(12) STANDARD PATENT (11) Application No. **AU 2012267006 B8**

(19) AUSTRALIAN PATENT OFFICE

(54) Title
Method and apparatus of scalable video coding

(51) International Patent Classification(s)
H04N 19/50 (2014.01)

(21) Application No: **2012267006** (22) Date of Filing: **2012.05.31**

(87) WIPO No: **WO12/167712**

(30) Priority Data

(31) Number	(32) Date	(33) Country
61/495,740	**2011.06.10**	**US**

(43) Publication Date: **2012.12.13**
(44) Accepted Journal Date: **2015.10.15**
(48) Corrigenda Journal Date: **2015.10.29**

(71) Applicant(s)
Mediatek Inc.

(72) Inventor(s)
Chuang, Tzu-Der;Chen, Ching-Yeh;Huang, Yu-Wen;Lei, Shaw-Min;Fu, Chih-Ming;Tsai, Chia-Yang

(74) Agent / Attorney
Watermark Patent and Trade Marks Attorneys, 302 Burwood Road, Hawthorn, VIC, 3122

(56) Related Art
REICHEL, J., et al., "Joint Scalable Video Model JSVM-5", Document: JVT-R202, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), 14-20 January 2006
WO 2007/128706

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 December 2012 (13.12.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/167712 A1

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

(21) International Application Number:
PCT/CN2012/076318

(22) International Filing Date:
31 May 2012 (31.05.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/495,740 10 June 2011 (10.06.2011) US

(71) Applicant *(for all designated States except US)*: **MEDIATEK INC.** [CN/CN]; No.1, Dusing Rd. 1st, Science-Based Industrial Park, Hsin-Chu, Taiwan (CN).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CHUANG, Tzu-Der** [CN/CN]; 5F., No.50, Zhuangjing 5th St., Zhubei City, Hsinchu County, Taiwan (CN). **CHEN, Ching-Yeh** [CN/CN]; 4F.,No.809, Sec.5,Zhongxiao E.Rd.,Nangang Dist., Taipei City, Taiwan (CN). **HUANG, Yu-Wen** [CN/CN]; 8F., No. 23, Ln. 298, Longjiang Rd., Zhongshan Dist., Taipei City, Taiwan (CN). **LEI, Shaw-Min** [CN/CN]; 6F., No.32, Liujia 6th St., Zhubei City, Hsinchu County, Taiwan (CN). **FU, Chin-Ming** [CN/CN]; 5, No. 115, Aly. 31, Ln. 647, Sec. 6, Zhonghua Rd., Xiangshan District, Hsinchu City, Taiwan (CN). **TSAI, Chin-Yang** [CN/CN]; 6, 2F., No. 10, Aly. 2, Ln. 35, Minsheng St., Zhonghe District, New Taipei City, Taiwan (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, 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, KM, 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, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) **Title**: METHOD AND APPARATUS OF SCALABLE VIDEO CODING

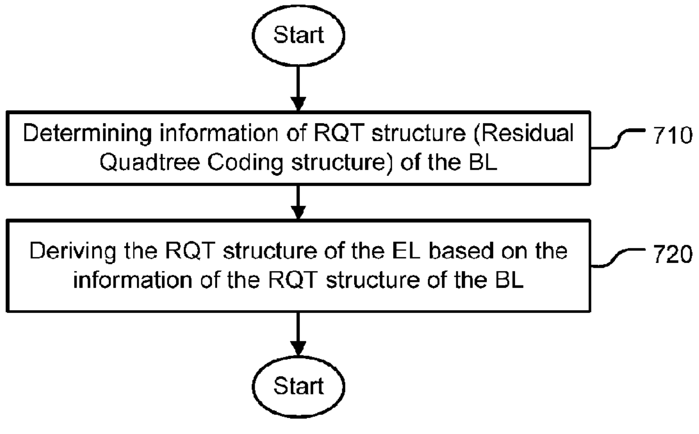

Start
Determining information of RQT structure (Residual Quadtree Coding structure) of the BL
710
Deriving the RQT structure of the EL based on the information of the RQT structure of the BL
720
Start

Fig. 7

(57) **Abstract**: A method and apparatus for scalable video coding are disclosed, wherein the video data is configured into a Base Layer (BL) and an Enhancement Layer (EL) and wherein the EL has higher spatial resolution or better video quality than the BL. According to embodiments of the present invention, information from the base layer is exploited for coding the enhancement layer. The information coding for the enhancement layer includes CU structure, motion information, motion information, MVP/merge candidates, intra prediction mode, residual quadtree information, texture information, residual information, context adaptive entropy coding, Adaptive Lop Filter (ALF), Sample Adaptive Offset (SAO), and deblocking filter.

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 SCALABLE VIDEO CODING

BACKGROUND OF THE INVENTION

Cross Reference To Related Applications

[0001] The present invention claims priority to U.S. Provisional Patent Application, Serial No. 61/495,740, filed June 10, 2011, entitled "Scalable Coding of High Efficiency Video Coding". The U.S. Provisional Patent Application is hereby incorporated by reference in its entirety.

Field of the Invention

[0002] The present invention relates to video coding. In particular, the present invention relates to scalable video coding that utilizes information of the base layer for enhancement layer coding.

Description of the Related Art

[0003] 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 service. 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 content may also be carried through cellular data network so that the content can be watch 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 typical low-resolution display on the smart phone or portable devices, 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 requirement 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 video at different spatial-temporal resolution and/or quality can be derived from the same compressed bitstream.

[0004] In the current H.264/AVC video standard, there is an extension of the H.264/AVC standard, named Scalable Video Coding (SVC). SVC provides temporal, spatial, and quality scalabilities based on a single bitstream. The SVC bitstream contains scalable video information from low frame-rate, low resolution, and low quality to high frame rate, high definition, and high quality respectively. Accordingly, SVC is suitable for various video applications such as video broadcasting, video streaming, and video surveillance to adapt to network infrastructure, traffic condition, user preference, and etc.

[0005] 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 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 that are required for decoding the intended video. NAL units for quality refinement can be additionally truncated in order to reduce the bit-rate and the associated video quality.

[0006] For example, temporal scalability can be derived from hierarchical coding structure based on B-pictures according to the H.264/AVC standard. Fig. 1 illustrates an example of hierarchical B-picture structure with 4 temporal layers and the Group of Pictures (GOP) with eight pictures. Pictures 0 and 8 in Fig. 1 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. 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. 1, picture 4 is first bi-directional 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-directional predicted using picture 0 and 4, and picture 6 is bi-directional predicted using picture 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. 1. 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. 1 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 enable 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.

[0007] In SVC, spatial scalability is supported based on the pyramid coding scheme as shown in Fig. 2. In a SVC system with spatial scalability, the video sequence is first down-sampled to obtain smaller pictures at different spatial resolutions (layers). For example, picture 210 at the original resolution can be processed by spatial decimation 220 to obtain resolution-reduced picture 211. The resolution-reduced picture 211 can be further processed by spatial decimation 221 to obtain further resolution-reduced picture 212 as shown in Fig. 2. In addition to dyadic spatial resolution, where the spatial resolution is reduced to half in each level, SVC also supports arbitrary resolution ratios, which is called extended spatial scalability (ESS). The SVC system in Fig. 2 illustrates an example of spatial scalable system with three layers, where layer 0 corresponds to the pictures with lowest spatial resolution and layer 2 corresponds to the pictures with the highest resolution. The layer-0 pictures are coded without reference to other layers, i.e., single-layer coding. For example, the lowest layer picture 212 is coded using motion-compensated and intra prediction 230.

[0008] The motion-compensated and intra prediction 230 will generate syntax elements as well as coding related information such as motion information for further entropy coding 240. Fig. 2 actually illustrates a combined SVC system that provides spatial scalability as well as quality scalability (also called SNR scalability). The system may also provide temporal scalability, which is not explicitly shown. For each single-layer coding, the residual coding errors can be refined using SNR enhancement layer coding 250. The SNR enhancement layer in Fig. 2 may provide multiple quality levels (quality scalability). Each supported resolution layer can be coded by respective single-layer motion-compensated and intra prediction like a non-scalable coding system. Each higher spatial layer may also be coded using inter-layer coding based on one or more lower spatial layers. For example, layer 1video can be adaptively coded using inter-layer prediction based on layer 0 video or a single-layer coding on a macroblock by macroblock basis or other block unit. Similarly, layer 2 video can be adaptively coded using inter-layer prediction based on reconstructed layer 1 video or a single-layer coding. As shown in Fig. 2, layer-1 pictures 211 can be coded by motion-compensated and intra prediction 231, base layer entropy coding 241 and SNR enhancement layer coding 251. Similarly, layer-2 pictures 210 can be coded by motion-compensated and intra prediction 232, base layer entropy coding 242 and SNR enhancement layer coding 252. The coding efficiency can be improved due to inter-layer coding. Furthermore, the information required to code spatial layer 1 may depend on reconstructed layer 0 (inter-layer prediction). The inter-layer differences are termed as the *enhancement layers*. The H.264 SVC provides three types of inter-layer prediction tools: inter-layer motion prediction, inter-layer intra prediction, and inter-layer residual prediction.

[0009] In SVC, the enhancement layer (EL) can reuse the motion information in the base layer (BL) to reduce the inter-layer motion data redundancy. For example, the EL macroblock coding may use a flag, such as base_mode_flag before mb_type is determined to indicate whether the EL motion information is directly derived from the BL. If base_mode_flag is equal to 1, the partitioning data of the EL macroblock together with the associated reference indexes and motion vectors are derived from the corresponding data of the co-located 8x8 block in the BL. The reference picture index of the BL is directly used in EL. The motion vectors of EL are scaled from the data associated with the BL. Besides, the scaled BL motion vector can be used as an additional motion vector predictor for the EL.

[0010] Inter-layer residual prediction uses the up-sampled BL residual information to reduce the information of EL residuals. The co-located residual of BL can be block-wise up-sampled using a bilinear filter and can be used as prediction for the residual of a current macroblock in the EL. The up-sampling of the reference layer residual is done on a transform block basis in order to ensure that no filtering is applied across transform block boundaries.

[0011] Similar to inter-layer residual prediction, the inter-layer intra prediction reduces the redundant texture information of the EL. The prediction in the EL is generated by block-wise up-sampling the co-located BL reconstruction signal. In the inter-layer intra prediction up-sampling procedure, 4-tap and 2-tap FIR filters are applied for luma and chroma components, respectively. Different from inter-layer residual prediction, filtering for the inter-layer intra prediction is always performed across sub-block boundaries. For decoding simplicity, inter-layer intra prediction can be restricted to only intra-coded macroblocks in the BL.

[0012] In SVC, quality scalability is realized by coding multiple quality ELs which are composed of refinement coefficients. The scalable video bitstream can be easily truncated or extracted to provide different video bitstreams with different video qualities or bitstream sizes. In SVC, the quality scalability, (also called SNR scalability) can be provided via two strategies, coarse grain scalability (CGS), and medium grain scalability (MGS). The CGS can be regarded as a special case of spatial scalability, where the spatial resolution of the BL and the EL are the same. However, the quality of the EL is better (the QP of the EL is smaller than the QP of the BL). The same inter-layer prediction mechanism for spatial scalable coding can be employed. However, no corresponding up-sampling or deblocking operations are performed. Furthermore, the inter-layer intra and residual prediction are directly performed in the transform domain. For the inter-layer prediction in CGS, a refinement of texture information is typically achieved by re-quantizing the residual signal in the EL with a smaller quantization step size than that used for the preceding CGS layer. CGS can provide multiple pre-defined quality points.

[0013] To provide finer bit rate granularity while maintaining reasonable complexity for quality scalability, MGS is used by H.264 SVC. MGS can be considered as an extension of CGS, where the quantized coefficients in one CGS slice can be divided into several MGS slices. The quantized coefficients in CGS are classified to 16 categories based on its scan position in the zig-zag scan order. These 16 categories of coefficients can be distributed into different slices to provide more quality extraction points than CGS.

[0014] In the current HEVC, it only provides single layer coding based on hierarchical-B coding structure without any spatial scalability and quality scalability. It is desirable to provide the capability of spatial scalability and quality scalability to the current HEVC. Furthermore, it is desirable to provide improved SVC over the H.264 SVC to achieve higher efficiency and/or more flexibility.

BRIEF SUMMARY OF THE INVENTION

[0014a] The present invention provides a method of Residual Quadtree Structure coding for scalable video coding, wherein video data is configured into a Base Layer (BL) and an Enhancement Layer (EL) and wherein the EL has higher spatial resolution or better video quality than the BL, the method comprising:

determining information of RQT structure (Residual Quadtree Coding structure) of the BL; and deriving the RQT structure of the EL based on the information of the RQT structure of the BL.

[0014b] Comprises/comprising and grammatical variations thereof when used in this specification are to be taken to specify the presence of stated features, integers, steps or components or groups thereof, but do not preclude the presence or addition of one or more other features, integers, steps, components or groups thereof.

[0014c] The present invention further provides an apparatus of Residual Quadtree Structure coding for scalable video coding, wherein video data is configured into a Base Layer (BL) and an Enhancement Layer (EL) and wherein the EL has higher spatial resolution or better video quality than the BL, the apparatus comprising: means for determining information of RQT structure (Residual Quadtree Coding structure) of the BL; and means for deriving the RQT structure of the EL based on the information of the RQT structure of the BL.

[0015] A method and apparatus for scalable video coding that exploits Base Layer (BL) information for Enhancement Layer (EL) are disclosed, where the EL has higher resolution and/or better quality than the BL. Embodiments of the present invention exploit various pieces of the BL information to improve coding efficiency of the EL. In one embodiment according to the present invention, the method and apparatus utilizes the CU structure information, the mode information, or the motion information of the BL to derive respective information for the EL. A combination of the CU structure, the mode, and the motion information may also be used to derive the respective information for the EL. In another embodiment according to the present invention, the method and apparatus derives Motion Vector Predictor (MVP) candidates or merge candidates of the EL based on MVP candidates or merge candidates of the BL. In yet another embodiment of the present invention, the method and apparatus derives intra prediction mode of the EL based on intra prediction mode of the BL.

[0016] An embodiment of the present invention utilizes Residual Quadtree Structure information of the BL to derive the Residual Quadtree Structure for the EL. Another embodiment of the present invention derives the texture of the EL by re-sampling the texture of the BL. A further embodiment of the present invention derives the predictor of residual of the EL by re-sampling the residual of the BL.

[0017] One aspect of the present invention addresses the coding efficiency of context-based adaptive entropy coding for the EL. An embodiment of the present invention determines context information for processing a syntax element of the EL using the information of the BL. Another aspect of the present invention addresses the coding efficiency related in-loop processing. An embodiment of the present invention derives the ALF information, the SAO information, or the DF information for the EL using the ALF information, the SAO information, or the DF information of the BL respectively.

BRIEF DESCRIPTION OF DRAWINGS

[0018] Fig. 1 illustrates an example of temporal scalable video coding using hierarchical B-pictures.

[0019] Fig. 2 illustrates an example of a combined scalable video coding system that provides spatial scalability as well as quality scalability where three spatial layers are provides.

[0020] Fig. 3 illustrates an example of CU structure reuse for scalable video coding where a CU structure for the base layer is scaled and used as an initial CU structure for the enhancement layer.

[0021] Fig. 4 illustrates an exemplary flow chart of CU structure coding or motion information coding for scalable video coding according to an embodiment of the present invention.

[0022] Fig. 5 illustrates an exemplary flow chart of MVP derivation or merge candidate derivation for scalable video coding according to an embodiment of the present invention.

[0023] Fig. 6 illustrates an exemplary flow chart of intra prediction mode derivation for scalable video coding according to an embodiment of the present invention.

[0024] Fig. 7 illustrates an exemplary flow chart of Residual Quadtree Structure coding for scalable video coding according to an embodiment of the present invention.

[0025] Fig. 8 illustrates an exemplary flow chart of texture prediction and re-sampling for scalable video coding according to an embodiment of the present invention.

[0026] Fig. 9 illustrates an exemplary flow chart of residual prediction and re-sampling for scalable video coding according to an embodiment of the present invention.

[0027] Fig. 10 illustrates an exemplary flow chart of context adaptive entropy coding for scalable video coding according to an embodiment of the present invention.

[0028] Fig. 11 illustrates an exemplary flow chart of ALF information coding, SAO information coding and DF information coding for scalable video coding according to an embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0029] In HEVC, coding unit (CU) structure was introduced as a new block structure for coding process. A picture is divided into largest CUs (LCUs) and each LCU is adaptively partitioned into CUs until a leaf CU is obtained or a minimum CU size is reached. The CU structure information has to be conveyed to the decoder side so that the same CU structure can be recovered at the decoder side. In order to improve coding efficiency associated with the CU structure for a scalable HEVC, an embodiment according to the present invention allows the CU structure of the BL reused by the EL. In the EL LCU or CU level, one flag is transmitted to indicate whether the CU structure is reused from corresponding CU of the BL. If the BL CU structure is reused, the BL CU structure is scaled to match the resolutions of the EL and the scaled BL CU structure is reused by the EL. Moreover, the leaf CU of scaled CU structures can be further split into sub-CUs. Fig. 3 illustrates an example of CU partition reuse. Partition 310 corresponds to the CU structure of the BL. The video resolution of the EL is two times of the video resolution of the BL horizontally and vertically. The CU structure of corresponding CU partition 315 of BL is scaled up by 2. The scaled CU structure 320 is then used as the initial CU structure for the EL LCU. The leaf CUs of the scaled CU in the EL can be further split into sub-CUs and the result is indicated by 330 in Fig. 3. A flag may be used to indicate whether the leaf CU is further divided into sub-CUs. While Fig. 3 illustrates an example of CU structure is reused, other information may also be reused. For example, the prediction type, prediction size, merge index, inter reference direction, reference picture index, motion vectors, MVP index, and intra mode. The information/data can be scaled when needed before the information/data is reused in the EL.

[0030] In another embodiment according to the present invention, the mode information for a leaf CU is reused. The mode information may include skip flag, prediction type, prediction size, inter reference direction, reference picture index, motion vectors, motion vector index, merge flag, merge index, and intra mode. The mode information of the leaf CU in the EL can share the same or scaled mode information of the corresponding CU in the BL. One flag can be used to indicate whether the EL will reuse the mode information from the BL or not. For one or more pieces of mode information, one flag can be used to indicate whether the EL will reuse this mode information from the BL or not.

[0031] In yet another embodiment according to the present invention, the motion information of corresponding Prediction Unit (PU) or Coding Unit (CU) in the BL is reused to derive the motion information of a PU or CU in the EL. The motion information may include inter prediction direction, reference picture index, motion vectors (MVs), Motion Vector Predictors (MVPs), MVP index, merge index, merge candidates, and intra mode. The motion information for the BL can be utilized as predictors or candidates for the motion information in the EL. For example, the BL MVs and BL MVPs can be added into the MVP list and/or merge list for EL MVP derivation. The aforementioned MVs of BL can be the MVs of the corresponding PU in the BL, the MVs of neighboring PUs of the corresponding PUs in the BL, the MVs of merge candidates of the corresponding PUs in the BL, the MVP of the corresponding PUs in the BL, or the co-located MVs of the corresponding PUs in the BL.

[0032] In another example, the merge candidate derivation for the EL can utilize the motion information of the BL. For example, the merge candidates of a corresponding PU in the BL can be added into the merge candidate list and/or MVP list. The aforementioned motion information of the BL can be the motion information of the corresponding PU in the BL, the motion information associated with a neighboring PU of the corresponding PU in the BL, merge candidates of the corresponding PUs in the BL, MVP of the corresponding PUs in the BL, or the co-located PU of the corresponding PU in the BL. In this case, the motion information includes inter prediction direction, reference picture index, and motion vectors.

[0033] In yet another example, the intra mode of a corresponding PU or CU in the BL can be reused for the EL. For example, the intra mode of a corresponding PU or CU in the BL can be added into the intra most probable mode list. An embodiment according to the present invention uses the motion information of the BL to predict the intra mode for the EL. The order for the most probable mode list in the EL can be adaptively changed according to the intra prediction mode information in the BL. Accordingly, the codeword lengths for codewords in the most probable mode list in the EL can be adaptively changed according to the intra prediction mode information in the BL. For example, the codewords of the intra remaining modes with prediction directions close to the prediction direction of coded BL intra mode are assigned a shorter length. As another example, the neighboring direction modes of BL intra mode can also be added into intra Most Probable Mode (MPM) list of the EL intra mode coding. The intra prediction mode information of the BL can be the intra prediction mode of the corresponding PU in the BL, or the neighboring direction modes of BL intra mode, or the intra prediction mode of a neighboring PU of the corresponding PU in the BL.

[0034] The selected MVP index, merge index, and intra mode index of BL motion information can be utilized to adaptively change the indices order in EL MVP list, merge index list, and intra most probable mode list. For example, in the HEVC Test Model Version 3.0 (HM-3.0), the order of the MVP list is {left MVP, above MVP, co-located MVP}. If the corresponding BL PU selects the above MVP, the order of the above MVP will be moved forward in the EL. Accordingly, the MVP list in the EL will become {above MVP, left MVP, co-located MVP}. Furthermore, the BL coded MV, scaled coded MV, MVP candidates, scaled MVP candidates, merge candidates, and scaled merge candidates can replace part of EL MVP candidates and/or merge candidates. The process of deriving the motion information for a PU or CU in the EL based on the motion information for a corresponding PU or CU in the BL is invoked when an MVP candidate or a merge candidate for a PU or CU in the EL is needed for encoding or decoding.

[0035] As mentioned earlier, the CU structure information for the BL can be used to determine the CU structure information for the EL. Furthermore, the CU structure information, the mode information and the motion information for the BL can be used jointly to determine the CU structure information, the mode information and the motion information for the EL. The mode information or the motion information for the BL may also be used to determine the mode information or the motion information for the EL. The process of deriving the CU structure information, the mode information, the motion information or any combination for the EL based on corresponding information for the BL can be invoked when the CU structure information, the mode information, the motion information or any combination for the EL needs to be encoded or decoded.

[0036] In HM-3.0, the prediction residual is further processed using quadtree partitioning and a coding type is selected for each block of results of residual quadtree partition. Both residual quadtree partition information and coding block pattern (CBP) information have to be incorporated into the bitstream so that the decoder can recover the residual quadtree information. An embodiment according to the present invention reuses the residual quadtree partition and CBP of a corresponding CU in the BL for the EL. The residual quadtree partition and CBP can be scaled and utilized as the predictor for the EL residual quadtree partition and CBP coding. In HEVC, the unit for block transform is termed as Transform Unit (TU) and a TU can be partitioned into smaller TUs. In an embodiment of the present invention, one flag for a root TU level or a TU level of the EL is transmitted to indicate that whether the Residual Quadtree Coding (RQT) structure of a corresponding TU in the BL is utilized to predict the RQT structure of the current TU in the EL. If the RQT structure of a corresponding TU in the BL is utilized to predict the RQT structure of the current TU in the EL, the RQT structure of the corresponding TU in the BL is scaled and used as the initial RQT structure of the current TU in the EL. In the leaf TU of the initial RQT structure for the EL, one split flag can be transmitted to indicate whether the TU is divided into sub-TUs. The process of deriving the RQT structure of the EL based on the information of the RQT structure of the BL is performed when an encoder needs to encode the RQT structure of the EL or a decoder needs to decode the RQT structure of the EL.

[0037] In H.264/AVC scalable extension, 4-tap and 2-tap FIR filters are adopted for the up-sampling operation of texture signal for luma and chroma components respectively. An embodiment according to the present invention re-samples the BL texture as the predictor of EL texture, where the re-sampling utilizes improved up-sampling methods to replace the 4-tap and 2-tap FIR filter in H.264/AVC scalable extension. The filter according to the present invention uses one of the following filters or a combination of the following filters: Discrete Cosine Transform Interpolation Filter (DCTIF), Discrete Sine Transform Interpolation Filter (DSTIF), Wiener filter, non-local mean filter, smoothing filter, and bilateral filter. The filter according to the present invention can cross TU boundaries or can be restricted within TU boundaries. An embodiment according to the present invention may skip the padding and deblocking procedures in inter-layer intra prediction to alleviate computation and data dependency problem. The Sample Adaptive Offset (SAO), Adaptive Loop Filter (ALF), non-local mean filter, and/or smoothing filter in the BL could also be skipped. The skipping of padding, deblocking, SAO, ALF, non-local mean filter, and smoothing filter can be applied to the entire LCU, leaf CU, PU, TU, pre-defined region, LCU boundary, leaf CU boundary, PU boundary, TU boundary, or boundary of a pre-defined region. In another embodiment, the texture of the BL is processed using a filter to produce filtered BL texture, and the BL texture has the same resolution as the EL texture and is used as the predictor of the texture of the EL. Wiener filter, ALF (Adaptive Loop Filter), non-local mean filter, smoothing filter, or SAO (Sample Adaptive Offset) can be applied to the texture of the BL before the texture of BL is utilized as the predictor of the texture of the EL.

[0038] To improve picture quality, an embodiment of the present invention applies Wiener filter or adaptive filter to the texture of the BL before the texture of the BL is re-sampled. Alternatively, the Wiener filter or adaptive filter can be applied to the texture of the BL after the texture of the BL is re-sampled. Furthermore, an embodiment of the present invention applies SAO or ALF to the texture of the BL before the texture of the BL is re-sampled.

[0039] Another embodiment according to the present invention utilizes LCU-based or CU-based Wiener filter and/or adaptive offset for inter-layer intra prediction. The filtering can be applied to BL texture data or up-sampled BL texture data.

[0040] In H.264 SVC, 2-tap FIR filter is adopted for the up-sampling operation of residual signal for both luma and chroma components. An embodiment according to the present invention uses improved up-sampling methods to replace the 2-tap FIR filter of H.264 SVC. The filter can be one of the following filters or a combination of the following filters: Discrete Cosine Transform Interpolation Filter (DCTIF), Discrete Sine Transform Interpolation Filter (DSTIF), Wiener filter, non-local mean filter, smoothing filter, and bilateral filter. When the EL has higher spatial resolution than the BL, the above filters can be applied to re-sample the BL residual. All the above filters can be restricted to cross or not to cross TU boundaries. Furthermore, the residual prediction can be performed in either the spatial domain or the frequency domain if the BL and the EL have the same resolution or the EL has a higher resolution than the BL. When the EL has higher spatial resolution than the BL, the residual of the BL can be re-sampled in frequency domain to form predictors for the EL residual. The process of deriving the predictor of residual of the EL by re-sampling the residual of the BL can be performed when an encoder or a decoder needs to derive the predictor of the residual of the EL based on the re-sampled residual of the BL.

[0041] An embodiment according to the present invention may utilize the BL information for context-based adaptive entropy coding in the EL. For example, the context formation or binarization of (Context-based Adaptive Binary Arithmetic Coding) CABAC can exploit the information of the BL. The EL can use different context models, different context formation methods, or different context sets based on corresponding information in the BL. For example, the EL PU can use different context models depending on whether the corresponding PU in the BL is coded in skip mode or not. In another embodiment of the present invention, the probability or most probable symbol (MPS) of part of context models for CABAC in the BL can be reused to derive the initial probability and MPS of part of context models for CABAC in the EL. The syntax element can be split flag, skip flag, merge flag, merge index, chroma intra mode, luma intra mode, partition size, prediction mode, inter prediction direction, motion vector difference, motion vector predictor index, reference index, delta quantization parameter, significant flag, last significant position, coefficient-greater-than-one, coefficient-magnitude-minus-one, ALF(Adaptive Loop Filter) control flag, ALF flag, ALF footprint size, ALF merge flag, ALF ON/OFF decision, ALF coefficient, sample adaptive offset (SAO) flag, SAO type, SAO offset, SAO merge flag, SAO run, SAO on/off decision, transform subdivision flags, residual quadtree CBF (Coded Block Flag), or residual quadtree root CBF. A codeword corresponding to the syntax elements can be adaptively changed according to the information of the BL and the codeword order corresponding to the syntax elements of the EL in a look-up codeword table can also be adaptively changed according to the information of the BL. The process of determining context information for processing the syntax element of the EL using the information of the BL is performed when the syntax element of the EL needs to be encoded or decoded.

[0042] An embodiment of the present invention uses some ALF information in the BL to derive the ALF information in the EL. The ALF information may include filter adaptation mode, filter coefficients, filter footprint, region partition, ON/OFF decision, enable flag, and merge results. For example, the EL can use part of ALF parameters in the BL as the ALF parameters or predictors of ALF parameters in the EL. When the ALF information is reused directly from the ALF information of the BL, there is no need to transmit the associated ALF parameters for the EL. A flag can be used to indicate whether the ALF information for the EL is predicted from the ALF information of the BL. If the flag indicates that the ALF information for the EL is predicted from the ALF information of the BL, the ALF information of the BL can be scaled and used as the predictor for the ALF information of the EL. A value can be used to denote the difference between the predictor of the ALF information and the ALF information of the EL. The process of deriving the ALF information for the EL using the ALF information of the BL is performed when an encoder or a decoder needs to derive the ALF information of the EL.

[0043] An embodiment of the present invention uses some SAO information in the BL to derive the SAO information in the EL. The SAO information may include offset type, offsets, region partition, ON/OFF decision, enable flag, and merge results. For example, the EL can use part of SAO parameters in the BL as the SAO parameters for the EL. When the SAO information is reused from the SAO information of the BL directly, there is no need to transmit the associated SAO parameters for the EL. A flag can be used to indicate whether the SAO information for the EL is predicted from the SAO information of the BL. If the flag indicates that the SAO information for the EL is predicted from the SAO information of the BL, the SAO information of the BL can be scaled and used as the predictor for the SAO information of the EL. A value can be used to denote the difference between the predictor of the SAO information and the SAO information of the EL. The process of deriving the SAO information for the EL using the SAO information of the BL is performed when an encoder or a decoder needs to derive the SAO information of the EL.

[0044] An embodiment of the present invention uses some Deblocking Filter (DF) information in the BL to derive the DF information in EL. The DF information may include threshold values, such as thresholds α, β, and t_C that are used to determine Boundary Strength (BS). The DF information may also include filter parameters, ON/OFF filter decision, Strong/Weak filter selection, or filter strength. When the DF information is reused from DF information of the BL directly, there is no need to transmit the associated DF parameters for the EL. A flag can be used to indicate whether the DF information for the EL is predicted from the DF information of the BL. If the flag indicates that the DF information for the EL is predicted from the DF information of the BL, the DF information of the BL can be scaled and used as the predictor for the DF information of the EL. A value can be used to denote the difference between the predictor of the DF information and the DF information of the EL. The process of deriving the DF information for the EL using the DF information of the BL is performed when an encoder or a decoder needs to derive the DF information of the EL.

[0045] Figs. 4 through 11 illustrate exemplary flow charts for scalable video coding according to various embodiments of the present invention. Fig. 4 illustrates an exemplary flow chart of CU structure coding or motion information coding for scalable video coding according to an embodiment of the present invention, wherein video data is configured into a Base Layer (BL) and an Enhancement Layer (EL) and wherein the EL has higher spatial resolution or better video quality than the BL. The CU structure (Coding Unit structure), motion information, or a combination of the CU structure and the motion information for a CU (Coding Unit) in the BL is determined in step 410. The CU structure, motion vector predictor (MVP) information, or a combination of the CU structure and the MVP information for a corresponding CU in the EL based on the CU structure, the motion information, or the combination of the CU structure and the motion information for the CU in the BL is respectively determined in step 420. Fig. 5 illustrates an exemplary flow chart of MVP derivation or merge candidate derivation for scalable video coding according to an embodiment of the present invention, wherein video data is configured into a Base Layer (BL) and an Enhancement Layer (EL) and wherein the EL has higher spatial resolution or better video quality than the BL. The motion information for in the BL is determined in step 510. The Motion Vector Predictor (MVP) candidates or merge candidates in the EL based on the motion information the BL is derived in step 520. Fig. 6 illustrates an exemplary flow chart of intra prediction mode derivation for scalable video coding according to an embodiment of the present invention, wherein video data is configured into a

Base Layer (BL) and an Enhancement Layer (EL) and wherein the EL has higher spatial resolution or better video quality than the BL. The information of intra prediction mode of the BL is determined in step 610. The intra prediction mode of the EL based on the information of the intra prediction mode of the BL is derived in step 620.

[0046] Fig. 7 illustrates an exemplary flow chart of Residual Quadtree Structure coding for scalable video coding according to an embodiment of the present invention, wherein video data is configured into a Base Layer (BL) and an Enhancement Layer (EL) and wherein the EL has higher spatial resolution or better video quality than the BL. The information of RQT structure (Residual Quadtree Coding structure) of the BL is determined in step 710. The RQT structure of the EL based on the information of the RQT structure of the BL is derived in step 720. Fig. 8 illustrates an exemplary flow chart of texture prediction and re-sampling for scalable video coding according to an embodiment of the present invention, wherein video data is configured into a Base Layer (BL) and an Enhancement Layer (EL) and wherein the EL has higher spatial resolution than the BL or better video quality than the BL. The information of texture of the BL is determined in step 810. A predictor of texture of the EL based on the information of the texture of the BL is derived in step 820. Fig. 9 illustrates an exemplary flow chart of residual prediction and re-sampling for scalable video coding according to an embodiment of the present invention, wherein video data is configured into a Base Layer (BL) and an Enhancement Layer (EL) and wherein the EL has higher spatial resolution than the BL or better video quality than the BL. The residual information of the BL is determined in step 910. A predictor of residual of the EL by re-sampling the residual of the BL is derived in step 920.

[0047] Fig. 10 illustrates an exemplary flow chart of context adaptive entropy coding for scalable video coding according to an embodiment of the present invention, wherein video data is configured into a Base Layer (BL) and an Enhancement Layer (EL) and wherein the EL has higher spatial resolution or better video quality than the BL. The information of the BL is determined in step 1010. The context information for processing a syntax element of the EL using the information of the BL is determined in step 1020. Fig. 11 illustrates an exemplary flow chart of ALF information coding, SAO information coding and DF information coding for scalable video coding according to an embodiment of the present invention, wherein video data is configured into a Base Layer (BL) and an Enhancement Layer (EL) and wherein the EL has higher spatial resolution or better video quality than the BL. The ALF information, SAO information or DF information of the BL is determined in step 1110. The ALF information, SAO information, or DF information for the EL using the ALF information, SAO information, or DF information of the BL is respectively derived in step 1120.

[0048] Embodiments of scalable video coding, where the enhancement layer coding exploits the information of the base layer, according to 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 codes integrated into video compression software to perform the processing described herein. An embodiment of the present invention may also be program codes 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 codes may be developed in different programming languages and different format or style. 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.

[0049] 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.

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A method of Residual Quadtree Structure coding for scalable video coding, wherein video data is configured into a Base Layer (BL) and an Enhancement Layer (EL) and wherein the EL has higher spatial resolution or better video quality than the BL, the method comprising:
determining information of RQT structure (Residual Quadtree Coding structure) of the BL; and
deriving the RQT structure of the EL based on the information of the RQT structure of the BL.

2. The method of Claim 1, wherein said deriving the RQT structure of the EL based on the information of the RQT structure of the BL is performed when an encoder needs to encode the RQT structure of the EL or a decoder needs to decode the RQT structure of the EL.

3. The method of Claim 1, wherein a first flag is used for a root TU (Transform Unit) or a TU of the EL, wherein the first flag indicates whether the RQT structure of the TU of the BL is utilized to derive the RQT structure of a corresponding TU of the EL.

4. The method of Claim 3, wherein the RQT structure of the TU in the BL is scaled and used as an initial RQT structure of the corresponding TU in the EL if the RQT structure of the TU in the BL is utilized to derive the RQT structure of the corresponding TU in the EL.

5. The method of Claim 4, wherein a split flag is used to indicate whether the corresponding TU of the EL is divided into sub-TUs for a leaf TU of the initial RQT structure of the corresponding TU in the EL.

6. An apparatus of Residual Quadtree Structure coding for scalable video coding, wherein video data is configured into a Base Layer (BL) and an Enhancement Layer (EL)

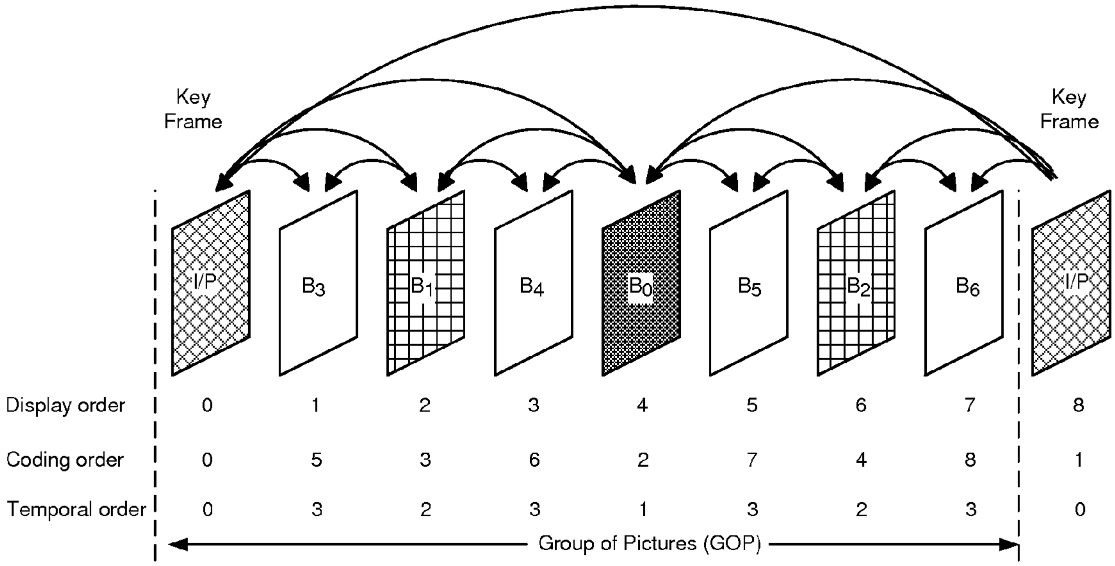
and wherein the EL has higher spatial resolution or better video quality than the BL, the apparatus comprising:
means for determining information of RQT structure (Residual Quadtree Coding structure) of the BL; and
means for deriving the RQT structure of the EL based on the information of the RQT structure of the BL.

7. The apparatus of Claim 6, wherein a first flag is used for a root TU (Transform Unit) or a TU of the EL, wherein the first flag indicates whether the RQT structure of the TU of the BL is utilized to derive the RQT structure of a corresponding TU of the EL.

MEDIATEK INC

WATERMARK PATENT AND TRADE MARKS ATTORNEYS

P37994AU00



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

Fig. 1

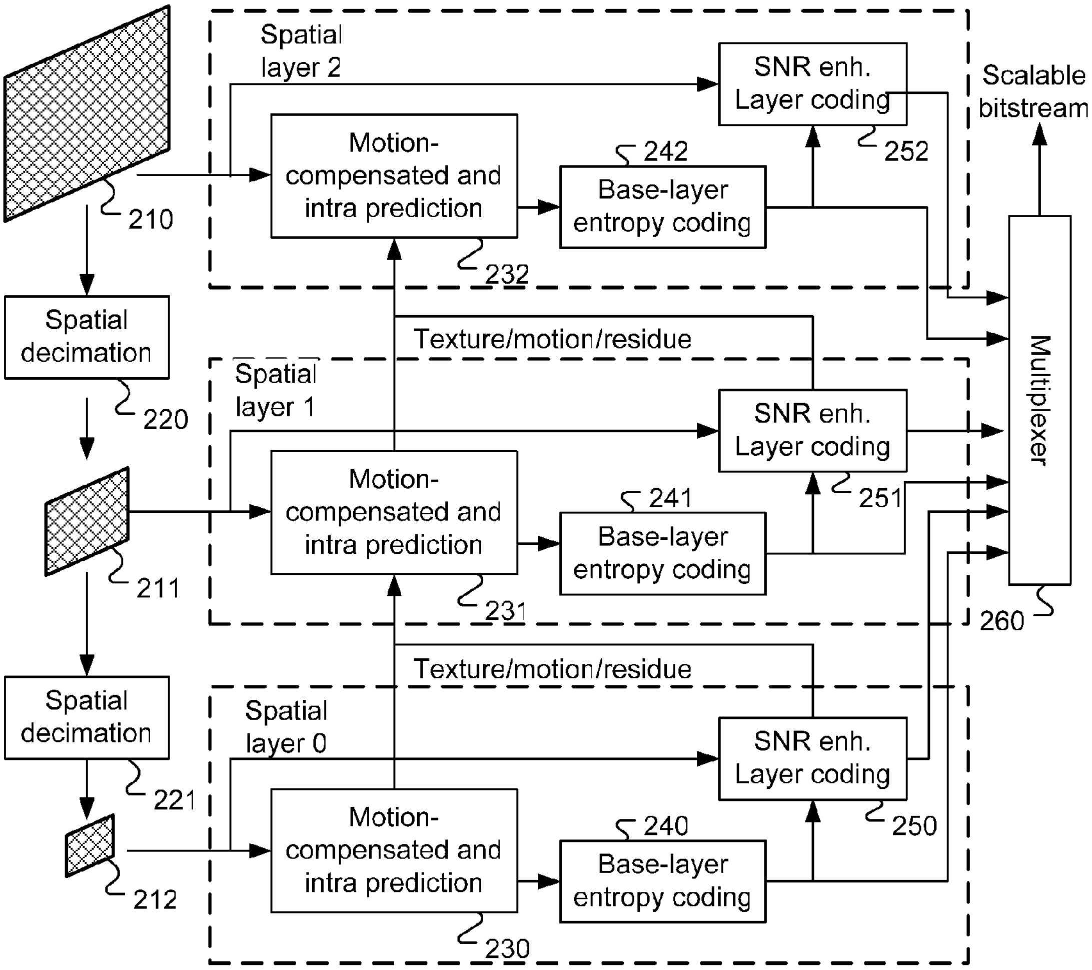

Spatial layer 2
SNR enh. Layer coding
Scalable bitstream
Motion-compensated and intra prediction
242
Base-layer entropy coding
252
210
232
Spatial decimation
Texture/motion/residue
Spatial layer 1
SNR enh. Layer coding
Multiplexer
220
Motion-compensated and intra prediction
241
251
Base-layer entropy coding
211
231
260
Texture/motion/residue
Spatial decimation
Spatial layer 0
SNR enh. Layer coding
221
Motion-compensated and intra prediction
240
250
Base-layer entropy coding
212
230

Fig. 2

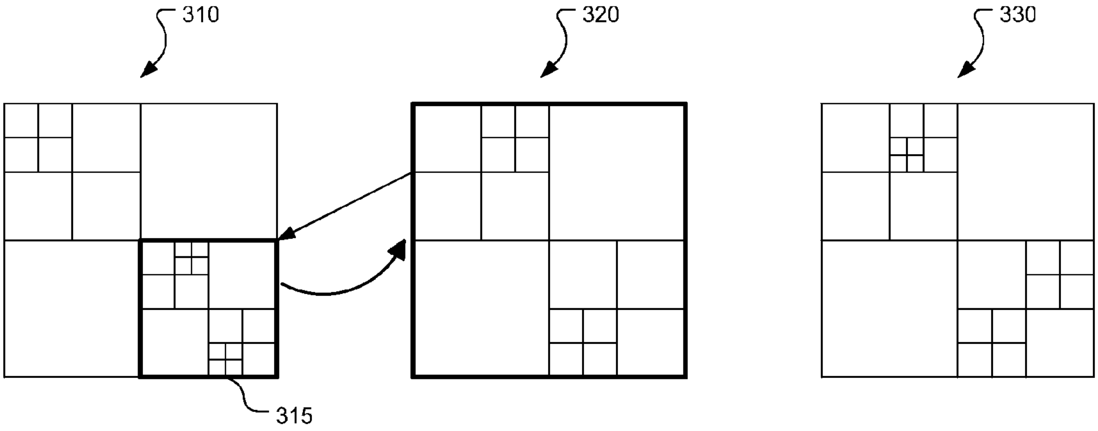


Fig. 3

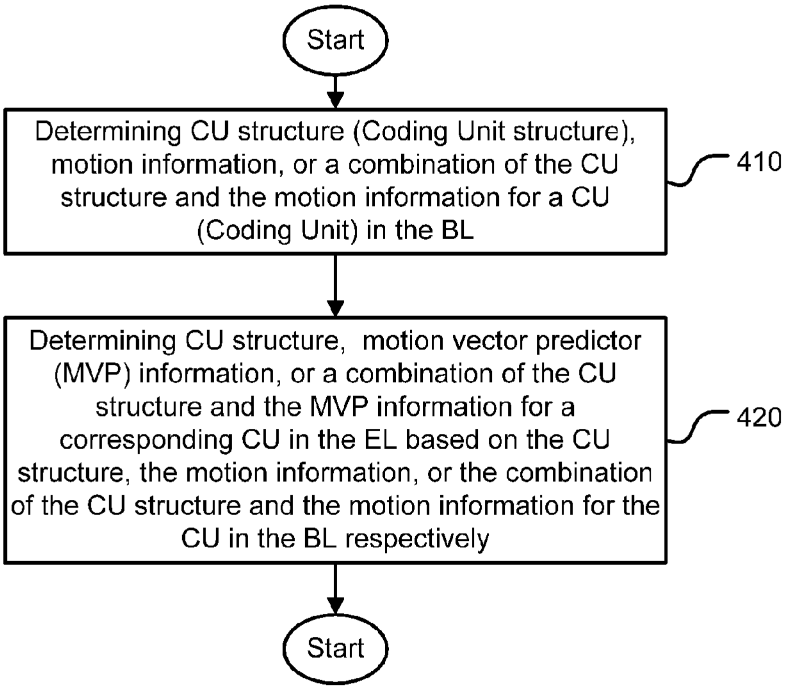

Start
Determining CU structure (Coding Unit structure), motion information, or a combination of the CU structure and the motion information for a CU (Coding Unit) in the BL
410
Determining CU structure, motion vector predictor (MVP) information, or a combination of the CU structure and the MVP information for a corresponding CU in the EL based on the CU structure, the motion information, or the combination of the CU structure and the motion information for the CU in the BL respectively
420
Start

Fig. 4

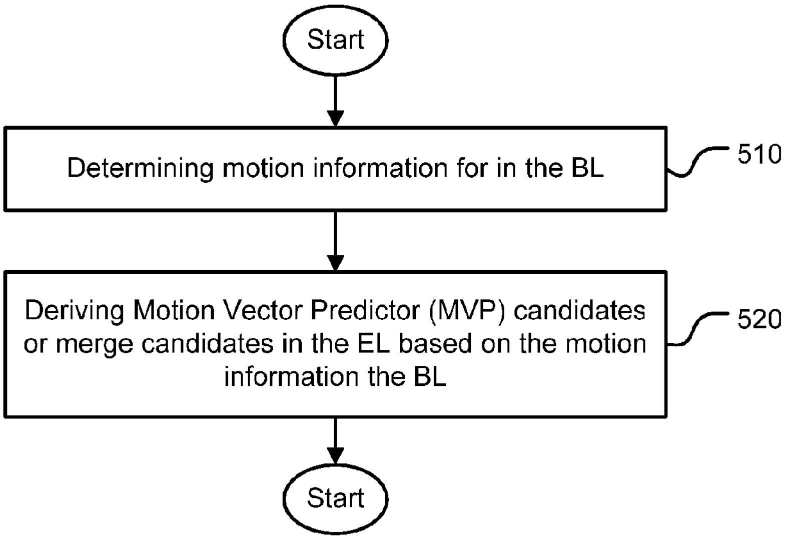

Start
Determining motion information for in the BL
510
Deriving Motion Vector Predictor (MVP) candidates or merge candidates in the EL based on the motion information the BL
520
Start

Fig. 5

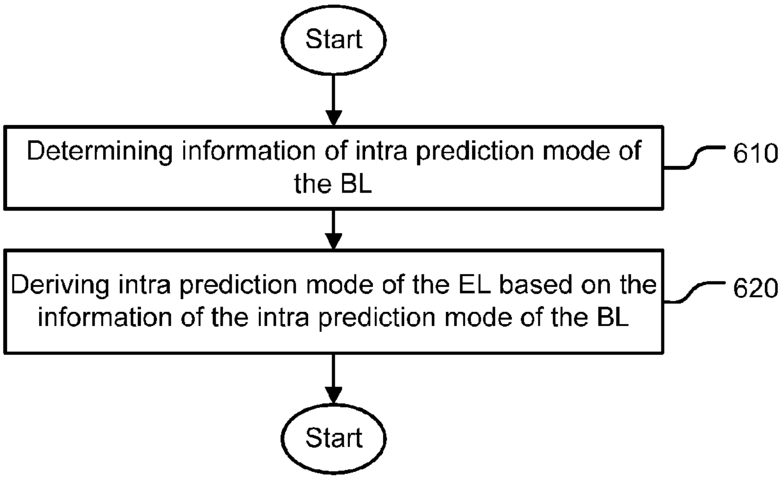

Start
Determining information of intra prediction mode of the BL
610
Deriving intra prediction mode of the EL based on the information of the intra prediction mode of the BL
620
Start

Fig. 6

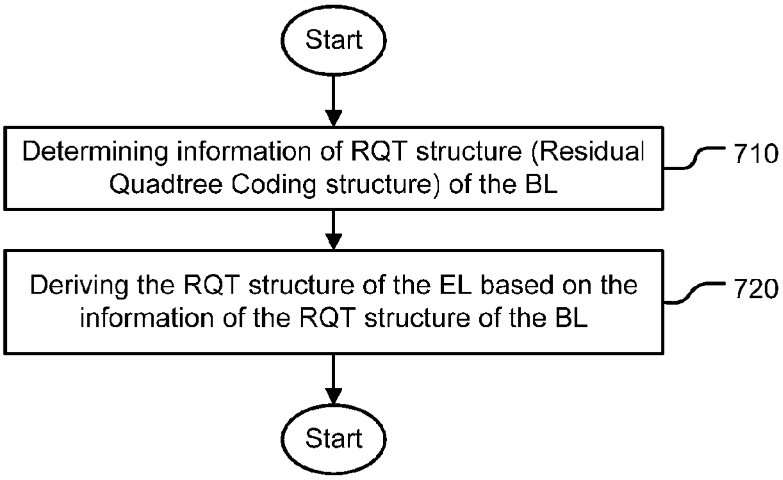

Start
Determining information of RQT structure (Residual Quadtree Coding structure) of the BL
710
Deriving the RQT structure of the EL based on the information of the RQT structure of the BL
720
Start

Fig. 7

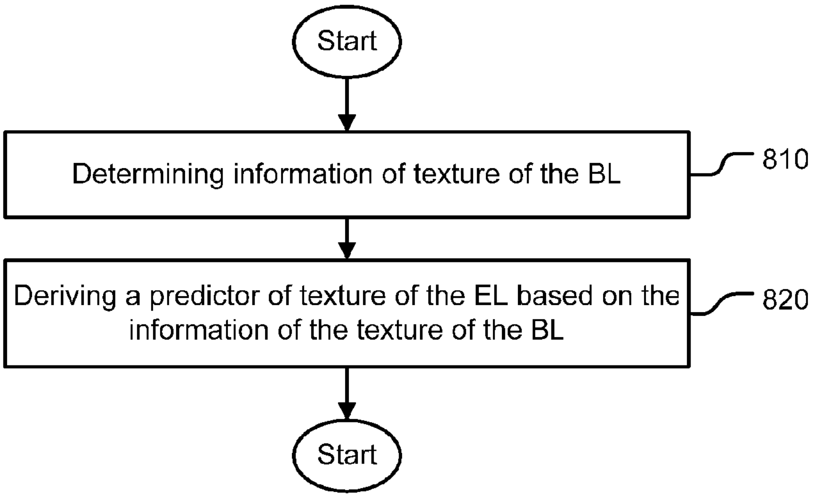

Start
Determining information of texture of the BL
810
Deriving a predictor of texture of the EL based on the information of the texture of the BL
820
Start

Fig. 8

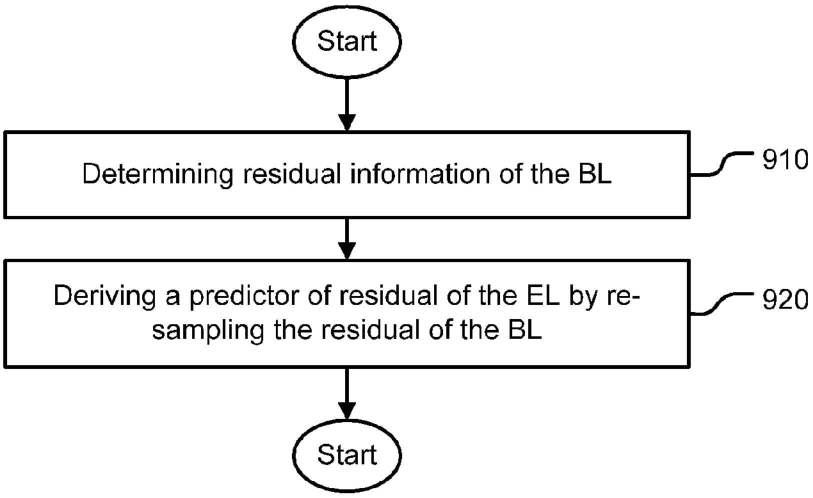

Start
Determining residual information of the BL
910
Deriving a predictor of residual of the EL by re-sampling the residual of the BL
920
Start

Fig. 9

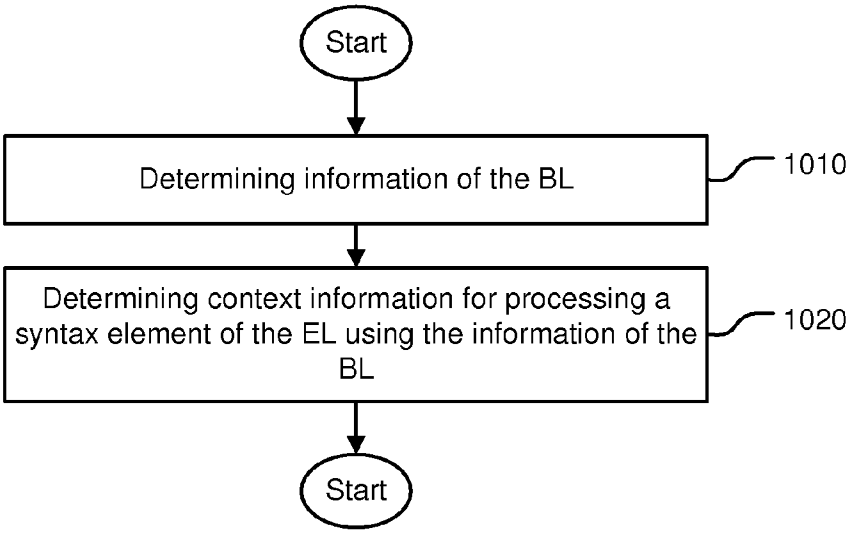

Start
Determining information of the BL
1010
Determining context information for processing a syntax element of the EL using the information of the BL
1020
Start

Fig. 10

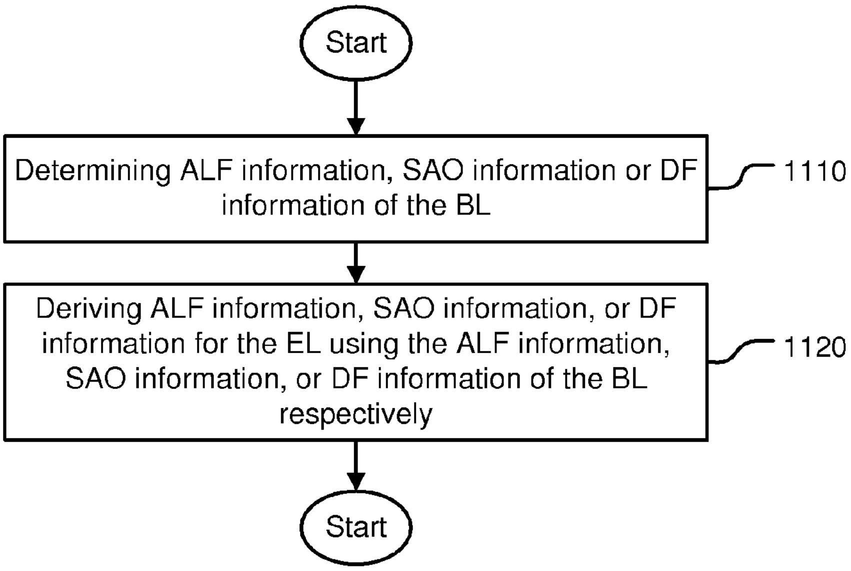

Start
Determining ALF information, SAO information or DF information of the BL
1110
Deriving ALF information, SAO information, or DF information for the EL using the ALF information, SAO information, or DF information of the BL respectively
1120
Start

Fig. 11