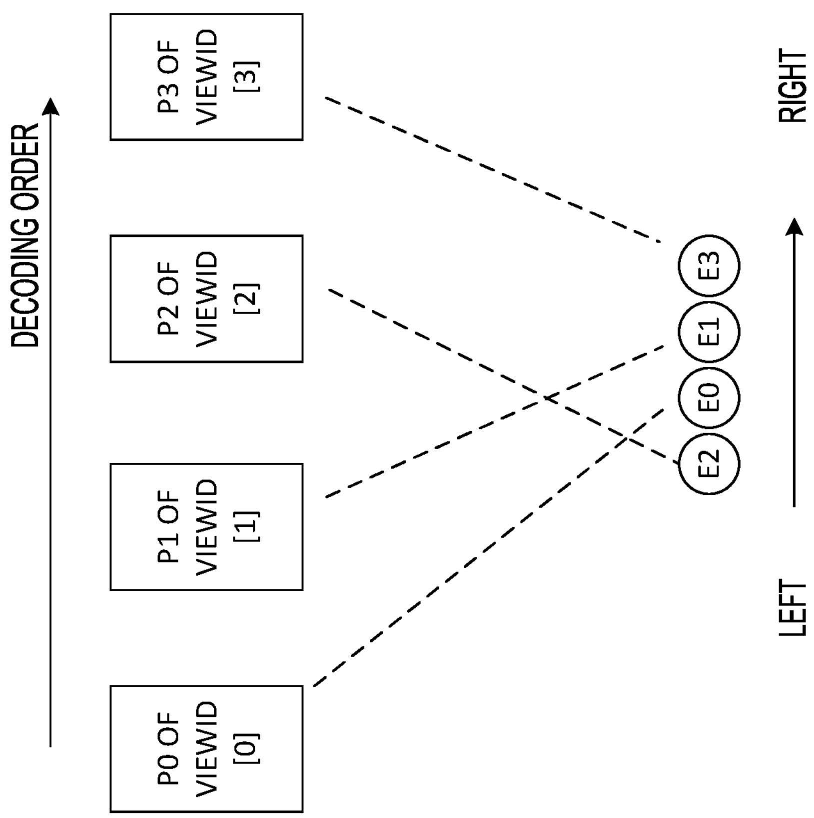

FIG. 10A

1100

	multiview_view_position(payloadSize) {	Descriptor
1110 →	num_views_minus1	ue(v)
	for(i = 0; i <= num_views_minus1; i++)	
1120 →	view_position[i]	ue(v)
	}	

FIG. 11

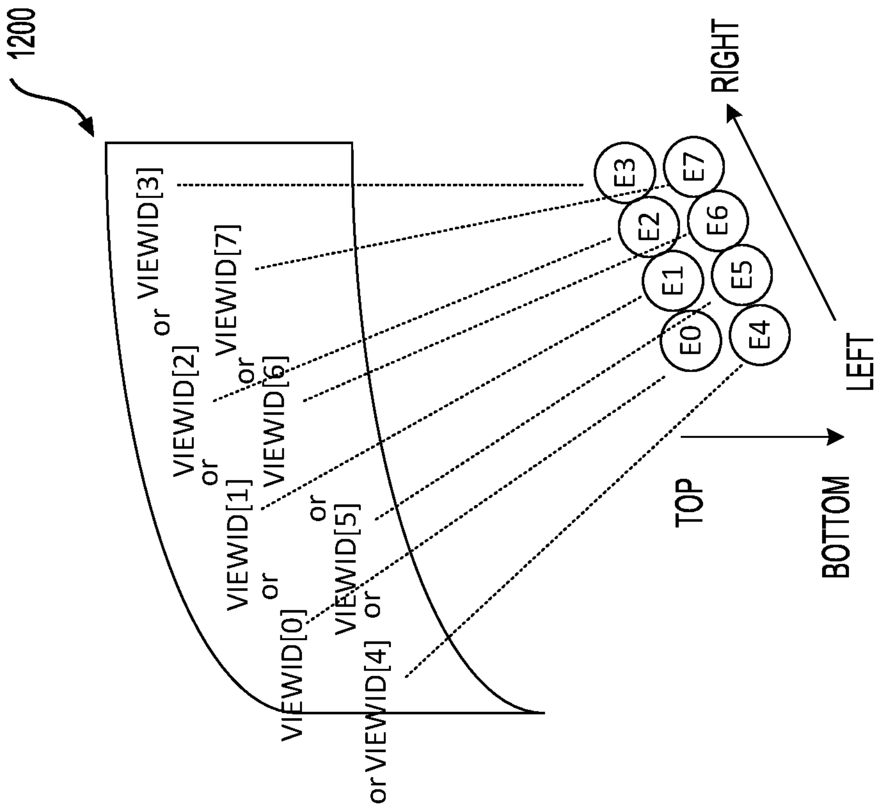
1200
VIEWID[3]
or VIEWID[7]
VIEWID[2]
or VIEWID[6]
VIEWID[1]
or VIEWID[5]
VIEWID[0]
or VIEWID[4]
E0
E1
E2
E3
E4
E5
E6
E7
RIGHT
LEFT
TOP
BOTTOM

FIG. 12

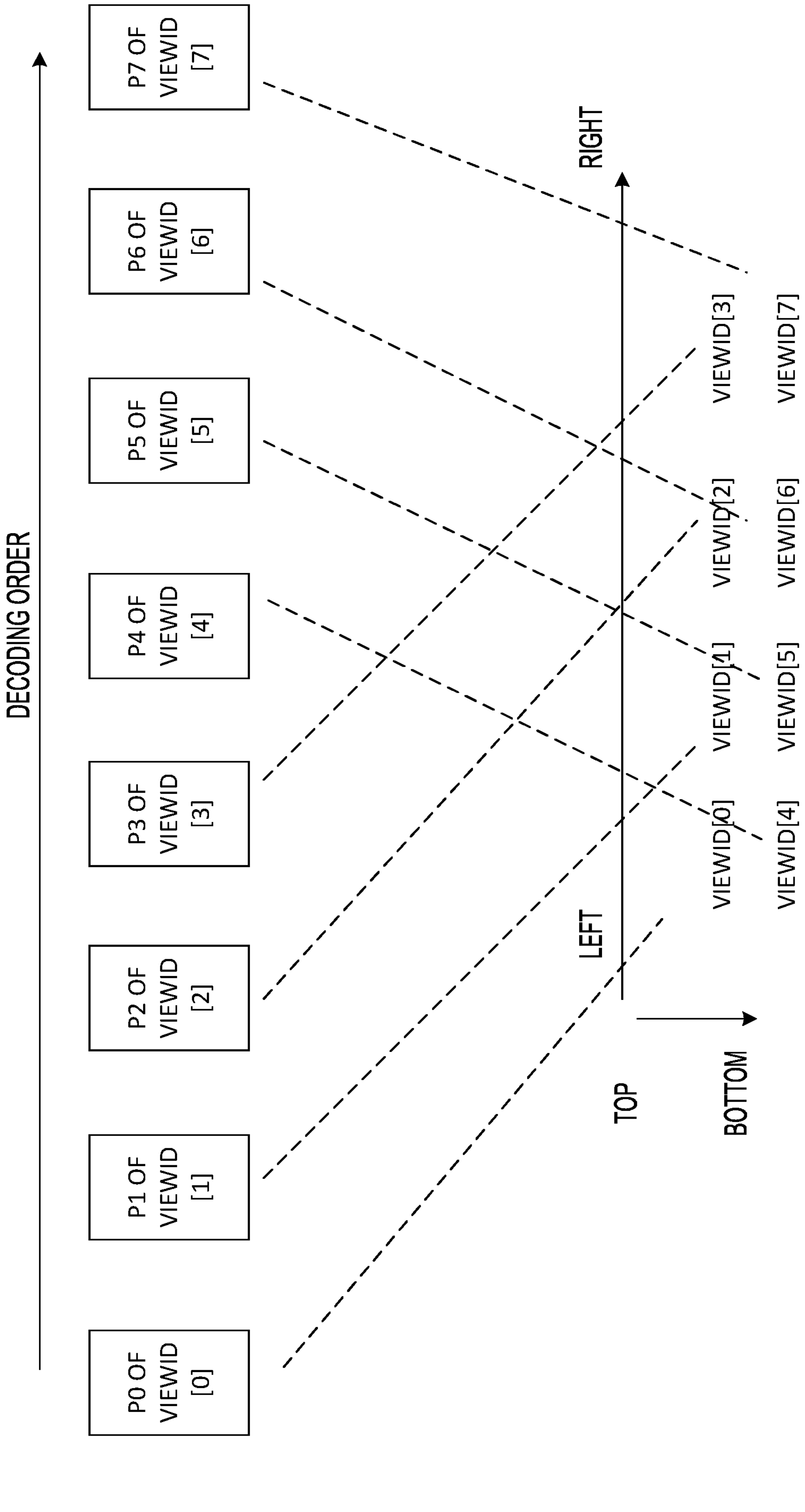
DECODING ORDER
P0 OF VIEWID [0]
P1 OF VIEWID [1]
P2 OF VIEWID [2]
P3 OF VIEWID [3]
P4 OF VIEWID [4]
P5 OF VIEWID [5]
P6 OF VIEWID [6]
P7 OF VIEWID [7]
LEFT
RIGHT
TOP
BOTTOM
VIEWID[0]
VIEWID[1]
VIEWID[2]
VIEWID[3]
VIEWID[4]
VIEWID[5]
VIEWID[6]
VIEWID[7]

FIG. 13

1400

multiview_view_position(payloadSize) {	**Descriptor**
num_views_minus1 (1410)	ue(v)
num_vertical_view_positions_minus1 (1420)	ue(v)
num_horizontal_view_positions_minus1 (1430)	ue(v)
for(i = 0; i <= num_views_minus1; i++) { (1460)	
view_position_y[i] (1440)	ue(v)
view_position_x[i] (1450)	ue(v)
}	
}	

FIG. 14

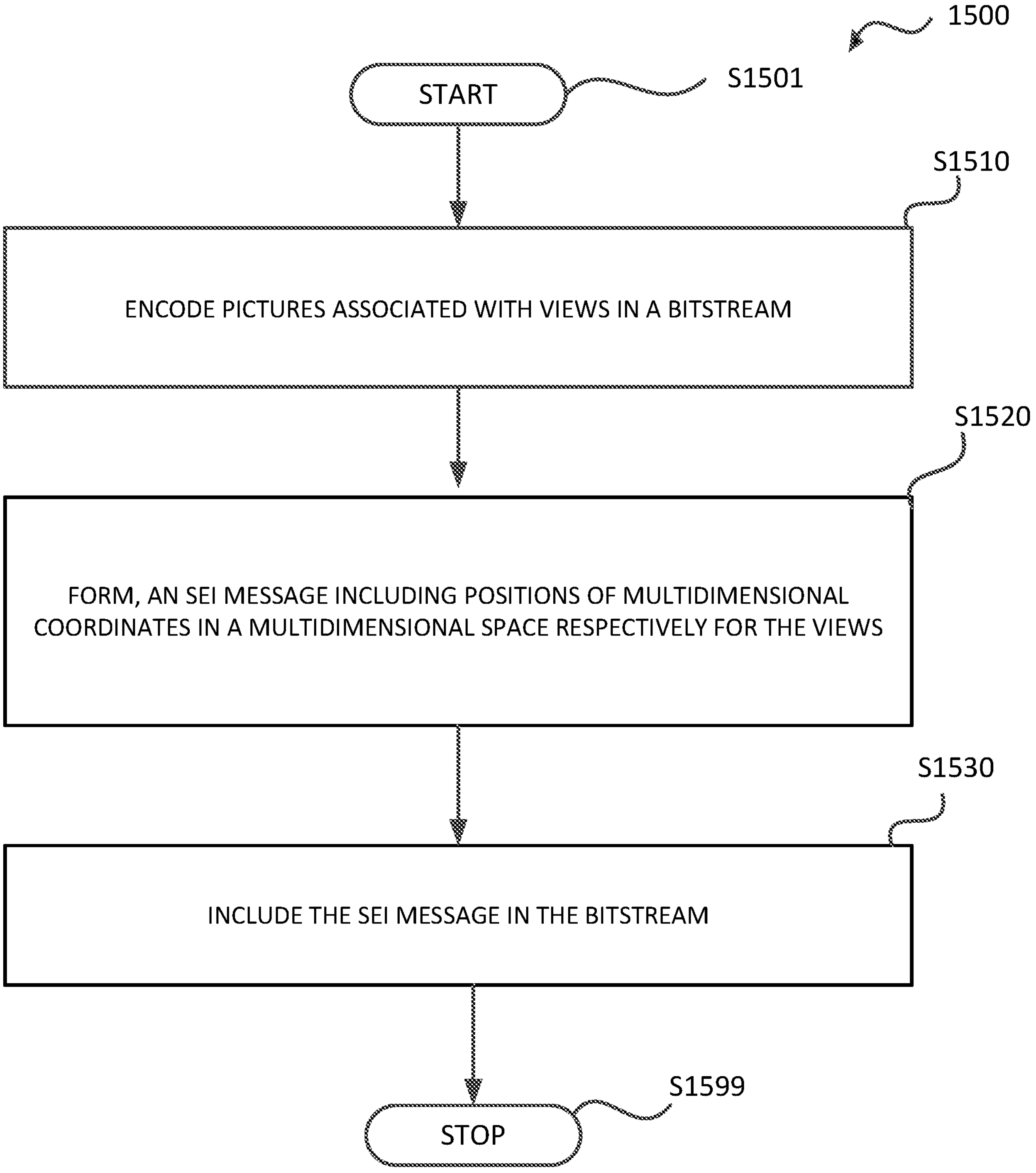
1500
START
S1501
S1510
ENCODE PICTURES ASSOCIATED WITH VIEWS IN A BITSTREAM
S1520
FORM, AN SEI MESSAGE INCLUDING POSITIONS OF MULTIDIMENSIONAL COORDINATES IN A MULTIDIMENSIONAL SPACE RESPECTIVELY FOR THE VIEWS
S1530
INCLUDE THE SEI MESSAGE IN THE BITSTREAM
S1599
STOP

FIG. 15

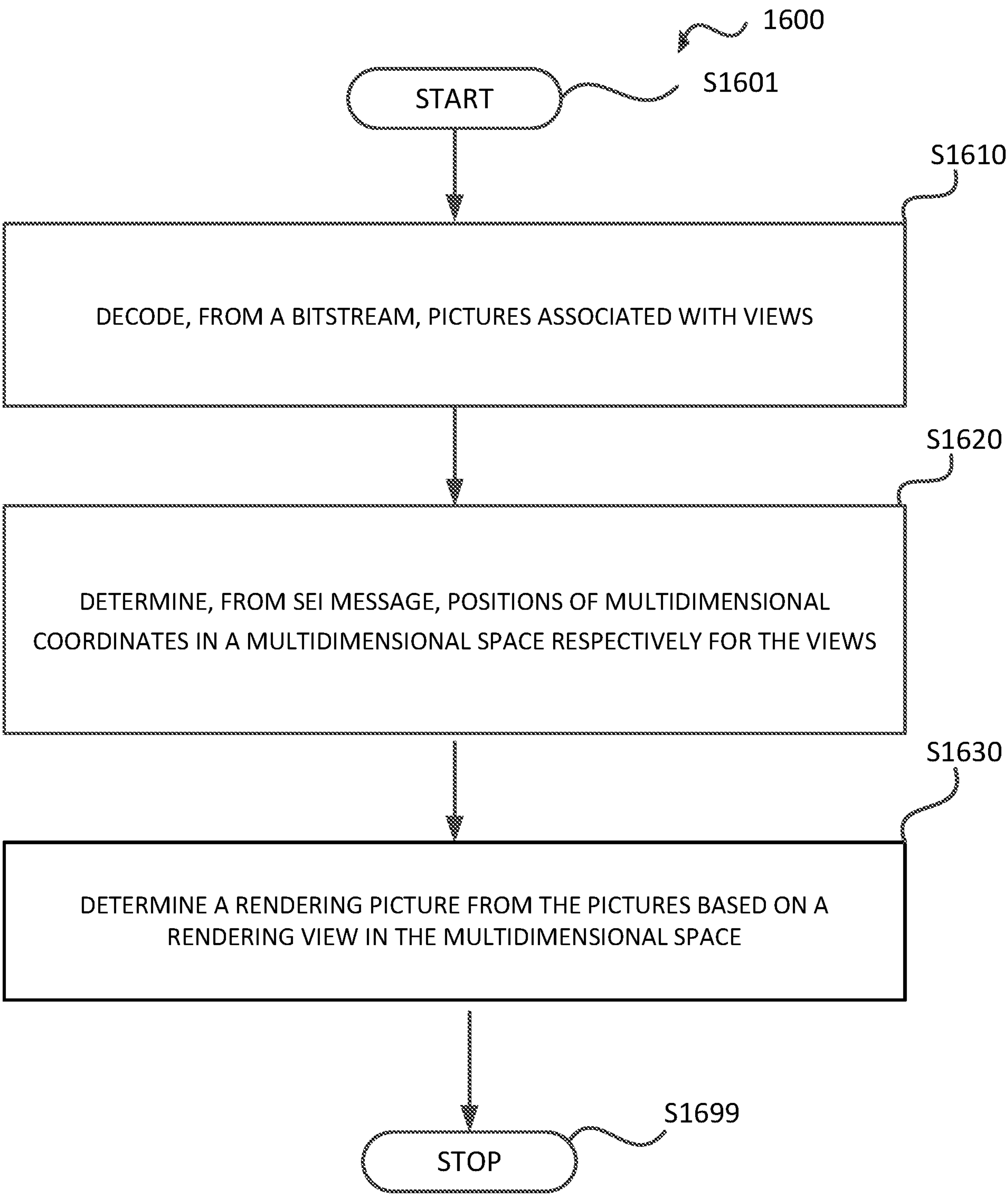
1600
START
S1601
S1610
DECODE, FROM A BITSTREAM, PICTURES ASSOCIATED WITH VIEWS
S1620
DETERMINE, FROM SEI MESSAGE, POSITIONS OF MULTIDIMENSIONAL COORDINATES IN A MULTIDIMENSIONAL SPACE RESPECTIVELY FOR THE VIEWS
S1630
DETERMINE A RENDERING PICTURE FROM THE PICTURES BASED ON A RENDERING VIEW IN THE MULTIDIMENSIONAL SPACE
S1699
STOP

FIG. 16

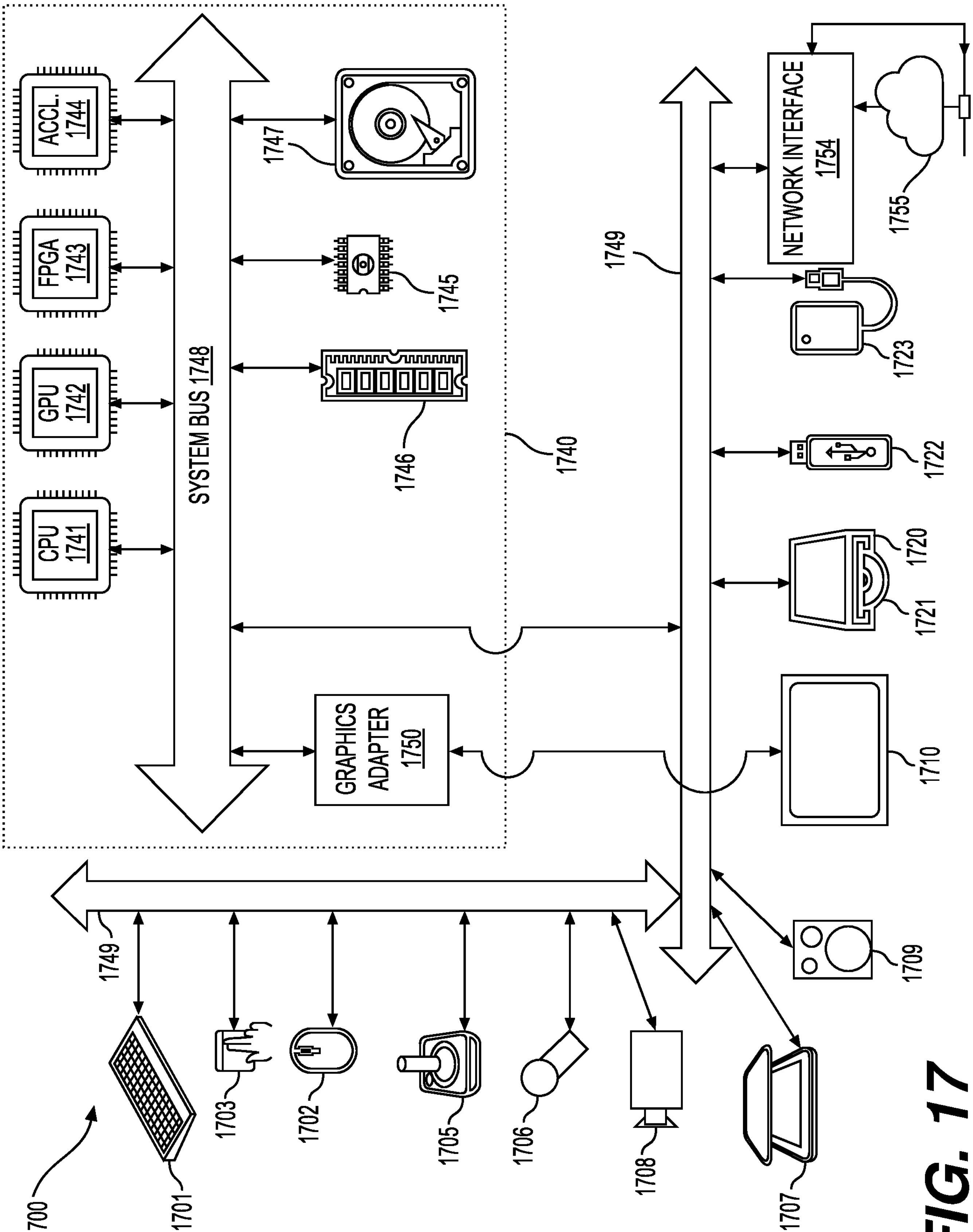
ACCL.
1744
FPGA
1743
GPU
1742
CPU
1741
SYSTEM BUS 1748
1747
1745
1746
1740
GRAPHICS
ADAPTER
1750
1749
NETWORK INTERFACE
1754
1755
1723
1722
1720
1721
1710
1709
1749
1700
1701
1703
1702
1705
1706
1708
1707

FIG. 17

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US22/72712

A. CLASSIFICATION OF SUBJECT MATTER
IPC - INV. H04N 19/597; H04N 19/30; H04N 19/44; H04N 19/70 (2022.01)

ADD.

CPC - INV. H04N 19/597; H04N 19/30; H04N 19/44; H04N 19/70

ADD.

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)
See Search History document

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2019/0253727 A1 (DOLBY LABORATORIES LICENSING CORPORATION) 15 August 2019; the entire document	1-20
A	US 2015/0304665 A1 (NOKIA CORPORATION) 22 October 2015; the entire document	1-20
A	US 2014/0086317 A1 (QUALCOMM INCORPORATED) 27 March 2014; the entire document	1-20

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

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"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)
"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
"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
"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
"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
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
07 August 2022 (07.08.2022)	AUG 19 2022

Name and mailing address of the ISA/US	Authorized officer
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Shane Thomas Telephone No. PCT Helpdesk: 571-272-4300