DISALLOWING UNNECESSARY LAYERS IN MULTI-LAYER VIDEO BITSTREAMS

TECHNICAL FIELD

[0001] In general, this disclosure describes techniques for multi-layer video bitstreams in video coding. More specifically, this disclosure ensures that unnecessary and/or unused layers are disallowed within a multi-layer bitstream in video coding.

BACKGROUND

[0002] The amount of video data needed to depict even a relatively short video can be substantial, which may result in difficulties when the data is to be streamed or otherwise communicated across a communications network with limited bandwidth capacity. Thus, video data is generally compressed before being communicated across modern day telecommunications networks. The size of a video could also be an issue when the video is stored on a storage device because memory resources may be limited. Video compression devices often use software and/or hardware at the source to code the video data prior to transmission or storage, thereby decreasing the quantity of data needed to represent digital video images. The compressed data is then received at the destination by a video decompression device that decodes the video data. With limited network resources and ever increasing demands of higher video quality, improved compression and decompression techniques that improve compression ratio with little to no sacrifice in image quality are desirable.

SUMMARY

[0003] A first aspect relates to a method of decoding implemented by a video decoder, comprising receiving, by the video decoder, a video bitstream including a video parameter set (VPS) and a plurality of layers, wherein no layer is neither an output layer of at least one output layer sets (OLS) nor a direct reference layer of any other layer; and decoding, by the video decoder, a picture from one of the plurality of layers.

[0004] The method provides techniques that disallow unnecessary layers in multi-layer video bitstreams. That is, every layer included in an output layer set (OLS) is either an output layer or used as a direct or indirect reference layer of an output layer. This avoids having extraneous information in the coding process and increases coding efficiency. Thus, the coder / decoder

(a.k.a., “codec”) in video coding is improved relative to current codecs. As a practical matter, the improved video coding process offers the user a better user experience when videos are sent, received, and/or viewed.

[0005] Optionally, in any of the preceding aspects, another implementation of the aspect provides that the picture is included in an output layer of the at least one OLS.

[0006] Optionally, in any of the preceding aspects, another implementation of the aspect provides selecting an output layer from the at least one OLS prior to the decoding.

[0007] Optionally, in any of the preceding aspects, another implementation of the aspect provides selecting the picture from the output layer after the output layer has been selected.

[0008] Optionally, in any of the preceding aspects, another implementation of the aspect provides that each layer in the plurality of layers comprises a set of video coding layer (VCL) network abstraction layer (NAL) units that all have a particular value of a layer identifier (ID) and associated non-VCL NAL units.

[0009] Optionally, in any of the preceding aspects, another implementation of the aspect provides that each layer of the at least one OLS comprises the output layer or the direct reference layer of the any other layer.

[0010] Optionally, in any of the preceding aspects, another implementation of the aspect provides that at least one OLS contains one or more output layers.

[0011] Optionally, in any of the preceding aspects, another implementation of the aspect provides that the VPS includes a layer used as a reference flag and a layer used as an output layer flag, and wherein a value of the layer used as a reference flag and a value of the layer used as an output layer flag are not both zero.

[0012] Optionally, in any of the preceding aspects, another implementation of the aspect provides displaying the decoded picture on a display of an electronic device.

[0013] Optionally, in any of the preceding aspects, another implementation of the aspect provides receiving, by the video decoder, a second video bitstream including a second video parameter set (VPS) and a second plurality of layers, wherein at least one layer is neither an output layer of at least one output layer sets (OLS) nor a direct reference layer of any other layer; and in response to the receiving, taking some other corrective measures to ensure that a conforming bitstream corresponding to the second video bitstream is received prior to decoding a picture from one of the second plurality of layers.

[0014] A second aspect relates to a method of encoding implemented by a video encoder, the method comprising generating, by the video encoder, a plurality of layers and a video parameter set (VPS) specifying at least one output layer set (OLS), wherein the video encoder is constrained such that no layer is neither an output layer of the at least one OLS nor a direct reference layer of any other layer; encoding, by the video encoder, the plurality of layers and the VPS into a video bitstream; and storing, by the video encoder, the video bitstream for communication toward a video decoder.

[0015] The method provides techniques that disallow unnecessary layers in multi-layer video bitstreams. That is, every layer included in an output layer set (OLS) is either an output layer or used as a direct or indirect reference layer of an output layer. This avoids having extraneous information in the coding process and increases coding efficiency. Thus, the coder / decoder (a.k.a., "codec") in video coding is improved relative to current codecs. As a practical matter, the improved video coding process offers the user a better user experience when videos are sent, received, and/or viewed.

[0016] Optionally, in any of the preceding aspects, another implementation of the aspect provides that each of the one or more OLSs contains one or more output layers, and wherein each of the output layers comprises one or more pictures.

[0017] Optionally, in any of the preceding aspects, another implementation of the aspect provides that each layer in the plurality of layers comprises a set of video coding layer (VCL) network abstraction layer (NAL) units that all have a particular value of a layer identifier (ID) and associated non-VCL NAL units.

[0018] Optionally, in any of the preceding aspects, another implementation of the aspect provides that one of the OLSs comprises two output layers, and wherein one of the two output layers references another of the two output layers.

[0019] Optionally, in any of the preceding aspects, another implementation of the aspect provides that the VPS includes a layer used as a reference flag and a layer used as an output layer flag, and wherein a value of the layer used as a reference flag and a value of the layer used as an output layer flag are not both zero.

[0020] A third aspect relates to decoding device, comprising a receiver configured to receive a video bitstream including a video parameter set (VPS) and a plurality of layers, wherein no layer is neither an output layer of at least one OLS nor a direct reference layer of any other layer; a memory coupled to the receiver, the memory storing instructions; and a processor coupled to the

memory, the processor configured to execute the instructions to cause the decoding device to decode a picture from one of the plurality of layers to obtain a decoded picture.

[0021] The decoding device provides techniques that disallow unnecessary layers in multi-layer video bitstreams. That is, every layer included in an output layer set (OLS) is either an output layer or used as a direct or indirect reference layer of an output layer. This avoids having extraneous information in the coding process and increases coding efficiency. Thus, the coder / decoder (a.k.a., "codec") in video coding is improved relative to current codecs. As a practical matter, the improved video coding process offers the user a better user experience when videos are sent, received, and/or viewed.

[0022] Optionally, in any of the preceding aspects, another implementation of the aspect provides that the picture is included in an output layer of the at least one OLS.

[0023] Optionally, in any of the preceding aspects, another implementation of the aspect provides that the processor is further configured to select an output layer from the at least one OLS prior to the picture being decoded.

[0024] Optionally, in any of the preceding aspects, another implementation of the aspect provides that the processor is further configured to select the picture from the output layer after the output layer has been selected.

[0025] Optionally, in any of the preceding aspects, another implementation of the aspect provides that each layer in the plurality of layers comprises a set of video coding layer (VCL) network abstraction layer (NAL) units that all have a particular value of a layer identifier (ID) and associated non-VCL NAL units.

[0026] Optionally, in any of the preceding aspects, another implementation of the aspect provides that each layer of the at least one OLS comprises the output layer or the direct reference layer of the any other layer.

[0027] Optionally, in any of the preceding aspects, another implementation of the aspect provides that the at least one OLS contains one or more output layers.

[0028] Optionally, in any of the preceding aspects, another implementation of the aspect provides that a display configured to display the decoded picture.

[0029] Optionally, in any of the preceding aspects, another implementation of the aspect provides that the processor is configured to execute the instructions to further cause the decoding device receive a second video bitstream including a second video parameter set (VPS) and a second plurality of layers, wherein at least one layer is neither an output layer of at least one

output layer sets (OLS) nor a direct reference layer of any other layer; and in response to the receipt of the second video bitstream, taking some other corrective measures to ensure that a conforming bitstream corresponding to the second video bitstream is received prior to decoding a picture from one of the second plurality of layers.

[0030] A fourth aspect relates to an encoding device, comprising a memory containing instructions; a processor coupled to the memory, the processor configured to implement the instructions to cause the encoding device to generate a plurality of layers and a video parameter set (VPS) specifying one or more output layer sets (OLSs), wherein the encoding device is constrained such that no layer is neither an output layer of at least one OLS nor a direct reference layer of any other layer; encode the plurality of layers and the VPS into a video bitstream; and a transmitter coupled to the processor, the transmitter configured to transmit the video bitstream toward a video decoder.

[0031] The encoding device provides techniques that disallow unnecessary layers in multi-layer video bitstreams. That is, every layer included in an output layer set (OLS) is either an output layer or used as a direct or indirect reference layer of an output layer. This avoids having extraneous information in the coding process and increases coding efficiency. Thus, the coder / decoder (a.k.a., “codec”) in video coding is improved relative to current codecs. As a practical matter, the improved video coding process offers the user a better user experience when videos are sent, received, and/or viewed.

[0032] Optionally, in any of the preceding aspects, another implementation of the aspect provides that each of the one or more OLSs contains one or more output layers, and wherein each of the output layers comprises one or more pictures.

[0033] Optionally, in any of the preceding aspects, another implementation of the aspect provides that each layer in the plurality of layers comprises a set of video coding layer (VCL) network abstraction layer (NAL) units that all have a particular value of a layer identifier (ID) and associated non-VCL NAL units.

[0034] Optionally, in any of the preceding aspects, another implementation of the aspect provides that each layer of the at least one OLS comprises the output layer or the direct reference layer of the any other layer.

[0035] Optionally, in any of the preceding aspects, another implementation of the aspect provides that the VPS includes a layer used as a reference flag and a layer used as an output layer

flag, and wherein a value of the layer used as a reference flag and a value of the layer used as an output layer flag are not both zero.

[0036] A fifth aspect relates to a coding apparatus. The coding apparatus includes a receiver configured to receive a picture to encode or to receive a bitstream to decode; a transmitter coupled to the receiver, the transmitter configured to transmit the bitstream to a decoder or to transmit a decoded image to a display; a memory coupled to at least one of the receiver or the transmitter, the memory configured to store instructions; and a processor coupled to the memory, the processor configured to execute the instructions stored in the memory to perform any of the methods disclosed herein.

[0037] The coding apparatus provides techniques that disallow unnecessary layers in multi-layer video bitstreams. That is, every layer included in an output layer set (OLS) is either an output layer or used as a direct or indirect reference layer of an output layer. This avoids having extraneous information in the coding process and increases coding efficiency. Thus, the coder / decoder (a.k.a., "codec") in video coding is improved relative to current codecs. As a practical matter, the improved video coding process offers the user a better user experience when videos are sent, received, and/or viewed.

[0038] Optionally, in any of the preceding aspects, another implementation of the aspect provides a display configured to display a decoded picture.

[0039] A sixth aspect relates to a system. The system includes an encoder; and a decoder in communication with the encoder, wherein the encoder or the decoder includes the decoding device, the encoding device, or the coding apparatus disclosed herein.

[0040] The system provides techniques that disallow unnecessary layers in multi-layer video bitstreams. That is, every layer included in an output layer set (OLS) is either an output layer or used as a direct or indirect reference layer of an output layer. This avoids having extraneous information in the coding process and increases coding efficiency. Thus, the coder / decoder (a.k.a., "codec") in video coding is improved relative to current codecs. As a practical matter, the improved video coding process offers the user a better user experience when videos are sent, received, and/or viewed.

[0041] A seventh aspect relates to a means for coding. The means for coding includes receiving means configured to receive a picture to encode or to receive a bitstream to decode; transmission means coupled to the receiving means, the transmission means configured to transmit the bitstream to a decoding means or to transmit a decoded image to a display means; storage means

coupled to at least one of the receiving means or the transmission means, the storage means configured to store instructions; and processing means coupled to the storage means, the processing means configured to execute the instructions stored in the storage means to perform any of the methods disclosed herein.

[0042] The means for coding provides techniques that disallow unnecessary layers in multi-layer video bitstreams. That is, every layer included in an output layer set (OLS) is either an output layer or used as a direct or indirect reference layer of an output layer. This avoids having extraneous information in the coding process and increases coding efficiency. Thus, the coder / decoder (a.k.a., "codec") in video coding is improved relative to current codecs. As a practical matter, the improved video coding process offers the user a better user experience when videos are sent, received, and/or viewed.

[0043] For the purpose of clarity, any one of the foregoing embodiments may be combined with any one or more of the other foregoing embodiments to create a new embodiment within the scope of the present disclosure.

[0044] These and other features will be more clearly understood from the following detailed description taken in conjunction with the accompanying drawings and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0045] For a more complete understanding of this disclosure, reference is now made to the following brief description, taken in connection with the accompanying drawings and detailed description, wherein like reference numerals represent like parts.

[0046] FIG. 1 is a flowchart of an example method of coding a video signal.

[0047] FIG. 2 is a schematic diagram of an example coding and decoding (codec) system for video coding.

[0048] FIG. 3 is a schematic diagram illustrating an example video encoder.

[0049] FIG. 4 is a schematic diagram illustrating an example video decoder.

[0050] FIG. 5 illustrates an example of multi-layer coding for spatial scalability.

[0051] FIG. 6 illustrates an example of multi-layer coding using output layer sets (OLSs).

[0052] FIG. 7 illustrates an embodiment of a video bitstream.

[0053] FIG. 8 is an embodiment of a method of decoding a coded video bitstream.

[0054] FIG. 9 is an embodiment of a method of encoding a coded video bitstream.

[0055] FIG. 10 is a schematic diagram of a video coding device.

[0056] FIG. 11 is a schematic diagram of an embodiment of a means for coding.

DETAILED DESCRIPTION

[0057] It should be understood at the outset that although an illustrative implementation of one or more embodiments are provided below, the disclosed systems and/or methods may be implemented using any number of techniques, whether currently known or in existence. The disclosure should in no way be limited to the illustrative implementations, drawings, and techniques illustrated below, including the exemplary designs and implementations illustrated and described herein, but may be modified within the scope of the appended claims along with their full scope of equivalents.

[0058] The following terms are defined as follows unless used in a contrary context herein. Specifically, the following definitions are intended to provide additional clarity to the present disclosure. However, terms may be described differently in different contexts. Accordingly, the following definitions should be considered as a supplement and should not be considered to limit any other definitions of descriptions provided for such terms herein.

[0059] A bitstream is a sequence of bits including video data that is compressed for transmission between an encoder and a decoder. An encoder is a device that is configured to employ encoding processes to compress video data into a bitstream. A decoder is a device that is configured to employ decoding processes to reconstruct video data from a bitstream for display. A picture is an array of luma samples and/or an array of chroma samples that create a frame or a field thereof. A picture that is being encoded or decoded can be referred to as a current picture for clarity of discussion. A reference picture is a picture that contains reference samples that can be used when coding other pictures by reference according to inter-prediction and/or inter-layer prediction. A reference picture list is a list of reference pictures used for inter-prediction and/or inter-layer prediction. Some video coding systems utilize two reference picture lists, which can be denoted as reference picture list one and reference picture list zero. A reference picture list structure is an addressable syntax structure that contains multiple reference picture lists. Inter-prediction is a mechanism of coding samples of a current picture by reference to indicated samples in a reference picture that is different from the current picture where the reference picture and the current picture are in the same layer. A reference picture list structure entry is an addressable location in a reference picture list structure that indicates a reference picture associated with a reference picture list. A slice header is a part of a coded slice containing data

elements pertaining to all video data within a tile represented in the slice. A picture parameter set (PPS) is a parameter set that contains data related to an entire picture. More specifically, the PPS is a syntax structure containing syntax elements that apply to zero or more entire coded pictures as determined by a syntax element found in each picture header. A sequence parameter set (SPS) is a parameter set that contains data related to a sequence of pictures. An access unit (AU) is a set of one or more coded pictures associated with the same display time (e.g., the same picture order count) for output from a decoded picture buffer (DPB) (e.g., for display to a user). An access unit delimiter (AUD) is an indicator or data structure used to indicate the start of an AU or the boundary between AUs. A decoded video sequence is a sequence of pictures that have been reconstructed by a decoder in preparation for display to a user.

[0060] A network abstraction layer (NAL) unit is a syntax structure containing data in the form of a Raw Byte Sequence Payload (RBSP), an indication of the type of data, and interspersed as desired with emulation prevention bytes. A video coding layer (VCL) NAL unit is a NAL unit coded to contain video data, such as a coded slice of a picture. A non-VCL NAL unit is a NAL unit that contains non-video data such as syntax and/or parameters that support decoding the video data, performance of conformance checking, or other operations. A layer is a set of VCL NAL units that share a specified characteristic (e.g., a common resolution, frame rate, image size, etc.) and associated non-VCL NAL units. The VCL NAL units of a layer may share a particular value of a NAL unit header layer identifier (nuh_layer_id). A coded picture is a coded representation of a picture comprising VCL NAL units with a particular value of a NAL unit header layer identifier (nuh_layer_id) within an access unit (AU) and containing all coding tree units (CTUs) of the picture. A decoded picture is a picture produced by applying a decoding process to a coded picture.

[0061] An output layer set (OLS) is a set of layers for which one or more layers are specified as output layer(s). An output layer is a layer that is designated for output (e.g., to a display). A zeroth (0-th) OLS is an OLS that contains only a lowest layer (layer with a lowest layer identifier) and hence contains only an output layer. A video parameter set (VPS) is a data unit that contains parameters related to an entire video. Inter-layer prediction is a mechanism of coding a current picture in a current layer by reference to a reference picture in a reference layer, where the current picture and the reference picture are included in the same AU and the reference layer includes a lower nuh_layer_id than the current layer.

[0062] The following acronyms are used herein, Coding Tree Block (CTB), Coding Tree Unit (CTU), Coding Unit (CU), Coded Video Sequence (CVS), Joint Video Experts Team (JVET), Motion-Constrained Tile Set (MCTS), Maximum Transfer Unit (MTU), Network Abstraction Layer (NAL), Picture Order Count (POC), Picture Parameter Set (PPS), Raw Byte Sequence Payload (RBSP), Sequence Parameter Set (SPS), Versatile Video Coding (VVC), and Working Draft (WD).

[0063] FIG. 1 is a flowchart of an example operating method 100 of coding a video signal. Specifically, a video signal is encoded at an encoder. The encoding process compresses the video signal by employing various mechanisms to reduce the video file size. A smaller file size allows the compressed video file to be transmitted toward a user, while reducing associated bandwidth overhead. The decoder then decodes the compressed video file to reconstruct the original video signal for display to an end user. The decoding process generally mirrors the encoding process to allow the decoder to consistently reconstruct the video signal.

[0064] At step 101, the video signal is input into the encoder. For example, the video signal may be an uncompressed video file stored in memory. As another example, the video file may be captured by a video capture device, such as a video camera, and encoded to support live streaming of the video. The video file may include both an audio component and a video component. The video component contains a series of image frames that, when viewed in a sequence, gives the visual impression of motion. The frames contain pixels that are expressed in terms of light, referred to herein as luma components (or luma samples), and color, which is referred to as chroma components (or color samples). In some examples, the frames may also contain depth values to support three dimensional viewing.

[0065] At step 103, the video is partitioned into blocks. Partitioning includes subdividing the pixels in each frame into square and/or rectangular blocks for compression. For example, in High Efficiency Video Coding (HEVC) (also known as H.265 and MPEG-H Part 2) the frame can first be divided into coding tree units (CTUs), which are blocks of a predefined size (e.g., sixty-four pixels by sixty-four pixels). The CTUs contain both luma and chroma samples. Coding trees may be employed to divide the CTUs into blocks and then recursively subdivide the blocks until configurations are achieved that support further encoding. For example, luma components of a frame may be subdivided until the individual blocks contain relatively homogenous lighting values. Further, chroma components of a frame may be subdivided until

the individual blocks contain relatively homogenous color values. Accordingly, partitioning mechanisms vary depending on the content of the video frames.

[0066] At step 105, various compression mechanisms are employed to compress the image blocks partitioned at step 103. For example, inter-prediction and/or intra-prediction may be employed. Inter-prediction is designed to take advantage of the fact that objects in a common scene tend to appear in successive frames. Accordingly, a block depicting an object in a reference frame need not be repeatedly described in adjacent frames. Specifically, an object, such as a table, may remain in a constant position over multiple frames. Hence the table is described once and adjacent frames can refer back to the reference frame. Pattern matching mechanisms may be employed to match objects over multiple frames. Further, moving objects may be represented across multiple frames, for example due to object movement or camera movement. As a particular example, a video may show an automobile that moves across the screen over multiple frames. Motion vectors can be employed to describe such movement. A motion vector is a two-dimensional vector that provides an offset from the coordinates of an object in a frame to the coordinates of the object in a reference frame. As such, inter-prediction can encode an image block in a current frame as a set of motion vectors indicating an offset from a corresponding block in a reference frame.

[0067] Intra-prediction encodes blocks in a common frame. Intra-prediction takes advantage of the fact that luma and chroma components tend to cluster in a frame. For example, a patch of green in a portion of a tree tends to be positioned adjacent to similar patches of green. Intra-prediction employs multiple directional prediction modes (e.g., thirty-three in HEVC), a planar mode, and a direct current (DC) mode. The directional modes indicate that a current block is similar/the same as samples of a neighbor block in a corresponding direction. Planar mode indicates that a series of blocks along a row/column (e.g., a plane) can be interpolated based on neighbor blocks at the edges of the row. Planar mode, in effect, indicates a smooth transition of light/color across a row/column by employing a relatively constant slope in changing values. DC mode is employed for boundary smoothing and indicates that a block is similar/the same as an average value associated with samples of all the neighbor blocks associated with the angular directions of the directional prediction modes. Accordingly, intra-prediction blocks can represent image blocks as various relational prediction mode values instead of the actual values. Further, inter-prediction blocks can represent image blocks as motion vector values instead of the actual values. In either case, the prediction blocks may not exactly represent the image blocks in some

cases. Any differences are stored in residual blocks. Transforms may be applied to the residual blocks to further compress the file.

[0068] At step 107, various filtering techniques may be applied. In HEVC, the filters are applied according to an in-loop filtering scheme. The block based prediction discussed above may result in the creation of blocky images at the decoder. Further, the block based prediction scheme may encode a block and then reconstruct the encoded block for later use as a reference block. The in-loop filtering scheme iteratively applies noise suppression filters, de-blocking filters, adaptive loop filters, and sample adaptive offset (SAO) filters to the blocks/frames. These filters mitigate such blocking artifacts so that the encoded file can be accurately reconstructed. Further, these filters mitigate artifacts in the reconstructed reference blocks so that artifacts are less likely to create additional artifacts in subsequent blocks that are encoded based on the reconstructed reference blocks.

[0069] Once the video signal has been partitioned, compressed, and filtered, the resulting data is encoded in a bitstream at step 109. The bitstream includes the data discussed above as well as any signaling data desired to support proper video signal reconstruction at the decoder. For example, such data may include partition data, prediction data, residual blocks, and various flags providing coding instructions to the decoder. The bitstream may be stored in memory for transmission toward a decoder upon request. The bitstream may also be broadcast and/or multicast toward a plurality of decoders. The creation of the bitstream is an iterative process. Accordingly, steps 101, 103, 105, 107, and 109 may occur continuously and/or simultaneously over many frames and blocks. The order shown in FIG. 1 is presented for clarity and ease of discussion, and is not intended to limit the video coding process to a particular order.

[0070] The decoder receives the bitstream and begins the decoding process at step 111. Specifically, the decoder employs an entropy decoding scheme to convert the bitstream into corresponding syntax and video data. The decoder employs the syntax data from the bitstream to determine the partitions for the frames at step 111. The partitioning should match the results of block partitioning at step 103. Entropy encoding/decoding as employed in step 111 is now described. The encoder makes many choices during the compression process, such as selecting block partitioning schemes from several possible choices based on the spatial positioning of values in the input image(s). Signaling the exact choices may employ a large number of bins. As used herein, a bin is a binary value that is treated as a variable (e.g., a bit value that may vary depending on context). Entropy coding allows the encoder to discard any options that are clearly

not viable for a particular case, leaving a set of allowable options. Each allowable option is then assigned a code word. The length of the code words is based on the number of allowable options (e.g., one bin for two options, two bins for three to four options, etc.) The encoder then encodes the code word for the selected option. This scheme reduces the size of the code words as the code words are as big as desired to uniquely indicate a selection from a small sub-set of allowable options as opposed to uniquely indicating the selection from a potentially large set of all possible options. The decoder then decodes the selection by determining the set of allowable options in a similar manner to the encoder. By determining the set of allowable options, the decoder can read the code word and determine the selection made by the encoder.

[0071] At step 113, the decoder performs block decoding. Specifically, the decoder employs reverse transforms to generate residual blocks. Then the decoder employs the residual blocks and corresponding prediction blocks to reconstruct the image blocks according to the partitioning. The prediction blocks may include both intra-prediction blocks and inter-prediction blocks as generated at the encoder at step 105. The reconstructed image blocks are then positioned into frames of a reconstructed video signal according to the partitioning data determined at step 111. Syntax for step 113 may also be signaled in the bitstream via entropy coding as discussed above.

[0072] At step 115, filtering is performed on the frames of the reconstructed video signal in a manner similar to step 107 at the encoder. For example, noise suppression filters, de-blocking filters, adaptive loop filters, and SAO filters may be applied to the frames to remove blocking artifacts. Once the frames are filtered, the video signal can be output to a display at step 117 for viewing by an end user.

[0073] FIG. 2 is a schematic diagram of an example coding and decoding (codec) system 200 for video coding. Specifically, codec system 200 provides functionality to support the implementation of operating method 100. Codec system 200 is generalized to depict components employed in both an encoder and a decoder. Codec system 200 receives and partitions a video signal as discussed with respect to steps 101 and 103 in operating method 100, which results in a partitioned video signal 201. Codec system 200 then compresses the partitioned video signal 201 into a coded bitstream when acting as an encoder as discussed with respect to steps 105, 107, and 109 in method 100. When acting as a decoder, codec system 200 generates an output video signal from the bitstream as discussed with respect to steps 111, 113, 115, and 117 in operating method 100. The codec system 200 includes a general coder control

component 211, a transform scaling and quantization component 213, an intra-picture estimation component 215, an intra-picture prediction component 217, a motion compensation component 219, a motion estimation component 221, a scaling and inverse transform component 229, a filter control analysis component 227, an in-loop filters component 225, a decoded picture buffer component 223, and a header formatting and context adaptive binary arithmetic coding (CABAC) component 231. Such components are coupled as shown. In FIG. 2, black lines indicate movement of data to be encoded/decoded while dashed lines indicate movement of control data that controls the operation of other components. The components of codec system 200 may all be present in the encoder. The decoder may include a subset of the components of codec system 200. For example, the decoder may include the intra-picture prediction component 217, the motion compensation component 219, the scaling and inverse transform component 229, the in-loop filters component 225, and the decoded picture buffer component 223. These components are now described.

[0074] The partitioned video signal 201 is a captured video sequence that has been partitioned into blocks of pixels by a coding tree. A coding tree employs various split modes to subdivide a block of pixels into smaller blocks of pixels. These blocks can then be further subdivided into smaller blocks. The blocks may be referred to as nodes on the coding tree. Larger parent nodes are split into smaller child nodes. The number of times a node is subdivided is referred to as the depth of the node/coding tree. The divided blocks can be included in coding units (CUs) in some cases. For example, a CU can be a sub-portion of a CTU that contains a luma block, red difference chroma (Cr) block(s), and a blue difference chroma (Cb) block(s) along with corresponding syntax instructions for the CU. The split modes may include a binary tree (BT), triple tree (TT), and a quad tree (QT) employed to partition a node into two, three, or four child nodes, respectively, of varying shapes depending on the split modes employed. The partitioned video signal 201 is forwarded to the general coder control component 211, the transform scaling and quantization component 213, the intra-picture estimation component 215, the filter control analysis component 227, and the motion estimation component 221 for compression.

[0075] The general coder control component 211 is configured to make decisions related to coding of the images of the video sequence into the bitstream according to application constraints. For example, the general coder control component 211 manages optimization of bitrate/bitstream size versus reconstruction quality. Such decisions may be made based on storage space/bandwidth availability and image resolution requests. The general coder control

component 211 also manages buffer utilization in light of transmission speed to mitigate buffer underrun and overrun issues. To manage these issues, the general coder control component 211 manages partitioning, prediction, and filtering by the other components. For example, the general coder control component 211 may dynamically increase compression complexity to increase resolution and increase bandwidth usage or decrease compression complexity to decrease resolution and bandwidth usage. Hence, the general coder control component 211 controls the other components of codec system 200 to balance video signal reconstruction quality with bit rate concerns. The general coder control component 211 creates control data, which controls the operation of the other components. The control data is also forwarded to the header formatting and CABAC component 231 to be encoded in the bitstream to signal parameters for decoding at the decoder.

[0076] The partitioned video signal 201 is also sent to the motion estimation component 221 and the motion compensation component 219 for inter-prediction. A frame or slice of the partitioned video signal 201 may be divided into multiple video blocks. Motion estimation component 221 and the motion compensation component 219 perform inter-predictive coding of the received video block relative to one or more blocks in one or more reference frames to provide temporal prediction. Codec system 200 may perform multiple coding passes, e.g., to select an appropriate coding mode for each block of video data.

[0077] Motion estimation component 221 and motion compensation component 219 may be highly integrated, but are illustrated separately for conceptual purposes. Motion estimation, performed by motion estimation component 221, is the process of generating motion vectors, which estimate motion for video blocks. A motion vector, for example, may indicate the displacement of a coded object relative to a predictive block. A predictive block is a block that is found to closely match the block to be coded, in terms of pixel difference. A predictive block may also be referred to as a reference block. Such pixel difference may be determined by sum of absolute difference (SAD), sum of square difference (SSD), or other difference metrics. HEVC employs several coded objects including a CTU, coding tree blocks (CTBs), and CUs. For example, a CTU can be divided into CTBs, which can then be divided into CBs for inclusion in CUs. A CU can be encoded as a prediction unit (PU) containing prediction data and/or a transform unit (TU) containing transformed residual data for the CU. The motion estimation component 221 generates motion vectors, PUs, and TUs by using a rate-distortion analysis as part of a rate distortion optimization process. For example, the motion estimation component

221 may determine multiple reference blocks, multiple motion vectors, etc. for a current block/frame, and may select the reference blocks, motion vectors, etc. having the best rate-distortion characteristics. The best rate-distortion characteristics balance both quality of video reconstruction (e.g., amount of data loss by compression) with coding efficiency (e.g., size of the final encoding).

[0078] In some examples, codec system 200 may calculate values for sub-integer pixel positions of reference pictures stored in decoded picture buffer component 223. For example, video codec system 200 may interpolate values of one-quarter pixel positions, one-eighth pixel positions, or other fractional pixel positions of the reference picture. Therefore, motion estimation component 221 may perform a motion search relative to the full pixel positions and fractional pixel positions and output a motion vector with fractional pixel precision. The motion estimation component 221 calculates a motion vector for a PU of a video block in an inter-coded slice by comparing the position of the PU to the position of a predictive block of a reference picture. Motion estimation component 221 outputs the calculated motion vector as motion data to header formatting and CABAC component 231 for encoding and motion to the motion compensation component 219.

[0079] Motion compensation, performed by motion compensation component 219, may involve fetching or generating the predictive block based on the motion vector determined by motion estimation component 221. Again, motion estimation component 221 and motion compensation component 219 may be functionally integrated, in some examples. Upon receiving the motion vector for the PU of the current video block, motion compensation component 219 may locate the predictive block to which the motion vector points. A residual video block is then formed by subtracting pixel values of the predictive block from the pixel values of the current video block being coded, forming pixel difference values. In general, motion estimation component 221 performs motion estimation relative to luma components, and motion compensation component 219 uses motion vectors calculated based on the luma components for both chroma components and luma components. The predictive block and residual block are forwarded to transform scaling and quantization component 213.

[0080] The partitioned video signal 201 is also sent to intra-picture estimation component 215 and intra-picture prediction component 217. As with motion estimation component 221 and motion compensation component 219, intra-picture estimation component 215 and intra-picture prediction component 217 may be highly integrated, but are illustrated separately for conceptual purposes. The intra-picture estimation component 215 and intra-picture prediction component

217 intra-predict a current block relative to blocks in a current frame, as an alternative to the inter-prediction performed by motion estimation component 221 and motion compensation component 219 between frames, as described above. In particular, the intra-picture estimation component 215 determines an intra-prediction mode to use to encode a current block. In some examples, intra-picture estimation component 215 selects an appropriate intra-prediction mode to encode a current block from multiple tested intra-prediction modes. The selected intra-prediction modes are then forwarded to the header formatting and CABAC component 231 for encoding.

[0081] For example, the intra-picture estimation component 215 calculates rate-distortion values using a rate-distortion analysis for the various tested intra-prediction modes, and selects the intra-prediction mode having the best rate-distortion characteristics among the tested modes. Rate-distortion analysis generally determines an amount of distortion (or error) between an encoded block and an original unencoded block that was encoded to produce the encoded block, as well as a bitrate (e.g., a number of bits) used to produce the encoded block. The intra-picture estimation component 215 calculates ratios from the distortions and rates for the various encoded blocks to determine which intra-prediction mode exhibits the best rate-distortion value for the block. In addition, intra-picture estimation component 215 may be configured to code depth blocks of a depth map using a depth modeling mode (DMM) based on rate-distortion optimization (RDO).

[0082] The intra-picture prediction component 217 may generate a residual block from the predictive block based on the selected intra-prediction modes determined by intra-picture estimation component 215 when implemented on an encoder or read the residual block from the bitstream when implemented on a decoder. The residual block includes the difference in values between the predictive block and the original block, represented as a matrix. The residual block is then forwarded to the transform scaling and quantization component 213. The intra-picture estimation component 215 and the intra-picture prediction component 217 may operate on both luma and chroma components.

[0083] The transform scaling and quantization component 213 is configured to further compress the residual block. The transform scaling and quantization component 213 applies a transform, such as a discrete cosine transform (DCT), a discrete sine transform (DST), or a conceptually similar transform, to the residual block, producing a video block comprising residual transform coefficient values. Wavelet transforms, integer transforms, sub-band transforms or other types of transforms could also be used. The transform may convert the residual information from a

pixel value domain to a transform domain, such as a frequency domain. The transform scaling and quantization component 213 is also configured to scale the transformed residual information, for example based on frequency. Such scaling involves applying a scale factor to the residual information so that different frequency information is quantized at different granularities, which may affect final visual quality of the reconstructed video. The transform scaling and quantization component 213 is also configured to quantize the transform coefficients to further reduce bit rate. The quantization process may reduce the bit depth associated with some or all of the coefficients. The degree of quantization may be modified by adjusting a quantization parameter. In some examples, the transform scaling and quantization component 213 may then perform a scan of the matrix including the quantized transform coefficients. The quantized transform coefficients are forwarded to the header formatting and CABAC component 231 to be encoded in the bitstream.

[0084] The scaling and inverse transform component 229 applies a reverse operation of the transform scaling and quantization component 213 to support motion estimation. The scaling and inverse transform component 229 applies inverse scaling, transformation, and/or quantization to reconstruct the residual block in the pixel domain, e.g., for later use as a reference block which may become a predictive block for another current block. The motion estimation component 221 and/or motion compensation component 219 may calculate a reference block by adding the residual block back to a corresponding predictive block for use in motion estimation of a later block/frame. Filters are applied to the reconstructed reference blocks to mitigate artifacts created during scaling, quantization, and transform. Such artifacts could otherwise cause inaccurate prediction (and create additional artifacts) when subsequent blocks are predicted.

[0085] The filter control analysis component 227 and the in-loop filters component 225 apply the filters to the residual blocks and/or to reconstructed image blocks. For example, the transformed residual block from the scaling and inverse transform component 229 may be combined with a corresponding prediction block from intra-picture prediction component 217 and/or motion compensation component 219 to reconstruct the original image block. The filters may then be applied to the reconstructed image block. In some examples, the filters may instead be applied to the residual blocks. As with other components in FIG. 2, the filter control analysis component 227 and the in-loop filters component 225 are highly integrated and may be implemented together, but are depicted separately for conceptual purposes. Filters applied to the

reconstructed reference blocks are applied to particular spatial regions and include multiple parameters to adjust how such filters are applied. The filter control analysis component 227 analyzes the reconstructed reference blocks to determine where such filters should be applied and sets corresponding parameters. Such data is forwarded to the header formatting and CABAC component 231 as filter control data for encoding. The in-loop filters component 225 applies such filters based on the filter control data. The filters may include a deblocking filter, a noise suppression filter, a SAO filter, and an adaptive loop filter. Such filters may be applied in the spatial/pixel domain (e.g., on a reconstructed pixel block) or in the frequency domain, depending on the example.

[0086] When operating as an encoder, the filtered reconstructed image block, residual block, and/or prediction block are stored in the decoded picture buffer component 223 for later use in motion estimation as discussed above. When operating as a decoder, the decoded picture buffer component 223 stores and forwards the reconstructed and filtered blocks toward a display as part of an output video signal. The decoded picture buffer component 223 may be any memory device capable of storing prediction blocks, residual blocks, and/or reconstructed image blocks.

[0087] The header formatting and CABAC component 231 receives the data from the various components of codec system 200 and encodes such data into a coded bitstream for transmission toward a decoder. Specifically, the header formatting and CABAC component 231 generates various headers to encode control data, such as general control data and filter control data. Further, prediction data, including intra-prediction and motion data, as well as residual data in the form of quantized transform coefficient data are all encoded in the bitstream. The final bitstream includes all information desired by the decoder to reconstruct the original partitioned video signal 201. Such information may also include intra-prediction mode index tables (also referred to as codeword mapping tables), definitions of encoding contexts for various blocks, indications of most probable intra-prediction modes, an indication of partition information, etc. Such data may be encoded by employing entropy coding. For example, the information may be encoded by employing context adaptive variable length coding (CAVLC), CABAC, syntax-based context-adaptive binary arithmetic coding (SBAC), probability interval partitioning entropy (PIPE) coding, or another entropy coding technique. Following the entropy coding, the coded bitstream may be transmitted to another device (e.g., a video decoder) or archived for later transmission or retrieval.

[0088] FIG. 3 is a block diagram illustrating an example video encoder 300. Video encoder 300 may be employed to implement the encoding functions of codec system 200 and/or implement steps 101, 103, 105, 107, and/or 109 of operating method 100. Encoder 300 partitions an input video signal, resulting in a partitioned video signal 301, which is substantially similar to the partitioned video signal 201. The partitioned video signal 301 is then compressed and encoded into a bitstream by components of encoder 300.

[0089] Specifically, the partitioned video signal 301 is forwarded to an intra-picture prediction component 317 for intra-prediction. The intra-picture prediction component 317 may be substantially similar to intra-picture estimation component 215 and intra-picture prediction component 217. The partitioned video signal 301 is also forwarded to a motion compensation component 321 for inter-prediction based on reference blocks in a decoded picture buffer component 323. The motion compensation component 321 may be substantially similar to motion estimation component 221 and motion compensation component 219. The prediction blocks and residual blocks from the intra-picture prediction component 317 and the motion compensation component 321 are forwarded to a transform and quantization component 313 for transform and quantization of the residual blocks. The transform and quantization component 313 may be substantially similar to the transform scaling and quantization component 213. The transformed and quantized residual blocks and the corresponding prediction blocks (along with associated control data) are forwarded to an entropy coding component 331 for coding into a bitstream. The entropy coding component 331 may be substantially similar to the header formatting and CABAC component 231.

[0090] The transformed and quantized residual blocks and/or the corresponding prediction blocks are also forwarded from the transform and quantization component 313 to an inverse transform and quantization component 329 for reconstruction into reference blocks for use by the motion compensation component 321. The inverse transform and quantization component 329 may be substantially similar to the scaling and inverse transform component 229. In-loop filters in an in-loop filters component 325 are also applied to the residual blocks and/or reconstructed reference blocks, depending on the example. The in-loop filters component 325 may be substantially similar to the filter control analysis component 227 and the in-loop filters component 225. The in-loop filters component 325 may include multiple filters as discussed with respect to in-loop filters component 225. The filtered blocks are then stored in a decoded picture buffer component 323 for use as reference blocks by the motion compensation

component 321. The decoded picture buffer component 323 may be substantially similar to the decoded picture buffer component 223.

[0091] FIG. 4 is a block diagram illustrating an example video decoder 400. Video decoder 400 may be employed to implement the decoding functions of codec system 200 and/or implement steps 111, 113, 115, and/or 117 of operating method 100. Decoder 400 receives a bitstream, for example from an encoder 300, and generates a reconstructed output video signal based on the bitstream for display to an end user.

[0092] The bitstream is received by an entropy decoding component 433. The entropy decoding component 433 is configured to implement an entropy decoding scheme, such as CAVLC, CABAC, SBAC, PIPE coding, or other entropy coding techniques. For example, the entropy decoding component 433 may employ header information to provide a context to interpret additional data encoded as codewords in the bitstream. The decoded information includes any desired information to decode the video signal, such as general control data, filter control data, partition information, motion data, prediction data, and quantized transform coefficients from residual blocks. The quantized transform coefficients are forwarded to an inverse transform and quantization component 429 for reconstruction into residual blocks. The inverse transform and quantization component 429 may be similar to inverse transform and quantization component 329.

[0093] The reconstructed residual blocks and/or prediction blocks are forwarded to intra-picture prediction component 417 for reconstruction into image blocks based on intra-prediction operations. The intra-picture prediction component 417 may be similar to intra-picture estimation component 215 and an intra-picture prediction component 217. Specifically, the intra-picture prediction component 417 employs prediction modes to locate a reference block in the frame and applies a residual block to the result to reconstruct intra-predicted image blocks. The reconstructed intra-predicted image blocks and/or the residual blocks and corresponding inter-prediction data are forwarded to a decoded picture buffer component 423 via an in-loop filters component 425, which may be substantially similar to decoded picture buffer component 223 and in-loop filters component 225, respectively. The in-loop filters component 425 filters the reconstructed image blocks, residual blocks and/or prediction blocks, and such information is stored in the decoded picture buffer component 423. Reconstructed image blocks from decoded picture buffer component 423 are forwarded to a motion compensation component 421 for inter-prediction. The motion compensation component 421 may be substantially similar to motion

estimation component 221 and/or motion compensation component 219. Specifically, the motion compensation component 421 employs motion vectors from a reference block to generate a prediction block and applies a residual block to the result to reconstruct an image block. The resulting reconstructed blocks may also be forwarded via the in-loop filters component 425 to the decoded picture buffer component 423. The decoded picture buffer component 423 continues to store additional reconstructed image blocks, which can be reconstructed into frames via the partition information. Such frames may also be placed in a sequence. The sequence is output toward a display as a reconstructed output video signal.

[0094] Keeping the above in mind, video compression techniques perform spatial (intra-picture) prediction and/or temporal (inter-picture) prediction to reduce or remove redundancy inherent in video sequences. For block-based video coding, a video slice (i.e., a video picture or a portion of a video picture) may be partitioned into video blocks, which may also be referred to as treeblocks, coding tree blocks (CTBs), coding tree units (CTUs), coding units (CUs), and/or coding nodes. Video blocks in an intra-coded (I) slice of a picture are encoded using spatial prediction with respect to reference samples in neighboring blocks in the same picture. Video blocks in an inter-coded (P or B) slice of a picture may use spatial prediction with respect to reference samples in neighboring blocks in the same picture or temporal prediction with respect to reference samples in other reference pictures. Pictures may be referred to as frames, and reference pictures may be referred to as reference frames. The POC is a variable associated with each picture that uniquely identifies the associated picture among all pictures in the coded layer video sequence (CLVS), indicates when the associated picture is to be output from the DPB, and indicates the position of the associated picture in output order relative to the output order positions of the other pictures in the same CLVS that are to be output from the DPB. A flag is a variable or single-bit syntax element that can take one of the two possible values: 0 and 1.

[0095] Spatial or temporal prediction results in a predictive block for a block to be coded. Residual data represents pixel differences between the original block to be coded and the predictive block. An inter-coded block is encoded according to a motion vector that points to a block of reference samples forming the predictive block, and the residual data indicating the difference between the coded block and the predictive block. An intra-coded block is encoded according to an intra-coding mode and the residual data. For further compression, the residual data may be transformed from the pixel domain to a transform domain, resulting in residual transform coefficients, which then may be quantized. The quantized transform coefficients,

initially arranged in a two-dimensional array, may be scanned in order to produce a one-dimensional vector of transform coefficients, and entropy coding may be applied to achieve even more compression.

[0096] Image and video compression has experienced rapid growth, leading to various coding standards. Such video coding standards include ITU-T H.261, International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) MPEG-1 Part 2, ITU-T H.262 or ISO/IEC MPEG-2 Part 2, ITU-T H.263, ISO/IEC MPEG-4 Part 2, Advanced Video Coding (AVC), also known as ITU-T H.264 or ISO/IEC MPEG-4 Part 10, and High Efficiency Video Coding (HEVC), also known as ITU-T H.265 or MPEG-H Part 2. AVC includes extensions such as Scalable Video Coding (SVC), Multiview Video Coding (MVC) and Multiview Video Coding plus Depth (MVC+D), and 3D AVC (3D-AVC). HEVC includes extensions such as Scalable HEVC (SHVC), Multiview HEVC (MV-HEVC), and 3D HEVC (3D-HEVC).

[0097] There is also a new video coding standard, named Versatile Video Coding (VVC), being developed by the joint video experts team (JVET) of ITU-T and ISO/IEC. While the VVC standard has several working drafts, one Working Draft (WD) of VVC in particular, namely B. Bross, J. Chen, and S. Liu, "Versatile Video Coding (Draft 5)," JVET-N1001-v3, 13th JVET Meeting, March 27, 2019 (VVC Draft 5) is referenced herein.

[0098] Scalability in video coding usually is supported by using multi-layer coding techniques. A multi-layer bitstream comprises a base layer (BL) and one or more enhancement layers (ELs). An example of scalabilities includes spatial scalability, quality / signal-to-noise (SNR) scalability, multi-view scalability, etc. When a multi-layer coding technique is used, a picture or a part thereof may be coded (1) without using a reference picture, i.e., using intra prediction; (2) by referencing to reference pictures that are in the same layer, i.e., using inter prediction; or (3) by referencing to reference pictures that are in other layer(s), i.e., using inter-layer prediction. A reference picture used for inter-layer prediction of the current picture is referred to as an inter-layer reference picture (ILRP).

[0099] FIG. 5 is a schematic diagram illustrating an example of layer based prediction 500, for example as performed to determine MVs at block compression step 105, block decoding step 113, motion estimation component 221, motion compensation component 219, motion compensation component 321, and/or motion compensation component 421. Layer based

prediction 500 is compatible with unidirectional inter-prediction and/or bidirectional inter-prediction, but is also performed between pictures in different layers.

[00100] Layer based prediction 500 is applied between pictures 511, 512, 513, and 514 and pictures 515, 516, 517, and 518 in different layers. In the example shown, pictures 511, 512, 513, and 514 are part of layer N+1 532 and pictures 515, 516, 517, and 518 are part of layer N 531. A layer, such as layer N 531 and/or layer N+1 532, is a group of pictures that are all associated with a similar value of a characteristic, such as a similar size, quality, resolution, signal to noise ratio, capability, etc. In the example shown, layer N+1 532 is associated with a larger image size than layer N 531. Accordingly, pictures 511, 512, 513, and 514 in layer N+1 532 have a larger picture size (e.g., larger height and width and hence more samples) than pictures 515, 516, 517, and 518 in layer N 531 in this example. However, such pictures can be separated between layer N+1 532 and layer N 531 by other characteristics. While only two layers, layer N+1 532 and layer N 531, are shown, a set of pictures can be separated into any number of layers based on associated characteristics. Layer N+1 532 and layer N 531 may also be denoted by a layer ID. A layer ID is an item of data that is associated with a picture and denotes the picture is part of an indicated layer. Accordingly, each picture 511-518 may be associated with a corresponding layer ID to indicate which layer N+1 532 or layer N 531 includes the corresponding picture.

[00101] Pictures 511-518 in different layers 531-532 are configured to be displayed in the alternative. As such, pictures 511-518 in different layers 531-532 can share the same temporal identifier (ID) and can be included in the same AU. As used herein, an AU is a set of one or more coded pictures associated with the same display time for output from a DPB. For example, a decoder may decode and display picture 515 at a current display time if a smaller picture is desired or the decoder may decode and display picture 511 at the current display time if a larger picture is desired. As such, pictures 511-514 at higher layer N+1 532 contain substantially the same image data as corresponding pictures 515-518 at lower layer N 531 (notwithstanding the difference in picture size). Specifically, picture 511 contains substantially the same image data as picture 515, picture 512 contains substantially the same image data as picture 516, etc.

[00102] Pictures 511-518 can be coded by reference to other pictures 511-518 in the same layer N 531 or N+1 532. Coding a picture in reference to another picture in the same layer results in inter-prediction 523, which is compatible unidirectional inter-prediction and/or bidirectional inter-prediction. Inter-prediction 523 is depicted by solid line arrows. For

example, picture 513 may be coded by employing inter-prediction 523 using one or two of pictures 511, 512, and/or 514 in layer N+1 532 as a reference, where one picture is referenced for unidirectional inter-prediction and/or two pictures are reference for bidirectional inter-prediction. Further, picture 517 may be coded by employing inter-prediction 523 using one or two of pictures 515, 516, and/or 518 in layer N 531 as a reference, where one picture is referenced for unidirectional inter-prediction and/or two pictures are reference for bidirectional inter-prediction. When a picture is used as a reference for another picture in the same layer when performing inter-prediction 523, the picture may be referred to as a reference picture. For example, picture 512 may be a reference picture used to code picture 513 according to inter-prediction 523. Inter-prediction 523 can also be referred to as intra-layer prediction in a multi-layer context. As such, inter-prediction 523 is a mechanism of coding samples of a current picture by reference to indicated samples in a reference picture that are different from the current picture where the reference picture and the current picture are in the same layer.

[00103] Pictures 511-518 can also be coded by reference to other pictures 511-518 in different layers. This process is known as inter-layer prediction 521, and is depicted by dashed arrows. Inter-layer prediction 521 is a mechanism of coding samples of a current picture by reference to indicated samples in a reference picture where the current picture and the reference picture are in different layers and hence have different layer IDs. For example, a picture in a lower layer N 531 can be used as a reference picture to code a corresponding picture at a higher layer N+1 532. As a specific example, picture 511 can be coded by reference to picture 515 according to inter-layer prediction 521. In such a case, the picture 515 is used as an inter-layer reference picture. An inter-layer reference picture is a reference picture used for inter-layer prediction 521. In most cases, inter-layer prediction 521 is constrained such that a current picture, such as picture 511, can only use inter-layer reference picture(s) that are included in the same AU and that are at a lower layer, such as picture 515. When multiple layers (e.g., more than two) are available, inter-layer prediction 521 can encode/decode a current picture based on multiple inter-layer reference picture(s) at lower levels than the current picture.

[00104] A video encoder can employ layer based prediction 500 to encode pictures 511-518 via many different combinations and/or permutations of inter-prediction 523 and inter-layer prediction 521. For example, picture 515 may be coded according to intra-prediction. Pictures 516-518 can then be coded according to inter-prediction 523 by using picture 515 as a reference picture. Further, picture 511 may be coded according to inter-layer prediction 521 by using

picture 515 as an inter-layer reference picture. Pictures 512-514 can then be coded according to inter-prediction 523 by using picture 511 as a reference picture. As such, a reference picture can serve as both a single layer reference picture and an inter-layer reference picture for different coding mechanisms. By coding higher layer N+1 532 pictures based on lower layer N 531 pictures, the higher layer N+1 532 can avoid employing intra-prediction, which has much lower coding efficiency than inter-prediction 523 and inter-layer prediction 521. As such, the poor coding efficiency of intra-prediction can be limited to the smallest/lowest quality pictures, and hence limited to coding the smallest amount of video data. The pictures used as reference pictures and/or inter-layer reference pictures can be indicated in entries of reference picture list(s) contained in a reference picture list structure.

[00105] Each AU 506 in FIG. 5 may contain several pictures. For example, one AU 506 may contain pictures 511 and 515. Another AU 506 may contain pictures 512 and 516. Indeed, each AU 506 is a set of one or more coded pictures associated with the same display time (e.g., the same temporal ID) for output from a decoded picture buffer (DPB) (e.g., for display to a user). Each AUD 508 is an indicator or data structure used to indicate the start of an AU (e.g., AU 508) or the boundary between AUs.

[00106] Previous H.26x video coding families have provided support for scalability in separate profile(s) from the profile(s) for single-layer coding. Scalable video coding (SVC) is the scalable extension of the AVC/H.264 that provides support for spatial, temporal, and quality scalabilities. For SVC, a flag is signaled in each macroblock (MB) in EL pictures to indicate whether the EL MB is predicted using the collocated block from a lower layer. The prediction from the collocated block may include texture, motion vectors, and/or coding modes. Implementations of SVC cannot directly reuse unmodified H.264/AVC implementations in their design. The SVC EL macroblock syntax and decoding process differs from H.264/AVC syntax and decoding process.

[00107] Scalable HEVC (SHVC) is the extension of the HEVC/H.265 standard that provides support for spatial and quality scalabilities, multiview HEVC (MV-HEVC) is the extension of the HEVC/H.265 that provides support for multi-view scalability, and 3D HEVC (3D-HEVC) is the extension of the HEVC/H.264 that provides supports for three dimensional (3D) video coding that is more advanced and more efficient than MV-HEVC. Note that the temporal scalability is included as an integral part of the single-layer HEVC codec. The design of the multi-layer extension of HEVC employs the idea where the decoded pictures used for

inter-layer prediction come only from the same access unit (AU) and are treated as long-term reference pictures (LTRPs), and are assigned reference indices in the reference picture list(s) along with other temporal reference pictures in the current layer. Inter-layer prediction (ILP) is achieved at the prediction unit (PU) level by setting the value of the reference index to refer to the inter-layer reference picture(s) in the reference picture list(s).

[00108] Notably, both reference picture resampling and spatial scalability features call for resampling of a reference picture or part thereof. Reference picture resampling (RPR) can be realized at either the picture level or coding block level. However, when RPR is referred to as a coding feature, it is a feature for single-layer coding. Even so, it is possible or even preferable from a codec design point of view to use the same resampling filter for both the RPR feature of single-layer coding and the spatial scalability feature for multi-layer coding.

[00109] FIG. 6 illustrates an example of layer based prediction 600 utilizing output layer sets (OLSs), for example as performed to determine MVs at block compression step 105, block decoding step 113, motion estimation component 221, motion compensation component 219, motion compensation component 321, and/or motion compensation component 421. Layer based prediction 500 is compatible with unidirectional inter-prediction and/or bidirectional inter-prediction, but is also performed between pictures in different layers. The layer based prediction of FIG. 6 is similar to that of FIG. 5. Therefore, for the sake of brevity, a full description of layer based prediction is not repeated.

[00110] Some of the layers in the coded video sequence (CVS) 690 of FIG. 6 are included in an OLS. An OLS is set of layers for which one or more layers are specified as the output layers. An output layer is a layer of an OLS that is output. FIG. 6 depicts three different OLSs, namely OLS 1, OLS 2, and OLS 3. As shown, OLS 1 includes Layer N 631 and Layer N+1 632. OLS 2 includes Layer N 631, Layer N+1 632, Layer N+2 633, and Layer N+1 634. OLS 3 includes Layer N 631, Layer N+1 632, and Layer N+2 633. Despite three OLSs being shown, a different number of OLSs may be used in practical applications.

[00111] Each of the different OLSs may contain any number of layers. The different OLSs are generated in an effort to accommodate the coding capabilities of a variety of different devices having varying coding capabilities. For example, OLS 1, which contains only two layers, may be generated to accomoate a mobile phone with relatively limited coding capabilities. On the other hand, OLS 2, which contains four layers, may be generated to accommodate a big screen television, which is able to decode higher layers than the mobile

phone. OLS 3, which contains three layers, may be generated to accommodate a personal computer, laptop computer, or a tablet computer, which may be able to decode higher layers than the mobile phone but cannot decode the highest layers like the big screen television.

[00112] The layers in FIG. 6 can be all independent from each other. That is, each layer can be coded without using inter-layer prediction (ILP). In this case, the layers are referred to as simulcast layers. One or more of the layers in FIG. 6 may also be coded using ILP. Whether the layers are simulcast layers or whether some of layers are coded using ILP is signaled by a flag in a video parameter set (VPS), which will be more fully discussed below. When some layers use ILP, the layer dependency relationship among layers is also signaled in the VPS.

[00113] In an embodiment, when the layers are simulcast layers, only one layer is selected for decoding and output. In an embodiment, when some layers use ILP, all of the layers (e.g., the entire bitstream) are specified to be decoded, and certain layers among the layers are specified to be output layers. The output layer or layers may be, for example, 1) only the highest layer, 2) all the layers, or 3) the highest layer plus a set of indicated lower layers. For example, when the highest layer plus a set of indicated lower layers are designated for output by a flag in the VPS, layer N+3 634 (which is the highest layer) and layers N 631 and N+1 632 (which are lower layers) from OLS 2 are output.

[00114] Still referring to FIG. 6, some layers are included in an OLS, but not output or used for reference by an output layer. For example, Layer N + 2 633 is contained in OLS 2, but is not output. Rather, Layer N+3 634 is the highest layer in OLS 2 and is, therefore, output as the output layer. In addition, Layer N + 2 633 in OLS 2 is not used for reference by the output layer, Layer N+3 634. Indeed, Layer N + 3 634 in OLS 2 relies on Layer N + 2 633, Layer N+1 632, and/or Layer N 631 in OLS 2. Thus, Layer N+2 633 in OLS 2 is referred to as an unnecessary layer. Unfortunately, SHVC and MV-HEVC allow such unnecessary layers to be included in multi-layer video bitstreams, which unnecessarily burdens coding resources and decreases coding efficiency.

[00115] Disclosed herein are techniques that disallow unnecessary layers in multi-layer video bitstreams. That is, every layer included in an output layer set (OLS) is either an output layer or used as a direct or indirect reference layer of an output layer. This avoids having extraneous information in the coding process and increases coding efficiency. Thus, the coder / decoder (a.k.a., “codec”) in video coding is improved relative to current codecs. As a practical

matter, the improved video coding process offers the user a better user experience when videos are sent, received, and/or viewed.

[00116] FIG. 7 illustrates an embodiment of a video bitstream 700. As used herein the video bitstream 700 may also be referred to as a coded video bitstream, a bitstream, or variations thereof. As shown in FIG. 7, the bitstream 700 comprises at least one picture unit (PU) 701. While three of the PUs 701 are shown in FIG. 7, a different number of PUs 701 may be present in the bitstream 700 in practical applications. Each PU 701 is a set of NAL units that are associated with each other according to a specified classification rule, are consecutive in decoding order, and contain exactly one coded picture (e.g., picture 714).

[00117] In an embodiment, each PU 701 contains one or more of the following: decoding capability information (DCI) 702, a video parameter set (VPS) 704, a sequence parameter set (SPS) 706, a picture parameter set (PPS) 708, a picture header (PH) 712, and a picture 714. Each of the DCI 702, the VPS 704, the SPS 706, and the PPS 708 may be generically referred to as a parameter set. In an embodiment, other parameter sets not shown in FIG. 7 may also be included in the bitstream 700 such as, for example, an adaption parameter set (APS), which is a syntax structure containing syntax elements that apply to zero or more slices as determined by zero or more syntax elements found in slice headers.

[00118] The DCI 702, which may also be referred to a decoding parameter set (DPS) or decoder parameter set, is a syntax structure containing syntax elements that apply to the entire bitstream. The DCI 702 includes parameters that stay constant for the lifetime of the video bitstream (e.g., bitstream 700), which can translate to the lifetime of a session. The DCI 702 can include profile, level, and sub-profile information to determine a maximum complexity interop point that is guaranteed to be never exceeded, even if splicing of video sequences occurs within a session. It further optionally includes constraint flags, which indicate that the video bitstream will be constraint of the use of certain features as indicated by the values of those flags. With this, a bitstream can be labelled as not using certain tools, which allows among other things for resource allocation in a decoder implementation. Like all parameter sets, the DCI 702 is present when first referenced, and referenced by the very first picture in a video sequence, implying that it has to be sent among the first NAL units in the bitstream. While multiple DCIs 702 can be in the bitstream, the value of the syntax elements therein cannot be inconsistent when being referenced.

[00119] The VPS 704 includes decoding dependency or information for reference picture set construction of enhancement layers. The VPS 704 provides an overall perspective or view of a scalable sequence, including what types of operation points are provided, the profile, tier, and level of the operation points, and some other high-level properties of the bitstream that can be used as the basis for session negotiation and content selection, etc.

[00120] In an embodiment, when it is indicated that some of the layers use ILP, the VPS 704 indicates that a total number of OLSs specified by the VPS is equal to the number of layers, indicates that the i-th OLS includes the layers with layer indices from 0 to i, inclusive, and indicates that for each OLS only the highest layer in the OLS is output.

[00121] In an embodiment, the VPS 704 includes syntax and semantics corresponding to the OLSs in the CLVS and/or video bitstream. The following syntax and semantics corresponding to the VPS 704 may be employed to implement the embodiments disclosed herein.

[00122] The syntax for the VPS 704 may be as follows.

video_parameter_set_rbsp( ) {	**Descriptor**
vps_video_parameter_set_id	u(4)
vps_max_layers_minus1	u(6)
vps_max_sub_layers_minus1	u(3)
if(vps_max_layers_minus1 > 0)	
vps_all_independent_layers_flag	u(1)
for(i = 0; i <= vps_max_layers_minus1; i++) {	
vps_layer_id[i]	u(6)
if(i > 0 && !vps_all_independent_layers_flag) {	
vps_independent_layer_flag[i]	u(1)
if(!vps_independent_layer_flag[i])	
for(j = 0; j < i; j++)	
vps_direct_dependency_flag[i][j]	u(1)
}	
}	
if(vps_max_layers_minus1 > 0) {	

if(vps_all_independent_layers_flag)	
each_layer_is_an_ols_flag	u(1)
if(!each_layer_is_an_ols_flag) {	
if(!vps_all_independent_layers_flag)	
ols_mode_idc	u(2)
if(ols_mode_idc == 2) {	
num_output_layer_sets_minus1	u(8)
for(i = 1; i < num_output_layer_sets_minus1 + 1; i ++) {	
for(j = 0; j <= vps_max_layers_minus1; j++)	
layer_included_flag[i][j]	u(1)
if(!vps_all_independent_layers_flag)	
for(j = 0; j < NumLayersInOls[i] − 1; j++)	
vps_output_layer_flag[i][j]	u(1)
}	
}	
}	
}	
vps_constraint_info_present_flag	u(1)
vps_reserved_zero_7bits	u(7)
if(vps_constraint_info_present_flag)	
general_constraint_info( )	
general_hrd_params_present_flag	u(1)
if(general_hrd_params_present_flag) {	
num_units_in_tick	u(32)
time_scale	u(32)
general_hrd_parameters( )	
}	
vps_extension_flag	u(1)
if(vps_extension_flag)	
while(more_rbsp_data( ))	
vps_extension_data_flag	u(1)

rbsp_trailing_bits( )	
}	

[00123] The semantics for the VPS 704 may be as follows. In an embodiment, the VPS 704 includes one or more of the flags and parameters noted below.

[00124] A VPS raw byte sequence payload (RBSP) shall be available to the decoding process prior to it being referenced, included in at least one access unit with TemporalId equal to 0 or provided through external means, and the VPS NAL unit containing the VPS RBSP shall have nuh_layer_id equal to vps_layer_id[0]. The nuh_layer_id specifies the identifier of the layer to which a VCL NAL unit belongs or the identifier of a layer to which a non-VCL NAL unit applies. The value of nuh_layer_id shall be in the range of 0 to 55, inclusive. The TemporalId is an ID used to uniquely identify a particular NAL unit relative to other NAL units. The value of TemporalId shall be the same for all VCL NAL units of an AU. The value of TemporalId of a coded picture, a PU, or an AU is the value of the TemporalId of the VCL NAL units of the coded picture, PU, or AU.

[00125] All VPS NAL units with a particular value of vps_video_parameter_set_id in a CVS shall have the same content. vps_video_parameter_set_id provides an identifier for the VPS for reference by other syntax elements. vps_max_layers_minus1 plus 1 specifies the maximum allowed number of layers in each CVS referring to the VPS. vps_max_sub_layers_minus1 plus 1 specifies the maximum number of temporal sub-layers that may be present in each CVS referring to the VPS. The value of vps_max_sub_layers_minus1 shall be in the range of 0 to 6, inclusive.

[00126] vps_all_independent_layers_flag equal to 1 specifies that all layers in the CVS are independently coded without using inter-layer prediction. vps_all_independent_layers_flag equal to 0 specifies that one or more of the layers in the CVS may use inter-layer prediction. When not present, the value of vps_all_independent_layers_flag is inferred to be equal to 1. When vps_all_independent_layers_flag is equal to 1, the value of vps_independent_layer_flag[i] is inferred to be equal to 1. When vps_all_independent_layers_flag is equal to 0, the value of vps_independent_layer_flag[0] is inferred to be equal to 1.

[00127] vps_layer_id[i] specifies the nuh_layer_id value of the i-th layer. For any two non-negative integer values of m and n, when m is less than n, the value of vps_layer_id[m]

shall be less than vps_layer_id[n]. vps_independent_layer_flag[i] equal to 1 specifies that the layer with index i does not use inter-layer prediction. vps_independent_layer_flag[i] equal to 0 specifies that the layer with index i may use inter-layer prediction and vps_layer_dependency_flag[i] is present in VPS. When not present, the value of vps_independent_layer_flag[i] is inferred to be equal to 1.

[00128] vps_direct_dependency_flag[i][j] equal to 0 specifies that the layer with index j is not a direct reference layer for the layer with index i. vps_direct_dependency_flag [i][j] equal to 1 specifies that the layer with index j is a direct reference layer for the layer with index i. When vps_direct_dependency_flag[i][j] is not present for i and j in the range of 0 to vps_max_layers_minus1, inclusive, it is inferred to be equal to 0.

[00129] The variable DirectDependentLayerIdx[i][j], specifying the j-th direct dependent layer of the i-th layer, is derived as follows.

```
for( i = 1; i < vps_max_layers_minus1; i++ )
    if( !vps_independent_layer_flag[ i ] )
        for( j = i, k = 0; j >= 0; j− − )
            if( vps_direct_dependency_flag[ i ][ j ] )
                DirectDependentLayerIdx[ i ][ k++ ] = j
```

[00130] The variable GeneralLayerIdx[i], specifying the layer index of the layer with nuh_layer_id equal to vps_layer_id[i], is derived as follows.

```
for( i = 0; i <= vps_max_layers_minus1; i++ )
    GeneralLayerIdx[ vps_layer_id[ i ] ] = i
```

[00131] each_layer_is_an_ols_flag equal to 1 specifies that each output layer set contains only one layer and each layer itself in the bitstream is an outpu layer set with the single included layer being the only output layer. each_layer_is_an_ols_flag equal to 0 that an outpu layer set may contain more than one layer. If vps_max_layers_minus1 is equal to 0, the vlaue of the value of each_layer_is_an_ols_flag is inferred to be equal to 1. Otherwise, when vps_all_independent_layers_flag is equal to 0, the value of each_layer_is_an_ols_flag is inferred to be equal to 0.

[00132] ols_mode_idc equal to 0 specifies that the total number of OLSs specified by the VPS is equal to vps_max_layers_minus1 + 1, the i-th OLS includes the layers with layer indices from 0 to i, inclusive, and for each OLS only the highest layer in the OLS is output. ols_mode_idc equal to 1 specifies that the total number of OLSs specified by the VPS is equal to

vps_max_layers_minus1 + 1, the i-th OLS includes the layers with layer indices from 0 to i, inclusive, and for each OLS all layers in the OLS are output. ols_mode_idc equal to 2 specifies that the total number of OLSs specified by the VPS is explicitly signalled and for each OLS the highest layer and an explicitly signalled set of lower layers in the OLS are output. The value of ols_mode_idc shall be in the range of 0 to 2, inclusive. The value 3 of ols_mode_idc is reserved for future use by ITU-T | ISO/IEC. When vps_all_independent_layers_flag is equal to 1 and each_layer_is_an_ols_flag is equal to 0, the value of ols_mode_idc is inferred to be equal to 2.

[00133] num_output_layer_sets_minus1 plus 1 specifies the total number of OLSs specified by the VPS when ols_mode_idc is equal to 2.

[00134] The variable TotalNumOlss, specifying the total number of OLSs specified by the VPS, is derived as follows:

```
if( vps_max_layers_minus1 == 0 )
    TotalNumOlss = 1
else if( each_layer_is_an_ols_flag || ols_mode_idc == 0 || ols_mode_idc == 1 )
    TotalNumOlss = vps_max_layers_minus1 + 1
else if( ols_mode_idc == 2 )
    TotalNumOlss = num_output_layer_sets_minus1 + 1
```

[00135] layer_included_flag[i][j] specifies whether the j-th layer (i.e., the layer with nuh_layer_id equal to vps_layer_id[j]) is included in the i-th OLS when ols_mode_idc is equal to 2. layer_included_flag[i][j] equal to 1 specifies that the j-th layer is included in the i-th OLS. layer_included_flag[i][j] equal to 0 specifies the j-th layer is not included in the i-th OLS.

[00136] The variable NumLayersInOls[i], specifying the number of layers in the i-th OLS, and the variable LayerIdInOls[i][j], specifying the nuh_layer_id value of the j-th layer in the i-th OLS, are derived as follows.

```
NumLayersInOls[ 0 ] = 1
LayerIdInOls[ 0 ][ 0 ] = vps_layer_id[ 0 ]
for( i = 1, i < TotalNumOlss; i++ ) {
 if( each_layer_is_an_ols_flag ) {
     NumLayersInOls[ i ] = 1
     LayerIdInOls[ i ][ 0 ] = vps_layer_id[ i ]
 } else if( ols_mode_idc == 0 || ols_mode_idc == 1 ) {
```

```
        NumLayersInOls[ i ] = i + 1
        for( j = 0; j < NumLayersInOls[ i ]; j++ )
            LayerIdInOls[ i ][ j ] = vps_layer_id[ j ]
    } else if( ols_mode_idc  = =  2 ) {
        for( k = 0, j = 0; k  <=  vps_max_layers_minus1; k++ )
            if( layer_included_flag[ i ][ k ] )
                LayerIdInOls[ i ][ j++ ] = vps_layer_id[ k ]
        NumLayersInOls[ i ] = j
    }
```

[00137] The variable OlsLayeIdx[i][j], specifying the OLS layer index of the layer with nuh_layer_id equal to LayerIdInOls[i][j], is derived as follows.

```
for( i = 0, i < TotalNumOlss; i++ )
    for j = 0; j < NumLayersInOls[ i ]; j++ )
        OlsLayeIdx[ i ][ LayerIdInOls[ i ][ j ] ] = j
```

[00138] The lowest layer in each OLS shall be an independent layer. In other words, for each i in the range of 0 to TotalNumOlss − 1, inclusive, the value of vps_independent_layer_flag[GeneralLayerIdx[LayerIdInOls[i][0]]] shall be equal to 1.

[00139] Each layer shall be included in at least one OLS specified by the VPS. That is, for each layer with a particular value of nuh_layer_id nuhLayerId equal to one of vps_layer_id[k] for k in the range of 0 to vps_max_layers_minus1, inclusive, there shall be at least one pair of values of i and j, where i is in the range of 0 to TotalNumOlss − 1, inclusive, and j is in the range of NumLayersInOls[i] − 1, inclusive, such that the value of LayerIdInOls[i][j] is equal to nuhLayerId.

[00140] Any layer in an OLS shall be an output layer of the OLS or a (direct or indirect) reference layer of an output layer of the OLS.

[00141] vps_output_layer_flag[i][j] specifies whether the j-th layer in the i-th OLS is output when ols_mode_idc is equal to 2. vps_output_layer_flag[i] equal to 1 specifies that the j-th layer in the i-th OLS is output. vps_output_layer_flag[i] equal to 0 specifies that the j-th layer in the i-th OLS is not output. When vps_all_independent_layers_flag is equal to 1 and

each_layer_is_an_ols_flag is equal to 0, the value of vps_output_layer_flag[i] is inferred to be equal to 1.

[00142] The variable OutputLayerFlag[i][j], for which the value 1 specifies that the j-th layer in the i-th OLS is output and the value 0 specifies that the j-th layer in the i-th OLS is not output, is derived as follows.

```
for( i = 0, i < TotalNumOlss; i++ ) {
    OutputLayerFlag[ i ][ NumLayersInOls[ i ] − 1 ] = 1
    for( j = 0; j < NumLayersInOls[ i ] − 1; j++ )
        if( ols_mode_idc[ i ]  = =  0 )
            OutputLayerFlag[ i ][ j ] = 0
        else if( ols_mode_idc[ i ]  = =  1 )
            OutputLayerFlag[ i ][ j ] = 1
        else if( ols_mode_idc[ i ]  = =  2 )
            OutputLayerFlag[ i ][ j ] = vps_output_layer_flag[ i ][ j ]
}
```

[00143] NOTE – The 0-th OLS contains only the lowest layer (i.e., the layer with nuh_layer_id equal to vps_layer_id[0]) and for the 0-th OLS the only included layer is output.

[00144] vps_constraint_info_present_flag equal to 1 specifies that the general_constraint_info( ) syntax structure is present in the VPS. vps_constraint_info_present_flag equal to 0 specifies that the general_constraint_info( ) syntax structure is not present in the VPS.

[00145] vps_reserved_zero_7bits shall be equal to 0 in bitstreams conforming to this version of this VVC draft. Other values for vps_reserved_zero_7bits are reserved for future use by ITU-T | ISO/IEC. In an embodiment, decoders shall ignore the value of vps_reserved_zero_7bits.

[00146] general_hrd_params_present_flag equal to 1 specifies that the syntax elements num_units_in_tick and time_scale and the syntax structure general_hrd_parameters( ) are present in the SPS RBSP syntax structure. general_hrd_params_present_flag equal to 0 specifies that the syntax elements num_units_in_tick and time_scale and the syntax structure general_hrd_parameters( ) are not present in the SPS RBSP syntax structure.

[00147] num_units_in_tick is the number of time units of a clock operating at the frequency time_scale Hz that corresponds to one increment (called a clock tick) of a clock tick

counter. num_units_in_tick shall be greater than 0. A clock tick, in units of seconds, is equal to the quotient of num_units_in_tick divided by time_scale. For example, when the picture rate of a video signal is 25 Hz, time_scale may be equal to 27,000,000 and num_units_in_tick may be equal to 1,080,000, and consequently a clock tick may be equal to 0.04 seconds.

[00148] time_scale is the number of time units that pass in one second. For example, a time coordinate system that measures time using a 27 MHz clock has a time_scale of 27,000,000. The value of time_scale shall be greater than 0.

[00149] vps_extension_flag equal to 0 specifies that no vps_extension_data_flag syntax elements are present in the VPS RBSP syntax structure. vps_extension_flag equal to 1 specifies that there are vps_extension_data_flag syntax elements present in the VPS RBSP syntax structure.

[00150] vps_extension_data_flag may have any value. Its presence and value do not affect decoder conformance to profiles specified in this version of this Specification. Decoders conforming to this version of this Specification shall ignore all vps_extension_data_flag syntax elements.

[00151] The SPS 706 contains data that is common to all the pictures in a sequence of pictures (SOP). The SPS 706 is a syntax structure containing syntax elements that apply to zero or more entire CLVSs as determined by the content of a syntax element found in the PPS referred to by a syntax element found in each picture header. In contrast, the PPS 708 contains data that is common to the entire picture. The PPS 708 is a syntax structure containing syntax elements that apply to zero or more entire coded pictures as determined by a syntax element found in each picture header (e.g., PH 712).

[00152] The DCI 702, the VPS 704, the SPS 706, and the PPS 708 are contained in different types of Network Abstraction Layer (NAL) units. A NAL unit is a syntax structure containing an indication of the type of data to follow (e.g., coded video data). NAL units are classified into video coding layer (VCL) and non-VCL NAL units. The VCL NAL units contain the data that represents the values of the samples in the video pictures, and the non-VCL NAL units contain any associated additional information such as parameter sets (important data that can apply to a number of VCL NAL units) and supplemental enhancement information (timing information and other supplemental data that may enhance usability of the decoded video signal but are not necessary for decoding the values of the samples in the video pictures).

[00153] In an embodiment, the DCI 702 is contained in a non-VCL NAL unit designated as a DCI NAL unit or a DPS NAL unit. That is, the DCI NAL unit has a DCI NAL unit type (NUT) and the DPS NAL unit has a DPS NUT. In an embodiment, the VPS 704 is contained in a non-VCL NAL unit designated as a VPS NAL unit. Therefore, the VPS NAL unit has a VPS NUT. In an embodiment, the SPS 706 is a non-VCL NAL unit designated as a SPS NAL unit. Therefore, the SPS NAL unit has an SPS NUT. In an embodiment, the PPS 708 is contained in a non-VCL NAL unit designated as a PPS NAL unit. Therefore, the PPS NAL unit has a PPS NUT.

[00154] The PH 712 is a syntax structure containing syntax elements that apply to all slices (e.g., slices 718) of a coded picture (e.g., picture 714). In an embodiment, the PH 712 is in a new type of non-VCL NAL unit designated a PH NAL unit. Therefore, the PH NAL unit has a PH NUT (e.g., PH_NUT). In an embodiment, there is one and only one PH 712 included in each PU 701. That is, the PU 701 contains a single or lone PH 712. In an embodiment, exactly one PH NAL unit is present for each picture 701 in the bitstream 700.

[00155] In an embodiment, the PH NAL unit associated with the PH 712 has a temporal ID and a layer ID. The temporal ID identifier indicates the position of the PH NAL unit, in time, relative to the other PH NAL units in the bitstream (e.g., bitstream 701). The layer ID indicates the layer (e.g., layer 531 or layer 532) that contains the PH NAL unit. In an embodiment, the temporal ID is similar to, but different from, the POC. The POC uniquely identifies each picture in order. In a single layer bitstream, temporal ID and POC would be the same. In a multi-layer bitstream (e.g., see FIG. 5), pictures in the same AU would have different POCs, but the same temporal ID.

[00156] In an embodiment, the PH NAL unit precedes the VCL NAL unit containing the first slice 718 of the associated picture 714. This establishes the association between the PH 712 and the slices 718 of the picture 714 associated with the PH 712 without the need to have a picture header ID signaled in the PH 712 and referred to from the slice header 720. Consequently, it can be inferred that all VCL NAL units between two PHs 712 belong to the same picture 714 and that the picture 714 is associated with the first PH 712 between the two PHs 712. In an embodiment, the first VCL NAL unit that follows a PH 712 contains the first slice 718 of the picture 714 associated with the PH 712.

[00157] In an embodiment, the PH NAL unit follows picture level parameter sets (e.g., the PPS) or higher level parameter sets such as the DCI (a.k.a., the DPS), the VPS, the SPS, the PPS,

etc. having both a temporal ID and a layer ID less than the temporal ID and layer ID of the PH NAL unit, respectively. Consequently, those parameter sets are not repeated within a picture or an access unit. Because of this ordering, the PH 712 can be resolved immediately. That is, parameter sets that contain parameters relevant to an entire picture are positioned in the bitstream before the PH NAL unit. Anything that contains parameters for part of a picture is positioned after the PH NAL unit.

[00158] In one alternative, the PH NAL unit follows picture level parameter sets and prefix supplemental enhancement information (SEI) messages, or higher level parameter sets such as the DCI (a.k.a., the DPS), the VPS, the SPS, the PPS, the APS, the SEI message, etc.

[00159] The picture 714 is an array of luma samples in monochrome format or an array of luma samples and two corresponding arrays of chroma samples in 4:2:0, 4:2:2, and 4:4:4 colour format. In an embodiment, there is one and only one picture 714 included in each PU 701. As such, there is only one PH 712 and only one picture 714 corresponding to that PH 712 in each PU 701. That is, the PU 701 contains a single or lone picture 714.

[00160] The picture 714 may be either a frame or a field. However, in one CVS 716, either all pictures 714 are frames or all pictures 714 are fields. The CVS 716 is a coded video sequence for every coded layer video sequence (CLVS) in the video bitstream 700. Notably, the CVS 716 and the CLVS are the same when the video bitstream 700 includes a single layer. The CVS 716 and the CLVS are only different when the video bitstream 700 includes multiple layers (e.g., as shown in FIGS. 5 and 6).

[00161] Each picture 714 contains one or more slices 718. A slice 718 is an integer number of complete tiles or an integer number of consecutive complete CTU rows within a tile of a picture (e.g., picture 714). Each slice 718 is exclusively contained in a single NAL unit (e.g., a VCL NAL unit). A tile (not shown) is a rectangular region of CTUs within a particular tile column and a particular tile row in a picture (e.g., picture 714). A CTU (not shown) is a CTB of luma samples, two corresponding CTBs of chroma samples of a picture that has three sample arrays, or a CTB of samples of a monochrome picture or a picture that is coded using three separate color planes and syntax structures used to code the samples. A CTB (not shown) is an N×N block of samples for some value of N such that the division of a component into CTBs is a partitioning. A block (not shown) is an MxN (M-column by N-row) array of samples (e.g., pixels), or an MxN array of transform coefficients.

[00162] In an embodiment, each slice 718 contains a slice header 720. A slice header 720 is the part of the coded slice 718 containing the data elements pertaining to all tiles or CTU rows within a tile represented in the slice 718. That is, the slice header 720 contains information about the slice 718 such as, for example, the slice type, which of the reference pictures will be used, and so on.

[00163] The pictures 714 and their slices 718 comprise data associated with the images or video being encoded or decoded. Thus, the pictures 714 and their slices 718 may be simply referred to as the payload or data being carried in the bitstream 700.

[00164] Those skilled in the art will appreciate that the bitstream 700 may contain other parameters and information in practical applications.

[00165] FIG. 8 is an embodiment of a method 800 of decoding implemented by a video decoder (e.g., video decoder 400). The method 800 may be performed after a bitstream has been directly or indirectly received from a video encoder (e.g., video encoder 300). The method 800 improves the decoding process by disallowing unnecessary layers in multi-layer video bitstreams. That is, every layer included in an output layer set (OLS) is either an output layer or used as a direct or indirect reference layer of an output layer. This avoids having extraneous information in the coding process and increases coding efficiency. Thus, the coder / decoder (a.k.a., "codec") in video coding is improved relative to current codecs. As a practical matter, the improved video coding process offers the user a better user experience when videos are sent, received, and/or viewed.

[00166] In block 802, the video decoder receives a video bitstream including a VPS (e.g., VPS 704) and a plurality of layers (e.g., layer N 631, layer N+1 632, etc.). In an embodiment, no layer is neither an output layer of at least one OLS (e.g., OLS 1, OLS 2, etc.) nor a direct reference layer of any other layer. That is, each layer of the at least one OLS is the output layer (e.g., Layer N+2 633 in OLS 3) or the direct reference layer of the any other layer (e.g., Layer N+1 632 and Layer N 631 in OLS 3). Any layer that is not an output layer or a layer that is directly referenced by another layer is unnecessary and, therefore, excluded from the CVS (e.g., 690) and/or the video bitstream.

[00167] In an embodiment, the video decoder expects no layer to be neither an output layer of at least one output layer set (OLS) nor a direct reference layer of any other layer as described above based on VVC or some other standard. If, however, the decoder determines that this condition is not true, the decoder may detect an error, signal an error, request that a revised

bitstream (or a portion thereof) be resent, or take some other corrective measures to ensure that a conforming bitstream is received.

[00168] In an embodiment, the VPS includes a layer used as a reference flag (e.g., LayerUsedAsRefLayerFlag) and a layer used as an output layer flag (e.g., LayerUsedAsOutputLayerFlag), and wherein a value (e.g., [i]) of the layer used as a reference flag and a value (e.g., [i]) of the layer used as an output layer flag are not both zero.

[00169] In an embodiment, each layer in the plurality of layers comprises a set of video coding layer (VCL) network abstraction layer (NAL) units that all have a particular value of a layer identifier (ID) and associated non-VCL NAL units. In an embodiment, the at least one OLS contains one or more output layers. In an embodiment, for each of the plurality of layers with a particular value of layer ID specified in the VPS, one of the layers in the at least one OLS shall also have the particular value of layer ID.

[00170] In block 804, the video decoder decodes a picture (e.g., picture 615) from one of the plurality of layers. In an embodiment, the picture is included in an output layer of the at least one OLS. In an embodiment, the method 800 further comprises selecting an output layer from the at least one OLS prior to the decoding. In an embodiment, the method 800 further includes selecting the picture from the output layer after the output layer has been selected.

[00171] Once the picture has been decoded, the picture may be used to generate or produce an image or video sequence for display to a user on the display or screen of an electronic device (e.g., a smart phone, tablet, laptop, personal computer, etc.).

[00172] FIG. 9 is an embodiment of a method 900 of encoding a video bitstream implemented by a video encoder (e.g., video encoder 300). The method 900 may be performed when a picture (e.g., from a video) is to be encoded into a video bitstream and then transmitted toward a video decoder (e.g., video decoder 400). The method 900 improves the decoding process by disallowing unnecessary layers in multi-layer video bitstreams. That is, every layer included in an output layer set (OLS) is either an output layer or used as a direct or indirect reference layer of an output layer. This avoids having extraneous information in the coding process and increases coding efficiency. Thus, the coder / decoder (a.k.a., “codec”) in video coding is improved relative to current codecs. As a practical matter, the improved video coding process offers the user a better user experience when videos are sent, received, and/or viewed.

[00173] In block 902, the video encoder generates a plurality of layers (e.g., layer N 631, layer N+1 632, etc.) and a VPS (e.g., VPS 704) specifying one or more OLSs (e.g., OLS 1, OLS

2, etc.). In an embodiment, each layer from the plurality of layers is included in at least one of the OLSs specified by the VPS. In an embodiment, no layer is neither an output layer of at least one OLS (e.g., OLS 1, OLS 2, etc.) nor a direct reference layer of any other layer. That is, each layer of the at least one OLS is the output layer (e.g., Layer N+2 633 in OLS 3) or the direct reference layer of the any other layer (e.g., Layer N+1 632 and Layer N 631 in OLS 3). Any layer that is not an output layer or a layer that is directly referenced by another layer is unnecessary and, therefore, excluded from the CVS (e.g., 690) and/or the video bitstream. In an embodiment, the video encoder is constrained so that no layer is neither an output layer of at least one output layer sets (OLS) nor a direct reference layer of any other layer. That is, the video encoder is required to encode such that no layer is neither an output layer of at least one output layer sets (OLS) nor a direct reference layer of any other layer. Such a constraint or requirement ensures that the bitstream conforms with, for example, VVC or some other standard.

[00174] In an embodiment, each of the one or more OLSs contains one or more output layers, and wherein each of the output layers comprises one or more pictures. In an embodiment, each layer in the plurality of layers comprises a set of video coding layer (VCL) network abstraction layer (NAL) units that all have a particular value of a layer identifier (ID) and associated non-VCL NAL units.

[00175] In an embodiment, one of the OLSs comprises two output layers, and wherein one of the two output layers references another of the two output layers. In an embodiment, wherein the VPS includes a layer used as a reference flag and a layer used as an output layer flag, and wherein a value of the layer used as a reference flag and a value of the layer used as an output layer flag are not both zero.

[00176] In an embodiment, a hypothetical reference decoder (HRD) disposed in the encoder will check to ensure that every layer included in an output layer set (OLS) is either an output layer or used as a direct or indirect reference layer of an output layer. When the HRD finds a layer that is unnecessary as described herein, then the HRD will return a conformance test error. That is, the HRD conformance test ensures that there are no unnecessary layers. Hence, the encoder encodes according to the no unused layer requirement, but the HRD enforces this requirement.

[00177] In block 904, the video encoder encodes the plurality of layers and the VPS into a video bitstream. In block 906, the video encoder stores the video bitstream for communication toward a video decoder. The video bitstream may be stored in memory until the video bitstream

is transmitted toward the video decoder. Once received by the video decoder, the encoded video bitstream may be decoded (e.g., as described above) to generate or produce an image or video sequence for display to a user on the display or screen of an electronic device (e.g., a smart phone, tablet, laptop, personal computer, etc.).

[00178] FIG. 10 is a schematic diagram of a video coding device 1000 (e.g., a video encoder 300 or a video decoder 400) according to an embodiment of the disclosure. The video coding device 1000 is suitable for implementing the disclosed embodiments as described herein. The video coding device 1000 comprises ingress ports 1010 and receiver units (Rx) 1020 for receiving data; a processor, logic unit, or central processing unit (CPU) 1030 to process the data; transmitter units (Tx) 1040 and egress ports 1050 for transmitting the data; and a memory 1060 for storing the data. The video coding device 1000 may also comprise optical-to-electrical (OE) components and electrical-to-optical (EO) components coupled to the ingress ports 1010, the receiver units 1020, the transmitter units 1040, and the egress ports 1050 for egress or ingress of optical or electrical signals.

[00179] The processor 1030 is implemented by hardware and software. The processor 1030 may be implemented as one or more CPU chips, cores (e.g., as a multi-core processor), field-programmable gate arrays (FPGAs), application specific integrated circuits (ASICs), and digital signal processors (DSPs). The processor 1030 is in communication with the ingress ports 1010, receiver units 1020, transmitter units 1040, egress ports 1050, and memory 1060. The processor 1030 comprises a coding module 1070. The coding module 1070 implements the disclosed embodiments described above. For instance, the coding module 1070 implements, processes, prepares, or provides the various codec functions. The inclusion of the coding module 1070 therefore provides a substantial improvement to the functionality of the video coding device 1000 and effects a transformation of the video coding device 1000 to a different state. Alternatively, the coding module 1070 is implemented as instructions stored in the memory 1060 and executed by the processor 1030.

[00180] The video coding device 1000 may also include input and/or output (I/O) devices 1080 for communicating data to and from a user. The I/O devices 1080 may include output devices such as a display for displaying video data, speakers for outputting audio data, etc. The I/O devices 1080 may also include input devices, such as a keyboard, mouse, trackball, etc., and/or corresponding interfaces for interacting with such output devices.

[00181] The memory 1060 comprises one or more disks, tape drives, and solid-state drives and may be used as an over-flow data storage device, to store programs when such programs are selected for execution, and to store instructions and data that are read during program execution. The memory 1060 may be volatile and/or non-volatile and may be read-only memory (ROM), random access memory (RAM), ternary content-addressable memory (TCAM), and/or static random-access memory (SRAM).

[00182] FIG. 11 is a schematic diagram of an embodiment of a means for coding 1100. In an embodiment, the means for coding 1100 is implemented in a video coding device 1102 (e.g., a video encoder 300 or a video decoder 400). The video coding device 1102 includes receiving means 1101. The receiving means 1101 is configured to receive a picture to encode or to receive a bitstream to decode. The video coding device 1102 includes transmission means 1107 coupled to the receiving means 1101. The transmission means 1107 is configured to transmit the bitstream to a decoder or to transmit a decoded image to a display means (e.g., one of the I/O devices 1080).

[00183] The video coding device 1102 includes a storage means 1103. The storage means 1103 is coupled to at least one of the receiving means 1101 or the transmission means 1107. The storage means 1103 is configured to store instructions. The video coding device 1102 also includes processing means 1105. The processing means 1105 is coupled to the storage means 1103. The processing means 1105 is configured to execute the instructions stored in the storage means 1103 to perform the methods disclosed herein.

[00184] It should also be understood that the steps of the exemplary methods set forth herein are not necessarily required to be performed in the order described, and the order of the steps of such methods should be understood to be merely exemplary. Likewise, additional steps may be included in such methods, and certain steps may be omitted or combined, in methods consistent with various embodiments of the present disclosure.

[00185] While several embodiments have been provided in the present disclosure, it should be understood that the disclosed systems and methods might be embodied in many other specific forms without departing from the spirit or scope of the present disclosure. The present examples are to be considered as illustrative and not restrictive, and the intention is not to be limited to the details given herein. For example, the various elements or components may be combined or integrated in another system or certain features may be omitted, or not implemented.

[00186] In addition, techniques, systems, subsystems, and methods described and illustrated in the various embodiments as discrete or separate may be combined or integrated with other systems, modules, techniques, or methods without departing from the scope of the present disclosure. Other items shown or discussed as coupled or directly coupled or communicating with each other may be indirectly coupled or communicating through some interface, device, or intermediate component whether electrically, mechanically, or otherwise. Other examples of changes, substitutions, and alterations are ascertainable by one skilled in the art and could be made without departing from the spirit and scope disclosed herein.

CLAIMS

1. A method of decoding implemented by a video decoder, comprising:

receiving, by the video decoder, a video bitstream including a video parameter set (VPS) and a plurality of layers, wherein no layer is neither an output layer of at least one output layer set (OLS) nor a direct reference layer of any other layer; and

decoding, by the video decoder, a picture from one of the plurality of layers.

2. The method of claim 1, wherein the picture is included in an output layer of the at least one OLS.

3. The method of any of claims 1 to 2, further comprising selecting an output layer from the at least one OLS prior to the decoding.

4. The method of any of claims 1 to 3, further comprising selecting the picture from the output layer after the output layer has been selected.

5. The method of any of claims 1 to 4, wherein each layer in the plurality of layers comprises a set of video coding layer (VCL) network abstraction layer (NAL) units that all have a particular value of a layer identifier (ID) and associated non-VCL NAL units.

6. A method of encoding implemented by a video encoder, the method comprising:

generating, by the video encoder, a plurality of layers and a video parameter set (VPS) specifying at least one output layer set (OLS), wherein the video encoder is constrained such that no layer is neither an output layer of the at least one OLS nor a direct reference layer of any other layer;

encoding, by the video encoder, the plurality of layers and the VPS into a video bitstream; and

storing, by the video encoder, the video bitstream for communication toward a video decoder.

7. The method of claim 6, wherein each of the one or more OLSs contains one or more output layers, and wherein each of the output layers comprises one or more pictures.

8. A decoding device, comprising:

a receiver configured to receive a video bitstream including a video parameter set (VPS) and a plurality of layers, wherein no layer is neither an output layer of at least one OLS nor a direct reference layer of any other layer;

a memory coupled to the receiver, the memory storing instructions; and

a processor coupled to the memory, the processor configured to execute the instructions to cause the decoding device to decode a picture from one of the plurality of layers to obtain a decoded picture.

9. The decoding device of claim 8, wherein the picture is included in an output layer of the at least one OLS.

10. The decoding device of any of claims 8 to 9, wherein the processor is further configured to select an output layer from the at least one OLS prior to the picture being decoded.

11. The decoding device of any of claims 8 to 10, wherein the processor is further configured to select the picture from the output layer after the output layer has been selected.

12. The decoding device of any of claims 8 to 11, wherein each layer in the plurality of layers comprises a set of video coding layer (VCL) network abstraction layer (NAL) units that all have a particular value of a layer identifier (ID) and associated non-VCL NAL units.

13. An encoding device, comprising:

a memory containing instructions;

a processor coupled to the memory, the processor configured to implement the instructions to cause the encoding device to:

generate a plurality of layers and a video parameter set (VPS) specifying one or more output layer sets (OLSs), wherein the encoding device is constrained such that no

layer is neither an output layer of at least one OLS nor a direct reference layer of any other layer;

encode the plurality of layers and the VPS into a video bitstream; and

a transmitter coupled to the processor, the transmitter configured to transmit the video bitstream toward a video decoder.

14. The encoding device of claim 13, wherein each of the one or more OLSs contains one or more output layers, and wherein each of the output layers comprises one or more pictures.

15. A coding apparatus, comprising:

a receiver configured to receive a picture to encode or to receive a bitstream to decode;

a transmitter coupled to the receiver, the transmitter configured to transmit the bitstream to a decoder or to transmit a decoded image to a display;

a memory coupled to at least one of the receiver or the transmitter, the memory configured to store instructions; and

a processor coupled to the memory, the processor configured to execute the instructions stored in the memory to perform the method in any of claims 1 to 5 and any of claims 6 to 7.