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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 September 2012 (13.09.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/122421 A1

(51) International Patent Classification:
H04N 7/26 (2006.01) *H04N 7/36* (2006.01)

(21) International Application Number:
PCT/US2012/028366

(22) International Filing Date:
8 March 2012 (08.03.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/451,543 10 March 2011 (10.03.2011) US

(71) Applicant *(for all designated States except US)*: **DOLBY LABORATORIES LICENSING CORPORATION** [US/US]; 100 Potrero Avenue, San Francisco, California 94103-4813 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **TOURAPIS, Alexandros** [CY/US]; 939 Coventry Circle, Milpitas, California 95035 (US).

(74) Agents: **DOLBY LABORATORIES, INC.** et al.; Intellectual Property Group, 999 Brannan Street, San Francisco, California 94103-4938 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, 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, MD, 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).

Declarations under Rule 4.17:

— *as to the identity of the inventor (Rule 4.17(i))*

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

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

— *of inventorship (Rule 4.17(iv))*

Published:

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

WO 2012/122421 A1

(54) Title: JOINT RATE DISTORTION OPTIMIZATION FOR BITDEPTH COLOR FORMAT SCALABLE VIDEO CODING

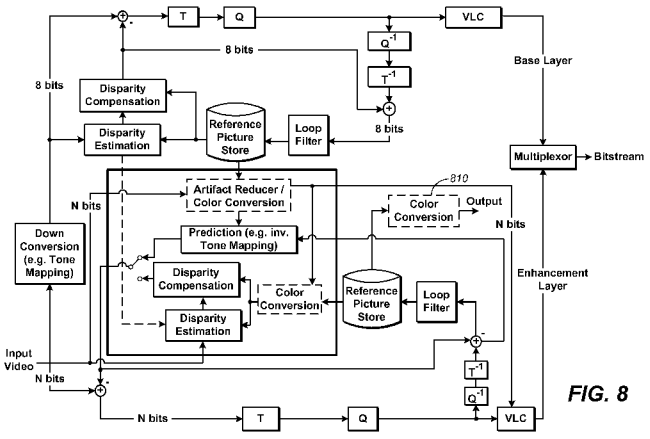


FIG. 8

(57) Abstract: Methods for scalable video coding are described. Such methods can be used to deliver video contents in Low Dynamic Range (LDR) and/or one color format and then converting the video contents to High Dynamic Range (HDR) and/or a different color format, respectively, while considering joint rate distortion optimization.

JOINT RATE DISTORTION OPTIMIZATION FOR BITDEPTH COLOR FORMAT SCALABLE VIDEO CODING

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Patent Application No. 61/451,543 filed on March 10, 2011.

[0002] The present application may be related to International Patent Application No. US2006/020633, filed on May 25, 2006, International Patent Application No. US2006/024528, filed on June 23, 2006, U.S. Patent Application No. 12/188,919, filed on August 8, 2008, U.S. Patent Application No. 12/999,419, filed on December 16, 2010, U.S. Patent Application No. 13/057,204, filed on February 2, 2011, U.S. Provisional Patent Application No. 61/380,111, filed on September 3, 2010, and U.S. Provisional Patent Application No. 61/223,027, filed on July 4, 2009, all of which are incorporated herein by reference in their entirety. In addition, the present application may be related to U.S. Provisional Patent Application No. 61/451,536 filed on March 10, 2011; U.S. Provisional Patent Application No. 61/451,541 filed on March 10, 2011; and U.S. Provisional Patent Application No. 61/451,551 filed on March 10, 2011, all of which are incorporated herein by reference in their entirety.

FIELD OF THE INVENTION

[0003] The present disclosure relates to scalable video coding. Moreover in particular, it relates to joint rate distortion optimization for bitdepth and color format scalable video coding.

BACKGROUND OF THE INVENTION

[0004] Scalable video coding (SVC) is an extension of H.264/AVC, which was developed by the Joint Video Team (JVT). Enhanced content applications such as High Dynamic Range (HDR), Wide Color Gamut (WCG), spatial scalability, and 3-D have become widely popular. Along with popularity, systems and methods for delivering such content to current generation consumer set-top box decoders have become increasingly important. However, there are drawbacks in delivering such content in enhanced format. For instance, higher amounts of bandwidth may be involved in delivery of the content in enhanced format. Additionally, content providers may have to upgrade or replace their infrastructure in order to receive and/or deliver the content in enhanced format.

BRIEF DESCRIPTION OF DRAWINGS

[0005] The accompanying drawings, which are incorporated into and constitute a part of this specification, illustrate one or more embodiments of the present disclosure and, together with the description of example embodiments, serve to explain the principles and implementations of the disclosure.

FIGS. 1A-1B show exemplary bit-depth and color format scalable encoders.

FIG. 2 shows an exemplary tree structure used for encoding a block or macroblock, where nodes of the tree structure denote motion and weighted prediction parameters.

FIG. 3 shows the bit representation corresponding to the tree structure presented in FIG. 2.

FIG. 4 shows an exemplary zero tree representation of a signaling process of the macroblock/block information in context of tone mapping/scalability.

FIG. 5 shows an exemplary diagram of coding dependencies between enhancement and base layer.

FIG. 6 shows an exemplary bit-depth scalable encoder with color space conversion.

FIG. 7 shows an exemplary Overlapped Block Motion Compensation (OBMC) consideration for inter prediction or inverse tone mapping.

FIG. 8 shows an exemplary bit-depth scalable encoder with adaptive color space conversion.

FIG. 9 shows an exemplary diagram of coding dependencies in a 3D system between enhancement and base layer.

FIG. 10 shows an exemplary block diagram of encoding and decoding dependencies for bit-depth scalability.

FIG. 11 shows exemplary decoded picture buffers (DPBs) of a base layer and an enhancement layer.

FIG. 12A shows an exemplary diagram of coding dependencies involving inter-layer and intra-layer prediction.

FIG. 12B shows an exemplary diagram of coding dependencies involving inter-layer, intra-layer and temporal prediction.

FIGS. 13A-13B shows a complex prediction structure that includes prediction of RPU information from one RPU to a next RPU. FIG. 13A shows an exemplary encoder system involving an enhancement layer pre-processing and synchronization between the enhancement layer and the base layer. FIG. 13B shows the exemplary encoder system of FIG. 13A with an additional, and optional, low-complexity base layer pre-processor.

FIGS. 14A-14B show an exemplary prediction method from the base layer to the enhancement layer using a reference processing (RPU) unit element in an encoder and a decoder.

DESCRIPTION OF EXAMPLE EMBODIMENTS

[0006] According to a first aspect, a method of mapping input video data from a first layer to a second layer is described, the method comprising: providing the input video data; providing a plurality of video blocks or macroblocks, each of the video blocks or macroblocks comprising a portion of the input video data; providing a plurality of prediction methods; selecting one or more prediction methods from among the plurality of prediction methods for each of the video blocks or macroblocks; and applying, for each video block or macroblock, the selected one or more prediction methods, wherein the applying maps the video data from the first layer to the second layer.

[0007] According to a second aspect, a method of mapping input video data from a first layer to a second layer is described, the method comprising: providing the input video data for the first layer, the input video data comprising input pictures; providing a plurality of reference pictures; selecting, for each input picture, one or more reference pictures from the plurality of reference pictures, wherein the selecting is a function of each reference picture in the plurality of reference pictures and the input picture; providing a plurality of prediction methods; selecting one or more prediction methods from the plurality of prediction methods for each reference picture; and applying, for each reference picture, the selected one or more prediction methods, wherein the applying maps the input video data from the first layer to the second layer.

[0008] According to a third aspect, a method of mapping input video data from a first layer to a second layer is described, the method comprising: providing the input video data for the first layer, the input video data comprising input pictures, wherein each input picture comprises at least one region; providing a plurality of reference pictures, wherein each reference picture comprises at least one region; selecting, for each region in each input picture, one or more reference pictures or regions thereof from the plurality of reference pictures, wherein the selecting is a function of each reference picture or region and each region in each input picture; providing a plurality of prediction methods; selecting one or more prediction methods from the plurality of prediction methods for each reference picture or region; and applying, for each reference picture or region, the selected one or more

prediction methods, wherein the applying maps the input video data from the first layer to the second layer.

[0009] According to a fourth aspect, a method for optimizing distortion of video data is described, the method comprising: providing input video data comprising a base layer input picture to a base layer and an enhancement layer input picture to an enhancement layer; providing a base layer reference picture and an enhancement layer reference picture; computing a first distortion based on a difference between the base layer reference picture and the base layer input picture; computing a second distortion based on a difference between the enhancement layer reference picture and the enhancement layer input picture; and optimizing distortion of the video data by jointly considering the first distortion and the second distortion.

[0010] According to a fifth aspect, a method of processing input video data is described, the method comprising: providing a first layer and at least one second layer; providing the input video data to the first layer and the at least one second layer; pre-processing the input video data in the first layer and pre-processing the input video data in the at least one second layer, the pre-processing of the input video data in the first layer being performed synchronously with the pre-processing of the input video data in the at least one second layer; and encoding the pre-processed input video data in the first layer and the at least one second layer.

[0011] According to a sixth aspect, a method of processing input video data is described, the method comprising: providing a base layer and at least one enhancement layer; applying the input video data to the base layer and the at least one enhancement layer; and pre-processing the input video data at the at least one enhancement layer and applying the pre-processed input video data to the at least one enhancement layer and the base layer.

[0012] According to a seventh aspect, a system for removing information from video data before encoding is described, the system comprising: a base layer pre-processor connected to a base layer encoder; an enhancement layer pre-processor connected to an enhancement layer encoder; and a reference processing unit (RPU) connected between the base layer encoder and the enhancement layer encoder, wherein the base layer pre-processor and the enhancement layer pre-processor are adapted for pre-processing the video data such that the pre-processing removes information from the video data, and wherein the base layer pre-processor is adapted to operate synchronously with the enhancement layer pre-processor.

[0013] According to an eighth aspect, a system for removing information from video data before encoding is described, the system comprising: a base layer pre-processor connected to a base layer encoder; an enhancement layer pre-processor connected to an enhancement layer encoder, the enhancement layer pre-processor adapted to receive higher dynamic range video data; and a tone mapping unit connected between the base layer pre-processor and the enhancement layer pre-processor, the tone mapping unit is adapted to tone map pre-processed video data from the enhancement layer pre-processor to the base layer pre-processor.

[0014] A compatible delivery system involves creation of a scalable system which supports a legacy base layer (e.g., MPEG-2, H.264/AVC, and possibly VC1 or AVS) and additional enhancement layers with enhanced capabilities such as increased resolution, High Dynamic Range (HDR), Wide Color Gamut (WCG), and 3-D, among others. A compatible delivery system considers complexity, cost, time to market, flexibility, expandability, and compression efficiency.

[0015] An increase in complexity over existing consumer level devices can be a factor in designing a system for compatible delivery. In particular, certain constraints should be considered when designing algorithms for such applications so as to keep memory, power consumption, and processing within appropriate limits. These could be accomplished by considering the characteristics and interactions of both the base and enhancement layer codec designs, as well as, optionally, the characteristics of other associated information such as audio.

[0016] Similarly, it would be highly desirable if components from an existing system (e.g., inverse transform and quantization elements, deblocking, entropy decoding etc.) can be reused if available from the base layer, thus potentially reducing further implementation costs of such schemes.

[0017] Costs are generally related with complexity. Decoding both base and enhancement layer data using higher end devices can incur high costs, both implementation and computational. Furthermore, cost may also be affected by the amount of resources and time required for developing compatible delivery systems.

[0018] Flexibility and expandability are also generally considered in designing compatible delivery systems. More specifically, it is desirable for a compatible delivery system to provide support within for multiple different codecs as the base layer. These different codecs may include H.264/AVC as well as legacy codecs such as MPEG-2, VC-1, AVS, VP-6, VP-7, and VP-8 among others. Next Generation codecs, such as High Efficiency Video Codec (HEVC), may also be considered. Codecs can be designed in such a way as to fit, or exist within existing compatible delivery systems. In essence, this allows devices designed to support a particular compatible delivery system to also support decoding of a more optimized but single layer enhanced content bitstream without significant (if any) modifications.

[0019] Coding performance/compression efficiency can also be considered in designing compatible delivery systems. In an example illustrating coding performance/compression efficiency, consider the bitdepth scalable method of references [3][10], which extend the concepts used for spatial scalability in the context of the Scalable Video Coding extension of MPEG-4 AVC to also support bitdepth scalability. Instead of utilizing a dual loop decoding system (e.g., two decoders: one decoder for the base layer and a second decoder utilizing information of the base layer as well as its own information to decode the enhancement layer), a single decoder is utilized that adjusts its behavior depending on whether base layer decoding or enhancement layer decoding is expected. If base layer decoding is performed, then only the base layer bitstream information is decoded. Thus, a lower bitdepth image will be decoded. If enhancement layer decoding is performed, then some of the information from the base layer can be considered and decoded. The considered and decoded information, such as mode/motion information and/or residual information, can assist in the decoding of the enhancement layer and additional data. The image or residual data decoded from the base layer are used for prediction by directly up-converting base layer macroblocks using bit shifts or inverse tone mapping.

[0020] For inter pictures, motion compensation (110) was performed directly on higher bit-depth content, while a base layer residual (120) was also considered after appropriate conversion (e.g., bitdepth scaling or tone mapping), of the residual. An additional residual signal was also transmitted when this prediction method was used to avoid drift issues. A diagram of this method is presented in FIG. 1B.

[0021] The bit-depth scalable method according to references [4] and [11], considered a particular method for performing bit-depth scalability. In this method, bit depth scalability was considered by always applying inverse tone mapping to a reconstructed base layer video. Color conversion (100) can be applied prior to considering any inverse tone mapping. In that scenario, inverse tone mapping information can be adjusted for all color components accordingly. On the other hand, it may be possible that encoding for High Dynamic Range (HDR) content and, more generally, content of different bit depth and/or color-format from the bit depth and color-format used for the base layer can remain in the same color space, commonly the YUV color space, and appropriate color conversion depending on the capabilities of a display are performed, given some color transform equation at a decoder. A diagram of this method is shown in FIG. 1A. In this method, motion compensation considers 8 bit samples. Therefore, existing implementations of H.264 decoders can still be used with minor modification, if any at all. The method resembles the Fine Granularity Scalability methods previously used in MPEG-4. A plurality of methods can be specified for the inverse tone mapping methods such as, for example, linear scaling and clipping, linear interpolation, look-up table mapping, color format conversion, Nth order polynomial, and splines. More specifically:

a) Linear scaling and clipping: the current sample predictor y with a bit depth of M is obtained from its corresponding sample x in the base layer which has a bit depth of N:

$$y = \min(2^{M-N}x, 2^M - 1)$$

b) Linear interpolation using an arbitrary number of interpolation points: for a low bit depth sample with value x and two given interpolation points (x_n, y_n) and (x_{n+1}, y_{n+1}), with $x_n \le x \le x_{n+1}$, the corresponding prediction sample y with a bit depth of M is obtained as:

$$y = \min\left(y_n + \frac{x - x_n}{x_{n+1} - x_n}(y_{n+1} - y_n), 2^M - 1 \right)$$

c) Look-up table mapping**:** for each possible low bit depth sample value, the corresponding high bit depth sample value is specified.

[0022] A similar approach was also presented in references [5] and [6]. Using the base layer, after performing a color space transform in logarithmic space and an appropriate inverse tone mapping process, a residual image was generated. This residual image was then filtered, quantized from a higher bit depth space to 8 bits, and the residual image was encoded using an MPEG-4 Advanced Simple Profile (ASP) encoder. One of the primary differences

from other methods was that in this system, color space transform and log encoding were considered. Additionally, the enhancement layer was constrained to fit in 8 bits to allow the reuse of existing MPEG-4 ASP encoder and decoder implementations. Finally, this approach can also use other tools available within the MPEG-4 implementation such as inter prediction.

[0023] An enhancement of reference [11] was presented in reference [12] where weighting parameters were estimated at a macroblock level to better handle localized tone mapping. More specifically, scale s and offset o parameters for each color component can be predicted from the macroblocks above or to the left of the current macroblock, and are used in place of the inverse tone mapping information used in reference [11]. The scaling factor s and offset o can then be differentially and entropy encoded within the bitstream. Prediction y for the current sample x from the lower bit-depth image can be generated as $y = s \times x + o$. This method retained the "8-bit only motion compensation" principle of the original method. A similar method was presented in reference [9]. This method was implemented in the context of reference [3] and considered a restricted weighted prediction process to predict samples in the higher bit-depth images from the base layer.

[0024] The methods presented in references [7] and [8] are also similar to the weighted prediction methods discussed in the previous paragraphs. Reference [7] proposed encoding a log encoded lower resolution ratio image with a Low Dynamic Range (LDR) 8 bit image, which was then used to reconstruct an image of higher dynamic range, such as an HDR image. This ratio image was encoded using basic image encoding methods (e.g., using the 8×8 DCT used in JPEG and quantization) instead of performing prediction as in reference [12]. On the other hand, an offset, unlike the previous method, was not considered while no other residual signals were provided. Using operations more appropriate for linear space samples such as transform and quantization in log encoded images may have some impact on performance.

[0025] Similar methods were used also in reference [8], but instead of encoding the ratio image, a lower resolution HDR image was encoded and signaled. Using full resolution LDR and lower resolution HDR information, a decoder is used to derive full resolution HDR images. However, such a process may involve additional processing at the decoder and does not fully exploit correlations that may exist between LDR and HDR images. This could therefore potentially degrade coding efficiency. Additionally, coding efficiency and quality

can be further affected from the quantization parameters and coding decisions applied at either layer.

[0026] By examining the methods presented in the previous paragraphs, further enhancements can be made to better handle region based tone mapping. Specifically, the methods described in the present disclosure are based on either a single inverse tone mapping method as in reference [1] or a weighted prediction based method as in references [9] and [12]. A technique to extend such methods is to consider signaling of multiple inverse mapping tables or methods. More specifically, N (up to 16) inverse mapping mechanisms can be signaled simultaneously within the Sequence Parameter Sets (SPS) and/or Picture Parameter Sets (PPS), as well as within other mechanisms provided within a bitstream, such as the "reference processing unit (RPU)" as described in U.S. Provisional Patent Application No. 61/223,027. An SPS, for example, can be defined as a parameter set or coding unit comprising parameters to be applied to a video sequence, and a PPS can be defined as a parameter set or coding unit comprising parameters to be applied to one or more pictures within a sequence. An RPU can also provide signaling parameters at a similar level as the PPS, but needs not be associated with any particular codec design and can be more flexible on how information is processed or used. Such inverse mapping process can also be extended for slice headers as well. For each block or macroblock, if more than one inverse tone mapping mechanism is allowed for coding a slice/picture, then a parameter is signaled to by a selector to select the inverse tone mapping method that is used for prediction.

[0027] Syntax examples for implementing such methods are shown in Tables 1 through 4. Further details of some of the parameters can be found, for example in reference [4]. Extensions of the methods according to the present disclosure can be made to allow bi-prediction, which could allow for additional tone mapping considerations beyond those specified using a single prediction. That is, assuming N inverse mapping methods are signaled, then for every macroblock that is signaled, a prediction mode (e.g., single list prediction or bi-prediction) is also selected. If single list prediction is selected, then only one inverse mapping method is signaled. If bi-prediction is selected, then two inverse mapping methods are signaled. For the bi-prediction case, a final prediction is created as $y = (y_0 + y_1 + 1) >> 1$ where y_0 and y_1 correspond to predictions generated independently by the two inverse mapping methods. If weighted prediction is also used, the final prediction may be of the form $y = ((a_0*y_0 + a_1*y_1 + 2^{N-1}) >> N) + o$.

[0028] In another embodiment of the present disclosure, the method described above can be extended with an addition of a "skip" type prediction mode, which determines the inverse mapping method based on neighbors of the macroblock to be predicted (e.g., majority vote or smallest index in neighborhood) without signaling residuals. Additionally, modes can be signaled separately from residuals to exploit entropy coding behavior. Determining a set of efficient inverse mapping parameters can have great impact on performance. Furthermore, macroblocks can be of any size. However, 8×8 blocks may be preferred over 16×16 blocks when consideration is given to existing microprocessors.

[0029] In an alternative embodiment of the present disclosure, adaptive inverse mapping (e.g., inverse tone mapping), tables can be considered. Similarly to the methods described in reference [12], when determining an inverse mapping method to be applied to a particular block or macroblock, neighboring macroblocks of the particular macroblock can be considered. However, instead of using neighboring macroblocks/blocks to determine weighting and offset parameters, sample values in the neighboring macroblocks are considered to update a default lookup table. All pixels in all neighbors can be considered if desired, although updating the default lookup table can consider the samples of the lines above and/or on the left only. The method can also be extended for use with multiple lookup tables as well. For example, a fixed table can be used initially. A copy of the initial table is also created. However, the created copy of the initial table is adaptive instead of fixed. For every macroblock that is encoded, the adaptive table is then updated with a true relationship between base and enhancement images. The bitstream can contain a signal on whether to use the fixed or adaptive tables (maps). Furthermore, a signal can be provided that resets the adaptive table to the initial table. Again, multiple tables can also be used.

[0030] Consideration of values within the neighboring macroblocks may be unnecessary and could possibly make optimization techniques more difficult (e.g., Trellis based decision of the weighting parameters and the quantization process of the remaining residual). Therefore, weighting parameters can be differentially encoded using the weighting parameters of the neighbors directly. That is, the weighting parameters of the left, top, and top-right macroblocks can be used to predict the weighting parameters of the current macroblock directly. For example, weight' = median($weight_L$, $weight_T$, $weight_{TR}$) and offset' = median($offset_L$, $offset_T$, $offset_{TR}$). This method can be combined with the multiple inverse

tone mapping methods as described above, while deblocking can also be considered to reduce blockiness in an enhanced bit-depth image.

[0031] Weighting can also be used in combination with inverse mapping tables. Thus, instead of the weighting parameters being applied on the base layer samples directly, the weighting parameters are applied to the inverse mapped samples. The methods which consider just the base layer for prediction are more or less independent of the base layer codec. Note that similar considerations can be made for color parameters, or in predicting other color parameters using information from a first color parameter. In an example, given the method according to reference [12], and the method of the present disclosure, the weighting parameters for all components can be predicted separately, yet the same residual weighting parameters can be applied in all three components. In another example, assuming that an 8-bit YUV color space is being used where chroma components are normalized to around 128 and a weight α corresponds with the luma component, the weighted prediction for other components can be performed as described in U.S. Provisional Patent Application No. 61/380,111, where:

$$U' = \alpha \times U + 128 \times (1 - \alpha)$$
$$V' = \alpha \times V + 128 \times (1 - \alpha)$$

[0032] As shown in reference [13], considering temporal prediction within the bit-depth scalability framework can be valuable. However, if prediction directly from the enhancement layer is not provided, then the methods described herein can be difficult for a single layer approach. Similarly to the methods presented for fine granularity scalability in reference [2], for each macroblock (e.g., block of size 8×8 or 16×16), use of different coding modes and/or motion information can be specified for prediction. Specifically, the following coding modes for a macroblock can be considered:

a) Base layer prediction using inverse mapping methods as described previously

b) Base layer prediction using inverse mapping while mapping is generated by considering relationships of base and enhancement motion compensated predictions

c) Base layer skip (no additional parameter signaling or residual)

d) Inter layer prediction directly from the enhancement layer using motion information from the base layer. Correction motion vector/weighting parameter information can also be signaled to permit encoding without the presence of a base layer

e) Inter layer skip mode where motion information can be derived from the base and/or from the enhancement layer

f) Bi-prediction using inverse mapped, e.g. inverse tone mapped, base layer and temporal prediction

g) Intra prediction from the enhancement layer

h) Combined intra prediction with inter and/or base layer prediction

[0033] International Patent Application No. US2006/020633 describes an efficient scheme for coding mode and motion information based on a zero tree representation where prediction related parameters (e.g., motion vectors and weighting parameters) are differentially encoded given an easy to determine prediction (e.g., the values of the neighboring blocks). The differential parameters are then grouped together based on their relationships. For example, for a bi-predicted block, motion vectors can be grouped together based on their orientation or the list to which they belong, while weighting parameters belong in a different group. Signaling is then performed by examining which nodes contain non-zero values. For example, for the motion representation provided in the tree structure (200) of FIG. 2, if only MVD^{l0}_{x} (210) (horizontal motion vector difference component of list0) and OD (220) (offset difference) are non-zero, then eight bits are needed to signal the representation (300) (FIG. 3) in addition to the values of MVD^{l0}_{x} and OD. However, if only MVD^{l0}_{x} is non-zero, then only six bits are needed to signal the representation.

[0034] A possible representation (400) for performing the signaling within the context of bit-depth scalability is presented in FIG. 4. A prediction mode order can be established through experimentation even if the mode order is needed. Furthermore, it is possible to define slice/picture types that consider one or a subset of modes. For example, a slice type can be defined to consider inverse mapping, e.g. tone mapping, prediction. Then, a different slice type can consider intra prediction (410), while a third slice type can consider intra, single list prediction, bi-prediction (420), or single list and inverse tone mapping prediction. Other combinations are also possible depending on whether or not coding advantages can be determined due to reduced overhead representation versus a generic method. Such coding types can also be useful in the case of single layer coding since inverse tone mapping would not be useful in such cases.

[0035] Another possible method to consider inverse mapping within existing frameworks for motion compensated prediction is to add the base layer images as additional prediction references within available reference prediction lists. A base layer image is assigned one or multiple reference indices within each available list (e.g., LIST_0 and LIST_1) and also associated with a different inverse mapping process. Specifically, FIG. 5 shows the coding structure of the base layer (500) where the picture at time $t = 0$ (510) (denoted C_0) is intra encoded (I_0) (520). The picture C_0 (530) can be used to predict the enhancement layer (540) using inverse mapping when it is desired to synchronize decoding of the base and enhancement layer. Specifically, this prediction can be accomplished by encoding the enhancement layer picture E_0 (550) as an inter coded (P or B) picture, and adding C_0 as a reference within the available list. FIG. 9 shows the coding structure of FIG. 5 in a 3-D system between a left view (910), used as the base layer, and a right view (920), used as the enhancement layer.

[0036] Assuming that two different inverse mapping tables or methods are sufficient to predict E_0, then using reordering or reference picture list modification commands, C_0 can be added as a reference with indices 0 and 1 in the LIST_0 reference list, and each of the two mapping tables can then be assigned to C_0. Motion estimation and compensation can then be performed using the two references for prediction. As an additional example, for the encoding of E_1; E_0, E_2, and C_1 can be considered for prediction. C_1 can be placed as a reference in both LIST_0 and LIST_1 reference lists as reference with index 0, and E_0 and E_1 placed in LIST_0 and LIST_1 respectively, with index 1. Note that in such scenario, bi-prediction can result in combinations of different inverse mapping tables or methods as described previously. Motion estimation can be performed from the base layer to the enhancement layer to potentially provide additional performance benefits. Such concepts are reminiscent of fractal encoding as described in references [16] and [17].

[0037] FIG. 11 shows exemplary decoded picture buffers (DPBs) of a base layer and an enhancement layer. The base layer DPB (1100) comprises previously decoded base layer pictures (1130) (or previously decoded regions of base layer pictures). The enhancement layer DPB (1120) comprises previously decoded enhancement layer pictures (1140) (or previously decoded regions of enhancement layer pictures) as well as inter-layer reference pictures (1150). Specifically, the RPU can create one or more inter-layer reference pictures

given certain mapping criteria, which are specified in the RPU syntax that can then be used for predicting the enhancement layer.

[0038] By way of example and not of limitation, the RPU (1400) can contain information of how an entire picture, or regions within a picture, can be mapped (1410) from one bit depth, color space, and/or color format to another bit depth, color space, and/or color format as shown in FIGS. 14A-14B. Information contained in the RPU on one region of a picture can be used to predict other regions in the same RPU as well as predict regions in another RPU. FIG. 12A shows an exemplary diagram of coding dependencies involving inter-layer prediction (1200), where inter-layer references within a DPB can be used for prediction of the enhancement layer from the base layer. FIG. 12B shows another exemplary diagram of coding dependencies involving inter-layer prediction (1220) and temporal prediction (1210). In addition to those coding dependencies shown in FIG. 12A, temporal prediction (1210) and samples previously reconstructed from previously decoded pictures can also be utilized in prediction. Additionally, information concerning one picture or region of a picture within one RPU (1230) can be utilized in prediction of a picture or region of a picture within another RPU (1240).

[0039] A coding scheme such as that shown in FIG. 6 can be used for encoding of the enhanced content in the enhancement layer. Although such a coding scheme may appear similar to those described in reference [13], several enhancements are introduced in various elements of the system in the present disclosure, including the inverse mapping process (620), motion compensation, residual coding, and other components.

[0040] In another embodiment of the present disclosure, additional concepts can be considered to further improve performance. For example, in U.S. Patent Application 13/057,204, a simpler architecture than the methods presented in reference [14] were identified for performing Overlapped Block Motion Compensation. This method can be extended to also consider inverse mapping. Prediction on the top (710) and left (720) boundaries of a block can be altered based on the coding parameters of its neighbors as shown in FIG. 7. If the current block uses weighted prediction parameters (w_x, o_x) to perform mapping from the base layer representation to the enhancement layer representation and the blocks on the top and left uses parameters (w_T, o_T) and (w_L, o_L) respectively, then samples on the left and top of this block can use weighting parameters in the form of:

$$(d_{x,w}\times w_x + d_{L,w}\times w_L + d_{T,w}\times w_T,\ d_{x,o}\times o_x + d_{L,o}\times o_L + d_{T,o}\times o_T),$$

where parameters d specify influence of each weight to the prediction process and relate to the sample distance from each neighbor. However, since OBMC can be complicated and expensive for inter prediction, benefits should be carefully evaluated to determine whether using OBMC is justifiable within the application.

[0041] In addition to high correlation between samples from the base and enhancement layers, high correlation can also exist within the motion of the base and enhancement layers. However, encoding decisions, for example, usage of Rate Distortion Optimization at the base layer, can result in suboptimal motion vectors for the enhancement layer. Furthermore, using motion vectors directly from the base layer can affect certain implementations, particularly in cases involving hardware, where existing decoding architectures may not be reusable due to different codecs being handled differently, since motion compensation is considered in that framework. Conversely, high correlation also exists between the motion vectors of adjacent macroblocks, while inverse mapping can be the dominant prediction mode in bit-depth scalability applications.

[0042] Similarly, correlation can exist between the multiple inverse mapping tables or mechanisms used for prediction as described in previous paragraphs. Specifically, correlation can exist between same values in different tables, or between a current value and its previously encoded neighbor. Although such parameters can be transmitted once per SPS, PPS, or slice header, or within another coding unit such as an RPU, efficient encoding of these parameters can result in some coding gain. For example, one inverse tone mapping method could be described as:

$$y = [((w+\varepsilon_w)\times x + (1 << (N-1))) >> N] + (o+\varepsilon_o),$$

where weighting parameters w and o only need to be signaled once, while ε_w and ε_o are signaled for every possible x value. N allows integer only operations for the inverse tone mapping process. Since the value of ε_w and ε_o is likely to be close to or equal to 0, they can be differentially encoded and then entropy encoded, ultimately resulting in fewer bits.

[0043] In another embodiment of the present disclosure, color conversion with the SVC framework can also be considered to encode HDR content in a way that the dynamic range of the content is retained while achieving the smallest possible loss in fidelity. The encoding process can be performed in any color space, aside from any color space constraints

imposed on the base layer. Thus, the use of variable and dynamic color spaces for encoding can be implemented instead of fixing the color space for encoding of the enhancement layer in the present disclosure.

[0044] For every sequence, group of pictures (GOP), or every single picture or slice, Applicants can determine and use a color space transform which would lead to best coding efficiency. The color space transform that is applied to the base layer and the inverse color transform that is applied to the reconstructed image to achieve the appropriate HDR space can be signaled through the SPS, PPS, or for every slice header, or within a similar coding unit such as an RPU. This can be a preliminary transform process which best de-correlates the color components for compression purposes. The transform can be similar to existing transforms such as YUV to RGB or XYZ, but can also include nonlinear operations such as gamma correction.

[0045] The color transform can remain the same for a single video sequence, or can be changed and/or updated for every Instantaneous Intra Refresh (IDR) picture or at fixed or predefined intervals since the content characteristics are not likely to change rapidly. The conversion process (810) from, and to, any possible color space used by the pictures within the video bitstream may need to be specified if unknown to allow for predicting a picture of a certain color space C_1 with motion compensated prediction from pictures of a different color space C_2. An example of such process is shown in FIG. 8. Such a process can also be applicable to other applications such as encoding of infrared or thermal images, or other spaces where an original color space used for capture and/or representation may not provide the best color space for compression purposes.

[0046] As described in reference [15], encoding decisions within the base layer can affect the performance of the enhancement layer. Therefore, design aspects of normative tools within the system of the present disclosure are considered as well as methods to best design encoding and/or non-normative algorithms. For example, a system can reuse motion information for both base and enhancement layer when considering complexity decisions, while the design of joint algorithms for rate distortion optimization and rate control can result in improved performance for both layers. In particular, a lagrangian optimization can be used for Rate Distortion Optimization by minimizing the equation:

$$J = w_{base} \times D_{base} + w_{enhanced} \times D_{enhanced} + R_{total},$$

where w_{base} and $w_{enhanced}$ are the lagrangian parameters, D_{base} and $D_{enhanced}$ are the distortion at each level, and R_{total} is the total bitrate for encoding both layers. Such a process can be extended to take into account encoding of multiple pictures considering interdependencies that can exist in between the pictures. Distortion can be based on simple metrics such as, for example, the Sum of Square Errors (SSE), Sum of Absolute Differences (SAD), Structure Similarity Index Metric (SSIM), Weighted SSE, Weighted SAD, or Sum of Transformed Absolute Differences (STAD). However, different distortion metrics can also be considered to satisfy the human visual model, or for display of content on a certain display device.

[0047] Alternatively, decisions can be made for both layers for rate control/quantization, including selection of quantization parameters, adaptive rounding or trellis optimization of coded coefficients so as to satisfy all bitrate target requirements that may have been imposed while achieving best possible quality. Mode decision and/or motion parameter trellis can also be applied to determine affine parameters using, for example, a True Motion Estimation (TME) method.

[0048] Encoding performance and subjective quality can be affected by consideration of pre-processing algorithms. Pre-processing methods as shown in FIGS. 10, 13A, and 13B, attempt to remove information prior to encoding that is likely to be removed during the encoding process (e.g., noise) but are not constrained by syntax and constraints of the codec. Such methods can result in improved spatial and temporal correlation of the signal to be compressed, resulting in improved subjective quality.

[0049] FIG. 13A shows an exemplary encoder system involving enhancement layer pre-processing. Higher bit depth content input into the enhancement layer can be processed using, for example, motion compensated temporal filtering (MCTF) (1310) to produce pre-processed enhancement layer pictures. In FIG. 13A, these pre-processed enhancement layer pictures serve as inputs to an enhancement layer encoder (1320) and a tone mapping and/or color conversion module (1330) (for tone mapping and/or color converting from the enhancement layer to the base layer). Base layer pictures, formed from information from the original higher bit depth content (1350) and the pre-processed enhancement layer pictures, can then be input into a base layer encoder (1340).

[0050] In the exemplary encoder system of FIG. 13B, synchronization of the pre-processor is not necessarily required since the pre-processing that is applied to both the base layer encoder (1335) and the enhancement layers encoder (1345) take place in the

enhancement layer pre-processor. In such cases, complex pre-processing methods by way of filters such as MCTF can be utilized. FIG. 13B shows an encoder system comprising an additional, optional pre-processor (1315) in the base layer. This pre-processing takes place after the first pre-processing (1325) in the enhancement layer. Since the pre-processing in this case is not synchronized, the complexity of this additional pre-processing is constrained to further pre-processing based on information from the pre-processing method performed for the first layer, or limited to low complexity filters such as spatial filters that will not introduce or will introduce limited/controlled desynchronization.

[0051] An MCTF can be described specifically, such that a frame 2 (at t_2) can be predicted using reference pictures from the past (t_0, t_1), current (t_2), or/and future (t_3, t_4). Predictions t_{20}, t_{21}, t_{22}, t_{23}, and t_{24} (where, for example, t_{21}, denotes a prediction of a frame 2 using information from frame 1) can be used to remove noise by utilizing temporal information and form a final prediction for t_2.

[0052] For scalable systems, joint pre-processing considerations for the base and enhancement layers can be used to eliminate cases which can be difficult to predict from, and also increase layer correlation, which can result in improved coding efficiency. Pre-processing can be particularly useful when using less efficient codecs, such as MPEG-2. As an example, in 3D systems, pre-processing can help eliminate camera color misalignment issues and noise that may have been introduced on each view. Similar considerations can also apply to post-processing. The tools that have been used for content creation, such as, pre-processing, and encoding, can be used, given a specific display device, to select different post-processing methods for each layer. Such methods can be signaled either through external mechanisms (e.g., SEI message or directly through the bitstream as in U.S. Patent Application 12/999,419. FIG. 10 shows the dependencies that can exist within an entire encoding (preparation) and decoding (delivery) chain of enhanced content.

[0053] The methods and systems described in the present disclosure may be implemented in hardware, software, firmware or combination thereof. Features described as blocks, modules or components may be implemented together (e.g., in a logic device such as an integrated logic device) or separately (e.g., as separate connected logic devices). The software portion of the methods of the present disclosure may comprise a computer-readable medium which comprises instructions that, when executed, perform, at least in part, the described methods. The computer-readable medium may comprise, for example, a random

access memory (RAM) and/or a read-only memory (ROM). The instructions may be executed by a processor (e.g., a digital signal processor (DSP), an application specific integrated circuit (ASIC), or a field programmable logic array (FPGA)).

[0054] The examples set forth above are provided to give those of ordinary skill in the art a complete disclosure and description of how to make and use the embodiments of the joint rate distortion optimization for bitdepth and color format scalable video coding of the disclosure, and are not intended to limit the scope of what the inventors regard as their disclosure. Modifications of the above-described modes for carrying out the disclosure may be used by persons of skill in the art, and are intended to be within the scope of the following claims. All patents and publications mentioned in the specification may be indicative of the levels of skill of those skilled in the art to which the disclosure pertains. All references cited in this disclosure are incorporated by reference to the same extent as if each reference had been incorporated by reference in its entirety individually.

[0055] It is to be understood that the disclosure is not limited to particular methods or systems, which can, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting. As used in this specification and the appended claims, the singular forms “a,” “an,” and “the” include plural referents unless the content clearly dictates otherwise. The term “plurality” includes two or more referents unless the content clearly dictates otherwise. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the disclosure pertains.

[0056] A number of embodiments of the disclosure have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the present disclosure. Accordingly, other embodiments are within the scope of the following claims.

LIST OF REFERENCES

[1] *Advanced Video Coding for Generic Audiovisual Services*, ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), ITU-T and ISO/IEC JTC 1, Version 1: May 2003, Version 2: May 2004, Version 3: Mar. 2005, Version 4: Sept. 2005, Version 5 and Version 6: June 2006, Version 7: Apr. 2007, Version 8 (including SVC extension): Consented in July 2007, http://www.itu.int/rec/recommendation.asp?type=folders&lang=e&parent=T-REC-H.264.

[2] A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G. B. Akar, G. Triantafyllidis, and A. Koz, "Coding Algorithms for 3DTV – A Survey," *in IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1606–1621, Nov. 2007.

[3] Y. Gao and Y. Wu, "Applications and Requirement for Color Bit Depth Scalability," Joint Video Team, Doc. JVT-U049, Hangzhou, China, October 2006.

[4] M. Winken, H. Schwarz, D. Marpe, and T. Wiegand, "SVC bit depth scalability," Joint Video Team, Doc. JVT-V078, Marrakech, Morocco, January 2007.

[5] R. Mantiuk, A. Efremov, K. Myszkowski, and H. P. Seidel, "Backward Compatible High Dynamic Range MPEG Video Compression," in Proc. of SIGGRAPH '06 (Special issue of ACM Transactions on Graphics), 25 (3), pp. 713-723, 2006.

[6] R. Mantiuk, G. Krawczyk, K. Myszkowski, and H. P.Seidel, "High Dynamic Range Image and Video Compression – Fidelity Matching Human Visual Performance," in Proc. of IEEE International Conference on Image Processing 2007, pp. 9-12.

[7] G. Ward, and M. Simmons, "JPEG-HDR: A Backwards-Compatible, High Dynamic Range Extension to JPEG," Proceedings of the Thirteenth Color Imaging Conference, November 2005.

[8] G. Ward, "A General Approach to Backwards-Compatible Delivery of High Dynamic Range Images and Video," Proceedings of the Fourteenth Color Imaging Conference, November 2006.

[9] A. Segall and Y. Su, "System for bit-depth scalable coding," Joint Video Team, Doc. JVT-W113, San Jose, CA, April 2007.

[10] Y. Wu and Y. Gao, "Study on Inter-layer Prediction in Bit-Depth Scalability," Joint Video Team, JVT-X052, Geneva, Switzerland, June 2007.

[11] M. Winken, H. Schwarz, D. Marpe, and T. Wiegand, "CE2: SVC bit-depth scalability," Joint Video Team, JVT-X057, Geneva, Switzerland, June 2007.

[12] S. Liu, A. Vetro and W.-S. Kim, "Inter-layer Prediction for SVC Bit-Depth Scalable Coding," Joint Video Team, JVT-X075, Geneva, Switzerland, June 2007.

[13] Y. Ye, H. Chung, M. Karczewicz, and I. S. Chong, "Improvements to Bit Depth Scalability Coding," Joint Video Team, JVT-Y048, Shenzhen, China, October 2007.

[14] M.T. Orchard and G.J. Sullivan, "Overlapped block motion compensation: an estimation-theoretic approach," IEEE Trans. on Image Processing, vol. 3, no. 5, pp. 693-699, September 1994.

[15] H. Schwarz and T. Wiegand, "R-D optimized multilayer encoder control for SVC," in Proceedings of the IEEE International Conference on Image Processing (ICIP) 2007, San Antonio, TX, Sep. 2007.

[16] M. F. Barnsley and L. P. Hurd, Fractal Image Compression, AK Peters, Ltd., Wellesley, 1993.

[17] N. Lu, Fractal Imaging, Academic Press, USA, 1997.

CLAIMS

1. A method for optimizing distortion of video data, the method comprising:

providing input video data comprising a base layer input picture to a base layer and an enhancement layer input picture to an enhancement layer;

providing a base layer reference picture and an enhancement layer reference picture;

computing a first distortion based on a difference between the base layer reference picture and the base layer input picture;

computing a second distortion based on a difference between the enhancement layer reference picture and the enhancement layer input picture; and

optimizing distortion of the video data by jointly considering the first distortion and the second distortion.

2. The method of claim 1, wherein the joint consideration of the first distortion and the second distortion is performed using weight factors in a Lagrangian equation.
3. The method of any one of claims 1 or 2, wherein the first distortion and the second distortion are selected from the group consisting of sum of square errors (SSE), sum of absolute differences (SAD), structure similarity index metric (SSIM), weighted SSE, weighted SAD, and sum of transformed absolute differences.
4. An encoder for optimizing distortion of video data according to the method recited in any one of claims 1-3.
5. An apparatus for optimizing distortion of video data according to the method recited in any one of claims 1-3.
6. A system for optimizing distortion of video data according to the method recited in any one of claims 1-3.
7. A computer-readable medium containing a set of instructions that causes a computer to perform the method recited in any one of claims 1-3.
8. Use of the method recited in any one of claims 1-3 to optimize distortion of video data.

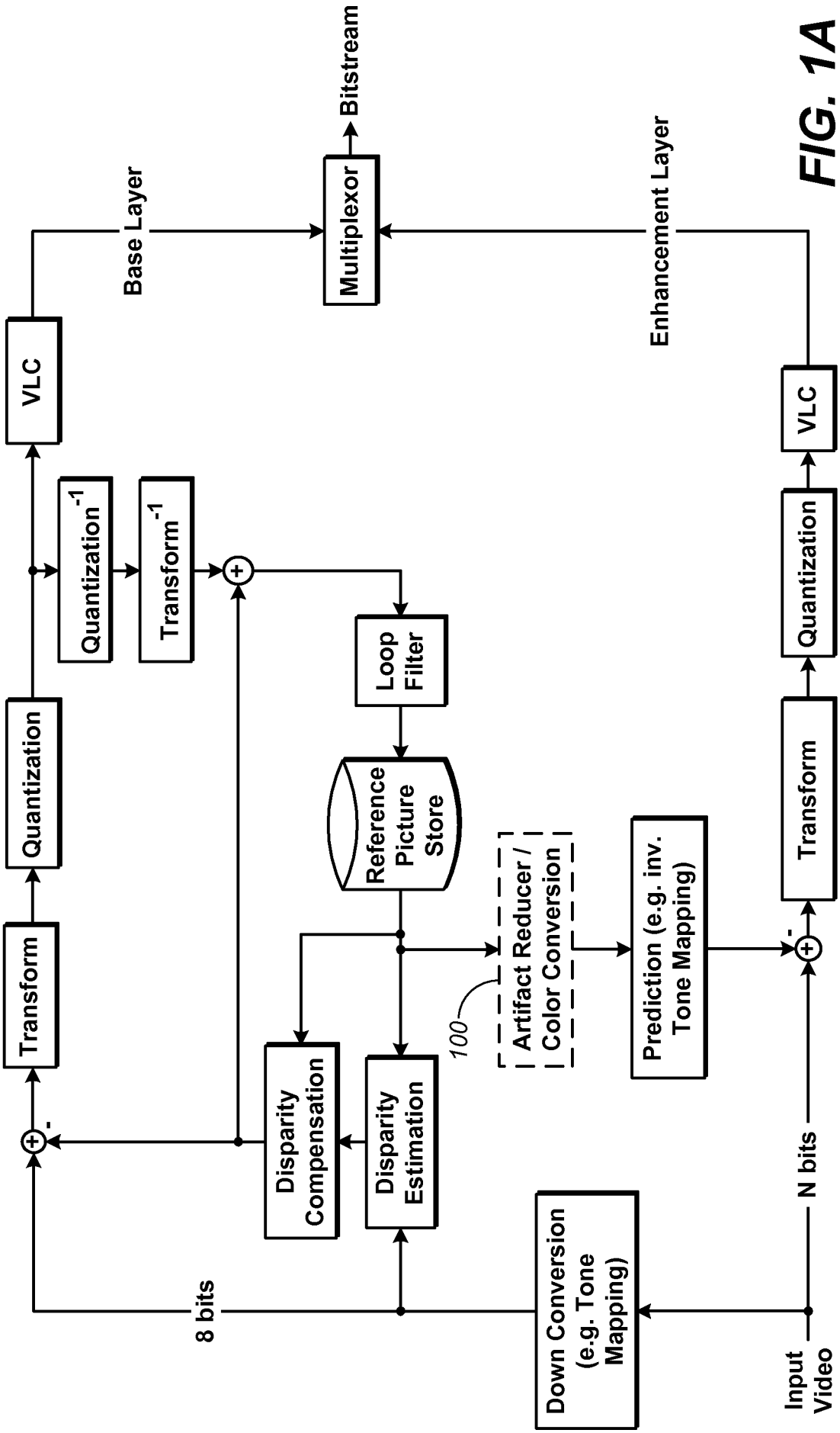

Bitstream
Multiplexor
Base Layer
Enhancement Layer
VLC
Quantization
Transform
Quantization^{-1}
Transform^{-1}
Loop Filter
Reference Picture Store
Artifact Reducer / Color Conversion
100
Prediction (e.g. inv. Tone Mapping)
Disparity Compensation
Disparity Estimation
Down Conversion (e.g. Tone Mapping)
8 bits
N bits
Input Video
Transform
Quantization
VLC

FIG. 1A

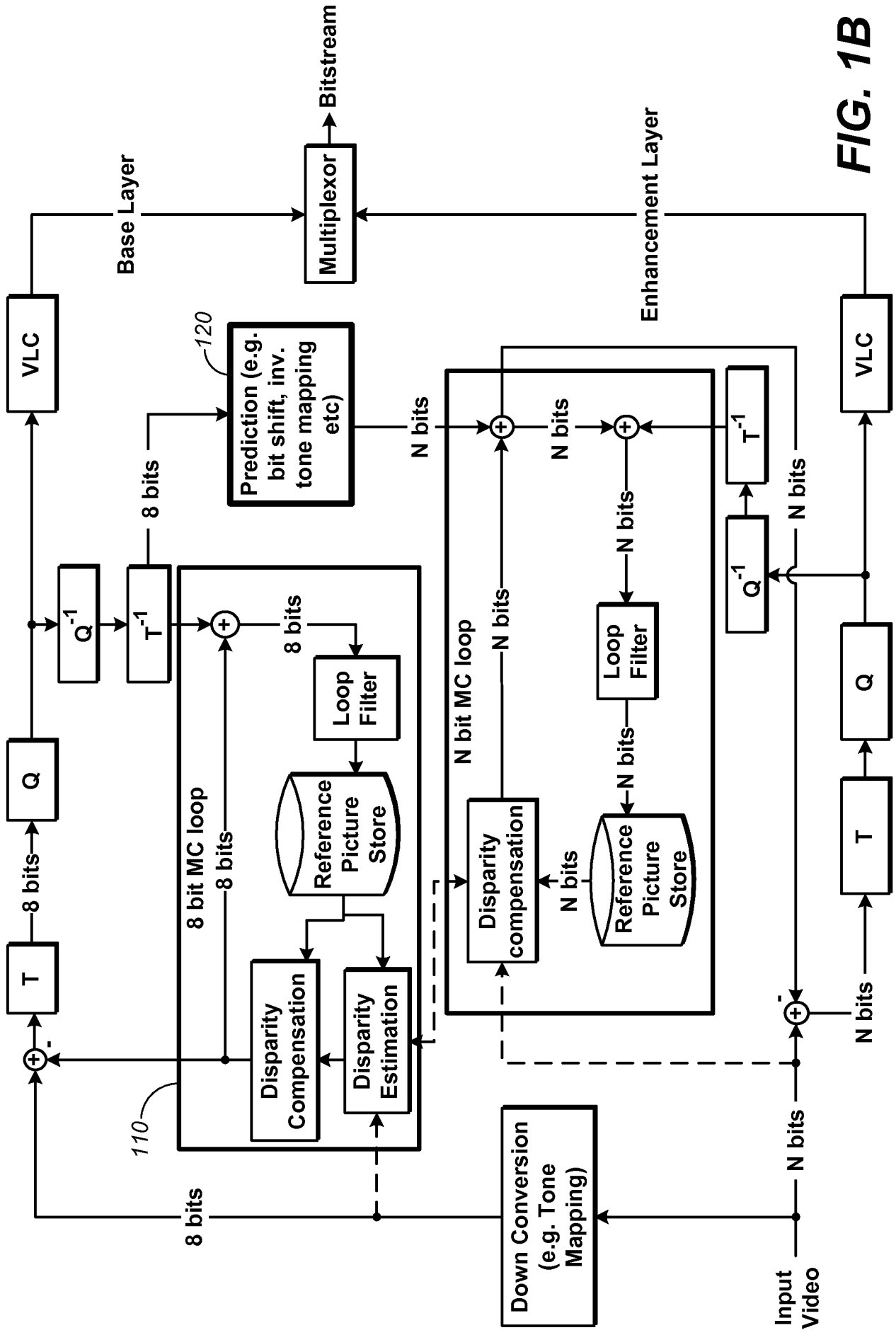

Base Layer
Multiplexor
Bitstream
Enhancement Layer
VLC
120
Prediction (e.g. bit shift, inv. tone mapping etc)
8 bits
N bits
Q
T
Q-1
T-1
8 bit MC loop
N bit MC loop
Loop Filter
Reference Picture Store
Disparity Compensation
Disparity Estimation
Disparity compensation
110
Down Conversion (e.g. Tone Mapping)
Input Video

FIG. 1B

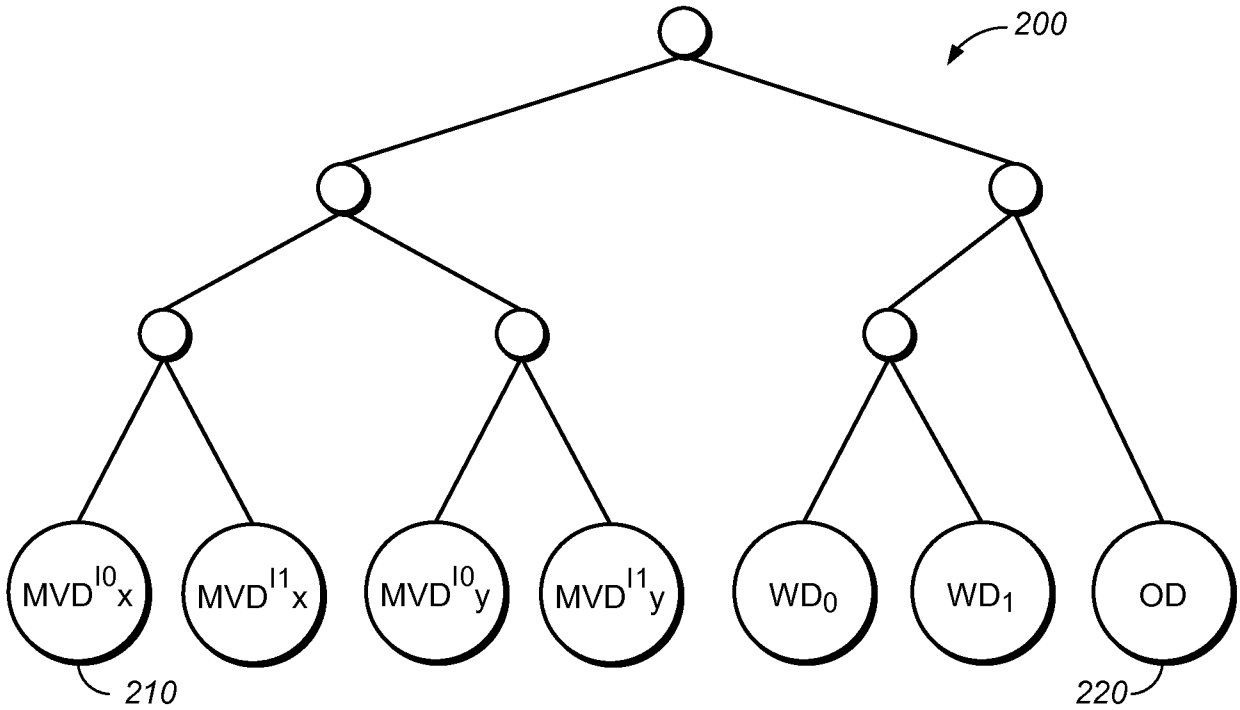

200
MVD^l0x
MVD^l1x
MVD^l0y
MVD^l1y
WD0
WD1
OD
210
220

FIG. 2

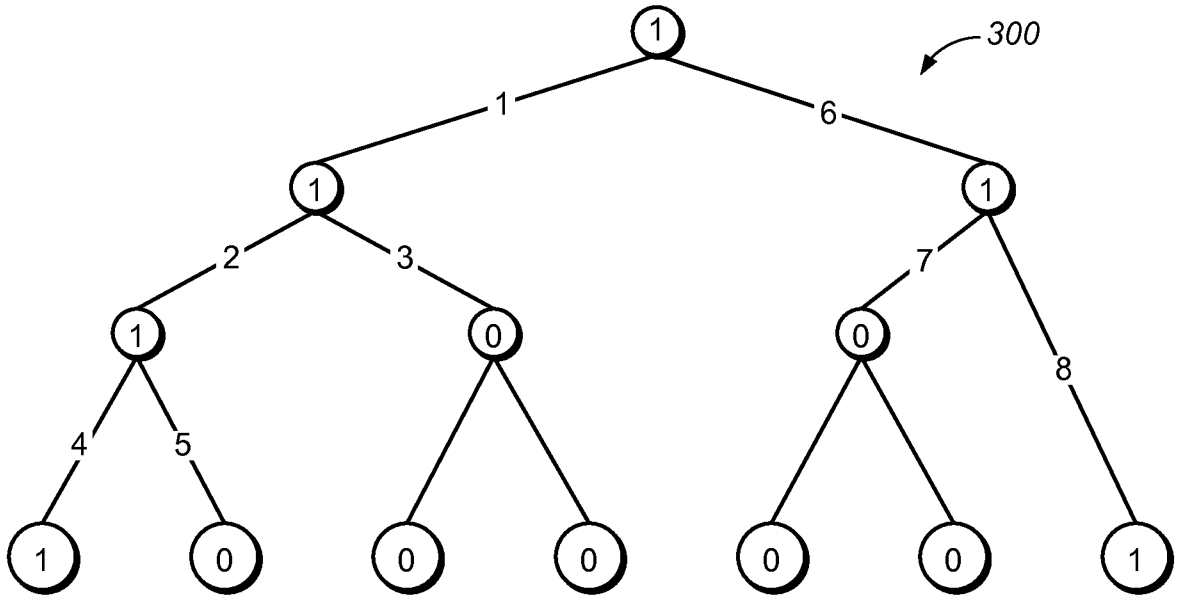

300
1
1
6
1
1
2
3
7
1
0
0
8
4
5
1
0
0
0
0
0
1

FIG. 3

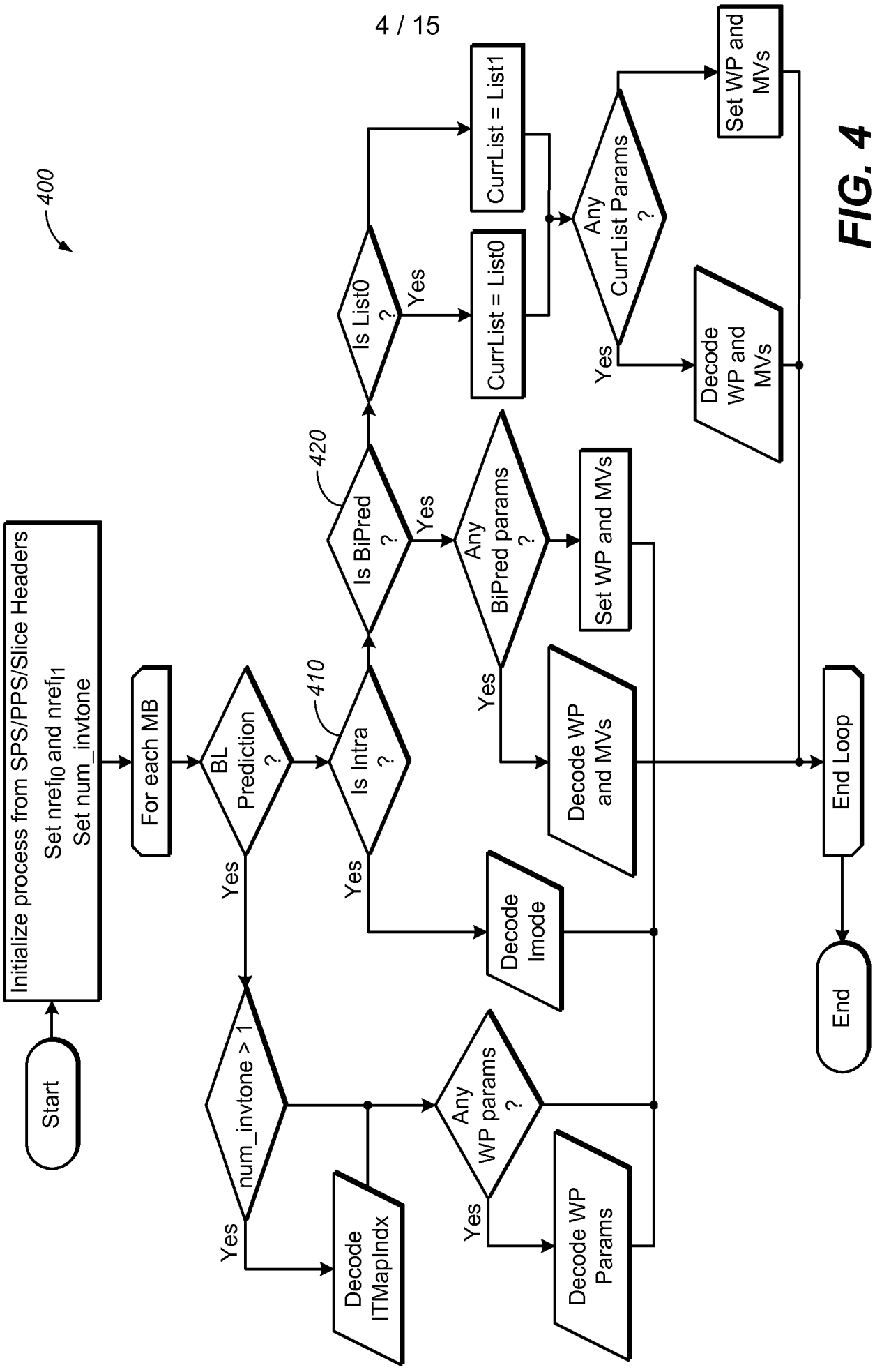

400
Start
Initialize process from SPS/PPS/Slice Headers
Set nref$_{l0}$ and nref$_{l1}$
Set num_invtone
For each MB
BL Prediction ?
Yes
num_invtone > 1
Yes
Decode ITMapIndx
Any WP params ?
Yes
Decode WP Params
410
Is Intra ?
Yes
Decode Imode
420
Is BiPred ?
Yes
Any BiPred params ?
Yes
Decode WP and MVs
Set WP and MVs
Is List0 ?
Yes
CurrList = List0
CurrList = List1
Any CurrList Params ?
Yes
Decode WP and MVs
Set WP and MVs
End Loop
End

FIG. 4

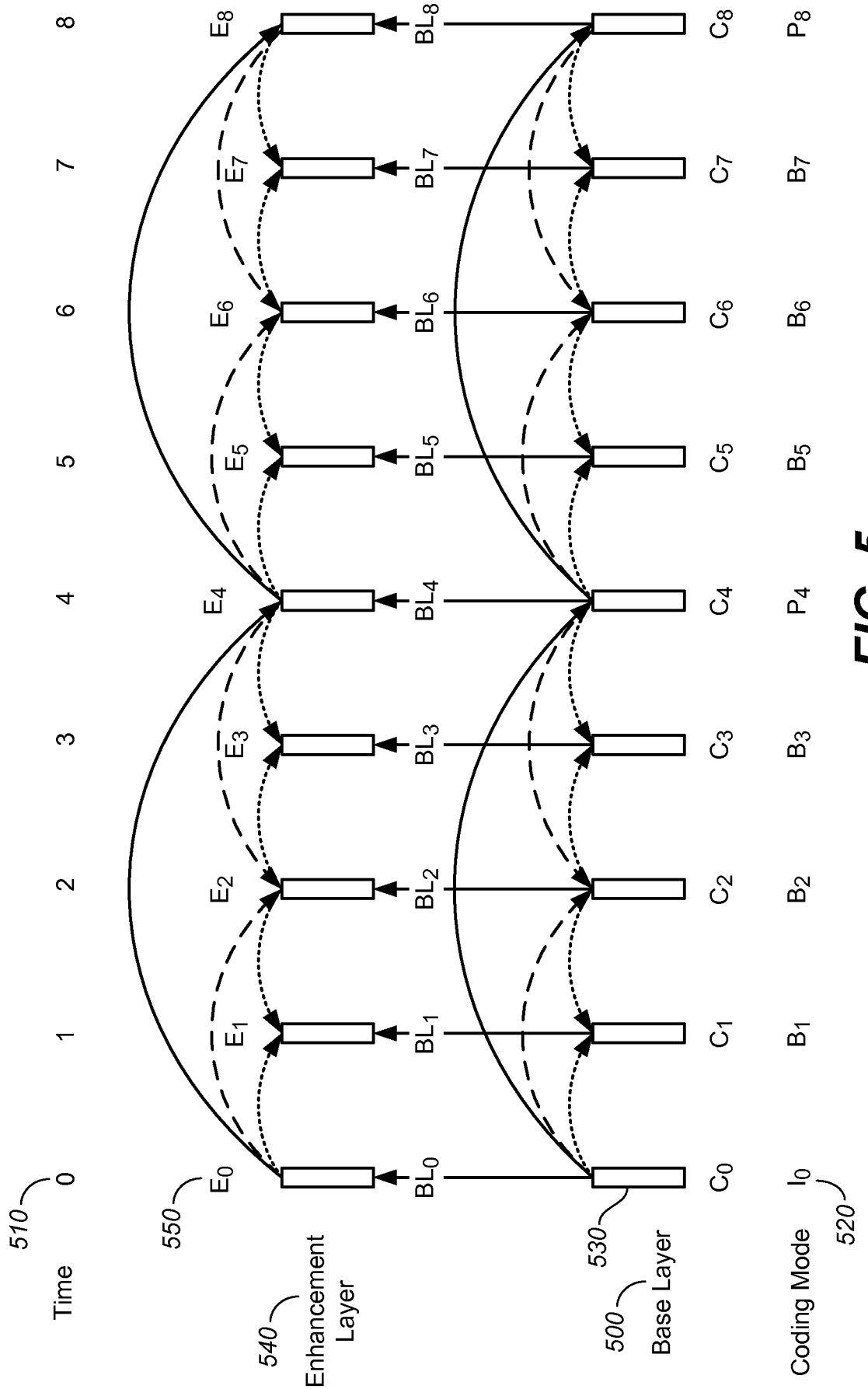

Time
0
1
2
3
4
5
6
7
8
510
Coding Mode
I0
520
Base Layer
500
530
Enhancement Layer
540
550
C0
C1
C2
C3
C4
C5
C6
C7
C8
B1
B2
B3
P4
B5
B6
B7
BL0
BL1
BL2
BL3
BL4
BL5
BL6
BL7
BL8
E0
E1
E2
E3
E4
E5
E6
E7
E8

FIG. 5

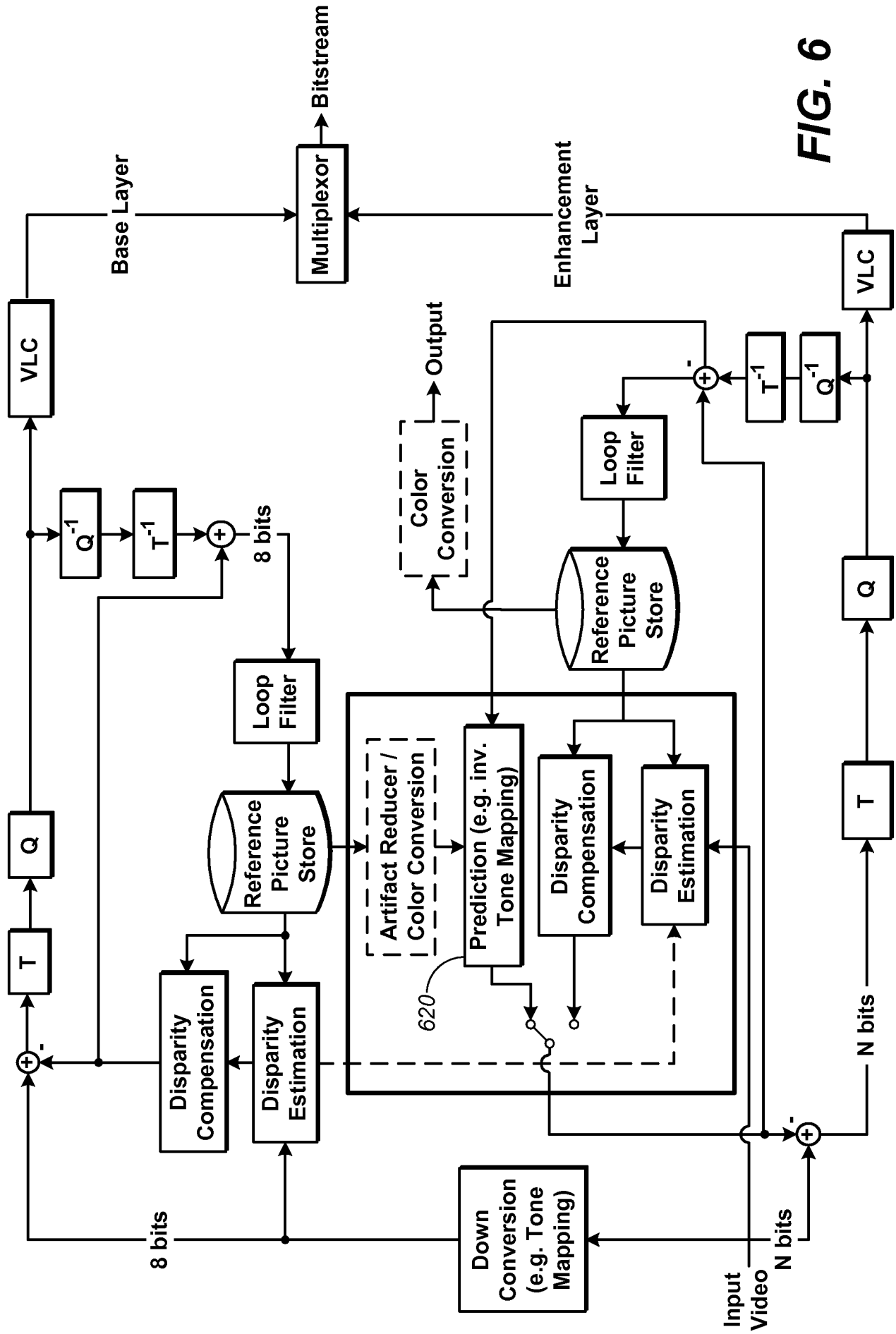

Base Layer
Multiplexor
Bitstream
Enhancement Layer
VLC
VLC
Output
Color Conversion
Loop Filter
Reference Picture Store
Q
T
Q
T
Q-1
T-1
8 bits
Loop Filter
Reference Picture Store
Artifact Reducer / Color Conversion
Prediction (e.g. inv. Tone Mapping)
Disparity Compensation
Disparity Estimation
620
Disparity Compensation
Disparity Estimation
8 bits
Down Conversion (e.g. Tone Mapping)
Input Video
N bits
N bits

FIG. 6

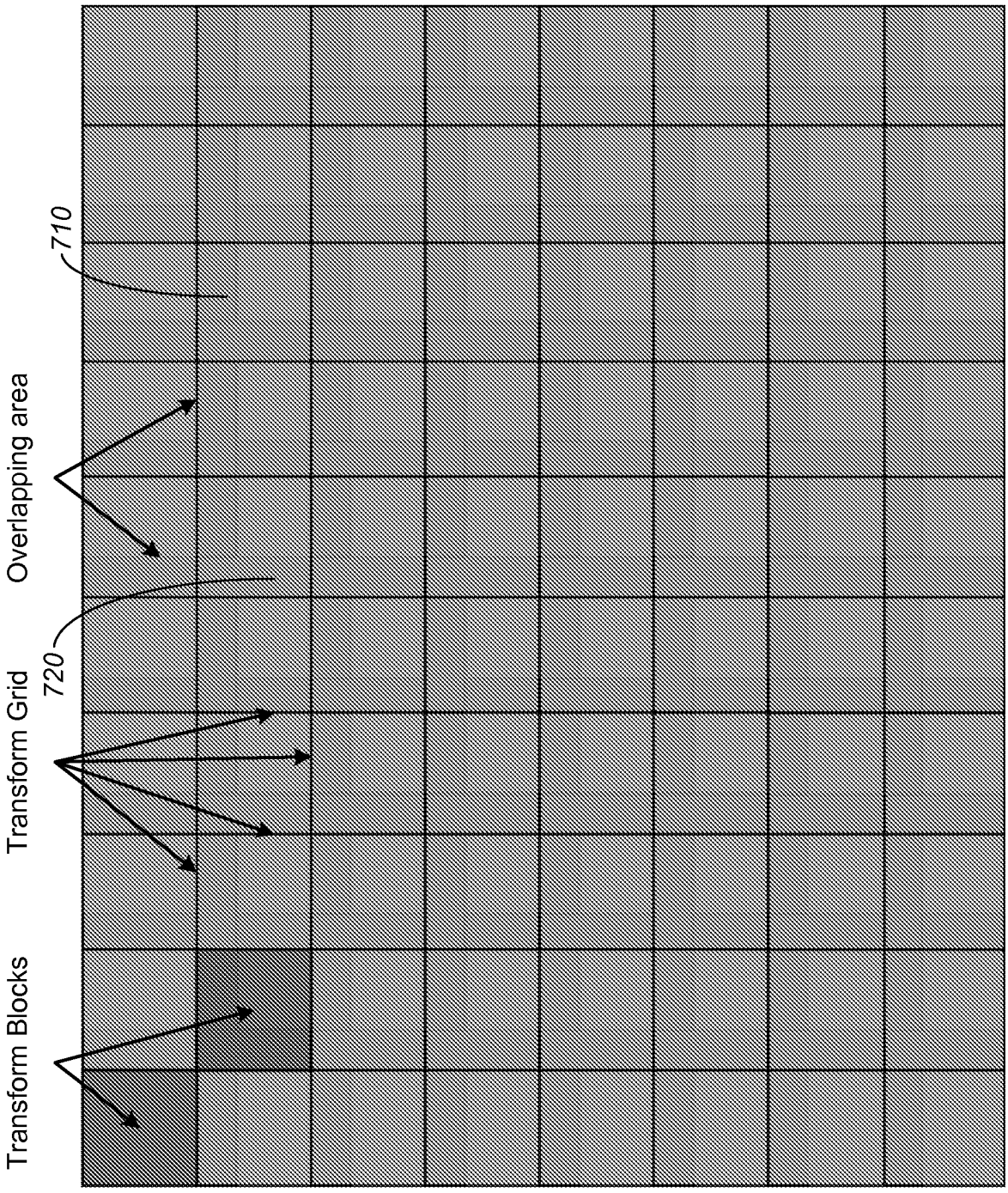


FIG. 7

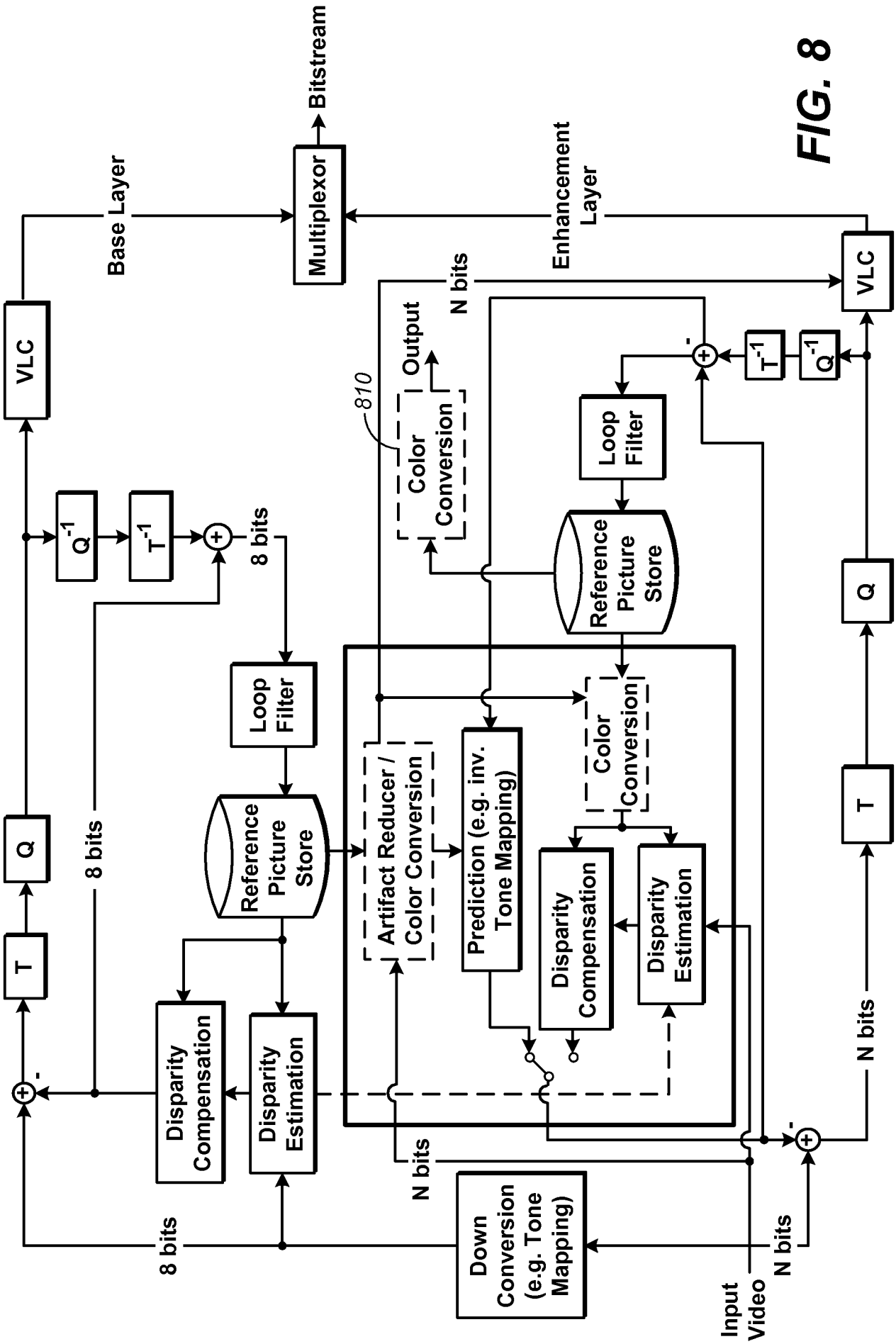

Bitstream
Multiplexor
Base Layer
Enhancement Layer
VLC
VLC
Q
T
Q
T
Q-1
T-1
Q-1
T-1
8 bits
8 bits
N bits
N bits
Output
Color Conversion
810
Loop Filter
Loop Filter
Reference Picture Store
Reference Picture Store
Artifact Reducer / Color Conversion
Prediction (e.g. inv. Tone Mapping)
Color Conversion
Disparity Compensation
Disparity Estimation
Disparity Compensation
Disparity Estimation
Down Conversion (e.g. Tone Mapping)
Input Video
N bits
8 bits
N bits

FIG. 8

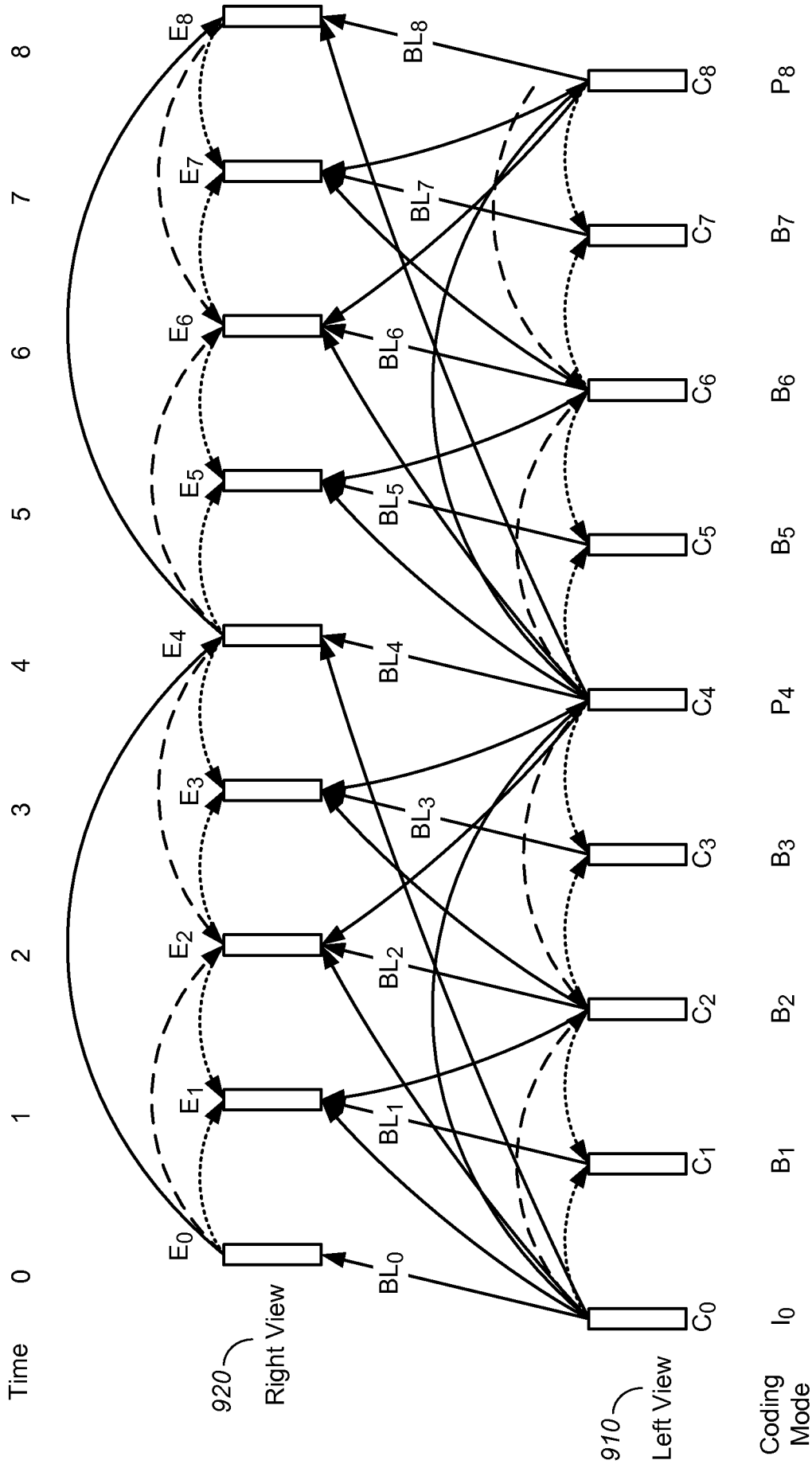

Time
0
1
2
3
4
5
6
7
8
920
Right View
910
Left View
Coding Mode
I_0
P_4
P_8
B_1
B_2
B_3
B_5
B_6
B_7
E_0
E_1
E_2
E_3
E_4
E_5
E_6
E_7
E_8
BL_0
BL_1
BL_2
BL_3
BL_4
BL_5
BL_6
BL_7
BL_8
C_0
C_1
C_2
C_3
C_4
C_5
C_6
C_7
C_8

FIG. 9

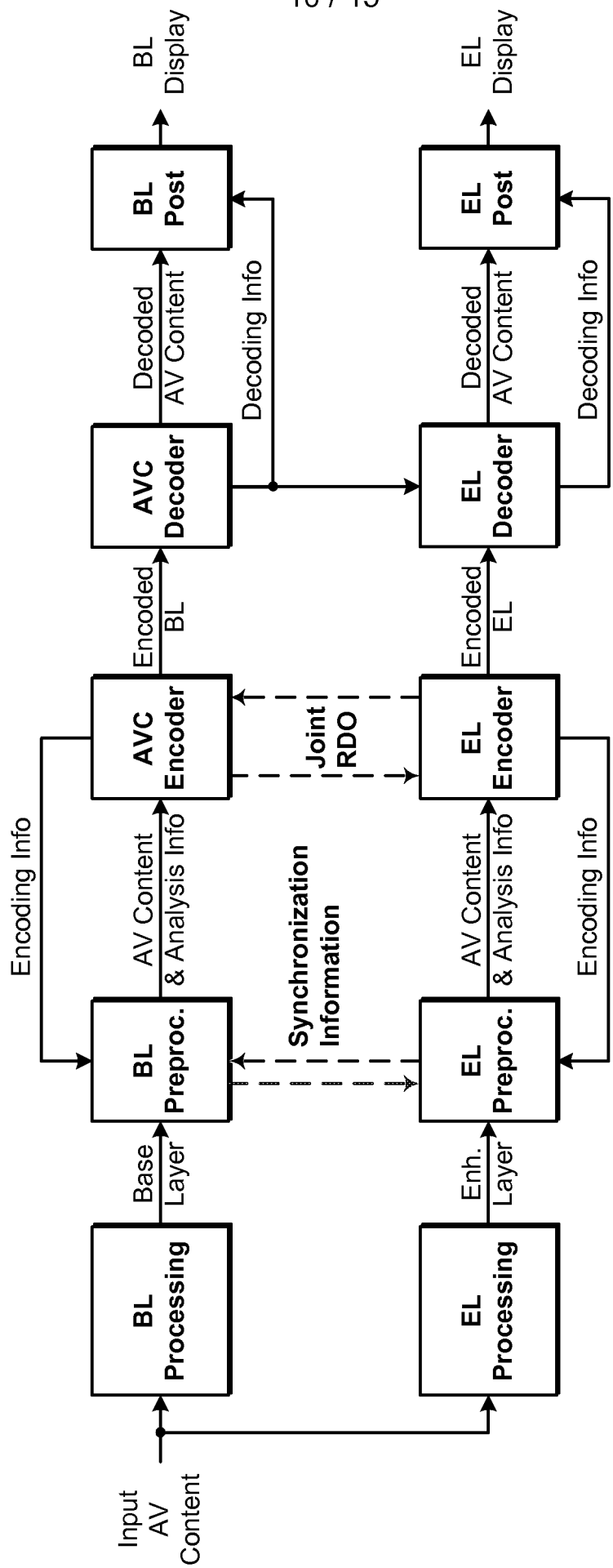

Input AV Content
BL Processing
Base Layer
EL Processing
Enh. Layer
BL Preproc.
EL Preproc.
Synchronization Information
AV Content & Analysis Info
Encoding Info
AVC Encoder
EL Encoder
Joint RDO
Encoded BL
Encoded EL
AVC Decoder
EL Decoder
Decoded AV Content
Decoding Info
BL Post
EL Post
BL Display
EL Display

FIG. 10

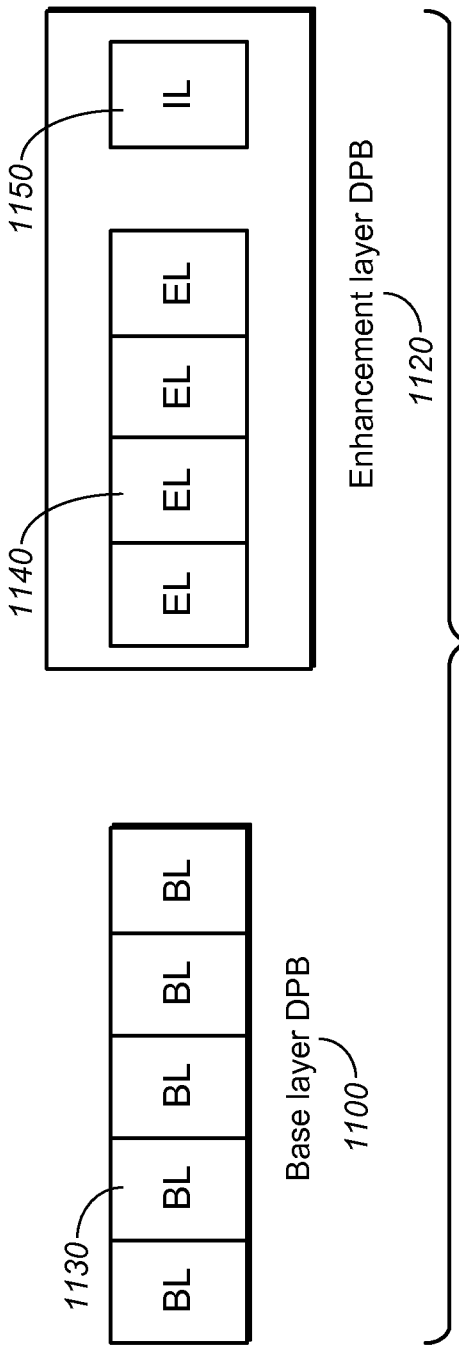


FIG. 11

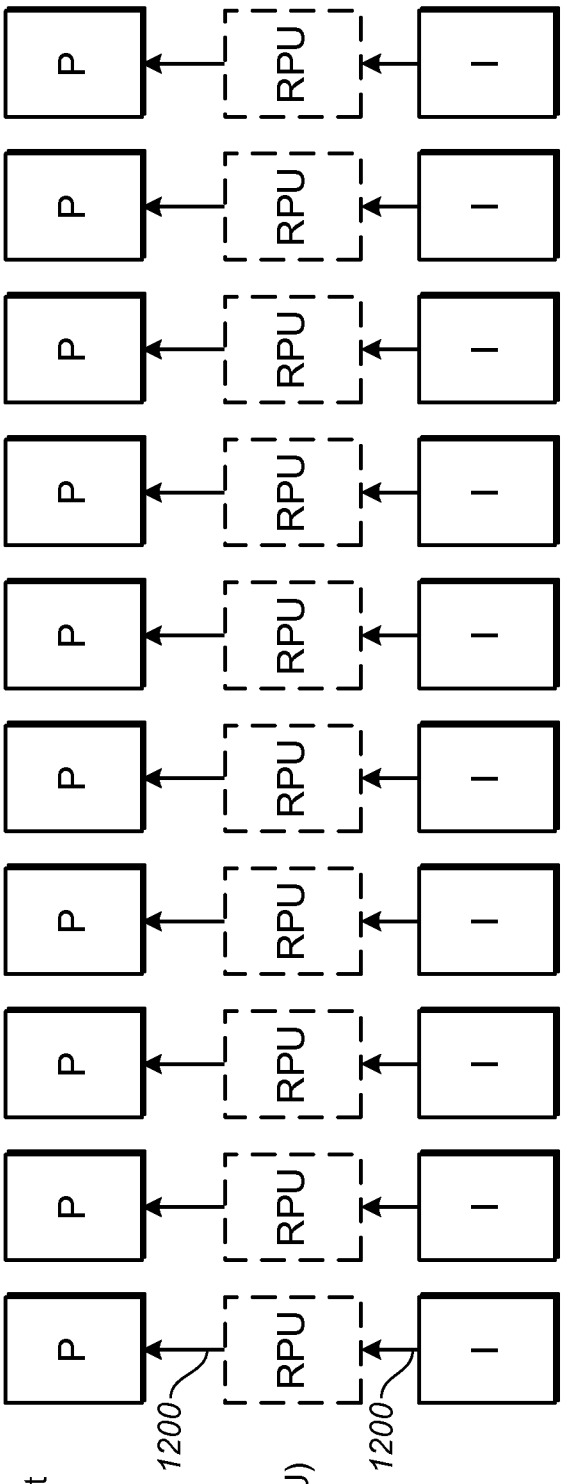


FIG. 12A

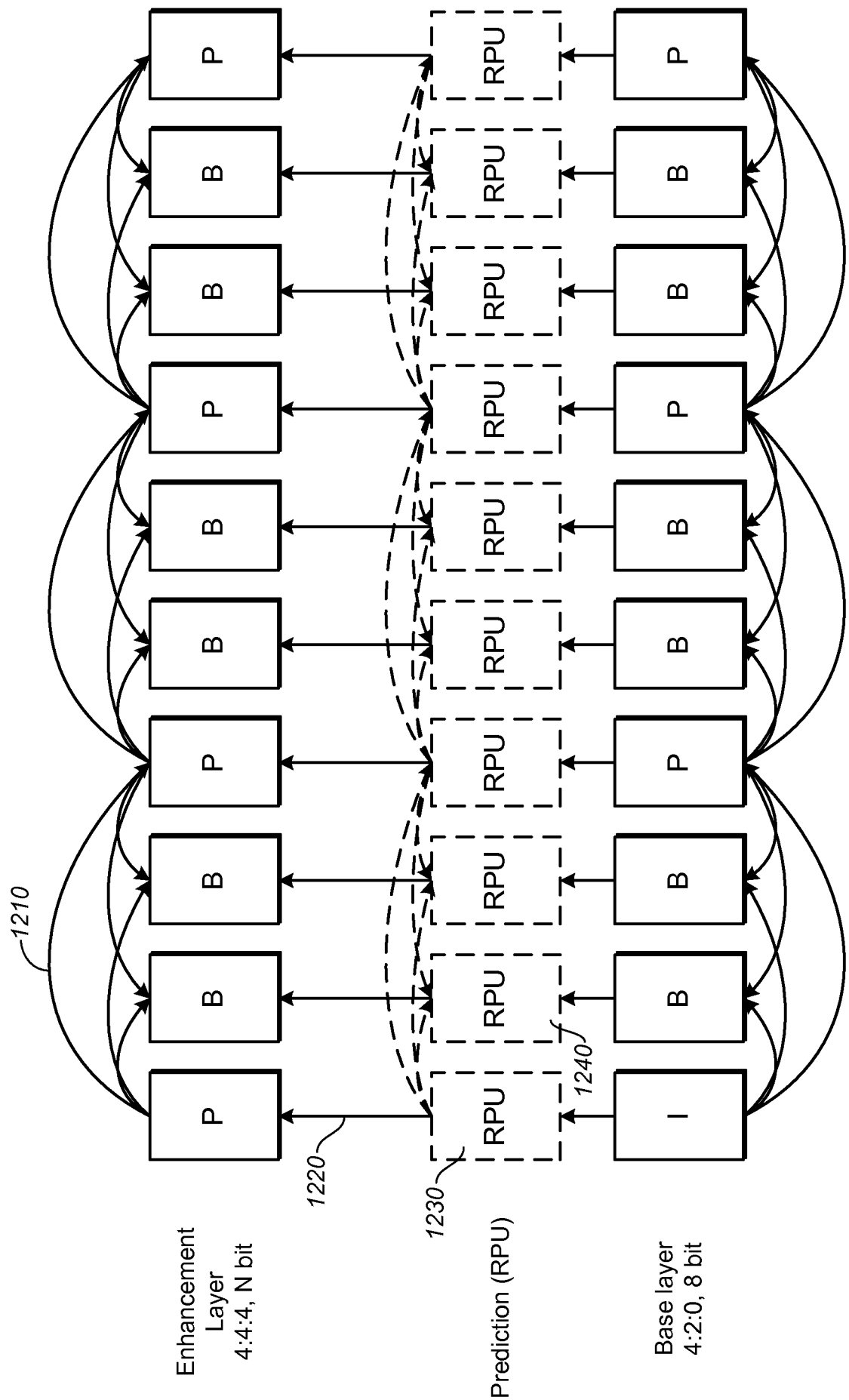

1210
P
B
B
P
B
B
P
B
B
P
1220
RPU
RPU
RPU
RPU
RPU
RPU
RPU
RPU
1230
1240
I
B
B
P
B
B
P
B
B
P
Enhancement Layer 4:4:4, N bit
Prediction (RPU)
Base layer 4:2:0, 8 bit

FIG. 12B

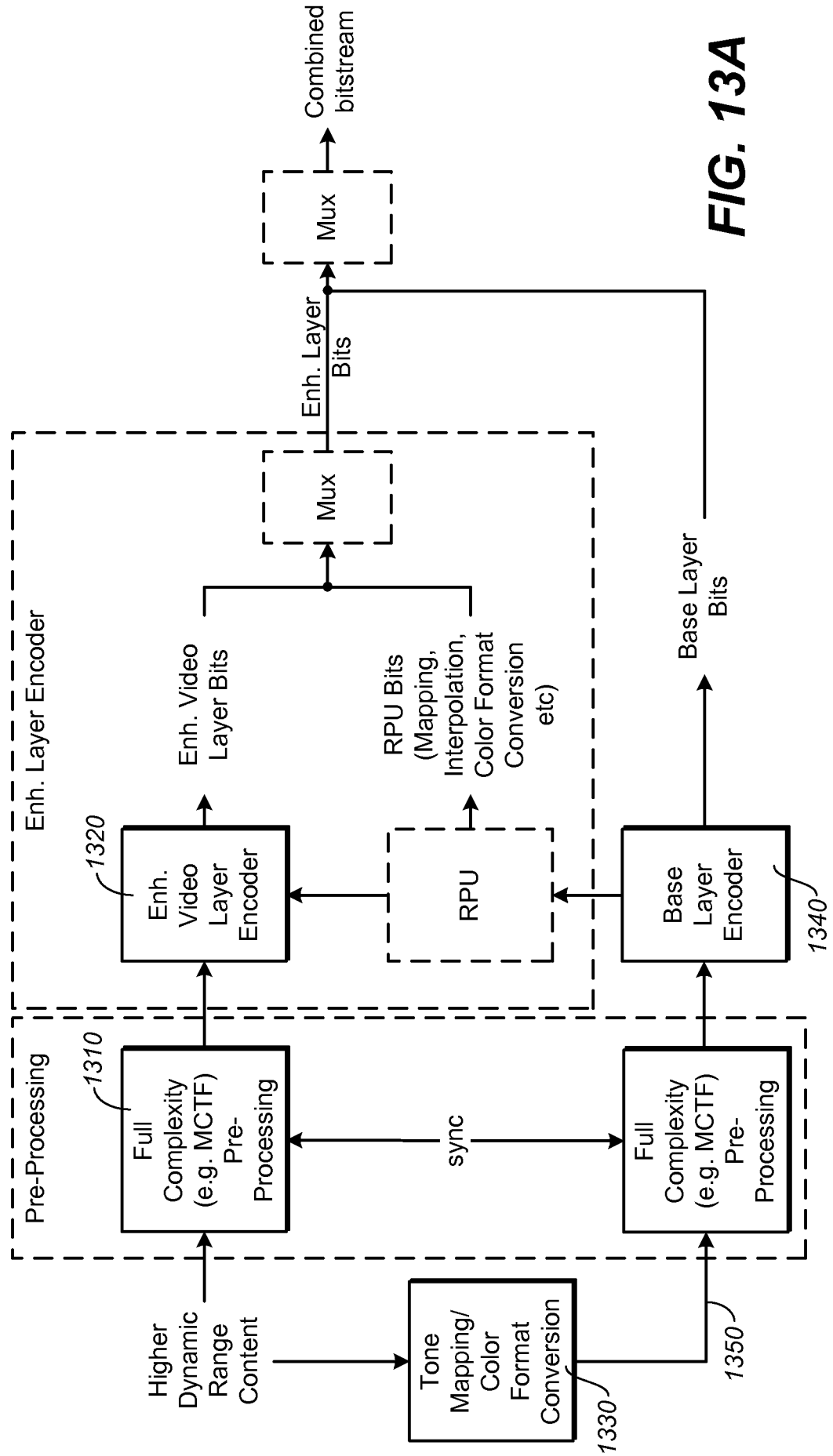

Higher Dynamic Range Content
Tone Mapping/ Color Format Conversion
1330
1350
Pre-Processing
1310
Full Complexity (e.g. MCTF) Pre-Processing
sync
Full Complexity (e.g. MCTF) Pre-Processing
Enh. Layer Encoder
1320
Enh. Video Layer Encoder
RPU
Base Layer Encoder
1340
Enh. Video Layer Bits
RPU Bits (Mapping, Interpolation, Color Format Conversion etc)
Base Layer Bits
Mux
Enh. Layer Bits
Mux
Combined bitstream

FIG. 13A

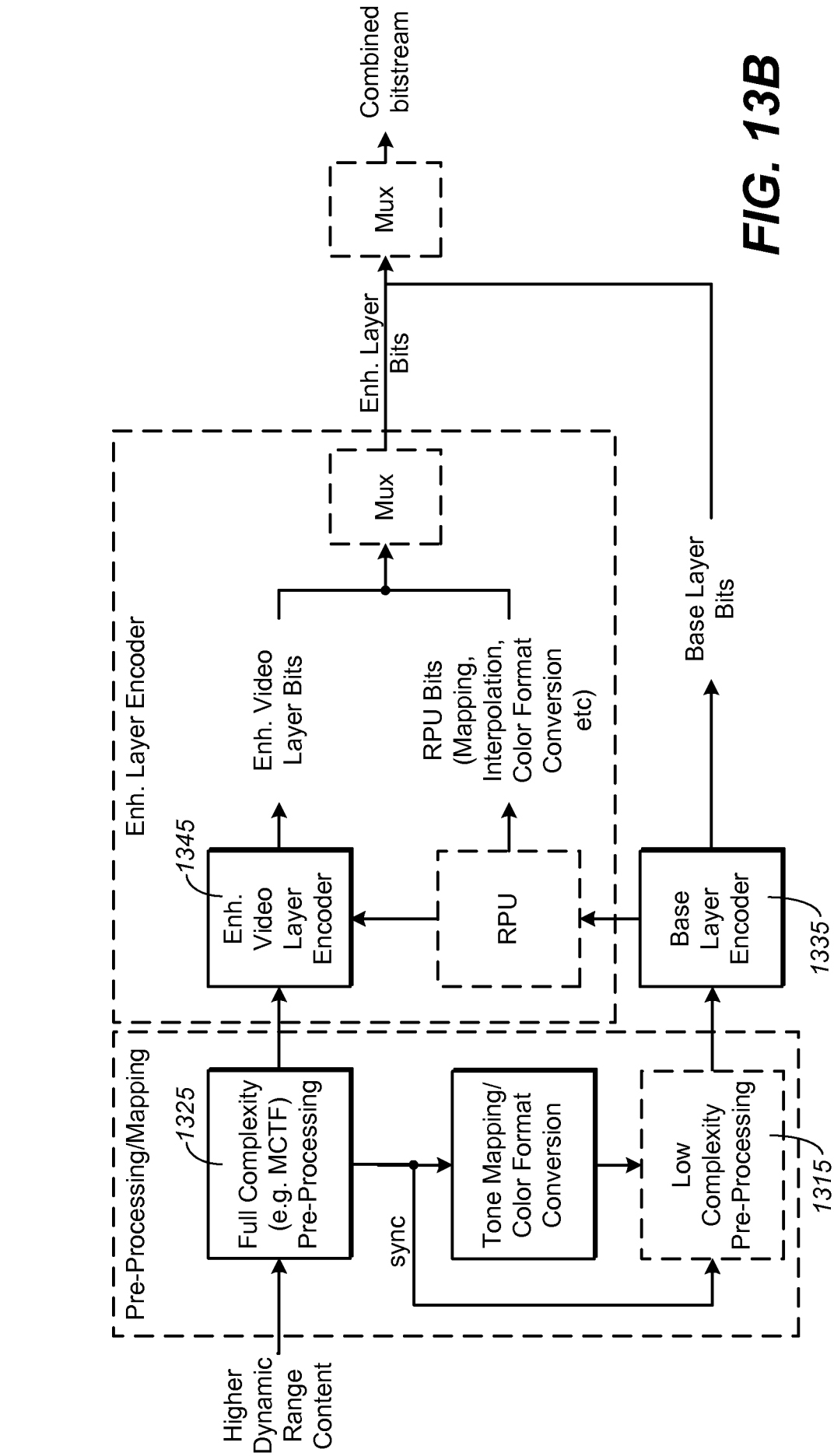

Higher Dynamic Range Content
Pre-Processing/Mapping
1325
Full Complexity (e.g. MCTF) Pre-Processing
sync
Tone Mapping/ Color Format Conversion
Low Complexity Pre-Processing
1315
Enh. Layer Encoder
1345
Enh. Video Layer Encoder
RPU
RPU Bits (Mapping, Interpolation, Color Format Conversion etc)
Base Layer Encoder
1335
Base Layer Bits
Enh. Video Layer Bits
Mux
Enh. Layer Bits
Mux
Combined bitstream

FIG. 13B

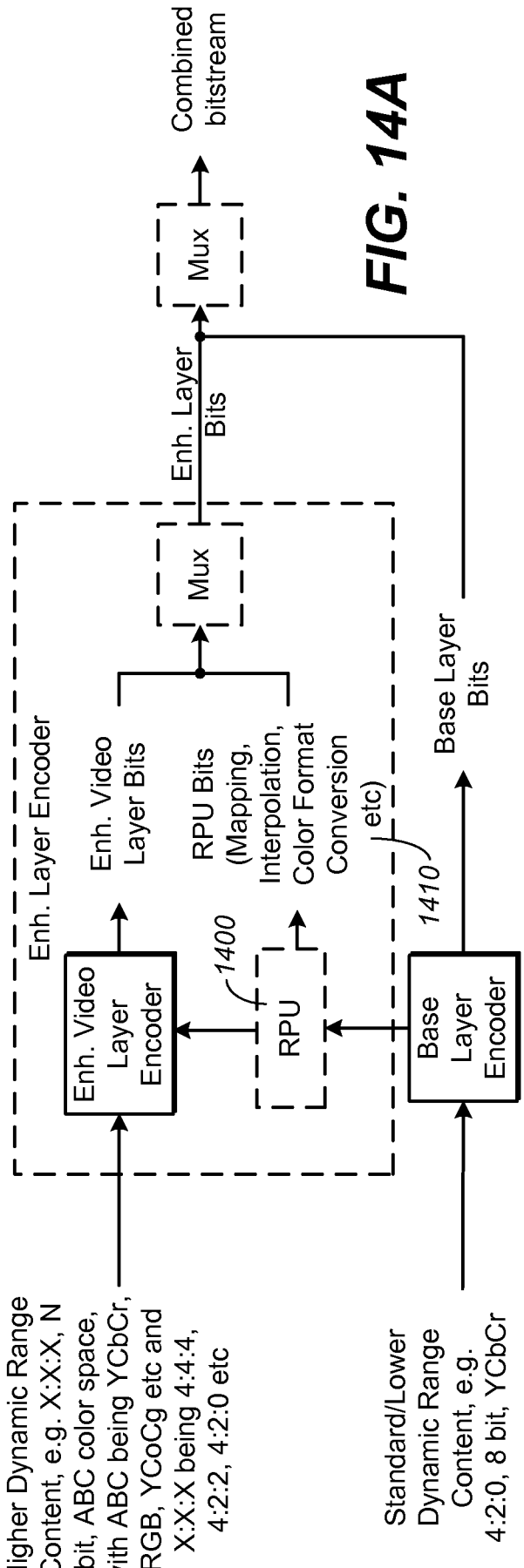


FIG. 14A

FIG. 14B

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2012/028366

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04N7/26 H04N7/36
ADD.

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)
H04N

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)
EPO-Internal, COMPENDEX, INSPEC, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HEIKO SCHWARZ ET AL: "R-D Optimized Multi-Layer Encoder Control for SVC", IMAGE PROCESSING, 2007. ICIP 2007. IEEE INTERNATIONAL CONFERENCE ON, IEEE, PI, 1 September 2007 (2007-09-01), pages II-281, XP031157916, ISBN: 978-1-4244-1436-9 cited in the application the whole document -----	1-8
X	LI XIANG ET AL: "Lagrange multiplier selection for rate-distortion optimization in SVC Ã Â", PICTURE CODING SYMPOSIUM 2009; 6-5-2009 - 8-5-2009; CHICAGO,, 6 May 2009 (2009-05-06), XP030081854, the whole document -----	1-8
	-/--	

[X] Further documents are listed in the continuation of Box C.

[X] 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
11 May 2012	22/05/2012

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Martinière, Anthony

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2012/028366

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	WO 2012/050758 A1 (DOLBY LAB LICENSING CORP [US]; LEONTARIS ATHANASIOS [US]; TOURAPIS ALE) 19 April 2012 (2012-04-19) figure 11 figures 13A,13B figures 22,23	1-8
A	WO 2010/127692 A1 (ERICSSON TELEFON AB L M [SE]; RUSERT THOMAS [DE]; ANDERSSON KENNETH [S) 11 November 2010 (2010-11-11) figure 10	1-8
A	WO 2009/155398 A1 (DOLBY LAB LICENSING CORP [US]; LI ZHEN [US]; TOURAPIS ALEXANDROS [US]) 23 December 2009 (2009-12-23) figure 7A figure 10A	1-8
A	ZHEN LI ET AL: "An Estimation-Theoretic Interpretation of Video Rate Distortion Optimization with Lagrangian Formulation", DATA COMPRESSION CONFERENCE, 2008. DCC 2008, IEEE, PISCATAWAY, NJ, USA, 25 March 2008 (2008-03-25), pages 222-231, XP031241179, ISBN: 978-0-7695-3121-2 the whole document	1-8
A	WO 2010/033565 A1 (DOLBY LAB LICENSING CORP [US]; LI ZHEN [US]; TOURAPIS ALEXANDROS [US]) 25 March 2010 (2010-03-25)	1-8
A	XIANG LI ET AL: "Laplace Distribution Based Lagrangian Rate Distortion Optimization for Hybrid Video Coding", IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE SERVICE CENTER, PISCATAWAY, NJ, US, vol. 19, no. 2, 1 February 2009 (2009-02-01), pages 193-205, XP011239534, ISSN: 1051-8215, DOI: 10.1109/TCSVT.2008.2009255 the whole document	1-8

-/--

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2012/028366

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WINKEN M ET AL: "SVC bit depth scalability", 22. JVT MEETING; 79. MPEG MEETING; 13-01-2007 - 20-01-2007; MARRAKECH,MA; (JOINT VIDEO TEAM OF ISO/IEC JTC1/SC29/WG11 AND ITU-T SG.16),, no. JVT-V078, 14 January 2007 (2007-01-14) , XP030006886, ISSN: 0000-0156 the whole document -----	1-8

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2012/028366

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2012050758 A1	19-04-2012	NONE	
WO 2010127692 A1	11-11-2010	EP 2428042 A1	14-03-2012
		WO 2010127692 A1	11-11-2010
WO 2009155398 A1	23-12-2009	AT 536048 T	15-12-2011
		CN 102067603 A	18-05-2011
		EP 2297963 A1	23-03-2011
		EP 2439951 A2	11-04-2012
		EP 2439952 A2	11-04-2012
		JP 2011525332 A	15-09-2011
		US 2011103473 A1	05-05-2011
		WO 2009155398 A1	23-12-2009
WO 2010033565 A1	25-03-2010	CN 102165772 A	24-08-2011
		EP 2338282 A1	29-06-2011
		JP 2012503367 A	02-02-2012
		KR 20110056326 A	26-05-2011
		US 2011170591 A1	14-07-2011
		WO 2010033565 A1	25-03-2010