



(51) International Patent Classification:

H04N 19/70 (2014.01) *H04N 21/4728* (2011.01)
H04N 19/167 (2014.01)

(21) International Application Number:

PCT/JP2017/045270

(22) International Filing Date:

18 December 2017 (18.12.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/443,767 08 January 2017 (08.01.2017) US

(71) Applicant: SHARP KABUSHIKI KAISHA [JP/JP]; 1,
Takumi-cho, Sakai-ku, Sakai City, Osaka, 5908522 (JP).

(72) Inventor: DESHPANDE, Sachin G..

(74) Agent: HARA KENZO WORLD PATENT & TRADE-
MARK; Daiwa Minamimorimachi Building, 2-6, Tenjin-
bashi 2-chome Kita, Kita-ku, Osaka-shi, Osaka, 5300041
(JP).

(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, IR, IS, 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, NO, NZ,
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,

SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, 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).

Declarations under Rule 4.17:

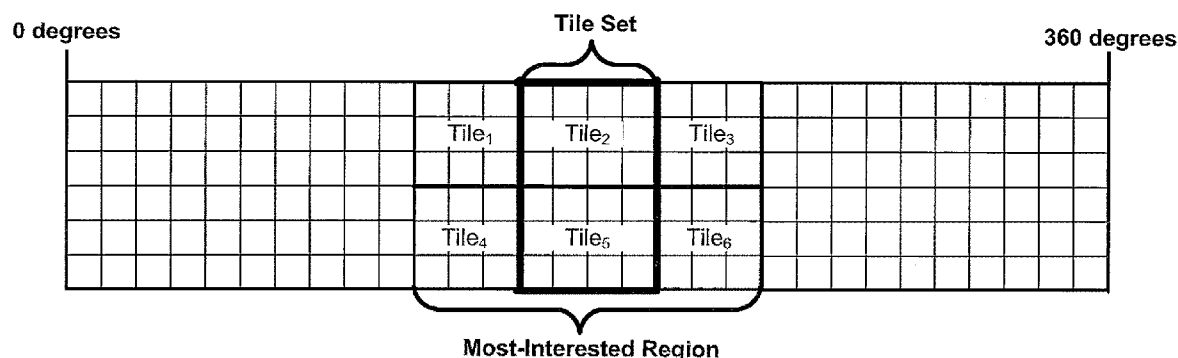
- as to applicant's entitlement to apply for and be granted a
patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the
earlier application (Rule 4.17(iii))

Published:

- with international search report (Art. 21(3))

(54) Title: SYSTEMS AND METHODS FOR SIGNALING OF TILES IN MOST-INTERESTED REGIONS FOR VIRTUAL REALITY APPLICATIONS

FIG. 2C



(57) Abstract: A device may be configured to signal tiles according to one or more of the techniques described herein.

Description

Title of Invention: SYSTEMS AND METHODS FOR SIGNALING OF TILES IN MOST-INTERESTED REGIONS FOR VIRTUAL REALITY APPLICATIONS

Technical Field

[0001] This disclosure relates to video coding and more particularly to techniques for signaling of tiles in most-interested regions of video.

Background Art

[0002] Digital video capabilities can be incorporated into a wide range of devices, including digital televisions, laptop or desktop computers, tablet computers, digital recording devices, digital media players, video gaming devices, cellular telephones, including so-called smartphones, medical imaging devices, head-mounted displays, and the like. Digital video may be coded according to a video coding standard. Video coding standards may incorporate video compression techniques. Examples of video coding standards include ISO/IEC MPEG-4 Visual and ITU-T H.264 (also known as ISO/IEC MPEG-4 AVC) and High-Efficiency Video Coding (HEVC). HEVC is described in High Efficiency Video Coding (HEVC), Rec. ITU-T H.265 April 2015, which is incorporated by reference, and referred to herein as ITU-T H.265. Video compression techniques enable data requirements for storing and transmitting video data to be reduced.

[0003] Video compression techniques may reduce data requirements by exploiting the inherent redundancies in a video sequence. Video compression techniques may subdivide a video sequence into successively smaller portions (i.e., groups of frames within a video sequence, a frame within a group of frames, slices within a frame, coding tree units (e.g., macroblocks) within a slice, coding blocks within a coding tree unit, etc.). Intra prediction coding techniques (e.g., intra-picture (spatial)) and inter prediction techniques (i.e., inter-picture (temporal)) may be used to generate difference values between a unit of video data to be coded and a reference unit of video data. The difference values may be referred to as residual data. Residual data may be coded as quantized transform coefficients. Syntax elements may relate residual data and a reference coding unit (e.g., intra-prediction mode indices, motion vectors, and block vectors). Residual data and syntax elements may be entropy coded. Entropy encoded residual data and syntax elements may be included in a compliant bitstream. Compliant bitstreams and associated metadata may be formatted according to data structures.

Summary of Invention

[0004] An aspect of the invention is a method of signaling of a most-interested region of a

omnidirectional video, the method comprising: signaling a most-interested region; and signaling a number of tile groups corresponding to the most interested region.

- [0005] An aspect of the invention is a method of performing sub-bitstream extraction of a motion-constrained tile set, the method comprising: receiving a message identifying a most-interested region of a omnidirectional video, wherein the message includes an identifier for a motion-constrained tile set corresponding to the most interested region; and performing sub-bitstream extraction based on the identified motion-constrained tile set.

Brief Description of Drawings

- [0006] [fig.1]FIG. 1 is a block diagram illustrating an example of a system that may be configured to encode and decode video data according to one or more techniques of this disclosure.
- [fig.2A]FIG. 2A is a conceptual diagram illustrating coded video data and corresponding data structures according to one or more techniques of this disclosure.
- [fig.2B]FIG. 2B is a conceptual diagram illustrating coded video data and corresponding data structures according to one or more techniques of this disclosure.
- [fig.2C]FIG. 2C is a conceptual diagram illustrating coded video data and corresponding data structures according to one or more techniques of this disclosure.
- [fig.3]FIG. 3 is a conceptual diagram illustrating a data structure encapsulating coded video data and corresponding metadata according to one or more techniques of this disclosure.
- [fig.4]FIG. 4 is a conceptual drawing illustrating an example of components that may be included in an implementation of a system that may be configured to encode and decode video data according to one or more techniques of this disclosure.
- [fig.5]FIG. 5 is a block diagram illustrating an example of a video encoder that may be configured to encode video data according to one or more techniques of this disclosure.
- [fig.6]FIG. 6 is a block diagram illustrating an example of a video decoder that may be configured to decode video data according to one or more techniques of this disclosure.

Description of Embodiments

- [0007] In general, this disclosure describes various techniques for coding video data. In particular, this disclosure describes techniques for signaling of tile groups, including, for example, individual tiles and/or motion-constrained tile sets (MCTS) in most-interested regions of omnidirectional video. Signaling of tiles according to the techniques described herein may be particularly useful for improving video distribution system performance by lowering transmission bandwidth and/or lowering coding complexity. It should be noted that although techniques of this disclosure are described with respect to ITU-T H.264 and ITU-T H.265, the techniques of this disclosure are

generally applicable to video coding. For example, the coding techniques described herein may be incorporated into video coding systems, (including video coding systems based on future video coding standards) including block structures, intra prediction techniques, inter prediction techniques, transform techniques, filtering techniques, and/or entropy coding techniques other than those included in ITU-T H.265. Thus, reference to ITU-T H.264 and ITU-T H.265 is for descriptive purposes and should not be construed to limit the scope of the techniques described herein. Further, it should be noted that incorporation by reference of documents herein should not be construed to limit or create ambiguity with respect to terms used herein. For example, in the case where an incorporated reference provides a different definition of a term than another incorporated reference and/or as the term is used herein, the term should be interpreted in a manner that broadly includes each respective definition and/or in a manner that includes each of the particular definitions in the alternative.

[0008] Video content typically includes video sequences comprised of a series of frames. A series of frames may also be referred to as a group of pictures (GOP). Each video frame or picture may include a one or more slices, where a slice includes a plurality of video blocks. A video block may be defined as the largest array of pixel values (also referred to as samples) that may be predictively coded. Video blocks may be ordered according to a scan pattern (e.g., a raster scan). A video encoder performs predictive encoding on video blocks and sub-divisions thereof. ITU-T H.264 specifies a macroblock including 16 x 16 luma samples. ITU-T H.265 specifies an analogous Coding Tree Unit (CTU) structure where a picture may be split into CTUs of equal size and each CTU may include Coding Tree Blocks (CTB) having 16 x 16, 32 x 32, or 64 x 64 luma samples. As used herein, the term video block may generally refer to an area of a picture or may more specifically refer to the largest array of pixel values that may be predictively coded, sub-divisions thereof, and/or corresponding structures. Further, according to ITU-T H.265, each video frame or picture may be partitioned to include one or more tiles, where a tile is a sequence of coding tree units corresponding to a rectangular area of a picture.

[0009] In ITU-T H.265, the CTBs of a CTU may be partitioned into Coding Blocks (CB) according to a corresponding quadtree block structure. According to ITU-T H.265, one luma CB together with two corresponding chroma CBs and associated syntax elements are referred to as a coding unit (CU). A CU is associated with a prediction unit (PU) structure defining one or more prediction units (PU) for the CU, where a PU is associated with corresponding reference samples. That is, in ITU-T H.265 the decision to code a picture area using intra prediction or inter prediction is made at the CU level and for a CU one or more predictions corresponding to intra prediction or inter prediction may be used to generate reference samples for CBs of the CU. In ITU-T

H.265, a PU may include luma and chroma prediction blocks (PBs), where square PBs are supported for intra prediction and rectangular PBs are supported for inter prediction. Intra prediction data (e.g., intra prediction mode syntax elements) or inter prediction data (e.g., motion data syntax elements) may associate PUs with corresponding reference samples. Residual data may include respective arrays of difference values corresponding to each component of video data (e.g., luma (Y) and chroma (Cb and Cr)). Residual data may be in the pixel domain. A transform, such as, a discrete cosine transform (DCT), a discrete sine transform (DST), an integer transform, a wavelet transform, or a conceptually similar transform, may be applied to pixel difference values to generate transform coefficients. It should be noted that in ITU-T H.265, CUs may be further sub-divided into Transform Units (TUs). That is, an array of pixel difference values may be sub-divided for purposes of generating transform coefficients (e.g., four 8 x 8 transforms may be applied to a 16 x 16 array of residual values corresponding to a 16 x 16 luma CB), such sub-divisions may be referred to as Transform Blocks (TBs). Transform coefficients may be quantized according to a quantization parameter (QP). Quantized transform coefficients (which may be referred to as level values) may be entropy coded according to an entropy encoding technique (e.g., content adaptive variable length coding (CAVLC), context adaptive binary arithmetic coding (CABAC), probability interval partitioning entropy coding (PIPE), etc.). Further, syntax elements, such as, a syntax element indicating a prediction mode, may also be entropy coded. Entropy encoded quantized transform coefficients and corresponding entropy encoded syntax elements may form a compliant bitstream that can be used to reproduce video data. A binarization process may be performed on syntax elements as part of an entropy coding process. Binarization refers to the process of converting a syntax value into a series of one or more bits. These bits may be referred to as “bins.”

Virtual Reality (VR) applications may include video content that may be rendered with a head-mounted display, where only the area of the spherical video that corresponds to the orientation of the user's head is rendered. VR applications may be enabled by omnidirectional video, which is also referred to as 360° spherical video. Omnidirectional video is typically captured by multiple cameras that cover up to 360° of a scene. A distinct feature of omnidirectional video compared to normal video is that, typically only a subset of the entire captured video region is displayed, i.e., the area corresponding to the current user's field of view (FOV) is displayed. A FOV is sometimes also referred to as viewport. In other cases, a viewport may be part of the spherical video that is currently displayed and viewed by the user. It should be noted that the size of the viewport can be smaller than or equal to the field of view.

[0010] A most-interested region in an omnidirectional video picture may refer to a subset of

the entire video region that is statistically the most likely to be rendered to the user at the presentation time of that picture (i.e., most likely to be in a FOV). It should be noted that most-interested regions of an omnidirectional video may be determined by the intent of a director or producer, or derived from user statistics by a service or content provider, e.g., through the statistics of which regions have been requested/seen the most by users when the omnidirectional video content was provided through a streaming service. Most-interested regions may be used for data pre-fetching in omnidirectional video adaptive streaming by edge servers or clients, and/or transcoding optimization when an omnidirectional video is transcoded, e.g., to a different codec or projection mapping. Thus, signaling most-interested regions in an omnidirectional video picture may improve system performance by lowering transmission bandwidth and lowering decoding complexity.

- [0011] Choi et al., ISO/IEC JTC1/SC29/WG11 N16439, “MPEG-A Part 20 (WD on ISO/IEC 23000-20): Omnidirectional Media Application Format,” October 2016, Chengdu, which is incorporated by reference and herein referred to as Choi, defines a media application format that enables omnidirectional media applications. Choi specifies how to store omnidirectional media and the associated metadata using the International Organization for Standardization (ISO) base media file format (ISOBMFF) and how to encapsulate, signal, and stream omnidirectional media using dynamic adaptive streaming over Hypertext Transfer Protocol (HTTP) (DASH). Choi describes metadata signaling for most-interested regions. Choi provides the following syntax for signaling of most-interested regions:

```

unsigned int(32) regionbase_id;
unsigned int(16) entry_count;
for (i=1; i<= entry_count; i++) {
    unsigned int(16) left_horizontal_offset;
    unsigned int(16) top_vertical_offset;
    unsigned int(16) region_width;
    unsigned int(16) region_height;
}

```

- [0012] In Choi, the following semantics are provided:
regionbase_id specifies the base region against which the positions and sizes of the most-interested regions are specified.
- [0013] **entry_count** specifies the number of entries.
- [0014] **left_horizontal_offset**, **top_vertical_offset**, **region_width**, and **region_height** are integer values that indicate the position and size of the most-interested region. **left_horizontal_offset** and **top_vertical_offset** indicate the horizontal and vertical co-

ordinates, respectively, in luma samples, of the upper left corner of the most-interested region in relative to the base region. `region_width` and `region_height` indicate the width and height, respectively, in luma samples, of the most-interested region in relative to the base region.

- [0015] It should be noted that a base region generally refers to an overall region of coded video data, e.g., the entire video region. Further, it should be noted that in the syntax above and the syntax used herein, `unsigned int(n)` refers to an unsigned integer having `n`-bits. As illustrated by the syntax above, although the size and position of a most-interested region may be signaled in Choi, the signaling in Choi fails to provide any additional information about the video included in a most-interested region, such signaling may be less than ideal.
- [0016] As described above, according to ITU-T H.265, each video frame or picture may be partitioned to include one or more slices and further partitioned to include one or more tiles. FIGS. 2A-2C are conceptual diagrams illustrating an example of a group of pictures including slices and further partitioning pictures into tiles. In the example illustrated in FIG. 2A, `Pic4` is illustrated as including two slices (i.e., `Slice1` and `Slice2`) where each slice includes a sequence of CTUs (e.g., in raster scan order). In the example illustrated in FIG. 2B, `Pic4` is illustrated as including six tiles (i.e., `Tile1` to `Tile6`), where each tile is rectangular and includes a sequence of CTUs. It should be noted that in ITU-T H.265, a tile may consist of coding tree units contained in more than one slice and a slice may consist of coding tree units contained in more than one tile. However, ITU-T H.265 provides that one or both of the following conditions shall be fulfilled: (1) All coding tree units in a slice belong to the same tile; and (2) All coding tree units in a tile belong to the same slice. Thus, with respect to FIG. 2B, each of the tiles may belong to a respective slice (e.g., `Tile1` to `Tile6` may respectively belong to slices, `Slice1` to `Slice6`) or multiple tiles may belong to a slice (e.g., `Tile1` to `Tile3` may belong to `Slice1` and `Tile4` to `Tile6` may belong to `Slice2`).
- [0017] Further, as illustrated in FIG. 2B, tiles may form tile sets (i.e., `Tile2` and `Tile5` form a tile set). Tile sets may be used to define boundaries for coding dependencies (e.g., intra-prediction dependencies, entropy encoding dependencies, etc.) and as such, may enable parallelism in coding and region-of-interest coding. For example, if the video sequence in the example illustrated in FIG. 2B corresponds to a nightly news program, the tile set formed by `Tile2` and `Tile5` may correspond to a visual region-of-interest including a news anchor reading the news. ITU-T H.265 defines signaling that enables motion-constrained tile sets (MCTS). A motion-constrained tile set may include a tile set for which inter-picture prediction dependencies are limited to the collocated tile sets in reference pictures. Thus, it is possible to perform motion compensation for a given MCTS independent of the decoding of other tile sets outside the MCTS. For

example, referring to FIG. 2B, if the tile set formed by Tile₂ and Tile₅ is a MCTS and each of Pic₁ to Pic₃ include collocated tile sets, motion compensation may be performed on Tile₂ and Tile₅ independent of coding Tile₁, Tile₃, Tile₄, and Tile₆ in Pic₄ and tiles collocated with tiles Tile₁, Tile₃, Tile₄, and Tile₆ in each of Pic₁ to Pic₃. Coding video data according to MCTS may be useful for video applications including omnidirectional video presentations.

[0018] As illustrated in FIG. 2C, Tile₁ to Tile₆ may form a most-interested region of an omnidirectional video. Further, the tile set formed by Tile₂ and Tile₅ may be a MCTS included within the most-interested region. Viewport dependent video coding, which may also be referred to as viewport dependent partial video coding, may be used to enable coding of only part of an entire video region. That is, for example, viewport dependent video coding may be used to provide sufficient information for rendering of a current FOV. For example, omnidirectional video may be coded using MCTS, such that each potential region covering a viewport can be independently coded from other regions across time. In this case, for example, for a particular current viewport, a minimum set of tiles that cover a viewport may be sent to the client, decoded, and/or rendered. This process may be referred to as simple tile based partial decoding (STPD).

[0019] In ITU-T H.265, a coded video sequence (CVS) may be encapsulated (or structured) as a sequence of access units, where each access unit includes video data structured as network abstraction layer (NAL) units. In ITU-T H.265, a bitstream is described as including a sequence of NAL units forming one or more CVSs. It should be noted that ITU-T H.265 supports multi-layer extensions, including format range extensions (RExt), scalability (SHVC), multi-view (MV-HEVC), and 3-D (3D-HEVC). Multi-layer extensions enable a video presentation to include a base layer and one or more additional enhancement layers. For example, a base layer may enable a video presentation having a basic level of quality (e.g., High Definition rendering) to be presented and an enhancement layer may enable a video presentation having an enhanced level of quality (e.g., an Ultra High Definition rendering) to be presented. In ITU-T H.265, an enhancement layer may be coded by referencing a base layer. That is, for example, a picture in an enhancement layer may be coded (e.g., using inter prediction techniques) by referencing one or more pictures (including scaled versions thereof) in a base layer. In ITU-T H.265, each NAL unit may include an identifier indicating a layer of video data the NAL unit is associated with. Referring to the example illustrated in FIG. 2A, each slice of video data included in Pic₄ (i.e., Slice₁ and Slice₂) is illustrated as being encapsulated in a NAL unit. Further, in ITU-T H.265 each of a video sequence, a GOP, a picture, a slice, and CTU may be associated with metadata that describes video coding properties. ITU-T H.265 defines parameters sets that may

be used to describe video data and/or video coding properties. In ITU-T H.265, parameter sets may be encapsulated as a special type of NAL unit or may be signaled as a message. NAL units including coded video data (e.g., a slice) may be referred to as VCL (Video Coding Layer) NAL units and NAL units including metadata (e.g., parameter sets) may be referred to as non-VCL NAL units. Further, ITU-T H.265 enables supplemental enhancement information (SEI) messages to be signaled. In ITU-T H.265, SEI messages assist in processes related to decoding, display or other purposes, however, SEI messages may not be required for constructing the luma or chroma samples by the decoding process. In ITU-T H.265, SEI messages may be signaled in a bitstream using non-VCL NAL units. Further, SEI messages may be conveyed by some means other than by being present in the bitstream (i.e., signaled out-of-band).

[0020] FIG. 3 illustrates an example of a bitstream including multiple CVSs, where a CVS is represented by NAL units included in a respective access unit. In the example, illustrated in FIG. 3, non-VCL NAL units include respective parameter set units (i.e., Video Parameter Sets (VPS), Sequence Parameter Sets (SPS), and Picture Parameter Set (PPS) units) and an access unit delimiter NAL unit. It should be noted that ITU-T H.265 defines NAL unit header semantics that specify the type of Raw Byte Sequence Payload (RBSP) data structure included in the NAL unit. As described above, omnidirectional video may be coded using MCTS. Sub-bitstream extraction may refer to a process where a device receiving a ITU-T H.265 compliant bitstream forms a new ITU-T H.265 compliant bitstream by discarding and/or modifying data in the received bitstream. For example, as described above, for a particular current viewport, a minimum set of tiles that cover a viewport may be sent to the client. Sub-bitstream extraction may be used to form a new ITU-T H.265 compliant bitstream including the minimum set of tiles. For example, referring to FIG. 2C, if a viewport includes only Tile₂ and Tile₅ and an access unit in a bitstream includes VCL NAL units for Tile₁ to Tile₆, a sub-bitstream extraction process may include generating a new bitstream that only includes VCL NAL units for Tile₂ and Tile₅ (i.e., VCL NAL units for Tile₁, Tile₃, Tile₄, and Tile₆ are removed from the received bitstream). Skupin, “Motion-Constrained Tile Sets Extraction Information SEI Messages Draft 1,” JCTVC-Y1008, October 2016, Chengdu, CN, which is incorporated by reference, and referred to herein as Skupin, describes potential modifications to the ITU-T H.265 specification to enable extraction of a motion-constrained tile set (MCTS) as an individual ITU-T H.265 conformant bitstream. In particular, Skupin specifies a (1) An MCTS extraction information set SEI message that provides syntax for carriage of MCTS specific replacement parameter sets and defines the extraction process in the semantics (referred to as `mcts_extraction_info_set`); and (2) An MCTS extraction in-

formation nesting SEI message, provides syntax for MCTS specific nested SEI messages (`mcts_extraction_info_nesting`).

[0021] FIG. 1 is a block diagram illustrating an example of a system that may be configured to code (i.e., encode and/or decode) video data according to one or more techniques of this disclosure. System 100 represents an example of a system that may encapsulate video data according to one or more techniques of this disclosure. As illustrated in FIG. 1, system 100 includes source device 102, communications medium 110, and destination device 120. In the example illustrated in FIG. 1, source device 102 may include any device configured to encode video data and transmit encoded video data to communications medium 110. Destination device 120 may include any device configured to receive encoded video data via communications medium 110 and to decode encoded video data. Source device 102 and/or destination device 120 may include computing devices equipped for wired and/or wireless communications and may include, for example, set top boxes, digital video recorders, televisions, desktop, laptop or tablet computers, gaming consoles, medical imaging devices, and mobile devices, including, for example, smartphones, cellular telephones, personal gaming devices.

[0022] Communications medium 110 may include any combination of wireless and wired communication media, and/or storage devices. Communications medium 110 may include coaxial cables, fiber optic cables, twisted pair cables, wireless transmitters and receivers, routers, switches, repeaters, base stations, or any other equipment that may be useful to facilitate communications between various devices and sites. Communications medium 110 may include one or more networks. For example, communications medium 110 may include a network configured to enable access to the World Wide Web, for example, the Internet. A network may operate according to a combination of one or more telecommunication protocols. Telecommunications protocols may include proprietary aspects and/or may include standardized telecommunication protocols. Examples of standardized telecommunications protocols include Digital Video Broadcasting (DVB) standards, Advanced Television Systems Committee (ATSC) standards, Integrated Services Digital Broadcasting (ISDB) standards, Data Over Cable Service Interface Specification (DOCSIS) standards, Global System Mobile Communications (GSM) standards, code division multiple access (CDMA) standards, 3rd Generation Partnership Project (3GPP) standards, European Telecommunications Standards Institute (ETSI) standards, Internet Protocol (IP) standards, Wireless Application Protocol (WAP) standards, and Institute of Electrical and Electronics Engineers (IEEE) standards.

[0023] Storage devices may include any type of device or storage medium capable of storing data. A storage medium may include a tangible or non-transitory computer-readable

media. A computer readable medium may include optical discs, flash memory, magnetic memory, or any other suitable digital storage media. In some examples, a memory device or portions thereof may be described as non-volatile memory and in other examples portions of memory devices may be described as volatile memory. Examples of volatile memories may include random access memories (RAM), dynamic random access memories (DRAM), and static random access memories (SRAM). Examples of non-volatile memories may include magnetic hard discs, optical discs, floppy discs, flash memories, or forms of electrically programmable memories (EPROM) or electrically erasable and programmable (EEPROM) memories. Storage device(s) may include memory cards (e.g., a Secure Digital (SD) memory card), internal/external hard disk drives, and/or internal/external solid state drives. Data may be stored on a storage device according to a defined file format.

[0024] FIG. 4 is a conceptual drawing illustrating an example of components that may be included in an implementation of system 100. In the example implementation illustrated in FIG. 4, system 100 includes one or more computing devices 402A-402N, television service network 404, television service provider site 406, wide area network 408, local area network 410, and one or more content provider sites 412A-412N. The implementation illustrated in FIG. 4 represents an example of a system that may be configured to allow digital media content, such as, for example, a movie, a live sporting event, etc., and data and applications and media presentations associated therewith to be distributed to and accessed by a plurality of computing devices, such as computing devices 402A-402N. In the example illustrated in FIG. 4, computing devices 402A-402N may include any device configured to receive data from one or more of television service network 404, wide area network 408, and/or local area network 410. For example, computing devices 402A-402N may be equipped for wired and/or wireless communications and may be configured to receive services through one or more data channels and may include televisions, including so-called smart televisions, set top boxes, and digital video recorders. Further, computing devices 402A-402N may include desktop, laptop, or tablet computers, gaming consoles, mobile devices, including, for example, “smart” phones, cellular telephones, and personal gaming devices.

[0025] Television service network 404 is an example of a network configured to enable digital media content, which may include television services, to be distributed. For example, television service network 404 may include public over-the-air television networks, public or subscription-based satellite television service provider networks, and public or subscription-based cable television provider networks and/or over the top or Internet service providers. It should be noted that although in some examples television service network 404 may primarily be used to enable television services to

be provided, television service network 404 may also enable other types of data and services to be provided according to any combination of the telecommunication protocols described herein. Further, it should be noted that in some examples, television service network 404 may enable two-way communications between television service provider site 406 and one or more of computing devices 402A-402N. Television service network 404 may comprise any combination of wireless and/or wired communication media. Television service network 404 may include coaxial cables, fiber optic cables, twisted pair cables, wireless transmitters and receivers, routers, switches, repeaters, base stations, or any other equipment that may be useful to facilitate communications between various devices and sites. Television service network 404 may operate according to a combination of one or more telecommunication protocols. Telecommunications protocols may include proprietary aspects and/or may include standardized telecommunication protocols. Examples of standardized telecommunications protocols include DVB standards, ATSC standards, ISDB standards, DTMB standards, DMB standards, Data Over Cable Service Interface Specification (DOCSIS) standards, HbbTV standards, W3C standards, and UPnP standards.

[0026] Referring again to FIG. 4, television service provider site 406 may be configured to distribute television service via television service network 404. For example, television service provider site 406 may include one or more broadcast stations, a cable television provider, or a satellite television provider, or an Internet-based television provider. For example, television service provider site 406 may be configured to receive a transmission including television programming through a satellite uplink/downlink. Further, as illustrated in FIG. 4, television service provider site 406 may be in communication with wide area network 408 and may be configured to receive data from content provider sites 412A-412N. It should be noted that in some examples, television service provider site 406 may include a television studio and content may originate therefrom.

[0027] Wide area network 408 may include a packet based network and operate according to a combination of one or more telecommunication protocols. Telecommunications protocols may include proprietary aspects and/or may include standardized telecommunication protocols. Examples of standardized telecommunications protocols include Global System Mobile Communications (GSM) standards, code division multiple access (CDMA) standards, 3rd Generation Partnership Project (3GPP) standards, European Telecommunications Standards Institute (ETSI) standards, European standards (EN), IP standards, Wireless Application Protocol (WAP) standards, and Institute of Electrical and Electronics Engineers (IEEE) standards, such as, for example, one or more of the IEEE 802 standards (e.g., Wi-Fi). Wide area network 408

may comprise any combination of wireless and/or wired communication media. Wide area network 480 may include coaxial cables, fiber optic cables, twisted pair cables, Ethernet cables, wireless transmitters and receivers, routers, switches, repeaters, base stations, or any other equipment that may be useful to facilitate communications between various devices and sites. In one example, wide area network 408 may include the Internet. Local area network 410 may include a packet based network and operate according to a combination of one or more telecommunication protocols. Local area network 410 may be distinguished from wide area network 408 based on levels of access and/or physical infrastructure. For example, local area network 410 may include a secure home network.

[0028] Referring again to FIG. 4, content provider sites 412A-412N represent examples of sites that may provide multimedia content to television service provider site 406 and/or computing devices 402A-402N. For example, a content provider site may include a studio having one or more studio content servers configured to provide multimedia files and/or streams to television service provider site 406. In one example, content provider sites 412A-412N may be configured to provide multimedia content using the IP suite. For example, a content provider site may be configured to provide multimedia content to a receiver device according to Real Time Streaming Protocol (RTSP), HTTP, or the like. Further, content provider sites 412A-412N may be configured to provide data, including hypertext based content, and the like, to one or more of receiver devices computing devices 402A-402N and/or television service provider site 406 through wide area network 408. Content provider sites 412A-412N may include one or more web servers. Data provided by data provider site 412A-412N may be defined according to data formats.

[0029] Referring again to FIG. 1, source device 102 includes video source 104, video encoder 106, data encapsulator 107, and interface 108. Video source 104 may include any device configured to capture and/or store video data. For example, video source 104 may include a video camera and a storage device operably coupled thereto. Video encoder 106 may include any device configured to receive video data and generate a compliant bitstream representing the video data. A compliant bitstream may refer to a bitstream that a video decoder can receive and reproduce video data therefrom. Aspects of a compliant bitstream may be defined according to a video coding standard. When generating a compliant bitstream video encoder 106 may compress video data. Compression may be lossy (discernible or indiscernible to a viewer) or lossless. FIG. 5 is a block diagram illustrating an example of video encoder 500 that may implement the techniques for encoding video data described herein. It should be noted that although example video encoder 500 is illustrated as having distinct functional blocks, such an illustration is for descriptive purposes and does not limit video encoder 500

and/or sub-components thereof to a particular hardware or software architecture.

Functions of video encoder 500 may be realized using any combination of hardware, firmware, and/or software implementations.

[0030] Video encoder 500 may perform intra prediction coding and inter prediction coding of picture areas, and, as such, may be referred to as a hybrid video encoder. In the example illustrated in FIG. 5, video encoder 500 receives source video blocks. In some examples, source video blocks may include areas of picture that has been divided according to a coding structure. For example, source video data may include macroblocks, CTUs, CBs, sub-divisions thereof, and/or another equivalent coding unit. In some examples, video encoder 500 may be configured to perform additional sub-divisions of source video blocks. It should be noted that the techniques described herein are generally applicable to video coding, regardless of how source video data is partitioned prior to and/or during encoding. In the example illustrated in FIG. 5, video encoder 500 includes summer 502, transform coefficient generator 504, coefficient quantization unit 506, inverse quantization and transform coefficient processing unit 508, summer 510, intra prediction processing unit 512, inter prediction processing unit 514, and entropy encoding unit 516. As illustrated in FIG. 5, video encoder 500 receives source video blocks and outputs a bitstream.

[0031] In the example illustrated in FIG. 5, video encoder 500 may generate residual data by subtracting a predictive video block from a source video block. The selection of a predictive video block is described in detail below. Summer 502 represents a component configured to perform this subtraction operation. In one example, the subtraction of video blocks occurs in the pixel domain. Transform coefficient generator 504 applies a transform, such as a discrete cosine transform (DCT), a discrete sine transform (DST), or a conceptually similar transform, to the residual block or sub-divisions thereof (e.g., four 8 x 8 transforms may be applied to a 16 x 16 array of residual values) to produce a set of residual transform coefficients. Transform coefficient generator 504 may be configured to perform any and all combinations of the transforms included in the family of discrete trigonometric transforms, including approximations thereof. Transform coefficient generator 504 may output transform coefficients to coefficient quantization unit 506. Coefficient quantization unit 506 may be configured to perform quantization of the transform coefficients. The quantization process may reduce the bit depth associated with some or all of the coefficients. The degree of quantization may alter the rate-distortion (i.e., bit-rate vs. quality of video) of encoded video data. The degree of quantization may be modified by adjusting a quantization parameter (QP). A quantization parameter may be determined based on slice level values and/or CU level values (e.g., CU delta QP values). QP data may include any data used to determine a QP for quantizing a particular set of transform coef-

ficients. As illustrated in FIG. 5, quantized transform coefficients (which may be referred to as level values) are output to inverse quantization and transform coefficient processing unit 508. Inverse quantization and transform coefficient processing unit 508 may be configured to apply an inverse quantization and an inverse transformation to generate reconstructed residual data. As illustrated in FIG. 5, at summer 510, reconstructed residual data may be added to a predictive video block. In this manner, an encoded video block may be reconstructed and the resulting reconstructed video block may be used to evaluate the encoding quality for a given prediction, transformation, and/or quantization. Video encoder 500 may be configured to perform multiple coding passes (e.g., perform encoding while varying one or more of a prediction, transformation parameters, and quantization parameters). The rate-distortion of a bitstream or other system parameters may be optimized based on evaluation of reconstructed video blocks. Further, reconstructed video blocks may be stored and used as reference for predicting subsequent blocks.

- [0032] Referring again to FIG. 5, intra prediction processing unit 512 may be configured to select an intra prediction mode for a video block to be coded. Intra prediction processing unit 512 may be configured to evaluate a frame and determine an intra prediction mode to use to encode a current block. As described above, possible intra prediction modes may include planar prediction modes, DC prediction modes, and angular prediction modes. Further, it should be noted that in some examples, a prediction mode for a chroma component may be inferred from a prediction mode for a luma prediction mode. Intra prediction processing unit 512 may select an intra prediction mode after performing one or more coding passes. Further, in one example, intra prediction processing unit 512 may select a prediction mode based on a rate-distortion analysis. As illustrated in FIG. 5, intra prediction processing unit 512 outputs intra prediction data (e.g., syntax elements) to entropy encoding unit 516 and transform coefficient generator 504. As described above, a transform performed on residual data may be mode dependent (e.g., a secondary transform matrix may be determined based on a predication mode).
- [0033] Referring again to FIG. 5, inter prediction processing unit 514 may be configured to perform inter prediction coding for a current video block. Inter prediction processing unit 514 may be configured to receive source video blocks and calculate a motion vector for PUs of a video block. A motion vector may indicate the displacement of a PU of a video block within a current video frame relative to a predictive block within a reference frame. Inter prediction coding may use one or more reference pictures. Further, motion prediction may be uni-predictive (use one motion vector) or bi-predictive (use two motion vectors). Inter prediction processing unit 514 may be configured to select a predictive block by calculating a pixel difference determined by,

for example, sum of absolute difference (SAD), sum of square difference (SSD), or other difference metrics. As described above, a motion vector may be determined and specified according to motion vector prediction. Inter prediction processing unit 514 may be configured to perform motion vector prediction, as described above. Inter prediction processing unit 514 may be configured to generate a predictive block using the motion prediction data. For example, inter prediction processing unit 514 may locate a predictive video block within a frame buffer (not shown in FIG. 5). It should be noted that inter prediction processing unit 514 may further be configured to apply one or more interpolation filters to a reconstructed residual block to calculate sub-integer pixel values for use in motion estimation. Inter prediction processing unit 514 may output motion prediction data for a calculated motion vector to entropy encoding unit 516.

[0034] Referring again to FIG. 5, entropy encoding unit 518 receives quantized transform coefficients and predictive syntax data (i.e., intra prediction data and motion prediction data). It should be noted that in some examples, coefficient quantization unit 506 may perform a scan of a matrix including quantized transform coefficients before the coefficients are output to entropy encoding unit 518. In other examples, entropy encoding unit 518 may perform a scan. Entropy encoding unit 518 may be configured to perform entropy encoding according to one or more of the techniques described herein. In this manner, video encoder 500 represents an example of a device configured to generate encoded video data according to one or more techniques of this disclosure. In one example, video encoder 500 may generate encoded video data including motion-constrained tile sets.

[0035] Referring again to FIG. 1, data encapsulator 107 may receive encoded video data and generate a compliant bitstream, e.g., a sequence of NAL units according to a defined data structure. A device receiving a compliant bitstream can reproduce video data therefrom. Further, as described above, sub-bitstream extraction may refer to a process where a device receiving a ITU-T H.265 compliant bitstream forms a new ITU-T H.265 compliant bitstream by discarding and/or modifying data in the received bitstream. It should be noted that the term conforming bitstream may be used in place of the term compliant bitstream. As described above, the signaling of a most-interested region as provided in Choi may be less than ideal. In one example, data encapsulator 107 may be configured to signal a most-interested region according to one or more techniques described herein. It should be noted that data encapsulator 107 need not necessarily be located in the same physical device as video encoder 106. For example, functions described as being performed by video encoder 106 and data encapsulator 107 may be distributed among devices illustrated in FIG. 4.

[0036] In one example, the information indicating which tile regions correspond to a

particular most-interested region may be signaled. Optionally, providing this information directly for each most-interested region enables a client that does not need to parse this information in ISOBMFF (e.g., enables a so-called simple client) and also avoids ambiguity in selection when multiple tile region groups correspond to the same most-interested region.

[0037] In one example, data encapsulator 107 may be configured to signal a most-interested region based on the following syntax:

```

unsigned int(32) regionbase_id;
unsigned int(16) entry_count;
for (i=1; i<= entry_count; i++) {
    unsigned int(16) left_horizontal_offset;
    unsigned int(16) top_vertical_offset;
    unsigned int(16) region_width;
    unsigned int(16) region_height;
    unsigned int(16) tile_info_entry_count;
    for(j = 0; j< tile_info_entry_count; j++) {
        unsigned int(16) tile_region_id;
    }
}

```

[0038] In the syntax above, semantics for **regionbase_id**, **entry_count**, **left_horizontal_offset**, **top_vertical_offset**, **region_width**, and **region_height** may be based on the definitions provided above. Semantics for **tile_info_entry_count** and **tile_region_id** may be based on the following definitions:

tile_info_entry_count specifies the number of tile groups corresponding to the most-interested region specified by **left_horizontal_offset**, **top_vertical_offset**, **region_width**, and **region_height**.

[0039] **tile_region_id** is the identifier of the tile region within the base region defined by **regionbase_id** which covers all or part of the region specified by **left_horizontal_offset**, **top_vertical_offset**, **region_width**, and **region_height**.

[0040] Thus, referring to the example illustrated in FIG. 2C, data encapsulator 107 may be configured to signal that the most-interested region illustrated in FIG. 2C includes one or more tile groups, where a tile group may include one or more tiles. For example, Tile₁ to Tile₆ may be signaled as one or more tile groups. Further, for each tile group in the most interested region, data encapsulator 107 may be configured to signal an identifier.

[0041] In ISOBMFF - ISO/ IEC 14496-15: 2014(E) (N16169, "Information technology - Coding of audio-visual objects - Part 15: Carriage of network abstraction layer (NAL)

unit structured video in the ISO base media file format,” August 2016, Geneva, which is incorporated by reference herein) a tile region is defined as one or more complete HEVC tiles in one or more complete slice segments that are within the same coded picture and that contain no other HEVC tiles and tile regions can be described through tile region visual sample group description entries (i.e., instances of TileRegionGroupEntry). Each TileRegionGroupEntry is assigned a unique identifier, called groupID. This identifier can be used to associate NAL units in a sample to a particular TileRegionGroupEntry.

[0042] In one example, in the context of ISOBMFF, the value of tile_region_id can be equal to the groupID value as defined in TileRegionGroupEntry of ISOBMFF - ISO/ IEC 14496-15: 2014(E).

[0043] In one example, tile_region_id may be instead specifically be signaled as a reference to ISOBMFF groupID and called tile_region_groupid and may be based on the following definition:

tile_region_groupid is the identifier of the tile region group within the base region defined by regionbase_id which covers all or part of the region specified by left_horizontal_offset, top_vertical_offset, region_width, and region_height .
tile_region_groupid value shall be equal to groupID value as defined in TileRegionGroupEntry of ISOBMFF - ISO/ IEC 14496-15: 2014(E).

[0044] In one example, instead of 16 bits, a tile_info_entry_count could instead include 8 bits or another number of bits.

[0045] In one example, instead of signaling tile_info_entry_count, the number of tile information entries could be signaled as unsigned int(16) tile_info_entry_count_minus1 which may be based on the following definition:

tile_info_entry_count_minus1 plus 1 specifies the number of tile groups corresponding to the most-interested region specified by left_horizontal_offset, top_vertical_offset, region_width, and region_height.

[0046] As described above, omnidirectional video may be coded using MCTS. An MCTS extraction information set SEI message provides supplemental information to carry out sub-bitstream MCTS extraction as specified in Skupin to derive a HEVC conformant bitstream from a motion-constrained tile set. In one example, information about which MCTS extraction set corresponds to a particular most-interested region may be signalled. Optionally, providing this information directly for each most-interested region allows a simple client to request MCTS sub-bitstream extraction for MCTS extraction set identifier corresponding to most-interested region.

[0047] In one example, data encapsulator 107 may be configured to signal a number of MCTS corresponding to a most-interested region and for each MCTS signal an identifier. Signaling of a number of MCTSs corresponding to a most-interested region

may be useful for performing sub-bitstream extraction. In one example, data encapsulator 107 may be configured to signal a most-interested region based on the following syntax:

```

    unsigned int(32) regionbase_id;
    unsigned int(16) entry_count;
    for (i=1; i<= entry_count; i++) {
        unsigned int(16) left_horizontal_offset;
        unsigned int(16) top_vertical_offset;
        unsigned int(16) region_width;
        unsigned int(16) region_height;
        unsigned int(32) mcts_extraction_set_info_entry_count;
        for(j = 0; j< mcts_extraction_set_info_entry_count; j++) {
            unsigned int(32) mcts_extraction_set_idx;
        }
    }

```

[0048] In the syntax above, semantics for **regionbase_id**, **entry_count**, **left_horizontal_offset**, **top_vertical_offset**, **region_width**, and **region_height** may be based on the definitions provided above. Semantics for **mcts_extraction_set_info_entry_count** and **mcts_extraction_set_idx** may be based on the following definitions:

mcts_extraction_set_info_entry_count specifies the number of extraction information sets corresponding to the most-interested region specified by **left_horizontal_offset**, **top_vertical_offset**, **region_width**, and **region_height**.

[0049] **mcts_extraction_set_idx** is the identifier of the motion-constrained tile sets (MCTS) extraction information set within the base region defined by **regionbase_id** which covers all or part of the region specified by **left_horizontal_offset**, **top_vertical_offset**, **region_width**, and **region_height**. **mcts_extraction_set_idx** shall be equal to one of the values of index *i* in MCTS extraction information set SEI message for the current picture as defined in Motion-Constrained Tile Sets Extraction Information SEI Messages Draft 1 [JCTVC-Y1008, “Skupin”].

[0050] In one example, following constraint may be required:

mcts_extraction_set_idx shall be equal to one of the values of MCTS extraction set identifiers in MCTS extraction information set SEI message for the current picture as defined in Motion-Constrained Tile Sets Extraction Information SEI Messages Draft 1 [JCTVC-Y1008, “Skupin”].

[0051] In one example, only one **mcts_extraction_set_idx** may be present in the signaling. In this example, data encapsulator 107 may be configured to signal a most-interested region based on the following syntax:

```

unsigned int(32) regionbase_id;
unsigned int(16) entry_count;
for (i=1; i<= entry_count; i++) {
    unsigned int(16) left_horizontal_offset;
    unsigned int(16) top_vertical_offset;
    unsigned int(16) region_width;
    unsigned int(16) region_height;
    unsigned int(32) mcts_extraction_set_idx;
}

```

[0052] In one example, a flag (e.g., syntax element `mcts_extraction_set_idx_present`) may be signaled to indicate if `mcts_extraction_set_idx` element is present. Semantics for `mcts_extraction_set_idx_present` may be based on the following definition:

`mcts_extraction_idx_present` equal to 1 indicates that the element `mcts_extraction_set_idx` is present. `mcts_extraction_idx_present` equal to 0 indicates that the element `mcts_extraction_set_idx` is not present.

[0053] In one example, data encapsulator 107 may be configured to signal that a most-interested region includes one or more tile groups and a number of MCTS corresponding to a most-interested region. In one example, data encapsulator 107 may be configured to signal a most-interested region based on the following syntax:

```

unsigned int(32) regionbase_id;
unsigned int(16) entry_count; // in one example, minus1 coding may be used.
for (i=1; i<= entry_count; i++) {
    unsigned int(16) left_horizontal_offset;
    unsigned int(16) top_vertical_offset;
    unsigned int(16) region_width;
    unsigned int(16) region_height;
    unsigned int(16) tile_info_entry_count;
    for(j = 0; j< tile_info_entry_count; j++) {
        unsigned int(16) tile_region_id;
    }
    unsigned int(32) mcts_extraction_set_info_entry_count;
    for(j = 0; j< mcts_extraction_info_entry_count; j++) {
        unsigned int(32) mcts_extraction_set_idx;
    }
}

```

[0054] In the syntax above, semantics for `regionbase_id`, `entry_count`, `left_horizontal_offset`, `top_vertical_offset`, `region_width`, `region_height`, `tile_info_entry_count`, `tile_region_id`, `mcts_extraction_set_info_entry_count`, and `mcts_extraction_set_idx` may be based on

the definitions provided above. Further, it should be noted that examples described with respect to signaling that a most-interested region includes one or more tile groups may be combined with examples described with respect to signaling a number of MCTS corresponding to a most-interested region in the example where data encapsulator 107 signals that a most-interested region includes one or more tile groups and a number of MCTS corresponding to a most-interested region.

[0055] Referring again to FIG. 1, interface 108 may include any device configured to receive data generated by data encapsulator 107 and transmit and/or store the data to a communications medium. Interface 108 may include a network interface card, such as an Ethernet card, and may include an optical transceiver, a radio frequency transceiver, or any other type of device that can send and/or receive information. Further, interface 108 may include a computer system interface that may enable a file to be stored on a storage device. For example, interface 108 may include a chipset supporting Peripheral Component Interconnect (PCI) and Peripheral Component Interconnect Express (PCIe) bus protocols, proprietary bus protocols, Universal Serial Bus (USB) protocols, I²C, or any other logical and physical structure that may be used to interconnect peer devices.

[0056] Referring again to FIG. 1, destination device 120 includes interface 122, data decapsulator 123, video decoder 124, and display 126. Interface 122 may include any device configured to receive data from a communications medium. Interface 122 may include a network interface card, such as an Ethernet card, and may include an optical transceiver, a radio frequency transceiver, or any other type of device that can receive and/or send information. Further, interface 122 may include a computer system interface enabling a compliant video bitstream to be retrieved from a storage device. For example, interface 122 may include a chipset supporting PCI and PCIe bus protocols, proprietary bus protocols, USB protocols, I²C, or any other logical and physical structure that may be used to interconnect peer devices. Data decapsulator 123 may be configured to receive a bitstream generated by data encapsulator 107 and perform sub-bitstream extraction according to one or more of the techniques described herein.

[0057] Video decoder 124 may include any device configured to receive a bitstream (e.g., a MCTS sub-bitstream extraction) and/or acceptable variations thereof and reproduce video data therefrom. Display 126 may include any device configured to display video data. Display 126 may comprise one of a variety of display devices such as a liquid crystal display (LCD), a plasma display, an organic light emitting diode (OLED) display, or another type of display. Display 126 may include a High Definition display or an Ultra High Definition display. It should be noted that although in the example illustrated in FIG. 1, video decoder 124 is described as outputting data to display 126, video decoder 124 may be configured to output video data to various types of devices

and/or sub-components thereof. For example, video decoder 124 may be configured to output video data to any communication medium, as described herein.

[0058] FIG. 6 is a block diagram illustrating an example of a video decoder that may be configured to decode video data according to one or more techniques of this disclosure. In one example, video decoder 600 may be configured to decode transform data and reconstruct residual data from transform coefficients based on decoded transform data. Video decoder 600 may be configured to perform intra prediction decoding and inter prediction decoding and, as such, may be referred to as a hybrid decoder. In the example illustrated in FIG. 6, video decoder 600 includes an entropy decoding unit 602, inverse quantization unit and transform coefficient processing unit 604, intra prediction processing unit 606, inter prediction processing unit 608, summer 610, post filter unit 612, and reference buffer 614. Video decoder 600 may be configured to decode video data in a manner consistent with a video coding system. It should be noted that although example video decoder 600 is illustrated as having distinct functional blocks, such an illustration is for descriptive purposes and does not limit video decoder 600 and/or sub-components thereof to a particular hardware or software architecture. Functions of video decoder 600 may be realized using any combination of hardware, firmware, and/or software implementations.

[0059] As illustrated in FIG. 6, entropy decoding unit 602 receives an entropy encoded bitstream. Entropy decoding unit 602 may be configured to decode syntax elements and quantized coefficients from the bitstream according to a process reciprocal to an entropy encoding process. Entropy decoding unit 602 may be configured to perform entropy decoding according any of the entropy coding techniques described above. Entropy decoding unit 602 may determine values for syntax elements in an encoded bitstream in a manner consistent with a video coding standard. As illustrated in FIG. 6, entropy decoding unit 602 may determine a quantization parameter, quantized coefficient values, transform data, and predication data from a bitstream. In the example, illustrated in FIG. 6, inverse quantization unit and transform coefficient processing unit 604 receives a quantization parameter, quantized coefficient values, transform data, and predication data from entropy decoding unit 602 and outputs reconstructed residual data.

[0060] Referring again to FIG. 6, reconstructed residual data may be provided to summer 610. Summer 610 may add reconstructed residual data to a predictive video block and generate reconstructed video data. A predictive video block may be determined according to a predictive video technique (i.e., intra prediction and inter frame prediction). Intra prediction processing unit 606 may be configured to receive intra prediction syntax elements and retrieve a predictive video block from reference buffer 614. Reference buffer 614 may include a memory device configured to store one or

more frames of video data. Intra prediction syntax elements may identify an intra prediction mode, such as the intra prediction modes described above. Inter prediction processing unit 608 may receive inter prediction syntax elements and generate motion vectors to identify a prediction block in one or more reference frames stored in reference buffer 814. Inter prediction processing unit 608 may produce motion compensated blocks, possibly performing interpolation based on interpolation filters. Identifiers for interpolation filters to be used for motion estimation with sub-pixel precision may be included in the syntax elements. Inter prediction processing unit 808 may use interpolation filters to calculate interpolated values for sub-integer pixels of a reference block. Post filter unit 612 may be configured to perform filtering on reconstructed video data. For example, post filter unit 612 may be configured to perform deblocking and/or Sample Adaptive Offset (SAO) filtering, e.g., based on parameters specified in a bitstream. Further, it should be noted that in some examples, post filter unit 612 may be configured to perform proprietary discretionary filtering (e.g., visual enhancements, such as, mosquito noise reduction). As illustrated in FIG. 6, a reconstructed video block may be output by video decoder 600. In this manner, video decoder 600 may be configured to generate reconstructed video data according to one or more of the techniques described herein.

[0061] In one or more examples, the functions described may be implemented in hardware, software, firmware, or any combination thereof. If implemented in software, the functions may be stored on or transmitted over as one or more instructions or code on a computer-readable medium and executed by a hardware-based processing unit. Computer-readable media may include computer-readable storage media, which corresponds to a tangible medium such as data storage media, or communication media including any medium that facilitates transfer of a computer program from one place to another, e.g., according to a communication protocol. In this manner, computer-readable media generally may correspond to (1) tangible computer-readable storage media which is non-transitory or (2) a communication medium such as a signal or carrier wave. Data storage media may be any available media that can be accessed by one or more computers or one or more processors to retrieve instructions, code and/or data structures for implementation of the techniques described in this disclosure. A computer program product may include a computer-readable medium.

[0062] By way of example, and not limitation, such computer-readable storage media can comprise RAM, ROM, EEPROM, CD-ROM or other optical disk storage, magnetic disk storage, or other magnetic storage devices, flash memory, or any other medium that can be used to store desired program code in the form of instructions or data structures and that can be accessed by a computer. Also, any connection is properly termed a computer-readable medium. For example, if instructions are transmitted from

a website, server, or other remote source using a coaxial cable, fiber optic cable, twisted pair, digital subscriber line (DSL), or wireless technologies such as infrared, radio, and microwave, then the coaxial cable, fiber optic cable, twisted pair, DSL, or wireless technologies such as infrared, radio, and microwave are included in the definition of medium. It should be understood, however, that computer-readable storage media and data storage media do not include connections, carrier waves, signals, or other transitory media, but are instead directed to non-transitory, tangible storage media. Disk and disc, as used herein, includes compact disc (CD), laser disc, optical disc, digital versatile disc (DVD), floppy disk and Blu-ray disc where disks usually reproduce data magnetically, while discs reproduce data optically with lasers. Combinations of the above should also be included within the scope of computer-readable media.

- [0063] Instructions may be executed by one or more processors, such as one or more digital signal processors (DSPs), general purpose microprocessors, application specific integrated circuits (ASICs), field programmable logic arrays (FPGAs), or other equivalent integrated or discrete logic circuitry. Accordingly, the term “processor,” as used herein may refer to any of the foregoing structure or any other structure suitable for implementation of the techniques described herein. In addition, in some aspects, the functionality described herein may be provided within dedicated hardware and/or software modules configured for encoding and decoding, or incorporated in a combined codec. Also, the techniques could be fully implemented in one or more circuits or logic elements.
- [0064] The techniques of this disclosure may be implemented in a wide variety of devices or apparatuses, including a wireless handset, an integrated circuit (IC) or a set of ICs (e.g., a chip set). Various components, modules, or units are described in this disclosure to emphasize functional aspects of devices configured to perform the disclosed techniques, but do not necessarily require realization by different hardware units. Rather, as described above, various units may be combined in a codec hardware unit or provided by a collection of interoperative hardware units, including one or more processors as described above, in conjunction with suitable software and/or firmware.
- [0065] Moreover, each functional block or various features of the base station device and the terminal device used in each of the aforementioned embodiments may be implemented or executed by a circuitry, which is typically an integrated circuit or a plurality of integrated circuits. The circuitry designed to execute the functions described in the present specification may comprise a general-purpose processor, a digital signal processor (DSP), an application specific or general application integrated circuit (ASIC), a field programmable gate array (FPGA), or other programmable logic devices, discrete gates or transistor logic, or a discrete hardware component, or a com-

bination thereof. The general-purpose processor may be a microprocessor, or alternatively, the processor may be a conventional processor, a controller, a microcontroller or a state machine. The general-purpose processor or each circuit described above may be configured by a digital circuit or may be configured by an analogue circuit. Further, when a technology of making into an integrated circuit superseding integrated circuits at the present time appears due to advancement of a semiconductor technology, the integrated circuit by this technology is also able to be used.

[0066] Various examples have been described. These and other examples are within the scope of the following claims.

[0067] <overview>

In one example, a method of signaling of a most-interested region of a omnidirectional video comprises signaling a most-interested region and signaling a number of tile groups corresponding to the most interested region.

[0068] In one example, a device comprises one or more processors configured to signal a most-interested region and signal a number of tile groups corresponding to the most interested region.

[0069] In one example, a non-transitory computer-readable storage medium comprises instructions stored thereon that, when executed, cause one or more processors of a device to signal a most-interested region and signal a number of tile groups corresponding to the most interested region

In one example, an apparatus comprises means for signaling a most-interested region and means for signaling a number of tile groups corresponding to the most interested region

In one example, a method of performing sub-bitstream extraction of a motion-constrained tile set comprises receiving a message identifying a most-interested region of a omnidirectional video, wherein the message includes an identifier for a motion-constrained tile set corresponding to the most interested region, and performing sub-bitstream extraction based on the identified motion-constrained tile set.

[0070] In one example, a device comprises one or more processors configured to receive a message identifying a most-interested region of a omnidirectional video, wherein the message includes an identifier for a motion-constrained tile set corresponding to the most interested region, and perform sub-bitstream extraction based on the identified motion-constrained tile set.

[0071] In one example, a non-transitory computer-readable storage medium comprises instructions stored thereon that, when executed, cause one or more processors of a device to receive a message identifying a most-interested region of a omnidirectional video, wherein the message includes an identifier for a motion-constrained tile set corresponding to the most interested region, and perform sub-bitstream extraction based on

the identified motion-constrained tile set.

[0072] In one example, an apparatus comprises means for receiving a message identifying a most-interested region of a omnidirectional video, wherein the message includes an identifier for a motion-constrained tile set corresponding to the most interested region, and means for performing sub-bitstream extraction based on the identified motion-constrained tile set.

[0073] The details of one or more examples are set forth in the accompanying drawings and the description below. Other features, objects, and advantages will be apparent from the description and drawings, and from the claims.

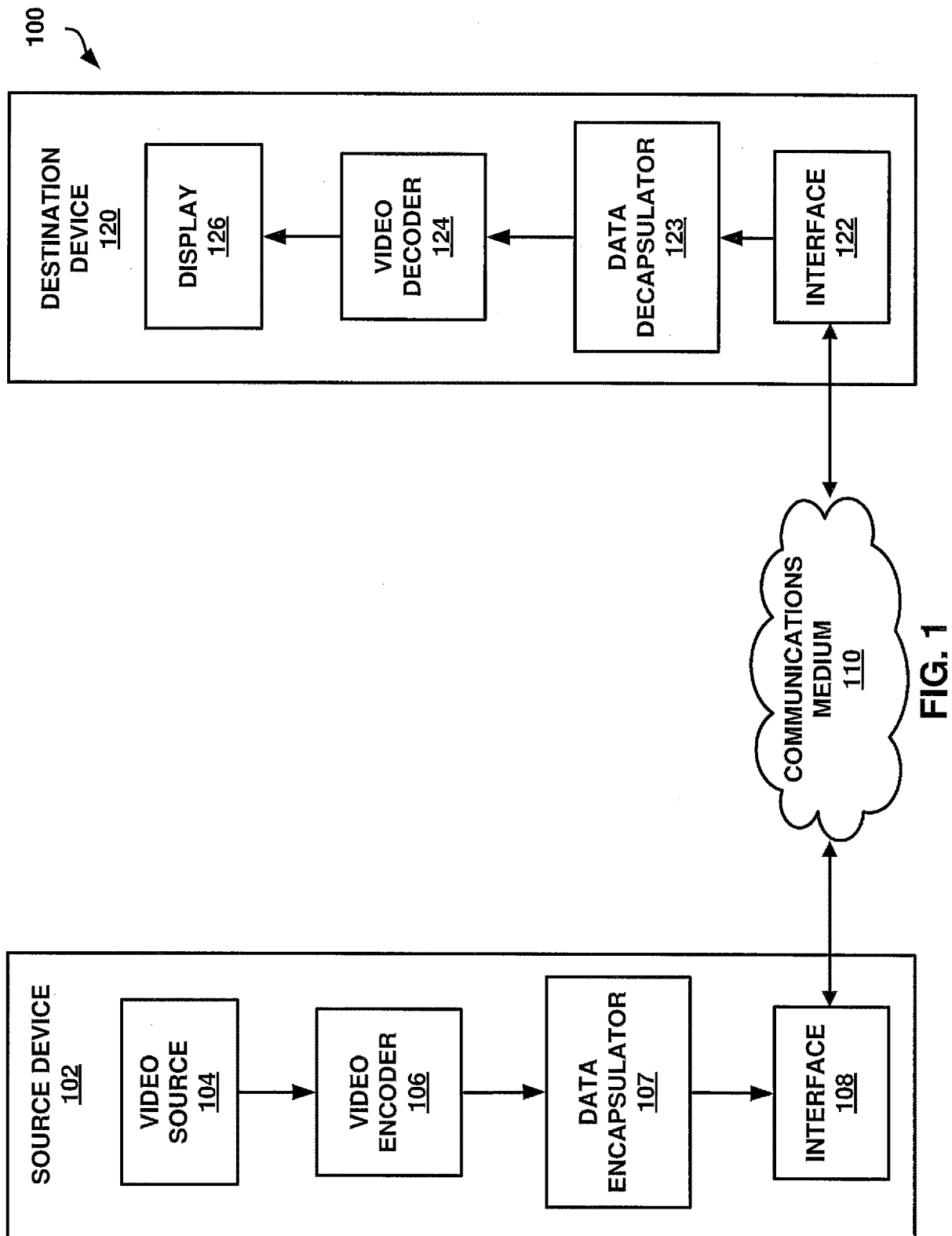
[0074] <Cross Reference>

This Nonprovisional application claims priority under 35 U.S.C. § 119 on provisional Application No. 62/443,767 on January 8, 2017, the entire contents of which are hereby incorporated by reference.

Claims

- [Claim 1] A method of signaling of a most-interested region of a omnidirectional video, the method comprising:
signaling a most-interested region; and
signaling a number of tile groups corresponding to the most interested region.
- [Claim 2] The method of claim 1, where signaling a number of tile groups corresponding to the most interested region includes signaling an identifier for each tile group corresponding to the most interested region.
- [Claim 3] The method of any of claims 1 or 2, wherein a tile group includes one or more motion-constrained tile sets.
- [Claim 4] A method of performing sub-bitstream extraction of a motion-constrained tile set, the method comprising:
receiving a message identifying a most-interested region of a omnidirectional video, wherein the message includes an identifier for a motion-constrained tile set corresponding to the most interested region;
and
performing sub-bitstream extraction based on the identified motion-constrained tile set.
- [Claim 5] A device comprising one or more processors configured to perform any and all combinations of the steps of claims 1-4.
- [Claim 6] The device of claim 5, wherein the device includes a video encoder.
- [Claim 7] The device of claim 5, wherein the device includes a video decoder.
- [Claim 8] A system comprising:
the device of claim 6; and
the device of claim 7.
- [Claim 9] An apparatus comprising means for performing any and all combinations of the steps of claims 1-4.
- [Claim 10] A non-transitory computer-readable storage medium comprising instructions stored thereon that, when executed, cause one or more processors of a device to perform any and all combinations of the steps of claims 1-4.

[Fig. 1]



[Fig. 2A]

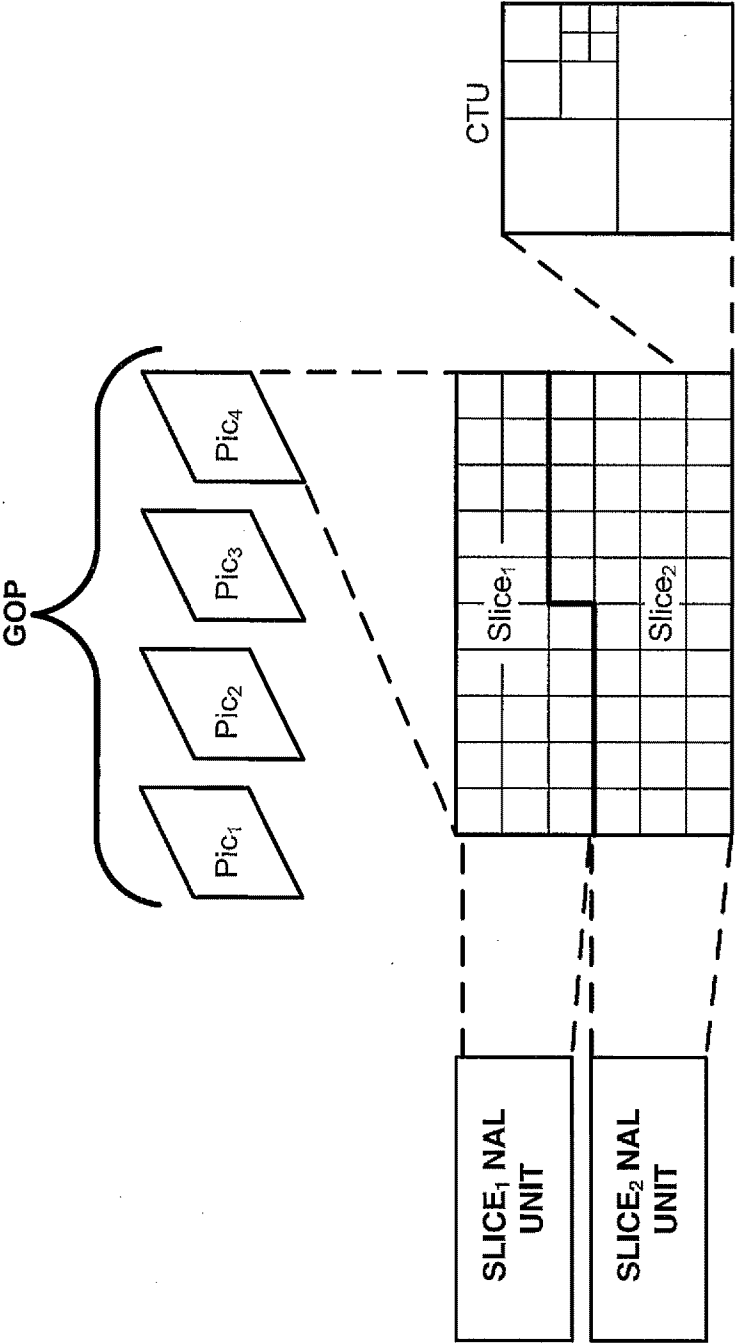
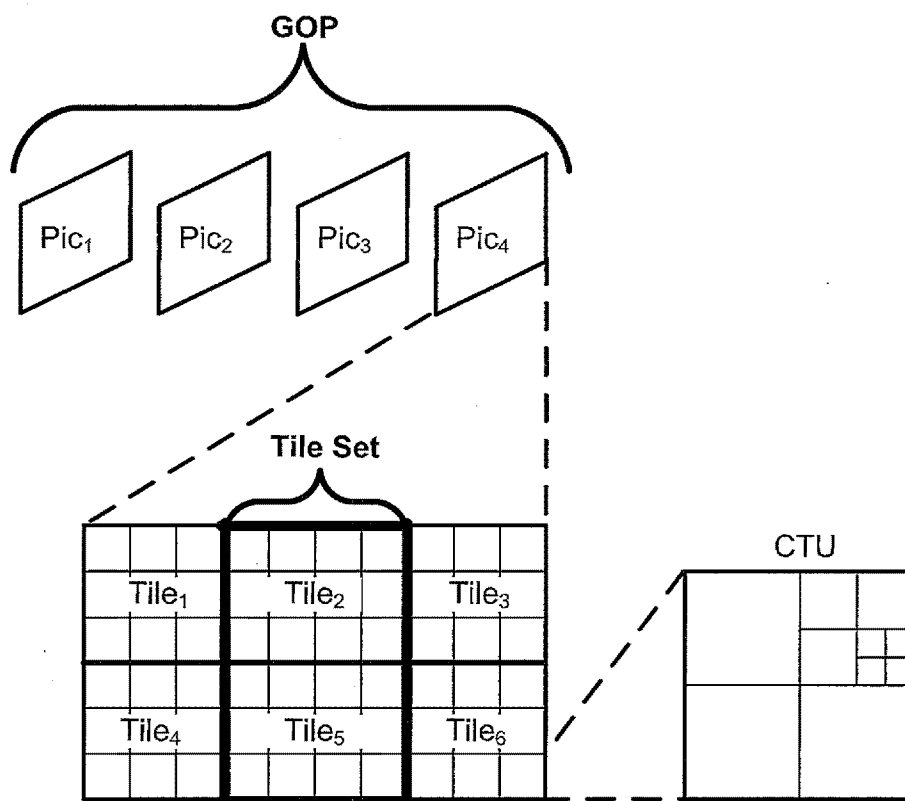


FIG. 2A

[Fig. 2B]

**FIG. 2B**

[Fig. 2C]

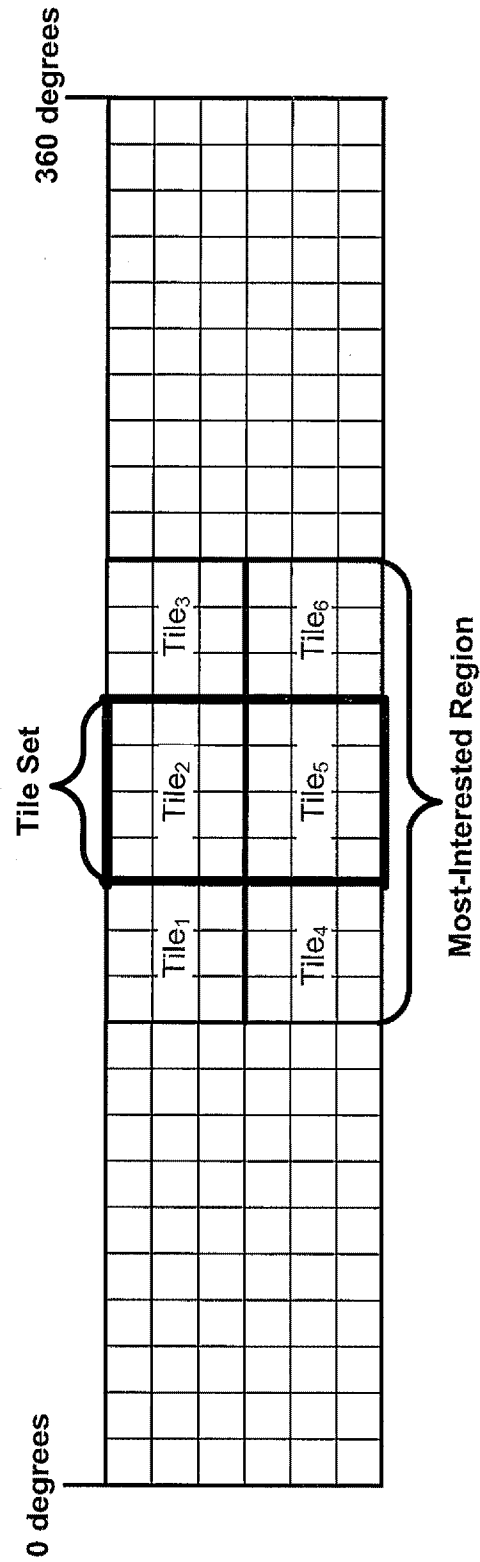


FIG. 2C

[Fig. 3]

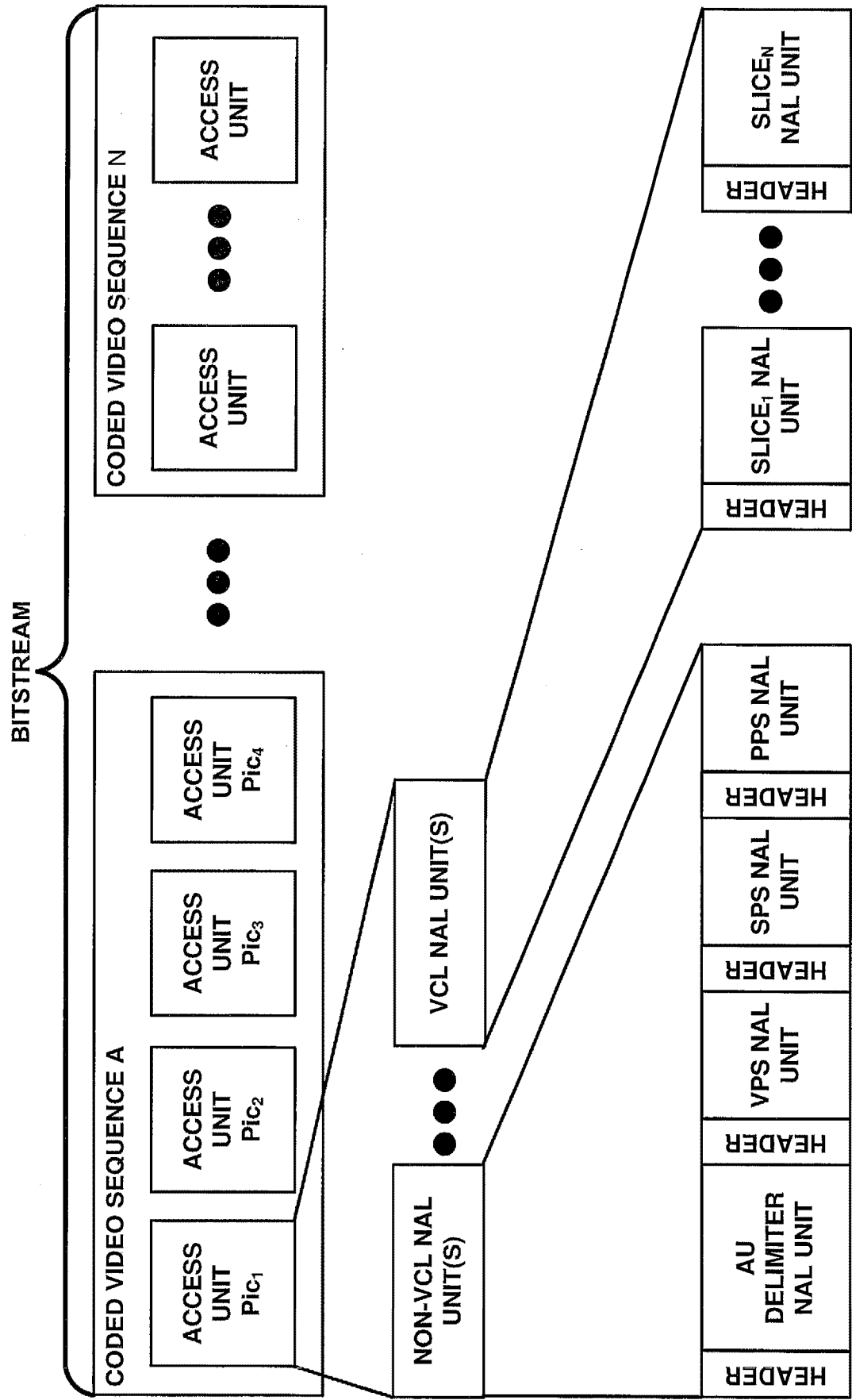


FIG. 3

[Fig. 4]

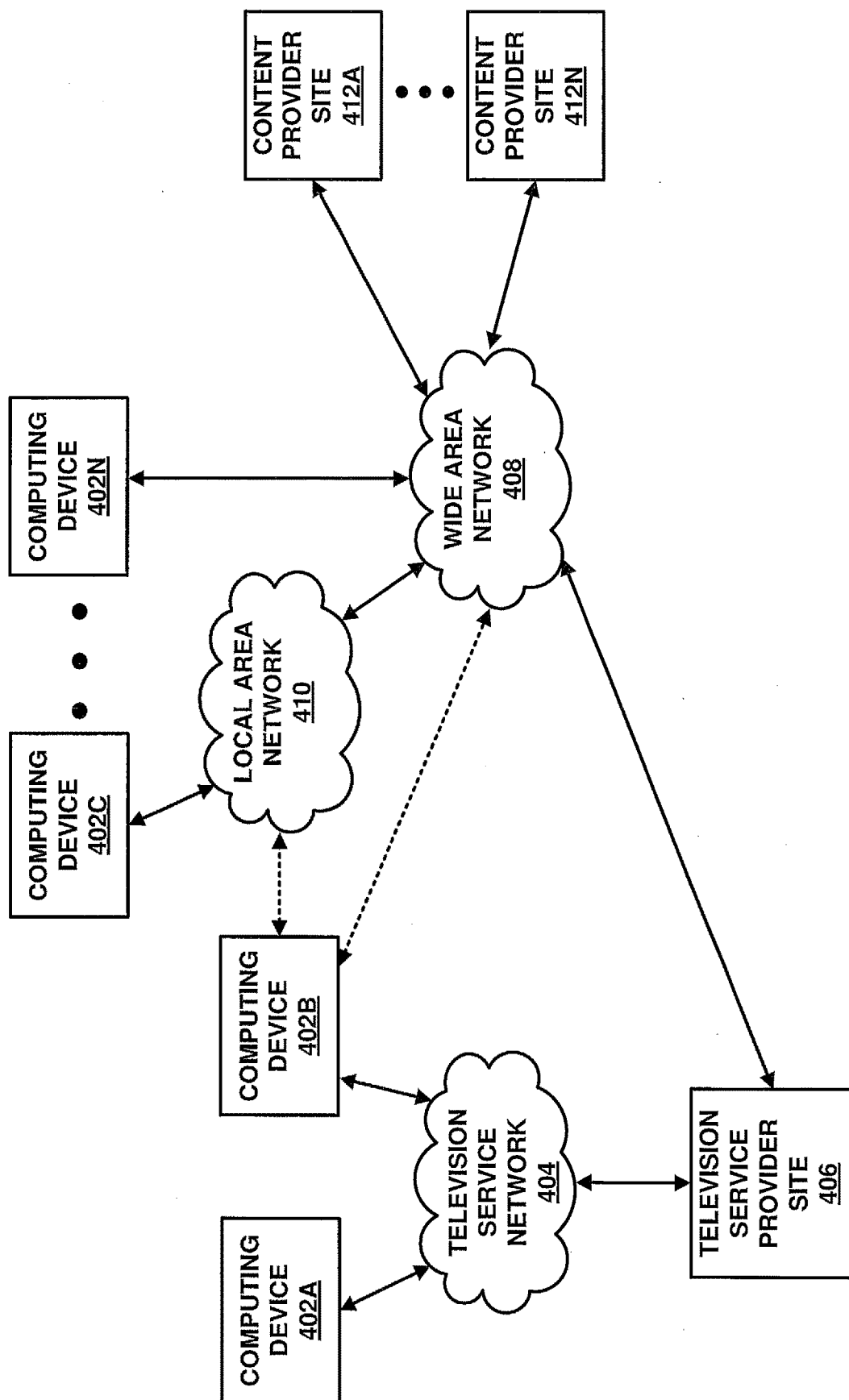
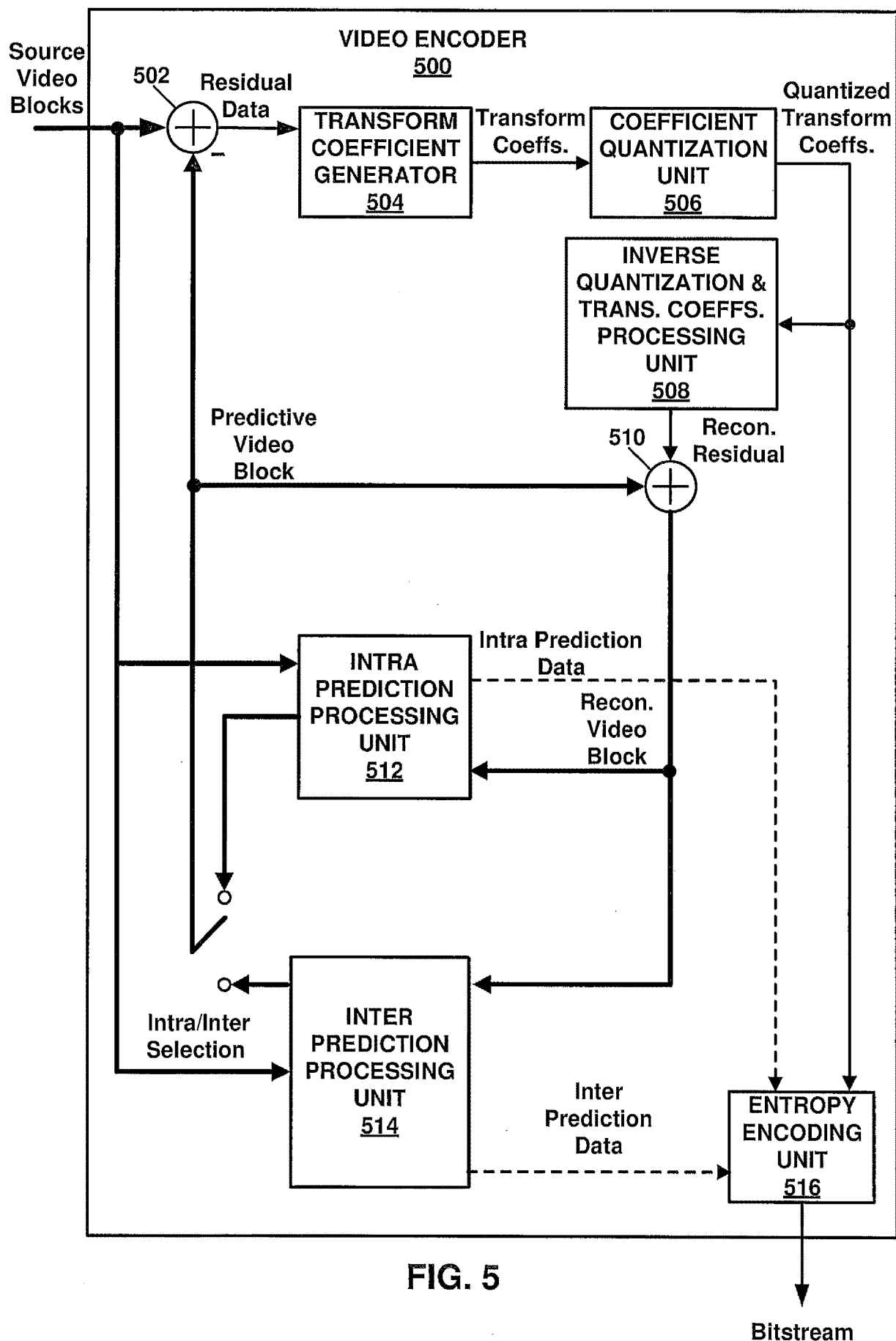


FIG. 4

FIG. 5



[Fig. 6]

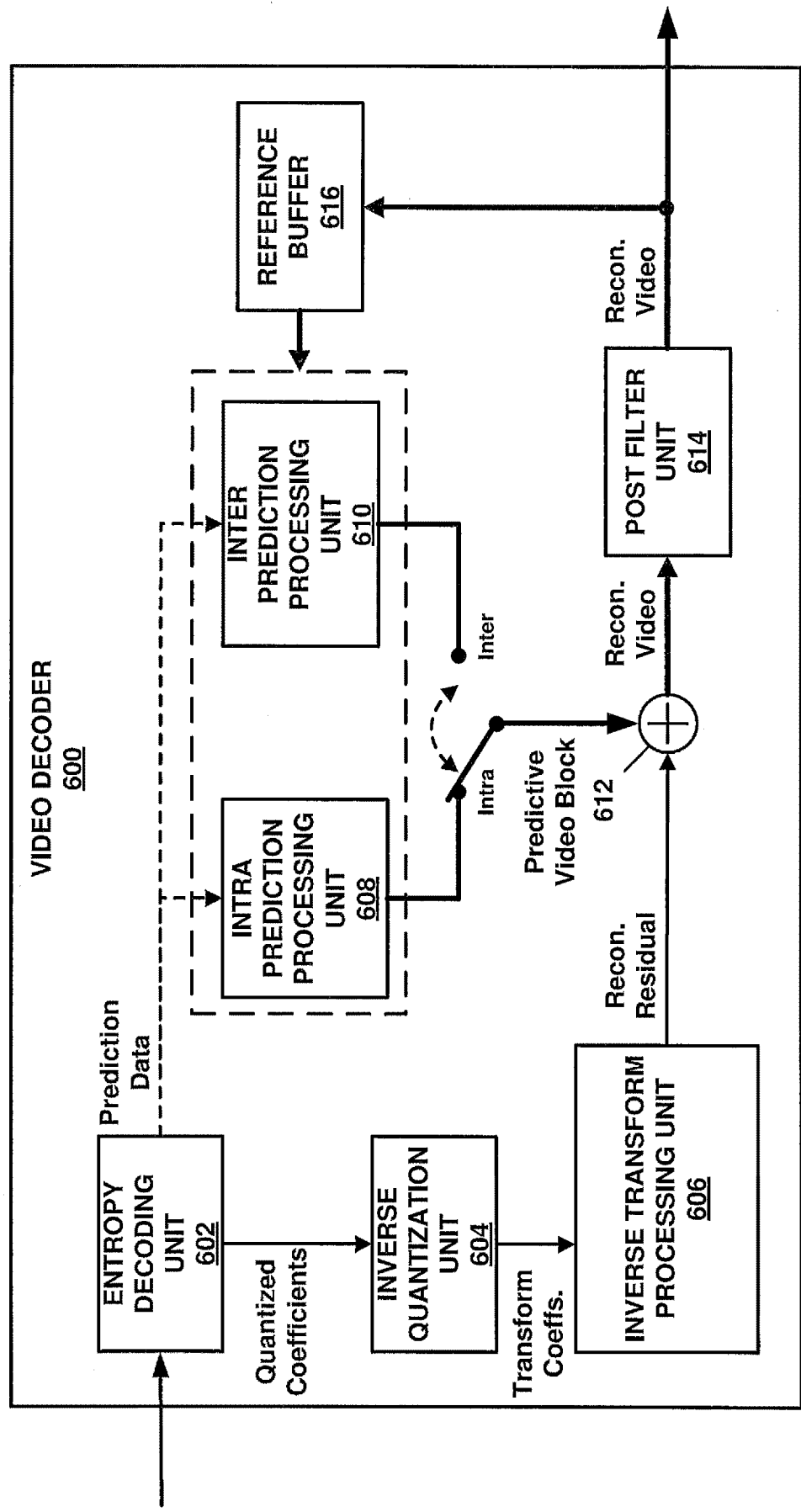


FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2017/045270

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. H04N19/70(2014.01)i, H04N19/167(2014.01)i, H04N21/4728(2011.01)i

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)

Int.Cl. H04N19/70, H04N19/167, H04N21/4728

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

Published examined utility model applications of Japan 1922-1996
 Published unexamined utility model applications of Japan 1971-2018
 Registered utility model specifications of Japan 1996-2018
 Published registered utility model applications of Japan 1994-2018

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2015/197815 A1 (KONINKLIJKE KPN N.V.) 2015.12.30, pages 1-3, 12-20, FIGs. 1-8 & KR 10-2017-0018352 A & CN 106664443 A	1-10
Y	Byeongdoo Choi et al., WD on ISO/IEC 23000-20 Omnidirectional Media Application Format, ISO/IEC JTC1/SC29/WG11 N16189 June 2016, Geneva, Switzerland, 2016.07.15	1-10
Y	WO 2014/006863 A1 (CANON KABUSHIKI KAISHA) 2014.01.09, pars 0163-0178, FIGs. 22-28 & JP 2014-30187 A	1-10
Y	JP 2016-19248 A (CANON KABUSHIKI KAISHA) 2016.02.01, pars 0050-0080, FIGs. 6-9 (No Family)	1-10



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

“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

22.02.2018

Date of mailing of the international search report

06.03.2018

Name and mailing address of the ISA/JP

Japan Patent Office

3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan

Authorized officer

AKO, Shuichiro

Telephone No. +81-3-3581-1101 Ext. 3541

5C

6297

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2017/045270

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, A	Sachin Deshpande, On Region-Wise Packing, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 1129th Meeting: Macao, CN, 19-24 October 2017, JCTVC-AC0037, 2017.10.24	1-10
A	Yan Ye et al., ROI tile sections, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11 11th Meeting: Shanghai, CN, 10-19 Oct. 2012, JCTVC-K0248r2, 2012.10.19	1-10
A	WO 2013/077236 A1 (CANON KABUSHIKI KAISHA) 2013.05.30, Whole Document & JP 2013-132048 A	1-10