

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
28 March 2002 (28.03.2002)

PCT

(10) International Publication Number
WO 02/25954 A2

(51) International Patent Classification⁷: **H04N 7/50**

(21) International Application Number: PCT/EP01/10874

(22) International Filing Date:
18 September 2001 (18.09.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/234,499 22 September 2000 (22.09.2000) US
09/887,743 21 June 2001 (21.06.2001) US

(71) Applicant: **KONINKLIJKE PHILIPS ELECTRONICS N.V.** [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) Inventors: **VAN DER SCHAAR, Mihaela**; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). **RADHA, Hayder**; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(74) Agent: **LANDOUSY, Christian**; Internationaal Octrooibureau B.V., Prof Holstlaan 6, NL-5656 AA Eindhoven (NL).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished upon receipt of that report
- entirely in electronic form (except for this front page) and available upon request from the International Bureau

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: DOUBLE-LOOP MOTION-COMPENSATION FINE GRANULAR SCALABILITY

(57) **Abstract:** A video coding technique having motion compensation within a fine granular scalable coded enhancement layer. In one embodiment, the video coding technique involves a two-loop prediction-based enhancement layer including non-motion-predicted enhancement layer I- and P-frames and motion-predicted enhancement layer B-frames. The motion-predicted enhancement layer B-frames are computed using: 1) motion-prediction from two temporally adjacent differential I- and P- or P- and P- frame residuals, and 2) the differential B-frame residuals obtained by subtracting the decoded base layer B-frame residuals from the original base layer B-frame residuals. In a second embodiment, the enhancement layer further includes motion-predicted enhancement layer P-frames. The motion-predicted enhancement layer P-frames are computed using: 1) motion-prediction from a temporally adjacent differential I- or P-frame residual, and 2) the differential P-frame residual obtained by subtracting the decoded base layer P-frame residual from the original base layer P-frame residual.



WO 02/25954 A2

Double-loop motion-compensation fine granular scalability

The present invention relates to video coding, and more particularly to a scalable enhancement layer video coding scheme that employs motion compensation within the enhancement layer for bi-directional predicted frames (B-frames) and predicted frames and bi-directional predicted frames and (P- and B-frames).

5

Scalable enhancement layer video coding has been used for compressing video transmitted over computer networks having a varying bandwidth, such as the Internet. A current enhancement layer video coding scheme employing fine granular scalable coding techniques (adopted by the ISO MPEG-4 standard) is shown in FIG. 1. As can be seen, the video coding scheme 10 includes a prediction-based base layer 11 coded at a bit rate R_{BL} , and an FGS enhancement layer 12 coded at R_{EL} .

The prediction-based base layer 11 includes intraframe coded I frames, interframe coded P frames which are temporally predicted from previous I- or P-frames using motion estimation-compensation, and interframe coded bi-directional B-frames which are temporally predicted from both previous and succeeding frames adjacent the B-frame using motion estimation-compensation. The use of predictive and/or interpolative coding i.e., motion estimation and corresponding compensation, in the base layer 11 reduces temporal redundancy therein.

The enhancement layer 12 includes FGS enhancement layer I-, P-, and B-frames derived by subtracting their respective reconstructed base layer frames from the respective original frames (this subtraction can also take place in the motion-compensated domain). Consequently, the FGS enhancement layer I-, P- and B-frames in the enhancement layer are not motion-compensated. (The FGS residual is taken from frames at the same time-instance.) The primary reason for this is to provide flexibility which allows truncation of each FGS enhancement layer frame individually depending on the available bandwidth at transmission time. More specifically, the fine granular scalable coding of the enhancement layer 12 permits an FGS video stream to be transmitted over any network session with an available bandwidth ranging from $R_{min} = R_{BL}$ to $R_{max} = R_{BL} + R_{EL}$. For example, if the

available bandwidth between the transmitter and the receiver is $B=R$, then the transmitter sends the base layer frames at the rate R_{BL} and only a portion of the enhancement layer frames at the rate $R_{EL} = R - R_{BL}$. As can be seen from FIG. 1, portions of the FGS enhancement layer frames in the enhancement layer can be selected in a fine granular scalable manner for transmission. Therefore, the total transmitted bit-rate is $R = R_{BL} + R_{EL}$.
5 Because of its flexibility in supporting a wide range of transmission bandwidth with a single enhancement layer.

FIG. 2 shows a block-diagram of a conventional FGS encoder for coding the base layer 11 and enhancement layer 12 of the video coding scheme of FIG. 1. As can be
10 seen, the enhancement layer residual of frame i ($FGSR(i)$) equals $MCR(i) - MCRQ(i)$, where $MCR(i)$ is the motion-compensated residual of frame i , and $MCRQ(i)$ is the motion-compensated residual of frame i after the quantization and the dequantization processes.

Although the current FGS enhancement layer video coding scheme 10 of FIG. 1 is very flexible, it has the disadvantage that its performance in terms of video image quality
15 is relatively low compared with that of a non-scalable coder functioning at the same transmission bit-rate. The decrease in image quality is not due to the fine granular scalable coding of the enhancement layer 12 but mainly due to the reduced exploitation of the temporal redundancy among the FGS residual frames within the enhancement layer 12. In particular, the FGS enhancement layer frames of the enhancement layer 12 are derived only
20 from the motion-compensated residual of their respective base layer I-, P-, and B-frames, no FGS enhancement layer frames are used to predict other FGS enhancement layer frames in the enhancement layer 12 or other frames in the base layer 11.

Accordingly, a scalable enhancement layer video coding scheme is needed that employs motion-compensation in the enhancement layer to improve image quality while
25 preserving most of the flexibility and attractive characteristics typical to the current FGS video coding scheme.

The present invention is directed to an enhancement layer video coding
30 scheme, and in particular an FGS enhancement layer video coding scheme that employs motion compensation within the enhancement layer for predicted and bi-directional predicted frames. One aspect of the invention involves a method comprising the steps of: coding an uncoded video with a non-scalable codec to generate base layer frames; computing differential frame residuals from the uncoded video and the base layer frames, at least

portions of certain ones of the differential frame residuals being operative as references; applying motion-compensation to the at least portions of the differential frame residuals that are operative as references to generate reference motion-compensated differential frame residuals; and subtracting the reference motion-compensated differential frame residuals
5 from respective ones of the differential frame residuals to generate motion-predicted enhancement layer frames.

Another aspect of the invention involves a method comprising the steps of: decoding a base layer stream to generate base layer video frames; decoding an enhancement layer stream to generate differential frame residuals, at least portions of certain ones of the
10 differential frame residuals being operative as references; applying motion-compensation to the at least portions of the differential frame residuals operative as references to generate reference motion-compensated differential frame residuals; adding the reference motion-compensated differential frame residuals with respective ones of the differential frame residuals to generate motion-predicted enhancement layer frames; and combining the motion-
15 predicted enhancement layer frames with respective ones of the base layer frames to generate an enhanced video.

Still another aspect of the invention involves a memory medium for encoding video, which comprises code for non-scalable encoding an uncoded original video into base layer frames; code for computing differential frame residuals from the uncoded original video
20 and the base layer frames, at least portions of certain ones of the differential frame residuals being operative as references; code for applying motion-compensation to the at least portions of the differential frame residuals that are operative as references to generate reference motion-compensated differential frame residuals; and code for subtracting the reference motion-compensated differential frame residuals from respective ones of the differential
25 frame residuals to generate motion-predicted enhancement layer frames.

A further aspect of the invention involves a memory medium for decoding a compressed video having a base layer stream and an enhancement layer stream, which comprises: code for decoding the base layer stream to generate base layer video frames; code for decoding the enhancement layer stream to generate differential frame residuals, at least
30 portions of certain ones of the differential frame residuals being operative as references; code for applying motion-compensation to the at least portions of the differential frame residuals operative as references to generate reference motion-compensated differential frame residuals; code for adding the reference motion-compensated differential frame residuals with respective ones of the differential frame residuals to generate motion-predicted enhancement

layer frames; and code for combining the motion-predicted enhancement layer frames with respective ones of the base layer frames to generate an enhanced video.

Still a further aspect of the invention involves an apparatus for coding video, which comprises: means for non-scalable coding an uncoded original video to generate base
5 layer frames; means for computing differential frame residuals from the uncoded original video and the base layer frames, at least portions of certain ones of the differential frame residuals being operative as references; means for applying motion-compensation to the at least portions of the differential frame residuals that are operative as references to generate reference motion-compensated differential frame residuals; and means for subtracting the
10 reference motion-compensated differential frame residuals from respective ones of the differential frame residuals to generate motion-predicted enhancement layer frames.

Still another aspect of the invention involves an apparatus for decoding a compressed video having a base layer stream and an enhancement layer stream, which comprises: means for decoding the base layer stream to generate base layer video frames;
15 means for decoding the enhancement layer stream to generate differential frame residuals, at least portions of certain ones of the differential frame residuals being operative as references; means for applying motion-compensation to the at least portions of the differential frame residuals operative as references to generate reference motion-compensated differential frame residuals; means for adding the reference motion-compensated differential frame residuals
20 with respective ones of the differential frame residuals to generate motion-predicted enhancement layer frames; and means for combining the motion-predicted enhancement layer frames with respective ones of the base layer frames to generate an enhanced video.

25 The advantages, nature, and various additional features of the invention will appear more fully upon consideration of the illustrative embodiments now to be described in detail in connection with accompanying drawings where like reference numerals identify like elements throughout the drawings:

FIG. 1 shows a current enhancement layer video coding scheme;

30 FIG. 2 shows a block-diagram of a conventional encoder for coding the base layer and enhancement layer of the video coding scheme of FIG. 1;

FIG. 3A shows an enhancement layer video coding scheme according to a first exemplary embodiment of the present invention;

FIG. 3B shows an enhancement layer video coding scheme according to a second exemplary embodiment of the present invention;

FIG. 4 shows a block-diagram of an encoder, according to an exemplary embodiment of the present invention, that may be used for generating the enhancement layer video coding scheme of FIG. 3A;

FIG. 5 shows a block-diagram of an encoder, according to an exemplary embodiment of the present invention, that may be used for generating the enhancement layer video coding scheme of FIG. 3B;

FIG. 6 shows a block-diagram of a decoder, according to an exemplary embodiment of the present invention, that may be used for decoding the compressed base layer and enhancement layer streams generated by the encoder of FIG. 4;

FIG. 7 shows a block-diagram of a decoder, according to an exemplary embodiment of the present invention, that may be used for decoding the compressed base layer and enhancement layer streams generated by the encoder of FIG. 5; and

FIG. 8 shows an exemplary embodiment of a system which may be used for implementing the principles of the present invention.

FIG. 3A shows an enhancement layer video coding scheme 30 according to a first exemplary embodiment of the present invention. As can be seen, the video coding scheme 30 includes a prediction-based base layer 31 and a two-loop prediction-based enhancement layer 32.

The prediction-based base layer 31 includes intraframe coded I frames, interframe coded predicted P-frames, and interframe coded bi-directional predicted B-frames, as in the conventional enhancement layer video scheme presented in FIG. 1. The base layer I-, P- and B-frames may be coded using conventional non-scalable frame-prediction coding techniques. (The base layer I-frames are of course not motion-predicted.).

The two-loop prediction-based enhancement layer 32 includes non-motion-predicted enhancement layer I- and P-frames and motion-predicted enhancement layer B-frames. The non-motion-predicted enhancement layer I- and P-frames are derived conventionally by subtracting their respective reconstructed (decoded) base layer I- and P-frame residuals from their respective original base layer I- and P-frame residuals.

In accordance with the present invention, the motion-predicted enhancement layer B-frames are each computed using: 1) motion-prediction from two temporally adjacent

differential I- and P- or P- and P- frame residuals (a.k.a. enhancement layer frames), and 2) the differential B-frame residual obtained by subtracting the decoded base layer B-frame residual from the original base layer B-frame residual. The difference between 2) the differential B-frame residual and 1) the B-frame motion prediction obtained from the two temporally adjacent motion-compensated differential frame residuals provide a motion-predicted enhancement layer B-frame in the Enhancement Layer 32. Both the motion-predicted enhancement layer B frames resulting from this process and the non-motion-predicted enhancement layer I- and P- frames may be coded with any suitable scalable codec, preferably a fine granular scalable (FGS) codec as shown in FIG. 3A.

The video coding scheme 30 of the present invention improves the video image quality because it reduces temporal redundancy in the enhancement layer B-frames of the enhancement layer 32. Since the enhancement layer B-frames account for 66% of the total bit-rate budget for the enhancement layer 32 in an IBBP group of pictures (GOP) structure, the loss in image quality associated with performing motion compensation only for the enhancement layer B-frames is very limited for most video sequences. (In conventional enhancement layer video coding schemes, a popular rate-control is mostly performed within the enhancement layer by allocating an equal number of bits to all enhancement layer I-, P-, and B-frames.)

Further, it is important to note that rate-control plays an important role for achieving good performance with the video coding scheme of the present invention. However, even a simplistic approach which allocates the total bit-budget B_{tot} for a GOP according to $B_{tot} = b_I * \text{No.}_I_frames + b_P * \text{No.}_P_frames + b_B * \text{No.}_B_frames$, where $b_I > b_P > b_B$, already provides very good results. Further note that a different number of enhancement layer bits/bitplanes (does not have to be an integer number of bits/bitplanes) can be considered for each enhancement layer reference frame used in the motion compensation loops. Moreover, if desired, only certain parts or frequencies within the enhancement layer reference frame need be incorporated in the enhancement layer motion-compensation loop.

The packet-loss robustness of the above scheme is similar to that of the current enhancement layer coding scheme of FIG. 1: if an error occurs in a motion-predicted enhancement layer B-frame, this error will not propagate beyond the next received I- or P-frame. Two packet-loss scenarios can occur:

- If an error occurs in the motion-predicted enhancement layer B-frame, the error is confined to that B-frame;

- If an error occurs in an enhancement layer I- or P-frame, the error will not go beyond the (two) motion-predicted enhancement layer B-frames using these enhancement layer frames as references. Then, either one of the motion-predicted enhancement layer B-frames can be discarded and frame-repetition applied or error concealment can be applied using the other uncorrupted reference enhancement layer frame.

FIG. 4 shows a block-diagram of an encoder 40, according to an exemplary embodiment of the present invention, that may be used for generating the enhancement layer video coding scheme of FIG. 3A. As can be seen, the encoder 40 includes a base layer encoder 41 and an enhancement layer encoder 42. The base layer encoder 41 is conventional and includes a motion estimator 43 that generates motion information (motion vectors and prediction modes) from the original video sequence and appropriate reference frame stored in memory 44. A first motion compensator 45 in a first motion compensation loop 62, processes the motion information and generates motion-compensated base layer reference frames (Ref(i)). A first subtractor 46 subtracts the motion-compensated base layer reference frames Ref(i) from the original video sequence to generate motion-compensated residuals of the base layer frames MCR(i). The motion-compensated residuals of the base layer frames MCR(i) are processed by a discrete cosine transform (DCT) encoder 47, a quantizer 48, and an entropy encoder 49 into a portion of a compressed base layer stream (base layer frames) from the original video sequence. The motion information generated by the motion estimator 43 is also combined, via a multiplexer 50, with the portion of the base layer stream processed by the first subtractor 46, DCT encoder 47, quantizer 48 and entropy encoder 49. The quantized motion-compensated residual of the base layer frames MCR(i) generated at the output of the quantizer 48 are dequantized by an inverse quantizer 51, and then inverse DCT transformed via an inverse DCT unit 52. This process generates quantized-and-dequantized versions of the motion-compensated residuals of the base layer frames MCRQ(i), at the output of the inverse DCT 52. The quantized-and-dequantized motion-compensated residuals of the base layer frames MCRQ(i) and their respective motion-compensated base layer reference frames Ref(i) are summed in an adder 53 to generate new reference frames that are stored in the first frame memory 44 and used by the motion estimator 43 and motion compensator 45 for processing other frames.

Still referring to FIG. 4, the enhancement layer encoder 42, which preferably comprises an FGS enhancement layer encoder (as shown in FIG. 4), includes a second subtractor 54 that computes the difference between the motion-compensated residuals of the base layer frames MCR(i) and the quantized-and-dequantized motion-compensated residuals

of the base layer frames $MCRQ(i)$ to generate differential I-, P-, and B-frame residuals $FGSR(i)$, which in the case of the I- and P-frame residuals, are the enhancement layer I- and P- frames. A frame flow control device 55 is provided for enabling the differential I- and P-frame residuals to be processed conventionally while the differential B-frame residuals are processed with motion-compensation in the enhancement layer in accordance with the principles of the present invention. The frame flow control device 55 accomplishes this task by causing the data flow at the output of the second subtractor 54 to stream in a different manner in accordance with the type of frame that is outputted by the second subtractor 54. More specifically, differential I- and P-frame residuals generated at the output of the second subtractor 54 are routed by the frame control device 55 to an FGS encoder 61 (or like scalable encoder) for FGS coding using conventional DCT encoding followed by bit-plane DCT scanning and entropy encoding to generate a portion (non-motion-predicted enhancement layer I- and P-frames) of a compressed enhancement layer stream. The differential I- and P-frame residuals generated at the output of the second subtractor 54 are also routed to a second frame memory 58 where they are used later on for motion-compensation. The differential B-frame residuals generated at the output of the second subtractor 54 are routed by the frame control device 55 to a third subtractor 60 and the second frame memory 58. A second motion compensator 59 in second motion compensation loop 63, reuses the motion information from the original video sequence (the output of the motion estimator 43 of the base layer encoder 41) and the differential I- and P-frame residuals stored in the second frame memory 58, which are used as references, to generate reference motion-compensated differential (I- and P- or P- and P-) frame residuals $MCFGSR(i)$. Note that only a portion of each reference differential I- and P-frame residual e.g. several bit-planes, is required, although the entire reference differential frame residual can be used if desired. The third subtractor 60 generates each motion-predicted enhancement layer B-frame $MCFG(i)$ by subtracting the reference motion-compensated differential (I- and P- or P- and P-) frame residual $MCFGSR(i)$ from its respective differential B-frame residual $FGSR(i)$. The frame flow control device 55 routes the motion-predicted enhancement layer B-frames $MCFG(i)$ to the FGS encoder 61 for FGS coding using conventional DCT encoding followed by bit-plane DCT scanning and entropy encoding where they are added to the compressed enhancement layer stream.

As should now be apparent, the base layer remains unchanged in the enhancement layer video coding scheme of FIG. 3A. Moreover, the enhancement layer I- and P-frames are processed in substantially the same manner as in the current FGS video coding

scheme of FIG. 1, therefore, these frames are not motion-predicted within the enhancement layer. In the case of the motion-predicted enhancement layer B-frames, it should be apparent now that the signal to be coded in the enhancement layer of the i^{th} frame MCFGSR equals:

$$\text{MCFGSR}(i) = \text{FGSR}(i) - \text{MCFGSR}(i) = \text{MCR}(i) - \text{MCRQ}(i) - \text{MCFGSR}(i)$$

5 where $\text{MCR}(i)$ is the motion-compensated residual of frame i after the quantization and the dequantization processes, $\text{FGSR}(i)$ is substantially identical to the current FGS video coding scheme of FIG. 1, i.e., $\text{FGSR}(i)$ equals $\text{MCR}(i) - \text{MCRQ}(i)$, and $\text{MCFGSR}(i)$ is the reference motion-compensated differential frame residual for frame (i) . It should be noted that enhancement layer B-frame processing method of the present invention merely requires an
10 additional motion-compensation loop in the enhancement layer for providing motion-predicted enhancement layer B-frames.

FIG. 6 shows a block-diagram of a decoder 70, according to an exemplary embodiment of the present invention, that may be used for decoding the compressed base layer and enhancement layer streams generated by the encoder 40 of FIG. 4. As can be seen,
15 the decoder 70 includes a base layer decoder 71 and an enhancement layer decoder 72. The base layer decoder 71 includes a demultiplexer 75 which receives the encoded base layer stream and demultiplexes the stream into first and second data streams 76 and 77. The first data stream 76, which includes motion information (motion vectors and motion prediction modes), is applied to a first motion compensator 78. The motion compensator 78 uses the
20 motion information and base layer reference video frames stored in an associated base layer frame memory 79 to generate motion-predicted base layer P- and B-frames that are applied to a first input 81 of a first adder 80. The second data stream 77 is applied to a base layer variable length code decoder 83 for decoding, and to an inverse quantizer 84 for dequantizing. The dequantized code is applied to an inverse DCT decoder 85 where the
25 dequantized code is transformed into base layer residual video I-, P- and B-frames which are applied to a second input 82 of the first adder 80. The base layer residual video frames and motion-predicted base layer frames generated by the motion compensator 78 are summed in the first adder 80 to generate base layer video I-, P-, and B-frames that are stored in the base layer frame memory 79 and optionally outputted as a base layer video.

30 The enhancement layer decoder 72 includes an FGS bit-plane decoder 86 or like scalable decoder that decodes the compressed enhancement layer stream to generate at first and second outputs 73 and 74 the differential I-, P-, and B-frame residuals which are respectively applied to first and second frame flow control devices 87 and 91. The first and second frame flow control devices 87 and 91 enable the differential I- and P-frame residuals

to be processed differently from the differential B-frame residuals by causing the data flow at the outputs 73 and 74 of the FGS bit-plane decoder 86 to stream in a different manner in accordance with the type of enhancement layer frame that is outputted by the decoder 86. The differential I- and P-frame residuals at the first output 73 of the FGS bit-plane decoder 86 are
5 routed by the first frame control device 87 to an enhancement layer frame memory 88 where they are stored and used later on for motion compensation. The differential B-frame residuals at the first output 73 of the FGS bit-plane decoder 86 are routed by the first frame control device 87 to a second adder 92 and processed as will be explained further on.

A second motion compensator 90 reuses the motion information received by
10 the base layer decoder 71 and the differential I- and P-frame residuals stored in the enhancement layer frame memory 88 to generate reference motion-compensated differential (I- and P- or P- and P-) frame residuals, which are used for predicting enhancement layer B-frames. The second adder 92 sums each reference motion-compensated differential frame residual and its respective differential B-frame residual to generate an enhancement layer B-
15 frame.

The second frame control device 91 sequentially routes the enhancement layer I- and P-frames (the differential I- and P-frame residuals) at the second output 74 of the FGS bit-plane decoder 86 and the motion-predicted enhancement layer B-frames at the output 93 of the second adder 92 to a third adder 89. The third adder 89 sums the enhancement layer I-,
20 P-, and B-frames together with their corresponding base layer I-, P-, and B-frames to generate an enhanced video.

FIG. 3B shows an enhancement layer video coding scheme 100 according to a second exemplary embodiment of the present invention. As can be seen, the video coding scheme 100 of the second embodiment is substantially identical to the first embodiment of
25 FIG. 3A except that the enhancement layer P-frames in the two-loop prediction-based enhancement layer 132 are motion-predicted like the enhancement layer B-frames.

The motion-predicted enhancement layer P-frames are computed in a manner similar to the enhancement B-frames i.e., each motion-predicted enhancement layer P-frame is computed using: 1) motion-prediction from a temporally adjacent differential I- or P-frame residual, and 2) the differential P-frame residual obtained by subtracting the decoded base
30 layer P-frame residual from the original base layer P-frame residual. The difference between 2) the differential P-frame residual and 1) the P-frame motion prediction obtained from the temporally adjacent motion-compensated differential frame residual provide a motion-predicted enhancement layer P-frame in the Enhancement Layer 132. Both the motion-

predicted enhancement layer P- and B-frames resulting from this process and the non-motion-predicted enhancement layer I-frames may be coded with any suitable scalable codec, preferably a fine granular scalable (FGS) codec as shown in FIG. 3B.

5 The video coding scheme 100 of FIG. 3B provides further improvements in the video image quality. This is because the video coding scheme 100 reduces temporal redundancy in both the P- and B-frames of the enhancement layer 132.

The video coding schemes of the present invention can be alternated with the current video coding scheme of FIG. 1 for the various portions of a video sequence or for various video sequences. Additionally, switching between all three video coding schemes i.e.,
10 current video coding scheme of FIG. 1 and the video coding schemes described in FIGS. 3A and 3B, can be done based on channel characteristics and can be performed at encoding or at transmission time. Further the video coding schemes of the present invention achieve a large gain in coding efficiency with only a limited increase in complexity.

FIG. 5 shows a block-diagram of an encoder 140, according to an exemplary
15 embodiment of the present invention, that may be used for generating the enhancement layer video coding scheme of FIG. 3B. As can be seen, the encoder 140 of FIG. 5 is substantially identical to the encoder 40 of FIG. 4 (which is used for generating the enhancement layer video coding scheme of FIG. 3A), except that the frame flow control device 55 used in the encoder 40 is omitted. The frame flow control device is not necessary in this encoder 140
20 because the differential I-frame residuals are not processed with motion-compensation and thus, do not need to be routed differently from the differential P- and B-frame residuals in the enhancement layer encoder 142.

Hence, the differential I-frame residuals generated at the output of the second subtractor 54 pass to an FGS encoder 61 for FGS coding using conventional DCT encoding
25 followed by bit-plane DCT scanning and entropy encoding to generate a portion (non-motion-predicted enhancement layer I- frames) of a compressed enhancement layer stream. The differential I- frame residuals also pass to a second frame memory 58 along with the differential P-frame residuals where they are used later on for motion-compensation. The differential P- and B-frame residuals generated at the output of the second subtractor 54 are
30 also passed to a third subtractor 60. A second motion compensator 59 in second motion compensation loop 63, reuses the motion information from the original video sequence (the output of the motion estimator 43 of the base layer encoder 41) and the differential I- and P-frame residuals stored in the second frame memory 58, which are used as references, to generate reference motion-compensated differential (I or P) frame residuals MCFGSR(i) for

motion-predicting enhancement layer P-frames and reference (I- and P- or P- and P-) frame residuals MCFGSR(i) for motion-predicting enhancement layer B-frames. The third subtractor 60 generates each motion-predicted enhancement layer P- or B-frame MCFGS(i) by subtracting the reference motion-compensated differential (I or P) or (I- and P- or P- and P-) frame residual MCFGSR(i) from its respective differential P- or B-frame residual FGSR(i). The motion-predicted enhancement layer P- and B-frames MCFGS(i) then pass to the FGS encoder 61 for FGS coding using conventional DCT encoding followed by bit-plane DCT scanning and entropy encoding where they are added to the compressed enhancement layer stream.

As in the video coding scheme of FIG. 3A, the base layer remains unchanged in the enhancement layer video coding scheme of FIG. 3B. Moreover, it should be noted that enhancement layer P- and B-frame processing method of the present invention merely requires an additional motion-compensation loop in the enhancement layer for providing motion-predicted enhancement layer P-and B-frames.

FIG. 7 shows a block-diagram of a decoder 170, according to an exemplary embodiment of the present invention, that may be used for decoding the compressed base layer and enhancement layer streams generated by the encoder 140 of FIG. 5. As can be seen, the decoder 170 of FIG. 7 is substantially identical to the decoder 70 of FIG. 6, except that the frame flow control devices 87 and 91 used in the decoder 70 are omitted. The frame flow control devices are not necessary in this decoder 170 because the differential I-frame residuals are not processed with motion-compensation and thus, do not need to be routed differently from the decoded differential P- and B-frame residuals in the enhancement layer decoder 172.

Accordingly, the differential I- and P-frame residuals at the first output 73 of the FGS bit-plane decoder 86 pass to the enhancement layer frame memory 88 where they are stored and used later on for motion compensation. The differential P- and B-frame residuals at the second output 74 of the FGS bit-plane decoder 86 pass to a second adder 92. The differential I-frame residuals (enhancement layer I-frames hereinafter) at the second output 74 of the FGS bit-plane decoder 86 pass to a third adder 89, the purpose of which will be explained further on. The second motion compensator 90 reuses the motion information received by the base layer decoder 71 and the differential I- and P-frame residuals stored in the enhancement layer frame memory 88 to generate 1) reference motion-compensated differential (I- and P- or P- and P-) frame residuals, which are used for predicting enhancement layer B-frames, and 2) reference motion-compensated differential (I-or P-)

frame residuals, which are used for predicting enhancement layer P-frames. The second adder 92 sums the reference motion-compensated differential frame residuals with their respective differential B-frame residuals or P-frame residuals to generate enhancement layer B- and P-frames. The third adder 89 sums the enhancement layer I-, P-, and B-frames together with
5 their corresponding base layer I-, P-, and B-frames to generate an enhanced video.

FIG. 8 shows an exemplary embodiment of a system 200 which may be used for implementing the principles of the present invention. The system 200 may represent a television, a set-top box, a desktop, laptop or palmtop computer, a personal digital assistant (PDA), a video/image storage device such as a video cassette recorder (VCR), a digital video
10 recorder (DVR), a TiVO device, etc., as well as portions or combinations of these and other devices. The system 200 includes one or more video/image sources 201, one or more input/output devices 202, a processor 203 and a memory 204. The video/image source(s) 201 may represent, e.g., a television receiver, a VCR or other video/image storage device. The source(s) 201 may alternatively represent one or more network connections for receiving
15 video from a server or servers over, e.g., a global computer communications network such as the Internet, a wide area network, a metropolitan area network, a local area network, a terrestrial broadcast system, a cable network, a satellite network, a wireless network, or a telephone network, as well as portions or combinations of these and other types of networks.

The input/output devices 202, processor 203 and memory 204 may
20 communicate over a communication medium 205. The communication medium 205 may represent, e.g., a bus, a communication network, one or more internal connections of a circuit, circuit card or other device, as well as portions and combinations of these and other communication media. Input video data from the source(s) 201 is processed in accordance with one or more software programs stored in memory 204 and executed by processor 203 in
25 order to generate output video/images supplied to a display device 206.

In a preferred embodiment, the coding and decoding employing the principles of the present invention may be implemented by computer readable code executed by the system. The code may be stored in the memory 204 or read/downloaded from a memory medium such as a CD-ROM or floppy disk. In other embodiments, hardware circuitry may be
30 used in place of, or in combination with, software instructions to implement the invention. For example, the elements shown in FIGS. 4-7 may also be implemented as discrete hardware elements.

While the present invention has been described above in terms of specific embodiments, it is to be understood that the invention is not intended to be confined or

limited to the embodiments disclosed herein. For example, other transforms besides DCT can be employed, including but not limited to wavelets or matching-pursuits. In another example, although motion-compensation is accomplished in the above embodiments by reusing motion data from the base layer, other embodiments of the invention can employ an additional
5 motion estimator in the enhancement layer, which would require sending additional motion vectors. In still another example, other embodiments of the invention may employ motion compensation in the enhancement layer for just the P-frames. These and all other such modifications and changes are considered to be within the scope of the appended claims.

CLAIMS:

1. A method of coding video, comprising the steps of:
coding (41, 42) an uncoded video with a non-scalable codec to generate base layer frames;
computing (54) differential frame residuals from the uncoded video and the
5 base layer frames, at least portions of certain ones of the differential frame residuals being operative as references;
applying motion-compensation (59) to the at least portions of the differential frame residuals that are operative as references to generate reference motion-compensated differential frame residuals; and
10 subtracting (60) the reference motion-compensated differential frame residuals from respective ones of the differential frame residuals to generate motion-predicted enhancement layer frames.
2. A method of coding video according to claim 1, further comprising the step of
15 coding (61) the motion-predicted enhancement layer frames with a scalable codec.
3. A method of coding video according to claim 1, further comprising the step of
coding (61) the motion-predicted enhancement layer frames with a fine granular scalable
20 codec.
4. A method of coding video according to claim 1, wherein the motion-predicted enhancement layer frames in the subtracting step (60) include motion-predicted enhancement layer B-frames, the reference motion-compensated differential frame residuals in the subtracting step include reference motion-compensated differential I- and P-frame residuals
25 or reference motion-compensated differential P- and P-frame residuals, and the respective ones of the differential frame residuals in the subtracting step include differential B-frames.
5. A method of coding video according to claim 4, wherein the motion-predicted enhancement layer frames in the subtracting step (60) further include motion-predicted

enhancement layer P-frames, the reference motion-compensated differential frame residuals in the subtracting step (60) further include reference motion-compensated differential I-frame residuals or reference motion-compensated P-frame residuals, and the respective ones of the differential frame residuals in the subtracting step (60) further include differential P-frames.

5

6. A method of coding video according to claim 1, wherein the motion-predicted enhancement layer frames in the subtracting step (60) include motion-predicted enhancement layer P-frames, the reference motion-compensated differential frame residuals in the subtracting step (60) include reference motion-compensated differential I-frame residuals or reference motion-compensated P-frame residuals, and the respective ones of the differential frame residuals in the subtracting step (60) include differential P-frames.

10

7. A method of decoding a compressed video having a base layer stream and an enhancement layer stream, the method comprising the steps of:

15

decoding (71, 171) the base layer stream to generate base layer video frames;
decoding (72, 172) the enhancement layer stream to generate differential frame residuals, at least portions of certain ones of the differential frame residuals being operative as references;

20

applying motion-compensation (90) to the at least portions of the differential frame residuals operative as references to generate reference motion-compensated differential frame residuals;

adding (92) the reference motion-compensated differential frame residuals with respective ones of the differential frame residuals to generate motion-predicted enhancement layer frames; and

25

combining (89) the motion-predicted enhancement layer frames with respective ones of the base layer frames to generate an enhanced video.

8. A method of decoding video according to claim 7, wherein the motion-predicted enhancement layer frames in the adding step (92) include motion-predicted enhancement layer B-frames, the reference motion-compensated differential frame residuals in the adding step (92) include reference motion-compensated differential I- and P-frame residuals or reference motion-compensated differential P- and P-frame residuals, and the respective ones of the differential frame residuals in the adding step (92) include differential B-frames.

30

9. A method of decoding video according to claim 8, wherein the motion-predicted enhancement layer frames in the adding step (92) further include motion-predicted enhancement layer P-frames, the reference motion-compensated differential frame residuals
5 in the adding step (92) further include reference motion-compensated differential I-frame residuals or reference motion-compensated P-frame residuals, and the respective ones of the differential frame residuals in the adding step (92) further include differential P-frames.

10. A method of decoding video according to claim 7, wherein the motion-predicted enhancement layer frames in the adding step (92) include motion-predicted enhancement layer P-frames, the reference motion-compensated differential frame residuals
10 in the adding step include reference motion-compensated differential I-frame residuals or reference motion-compensated P-frame residuals, and the respective ones of the differential frame residuals in the adding step include differential P-frames.

15

11. A memory medium for encoding video, the memory medium comprising:
code (41, 141) for non-scalable encoding an uncoded video into base layer frames;

code (54) for computing differential frame residuals from the uncoded video
20 and the base layer frames, at least portions of certain ones of the differential frame residuals being operative as references;

code for applying motion-compensation (59) to the at least portions of the differential frame residuals that are operative as references to generate reference motion-compensated differential frame residuals; and

25 code (60) for subtracting the reference motion-compensated differential frame residuals from respective ones of the differential frame residuals to generate motion-predicted enhancement layer frames.

12. A memory medium for encoding video according to claim 11, further
30 comprising code (61) for scalable encoding the motion-predicted enhancement layer frames.

13. A memory medium for encoding video according to claim 11, further comprising code (61) for fine granular scalable encoding the motion-predicted enhancement layer frames.

14. A memory medium for encoding video according to claim 11, wherein the motion-predicted enhancement layer frames include motion-predicted enhancement layer B-frames, the reference motion-compensated differential frame residuals include reference
5 motion-compensated differential I- and P-frame residuals or reference motion-compensated differential P- and P-frame residuals, and the respective ones of the differential frame residuals include differential B-frames.

15. A memory medium for encoding video according to claim 14, wherein the
10 motion-predicted enhancement layer frames further include motion-predicted enhancement layer P-frames, the reference motion-compensated differential frame residuals further include reference motion-compensated differential I-frame residuals or reference motion-compensated P-frame residuals, and the respective ones of the differential frame residuals further include differential P-frames.

16. A memory medium for encoding video according to claim 14, wherein the
15 motion-predicted enhancement layer frames include motion-predicted enhancement layer P-frames, the reference motion-compensated differential frame residuals include reference motion-compensated differential I-frame residuals or reference motion-compensated P-frame
20 residuals, and the respective ones of the differential frame residuals include differential P-frames.

17. A memory medium for decoding a compressed video having a base layer stream and an enhancement layer stream, the memory medium comprising:

25 code (71, 171) for decoding the base layer stream to generate base layer video frames;

code (72, 172) for decoding the enhancement layer stream to generate differential frame residuals, at least portions of certain ones of the differential frame residuals being operative as references;

30 code (90) for applying motion-compensation to the at least portions of the differential frame residuals operative as references to generate reference motion-compensated differential frame residuals;

code (92) for adding the reference motion-compensated differential frame residuals with respective ones of the differential frame residuals to generate motion-predicted enhancement layer frames; and

code (89) for combining the motion-predicted enhancement layer frames with
5 respective ones of the base layer frames to generate an enhanced video.

18. A memory medium for decoding a compressed video according to claim 17, wherein the motion-predicted enhancement layer frames include motion-predicted enhancement layer B-frames, the reference motion-compensated differential frame residuals
10 include reference motion-compensated differential I- and P-frame residuals or reference motion-compensated differential P- and P-frame residuals, and the respective ones of the differential frame residuals include differential B-frames.

19. A memory medium for decoding a compressed video according to claim 18,
15 wherein the motion-predicted enhancement layer frames further include motion-predicted enhancement layer P-frames, the reference motion-compensated differential frame residuals further include reference motion-compensated differential I-frame residuals or reference motion-compensated P-frame residuals, and the respective ones of the differential frame residuals further include differential P-frames.

20. A memory medium for decoding a compressed video according to claim 17, wherein the motion-predicted enhancement layer frames include motion-predicted enhancement layer P-frames, the reference motion-compensated differential frame residuals
20 include reference motion-compensated differential I-frame residuals or reference motion-compensated P-frame residuals, and the respective ones of the differential frame residuals include differential P-frames.

21. An apparatus for coding video, the apparatus comprising:
means (41, 141) for non-scalable coding an uncoded video to generate base
30 layer frames;
means (42, 142) for computing differential frame residuals from the uncoded video and the base layer frames, at least portions of certain ones of the differential frame residuals being operative as references;

means (59) for applying motion-compensation to the at least portions of the differential frame residuals that are operative as references to generate reference motion-compensated differential frame residuals; and

5 means (60) for subtracting the reference motion-compensated differential frame residuals from respective ones of the differential frame residuals to generate motion-predicted enhancement layer frames.

22. An apparatus for coding video according to claim 21, further comprising means (61) for scalable coding the motion-predicted enhancement layer frames.

10

23. An apparatus for coding video according to claim 21, further comprising means (61) for fine granular scalable coding the motion-predicted enhancement layer frames.

11

24. An apparatus for coding video according to claim 21, wherein the motion-predicted enhancement layer frames include motion-predicted enhancement layer B-frames, the reference motion-compensated differential frame residuals include reference motion-compensated differential I- and P-frame residuals or reference motion-compensated differential P- and P-frame residuals, and the respective ones of the differential frame residuals include differential B-frames.

20

25. An apparatus for coding video according to claim 24, wherein the motion-predicted enhancement layer frames further include motion-predicted enhancement layer P-frames, the reference motion-compensated differential frame residuals further include reference motion-compensated differential I-frame residuals or reference motion-compensated P-frame residuals, and the respective ones of the differential frame residuals further include differential P-frames.

26. An apparatus for coding video according to claim 21, wherein the motion-predicted enhancement layer frames include motion-predicted enhancement layer P-frames, the reference motion-compensated differential frame residuals include reference motion-compensated differential I-frame residuals or reference motion-compensated P-frame residuals, and the respective ones of the differential frame residuals include differential P-frames.

30

27. An apparatus for decoding a compressed video having a base layer stream and an enhancement layer stream, the apparatus comprising:

means (71, 171) for decoding the base layer stream to generate base layer video frames;

5 means (72, 172) for decoding the enhancement layer stream to generate differential frame residuals, at least portions of certain ones of the differential frame residuals being operative as references;

means (90) for applying motion-compensation to the at least portions of the differential frame residuals operative as references to generate reference motion-compensated
10 differential frame residuals;

means (92) for adding the reference motion-compensated differential frame residuals with respective ones of the differential frame residuals to generate motion-predicted enhancement layer frames; and

means (89) for combining the motion-predicted enhancement layer frames
15 with respective ones of the base layer frames to generate an enhanced video.

28. An apparatus for decoding a compressed video according to claim 27, wherein the motion-predicted enhancement layer frames include motion-predicted enhancement layer B-frames, the reference motion-compensated differential frame residuals include reference
20 motion-compensated differential I- and P-frame residuals or reference motion-compensated differential P- and P-frame residuals, and the respective ones of the differential frame residuals include differential B-frames.

29. An apparatus for decoding a compressed video according to claim 28, wherein
25 the motion-predicted enhancement layer frames further include motion-predicted enhancement layer P-frames, the reference motion-compensated differential frame residuals further include reference motion-compensated differential I-frame residuals or reference motion-compensated P-frame residuals, and the respective ones of the differential frame residuals further include differential P-frames.

30. An apparatus for decoding a compressed video according to claim 27, wherein
30 the motion-predicted enhancement layer frames include motion-predicted enhancement layer P-frames, the reference motion-compensated differential frame residuals include reference motion-compensated differential I-frame residuals or reference motion-compensated P-frame

residuals, and the respective ones of the differential frame residuals include differential P-frames.

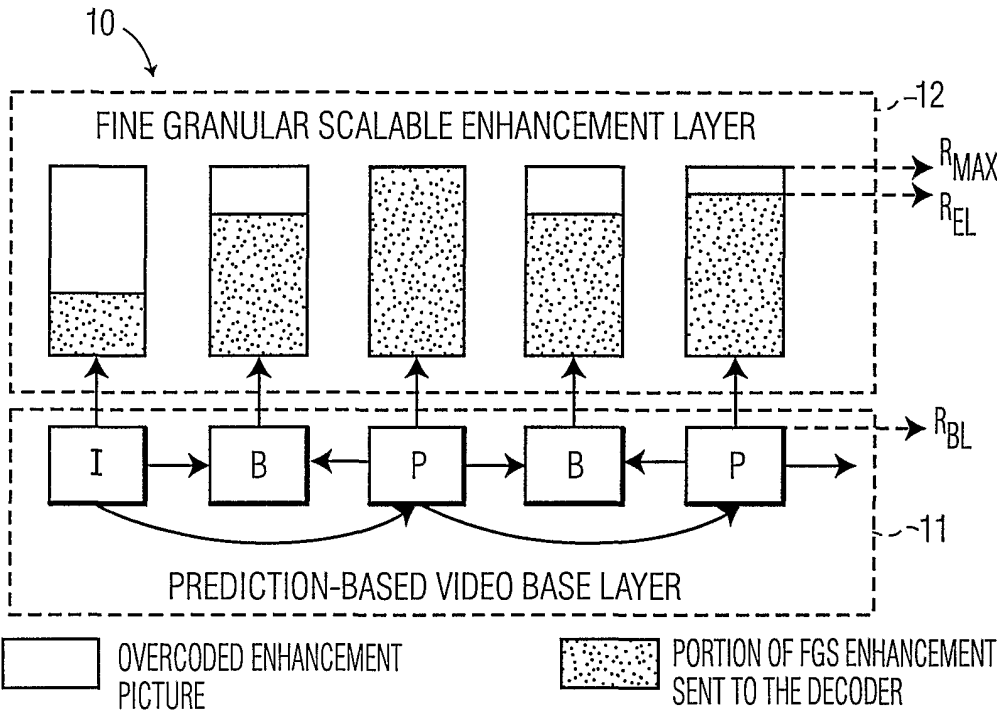


FIG. 1

2/8

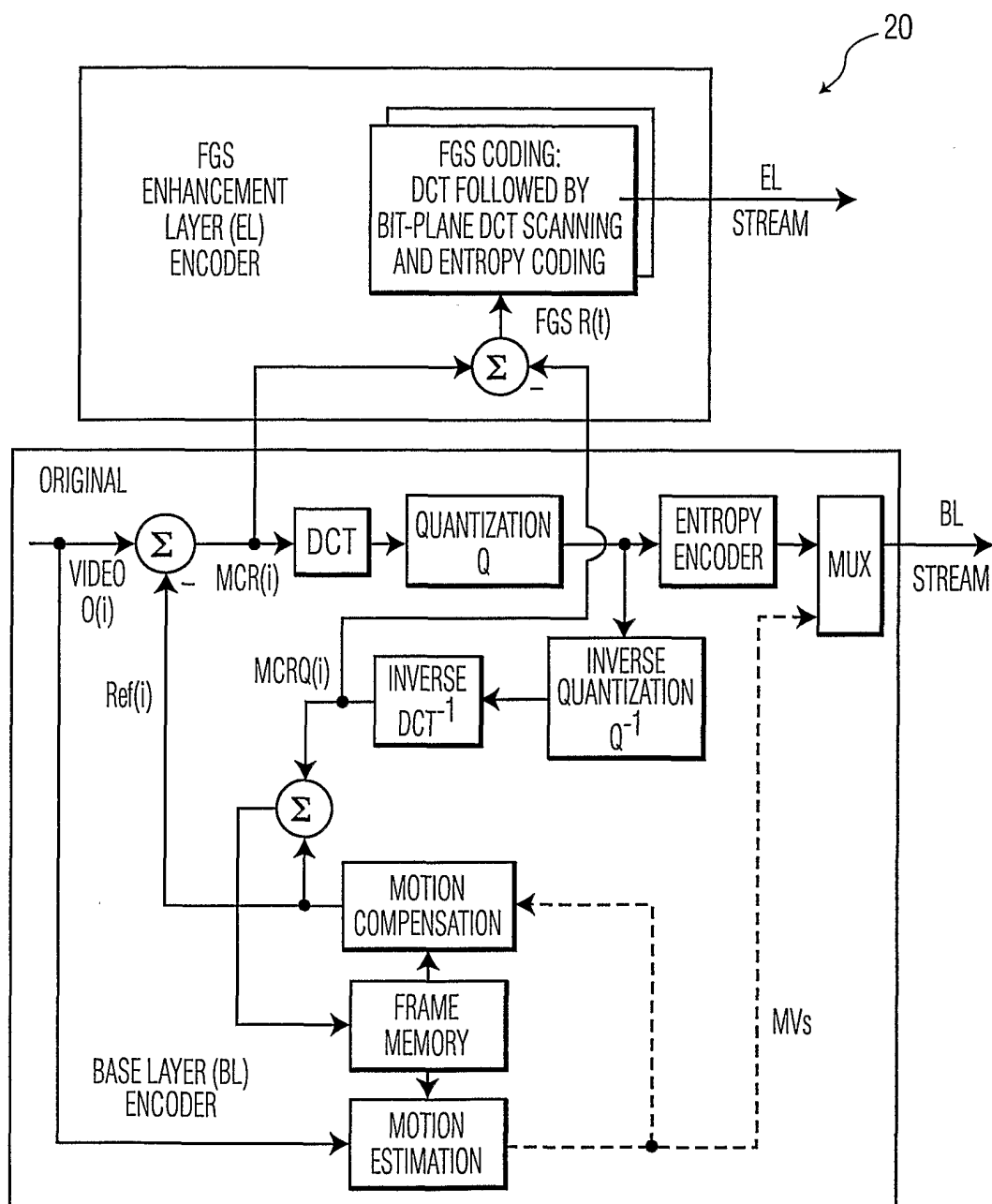


FIG. 2

3/8

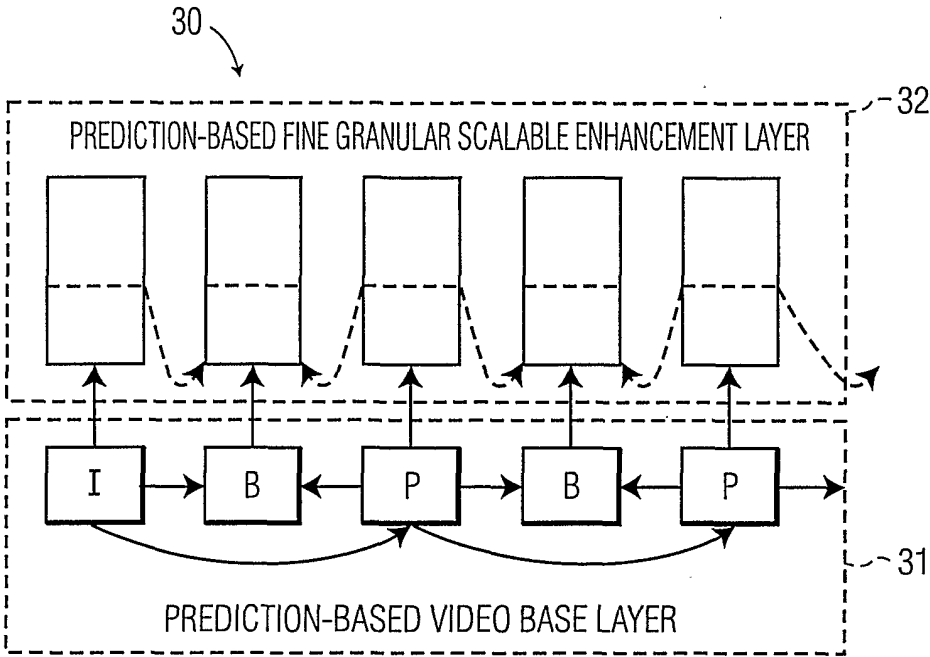


FIG. 3A

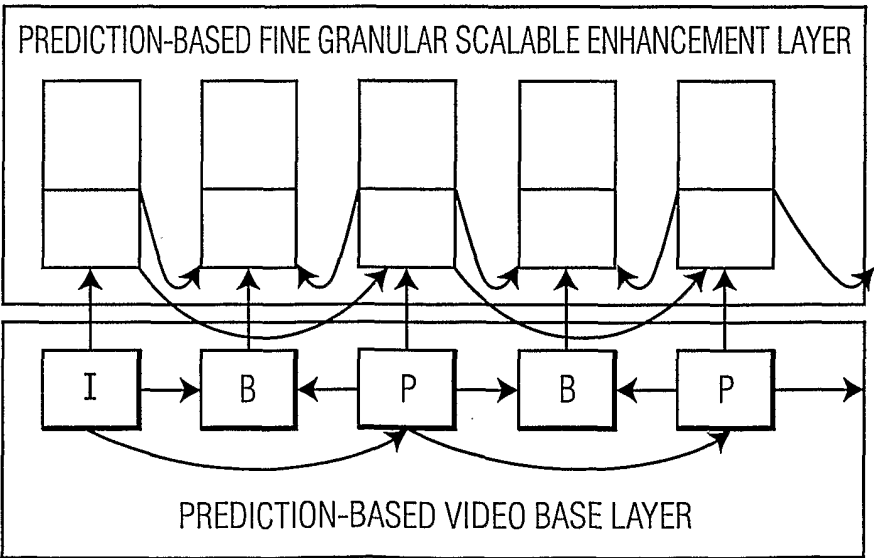


FIG. 3B

4/8

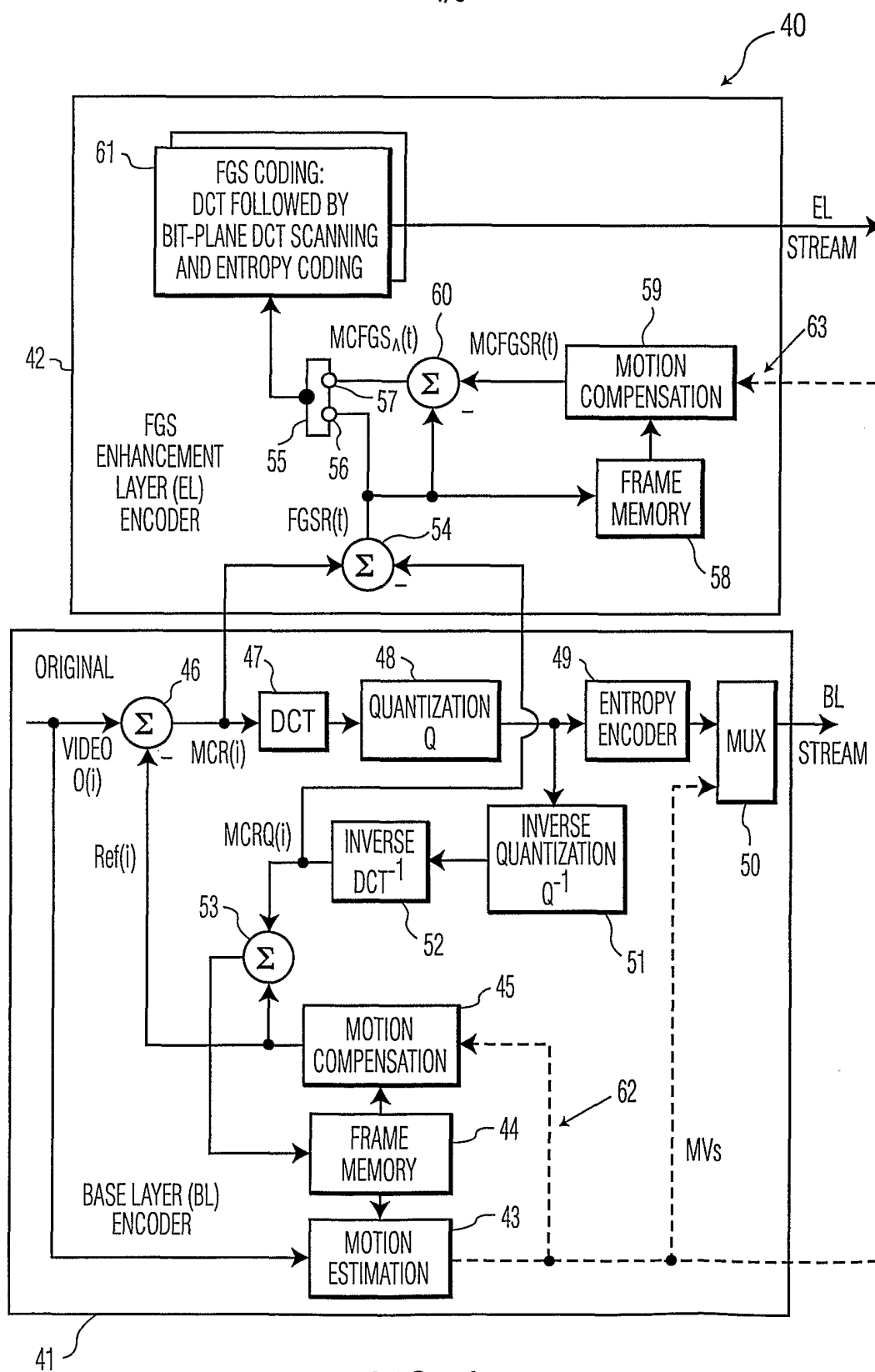


FIG. 4

5/8

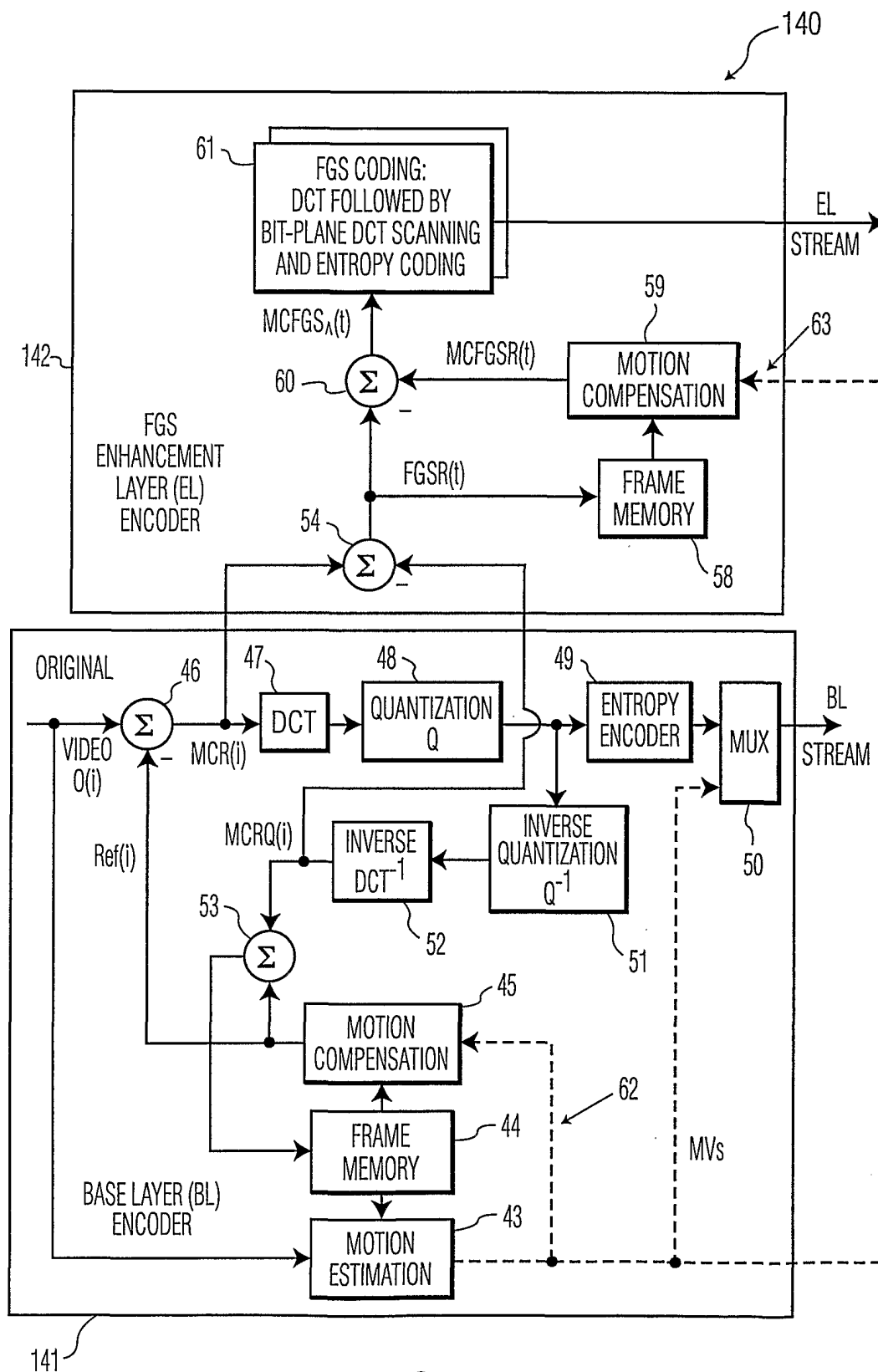


FIG. 5

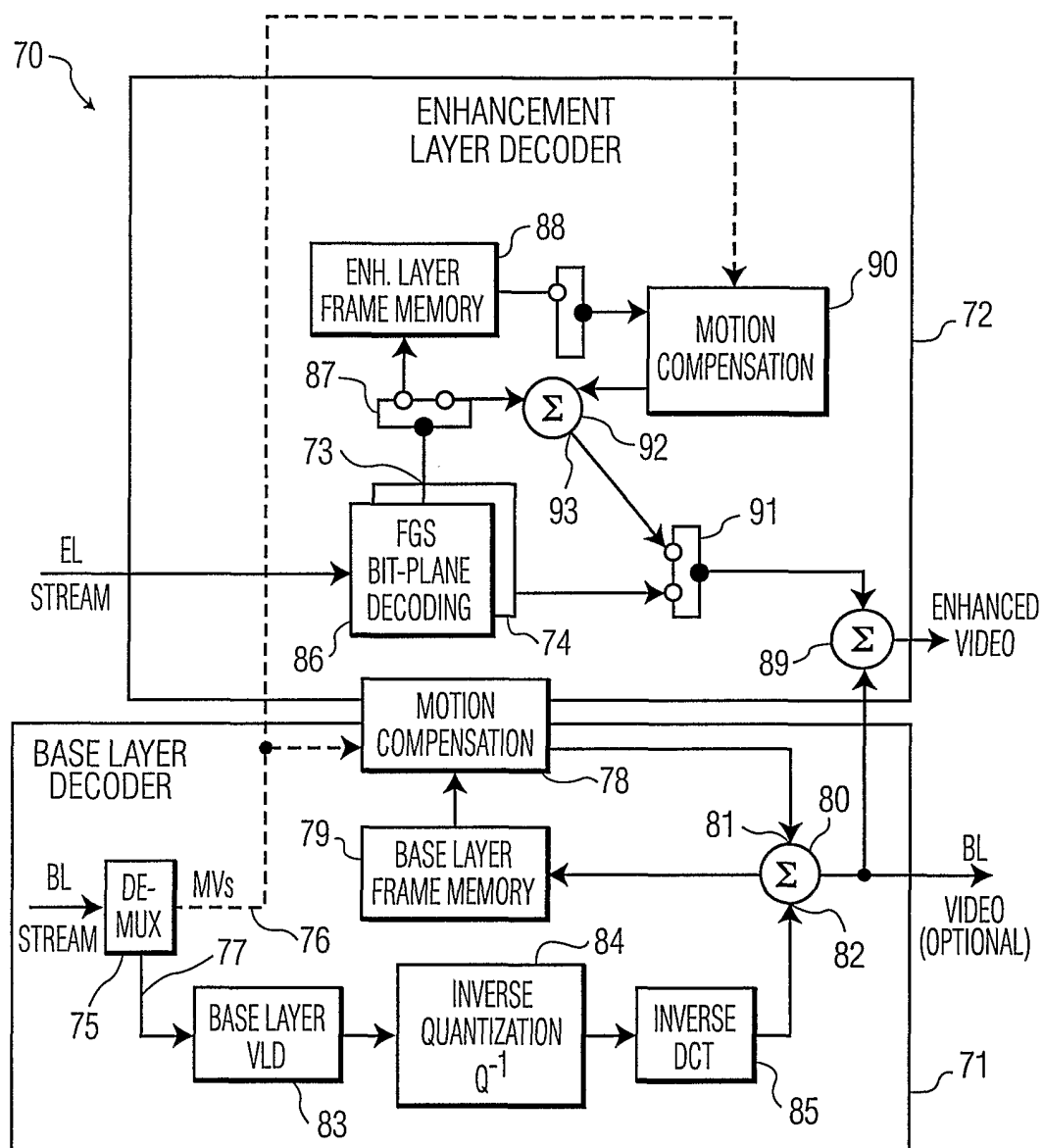


FIG. 6

7/8

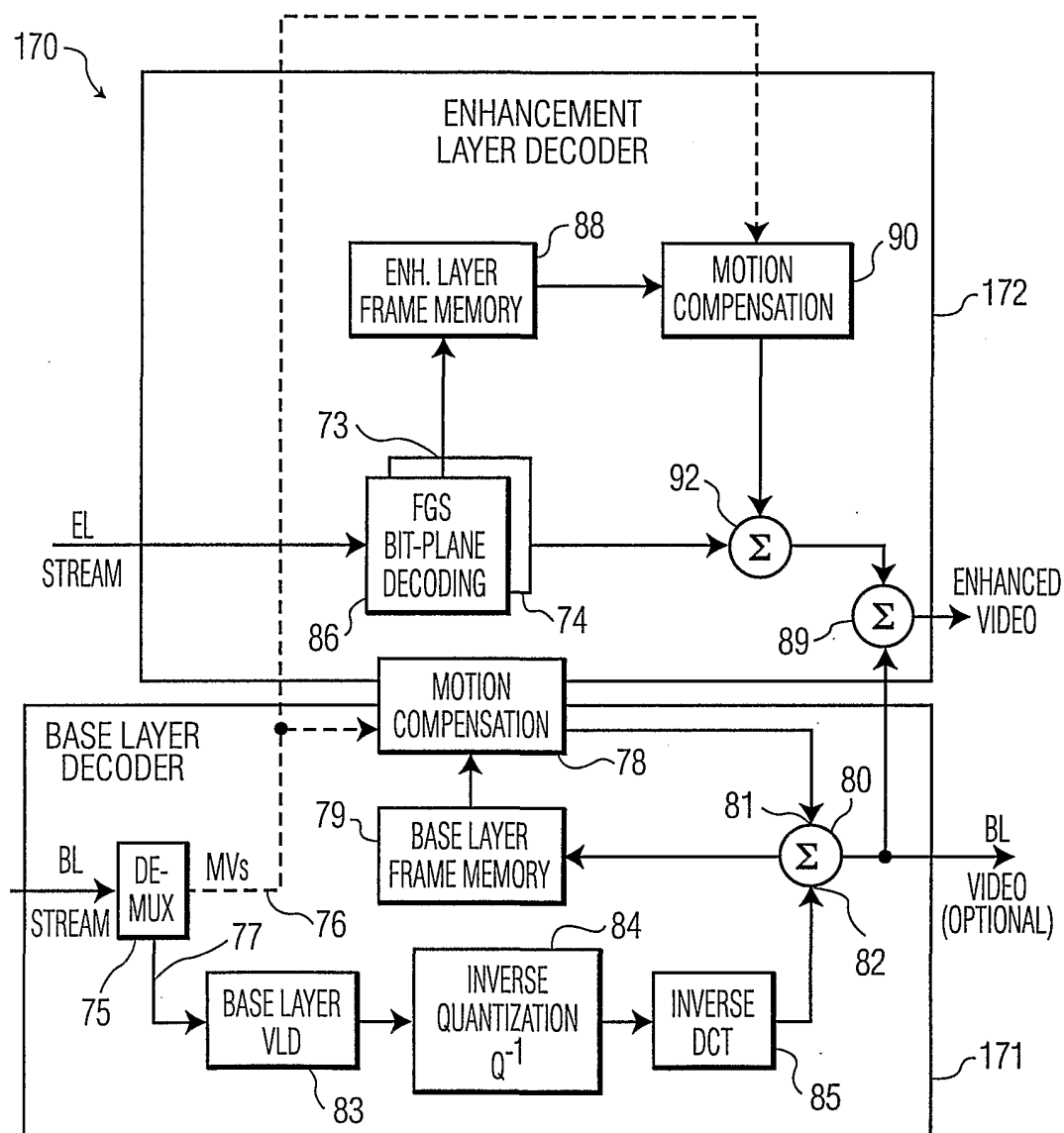


FIG. 7

8/8

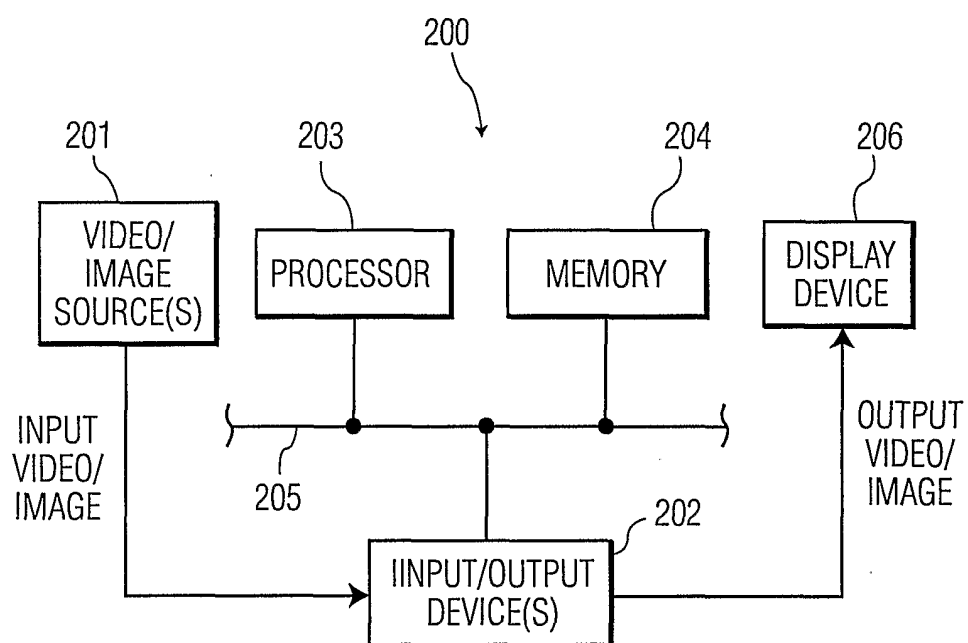


FIG. 8