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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 April 2014 (10.04.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/053085 A1

(51) International Patent Classification:
H04N 7/32 (2006.01)

(21) International Application Number:
PCT/CN2013/082569

(22) International Filing Date:
29 August 2013 (29.08.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/744,890 3 October 2012 (03.10.2012) US
61/714,359 16 October 2012 (16.10.2012) US

(71) Applicant: **MEDIATEK INC.** [CN/CN]; No. 1, Dusing Road 1st, Science-Based Industrial Park, Hsin-Chu, Taiwan 300 (CN).

(72) Inventors: **FU, Chih-Ming**; No.115, Aly.31, Ln.647,Sec.6, Zhonghua Rd., Xiangshan Dist., Hsinchu City, Taiwan 300 (CN). **CHEN, Yi-Wen**; 8F., No.66, Ln.211, Yongkang Rd., Fengyuan Dist., Taichung City, Taiwan 420 (CN). **LIN, Jian-Liang**; No.133, Neipi Rd., Nanjian Vil. Su'ao Township, Yilan County, Taiwan 270 (CN). **HUANG, Yu-Wen**; 8F, No. 23, Ln. 298, Longjiang Rd., Zhongshan Dist., Taipei City, Taiwan 104 (CN).

(74) Agent: **BEIJING SANYOU INTELLECTUAL PROPERTY AGENCY LTD.**; 16th Fl., Block A, Corporate Square, No.35 Jinrong Street, Beijing 100033 (CN).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

[Continued on next page]

(54) Title: METHOD AND APPARATUS OF MOTION INFORMATION MANAGEMENT IN VIDEO CODING

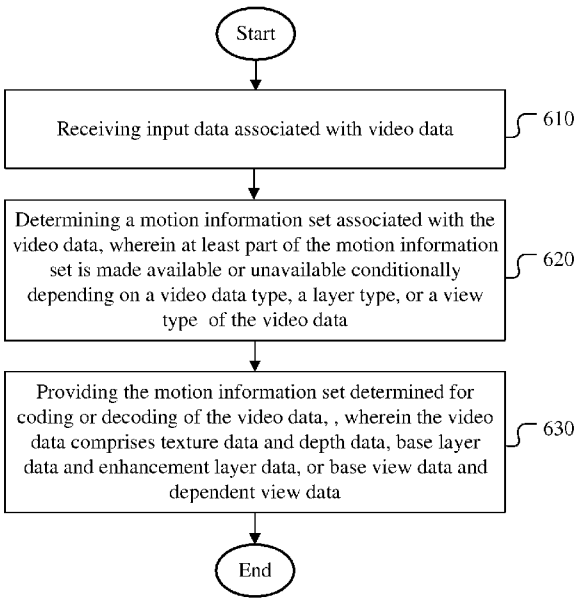

Fig. 6

(57) Abstract: A method and apparatus for three-dimensional and scalable video coding are disclosed. Embodiments according to the present invention determine a motion information set associated with the video data, wherein at least part of the motion information set is made available or unavailable conditionally depending on the video data type. The video data type may correspond to depth data, texture data, a view associated with the video data in three-dimensional video coding, or a layer associated with the video data in scalable video coding. The motion information set is then provided for coding or decoding of the video data, other video data, or both. At least a flag may be used to indicate whether part of the motion information set is available or unavailable. Alternatively, a coding profile for the video data may be used to determine whether the motion information is available or not based on the video data type.

WO 2014/053085 A1

Declarations under Rule 4.17:

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

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

METHOD AND APPARATUS OF MOTION INFORMATION MANAGEMENT IN VIDEO CODING

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] The present invention claims priority to U.S. Provisional Patent Application, Serial No. 61/744,890, filed October 3, 2012, entitled "Motion Information Management for Video Coding" and U.S. Provisional Patent Application, Serial No. 61/714,359, filed October 16, 2012, entitled "Motion Information Management for Video Coding". These U.S. Provisional Patent Applications are hereby incorporated by reference in their entireties.

TECHNICAL FIELD

[0002] The present invention relates to three-dimensional video coding (3DVC) and scalable video coding (SVC). In particular, the present invention relates to motion information management associated with temporal motion vector prediction (TMVP) in order to reduce the size of required buffer.

BACKGROUND

[0003] In advanced video coding, such as High Efficiency Video Coding (HEVC), temporal motion parameter (e.g. motion vectors (MVs), reference index, prediction mode) is used for MV prediction. Therefore, the motion parameters from previous pictures need to be stored in a motion parameter buffer. However, the size of motion parameter buffer may become quite significant since the granularity of motion representation can be as small as 4x4. There are two motion vectors per prediction unit (PU) that need to be stored for B-slices (bi-predicted slice). On the other hand, as the picture size continues to grow, the memory issue becomes even worse since not only more motion vectors need to store, but also more bits per vector need to use for representing the motion vector. For example, the estimated storage for MVs is approximately 26 Mbits/picture for video with picture size 4k by 2k and the actual size will depend on the precision and maximum MVs supported.

[0004] In order to reduce the size of the motion parameter buffer, a compression technique for motion parameters is being used in systems based on high efficiency video coding (HEVC), which stores the coded motion information from previous pictures at lower spatial resolution. It uses

decimation to reduce the number of motion vectors to be stored. The decimated motion vectors are associated with larger granularity instead of 4x4. The compression process for motion parameter buffer replaces the coded motion vector buffer with a reduced buffer to store motion vectors corresponding to lower spatial resolution (i.e., larger granularity). Each compressed vector is calculated as component-wise decimation.

[0005] In HEVC, the motion information compression is achieved using a decimation method as shown in Fig. 1, where each small square block consists of 4x4 pixels. In this example, the motion information compression is performed for each region consisting of 16x16 pixels (as indicated by a thick box). A representative block as indicated by a shaded area is selected and all the blocks within each 16x16 region share the same motion vectors, reference picture indices and prediction mode of the representative block. In Fig. 1, the top left 4x4 block is used as the representative block for the whole 16x16 region. In other words, 16 blocks share the same motion information. Accordingly, a 16:1 motion information compression is achieved in this example.

[0006] Three-dimensional (3D) video coding is developed for encoding/decoding video of multiple views simultaneously captured by cameras corresponding to different views. Since all cameras capture the same scene from different viewpoints, a multi-view video contains a large amount of inter-view redundancy. In order to share the previously encoded texture information of adjacent views, disparity-compensated prediction (DCP) has been added as an alternative to motion-compensated prediction (MCP). MCP refers to inter-picture prediction that uses previously coded pictures of the same view, while DCP refers to inter-picture prediction that uses previously coded pictures of other views in the same access unit. Fig. 2 illustrates an example of 3D video coding system incorporating MCP and DCP. The vector (210) used for DCP is termed as disparity vector (DV), which is analog to the motion vector (MV) used in MCP. Fig. 2 illustrates an example of three MVs (220, 230 and 240) associated with MCP. Furthermore, the DV of a DCP block can also be predicted by the disparity vector predictor (DVP) candidate derived from neighboring blocks or the temporal collocated blocks that also use inter-view reference pictures. In HTM3.1 (HEVC based test model version 3.1 for 3D video coding), when deriving an inter-view Merge candidate for Merge/Skip modes, if the motion information of a corresponding block is not available or not valid, the inter-view Merge candidate is replaced by a DV.

[0007] To share the previously coded residual information of adjacent views, the residual signal of the current block (PU) can be predicted by the residual signals of the corresponding blocks in the inter-view pictures as shown in Fig. 3. The corresponding blocks can be located by respective DVs. The video pictures and depth maps corresponding to a particular camera position are indicated by a view identifier (i.e., V0, V1 and V2 in Fig. 3). All video pictures and depth maps that belong to

the same camera position are associated with the same viewId (i.e., view identifier). The view identifiers are used for specifying the coding order within the access units and detecting missing views in error-prone environments. An access unit includes all video pictures and depth maps corresponding to the same time instant. Inside an access unit, the video picture and any associated depth map having viewId equal to 0 are coded first, followed by the video picture and depth map having viewId equal to 1, etc. The view with viewId equal to 0 (i.e., V0 in Fig. 3) is also referred to as *the base view* or the *independent view*. The base view video pictures can be coded using a conventional HEVC video coder without dependence on any other view.

[0008] As can be seen in Fig. 3, motion vector predictor (MVP)/ disparity vector predictor (DVP) for the current block can be derived from the inter-view blocks in the inter-view pictures. In the following, inter-view blocks in inter-view picture may be abbreviated as inter-view blocks. The derived candidates are termed as *inter-view candidates*, which can be inter-view MVPs or DVPs. Furthermore, a corresponding block in a neighboring view is termed as an *inter-view block* and the inter-view block is located using the disparity vector derived from the depth information of current block in current picture.

[0009] As described above, DV is critical in 3D video coding for disparity vector prediction, inter-view motion prediction, inter-view residual prediction, and disparity-compensated prediction (DCP) or any other coding tool that needs to indicate the correspondence between inter-view pictures.

[0010] Compressed digital video has been widely used in various applications such as video streaming over digital networks and video transmission over digital channels. Very often, a single video content may be delivered over networks with different characteristics. For example, a live sport event may be carried in a high-bandwidth streaming format over broadband networks for premium video services. In such applications, the compressed video usually preserves high resolution and high quality so that the video content is suited for high-definition devices such as an HDTV or a high resolution LCD display. The same contents may also be carried through cellular data network so that the contents can be watched on a portable device such as a smart phone or a network-connected portable media device. In such applications, due to the network bandwidth concerns as well as the typically lower resolution display on the smart phone or portable device, the video content usually is compressed into lower resolution and lower bitrates. Therefore, for different network environment and for different applications, the video resolution and video quality requirements are quite different. Even for the same type of network, users may experience different available bandwidths due to different network infrastructure and network traffic condition. Therefore, a user may desire to receive the video at higher quality when the available bandwidth is

high and receive a lower-quality, but smooth, video when the network congestion occurs. In another scenario, a high-end media player can handle high-resolution and high bitrate compressed video while a low-cost media player is only capable of handling low-resolution and low bitrate compressed video due to limited computational resources. Accordingly, it is desirable to construct the compressed video in a scalable manner so that videos at different spatial-temporal resolutions and/or quality can be derived based on the same compressed bitstream.

[0011] The joint video team (JVT) of ISO/IEC MPEG and ITU-T VCEG standardizes a Scalable Video Coding (SVC) extension of the H.264/AVC standard. An H.264/AVC SVC bitstream may contain video information from low frame-rate, low resolution, and low quality to high frame rate, high definition, and high quality. This single bitstream can be adapted to various applications and displayed on devices with different configurations. Accordingly, H.264/AVC SVC is suitable for various video applications such as video broadcasting, video streaming, and video surveillance to adapt to the network infrastructure, traffic condition, user preference, etc.

[0012] In SVC, three types of scalabilities, i.e., temporal scalability, spatial scalability, and quality scalability, are provided. SVC uses multi-layer coding structure to realize the three dimensions of scalability. A main goal of SVC is to generate one scalable bitstream that can be easily and rapidly adapted to the bit-rate requirement associated with various transmission channels, diverse display capabilities, and different computational resources without trans-coding or re-encoding. An important feature of the SVC design is that the scalability is provided at a bitstream level. In other words, bitstreams for deriving video with a reduced spatial and/or temporal resolution can be simply obtained by extracting Network Abstraction Layer (NAL) units (or network packets) from a scalable bitstream. NAL units for quality refinement can be additionally truncated in order to reduce the bit-rate and the associated video quality.

[0013] For temporal scalability, a video sequence can be hierarchically coded in the temporal domain. For example, temporal scalability can be achieved using hierarchical coding structure based on B-pictures according to the H.264/AVC standard. Fig. 4 illustrates an example of hierarchical B-picture structure with 4 temporal layers and the Group of Pictures (GOP) includes eight pictures. Pictures 0 and 8 in Fig. 4 are called *key pictures*. Inter prediction of key pictures only uses previous key pictures as references. Other pictures between two key pictures are predicted hierarchically. The video having only the key pictures forms the coarsest temporal resolution of the scalable system. Temporal scalability is achieved by progressively refining a lower-level (coarser) video by adding more B pictures corresponding to enhancement layers of the scalable system. In the example of Fig. 4, picture 4 (in the display order) is first bi-directionally predicted using key pictures (i.e., pictures 0 and 8) after the two key pictures are coded. After picture 4 is processed, pictures 2 and 6 are

processed. Picture 2 is bi-directionally predicted using pictures 0 and 4, and picture 6 is bi-directionally predicted using pictures 4 and 8. After pictures 2 and 6 are coded, remaining pictures, i.e., pictures 1, 3, 5 and 7 are processed bi-directionally using two respective neighboring pictures as shown in Fig. 4. Accordingly, the processing order for the GOP is 0, 8, 4, 2, 6, 1, 3, 5, and 7. The pictures processed according to the hierarchical process of Fig. 4 results in hierarchical four-level pictures, where pictures 0 and 8 belong to the first temporal order, picture 4 belongs the second temporal order, pictures 2 and 6 belong to the third temporal order and pictures 1, 3, 5, and 7 belong to the fourth temporal order. By decoding the base level pictures and adding higher temporal order pictures will be able to provide a higher level video. For example, base-level pictures 0 and 8 can be combined with second temporal-order picture 4 to form second-level pictures. By further adding the third temporal-order pictures to the second-level video can form the third-level video. Similarly, by adding the fourth temporal-order pictures to the third-level video can form the fourth-level video. Accordingly, the temporal scalability is achieved. If the original video has a frame rate of 30 frames per second, the base-level video has a frame rate of 30/8 = 3.75 frames per second. The second-level, third-level and fourth-level video correspond to 7.5, 15, and 30 frames per second. The first temporal-order pictures are also called base-level video or based-level pictures. The second temporal-order pictures through fourth temporal-order pictures are also called enhancement-level video or enhancement-level pictures. In addition to enabling temporal scalability, the coding structure of hierarchical B-pictures also improves the coding efficiency over the typical IBBP GOP structure at the cost of increased encoding-decoding delay.

[0014] In SVC, spatial scalability is supported based on the pyramid coding scheme. First, the video sequence is down-sampled to smaller pictures with coarser spatial resolutions (i.e., layers). In addition to dyadic spatial resolution, SVC also supports arbitrary resolution ratios, which is called extended spatial scalability (ESS). In order to improve the coding efficiency of the enhancement layers (video layers with coarser resolutions), the inter-layer prediction schemes are introduced. Three inter-layer prediction tools are adopted in SVC, namely inter-layer motion prediction, inter-layer intra prediction, and inter-layer residual prediction.

[0015] The inter-layer prediction process comprises identifying the collocated block in a lower layer (e.g. BL) based on the location of a corresponding EL block. The collocated lower layer block is then interpolated to generate prediction samples for the EL as shown in Fig. 5. In scalable video coding, the interpolation process is used for inter-layer prediction by using predefined coefficients to generate the prediction samples for the EL based on a lower layer pixels. The example in Fig. 5 consists of two layers. However, an SVC system may consist of more than two layers. The BL picture is formed by applying spatial decimation 510 to the input picture (i.e., an EL picture in

this example). The BL processing comprises BL prediction 520. The BL input is predicted by BL prediction 520, where subtractor 522 is used to form the difference between the BL input data and the BL prediction. The output of subtractor 522 corresponds to the BL prediction residues and the residues are processed by transform/quantization (T/Q) 530 and entropy coding 570 to generate compressed bitstream for the BL. Reconstructed BL data has to be generated at the BL in order to form BL prediction. Accordingly, inverse transform/inverse quantization (IT/IQ) 540 is used to recover the BL residues. The recovered BL residues and the BL prediction data are combined using reconstruction 550 to form reconstructed BL data. The reconstructed BL data is processed by in-loop filter 560 before it is stored in buffers inside the BL prediction. In the BL, BL prediction 520 uses Inter/Intra prediction 521. The EL processing consists of similar processing modules as the BL processing. The EL processing comprises EL prediction 525, subtractor 528, T/Q 535, entropy coding 575, IT/IQ 545, reconstruction 555 and in-loop filter 565. However, the EL prediction also utilizes reconstructed BL data as part of inter-layer prediction. Accordingly, EL prediction 525 comprises inter-layer prediction 527 in addition to Inter/Intra prediction 526. The reconstructed BL data is interpolated using interpolation 512 before it is used for inter-layer prediction. The compressed bitstreams from the BL and the EL are combined using multiplexer 580 to form a scalable bitstream.

[0016] In EL coding unit coding, a flag can be coded to indicate whether the EL motion information is directly derived from the BL. If the flag is equal to 1, the partitioning data of the EL coding unit together with the associated reference indexes and motion vectors are derived from the corresponding data of the collocated block in the BL. The reference picture index of BL is directly used in EL. The coding unit partitioning and motion vectors of EL correspond to the scaled coding unit partitioning and scaled motion vectors of the BL respectively. Besides, the scaled BL motion vector can be used as an additional motion vector predictor for the EL.

[0017] As illustrated in the above discussion, the DV and MV information is used for inter-view predictive coding in three-dimensional video coding systems and inter-layer predictive coding in scalable video coding systems. The DV and MV information may have to be stored for one or more reference pictures or depth maps. Therefore, the amount of storage may be substantial. It is desirable to reduce the required MV/DV information storage.

SUMMARY

[0018] A method and apparatus for three-dimensional and scalable video coding are disclosed. Embodiments according to the present invention determine a motion information set

associated with the video data, wherein at least part of the motion information set is made available or unavailable conditionally depending on the video data type, the layer type, or the view type of the video data. The motion information set determined is then provided for coding or decoding of the current video data, wherein the video data comprises texture data and depth data, base layer data and enhancement layer data, or base view data and dependent view data. At least a flag may be used to indicate whether said at least part of the motion information set is available or unavailable. The flag may be signaled in a sequence level, a picture level, a slice level, a video parameter set or an adaptation parameter set. Alternatively, the flag may be set to a value indicating whether said at least part of the motion information set is available or unavailable according to a coding profile for the video data, wherein the flag corresponds to a syntax element in a sequence level, a picture level, a slice level, a video parameter set or an adaptation parameter set.

[0019] A three-dimensional video coding or a scalable video coding system can be configured according to a coding profile, or a flag in the bitstream to make motion information available or unavailable depending on the video data type. In one embodiment, the temporal motion information is made unavailable for the depth sequence/picture, and the temporal motion information is made available for the texture sequence/picture in a dependent view or a base view. In another embodiment, the temporal motion information is made unavailable for the texture sequence/picture in a dependent view, and the temporal motion information is made available for the texture sequence/picture or the depth sequence/picture in a base view. In yet another embodiment, the temporal motion information is made unavailable for the texture sequence/picture or the depth sequence/picture in a dependent view, and the temporal motion information is made available for the texture sequence/picture in a base view. In yet another embodiment, the temporal motion information is made unavailable for the texture sequence/picture or the depth sequence/picture in an enhancement layer, and the temporal motion information is made available for the texture sequence/picture or the depth sequence/picture in a base layer.

BRIEF DESCRIPTION OF DRAWINGS

[0020] Fig. 1 illustrates an example of motion information compression using motion vector decimation according to a reference implementation of high efficiency video coding (HEVC).

[0021] Fig. 2 illustrates an example of three-dimensional video coding incorporating disparity-compensated prediction (DCP) as an alternative to motion-compensated prediction (MCP).

[0022] Fig. 3 illustrates an example of inter-view motion vector prediction (MVP) candidate and inter-view Merge candidate derivation according to high efficiency video coding (HEVC) based

three-dimensional video coding.

[0023] Fig. 4 illustrates an example of four-level hierarchical B-picture scheme for temporal scalable video coding.

[0024] Fig. 5 illustrates an example of two-layer spatial temporal scalable video coding based on a pyramid coding scheme.

[0025] Fig. 6 illustrates an exemplary flow chart for a three-dimensional video coding system or a scalable video coding system incorporating an embodiment of the present invention to conditionally make temporal motion information available or unavailable depending on the video data type.

DETAILED DESCRIPTION

[0026] In three-dimensional video coding systems and scalable video coding systems, the motion information can be used for depth map coding, dependent view video coding, and enhancement layer video coding. However, in order to utilize the motion information for three-dimensional video coding and scalable video coding, additional buffer may be needed to store the motion information. For each reference picture, the motion information of each block includes motion vector, reference picture index, and inter/intra mode information. For high definition video, the motion information buffer may become significant. For example, a 2048x1080 video sequence will require more than 2 Mbits memory to store the motion information for each picture having a block size of 64x64. Embodiments of the present invention further reduce the memory requirement of motion information at the decoder.

[0027] In 3D video coding, depth coding can refer to motion information of corresponding texture data for 3D video coding. However, since the motion information associated with texture coding usually is more accurate than the motion information associated with the depth. Therefore, motion information from corresponding texture data will be more useful than the motion information from depth coding. Accordingly, the temporal motion information for depth coding becomes less important when texture-based motion information is available. In the present disclosure, *video data* may refer to texture data or depth data. According to one embodiment, reference to the temporal motion information can be disable/enable conditionally according to the video data type. The *video data type* in this disclosure refers to whether the video data is texture data or depth data. For example, a flag, **sps_temporal_mvp_enable_flag** in the sequence level can be used to indicate whether temporal MVP is enabled. If the sequence level flag indicates that the temporal MVP for depth is disabled and the current picture corresponds to a depth map, any slice level syntax related to

temporal MVP can then be removed. Therefore, the buffer related to motion information at both the encoder side and the decoder side can be eliminated to reduce system cost. The slice level syntax related to temporal MVP may include the slice temporal MVP enable flag (e.g., **slice_temporal_mvp_enable_flag**), collocated picture index (e.g., **collocated_ref_idx**), and collocated picture direction (e.g., **collocated_from_l0_flag**). In an embodiment, the flag for enabling temporal MVP for depth data is set as disabled when the flag for enabling temporal MVP for texture data is enabled. The motion information obtained from the texture data is generally more reliable than the motion information obtained from the corresponding depth data. The temporal MVP for depth data may be disabled as the depth data may reuse the motion information obtained from the texture data if temporal MVP is used in coding texture data. In this case, the encoder and decoder systems only require the MV buffer for storing motion information related to temporal MVP for the texture data, where buffer for storing motion information related to temporal MVP for the depth data can be eliminated.

[0028] For 3D video coding, the dependent view coding can refer to the motion information of another view (e.g., base view) corresponding to a same time instance. However, since an inter-view reference picture and a current dependent view may correspond to a scene at the same time instance and different viewing positions, the motion information from inter-view reference picture may be more useful than the temporal motion information of current view. In other words, the temporal motion information for dependent view becomes less important when the motion information of inter-view reference picture is available. The above case is especially true if the depth map for the current view is also available since the depth map for the current view can be used to derive a disparity vector for locating a corresponding block in the inter-view reference picture. According to another embodiment of the present invention, reference to the temporal motion information is disable/enable conditionally based on the view type. The *view type* in this disclosure refers to whether the video data corresponds to a dependent view or a base view. For example, if a current picture being coded corresponds to a texture picture in a dependent view and the disparity vector is also available, the associated motion information can be removed. For example, any slice level syntax related to temporal MVP can be removed. Therefore, the buffer for storing motion information for the dependent view at both the encoder side and the decoder side can be eliminated to reduce system cost. The slice level syntax related to temporal MVP may include the slice temporal MVP enable flag (e.g., **slice_temporal_mvp_enable_flag**), collocated picture index (e.g., **collocated_ref_idx**), and collocated picture direction (e.g., **collocated_from_l0_flag**).

[0029] For scalable video coding, enhancement layer coding can refer to the motion information of the base layer. However, since the base layer and the enhancement layer correspond

to a scene at the same time instance,, the motion information from the base layer may be more useful than the temporal motion information of the enhancement layer. In other words, the temporal motion information for the enhancement layer becomes less important when the motion information of base layer is available especially for SNR scalability. In yet another embodiment of the present invention, reference to the temporal motion information can be enabled or disabled conditionally based on the layer type. The *layer type* in this disclosure refers to whether the video data corresponds to an enhancement layer or a base layer. For example, if current coding picture corresponds to an enhancement layer and the base-layer motion information is also available, the flag will indicate that the temporal MVP is removed conditionally. With temporal motion vector prediction disabled for the enhancement layer, the buffer for storing motion information for the enhancement layer at both the encoder side and the decoder side can be eliminated to reduce system cost. Other related flags, such as **collocated_from_l0_flag** and **collocated_ref_idx** can also be removed conditionally.

[0030] In another embodiment, the size of the candidate list for Merge mode or advanced motion vector prediction (AMVP) is modified according to the number of available motion vector information. For example, for depth coding in HTM version 4.0, the size of the candidate list for Merge mode is 5 or 6 (with MPI included) depending on whether motion parameter inheritance (MPI) for depth MV is included. The size of the MVP candidate list for AMVP is 2. If TMVP is disabled for depth coding according to an embodiment of the present invention, the size of the candidate list for Merge mode will be 4 or 5 (with MPI included) depending on whether MPI is included.

[0031] In addition to the conditional syntax changes to support aforementioned MV buffer reduction, other means can also be used to eliminate the MV buffer. In one embodiment, a coding profile specifying limitation is used to remove the MV buffer. In this case, the values of the **temporal_mvp** related syntax can be limited to indicate MV buffer reduction. For example, the TMVP enable flag in the slice level (i.e., **slice_temporal_mvp_enable_flag**) is set to 0, **collocated_from_l0_flag** can be equal to 0 or 1 by default, and **collocated_ref_idx** is set to 0 or any other default value. Also, the syntax element related to the size of candidate list for Merge mode (i.e., **five_minus_max_num_merge_cand**) is set to 1 or any other default value within a range for a specific profile. Accordingly, the MV buffer can be eliminated while keeping the syntax unchanged.

[0032] To reduce the memory requirement of motion information of decoder, an embodiment according to the present invention uses a sequence level flag (i.e., **sps_temporal_mvp_enable_flag**) to enable or disable TMVP for depth coding, dependent views or enhancement layer coding. In another embodiment, a slice level flag (i.e., **slice_temporal_mvp_enable_flag**) is used to enable or disable TMVP for depth coding, dependent views or enhancement layer coding. The

disabling/enabling of TMVP syntax may be signalled in other places in the bitstream such as picture parameter set (PPS) or any other parameter set (e.g. Video parameter set, Adaptive parameter set). Accordingly, an enable flag, **pps_temporal_mvp_enable_flag** can be signalled at the picture level to enable or disable TMVP for depth coding, dependent views or enhancement layer coding. When the flag is set to disable TMVP, the TMVP for the entire depth map sequence will be set to be unavailable while the TMVP for texture coding is still enabled.

[0033] Fig. 6 illustrates an exemplary flowchart for a three-dimensional video coding system or a scalable video coding system incorporating an embodiment of the present invention to conditionally make the temporal motion information available or unavailable depending on the video data type. The system receives input data associated with video data as shown in step 610. For encoding, the input data may correspond to original texture or depth data of a sequence, a picture, a slice, a largest coding unit (LCU) or a coding unit. For decoding, the input data corresponds to coded texture or depth data for a sequence, a picture, a slice, a largest coding unit (LCU) or a coding unit. The input data may be retrieved from storage such as a computer memory, buffer (RAM or DRAM) or other media. The input data may also be received from a processor such as a controller, a central processing unit, a digital signal processor or electronic circuits that produce the input data. A motion information set associated with the video data in step 620, wherein at least part of the motion information set is made available or unavailable conditionally depending on a video data type, a layer type, or a view type of the video data. For encoding, the motion information set is derived from the input video data. For decoding, the motion information set is determined from the bitstream. The motion information set determined is then provided for coding or decoding of the video data, other video data, or both.

[0034] A three-dimensional video coding or a scalable video coding system can be configured according to a coding profile or a flag in the bitstream to make motion information available or unavailable depending on the video data type. In one embodiment, the temporal motion information is made unavailable for the depth sequence/picture, and the temporal motion information is made available for the texture sequence/picture in a dependent view or a base view. In another embodiment, the temporal motion information is made unavailable for the texture sequence/picture in a dependent view, and the temporal motion information is made available for the texture sequence/picture or the depth sequence/picture in a base view. In yet another embodiment, the temporal motion information is made unavailable for the texture sequence/picture or the depth sequence/picture in a dependent view, and the temporal motion information is made available for the texture sequence/picture in a base view. In yet another embodiment, the temporal motion information is made unavailable for the sequence/picture in an enhancement layer, and the temporal motion

information is made available for the sequence/picture in a base layer.

[0035] The flowchart shown in Fig. 6 is intended to illustrate examples of temporal motion information management for video coding. A person skilled in the art may modify each step, re-arranges the steps, split a step, or combine steps to practice the present invention without departing from the spirit of the present invention.

[0036] The above description is presented to enable a person of ordinary skill in the art to practice the present invention as provided in the context of a particular application and its requirement. Various modifications to the described embodiments will be apparent to those with skill in the art, and the general principles defined herein may be applied to other embodiments. Therefore, the present invention is not intended to be limited to the particular embodiments shown and described, but is to be accorded the widest scope consistent with the principles and novel features herein disclosed. In the above detailed description, various specific details are illustrated in order to provide a thorough understanding of the present invention. Nevertheless, it will be understood by those skilled in the art that the present invention may be practiced.

[0037] Embodiment of the present invention as described above may be implemented in various hardware, software codes, or a combination of both. For example, an embodiment of the present invention can be a circuit integrated into a video compression chip or program code integrated into video compression software to perform the processing described herein. An embodiment of the present invention may also be program code to be executed on a Digital Signal Processor (DSP) to perform the processing described herein. The invention may also involve a number of functions to be performed by a computer processor, a digital signal processor, a microprocessor, or field programmable gate array (FPGA). These processors can be configured to perform particular tasks according to the invention, by executing machine-readable software code or firmware code that defines the particular methods embodied by the invention. The software code or firmware code may be developed in different programming languages and different formats or styles. The software code may also be compiled for different target platforms. However, different code formats, styles and languages of software codes and other means of configuring code to perform the tasks in accordance with the invention will not depart from the spirit and scope of the invention.

[0038] The invention may be embodied in other specific forms without departing from its spirit or essential characteristics. The described examples are to be considered in all respects only as illustrative and not restrictive. The scope of the invention is therefore, indicated by the appended claims rather than by the foregoing description. All changes which come within the meaning and range of equivalency of the claims are to be embraced within their scope.

CLAIMS

1. A method for a three-dimensional video coding system or a scalable video coding system, the method comprising:

receiving input data associated with video data;

determining a motion information set associated with the video data, wherein at least part of the motion information set is made available or unavailable conditionally depending on a video data type, a layer type, or a view type of the video data; and

providing the motion information set determined for coding or decoding of the video data, , wherein the video data comprises texture data and depth data, base layer data and enhancement layer data, or base view data and dependent view data.

2. The method of Claim 1, wherein a flag is used to indicate whether said at least part of the motion information set is available or unavailable.

3. The method of Claim 2, wherein the flag is signaled in a sequence level, a picture level, a slice level, a video parameter set or an adaptation parameter set.

4. The method of Claim 2, wherein the flag is set to a value indicating whether said at least part of the motion information set is available or unavailable according to a coding profile for the video data, wherein the flag corresponds to a syntax element in a sequence level, a picture level, a slice level, a video parameter set or an adaptation parameter set.

5. The method of Claim 1, wherein a sequence level flag is used for the video data corresponding to one sequence, or a picture level flag is used for the video data corresponding to one picture to indicate whether said at least part of the motion information set is available or not.

6. The method of Claim 1, wherein said at least part of the motion information set is made unavailable if the video data corresponds to a depth sequence or a depth picture, and said at least part of the motion information set is made available if the video data corresponds to a texture sequence or a texture picture in one dependent view or one base view, and wherein said at least part of the motion information set comprises temporal motion vector information.

7. The method of Claim 1, wherein said at least part of the motion information set is made unavailable if the video data corresponds to a first texture sequence or picture in one dependent view, and said at least part of the motion information set is made available if the video data corresponds to a second texture sequence or picture, or a depth sequence or picture in a base view, and wherein said at least part of the motion information set comprises temporal motion vector information.

8. The method of Claim 1, wherein said at least part of the motion information set is made unavailable if the video data corresponds to a first texture sequence or picture, or a depth sequence or

picture in one dependent view, and said at least part of the motion information set is made available if the video data corresponds to a second texture sequence or picture in a base view, and wherein said at least part of the motion information set comprises temporal motion vector information.

9. The method of Claim 1, wherein said at least part of the motion information set is made unavailable if the video data corresponds to a first sequence or picture in an enhancement layer, and said at least part of the motion information set is made available if the video data corresponds to a second sequence or picture in a base layer, and wherein said at least part of the motion information set comprises temporal motion vector information.

10. The method of Claim 1, wherein a coding profile is selected for the three-dimensional video coding system or the scalable video coding system, wherein the coding profile configures a syntax element to a desired value, and the syntax element is signaled in a sequence level, a picture level, a slice level, a video parameter set or an adaptation parameter set according to the video data type.

11. The method of Claim 10, wherein said at least part of the motion information set is made unavailable if the video data corresponds to a depth sequence or a depth picture, and said at least part of the motion information set is made available if the video data corresponds to a texture sequence or a texture picture in one dependent view or one base view, and wherein said at least part of the motion information set comprises temporal motion vector information.

12. The method of Claim 1, wherein said at least part of the motion information set is made unavailable for the depth data if said at least part of the motion information set is made available for the texture data, and wherein said at least part of the motion information set comprises temporal motion vector information.

13. The method of Claim 1, wherein candidate-list size associated with Merge mode or advanced motion vector prediction (AMVP) mode is dependent on whether said at least part of the motion information set is made available or unavailable, wherein the Merge mode or the AMVP mode is used to code the motion information set associated with the video data.

14. An apparatus for a three-dimensional video coding system or a scalable video coding system, the apparatus comprising:

means for receiving input data associated with video data;

means for determining a motion information set associated with the video data, wherein at least part of the motion information set is made available or unavailable conditionally depending on a video data type, a layer type, or a view type of the video data; and

means for providing the motion information set determined for coding or decoding of the video data, wherein the video data comprises texture data and depth data, base layer data and enhancement

layer data, or base view data and dependent view data.

15. The apparatus of Claim 14, wherein a flag is used to indicate whether said at least part of the motion information set is available or unavailable.

16. The apparatus of Claim 15, wherein the flag is signaled in a sequence level, a picture level, a slice level, a video parameter set or an adaptation parameter set.

17. The apparatus of Claim 15, wherein the flag is set to a value indicating whether said at least part of the motion information set is available or unavailable according to a coding profile for the video data, wherein the flag corresponds to a syntax element in a sequence level, a picture level, a slice level, a video parameter set or an adaptation parameter set.

18. The apparatus of Claim 14, wherein said at least part of the motion information set is made unavailable if the video data corresponds to a depth sequence or a depth picture, and said at least part of the motion information set is made available if the video data corresponds to a texture sequence or a texture picture in one dependent view or one base view, and wherein said at least part of the motion information set comprises temporal motion vector information.

19. The apparatus of Claim 14, wherein said at least part of the motion information set is made unavailable for the depth data if said at least part of the motion information set is made available for the texture data, and wherein said at least part of the motion information set comprises temporal motion vector information.

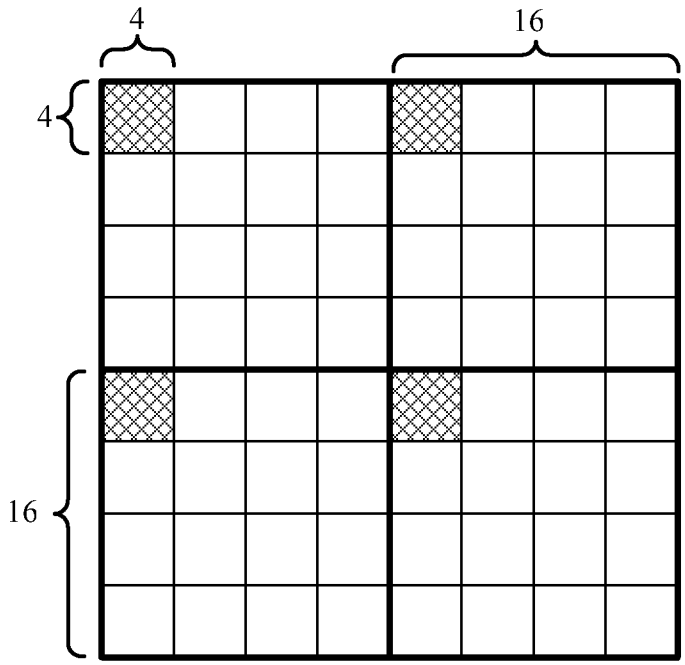

Fig. 1

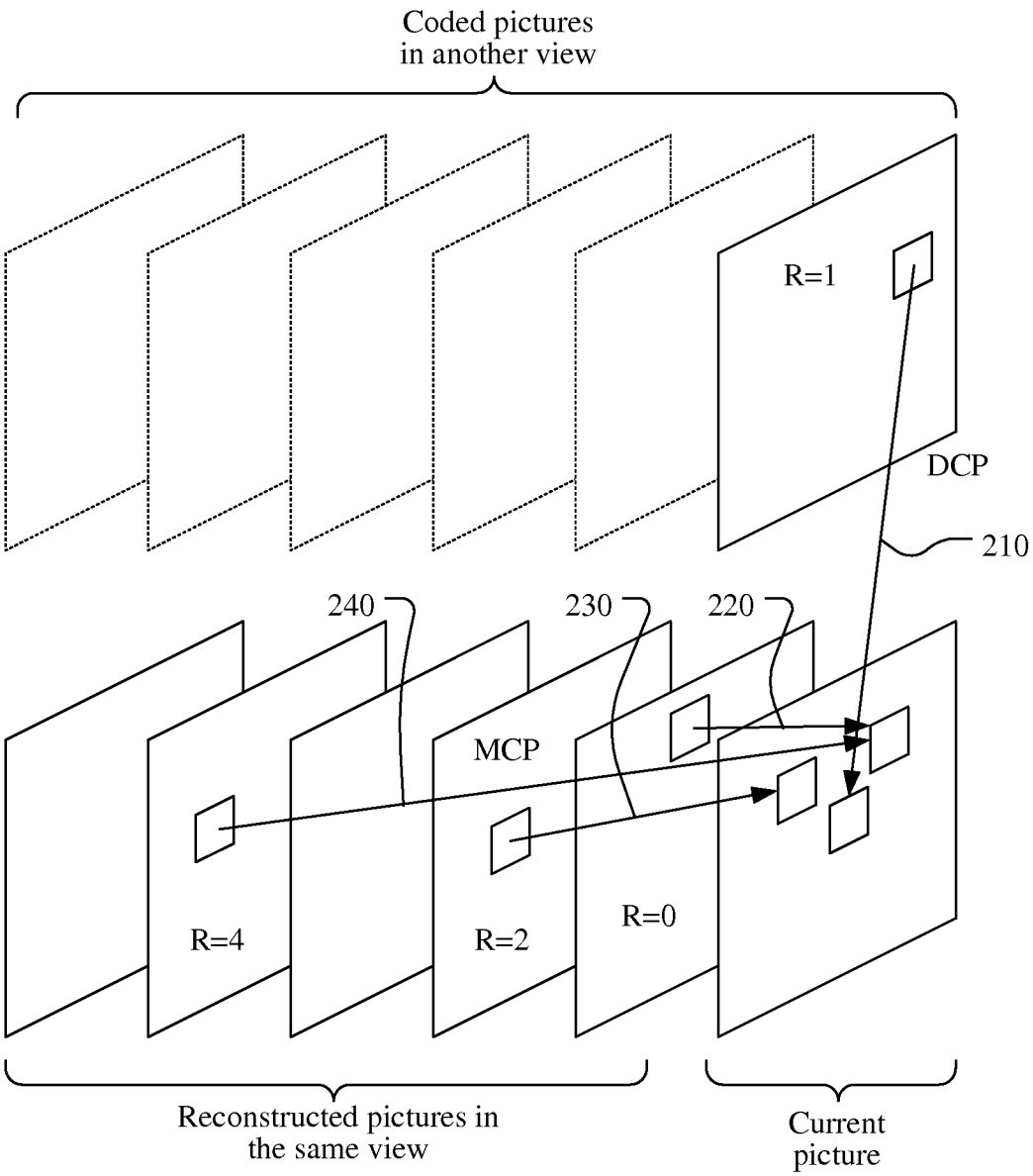
Coded pictures
in another view
R=1
DCP
210
240
230
220
MCP
R=4
R=2
R=0
Reconstructed pictures in
the same view
Current
picture

Fig. 2

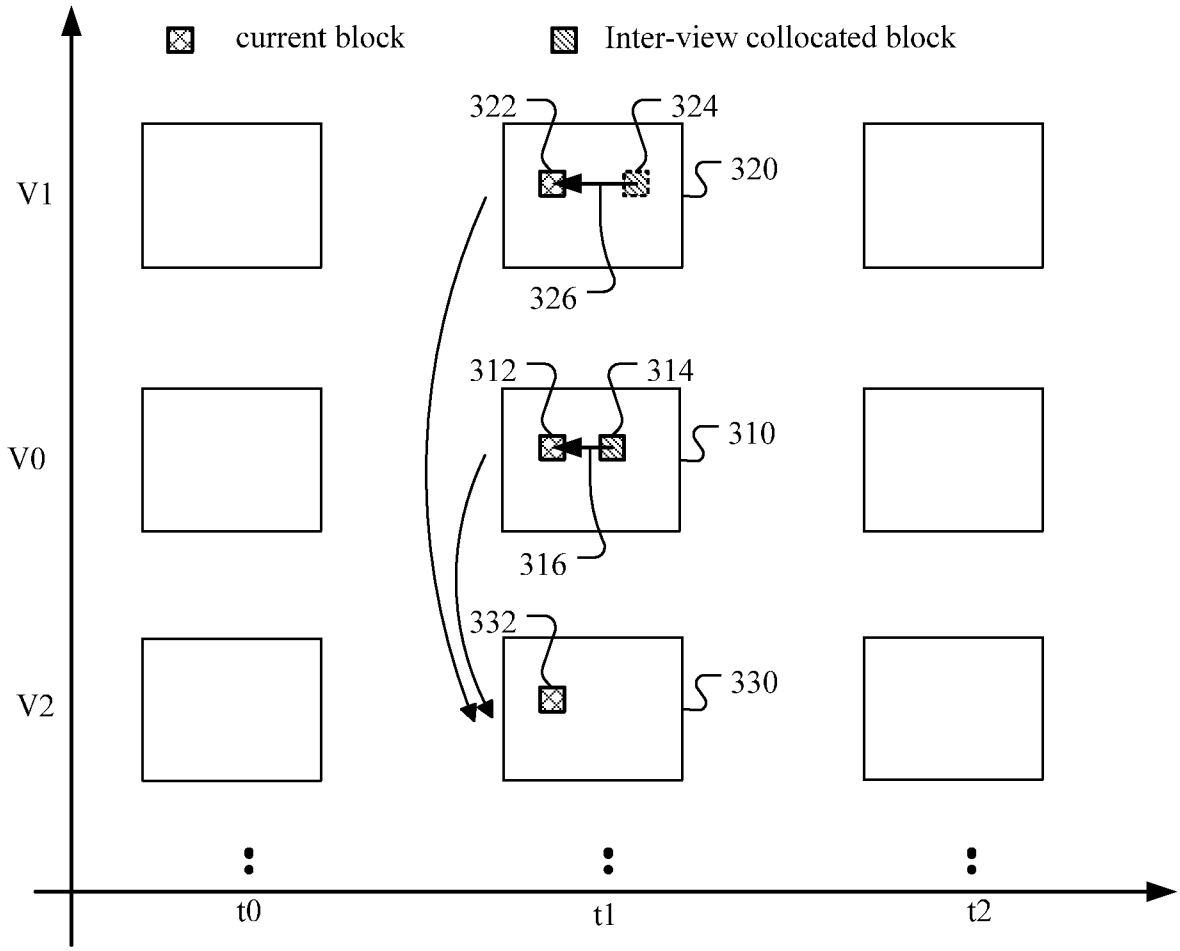
current block
Inter-view collocated block
322
324
320
326
312
314
310
316
332
330
V1
V0
V2
t0
t1
t2

Fig. 3

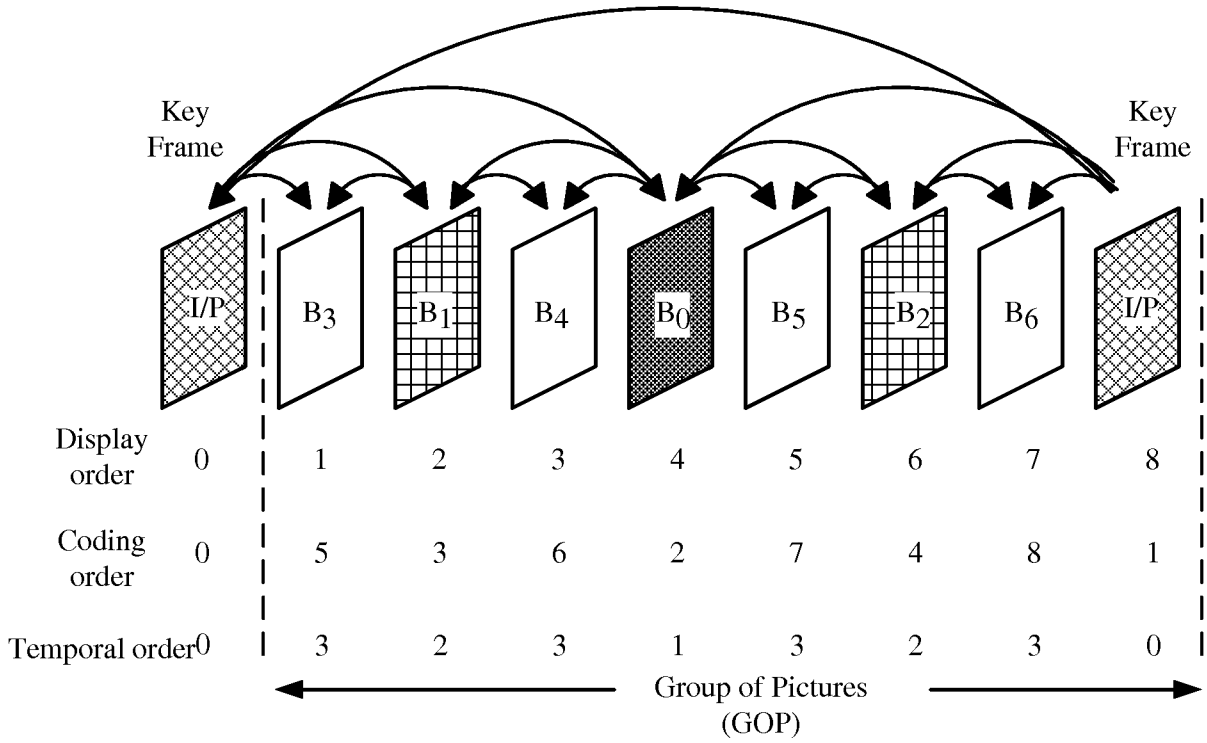
Key Frame
Key Frame
I/P
B3
B1
B4
B0
B5
B2
B6
I/P
Display order
0 1 2 3 4 5 6 7 8
Coding order
0 5 3 6 2 7 4 8 1
Temporal order
0 3 2 3 1 3 2 3 0
Group of Pictures (GOP)

Fig. 4

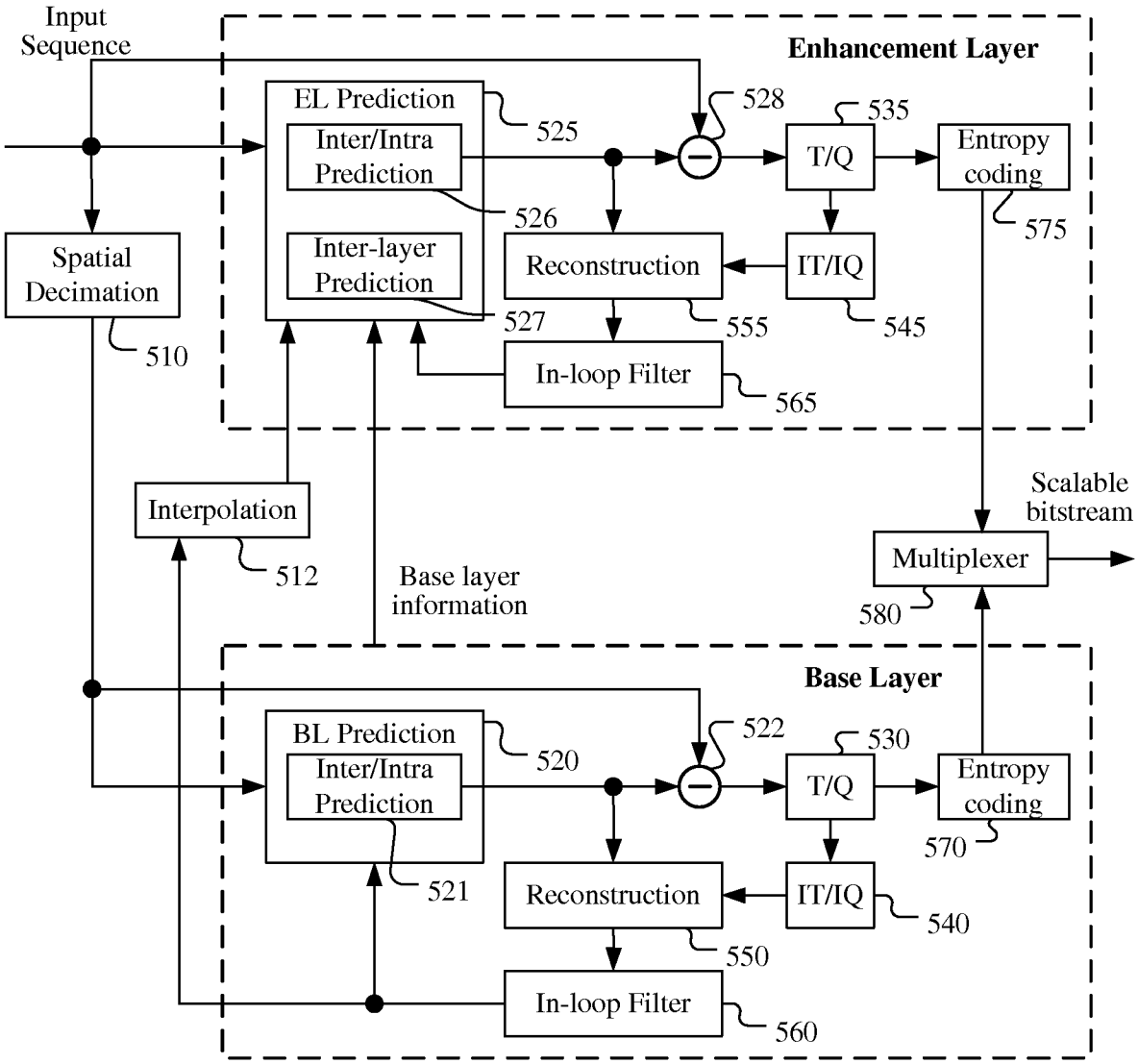
Input Sequence
Enhancement Layer
EL Prediction
Inter/Intra Prediction
525
528
535
T/Q
Entropy coding
575
526
Inter-layer Prediction
Reconstruction
IT/IQ
527
555
545
In-loop Filter
565
Spatial Decimation
510
Interpolation
512
Base layer information
Scalable bitstream
Multiplexer
580
Base Layer
BL Prediction
Inter/Intra Prediction
520
522
530
T/Q
Entropy coding
570
521
Reconstruction
IT/IQ
540
550
In-loop Filter
560

Fig. 5

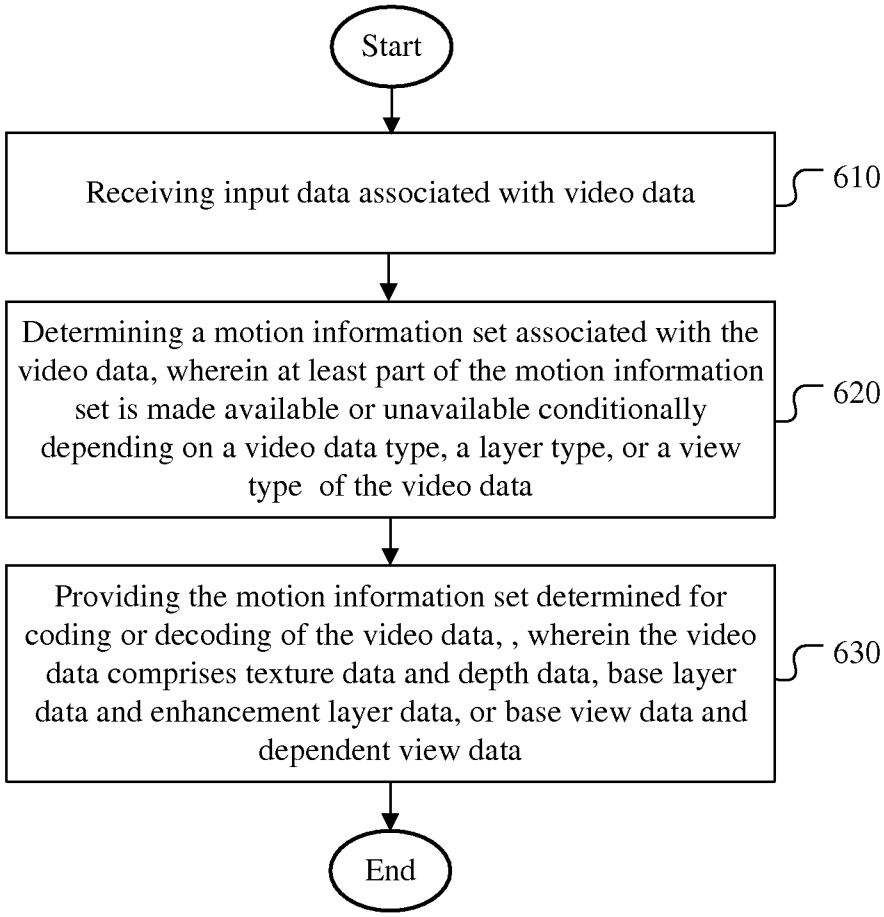
Start
Receiving input data associated with video data
610
Determining a motion information set associated with the video data, wherein at least part of the motion information set is made available or unavailable conditionally depending on a video data type, a layer type, or a view type of the video data
620
Providing the motion information set determined for coding or decoding of the video data, , wherein the video data comprises texture data and depth data, base layer data and enhancement layer data, or base view data and dependent view data
630
End

Fig. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2013/082569

A. CLASSIFICATION OF SUBJECT MATTER

H04N 7/32 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: H04N 7/-

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

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

CNPAT,WPI,EPODOC

three, dimension, 3D, scalable, video, code, decode, SVC, texture, depth, base, enhancement, layer, view, available, unavailable, able, disable, temporal, spatial, motion, information, vector

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2012097749 A1 (MEDIATEK INC.) 26 Jul. 2012 (26.07.2012) description, paragraph [0018] and claims 1,16	1-19
A	US 2005232358 A1 (STMICROELECTRONICS SA) 20 Oct. 2005 (20.10.2005) the whole document	1-19
A	EP 1775954 A1 (THOMSON LICENSING) 18 Apr. 2007 (18.04.2007) the whole document	1-19
A	WO 2006087609 A2 (NOKIA CORPORATION et al.) 24 Aug. 2006 (24.08.2006) the whole document	1-19
A	CN 102300087 A (PEKING UNIVERSITY et al.) 28 Dec. 2011 (28.12.2011) the whole document	1-19

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

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim (S) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 Oct. 2013 (17.10.2013)	**28 Nov. 2013 (28.11.2013)**
Name and mailing address of the ISA/CN The State Intellectual Property Office, the P.R.China 6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China 100088 Facsimile No. 86-10-62019451	Authorized officer WANG,Conglei Telephone No. (86-10)62413236

Form PCT/ISA /210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2013/082569

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
WO 2012097749 A1	26.07.2012	US 2013208804 A1	15.08.2013
		CN 103299630 A	11.09.2013
		AU 2012208842 A1	02.05.2013
US 2005232358 A1	20.10.2005	US 8144775 B2	27.03.2012
		EP 1591962 A2	02.11.2005
		DE 602005000425 T2	18.10.2007
		FR 2868579 A1	07.10.2005
EP 1775954 A1	18.04.2007	None	
WO 2006087609 A2	24.08.2006	EP 1851969 A2	07.11.2007
		US 2006153300 A1	13.07.2006
		TW 200642482 A	01.12.2006
CN 102300087 A	28.12.2011	None	