



(51) International Patent Classification:

H04B 1/66 (2006.01) H04N 11/02 (2006.01)  
H04N 7/32 (2006.01) H04N 11/04 (2006.01)

(21) International Application Number:

PCT/US2019/056269

(22) International Filing Date:

15 October 2019 (15.10.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/748,675 22 October 2018 (22.10.2018) US  
16/424,352 28 May 2019 (28.05.2019) US

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

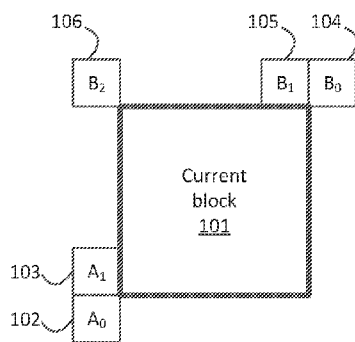
(72) Inventors: XU, Meng; c/o Tencent America LLC, 2747 Park Boulevard, Palo Alto, California 94306 (US). LI, Xi-ang; c/o Tencent America LLC, 2747 Park Boulevard, Palo Alto, California 94306 (US). LIU, Shan; c/o Tencent America LLC, 2747 Park Boulevard, Palo Alto, California 94306 (US).

(74) Agent: MA, Johnny et al.; Oblon, McClelland, Maier & Neustadt, L.L.P., 1940 Duke Street, Alexandria, Virginia 22314 (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, 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).

(54) Title: METHOD AND APPARATUS FOR VIDEO CODING



Related Art

FIG. 1

(57) Abstract: Aspects of the disclosure provide methods and apparatuses for video coding. In some examples, an apparatus includes processing circuitry. The processing circuitry determines first and second reference motion vectors for decoding a current block based on coding information of at least one previously decoded block, offsets a first component to obtain a shifted first component, and offsets a second component to obtain a shifted second component. The processing circuitry generates a third component, on the first coordinate axis, of a first averaged reference motion vector by adding the shifted first component and the shifted second component, and constructs a list of reference motion vectors incorporating the first and second reference motion vectors and the first averaged reference motion vector. The processing circuitry also determines a motion vector predictor using the list of reference motion vectors, and decodes the current block for output based on the determined motion vector predictor.



**Published:**

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

## METHOD AND APPARATUS FOR VIDEO CODING

### INCORPORATION BY REFERENCE

[0001] This present application claims the benefit of priority of U.S. Patent Application No. 16/424,352, "METHOD AND APPARATUS FOR VIDEO CODING" filed on May 28, 2019, which claims the benefit of priority to U.S. Provisional Application No. 62/748,675, "METHOD FOR MOTION VECTOR ROUNDING" filed on October 22, 2018. The entire 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] Video coding and decoding can be performed using inter-picture prediction with motion compensation. 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 significant 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 or storage space requirements, in some cases by two orders of magnitude or more. Both lossless 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]** 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).

**[0007]** 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.

**[0008]** 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".

[0009] Referring to FIG. 1, a current block (101) 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 (102 through 106, respectively). In H.265, the MV prediction can use predictors from the same reference picture that the neighboring block is using.

#### SUMMARY

[0010] Aspects of the disclosure provide methods and apparatuses for video encoding/decoding. In some examples, an apparatus includes processing circuitry that determines first and second reference motion vectors for decoding a current block based on coding information of at least one previously decoded block, offsets a first component, on a first coordinate axis and in a two's complement integer representation format, of the first reference motion vector by one, and right-shifts the offset first component by one bit to obtain a shifted first component, and offsets a second component, on the first coordinate axis and in the two's complement integer representation format, of the second reference motion vector by one, and right-shifts the offset second component by one bit to obtain a shifted second component. The processing circuitry generates a third component, on the first coordinate axis and in the two's complement integer representation format, of a first averaged reference motion vector by adding the shifted first component and the shifted second component, and constructs a list of reference motion vectors for decoding the current block, the list of reference motion vectors incorporating the first and second reference motion vectors and the first averaged reference motion vector. The processing circuitry also determines a motion vector predictor using the list of reference motion vectors, and decodes the current block for output based on the determined motion vector predictor.

[0011] In some embodiments, the processing circuitry further offsets a fourth component, on a second coordinate axis and in the two's complement integer representation format, of the first reference motion vector by one, and right-shifts the offset fourth component by one bit to obtain a shifted fourth component, offsets a fifth component, on the second coordinate axis and in the two's complement integer representation format, of the second reference motion vector by one, and right-shifts the offset fifth component by one bit to obtain a shifted fifth component, and

generates a sixth component, on the second coordinate axis and in the two's complement integer representation format, of the first averaged reference motion vector by adding the shifted fourth component and the shifted fifth component.

**[0012]** In some embodiments, the processing circuitry generates a second averaged reference motion vector based on the first reference motion vector and a third reference motion vector, generates a third averaged reference motion vector based on the second reference motion vector and the third reference motion vector, and constructs the list of reference motion vectors for decoding the current block by arranging the list of reference motion vectors to include, according to an ascending order of respective indices associated thereto, the first reference motion vector, the second reference motion vector, the third reference motion vector, the first averaged reference motion vector, the second averaged reference motion vector, and the third averaged reference motion vector.

**[0013]** In some examples, the processing circuitry generates a fourth averaged reference motion vector based on the first reference motion vector and a fourth reference motion vector, generates a fifth averaged reference motion vector based on the second reference motion vector and the fourth reference motion vector, and generates a sixth averaged reference motion vector based on the third reference motion vector and the fourth reference motion vector. In some examples, the processing circuitry constructs the list of reference motion vectors for decoding the current block by arranging the list of reference motion vectors to include, according to the ascending order of respective indices associated thereto, the first reference motion vector, the second reference motion vector, the third reference motion vector, the fourth reference motion vector, the first list reference motion vectors and adding the first averaged reference motion vector, the second averaged reference motion vector, the third averaged reference motion vector, the fourth averaged reference motion vector, the fifth averaged reference motion vector, and the sixth averaged reference motion vector.

**[0014]** In some examples, the processing circuitry sets a right-shift parameter as 1 and sets a left-shift parameter as 0. The processing circuitry further executes a rounding subroutine or operates a rounding circuit that is configured to receive the first component, the right-shift parameter, and the left-shift parameter, and calculates the shifted first component according to:

$$\text{offset} = 1 \ll (\text{rightShift} - 1), \text{ and}$$

$$y = (x \geq 0 ? (x + \text{offset}) \gg \text{rightShift} : -((-x + \text{offset}) \gg \text{rightShift})) \ll \text{leftShift},$$

where rightShift corresponds to the right-shift parameter, leftShift corresponds to the left-shift parameter, x corresponds to the first component, and y corresponds to the shifted first component.

[0015] In some embodiments, the processing circuitry determines the first and second reference motion vectors based on the at least one previously decoded block that includes a spatial neighboring block and a temporal neighboring block.

[0016] In some embodiments, the first and second reference motion vectors in the list of reference motion vectors correspond to 1/16 fractional-sample accuracy.

[0017] 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

[0018] 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:

[0019] FIG. 1 is a schematic illustration of a current block and its surrounding spatial merge candidates in one example.

[0020] FIG. 2 is a schematic illustration of a simplified block diagram of a communication system (200) in accordance with an embodiment.

[0021] FIG. 3 is a schematic illustration of a simplified block diagram of a communication system (300) in accordance with an embodiment.

[0022] FIG. 4 is a schematic illustration of a simplified block diagram of a decoder in accordance with an embodiment.

[0023] FIG. 5 is a schematic illustration of a simplified block diagram of an encoder in accordance with an embodiment.

[0024] FIG. 6 shows a block diagram of an encoder in accordance with another embodiment.

[0025] FIG. 7 shows a block diagram of a decoder in accordance with another embodiment.

[0026] FIG. 8 is a schematic illustration of a current block, its surrounding spatial candidates, and its collocated candidates in one embodiment.

[0027] FIG. 9 shows a flow chart outlining a decoding process (900) according to an embodiment of the disclosure.

[0028] FIG. 10 shows a flow chart outlining an encoding process (1000) according to an embodiment of the disclosure.

[0029] FIG. 11 is a schematic illustration of a computer system in accordance with an embodiment.

#### DETAILED DESCRIPTION OF EMBODIMENTS

[0030] FIG. 2 illustrates a simplified block diagram of a communication system (200) according to an embodiment of the present disclosure. The communication system (200) includes a plurality of terminal devices that can communicate with each other, via, for example, a network (250). For example, the communication system (200) includes a first pair of terminal devices (210) and (220) interconnected via the network (250). In the FIG. 2 example, the first pair of terminal devices (210) and (220) performs unidirectional transmission of data. For example, the terminal device (210) may code video data (e.g., a stream of video pictures that are captured by the terminal device (210)) for transmission to the other terminal device (220) via the network (250). The encoded video data can be transmitted in the form of one or more coded video bitstreams. The terminal device (220) may receive the coded video data from the network (250), 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.

[0031] In another example, the communication system (200) includes a second pair of terminal devices (230) and (240) 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 (230) and (240) 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 (230) and (240) via the network (250). Each terminal device of the terminal devices (230) and (240) also may receive the coded video data transmitted by the other terminal device of the terminal devices (230) and (240), 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.

[0032] In the FIG. 2 example, the terminal devices (210), (220), (230) and (240) 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 (250) represents any number of networks that convey coded video data among the terminal devices (210), (220), (230), and (240), including for example wireline (wired) and/or wireless communication networks. The communication network (250) 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 (250) may be immaterial to the operation of the present disclosure unless explained herein below.

**[0033]** FIG. 3 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.

**[0034]** A streaming system may include a capture subsystem (313), that can include a video source (301), for example a digital camera, creating for example a stream of video pictures (302) that are uncompressed. In an example, the stream of video pictures (302) includes samples that are taken by the digital camera. The stream of video pictures (302), depicted as a bold line to emphasize a high data volume when compared to encoded video data (304) (or coded video bitstreams), can be processed by an electronic device (320) that includes a video encoder (303) coupled to the video source (301). The video encoder (303) 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 (304) (or encoded video bitstream (304)), depicted as a thin line to emphasize the lower data volume when compared to the stream of video pictures (302), can be stored on a streaming server (305) for future use. One or more streaming client subsystems, such as client subsystems (306) and (308) in FIG. 3 can access the streaming server (305) to retrieve copies (307) and (309) of the encoded video data (304). A client subsystem (306) can include a video decoder (310), for example, in an electronic device (330). The video decoder (310) decodes the incoming copy (307) of the encoded video data and creates an outgoing stream of video pictures (311) that can be rendered on a display (312) (e.g., display screen) or other rendering device (not depicted). In some streaming systems, the encoded video

data (304), (307), and (309) (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.

**[0035]** It is noted that the electronic devices (320) and (330) can include other components (not shown). For example, the electronic device (320) can include a video decoder (not shown) and the electronic device (330) can include a video encoder (not shown) as well.

**[0036]** FIG. 4 shows a block diagram of a video decoder (410) according to an embodiment of the present disclosure. The video decoder (410) can be included in an electronic device (430). The electronic device (430) can include a receiver (431) (e.g., receiving circuitry). The video decoder (410) can be used in the place of the video decoder (310) in the FIG. 3 example.

**[0037]** The receiver (431) may receive one or more coded video sequences to be decoded by the video decoder (410); 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 (401), which may be a hardware/software link to a storage device which stores the encoded video data. The receiver (431) 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 (431) may separate the coded video sequence from the other data. To combat network jitter, a buffer memory (415) may be coupled in between the receiver (431) and an entropy decoder / parser (420) ("parser (420)" henceforth). In certain applications, the buffer memory (415) is part of the video decoder (410). In others, it can be outside of the video decoder (410) (not depicted). In still others, there can be a buffer memory (not depicted) outside of the video decoder (410), for example to combat network jitter, and in addition another buffer memory (415) inside the video decoder (410), for example to handle playout timing. When the receiver (431) is receiving data from a store/forward device of sufficient bandwidth and controllability, or from an isosynchronous network, the buffer memory (415) may not be needed, or can be small. For use on best effort packet networks such as the Internet, the buffer memory (415) 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 (410).

**[0038]** The video decoder (410) may include the parser (420) to reconstruct symbols (421) from the coded video sequence. Categories of those symbols include information used to manage operation of the video decoder (410), and potentially information to control a rendering device such as a render device (412) (e.g., a display screen) that is not an integral part of the electronic device (430) but can be coupled to the electronic device (430), as was shown in FIG. 4. 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 (420) 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 (420) 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 (420) may also extract from the coded video sequence information such as transform coefficients, quantizer parameter values, motion vectors, and so forth.

**[0039]** The parser (420) may perform an entropy decoding / parsing operation on the video sequence received from the buffer memory (415), so as to create symbols (421).

**[0040]** Reconstruction of the symbols (421) 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 (420). The flow of such subgroup control information between the parser (420) and the multiple units below is not depicted for clarity.

**[0041]** Beyond the functional blocks already mentioned, the video decoder (410) 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.

**[0042]** A first unit is the scaler / inverse transform unit (451). The scaler / inverse transform unit (451) 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) (421) from the parser (420). The scaler / inverse transform unit (451) can output blocks comprising sample values, that can be input into aggregator (455).

**[0043]** In some cases, the output samples of the scaler / inverse transform (451) 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 (452). In some cases, the intra picture prediction unit (452) 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 (458). The current picture buffer (458) buffers, for example, partly reconstructed current picture and/or fully reconstructed current picture. The aggregator (455), in some cases, adds, on a per sample basis, the prediction information the intra prediction unit (452) has generated to the output sample information as provided by the scaler / inverse transform unit (451).

**[0044]** In other cases, the output samples of the scaler / inverse transform unit (451) can pertain to an inter coded, and potentially motion compensated block. In such a case, a motion compensation prediction unit (453) can access reference picture memory (457) to fetch samples used for prediction. After motion compensating the fetched samples in accordance with the symbols (421) pertaining to the block, these samples can be added by the aggregator (455) to the output of the scaler / inverse transform unit (451) (in this case called the residual samples or residual signal) so as to generate output sample information. The addresses within the reference picture memory (457) from where the motion compensation prediction unit (453) fetches prediction samples can be controlled by motion vectors, available to the motion compensation prediction unit (453) in the form of symbols (421) 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 (457) when sub-sample exact motion vectors are in use, motion vector prediction mechanisms, and so forth.

**[0045]** The output samples of the aggregator (455) can be subject to various loop filtering techniques in the loop filter unit (456). 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 (456) as symbols (421) from the parser (420), 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.

**[0046]** The output of the loop filter unit (456) can be a sample stream that can be output to the render device (412) as well as stored in the reference picture memory (457) for use in future inter-picture prediction.

**[0047]** 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 (420)), the current picture buffer (458) can become a part of the reference picture memory (457), and a fresh current picture buffer can be reallocated before commencing the reconstruction of the following coded picture.

**[0048]** The video decoder (410) 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.

**[0049]** In an embodiment, the receiver (431) 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 (410) 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.

**[0050]** FIG. 5 shows a block diagram of a video encoder (503) according to an embodiment of the present disclosure. The video encoder (503) is included in an electronic device (520). The electronic device (520) includes a transmitter (540) (e.g., transmitting circuitry). The video encoder (503) can be used in the place of the video encoder (303) in the FIG. 3 example.

**[0051]** The video encoder (503) may receive video samples from a video source (501) (that is not part of the electronic device (520) in the FIG. 5 example) that may capture video image(s) to be coded by the video encoder (503). In another example, the video source (501) is a part of the electronic device (520).

**[0052]** The video source (501) may provide the source video sequence to be coded by the video encoder (503) 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 (501) may be a storage device storing previously prepared video. In a videoconferencing system, the video source (501) 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.

**[0053]** According to an embodiment, the video encoder (503) may code and compress the pictures of the source video sequence into a coded video sequence (543) in real time or under any other time constraints as required by the application. Enforcing appropriate coding speed is one function of a controller (550). In some embodiments, the controller (550) 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 (550) 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 (550) can be configured to have other suitable functions that pertain to the video encoder (503) optimized for a certain system design.

**[0054]** In some embodiments, the video encoder (503) is configured to operate in a coding loop. As an oversimplified description, in an example, the coding loop can include a source coder (530) (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 (533) embedded in the video encoder (503). The decoder (533) 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 (534). 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 (534) 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.

**[0055]** The operation of the "local" decoder (533) can be the same as of a "remote" decoder, such as the video decoder (410), which has already been described in detail above in conjunction with FIG. 4. Briefly referring also to FIG. 4, however, as symbols are available and encoding/decoding of symbols to a coded video sequence by an entropy coder (545) and the parser (420) can be lossless, the entropy decoding parts of the video decoder (410), including the buffer memory (415), and parser (420) may not be fully implemented in the local decoder (533).

**[0056]** 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.

**[0057]** During operation, in some examples, the source coder (530) 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 (532) 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.

**[0058]** The local video decoder (533) may decode coded video data of pictures that may be designated as reference pictures, based on symbols created by the source coder (530). Operations of the coding engine (532) may advantageously be lossy processes. When the coded video data may be decoded at a video decoder (not shown in FIG. 5), the reconstructed video sequence typically may be a replica of the source video sequence with some errors. The local video decoder (533) 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 (534). In this manner, the video encoder (503) 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).

**[0059]** The predictor (535) may perform prediction searches for the coding engine (532). That is, for a new picture to be coded, the predictor (535) may search the reference picture memory (534) 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 (535) 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 (535), an input picture may have prediction references drawn from multiple reference pictures stored in the reference picture memory (534).

**[0060]** The controller (550) may manage coding operations of the source coder (530), including, for example, setting of parameters and subgroup parameters used for encoding the video data.

**[0061]** Output of all aforementioned functional units may be subjected to entropy coding in the entropy coder (545). The entropy coder (545) 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.

**[0062]** The transmitter (540) may buffer the coded video sequence(s) as created by the entropy coder (545) to prepare for transmission via a communication channel (560), which may be a hardware/software link to a storage device which would store the encoded video data. The transmitter (540) may merge coded video data from the video coder (503) with other data to be transmitted, for example, coded audio data and/or ancillary data streams (sources not shown).

**[0063]** The controller (550) may manage operation of the video encoder (503). During coding, the controller (550) 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:

**[0064]** 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.

**[0065]** 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.

**[0066]** 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.

**[0067]** 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.

**[0068]** The video encoder (503) 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 (503) 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.

**[0069]** In an embodiment, the transmitter (540) may transmit additional data with the encoded video. The source coder (530) 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.

**[0070]** 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.

**[0071]** 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.

**[0072]** Further, a merge mode technique can be used in the inter-picture prediction to improve coding efficiency.

**[0073]** 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.

**[0074]** FIG. 6 shows a diagram of a video encoder (603) according to another embodiment of the disclosure. The video encoder (603) 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 (603) is used in the place of the video encoder (303) in the FIG. 3 example.

**[0075]** In an HEVC example, the video encoder (603) receives a matrix of sample values for a processing block, such as a prediction block of 8x8 samples, and the like. The video encoder (603) 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 (603) 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 (603) 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 (603) includes other components, such as a mode decision module (not shown) to determine the mode of the processing blocks.

**[0076]** In the FIG. 6 example, the video encoder (603) includes the inter encoder (630), an intra encoder (622), a residue calculator (623), a switch (626), a residue encoder (624), a general controller (621), and an entropy encoder (625) coupled together as shown in FIG. 6.

**[0077]** The inter encoder (630) 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.

**[0078]** The intra encoder (622) 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 (622) also calculates intra prediction results (e.g., predicted block) based on the intra prediction information and reference blocks in the same picture.

**[0079]** The general controller (621) is configured to determine general control data and control other components of the video encoder (603) based on the general control data. In an example, the general controller (621) determines the mode of the block, and provides a control signal to the switch (626) based on the mode. For example, when the mode is the intra mode, the general controller (621) controls the switch (626) to select the intra mode result for use by the residue calculator (623), and controls the entropy encoder (625) 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 (621) controls the switch (626) to select the inter prediction result for use by the residue calculator (623), and controls the entropy encoder (625) to select the inter prediction information and include the inter prediction information in the bitstream.

**[0080]** The residue calculator (623) is configured to calculate a difference (residue data) between the received block and prediction results selected from the intra encoder (622) or the inter encoder (630). The residue encoder (624) 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 (624) 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 (603) also includes a residue decoder (628). The residue decoder (628) is configured to perform inverse-transform, and generate the decoded residue data. The decoded residue data can be suitably used by the intra encoder (622) and the inter encoder (630). For example, the inter encoder (630) can generate decoded blocks based on the decoded residue data and inter prediction information, and the intra encoder (622) 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.

**[0081]** The entropy encoder (625) is configured to format the bitstream to include the encoded block. The entropy encoder (625) is configured to include various information according to a suitable standard, such as the HEVC standard. In an example, the entropy encoder (625) 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.

**[0082]** FIG. 7 shows a diagram of a video decoder (710) according to another embodiment of the disclosure. The video decoder (710) 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 (710) is used in the place of the video decoder (310) in the FIG. 3 example.

**[0083]** In the FIG. 7 example, the video decoder (710) includes an entropy decoder (771), an inter decoder (780), a residue decoder (773), a reconstruction module (774), and an intra decoder (772) coupled together as shown in FIG. 7.

**[0084]** The entropy decoder (771) 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 (772) or the inter decoder (780), 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 (780); and when the prediction type is the intra prediction type, the intra prediction information is provided to the intra decoder (772). The residual information can be subject to inverse quantization and is provided to the residue decoder (773).

**[0085]** The inter decoder (780) is configured to receive the inter prediction information, and generate inter prediction results based on the inter prediction information.

**[0086]** The intra decoder (772) is configured to receive the intra prediction information, and generate prediction results based on the intra prediction information.

**[0087]** The residue decoder (773) 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 (773) may also require certain control information (to include the Quantizer Parameter (QP)), and that information may be provided by the entropy decoder (771) (data path not depicted as this may be low volume control information only).

**[0088]** The reconstruction module (774) is configured to combine, in the spatial domain, the residual as output by the residue decoder (773) 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.

**[0089]** It is noted that the video encoders (303), (503), and (603), and the video decoders (310), (410), and (710) can be implemented using any suitable technique. In an embodiment, the video encoders (303), (503), and (603), and the video decoders (310), (410), and (710) can be implemented using one or more integrated circuits. In another embodiment, the

video encoders (303), (503), and (503), and the video decoders (310), (410), and (710) can be implemented using one or more processors that execute software instructions.

**[0090]** FIG. 8 is a schematic illustration of a current block (801), its spatial neighboring blocks, and its temporal neighboring blocks in one embodiment.

**[0091]** Referring to FIG. 8, the motion vector of a current block (801) can be predicted based on a list of reference motion information, which may include at least a list of reference motion vectors (also referred to as a “candidate list”) and identification information of corresponding reference pictures. The list of reference motion vectors can be derived based on motion vectors of neighboring blocks (also referred to as “candidates”), such as spatial neighboring blocks denoted A0, A1, B0, B1, and B2 (802 through 806, respectively) and/or temporal neighboring blocks denoted C0 and C1 (812 and 813, respectively). In some examples, spatial neighboring blocks A0, A1, B0, B1, and B2 and the current block belong to a same picture. In some examples, temporal neighboring blocks C0 and C1 belong to a reference picture, where block C0 corresponds to a position outside the current block (801) and adjacent to a lower-right corner of the current block (801), and block C1 corresponds to a position at the lower-right side of, and adjacent to, a center of the block (801).

**[0092]** In some examples, the list of reference motion vectors can be constructed using “merge” mode. For an inter prediction in the merge mode, in one example with reference to FIG. 8, neighboring blocks A1, B1, B0, A0, and B2 are sequentially checked. When any of the checked blocks includes a valid motion vector, the valid motion vector can be added to the list of reference motion vectors. A pruning operation can be performed to avoid duplicated motion vectors being included the list.

**[0093]** Temporal neighboring blocks can be checked to add reference motion information into the list. The temporal neighboring blocks can be checked after the spatial neighboring blocks. In some examples, the motion vector of block C0, if available, is added to the list as a temporal merge candidate. If block C0 is not coded in inter mode or not available, the motion vector of block C1 can be used as the temporal merge candidate instead.

**[0094]** In some examples, the list of reference motion vectors can be constructed using “Advanced Motion Vector Prediction” (AMVP) mode. For an inter prediction in the AMVP mode, at most one motion vector from the spatial neighboring blocks A0, A1, B0, B1, and B2 and at most one temporal neighboring blocks C0 and C1 can be included in the list of reference motion vectors. In one example with reference to FIG. 8, a first reference motion vector in the

list can be the first available motion vector from the left edge of the current block (801) by sequentially checking neighboring blocks A0 and A1; and a second reference motion vector in the list can be the first available motion vector from the upper edge of the current block (801) by sequentially checking neighboring blocks B0, B1, and B2. In some examples, if a valid motion vector cannot be found for the checked neighboring blocks, no reference motion vectors are added to the list. If the first and second reference motion vectors are available and are the same, only one will be kept in the list. If the list is not full (e.g., with two different reference motion vectors), the motion vector of the temporal neighboring block C0 can be included in the list. If the temporal neighboring block C0 does not have valid motion vector, the motion vector of the temporal neighboring block C1 can be included in the list instead. Moreover, in some examples, if there are still not enough motion vectors in the list, a zero motion vector can be used to fill up the list.

**[0095]** In some examples, the list of reference motion vectors can be constructed by determining various reference motion vectors based on coding information of at least one previously encoded or decoded block, such as spatial neighboring blocks A0, A1, B0, B1, and B2 and/or temporal neighboring blocks C0 and C1 using the merge mode, the AMVP mode, or any applicable list construction process. After reference motion vectors that are based on coding information of previously coded blocks have been determined, additional averaged reference motion vectors can be determined by pairwise averaging the determined reference motion vectors according to a predetermined pairing.

**[0096]** In some examples, there can be at most four reference motion vectors that have been determined to be included in a list of reference motion vectors for coding a current block based on at least one previously coded block. These four reference motion vectors may be arranged in the list in association with indices (e.g., 0, 1, 2, and 3) indicating an order of the reference motion vectors in the list. Additional averaged reference motion vectors may be respectively determined based on the pairs defined as  $\{(0, 1), (0, 2), (1, 2), (0, 3), (1, 3), (2, 3)\}$  for example. In some examples, if both reference motion vectors of a predefined pair are available in the list of reference motion vectors, these two reference motion vectors are averaged even when they point to different reference pictures.

**[0097]** In some examples, a process for determining pairwise averaged reference motion vectors can be performed, by executing software or operating hardware, as follows.

[0098] Initially, the process for determining pairwise averaged reference motion vectors can take the following parameters as input:

a merge candidate list **mergeCandList**, which is a list of reference motion vectors determined based on coding information of previously coded blocks, the reference indices **refIdxL0N** and **refIdxLIN** of every candidate **N** in **mergeCandList**,  
 the prediction list utilization flags **predFlagL0N** and **predFlagLIN** of every candidate **N** in **mergeCandList**,  
 the motion vectors in 1/16 fractional-sample accuracy **mvL0N** and **mvLIN** of every candidate **N** in **mergeCandList**, and  
 the number of elements **numCurrMergeCand** within **mergeCandList**.

[0099] The outputs of this process can include:

the merge candidate list **mergeCandList**, which is an updated list of reference motion vectors with additional averaged motion vectors added thereto, the number of elements **numCurrMergeCand** within **mergeCandList**, the reference indices **refIdxL0avgCand<sub>k</sub>** and **refIdxL1avgCand<sub>k</sub>** of every new candidate **avgCand<sub>k</sub>** added into **mergeCandList** during the invocation of this process,  
 the prediction list utilization flags **predFlagL0avgCand<sub>k</sub>** and **predFlagL1avgCand<sub>k</sub>** of every new candidate **avgCand<sub>k</sub>** added into **mergeCandList** during the invocation of this process, and  
 the motion vectors in 1/16 fractional-sample accuracy **mvL0avgCand<sub>k</sub>** and **mvL1avgCand<sub>k</sub>** of every new candidate **avgCand<sub>k</sub>** added into **mergeCandList** during the invocation of this process.

[0100] The variables **numInputMergeCand**, **numCutoff**, **avgIdx**, **avgStop**, **numRefLists** can be set as follows:

**numInputMergeCand** = **numCurrMergeCand**,  
**numCutoff** = **min(numCurrMergeCand, 4)**,  
**avgIdx** = 0,  
**avgStop** = (**avgIdx** >= (**numCutoff** \* (**numCutoff** - 1) / 2) ||  
**numCurrMergeCand** == **MaxNumMergeCand**), and  
**numRefLists** = (**slice\_type** == B) ? 2: 1.

[0101] Accordingly, the process can be executed in a recursive manner, using **avgIdx** as an index and **avgStop** as a stop condition, where **avgStop** can be determined based on a number of additional averaged reference motion vectors to be calculated and added to the original group of reference motion vectors that are determined based on coding information of previously coded blocks.

[0102] Here, the merge candidate list **mergeCandList** is a list of reference motion vectors constructed using the merge mode, for example. In some embodiments, the process for determining pairwise averaged reference motion vectors is applicable to process a list of reference motion vectors constructed using the AMVP mode, or any applicable list construction process.

[0103] In some embodiments, the process for determining pairwise averaged reference motion vectors can be performed according to the following steps and repeated until **avgStop** is equal to TRUE.

[0104] Step 1: The variables **p0CandIdx** and **p1CandIdx** are derived using **avgIdx** as specified in the following Table. Therefore, for determining the (**avgIdx**+1)-th averaged reference motion vector to be added to the list of reference motion vectors, a corresponding pair of determined reference motion vectors can be identified based on the associated indices, such that the averaged reference motion vectors can be respectively determined based on the pairs defined as {(0, 1), (0, 2), (1, 2), (0, 3), (1, 3), (2, 3)}.

Table – Specification of **p0CandIdx** and **p1CandIdx**

<b>avgIdx</b>	0	1	2	3	4	5
<b>p0CandIdx</b>	0	0	1	0	1	2
<b>p1CandIdx</b>	1	2	2	3	3	3

[0105] Step 2: The following assignments are made, with **p0Cand** being a first reference motion vector associated with index **p0CandIdx** and **p1Cand** being a second reference motion vector associated with index **p1CandIdx** in the merging candidate list **mergeCandList**:

**p0Cand** = **mergeCandList**[**p0CandIdx**], and

**p1Cand** = **mergeCandList**[**p1CandIdx**].

[0106] Step 3: The candidate **avgCand<sub>k</sub>**, with **k** equal to (**numCurrMergeCand** – **numInputMergeCand**) can be determined and added at the end of **mergeCandList**, i.e.,

**mergeCandList[numCurrMergeCand]** is set equal to **avgCand<sub>k</sub>**, and the reference indices, the prediction list utilization flags, and the motion vectors of **avgCand<sub>k</sub>** can be derived as follows:

For each reference picture list LX with X ranging from 0 to (**numRefLists** - 1):

In a first scenario when all of the following conditions are true:

**predFlagLXp0Cand** == 1, and

**predFlagLXp1Cand** == 1,

which indicates that a motion vector of **avgCand<sub>k</sub>** is to be determined by averaging the motion vectors of the corresponding predefined pair of candidates,

set

**refIdxLXavgCand<sub>k</sub>** = **refIdxLXp0Cand**, and

**predFlagLXavgCand<sub>k</sub>** = 1,

set

**mvLXp0Cand** as an input motion vector,

a right shift parameter **rightShift** as 1, and

a left shift parameter **leftShift** as 0,

execute a rounding subroutine or operate a rounding circuit that is configured to

receive **mvLXp0Cand[0]**, **mvLXp1Cand[0]**, **mvLXp0Cand[1]**,

**mvLXp1Cand[1]**, the right-shift parameter **rightShift**, and the left-shift

parameter **rightShift**, and to calculate shifted components according to:

**offset** = 1 << (**rightShift** - 1), and

**y** = ( **x** >= 0 ? ( **x** + **offset** ) >> **rightShift** : -(( -**x** + **offset** ) >> **rightShift** ) )  
<< **leftShift**,

where **x** corresponds to (**mvLXp0Cand[0]** + **mvLXp1Cand[0]**) or

(**mvLXp0Cand[1]** + **mvLXp1Cand[1]**), and **y** corresponds to

**mvLXavgCand<sub>k</sub>[0]** or **mvLXavgCand<sub>k</sub>[1]**, respectively.

**mvLXp0Cand** represents a first reference motion vector of the predefined pair of reference motion vectors,

**mvLXp1Cand** represents a second reference motion vector of the predefined pair of reference motion vectors,

**mvLXavgCand<sub>k</sub>** represents an averaged reference motion vector determined based on the predefined pair of reference motion vectors,

**mvLXp0Cand[0]** represents a first component, on a first coordinate axis (e.g., x-axis), of the first reference motion vector,

**mvLXp1Cand[0]** represents a second component, on the first coordinate axis, of the second reference motion vector,

**mvLXavgCand<sub>k</sub>[0]** represents a third component, on the first coordinate axis, of the averaged reference motion vector,

**mvLXp0Cand[1]** represents a fourth component, on a second coordinate axis (e.g., y-axis), of the first reference motion vector,

**mvLXp1Cand[1]** represents a fifth component, on the second coordinate axis, of the second reference motion vector, and

**mvLXavgCand<sub>k</sub>[1]** represents a sixth component, on the second coordinate axis, of the averaged reference motion vector.

In a second scenario when all of the following conditions are true:

**predFlagLXp0Cand** == 1, and

**predFlagLXp1Cand** == 0,

which indicates that a motion vector of **avgCand<sub>k</sub>** is to be determined by only one of the motion vectors of the corresponding predefined pair of candidates,

set

**refIdxLXavgCand<sub>k</sub>** = **refIdxLXp0Cand**,

**predFlagLXavgCand<sub>k</sub>** = 1,

**mvLXavgCand<sub>k</sub>[0]** = **mvLXp0Cand[0]**, and

**mvLXavgCand<sub>k</sub>[1]** = **mvLXp0Cand[1]**.

In a third scenario when all of the following conditions are true:

**predFlagLXp0Cand** == 0, and

**predFlagLXp1Cand** == 1,

which indicates that a motion vector of **avgCand<sub>k</sub>** is to be determined by only one of the motion vectors of the corresponding predefined pair of candidates,

set

**refIdxLXavgCand<sub>k</sub>** = **refIdxLXp1Cand**,  
**predFlagLXavgCand<sub>k</sub>** = 1,  
**mvLXavgCand<sub>k</sub>[0]** = **mvLXp1Cand[0]**, and  
**mvLXavgCand<sub>k</sub>[1]** = **mvLXp1Cand[1]**.

In a fourth scenario when all of the following conditions are true:

**predFlagLXp0Cand** == 0, and  
**predFlagLXp1Cand** == 0,

which indicates that a motion vector of **avgCand<sub>k</sub>** is not to be determined,

set

**refIdxLXavgCand<sub>k</sub>** = -1,  
**predFlagLXavgCand<sub>k</sub>** = 0,  
**mvLXavgCand<sub>k</sub>[0]** = 0, and  
**mvLXavgCand<sub>k</sub>[1]** = 0.

[0107] Step 4: The variable **numCurrMergeCand** and **avgIdx** are incremented by 1.

[0108] Step 5: When **avgIdx** is larger or equal to  $(\text{numCutoff} * (\text{numCutoff} - 1) / 2)$  or **numCurrMergeCand** is equal to **MaxNumMergeCand**, **avgStop** is set equal to TRUE, the process terminates. Otherwise, the process repeats Step 1 through Step 5 with updated indices to determine a next additional averaged reference motion vector.

[0109] In some examples, the rounding subroutine or the rounding circuit as described in Step 3, with respect to the first scenario, can be configured to perform a rounding operation on motion vectors or any given values based on the set parameters **rightShift** and **leftShift**.

Therefore, the rounding subroutine or the rounding circuit can be used not only for determining the averaged reference motion vector, but also rounding a given number with given settings in general. Accordingly, a decoder can reuse a same rounding subroutine or rounding circuit for various applications, including determining the averaged reference motion vectors, which can simplify the implementation of the program or circuit of the decoder, hence reducing software and/or hardware complexity of the decoder.

[0110] In some examples, various components of motion vectors, such as **mvLXp0Cand[0]**, **mvLXp1Cand[0]**, **mvLXavgCandk[0]**, **mvLXp0Cand[1]**, **mvLXp1Cand[1]**, and **mvLXavgCandk[1]**, are in a two's complement integer representation format. Also, in some examples, various components of motion vectors in the list of reference

motion vectors have 1/16 fractional-sample accuracy. In other examples, various components of motion vectors can be represented using a different binary number format with a different fractional-sample accuracy setting, and the rounding subroutine or the rounding circuit can be modified to accommodate such a binary number format and/or fractional-sample accuracy setting.

**[0111]** FIG. 9 shows a flow chart outlining a decoding process (900) according to an embodiment of the disclosure. The process (900) can be used in the reconstruction of a block (i.e., a current block) of a picture coded in inter mode. In some embodiments, one or more operations are performed before or after process (900), and some of the operations illustrated in FIG. 9 may be reordered or omitted.

**[0112]** In various embodiments, the process (900) is executed by processing circuitry, such as the processing circuitry in the terminal devices (210), (220), (230), and (240), the processing circuitry that performs functions of the video decoder (310), (410), or (710), and the like. In some embodiments, the process (900) is implemented by software instructions, thus when the processing circuitry executes the software instructions, the processing circuitry performs the process (900). The process starts at (S901) and proceeds to (S910).

**[0113]** At (S910), a first reference motion vector and a second reference motion vector to be included in a list of reference motion vectors for decoding a current block are determined based on coding information of at least one previously decoded block. In some embodiments, the first reference motion vector and the second reference motion vector can be determined based on at least the merge mode or the AMVP mode as described above with reference to FIG. 8, or other another applicable method for determining reference motion vectors based on coding information of a previously decoded block(s). In some examples, the first reference motion vector and the second reference motion vector can be determined using the system or decoders illustrated in FIGs. 3, 4, and 7.

**[0114]** At (S920), a first component, on a first coordinate axis (e.g., x-axis) and in a two's complement integer representation format, of the first reference motion vector is offset by one, and the offset first component is right-shifted by one bit to obtain a shifted first component. Also, at (S920), a fourth component, on a second coordinate axis (e.g., y-axis) and in the two's complement integer representation format, of the first reference motion vector is offset by one, and the offset fourth component is right-shifted by one bit to obtain a shifted fourth component.

**[0115]** At (S930), a second component, on the first coordinate axis and in the two's complement integer representation format, of the second reference motion vector is offset by one, and the offset first component is right-shifted by one bit to obtain a shifted second component. Also, at (S930), a fifth component, on the second coordinate axis and in the two's complement integer representation format, of the first reference motion vector is offset by one, and the offset fifth component is right-shifted by one bit to obtain a shifted fifth component.

**[0116]** In some embodiments, (S920) and (S930) can be performed by reusing a rounding subroutine or a rounding circuit. For example, the offsetting of the first, second, fourth, or fifth component and the right-shifting of the respective offset component can be performed by setting a right-shift parameter as 1, setting a left-shift parameter as 0, and then executing the rounding subroutine or operating the rounding circuit that is configured to receive the respective component, the right-shift parameter, and the left-shift parameter, and configured to calculate the shifted version of the corresponding component according to:

$$\text{offset} = 1 \ll (\text{rightShift} - 1), \text{ and}$$

$$y = (x \geq 0 ? (x + \text{offset}) \gg \text{rightShift} : -((-x + \text{offset}) \gg \text{rightShift})) \ll \text{leftShift},$$

where rightShift corresponds to the right-shift parameter, leftShift corresponds to the left-shift parameter, x corresponds to the corresponding component, and y corresponds to the shifted version of the corresponding component.

**[0117]** In some examples, the rounding subroutine or the rounding circuit can be incorporated in the system or decoders illustrated in FIGs. 3, 4, and 7.

**[0118]** At (S940), a third component, on the first coordinate axis and in the two's complement integer representation format, of a first averaged reference motion vector is generated by adding the shifted first component and the shifted second component. Also, at (S940), a sixth component, on the second coordinate axis and in the two's complement integer representation format, of the first averaged reference motion vector is generated by adding the shifted fourth component and the shifted fifth component. In some examples, (S940) can be performed using the system or decoders illustrated in FIGs. 3, 4, and 7.

**[0119]** At (S950), a list of reference motion vectors for decoding the current block is constructed, where the list of reference motion vectors incorporates the first and second reference motion vectors and the first averaged reference motion vector. In some embodiments, (S910), (S920), (S930), (S940), and (S950) may be repetitively performed according to predefined pairs of reference motion vectors to be used to generate corresponding averaged reference motion

vectors. The predefined pairs may be defined in the predefined tables and/or the Table described above. In some examples, (S950) can be performed using the system or decoders illustrated in FIGs. 3, 4, and 7.

**[0120]** In one example, (S910), (S920), (S930), (S940), and (S950) may be performed to generate a second averaged reference motion vector based on the first reference motion vector and a third reference motion vector and to generate a third averaged reference motion vector based on the second reference motion vector and the third reference motion vector. The list of reference motion vectors for decoding the current block can be constructed by arranging the list of reference motion vectors to include, according to an ascending order of respective indices associated thereto, the first reference motion vector, the second reference motion vector, the third reference motion vector, the first averaged reference motion vector, the second averaged reference motion vector, and the third averaged reference motion vector.

**[0121]** In one example, (S910), (S920), (S930), (S940), and (S950) may be further performed to generate a fourth averaged reference motion vector based on the first reference motion vector and a fourth reference motion vector, a fifth averaged reference motion vector based on the second reference motion vector and the fourth reference motion vector, and a sixth averaged reference motion vector based on the third reference motion vector and the fourth reference motion vector. The list of reference motion vectors for decoding the current block can be constructed by arranging the list of reference motion vectors to include, according to the ascending order of respective indices associated thereto, the first reference motion vector, the second reference motion vector, the third reference motion vector, the fourth reference motion vector, the first list reference motion vectors and adding the first averaged reference motion vector, the second averaged reference motion vector, the third averaged reference motion vector, the fourth averaged reference motion vector, the fifth averaged reference motion vector, and the sixth averaged reference motion vector.

**[0122]** At (S960), a motion vector predictor is determined using the list of reference motion vectors. At (S970), the current block is decoded for output based on the determined motion vector predictor. In some embodiments, (S960) and (S970) can be performed using the system or decoders illustrated in FIGs. 3, 4, and 7.

**[0123]** After (S970), the process proceeds to (S999) and terminates.

**[0124]** FIG. 10 shows a flow chart outlining an encoding process (1000) according to an embodiment of the disclosure. The process (1000) can be used to encode a block (i.e., a current

block) of a picture coded using inter mode. In some embodiments, one or more operations are performed before or after process (1000), and some of the operations illustrated in FIG. 10 may be reordered or omitted.

**[0125]** In various embodiments, the process (1000) is executed by processing circuitry, such as the processing circuitry in the terminal devices (210), (220), (230), and (240), the processing circuitry that performs functions of the video encoder (303), (503), or (603), and the like. In some embodiments, the process (1000) is implemented by software instructions, thus when the processing circuitry executes the software instructions, the processing circuitry performs the process (1000). The process starts at (S1001) and proceeds to (S1010).

**[0126]** At (S1010), a first reference motion vector and second reference motion vector to be included in a list of reference motion vectors for encoding a current block are determined based on coding information of at least one previously encoded block. In some embodiments, the first reference motion vector and the second reference motion vector are determined based on at least the merge mode or the AMVP mode as described above with reference to FIG. 8, or another applicable method for determining reference motion vectors based on coding information of a previously encoded block(s). In some examples, the first reference motion vector and the second reference motion vector are determined using the system or encoders illustrated in FIGs. 3, 5, and 6.

**[0127]** At (S1020), a first component, on a first coordinate axis (e.g., x-axis) and in a two's complement integer representation format, of the first reference motion vector is offset by one, and the offset first component is right-shifted by one bit to obtain a shifted first component. Also, at (S1020), a fourth component, on a second coordinate axis (e.g., y-axis) and in the two's complement integer representation format, of the first reference motion vector is offset by one, and the offset fourth component is right-shifted by one bit to obtain a shifted fourth component.

**[0128]** At (S1030), a second component, on the first coordinate axis and in the two's complement integer representation format, of the second reference motion vector is offset by one, and the offset first component is right-shifted by one bit to obtain a shifted second component. Also, at (S1030), a fifth component, on the second coordinate axis and in the two's complement integer representation format, of the first reference motion vector is offset by one, and the offset fifth component is right-shifted by one bit to obtain a shifted fifth component.

**[0129]** In some embodiments, (S1020) and (S1030) correspond to (S920) and (S930) in the decoding counterpart and can also be performed by reusing a rounding subroutine or a

rounding circuit. In some examples, the rounding subroutine or the rounding circuit can be incorporated in the system or encoders illustrated in FIGs. 3, 5, and 6. Detailed description of (S1020) and (S1030) is thus omitted.

**[0130]** At (S1040), a third component, on the first coordinate axis and in the two's complement integer representation format, of a first averaged reference motion vector is generated by adding the shifted first component and the shifted second component. Also, at (S1040), a sixth component, on the second coordinate axis and in the two's complement integer representation format, of the first averaged reference motion vector is generated by adding the shifted fourth component and the shifted fifth component.

**[0131]** At (S1050), a list of reference motion vectors for encoding the current block is constructed, where the list of reference motion vectors incorporates the first and second reference motion vectors and the first averaged reference motion vector. In some embodiments, (S1040) and (S1050) correspond to (S940) and (S950) in the decoding counterpart. In some examples, the rounding subroutine or the rounding circuit can be incorporated in the system or encoders illustrated in FIGs. 3, 5, and 6.

**[0132]** In some embodiments, (S1010), (S1020), (S1030), (S1040), and (S1050) may be repetitively performed according to predefined pairs of reference motion vectors to be used to generate corresponding averaged reference motion vectors in a manner similar to the example described above with reference to FIG. 9.

**[0133]** At (S1060), a motion vector predictor is determined using the list of reference motion vectors. At (S1070), the current block is encoded for output based on the determined motion vector predictor. In some embodiments, (S1060) and (S1070) can be performed in a manner as described above with reference to FIGs. 3, 5, and 6.

**[0134]** After (S1070), the process proceeds to (S1099) and terminates.

**[0135]** 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. 11 shows a computer system (1100) suitable for implementing certain embodiments of the disclosed subject matter.

**[0136]** 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.

**[0137]** 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.

**[0138]** The components shown in FIG. 11 for computer system (1100) 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 (1100).

**[0139]** Computer system (1100) 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).

**[0140]** Input human interface devices may include one or more of (only one of each depicted): keyboard (1101), mouse (1102), trackpad (1103), touch screen (1110), data-glove (not shown), joystick (1105), microphone (1106), scanner (1107), camera (1108).

**[0141]** Computer system (1100) 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 (1110), data-glove (not shown), or joystick (1105), but there can also be tactile feedback devices that do not serve as input devices), audio output devices (such as: speakers (1109), headphones (not depicted)), visual output devices (such as screens (1110) 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).

**[0142]** Computer system (1100) can also include human accessible storage devices and their associated media such as optical media including CD/DVD ROM/RW (1120) with CD/DVD or the like media (1121), thumb-drive (1122), removable hard drive or solid state drive (1123), 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.

**[0143]** 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.

**[0144]** Computer system (1100) can also include an interface to one or more communication networks. 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 (1149) (such as, for example USB ports of the computer system (1100)); others are commonly integrated into the core of the computer system (1100) 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 (1100) 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.

**[0145]** Aforementioned human interface devices, human-accessible storage devices, and network interfaces can be attached to a core (1140) of the computer system (1100).

**[0146]** The core (1140) can include one or more Central Processing Units (CPU) (1141), Graphics Processing Units (GPU) (1142), specialized programmable processing units in the form of Field Programmable Gate Areas (FPGA) (1143), hardware accelerators for certain tasks

(1144), and so forth. These devices, along with Read-only memory (ROM) (1145), Random-access memory (1146), internal mass storage such as internal non-user accessible hard drives, SSDs, and the like (1147), may be connected through a system bus (1148). In some computer systems, the system bus (1148) 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 (1148), or through a peripheral bus (1149). Architectures for a peripheral bus include PCI, USB, and the like.

**[0147]** CPUs (1141), GPUs (1142), FPGAs (1143), and accelerators (1144) can execute certain instructions that, in combination, can make up the aforementioned computer code. That computer code can be stored in ROM (1145) or RAM (1146). Transitional data can be also be stored in RAM (1146), whereas permanent data can be stored for example, in the internal mass storage (1147). 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 (1141), GPU (1142), mass storage (1147), ROM (1145), RAM (1146), and the like.

**[0148]** 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.

**[0149]** As an example and not by way of limitation, the computer system having architecture (1100), and specifically the core (1140) 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 (1140) that are of non-transitory nature, such as core-internal mass storage (1147) or ROM (1145). The software implementing various embodiments of the present disclosure can be stored in such devices and executed by core (1140). A computer-readable medium can include one or more memory devices or chips, according to particular needs. The software can cause the core (1140) 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 (1146) 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 (1144)), 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

SDR: Standard Dynamic Range

HDR: High Dynamic Range

**[0150]** 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 for video decoding in a decoder, comprising:
  - determining first and second reference motion vectors for decoding a current block based on coding information of at least one previously decoded block;
  - offsetting a first component, on a first coordinate axis and in a two's complement integer representation format, of the first reference motion vector by one, and right-shifting the offset first component by one bit to obtain a shifted first component;
  - offsetting a second component, on the first coordinate axis and in the two's complement integer representation format, of the second reference motion vector by one, and right-shifting the offset second component by one bit to obtain a shifted second component;
  - generating a third component, on the first coordinate axis and in the two's complement integer representation format, of a first averaged reference motion vector by adding the shifted first component and the shifted second component;
  - constructing a list of reference motion vectors for decoding the current block, the list of reference motion vectors incorporating the first and second reference motion vectors and the first averaged reference motion vector;
  - determining a motion vector predictor using the list of reference motion vectors; and
  - decoding the current block for output based on the determined motion vector predictor.
2. The method of claim 1, further comprising:
  - offsetting a fourth component, on a second coordinate axis and in the two's complement integer representation format, of the first reference motion vector by one, and right-shifting the offset fourth component by one bit to obtain a shifted fourth component,

offsetting a fifth component, on the second coordinate axis and in the two's complement integer representation format, of the second reference motion vector by one, and right-shifting the offset fifth component by one bit to obtain a shifted fifth component, and

generating a sixth component, on the second coordinate axis and in the two's complement integer representation format, of the first averaged reference motion vector by adding the shifted fourth component and the shifted fifth component.

3. The method of claim 1, further comprising:

generating a second averaged reference motion vector based on the first reference motion vector and a third reference motion vector; and

generating a third averaged reference motion vector based on the second reference motion vector and the third reference motion vector, wherein

the constructing the list of reference motion vectors for decoding the current block includes arranging the list of reference motion vectors to include, according to an ascending order of respective indices associated thereto, the first reference motion vector, the second reference motion vector, the third reference motion vector, the first averaged reference motion vector, the second averaged reference motion vector, and the third averaged reference motion vector.

4. The method of claim 3, further comprising:

generating a fourth averaged reference motion vector based on the first reference motion vector and a fourth reference motion vector;

generating a fifth averaged reference motion vector based on the second reference motion vector and the fourth reference motion vector; and

generating a sixth averaged reference motion vector based on the third reference motion vector and the fourth reference motion vector, wherein

the constructing the list of reference motion vectors for decoding the current block includes arranging the list of reference motion vectors to include, according to the ascending order of respective indices associated thereto, the first reference motion vector, the second reference motion vector, the third reference motion vector, the fourth reference motion vector, the first list reference motion vectors and adding the first averaged reference motion vector, the second averaged reference motion vector, the third averaged reference motion vector, the fourth averaged reference motion vector, the fifth averaged reference motion vector, and the sixth averaged reference motion vector.

5. The method of claim 1, wherein the offsetting the first component and the right-shifting the offset first component comprise:

setting a right-shift parameter as 1;

setting a left-shift parameter as 0; and

executing a rounding subroutine or operating a rounding circuit that is configured to receive the first component, the right-shift parameter, and the left-shift parameter, and to calculate the shifted first component according to:

$$\text{offset} = 1 \ll (\text{rightShift} - 1), \text{ and}$$

$$y = (x \geq 0 ? (x + \text{offset}) \gg \text{rightShift} : -((-x + \text{offset}) \gg \text{rightShift})) \ll$$

leftShift, wherein

rightShift corresponds to the right-shift parameter, leftShift corresponds to the left-shift parameter, x corresponds to the first component, and y corresponds to the shifted first component.

6. The method of claim 1, wherein the first and second reference motion vectors are determined based on the at least one previously decoded block that includes a spatial neighboring block and a temporal neighboring block.

7. The method of claim 1, wherein the first and second reference motion vectors in the list of reference motion vectors correspond to 1/16 fractional-sample accuracy.

8. An apparatus, comprising:

processing circuitry configured to:

determine first and second reference motion vectors for decoding a current block based on coding information of at least one previously decoded block;

offset a first component, on a first coordinate axis and in a two's complement integer representation format, of the first reference motion vector by one, and right-shift the offset first component by one bit to obtain a shifted first component;

offset a second component, on the first coordinate axis and in the two's complement integer representation format, of the second reference motion vector by one, and right-shift the offset second component by one bit to obtain a shifted second component;

generate a third component, on the first coordinate axis and in the two's complement integer representation format, of a first averaged reference motion vector by adding the shifted first component and the shifted second component;

construct a list of reference motion vectors for decoding the current block, the list of reference motion vectors incorporating the first and second reference motion vectors and the first averaged reference motion vector;

determine a motion vector predictor using the list of reference motion vectors; and

decode the current block for output based on the determined motion vector predictor.

9. The apparatus of claim 8, wherein the processing circuitry is further configured to:

offset a fourth component, on a second coordinate axis and in the two's complement integer representation format, of the first reference motion vector by one, and right-shift the offset fourth component by one bit to obtain a shifted fourth component,

offset a fifth component, on the second coordinate axis and in the two's complement integer representation format, of the second reference motion vector by one, and right-shift the offset fifth component by one bit to obtain a shifted fifth component, and

generate a sixth component, on the second coordinate axis and in the two's complement integer representation format, of the first averaged reference motion vector by adding the shifted fourth component and the shifted fifth component.

10. The apparatus of claim 8, wherein the processing circuitry is further configured to:

generate a second averaged reference motion vector based on the first reference motion vector and a third reference motion vector;

generate a third averaged reference motion vector based on the second reference motion vector and the third reference motion vector; and

construct the list of reference motion vectors for decoding the current block by arranging the list of reference motion vectors to include, according to an ascending order of respective indices associated thereto, the first reference motion vector, the second reference motion vector, the third reference motion vector, the first averaged reference motion vector, the second averaged reference motion vector, and the third averaged reference motion vector.

11. The apparatus of claim 10, wherein the processing circuitry is further configured to:  
generate a fourth averaged reference motion vector based on the first reference motion vector and a fourth reference motion vector;

generate a fifth averaged reference motion vector based on the second reference motion vector and the fourth reference motion vector;

generate a sixth averaged reference motion vector based on the third reference motion vector and the fourth reference motion vector; and

construct the list of reference motion vectors for decoding the current block by arranging the list of reference motion vectors to include, according to the ascending order of respective indices associated thereto, the first reference motion vector, the second reference motion vector, the third reference motion vector, the fourth reference motion vector, the first list reference motion vectors and adding the first averaged reference motion vector, the second averaged reference motion vector, the third averaged reference motion vector, the fourth averaged reference motion vector, the fifth averaged reference motion vector, and the sixth averaged reference motion vector.

12. The apparatus of claim 8, wherein the processing circuitry is further configured to:  
set a right-shift parameter as 1;

set a left-shift parameter as 0; and

execute a rounding subroutine or operate a rounding circuit that is configured to receive the first component, the right-shift parameter, and the left-shift parameter, and to calculate the shifted first component according to:

$$\text{offset} = 1 \ll (\text{rightShift} - 1), \text{ and}$$

$$y = (x \geq 0 ? (x + \text{offset}) \gg \text{rightShift} : -((-x + \text{offset}) \gg \text{rightShift})) \ll$$

leftShift, wherein

rightShift corresponds to the right-shift parameter, leftShift corresponds to the left-shift parameter, x corresponds to the first component, and y corresponds to the shifted first component.

13. The apparatus of claim 8, wherein the processing circuitry is further configured to:  
determine the first and second reference motion vectors based on the at least one previously decoded block that includes a spatial neighboring block and a temporal neighboring block.

14. The apparatus of claim 8, wherein the first and second reference motion vectors in the list of reference motion vectors correspond to 1/16 fractional-sample accuracy.

15. A non-transitory computer-readable medium storing instructions which when executed by a computer for video decoding cause the computer to perform:

determining first and second reference motion vectors for decoding a current block based on coding information of at least one previously decoded block;

offsetting a first component, on a first coordinate axis and in a two's complement integer representation format, of the first reference motion vector by one, and right-shifting the offset first component by one bit to obtain a shifted first component;

offsetting a second component, on the first coordinate axis and in the two's complement integer representation format, of the second reference motion vector by one, and right-shifting the offset second component by one bit to obtain a shifted second component;

generating a third component, on the first coordinate axis and in the two's complement integer representation format, of a first averaged reference motion vector by adding the shifted first component and the shifted second component;

constructing a list of reference motion vectors for decoding the current block, the list of reference motion vectors incorporating the first and second reference motion vectors and the first averaged reference motion vector;

determining a motion vector predictor using the list of reference motion vectors; and

decoding the current block for output based on the determined motion vector predictor.

16. The non-transitory computer-readable medium of claim 15, wherein the instructions when executed by the computer further cause the computer to perform:

offsetting a fourth component, on a second coordinate axis and in the two's complement integer representation format, of the first reference motion vector by one, and right-shifting the offset fourth component by one bit to obtain a shifted fourth component,

offsetting a fifth component, on the second coordinate axis and in the two's complement integer representation format, of the second reference motion vector by one, and right-shifting the offset fifth component by one bit to obtain a shifted fifth component, and

generating a sixth component, on the second coordinate axis and in the two's complement integer representation format, of the first averaged reference motion vector by adding the shifted fourth component and the shifted fifth component.

17. The non-transitory computer-readable medium of claim 15, wherein the instructions when executed by the computer further cause the computer to perform:

generating a second averaged reference motion vector based on the first reference motion vector and a third reference motion vector; and

generating a third averaged reference motion vector based on the second reference motion vector and the third reference motion vector, wherein

the constructing the list of reference motion vectors for decoding the current block includes arranging the list of reference motion vectors to include, according to an ascending order of respective indices associated thereto, the first reference motion vector, the second reference motion vector, the third reference motion vector, the first averaged reference motion vector, the second averaged reference motion vector, and the third averaged reference motion vector.

18. The non-transitory computer-readable medium of claim 17, wherein the instructions when executed by the computer further cause the computer to perform:

generating a fourth averaged reference motion vector based on the first reference motion vector and a fourth reference motion vector;

generating a fifth averaged reference motion vector based on the second reference motion vector and the fourth reference motion vector; and

generating a sixth averaged reference motion vector based on the third reference motion vector and the fourth reference motion vector, wherein

the constructing the list of reference motion vectors for decoding the current block includes arranging the list of reference motion vectors to include, according to the ascending order of respective indices associated thereto, the first reference motion vector, the second reference motion vector, the third reference motion vector, the fourth reference motion vector, the first list reference motion vectors and adding the first averaged reference motion vector, the second averaged reference motion vector, the third averaged reference motion vector, the fourth averaged reference motion vector, the fifth averaged reference motion vector, and the sixth averaged reference motion vector.

19. The non-transitory computer-readable medium of claim 15, wherein the offsetting the first component and the right-shifting the offset first component comprise:

setting a right-shift parameter as 1;

setting a left-shift parameter as 0; and

executing a rounding subroutine or operating a rounding circuit that is configured to receive the first component, the right-shift parameter, and the left-shift parameter, and to calculate the shifted first component according to:

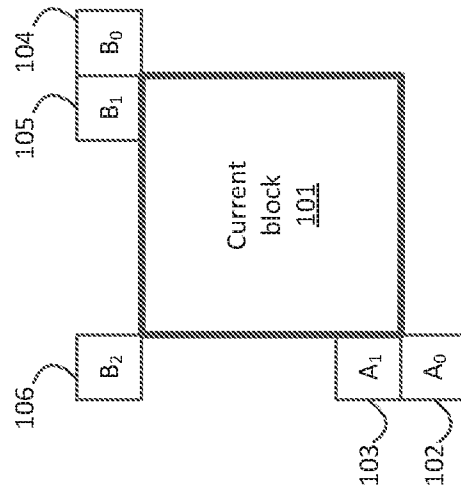
$$\text{offset} = 1 \ll (\text{rightShift} - 1), \text{ and}$$

$$y = (x \geq 0 ? (x + \text{offset}) \gg \text{rightShift} : -((-x + \text{offset}) \gg \text{rightShift})) \ll$$

leftShift, wherein

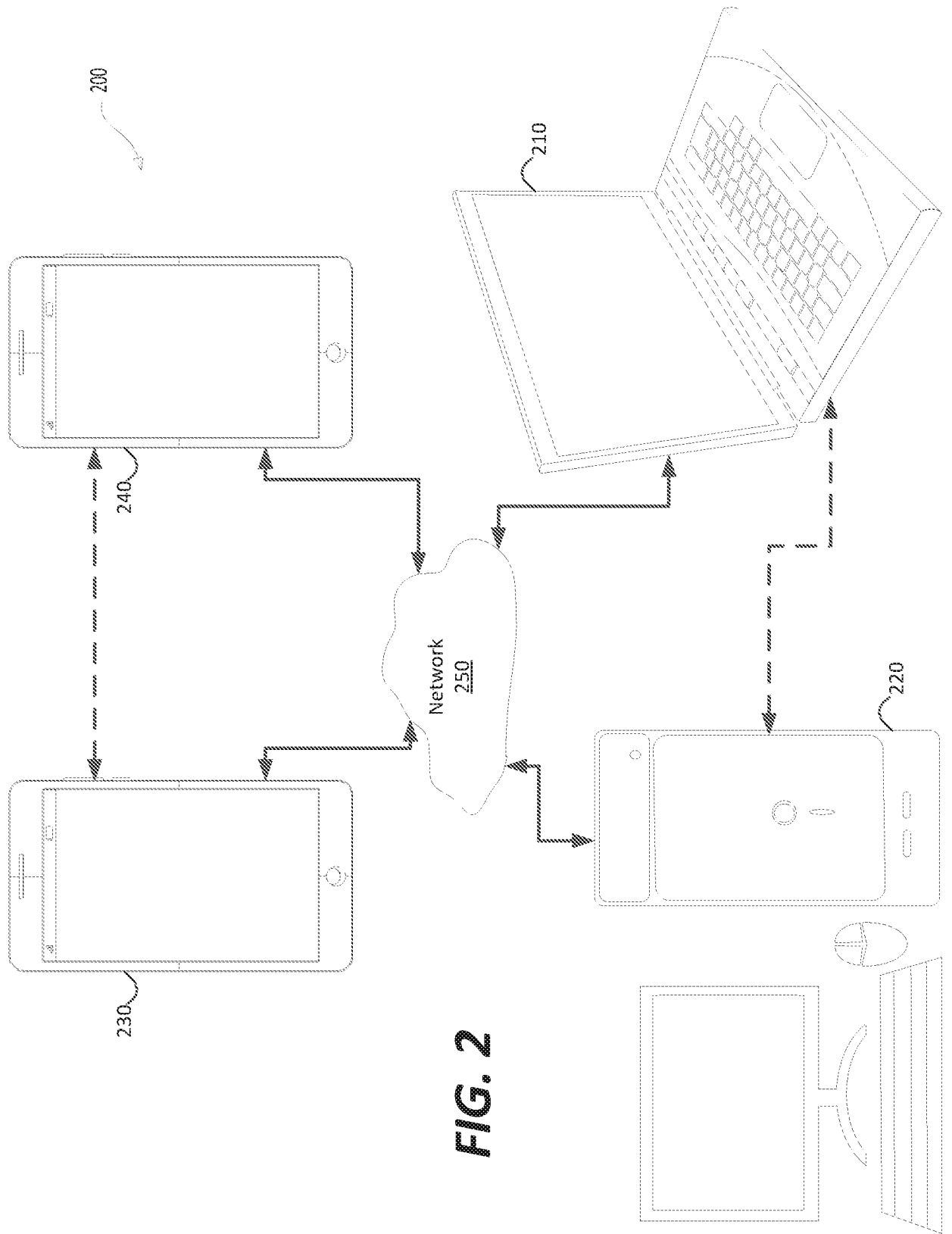
rightShift corresponds to the right-shift parameter, leftShift corresponds to the left-shift parameter, x corresponds to the first component, and y corresponds to the shifted first component.

20. The non-transitory computer-readable medium of claim 15, wherein the first and second reference motion vectors are determined based on the at least one previously decoded block that includes a spatial neighboring block and a temporal neighboring block.



Related Art

**FIG. 1**



**FIG. 2**

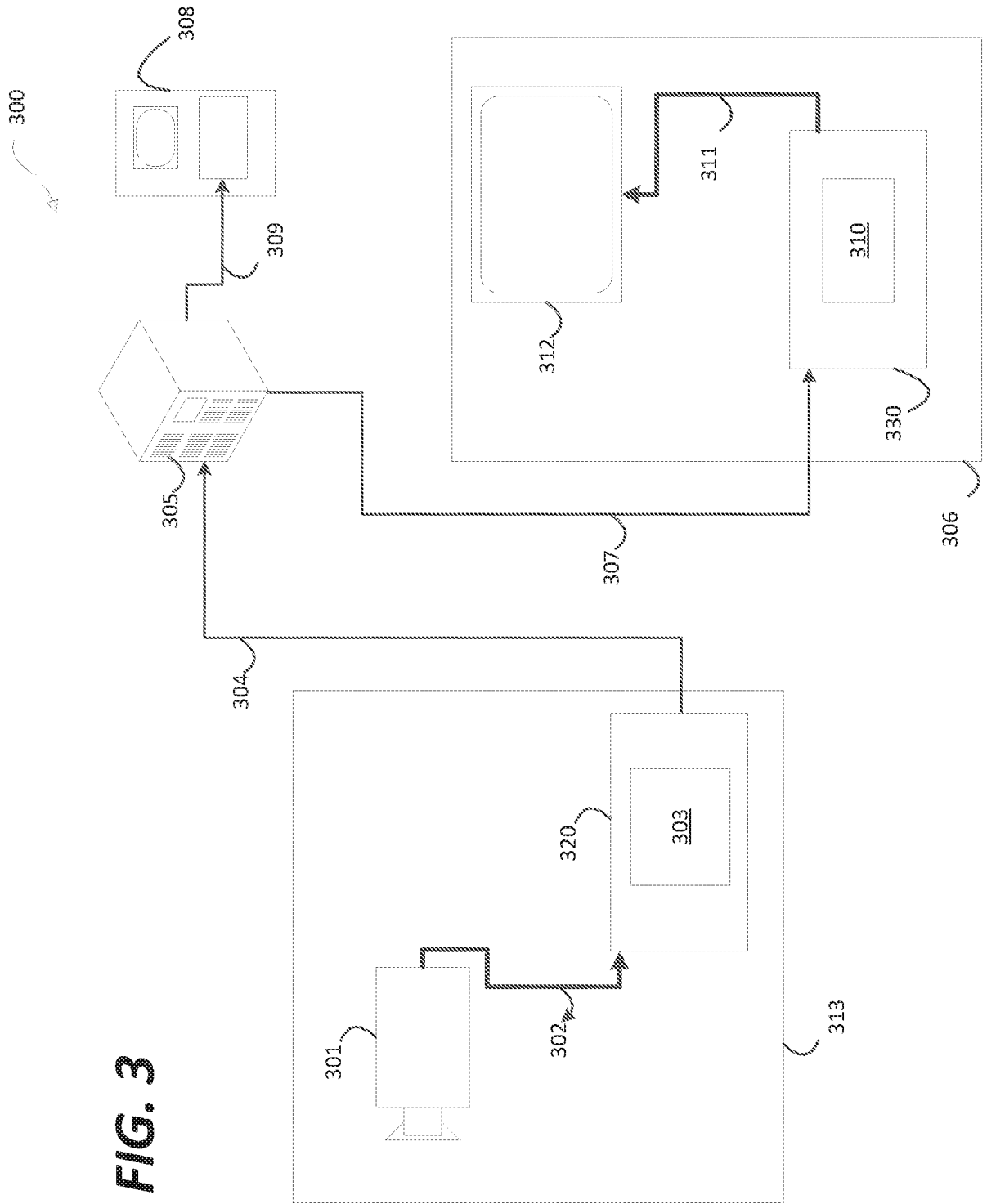
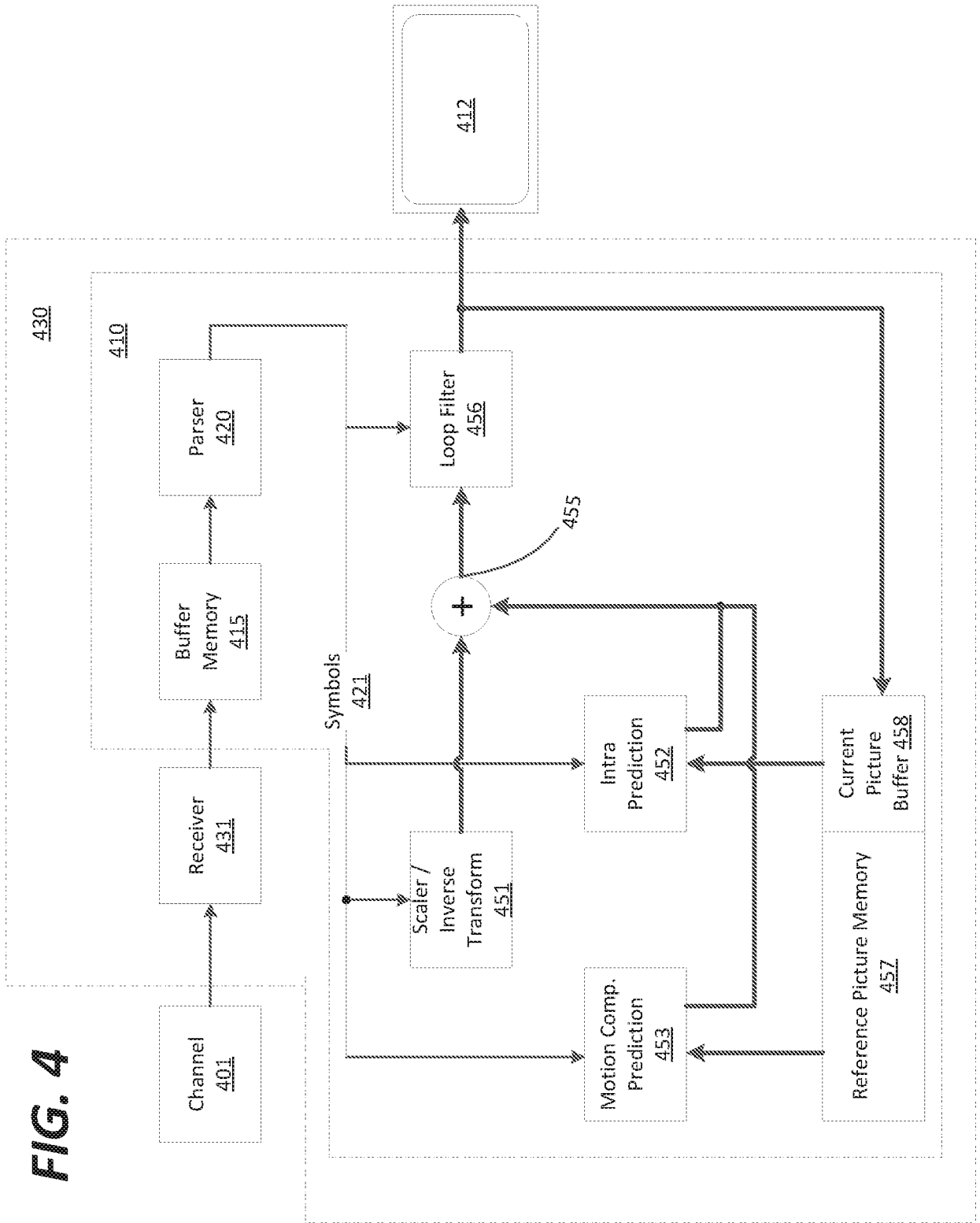


FIG. 3

**FIG. 4**



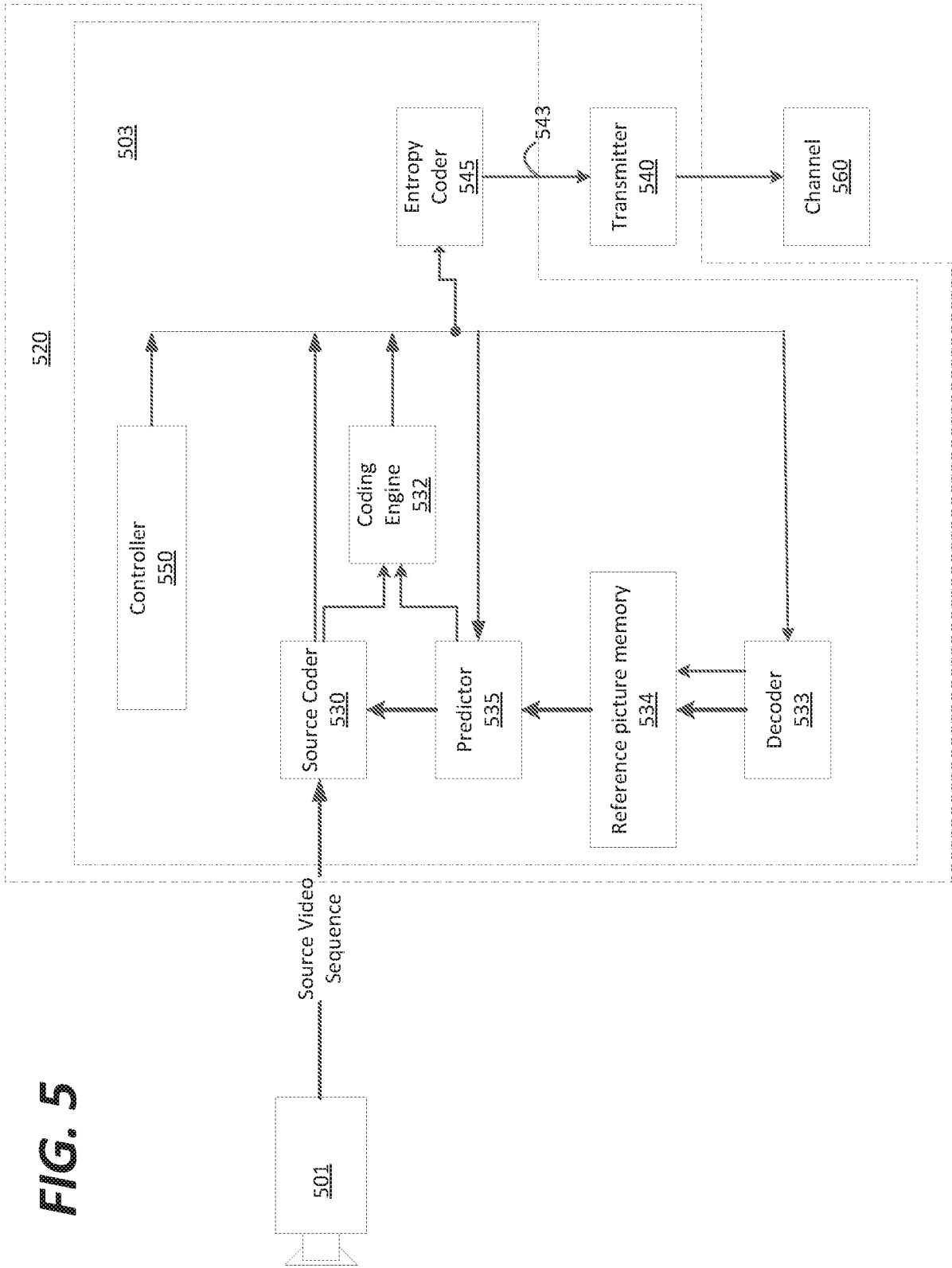
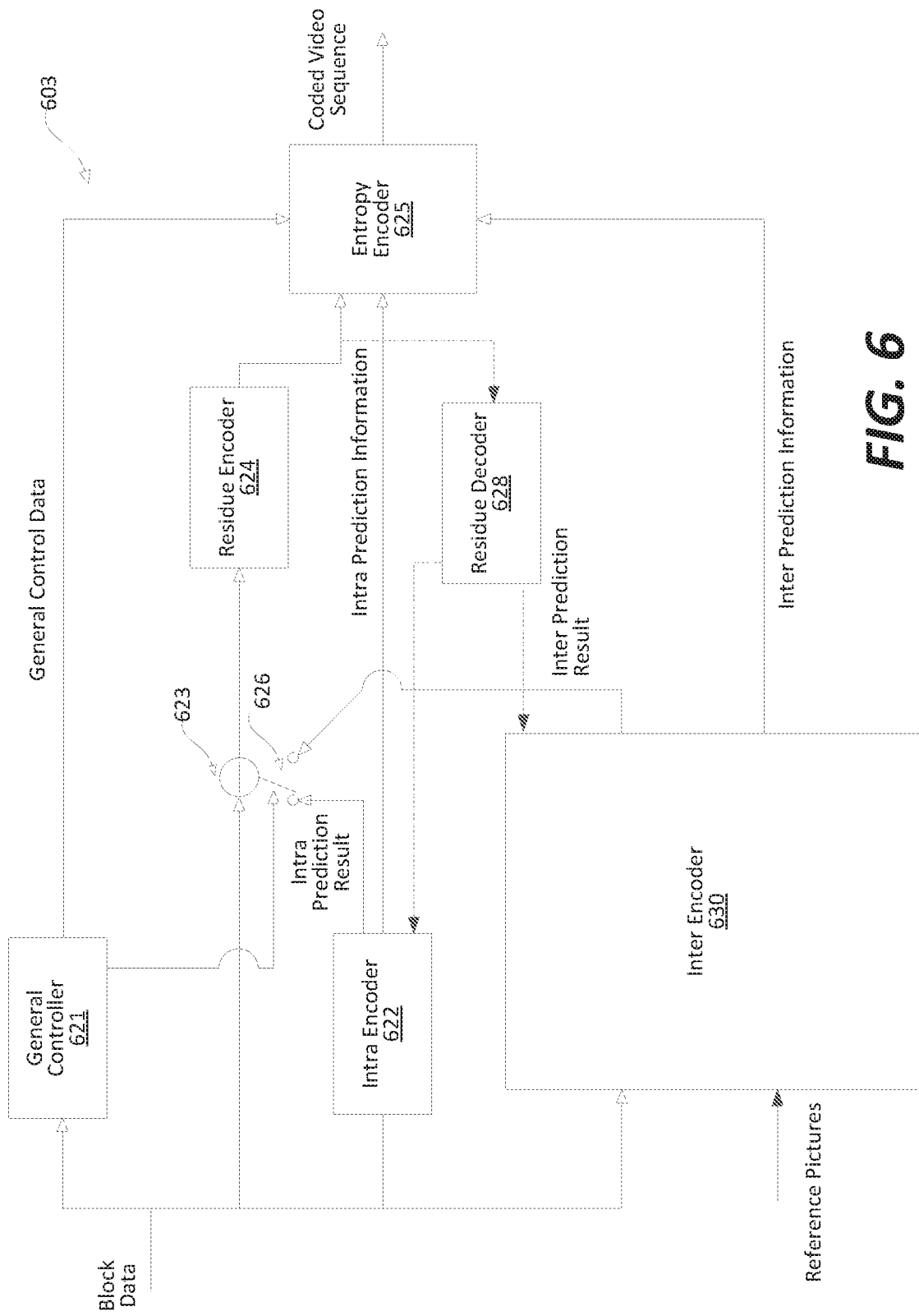


FIG. 5



**FIG. 6**

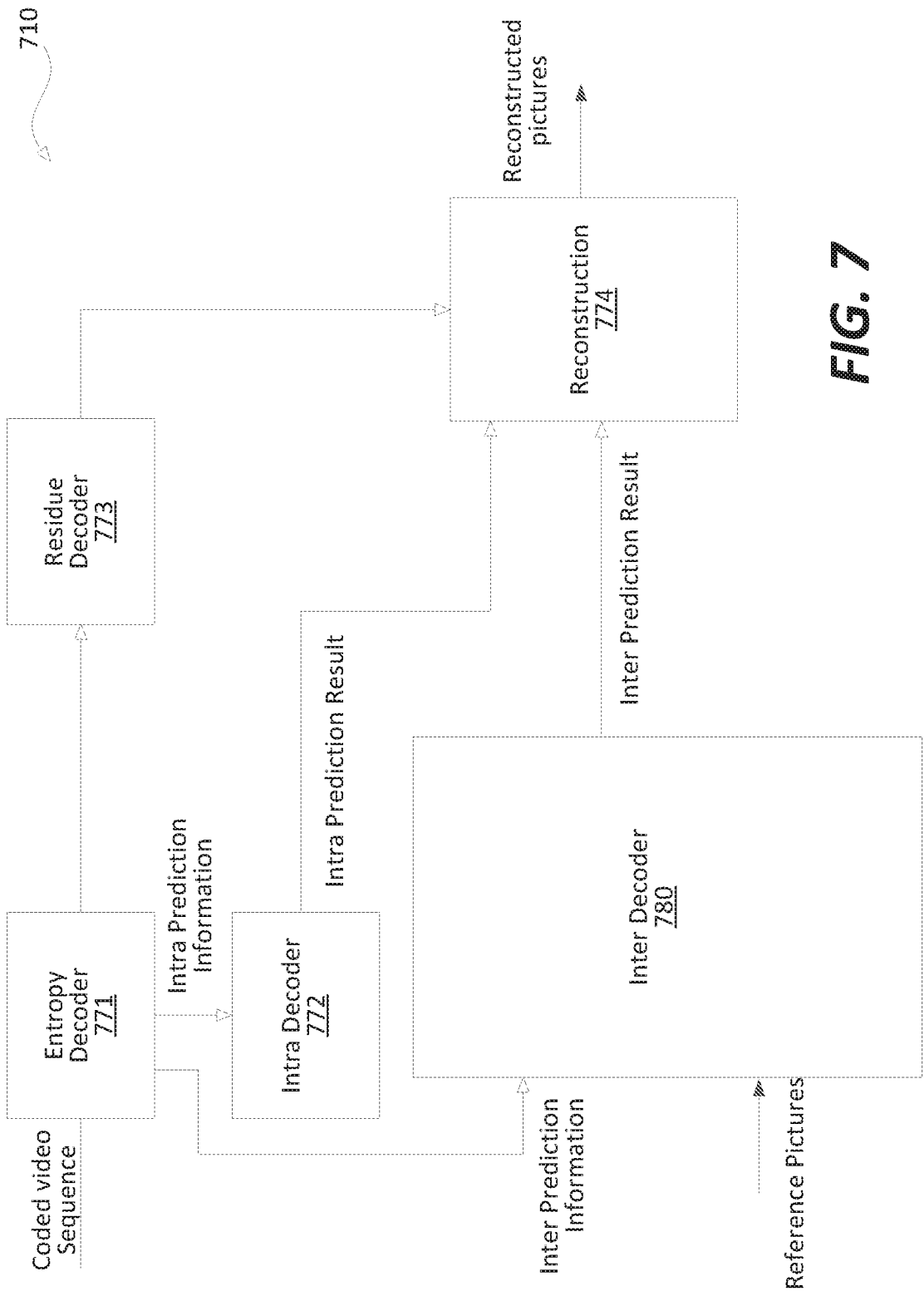


FIG. 7

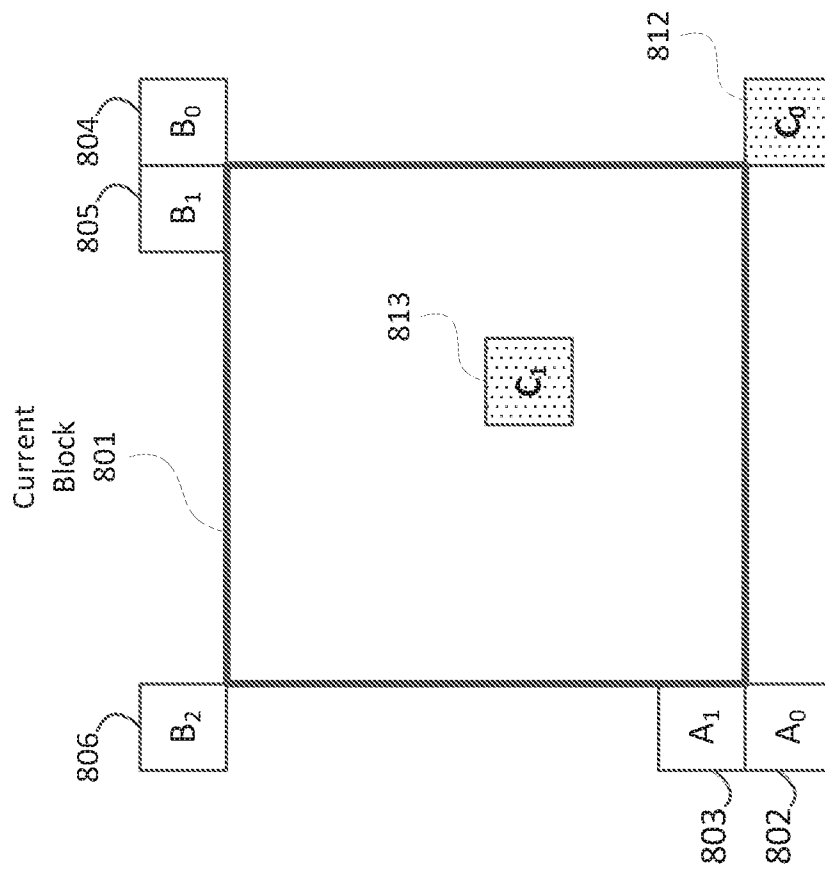
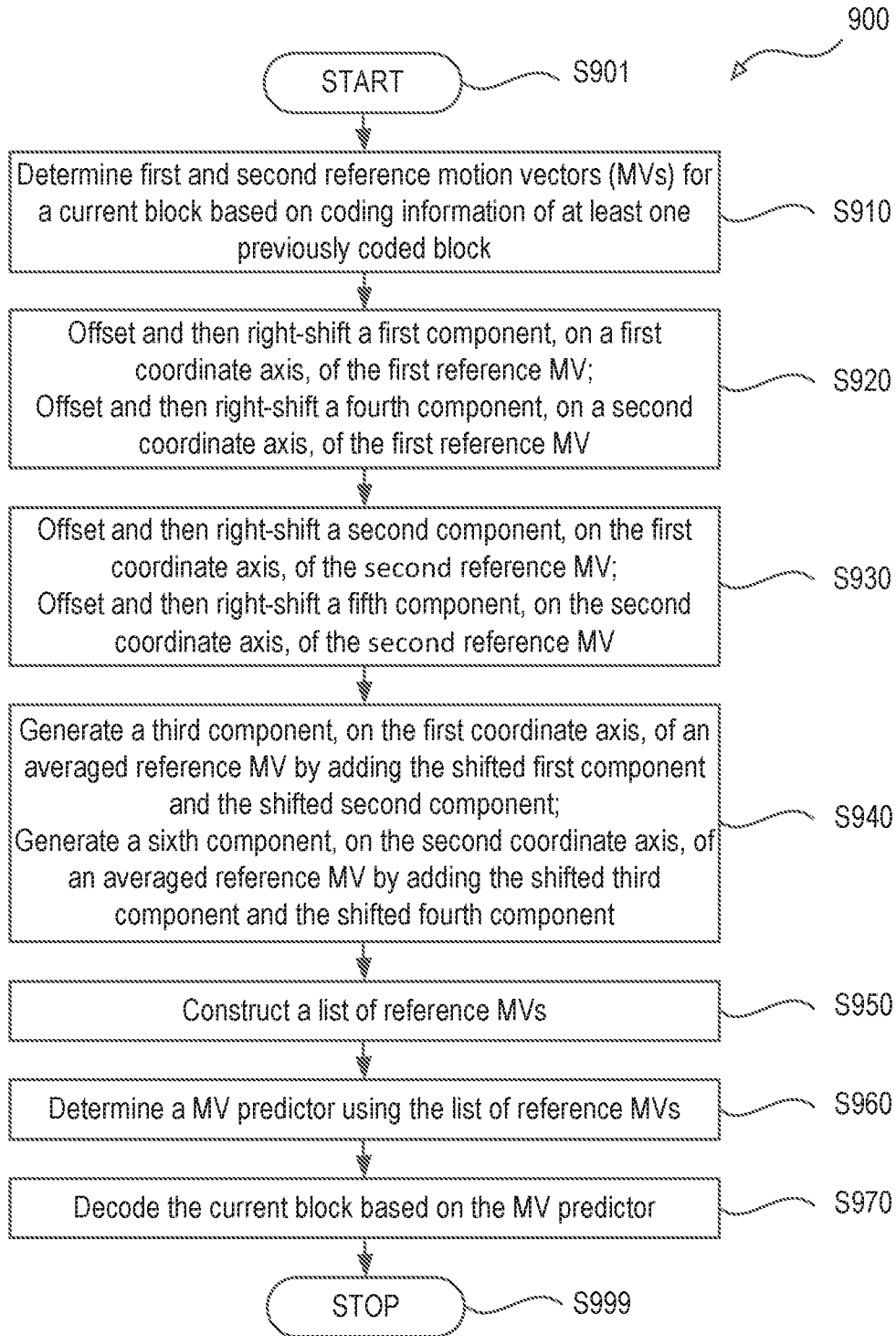
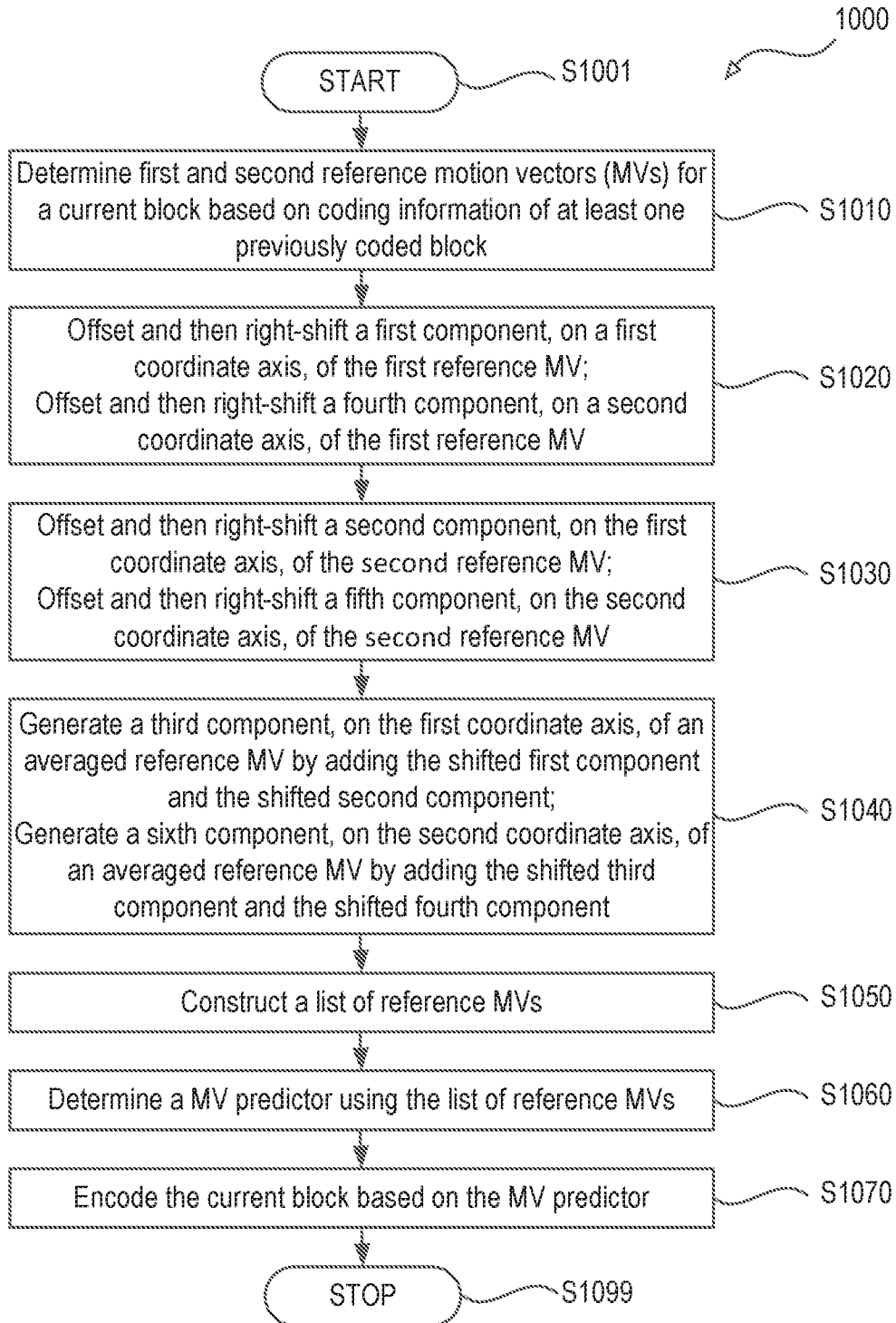


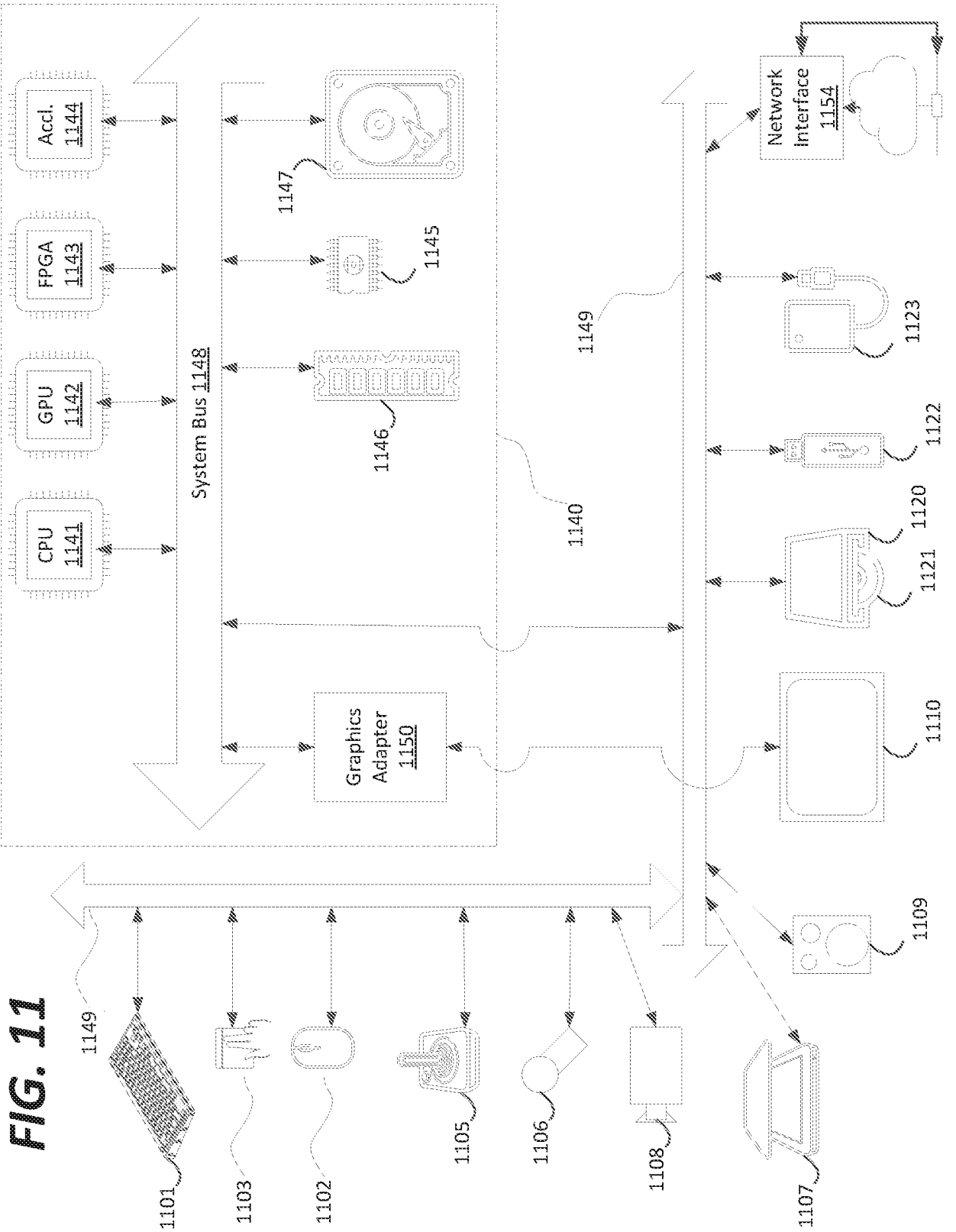
FIG. 8



**FIG. 9**



**FIG. 10**



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2019/056269

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - H04B 1/66; H04N 7/32; H04N 11/02; H04N 11/04 (2019.01)

CPC - H04N 19/176; H04N 19/51; H04N 19/513; H04N 19/577; H04N 19/61 (2019.08)

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

USPC - 375/240.16; 375/240.24; 375/240.15; 375/E7.25; 375/E7.243 (keyword delimited)

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 2010/0128792 A1 (SAITO et al) 27 May 2010 (27.05.2010) entire document	1-20
A	US 2008/0037645 A1 (JEON et al) 14 February 2008 (14.02.2008) entire document	1-20
A	US 2011/0080954 A1 (BOSSEN et al) 07 April 2011 (07.04.2011) entire document	1-20
A	US 2012/0294363 A1 (LEE et al) 22 November 2012 (22.11.2012) entire document	1-20
A	US 2007/0195885 A1 (LIANG) 23 August 2007 (23.08.2007) entire document	1-20
A	SEGALL et al. "Spatial scalability within the H. 264/AVC scalable video coding extension." In: IEEE Transactions on Circuits and Systems for Video Technology. October 2007 (10.2007) Retrieved on 02 December 2019 (02.12.2019) from < <a href="https://www.researchgate.net/profile/C_Segall/publication/3309168_Spatial_Scalability_Within_the_H264AVC_Scalable_Video_Coding_Extension/links/00b7d52e7442ac62fa000000.pdf">https://www.researchgate.net/profile/C_Segall/publication/3309168_Spatial_Scalability_Within_the_H264AVC_Scalable_Video_Coding_Extension/links/00b7d52e7442ac62fa000000.pdf</a> > 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

"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

"&amp;" document member of the same patent family

Date of the actual completion of the international search

03 December 2019

Date of mailing of the international search report

23 DEC 2019

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents  
P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300  
PCT OSP: 571-272-7774