other reference pictures. Pictures may be referred to as frames, and reference pictures may be referred to as reference frames.

[0005] Spatial or temporal prediction utilizes a predictive block. 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.

[0006] A given coded video sequence encoded to a bitstream includes an ordered sequence of coded pictures. In the H.264/AVC and HEVC standards, the decoding order of the coded pictures for a bitstream is equivalent to the ordered sequence. However, the standards also support an output order of decoded pictures that differs from the decoding order, and in such cases each of the coded pictures is associated with a picture order count (POC) value that specifies the output order for the picture in the video sequence.

[0007] Video timing information for a video sequence may be signaled in syntax elements of one or more syntax structures (alternatively referred to as "parameter set structures" or simply "parameter sets"). The syntax structures may include a sequence parameter set (SPS) that includes coding information that applied to all slices of a coded video sequence. The SPS may itself include parameters referred to as video usability information (VUI), which include hypothetical reference decoder (HRD) information as well as information for enhancing the use of the corresponding video sequence for various purposes. The HRD information may itself be signaled using a HRD syntax structure includable within other syntax structures such as the VUI syntax structure. The syntax structures may also include a video parameter set (VPS) that describes characteristics of a corresponding video sequence, such as common syntax elements shared by multiple layers or operation points as well as other operation point information that may be common to multiple sequence parameter sets, such as HRD information for various layers or sub-layers.

VUI part of the SPS or references references the syntax element timing_info_present_flag of an hrd_parameters( ) syntax structure in the VPS.

[0050] In addition, multiple layers or multiple possible bitstream subsets of a scalable video bitstream may share the common values of time scale and the number of units in a clock tick, which are specified in HEVC WD9 in the syntax elements time_scale and num_units_in_tick of an hrd_parameters( ) syntax structure, which may be repeatedly signaled in the VUI part of the SPS and in the VPS, for instance. Such replication, if present in the bitstream, may result in bit wastage.

[0051] Further, picture order count (POC) values are typically proportional to output times for all layers of a scalable video bitstream if the POC values are proportional to output times for any of the layers of a scalable video bitstream. However, the HEVC WD9 specification does not provide for signaling, in the scalable video bitstream, an indication that POC values are proportional to output times for all layers or all possible bitstream subsets of the scalable video bitstream. Reference to a "layer" of a scalable video bitstream may refer to a scalable layer, a texture view, and/or a depth view, for instance. In addition, although HEVC WD9 specifies that the flag poc_proportional_to_timing_flag is always signaled in the VUI syntax structure of the SPS, the flag poc_proportional_to_timing_flag has no utility if the syntax elements time_scale and num_units_in_tick are not also signaled in the bitstream.

[0052] The techniques of this disclosure may solve one or more of the above problems, as well as provide other improvements, to enable efficient signaling of parameters for HRD operations. Various examples of the techniques are described herein with reference to HEVC WD9 and potential improvements thereto. The solutions apply to any video coding standards, including AVC and HEVC, for example, that include a specification for a video buffering model, though for purposes of illustration the description is specific to the HRD parameters signaling defined in HEVC WD9 and modified in accordance with the techniques of this disclosure.

[0053] FIG. 1 is a block diagram illustrating an example video encoding and decoding system 10 that may utilize the techniques described in this disclosure. As shown in FIG. 1, system 10 includes a source device 12 that generates encoded video data to be decoded at a later time by a destination device 14. Source device 12 and destination device 14 may comprise any of a wide range of devices, including desktop computers, notebook (i.e., laptop) computers, tablet computers, set-top boxes, telephone handsets

that encodes the time scale and the number of units in a clock tick syntax elements at most once in the VPS syntax structure. In some instances, the video decoder 30 may test the encoded bitstream for conformance by decoding the encoded bitstream to determine the time scale and the number of units in a clock tick from a VUI syntax structure of the encoded bitstream that encodes the time scale and the number of units in a clock tick syntax elements at most once in the VUI syntax structure. The time scale and number of units in a clock tick may be signaled other than in an HRD parameters syntax structure incorporated within the VPS and/or VUI syntax structure. Video decoder 30 may use the determined time scale and the determined number of units in a clock tick as input for, e.g., determining CPB underflow or overflow during decoding of encoded pictures included in the encoded bitstream.

[0083] In some examples, the video encoder 20 may generate the encoded bitstream to signal, in a VPS syntax structure for one or more coded video sequences, a flag indicating whether the POC value for each picture in a coded video sequence that is not the first picture in the coded video sequence, in decoding order, is proportional to the output time of the picture relative to the output time of the first picture in the coded video sequence. This indication flag may alternatively be referred to as the POC proportional to timing indication flag. As a result, the video encoder 20 may reduce a number of instances of the indication in the timing information signaled for multiple layers of a coded video sequence and/or for a scalable video bitstream having multiple layers. In some instances, the video encoder 20 may include this flag in the VPS syntax structure only if the time scale and number of units in a clock tick syntax elements are also included. The video encoder 20 may in this way avoid signaling this particular timing information (i.e., whether the POC value for each picture in a coded video sequence that is not the first picture in the coded video sequence, in decoding order, is proportional to the output time of the picture relative to the output time of the first picture in the coded video sequence) if the clock tick information needed for using the POC proportional to timing indication is not also present.

[0084] According to techniques described herein, the video encoder 20 may test, for conformance, an encoded bitstream generated by video encoder 20 to signal, in a VPS syntax structure for one or more coded video sequences, the POC proportional to timing indication flag. The video encoder 20 may test the encoded bitstream for conformance by decoding the encoded bitstream to determine a value for the flag. The video encoder 20 may additionally, or alternatively, test an encoded bitstream, generated by video

encodes syntax elements for the number of units in a clock tick and the time scale directly, and at most once, to a VPS syntax structure for the coded video sequence (302). In some instances, even in cases in which the VPS syntax structure includes multiple instances of HRD parameters, by encoding the syntax elements directly to the VPS syntax structure (at most once) and not to the HRD parameter sets (or any other incorporated parameter set syntax structure), the VPS syntax structure may include a single syntax element for each of the number of units in a clock tick and the time scale. The video encoder 20 outputs the coded video sequence and the VPS syntax structure for the coded video sequence (304). In some cases, the video encoder 20 outputs these structures to a HRD of the video encoder 20.

[0142] In FIG. 6B, a video encoder 20 encodes pictures of a video sequence to generate a coded video sequence (310). The video encoder 20 additionally generates parameter sets for the coded video sequence. The parameter sets may include parameters encoded according to a video parameter set (SPS) syntax structure. According to techniques described herein, the video encoder 20 encodes syntax elements for the number of units in a clock tick and the time scale directly, and at most once, to a SPS syntax structure for the coded video sequence (312). In some instances, even in cases in which the SPS syntax structure includes multiple instances of HRD parameters, by encoding the syntax elements directly to the SPS syntax structure (at most once) and not to the HRD parameter sets (or any other incorporated parameter set syntax structure), the SPS syntax structure may include a single syntax element for each of the number of units in a clock tick and the time scale. The video encoder 20 outputs the coded video sequence and the SPS syntax structure for the coded video sequence (314). In some cases, the video encoder 20 outputs these structures to a HRD of the video encoder 20. In some instances, video encoder 20 may encode the syntax elements for the number of units in a clock tick and the time scale to both a VPS syntax structure and an SPS syntax structure for the coded video sequence.

[0143] FIG. 7 is a flowchart illustrating an example method of operation according to techniques described in this disclosure. A video encoder 20 encodes pictures of a video sequence to generate a coded video sequence (400). The video encoder 20 additionally generates parameter sets for the coded video sequence. The parameter sets may include parameters encoded according to a video parameter set (VPS) syntax structure. If timing information is to be included, e.g., for defining a HRD buffering model (YES branch of 402), the video encoder 20 encodes, directly to a VPS syntax structure for the

a difference of picture order count (POC) values equal to one (502). The condition may include one or more syntax elements that represent variables for a Boolean formula, in which case the decoder may process each such syntax element directly from the VPS syntax structure and/or SPS syntax structure of the coded video sequence.

[0149] The decoder additionally processes the VPS syntax structure and/or the SPS syntax structure to extract syntax elements for the number of units in a clock tick and the time scale directly from the VPS syntax structure and/or directly from the SPS syntax structure for the coded video sequence (504). The decoder may then verify conformance of the coded video sequence to a video buffering model that is defined, at least in part, by the values for the condition, the number of units in a clock tick, and the time scale as extracted from the VPS syntax structure and/or the SPS syntax structure and as read from the corresponding syntax elements (506).

[0150] FIGS. 9A–9B are flowcharts illustrating example methods of operation according to techniques described in this disclosure. In FIG. 9A, a video decoder device 30 or hypothetical reference decoder 57 of a video encoder device 20 (hereinafter, "the decoder") receives a coded video sequence and a video parameter set (VPS) syntax structure for the coded video sequence (600). The coded video sequence and/or the VPS syntax structure may be encoded to a bitstream that includes one or more encoded pictures.

[0151] According to techniques described herein, the decoder processes the VPS syntax structure to extract syntax elements for the number of units in a clock tick and the time scale that occur directly, and at most once, in the VPS syntax structure for the coded video sequence (602). The decoder may then verify conformance of the coded video sequence to a video buffering model that is defined, at least in part, by the values for the number of units in a clock tick and the time scale as extracted from the VPS syntax structure and as read from the corresponding syntax elements (604).

[0152] In FIG. 9B, the decoder receives a coded video sequence and a video parameter set (SPS) syntax structure for the coded video sequence (610). The coded video sequence and/or the SPS syntax structures may be encoded to a bitstream that includes one or more encoded pictures.

[0153] According to techniques described herein, the decoder processes the SPS syntax structure to extract syntax elements for the number of units in a clock tick and the time scale that occur directly, and at most once, in the SPS syntax structure for the coded video sequence (612). The decoder may then verify conformance of the coded video

CLAIMS

1. A method of processing video data, the method comprising:

receiving (700) a coded video sequence comprising encoded pictures of a video sequence; and

receiving (702) timing parameters for the coded video sequence that include an indication of whether a picture order count, POC, value for each picture in the coded video sequence that is not a first picture in the coded video sequence according to a decoding order is proportional to an output time of the picture relative to an output time of the first picture in the coded video sequence in a video parameter set, VPS, syntax structure referenced by the coded video sequence.

2. A method of encoding video data, the method comprising:

encoding (400) pictures of a video sequence to generate a coded video sequence comprising the encoded pictures; and

signaling (400,410) timing parameters for the coded video sequence by signaling an indication of whether a picture order count, POC, value for each picture in the coded video sequence that is not a first picture in the coded video sequence according to a decoding order is proportional to an output time of the picture relative to an output time of the first picture in the coded video sequence in a video parameter set, VPS, syntax structure referenced by the coded video sequence.

3. The method of claim 1 or 2, wherein the indication comprises a vps_poe_proportional_to_timing_flag syntax element.

4. The method of claim 1 or 2, wherein signaling the timing parameters for the coded video sequence further comprises:

only if the indication indicates the POC value for each picture in the coded video sequence that is not the first picture in the coded video sequence according to the decoding order is proportional to the output time of the picture relative to the output time of the first picture in the coded video sequence, respectively receiving or signaling a number of clock ticks that correspond to a difference of POC values equal to one in the VPS syntax structure.

5. The method of claim 1 or 2,

wherein respectively receiving or signaling the indication comprises respectively receiving or signaling the indication only if syntax elements for a time scale and a number of units in a clock tick are present in the VPS syntax structure.

6. The method of claim 1 or 2, wherein respectively receiving or signaling the timing parameters for the coded video sequence further comprises:

respectively receiving or signaling an indication in the VPS syntax structure of whether syntax elements for a time scale and a number of units in a clock tick are present in the VPS syntax structure.

7. The method of claim 6, wherein the indication in the VPS syntax structure of whether syntax elements for the time scale and the number of units in a clock tick are present in the VPS syntax structure comprises a vps_timing_info_present_flag syntax element.

8. The method of claim 1 or 2, wherein respectively receiving or signaling the timing parameters for the coded video sequence further comprises:

only if the POC value for each picture in the coded video sequence that is not the frrst picture in the coded video sequence according to the decoding order is proportional to the output time of the picture relative to the output time of the first picture in the coded video sequence, respectively receiving or signaling a number of clock ticks that correspond to a difference of POC values equal to one in a video usability information (VUI) part of a sequence parameter set (SPS) syntax structure referenced by the coded video sequence.

9. The method of claim 1 or 2, wherein the indication comprises a first indication, and wherein respectively receiving or signaling the timing parameters for the coded video sequence further comprises:

only if syntax elements for a time scale and a number of units in a clock tick are present in a video usability information, VUI, part of a sequence parameter set, SPS, syntax structure referenced by the coded video sequence, respectively receiving or signaling a second indication of whether the POC value for each picture in the coded video sequence that is not the first picture in the coded video sequence according to the decoding order is proportional to the output time of the picture relative to the output time of the first picture in the coded video sequence in the VUI part of the SPS syntax structure.

10. The method of claim 9, wherein the second indication comprises a sps_poe_proportional_to_timing_flag syntax element.

11. The method of claim 1 or 2, wherein respectively receiving or signaling the timing parameters for the coded video sequence further comprises:

respectively receiving or signaling an indication in a video usability information, VUI, part of a sequence parameter set, SPS, syntax structure referenced by the coded video sequence of whether syntax elements for a time scale and a number of units in a clock tick are present in the VUI part of the SPS syntax structure.

12. The method of claim 11, wherein the indication in the VUI part of SPS syntax structure referenced by the coded video sequence of whether syntax elements for the time scale and the number of units in a clock tick are present in the VUI part of the SPS syntax structure comprises a sps_timing_info_present_flag syntax element.

13. A non-transitory computer readable storage medium storing instructions for processing video data that upon execution by one or more processors cause the one or more processors to carry out the method of any of claims 1 to 12.

14. A device for processing video data comprising:

means for receiving a coded video sequence comprising encoded pictures of a video sequence; and

means for receiving timing parameters for the coded video sequence that include an indication of whether a picture order count, POC, value for each picture in the coded video sequence that is not a first picture in the coded video sequence according to a decoding order is proportional to an output time of the picture relative to an output time of the first picture in the coded video sequence in a video parameter set, VPS, syntax structure referenced by the coded video sequence.

15. A device for encoding video data, the device comprising:

means for encoding pictures of a video sequence to generate a coded video sequence comprising the encoded pictures; and

means for signaling timing parameters for the coded video sequence by signaling an indication of whether a picture order count, POC, value for each picture in the coded video sequence that is not a first picture in the coded video sequence according to a decoding order is proportional to an output time of the picture relative to an output time of the first picture in the coded video sequence in a video parameter set, VPS, syntax structure referenced by the coded video sequence.