

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
19 January 2006 (19.01.2006)

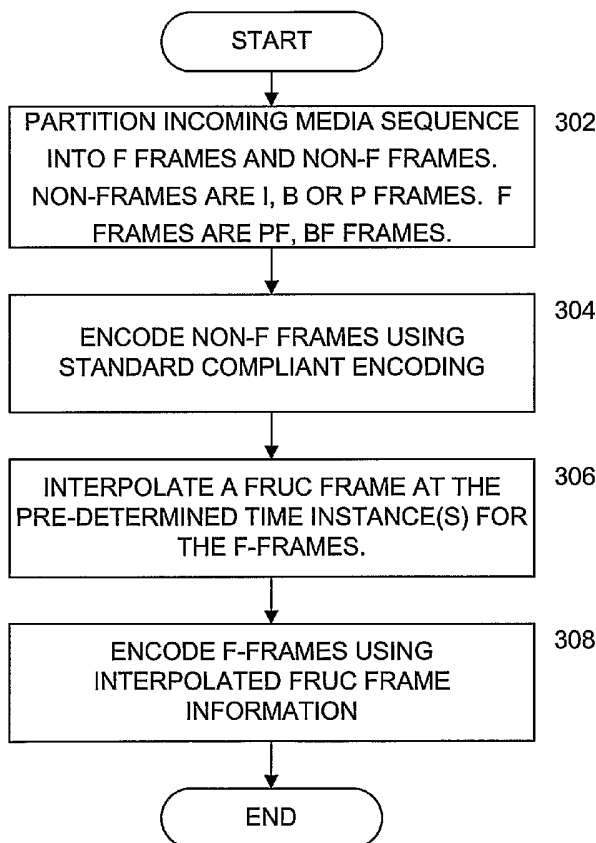
PCT

(10) International Publication Number  
WO 2006/007527 A2

- (51) International Patent Classification:  
H04N 7/01 (2006.01)
- (21) International Application Number:  
PCT/US2005/023333
- (22) International Filing Date: 1 July 2005 (01.07.2005)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/585,154 1 July 2004 (01.07.2004) US  
60/665,816 22 March 2005 (22.03.2005) US
- (71) Applicant (for all designated States except US): QUALCOMM Incorporated [US/US]; 5775 Morehouse Drive, San Diego, California 92121 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): RAVEENDRAN, Vijayalakshmi R. [IN/US]; 4272 Calle Mar De Ballenas, San Diego, California 92130 (US). WALKER, Gordon Kent [US/US]; 14484 Huntington Gate Drive, Poway, California 92064 (US).
- (74) Agents: WADSWORTH, Philip R. et al.; 5775 Morehouse Drive, San Diego, California 92121 (US).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: METHOD AND APPARATUS FOR USING FRAME RATE UP CONVERSION TECHNIQUES IN SCALABLE VIDEO CODING



(57) Abstract: A method for encoding a video stream includes partitioning the video stream into a main layer having a plurality of main layer frames, and an interpolated layer having a plurality of interpolated layer frames; interpolating a frame rate up conversion (FRUC) frame; and encoding the plurality of main layer frames in the interpolated layer with the assistance of the main FRUC frame. A video encoder implementing the method is also described.

WO 2006/007527 A2

**Declarations under Rule 4.17:**

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR,

GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations

**Published:**

- without international search report and to be republished upon receipt of that report

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.

## METHOD AND APPARATUS FOR USING FRAME RATE UP CONVERSION TECHNIQUES IN SCALABLE VIDEO CODING

### Claim of Priority under 35 U.S.C. §119

[0001] The present Application for Patent claims priority to Provisional Application No. 60/585,154 entitled "Method and Apparatus for Using Frame Rate up Conversion (FRUC) Technique in Scalable Video Coding" filed July 1, 2004, and Provisional Application No. 60/665,816 entitled "Method and Apparatus for Using Frame Rate Up Conversion Techniques in Scalable Video Coding" filed March 22, 2005 assigned to the assignee hereof and hereby expressly incorporated by reference herein.

### BACKGROUND

#### Field

[0002] The present invention relates generally to video compression, and more particularly, to a method and apparatus for using Frame Rate Up Conversion (FRUC) techniques in scalable video coding.

[0003]

#### Background

[0004] Rate adaptation in video and audio compression has typically been achieved through scalability (SNR, spatial, temporal) techniques, switching between bit streams coded at various bit-rates and multiple bit rate modes wherein the encoder delivers the content as a variety of media streams at variable bandwidths to the server. The server then delivers the appropriate stream based on the network conditions and/or the target audience.

[0005] Due to the limited bandwidth available in low bit-rate video application, some encoders apply the temporal sampling technique (also referred to as frame skipping) to meet the required compression ratios. Under this scenario, frames in the input sequence are periodically dropped from the encoding process and therefore not transmitted. Thus, instead of the whole input sequence, the decoder receives only partial information in the temporal direction. As a result, temporal artifacts such as motion jerkiness are introduced in the receiving side. Frame rate up conversion (FRUC) is used at the decoder to re-generate the skipped frames in order to reduce the temporal artifacts. Various techniques have been proposed for FRUC. Typically, motion

compensated interpolation (MCI) technique provides the best solution in temporal FRUC applications. However, it is desirable to optimize the implementation of FRUC techniques on the decoder while minimizing the bit-rate of the video stream.

### SUMMARY

[0006] Various approaches for using FRUC techniques in scalable video coding, and specifically, enabling the integration of FRUC compression techniques on the encoder side, also referred to as encoder assisted-FRUC (EA-FRUC), are described herein.

[0007] In one embodiment, a method for encoding a video stream includes the steps of partitioning the video stream into a main layer having a plurality of main layer frames, and an interpolated layer having a plurality of interpolated layer frames; interpolating an FRUC frame; and, encoding the plurality of main layer frames in the main layer with the assistance of the interpolated FRUC frame.

[0008] In another embodiment, EA-FRUC is implemented as a computer-readable medium having instructions stored thereon, the stored instructions, when executed by a processor, cause the processor to perform a method for encoding a video stream. The method includes the steps of partitioning the video stream into a main layer having a plurality of main layer frames, and an interpolated layer having a plurality of interpolated layer frames; interpolating a frame rate up conversion (FRUC) frame; and, encoding the plurality of main layer frames in the main layer with the assistance of the interpolated FRUC frame.

[0009] In still yet another embodiment, the system is implemented as a video encoder processor receiving a video stream. The video encoder processor including a frame rate up conversion (FRUC) module, the FRUC module generating a plurality of FRUC frames from the received video stream; a rate distortion cost (RD\_cost) calculation module coupled to the FRUC module, the RD\_cost calculation module receiving the plurality of FRUC frames and calculating a F frame macroblock RD\_cost based on a macroblock in one of the plurality of FRUC frames; and, a mode decision module coupled to the RD\_cost calculation module, the mode decision module configured to compare the FRUC frame macroblock RD\_cost based on the macroblock in one of the plurality of FRUC frames to a B frame macroblock RD\_cost for a corresponding macroblock in a corresponding B frame.

[0010] Other objects, features and advantages will become apparent to those skilled in the art from the following detailed description. It is to be understood, however, that the detailed description and specific examples, while indicating embodiments of the invention, are given by way of illustration and not limitation. Many changes and modifications within the scope of the description may be made without departing from the spirit thereof, and the invention includes all such modifications.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

[0011] The invention may be more readily understood by referring to the accompanying drawings in which:

[0012] FIG. 1 is a diagram illustrating a conventional two-layered encoding with a base layer and an enhancement layer;

[0013] FIG. 2 is a diagram illustrating an encoding scheme configured in accordance with one embodiment, where a main layer has been encoded with the assistance of FRUC frames;

[0014] FIG. 3 is a flow diagram of a method for encoding the base layer with the assistance of FRUC frames in accordance with one embodiment;

[0015] FIG. 4 is a block diagram of an encoding system configured in accordance with one embodiment;

[0016] FIG. 5 is a flow diagram of a method for mode selection for encoding in accordance with one embodiment;

[0017] FIG. 6 is a diagram illustrating a two-layered encoding scheme configured in accordance with one embodiment, where an enhancement layer has been encoded with the assistance of FRUC frames;

[0018] FIG. 7 is a flow diagram of a method for partitioning media data into two layers and to encode the enhancement layer with the assistance of FRUC frames in accordance with one embodiment; and,

[0019] FIG. 8 is a block diagram of an access terminal and an access point of a wireless system.

[0020] Like numerals refer to like elements throughout the drawings.

**DETAILED DESCRIPTION**

- [0021] Although the FRUC technique is typically used on the decoder side, it can be extended to the encoder side. Specifically, it is useful to utilize the FRUC frame as an additional prediction mode for the motion compensated frame. By adding an extra prediction mode for the inter-predicted frame, bit-rate savings can be achieved. In other words, compression ratios for the same reconstructed visual quality can be improved.
- [0022] For standards-based video encoding, the bit-stream syntax has to be augmented due to the introduction of the extra FRUC mode. However, compliance to standard syntax can be retained by flagging the use of FRUC in predicted frames in the user-data field or Supplemental Enhancement Information (SEI) fields as in the H.264 video coding standard promulgated by the International Telecommunications Union, Telecommunications Standardization Sector (ITU-T). This feature is particularly advantageous in video transmission systems where the FRUC function is performed by the decoder. Further, closed systems, where modifications to the video decoder are possible, stand to benefit significantly in terms of transmission bandwidth and compression efficiency.
- [0023] On the decoder side, upon receiving the transmitted bit-streams and the reconstructed Intra-Picture (I)/Prediction-Picture (P) (I/P) frames, the FRUC frames are interpolated utilizing the same techniques adopted on the encoder side. The inter-frame is then reconstructed based on both the reconstructed I/P frames and the FRUC frame of the same time instance. As long as the same technique is utilized in both the decoder and encoder sides, there will be no error or drift problems.
- [0024] An alternate application of the system would be in scalable video coding. When temporal scalability is employed, some predictive frames, particularly Bi-directional predicted/interpolated picture frames are transmitted in the enhancement layer; and I and P frames are transmitted in the base layer. In such a scenario, if the enhancement layer is not received or requested, the frame rate of the reconstructed video at the decoder is lower than the source frame rate. Sending the differential of a FRUC frame and a B frame in the base layer could enable reconstruction of a close approximation of the B frame through FRUC at the decoder. Results show that the cost of the differential (in bits) is significantly less and hence does not impact the bit rate of base layer video. The differential can be computed as a pure residual of FRUC and the true B frame, through thresholding (or quantization) to limit information to prominent regions (e.g., holes that cannot be reconstructed by FRUC), or as a refinement to known

FRUC methods (e.g., true motion vector data or quantized true motion vector data). Again, such differential information can be carried in "user data" fields of standard syntax or in SEI fields as in the H.264 standard.

[0025] Based on the availability of the decoder resources, different FRUC techniques can be selected based on their degree of computational complexity. A point to note is that FRUC typically expends less computational cycles than decoding a B frame since FRUC involves simple 3 or 5-tap filtering and motion estimation (ME) on a small percentage of the frame, in a relatively small search space; as compared to bi-directional motion compensation (MC) on all macroblocks (sometimes more than one motion compensation (MC)/Macroblock (MB)) of the B frame. FRUC offers granularity in complexity levels based on the level of refinement adopted to reconstruct the intermediate frame at the small cost of visual quality.

[0026] Typically, B frames are computationally more expensive than P frames, and I frames are the least expensive of the three types of frames. In this regard, an added advantage of the system and its extension to scalable video coding is in saving computational cycles and thus valuable power in handheld devices (which translates to longer standby and talk times in such devices). The power savings is realized irrespective of whether a B frame is received or not as, based on available cycles, the decoder can choose to decode the B frame or reconstruct an interpolated frame through FRUC and apply refinements from the differential data.

[0027] Block based FRUC algorithms can be classified based on increasing computational complexity:

[0028] 1. Non-motion compensated algorithms such as frame repetition and frame averaging;

[0029] 2. Bi-linear motion compensated interpolation (MCI) without MV processing;

[0030] 3. MCI with MV processing; and,

[0031] 4. Seeded ME assisted MCI with MV processing.

[0032] Any of the above FRUC assisted encoding technique can be integrated with layering coding techniques. Rate adaptation of the transmission channel can be achieved by layering coding. The base-layer of compressed bit stream usually comprises more important data, and the enhancement layer contains less important data that is subject to a larger transmission error rate.

[0033] In one embodiment, it is preferable that the base-layer bit stream is standards compliant, and the enhancement-layer bit stream is allowed to be dropped when the transmission bandwidth decrease below a certain threshold. Under this scenario, it is desirable to:

[0034] 1. partition the incoming frame sequence into base-layer frames and enhancement layer frames in such a way that there is a one-to-one relationship among each enhancement frame and each potential FRUC frame,

[0035] 2. encode the base-layer frame standard compliantly,

[0036] 3. interpolate a FRUC frame at the same time instance as the to-be-encoded enhancement frame, and,

[0037] 4. encode the enhancement frame based on the interpolated FRUC frame.

[0038] The benefits of this approach are two folds. First, by introducing one extra prediction mode for the enhancement layer frame, the compression ratio for the enhancement layer will increase. Thus, for a fixed transmission bandwidth, the possibility of dropping the enhancement layer bit stream is decreased as it becomes more unlikely that the enhancement layer bit stream will exceed the threshold of the transmission rate where the enhancement layer will be dropped. Second, however, even if the enhancement layer bit stream has to be dropped under degrading transmission conditions, due to the way the base-layer and enhancement-layer frames are partitioned, as explained herein, the decoder can still regenerate the missing enhancement layer frame by replacing them with interpolated FRUC frames.

[0039] FIG. 1 illustrates a conventional two-layered encoding scheme of a video sequence 106, including a base layer 104 and an enhancement layer 102. I and P frames such as P frames 112 and 114 are partitioned to base layer 104 due to their importance for the correct reconstruction of video sequence 106, while the less important B frames such as B frame 110 reside in enhancement layer 102. In this prior art approach, the reconstruction of enhancement frames in enhancement layer 102 only depends on the reconstructed base-layer frames in base layer 104, but does not depend on the reconstructed frames in enhancement layer 102. Specifically, B frames in enhancement layer 102 are not used as a reference for temporal prediction of the frame in base layer 104, although B frames can be used to predict future B frames.

[0040] FIG. 2 illustrates a video sequence 206 encoded in accordance with one embodiment of an encoding scheme, where an I frame 212 is a standard Intra frame (a frame that does not undergo temporal prediction) and a P frame 220 is a standard

Predicted frame. As seen in the figure, the frames in a main layer 202 is encoded through the use of an interpolated layer 204.

[0041] PF frame 216 are P-frames where a FRUC interpolated frame is used as one of the multiple reference pictures. For example, FRUC FRM 224 is a reference picture for PF FRM 216. In contrast, the standard approach uses I FRM 212 as the only reference picture for P frames. Thus, in one embodiment, a FRUC FRM is used as a reference picture for main layer decoding of interpolated frames.

[0042] BF frames are B-frames having FRUC interpolated frames as one of the multiple reference pictures. Thus, in one embodiment, reference pictures for BF frames may include one or more of I, B, P, PF or BF frames. For example, BF frame 214 is a B-frame having an FRUC interpolated frame 222 as one of the multiple reference pictures; and BF frame 218 is a B-frame having an FRUC interpolated frame 226 as one of the multiple reference pictures. In contrast, even though standard B (Bi-directional) predicted frames may have multiple reference pictures, these reference pictures only include I, B or P frames.

[0043] As referred to herein, the term "F frames" will refer to frames that are predicted frames with FRUC interpolated frames used as reference pictures. Thus, both PF frames and BF frames are both F frames. In one embodiment, multiple reference pictures are used for the prediction of F frames. In addition, FRUC interpolated frames may be the only reference picture used for the prediction of F frames. Further, the architecture described herein encompasses an approach where a portion of FRUC interpolated frames is used as references, such as using only specific macroblocks (which may be blocks of any size or shape) from the FRUC interpolated frame.

[0044] FIG. 3 illustrates a flow diagram of an exemplary FRUC encoding process for the bitstream described above in FIG. 2. In step 302, the incoming media (video) sequence is partitioned into F frames and non F frames, where F frames are PF and BF frames, as described above, and non-F frames are I, B or P frames. In one embodiment, there is a one-to-one relationship among each F frame and each potential FRUC frame. In other embodiments, the ratio of enhancement frames to potential FRUC frame does not have to have a one-to-one correspondence. Once the video sequence has been partitioned, operation continues with step 304, where the non-F frames are encoded based on the standard encoding scheme used. For example, the coding standard as promulgated in the H.264 standard may be used for these non-F frames. Then, in step 306, a FRUC frame is interpolated at the same time instance of the to-be-encoded

enhancement frame. In one embodiment, the FRUC frame may depend on both the reconstructed current frame and the stored previous frame. Other embodiments may use the other FRUC algorithms described above. In step 308, the frames in the enhancement layer are encoded with the assistance of the interpolated FRUC frame, as detailed in the description of FIG. 4. Thus, due to the availability of the FRUC frame, an extra prediction mode (FRUC mode) can be selected for the prediction frames in the main layer.

[0045] FIG. 4 illustrates an encoder block diagram configured in accordance with one embodiment. Initially, a P frame detector 402 determines if a new frame is an F frame. If not, an P or B frame encoder 404 encodes the frame and a variable length coding (VLC) encoder 406 generates a final bit stream for the encoded P or B frame to output in an output bit stream 404. The motion vectors for the P or B frame are stored in motion vector (MV) buffer 408, while reconstructed frame buffer 410 stores a reconstructed P or B frame. A FRUC unit 412 is coupled to MV buffer 408 and reconstructed frame buffer 410 to perform a FRUC algorithm and generate a FRUC frame, which is stored in FRUC frame buffer 414. As discussed above the FRUC frames are used to generate the various F (e.g., BF or PF) frames.

[0046] If F frame detector 402 determines that a new frame is an F frame, an F frame encoding unit 428 performs a macroblock by macroblock encoding of the new F frame. The encoded F frame macroblock are sent to a block 434 that determines the number of bits that need to be transmitted for the B frame, which includes the bits necessary to transmit the motion vectors. A B frame macroblock RD\_cost is then calculated in B-block RD\_cost calculator 436. Further, based on the B frame encoding, a comparator 430 compares (1) the value of the sum of absolute differences (SAD) of the F frame macroblock based on the FRUC frame retrieved from FRUC frame buffer 414 (SAD\_f) as subtracted from a corresponding B frame macroblock (SAD\_b) to (2) a predetermined threshold value (TH). If the results of the subtraction is greater than or equal to the threshold value, then the quality degradation would be too great and thus the encoder needs to send encoder assisted FRUC frame information. In one embodiment, the threshold value is six times the block sized being used (e.g., 16 x 16 block size). It should be noted that although the term "macroblock" is used herein, which typically refers to a standard block size of 16 X 16 pixel elements, any block size may be used. In addition, the blocks may be of any shape.

[0047] If difference between the SAD of the B frame macroblock and the SAD of the F frame macroblock is less than the threshold, then the encoder will indicate that the F frame macroblock should be used in module 432 and the bit to indicate the F-mode is placed into output bit stream 440. If not, a residual of the source (original) macroblock and collocated FRUC frame block is calculated in module 416, the results of which are quantized in discrete cosine transform (DCT)/Quantization/VLC module 418. The number of FRUC frame block bits necessary to transmit the FRUC frame information is determined in F-block bits module 420, and, based on that calculation, the FRUC frame macroblock RD\_cost is determined in F-block RD\_cost module 422. The RD\_cost calculation is described herein with regards to FIG. 5.

[0048] Based on the calculated B frame macroblock RD\_cost and F frame macroblock RD\_cost, comparator module 424 determines which bits to place into output bit stream 440. In one embodiment, if the B frame macroblock RD\_cost is less than the F frame macroblock RD\_cost, then the bits for the B frame macroblock, including the generated motion vectors, will be placed into output bit stream 440. Otherwise, the bits for the F frame macroblock will be placed into output bit stream 440.

[0049] FIG. 5 illustrates an algorithm implemented in accordance with one embodiment for a mode selection process of an enhancement layer B frame macroblock during the encoding the frames in the enhancement layer. In step 502, the forward and backward motion vectors are obtained for the B frame macroblock. In one embodiment, the vectors are obtained by performing conventional bi-directional motion estimation. Thus, no information from a FRUC frame is used. Then, in step 504, a best mode for the B frame macroblock is determined. In one embodiment, the best mode is selected from one of three modes: a forward prediction mode, a backward prediction mode and a bi-linear prediction mode. Further, the selection of the best mode is performed without the involvement of the FRUC frame.

[0050] In step 506, an error residue is obtained between the B frame macroblock based on the selected best mode and the original macroblock from the source frame (i.e., the to-be-encoded source frame). Error residue information between the original macroblock and the FRUC frame macroblock is then obtained in step 508. In one embodiment, error residue for the B frame macroblock and FRUC frame macroblock is calculated by subtracting the original macroblock from the B frame macroblock or FRUC frame macroblock, respectively. In addition, a variable referred to as the rate

distortion cost (RD\_cost) is calculated for both the B frame macroblock (B\_RD\_cost) and the FRUC frame macroblock (FRUC\_RD\_cost) given by the following formula:

[0051]  $RD\_cost = Distortion + \text{Lambda} * Rate$

[0052] In one embodiment, the Rate is the total bits used to encode a certain block (which uses the error residue of the respective macroblock), the Distortion is the distortion metric, and, Lambda is a predefined weighting factor that depends on the quantization of the macroblock. In one embodiment, Distortion is determined by a sum of absolute difference (SAD) calculation. In other embodiments, different distortion metrics may be used. Further, Lambda, in one embodiment, is an empirically derived formula dependent on the quantization parameter.

[0053] As described above, whether the FRUC frame or B frame mode of encoding is chosen is based on the sum-of-absolute-difference (SAD) value, which is effectively a measure of the distortion, and the number of bits required to encode the residual. These two values are summed, one being weighted by lambda, and the "best" mode is decided based on that which minimizes the result. The reason the summation of distortion and bit-rate is used is so the encoder can make a trade-off between image quality and bit-rate. So, for example, if the encoder encountered a highly complex macroblock that would suffer from a great deal of distortion if it used the same quantization parameter as before, it could lower the quantization value to allow for a tradeoff in distortion, even though it would require more bits to encode.

[0054] Once the error residue and, consequently, the B\_RD\_cost (the RD\_cost of the B frame macroblock) and the error residue and, consequently, the FRUC\_RD\_cost (the RD\_cost of the FRUC frame macroblock) have been determined, step 510 compares the costs and determines if FRUC\_RD\_cost is less than B\_RD\_cost. If so, then the FRUC mode is selected in step 512. Otherwise, the best mode as determined in step 504 is chosen in step 514.

[0055] FIG. 6 illustrates a video sequence 606 encoded in accordance with an encoding scheme of one embodiment, with a layer decision block 608 that partitions media data into two layers, including a base layer 604 and an enhancement layer 602. I and P frames such as P frames 612 and 614 are partitioned to base layer 604 due to their importance for the correct reconstruction of video sequence 606, while the less important B frames such as B frame 610 reside in enhancement layer 602. As illustrated in the figure, a FRUC frame 616 may also be in existence in base layer 604.

[0056] FIG. 7 illustrates a flow diagram of a FRUC encoding process in accordance with one embodiment, as described above. In step 702, the incoming media (video) sequence is partitioned into base layer frames and enhancement layer frames. In one embodiment, there is a one-to-one relationship among each enhancement frame and each potential FRUC frame. In other embodiments, the ratio of enhancement frames to potential FRUC frame does not have to have a one-to-one correspondence. Once the video sequence has been partitioned, operation continues with step 704, where the frames in the base-layer are encoded based on the standard encoding scheme used. For example, the coding standard as promulgated in the H.264 standard may be used. Then, in step 706, a FRUC frame is interpolated at the same time instance of the to-be-encoded enhancement frame. In one embodiment, the FRUC frame may depend on both the reconstructed current frame and the stored previous frame. Other embodiments may use the other FRUC algorithms described above. In step 708, the frames in the enhancement layer are encoded with the assistance of the interpolated FRUC frame.

[0057] It should be noted that the modules and processing blocks mentioned herein may be implemented in a variety of hardware/software implementations. Thus, one of ordinary skill in the art would understand that, for example, the same processor may perform the FRUC operation as well as the SAD calculations. Further, a digital signal processor (DSP) or other specialized processor may be used in conjunction with a general purpose processor to implement the functions described herein. References to modules or units performing a specific function or functions should not be limited to include a particular circuit for performing said function or functions, but may, again, include a processor configured to perform said function or functions.

[0058] The present invention achieves savings in the transmitted bit stream for very little to no compromise in visual quality, as based on the mode selection method described above. For example, there is a bit-rate/bandwidth reduction by augmentation of the standard (H.26X/MPEG-X) bit stream syntax at the same quality level. This will decrease the possibility of dropping the enhancement bit stream, and consequently, improve reconstructed video quality. In one embodiment, in encoding the enhancement layer, motion vector information need not be transmitted since that information can be recovered/computed at the decoder through simple interpolation functions. Thus all of the bit rate savings can be realized.

[0059] In addition, computational scalability may be achieved by the adaptive selection of an appropriate FRUC algorithm based on the optimal usage of the receiver's

hardware resources. For example, if the decoder has built-in motion estimation accelerator, the seeded ME assisted MCI with MV processing FRUC algorithm may be selected. Better temporal scalability may be achieved as, by using the FRUC features, the video will playback full frame rate where FRUC frames are inserted only when the base-layer bit stream is received. Power savings at the decoder may also be obtained, particularly in handhelds, where FRUC assisted reconstruction of the video stream requires fewer cycles than the reconstruction of a complete B frame.

[0060] It is predicted that B frames occupy up to 30% of the total bit rate when an IBP group of pictures (GOP) structure is used. Hence, the present invention, by reducing the data allocated for transmitting B frames, may decrease the overall bit rate. For example, for IBP GOP structures, the bit rate may be reduced by up to 15% when both the base and enhancement layers are put together. This percentage goes up when IBBP GOP or IBBBP GOP structures are used as these structures utilize more B frames. More for those with more number of B frames between I and P.

[0061] FIG. 8 shows a block diagram of an access terminal 802x and an access point 804x in a wireless system. An "access terminal," as discussed herein, refers to a device providing voice and/or data connectivity to a user. The access terminal may be connected to a computing device such as a laptop computer or desktop computer, or it may be a self contained device such as a personal digital assistant. The access terminal can also be referred to as a subscriber unit, mobile station, mobile, remote station, remote terminal, user terminal, user agent, or user equipment. The access terminal may be a subscriber station, wireless device, cellular telephone, PCS telephone, a cordless telephone, a Session Initiation Protocol (SIP) phone, a wireless local loop (WLL) station, a personal digital assistant (PDA), a handheld device having wireless connection capability, or other processing device connected to a wireless modem. An "access point," as discussed herein, refers to a device in an access network that communicates over the air-interface, through one or more sectors, with the access terminals. The access point acts as a router between the access terminal and the rest of the access network, which may include an IP network, by converting received air-interface frames to IP packets. The access point also coordinates the management of attributes for the air interface.

[0062] For the reverse link, at access terminal 802x, a transmit (TX) data processor 814 receives traffic data from a data buffer 812, processes (e.g., encodes, interleaves, and symbol maps) each data packet based on a selected coding and modulation scheme,

and provides data symbols. A data symbol is a modulation symbol for data, and a pilot symbol is a modulation symbol for pilot (which is known a priori). A modulator 816 receives the data symbols, pilot symbols, and possibly signaling for the reverse link, performs (e.g., OFDM) modulation and/or other processing as specified by the system, and provides a stream of output chips. A transmitter unit (TMTR) 818 processes (e.g., converts to analog, filters, amplifies, and frequency upconversion) the output chip stream and generates a modulated signal, which is transmitted from an antenna 820.

[0063] At access point 804x, the modulated signals transmitted by access terminal 802x and other terminals in communication with access point 804x are received by an antenna 852. A receiver unit (RCVR) 854 processes (e.g., conditions and digitizes) the received signal from antenna 852 and provides received samples. A demodulator (Demod) 856 processes (e.g., demodulates and detects) the received samples and provides detected data symbols, which are noisy estimate of the data symbols transmitted by the terminals to access point 804x. A receive (RX) data processor 858 processes (e.g., symbol demaps, deinterleaves, and decodes) the detected data symbols for each terminal and provides decoded data for that terminal.

[0064] For the forward link, at access point 804x, traffic data is processed by a TX data processor 860 to generate data symbols. A modulator 862 receives the data symbols, pilot symbols, and signaling for the forward link, performs (e.g., OFDM) modulation and/or other pertinent processing, and provides an output chip stream, which is further conditioned by a transmitter unit 864 and transmitted from antenna 852. The forward link signaling may include power control commands generated by a controller 870 for all terminals transmitting on the reverse link to access point 804x. At access terminal 802x, the modulated signal transmitted by access point 804x is received by antenna 820, conditioned and digitized by a receiver unit 822, and processed by a demodulator 824 to obtain detected data symbols. An RX data processor 826 processes the detected data symbols and provides decoded data for the terminal and the forward link signaling. Controller 830 receives the power control commands, and controls data transmission and transmit power on the reverse link to access point 804x. Controllers 830 and 870 direct the operation of access terminal 802x and access point 804x, respectively. Memory units 832 and 872 store program codes and data used by controllers 830 and 870, respectively.

[0065] The disclosed embodiments may be applied to any one or combinations of the following technologies: Code Division Multiple Access (CDMA) systems, Multiple-

Carrier CDMA (MC-CDMA), Wideband CDMA (W-CDMA), High-Speed Downlink Packet Access (HSDPA), Time Division Multiple Access (TDMA) systems, Frequency Division Multiple Access (FDMA) systems, and Orthogonal Frequency Division Multiple Access (OFDMA) systems.

[0066] It should be noted that the methods described herein may be implemented on a variety of communication hardware, processors and systems known by one of ordinary skill in the art. For example, the general requirement for the client to operate as described herein is that the client has a display to display content and information, a processor to control the operation of the client and a memory for storing data and programs related to the operation of the client. In one embodiment, the client is a cellular phone. In another embodiment, the client is a handheld computer having communications capabilities. In yet another embodiment, the client is a personal computer having communications capabilities. In addition, hardware such as a GPS receiver may be incorporated as necessary in the client to implement the various embodiments. The various illustrative logics, logical blocks, modules, and circuits described in connection with the embodiments disclosed herein may be implemented or performed with a general purpose processor, a digital signal processor (DSP), an application specific integrated circuit (ASIC), a field programmable gate array (FPGA) or other programmable logic device, discrete gate or transistor logic, discrete hardware components, or any combination thereof designed to perform the functions described herein. A general-purpose processor may be a microprocessor, but, in the alternative, the processor may be any conventional processor, controller, microcontroller, or state machine. A processor may also be implemented as a combination of computing devices, e.g., a combination of a DSP and a microprocessor, a plurality of microprocessors, one or more microprocessors in conjunction with a DSP core, or any other such configuration.

[0067] The disclosed embodiments may be applied to any one or combinations of the following technologies: Code Division Multiple Access (CDMA) systems, Multiple-Carrier CDMA (MC-CDMA), Wideband CDMA (W-CDMA), High-Speed Downlink Packet Access (HSDPA), Time Division Multiple Access (TDMA) systems, Frequency Division Multiple Access (FDMA) systems, and Orthogonal Frequency Division Multiple Access (OFDMA) systems.

[0068] It should be noted that the methods described herein may be implemented on a variety of communication hardware, processors and systems known by one of

ordinary skill in the art. For example, the general requirement for the client to operate as described herein is that the client has a display to display content and information, a processor to control the operation of the client and a memory for storing data and programs related to the operation of the client. In one embodiment, the client is a cellular phone. In another embodiment, the client is a handheld computer having communications capabilities. In yet another embodiment, the client is a personal computer having communications capabilities.

[0069] The various illustrative logics, logical blocks, modules, and circuits described in connection with the embodiments disclosed herein may be implemented or performed with a general purpose processor, a digital signal processor (DSP), an application specific integrated circuit (ASIC), a field programmable gate array (FPGA) or other programmable logic device, discrete gate or transistor logic, discrete hardware components, or any combination thereof designed to perform the functions described herein. A general-purpose processor may be a microprocessor, but, in the alternative, the processor may be any conventional processor, controller, microcontroller, or state machine. A processor may also be implemented as a combination of computing devices, e.g., a combination of a DSP and a microprocessor, a plurality of microprocessors, one or more microprocessors in conjunction with a DSP core, or any other such configuration.

[0070] The steps of a method or algorithm described in connection with the embodiments disclosed herein may be embodied directly in hardware, in a software module executed by a processor, or in a combination of the two. A software module may reside in RAM memory, flash memory, ROM memory, EPROM memory, EEPROM memory, registers, a hard disk, a removable disk, a CD-ROM, or any other form of storage medium known in the art. An exemplary storage medium is coupled to the processor, such that the processor can read information from, and write information to, the storage medium. In the alternative, the storage medium may be integral to the processor. The processor and the storage medium may reside in an ASIC. The ASIC may reside in a user terminal. In the alternative, the processor and the storage medium may reside as discrete components in a user terminal.

[0071] The description of the disclosed embodiments is provided to enable any person skilled in the art to make or use the invention. Various modifications to these embodiments may be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other embodiments, e.g., in an instant

messaging service or any general wireless data communication applications, without departing from the spirit or scope of the described embodiments. Thus, the scope of the description is not intended to be limited to the embodiments shown herein but is to be accorded the widest scope consistent with the principles and novel features disclosed herein. The word “exemplary” is used exclusively herein to mean “serving as an example, instance, or illustration.” Any embodiment described herein as “exemplary” is not necessarily to be construed as advantageous over other embodiments.

## CLAIMS

### WHAT IS CLAIMED IS:

1. A method for decoding a video stream, the method comprising:  
partitioning the video stream into a main layer having a plurality of main layer frames, and an interpolated layer having a plurality of interpolated layer frames;  
interpolating a frame rate up conversion (FRUC) frame; and  
encoding the plurality of main layer frames in the main layer with the assistance of the interpolated FRUC frame.
2. The method of claim 1, wherein there is a one-to-one relationship among each main layer frame and each potential FRUC frame.
3. The method of claim 1, wherein the FRUC frame depends on a reconstructed current frame and a stored previous frame.
4. The method of claim 1, wherein said interpolating the FRUC frame comprises interpolating the FRUC frame at the same time instance of the to-be-encoded main layer frame.
5. The method of claim 1, wherein said encoding plurality of main layer frames in the main layer with the assistance of the interpolated FRUC frame comprises performing a mode-selection process of a macroblock of one interpolated layer frame in the plurality of interpolated layer frames.
6. A computer-readable medium having instructions stored thereon, the stored instructions, when executed by a processor, cause the processor to perform a method for encoding a video stream, the method comprising:  
partitioning the video stream into a main layer having a plurality of main layer frames, and an interpolated layer having a plurality of interpolated layer frames;  
interpolating a frame rate up conversion (FRUC) frame; and  
encoding the plurality of main layer frames in the main layer with the assistance of the interpolated FRUC frame.

7. The computer-readable medium of claim 6, wherein there is a one-to-one relationship among each main layer frame and each potential FRUC frame.
8. The computer-readable medium of claim 6, wherein the FRUC frame depends on a reconstructed current frame and a stored previous frame.
9. The computer-readable medium of claim 6, wherein said interpolating the FRUC frame comprises interpolating the FRUC frame at the same time instance of the to-be-encoded main layer frame.
10. The computer readable medium of claim 6, wherein of encoding plurality of main layer frames in the main layer with the assistance of the interpolated FRUC frame comprises performing a mode-selection process of a macroblock of one interpolated layer frame in the plurality of interpolated layer frames.
11. An apparatus for encoding a video stream comprising:
  - means for partitioning the video stream into a main layer having a plurality of main layer frames, and an interpolated layer having a plurality of interpolated layer frames;
  - means for interpolating a frame rate up conversion (FRUC) frame; and
  - means for encoding the plurality of main layer frames in the main layer with the assistance of the interpolated FRUC frame.
12. The apparatus of claim 11, wherein there is a one-to-one relationship among each main layer frame and each potential FRUC frame.
13. The apparatus of claim 11, wherein the FRUC frame depends on a reconstructed current frame and a stored previous frame.
14. The apparatus of claim 11, wherein a means for interpolating the FRUC frame comprises means for interpolating the FRUC frame at the same time instance of the to-be-encoded main layer frame.
15. The apparatus of claim 11, wherein the means for encoding plurality of main layer frames in the main layer with the assistance of the interpolated FRUC frame comprises means for performing a mode-selection process of a macroblock of one interpolated layer frame in the plurality of interpolated layer frames.

16. A video-encoder processor receiving a video stream, the video encoder processor comprising:

a frame rate-up-conversion (FRUC) module, the FRUC module generating a plurality of FRUC frames from the received video stream;

a rate distortion cost (RD\_cost) calculation module coupled to the FRUC module, the RD\_cost calculation module receiving the plurality of FRUC frames and calculating a F frame macroblock RD\_cost based on a macroblock in one of the plurality of FRUC frames; and

a mode decision module coupled to the RD\_cost calculation module, the mode decision module configured to compare the F frame macroblock RD\_cost based on the macroblock in one of the plurality of FRUC frames to a B frame macroblock RD\_cost for a corresponding macroblock in a corresponding B frame.

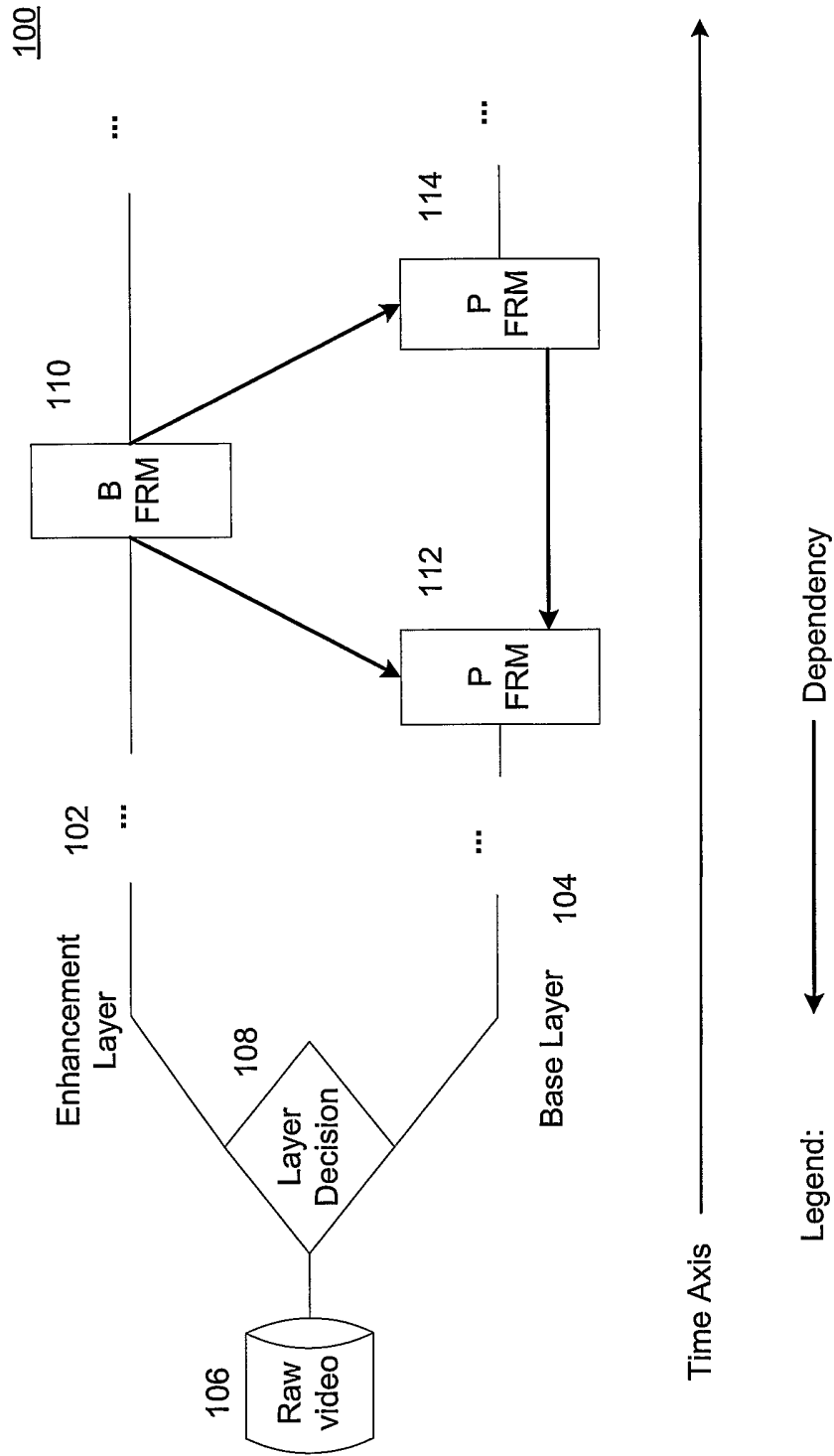
17. The video-encoder processor of claim 16, wherein the mode-decision module is further configured to select a F frame encoding mode for an output video stream when the F frame macroblock RD\_cost based on the macroblock in one of the plurality of FRUC frames is lower than the B frame macroblock RD\_cost for the corresponding macroblock in the corresponding B frame.

18. The video-encoder processor of claim 16, further comprising a B frame encoder, the B frame encoder generating the corresponding B frame.

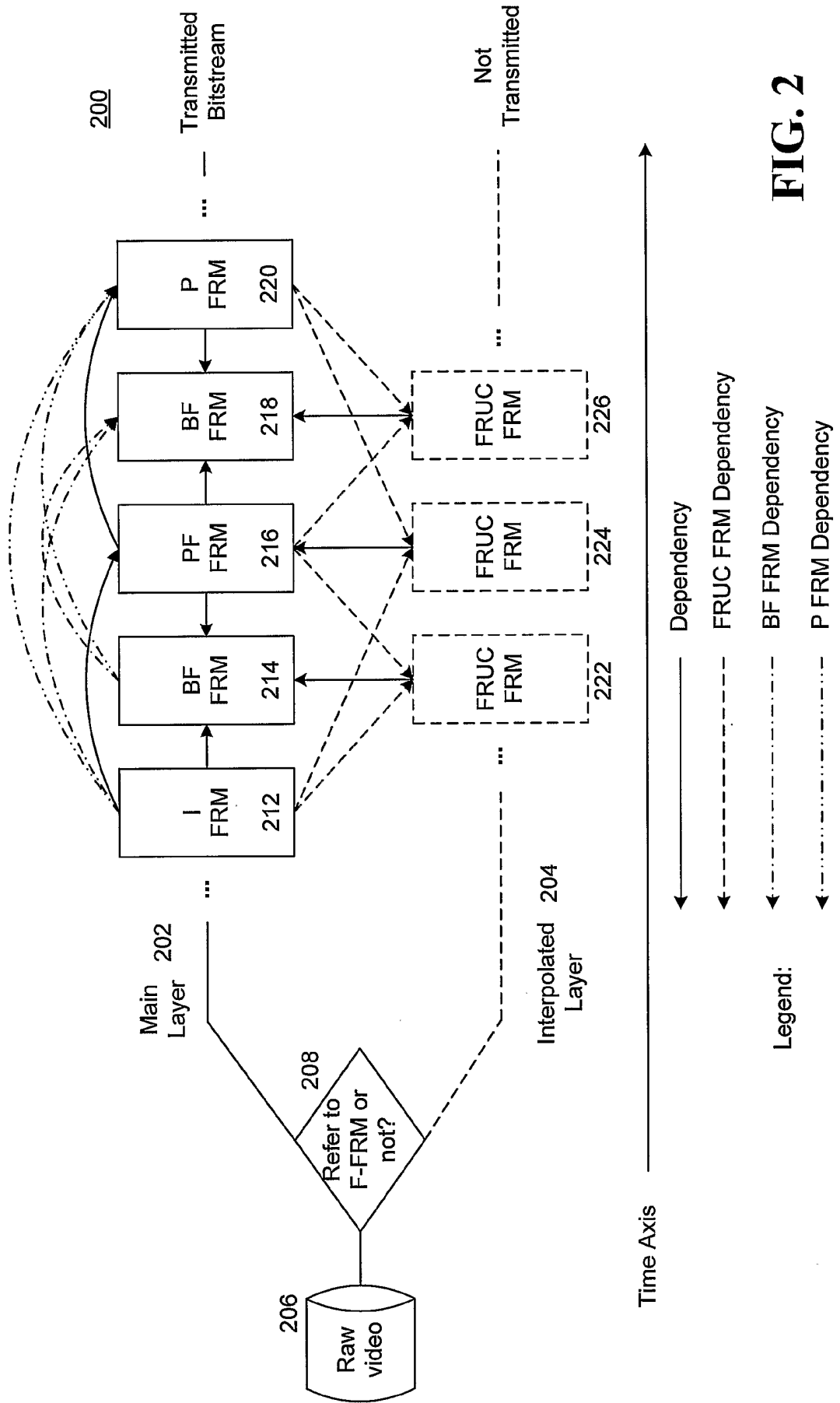
19. The video-encoder processor of claim 16, further comprising a sum-of-absolute difference (SAD) calculation module coupled to the FRUC module, the SAD calculation module configured to calculate a SAD value based on the macroblock in one of the plurality of FRUC frames and a SAD value for the corresponding macroblock in the corresponding B frame.

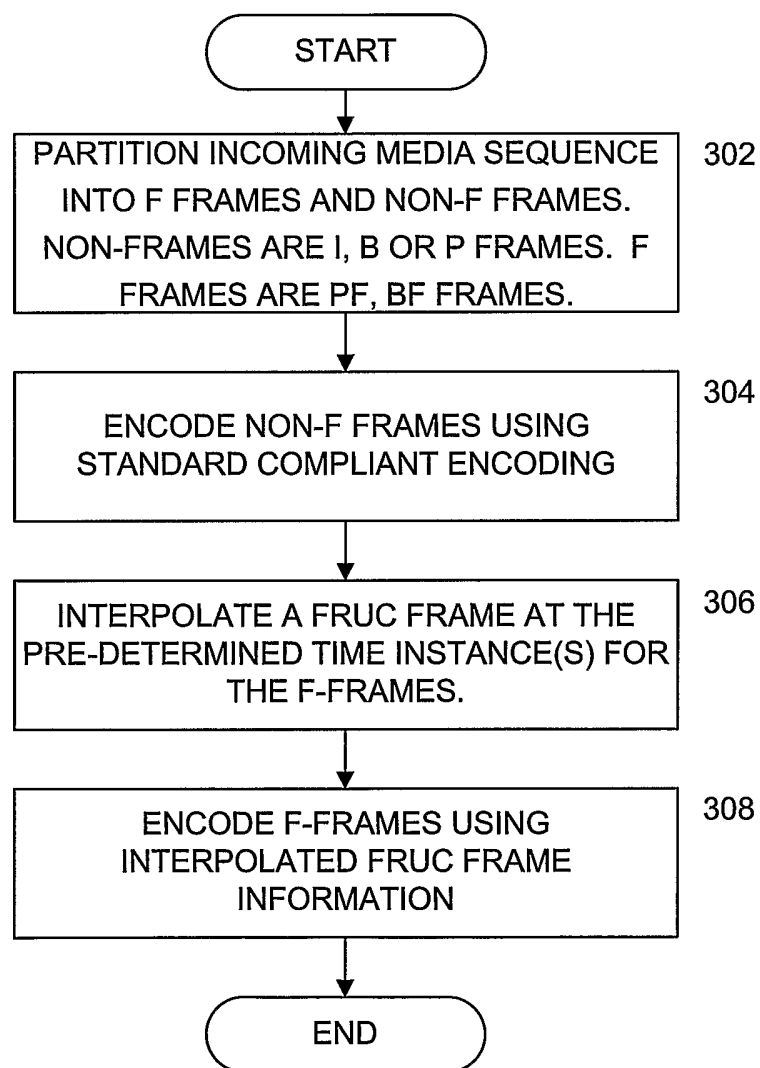
20. The video encoder processor of claim 19, wherein the SAD calculation module is configured to calculate a difference of the SAD value based on the macroblock in one of the plurality of FRUC frames and the SAD value for the corresponding macroblock in the corresponding B frame, and to compare the difference to a predetermined threshold, the SAD calculation module generating a single bit to indicate a F frame mode if the difference is lower than the predetermined threshold.

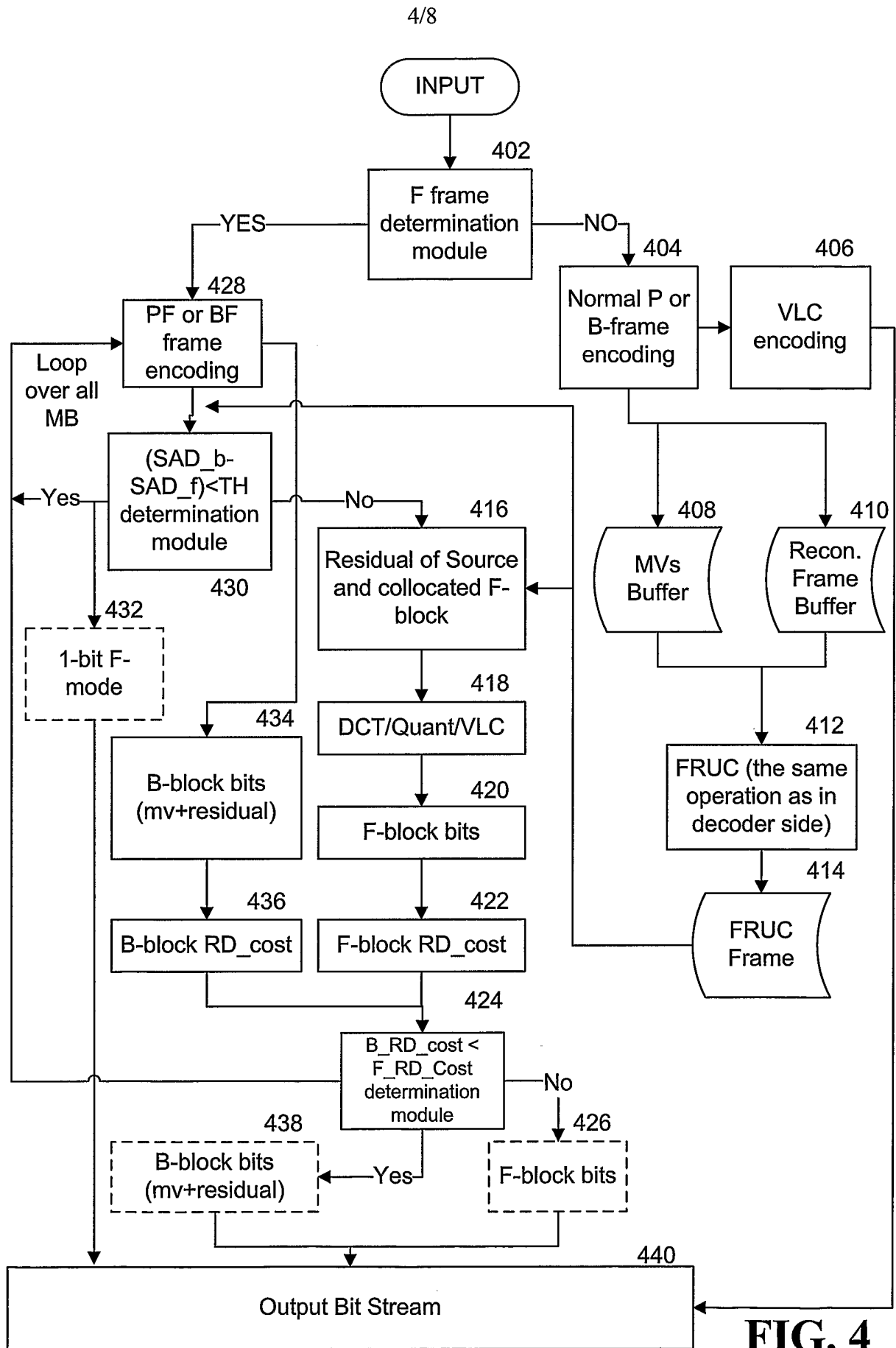
21. A method for encoding a video stream, the method comprising:
  - generating a plurality of a frame rate up conversion (FRUC) frames from the received video stream;
  - calculating a F frame macroblock RD\_cost based on a macroblock in one of the plurality of FRUC frames;
  - comparing the calculated F frame macroblock RD\_cost to a B frame macroblock RD\_cost for a corresponding macroblock in a corresponding B frame; and
  - selecting an encoding scheme based on the result of the comparison.
  
22. An apparatus for encoding a video stream, comprising:
  - means for generating a plurality of a frame rate up conversion (FRUC) frames from the received video stream;
  - means for calculating a F frame macroblock RD\_cost based on a macroblock in one of the plurality of FRUC frames;
  - means for comparing the calculated F frame macroblock RD\_cost to a B frame macroblock RD\_cost for a corresponding macroblock in a corresponding B frame; and
  - means for selecting an encoding scheme based on the result of the comparison.
  
23. A computer-readable medium having instructions stored thereon, the stored instructions, when executed by a processor, cause the processor to perform a method for encoding a video stream, the method comprising:
  - generating a plurality of a frame rate up conversion (FRUC) frames from the received video stream;
  - calculating a F frame macroblock RD\_cost based on a macroblock in one of the plurality of FRUC frames;
  - comparing the calculated F frame macroblock RD\_cost to a B frame macroblock RD\_cost for a corresponding macroblock in a corresponding B frame; and
  - selecting an encoding scheme based on the result of the comparison.



**FIG. 1**  
(prior art)

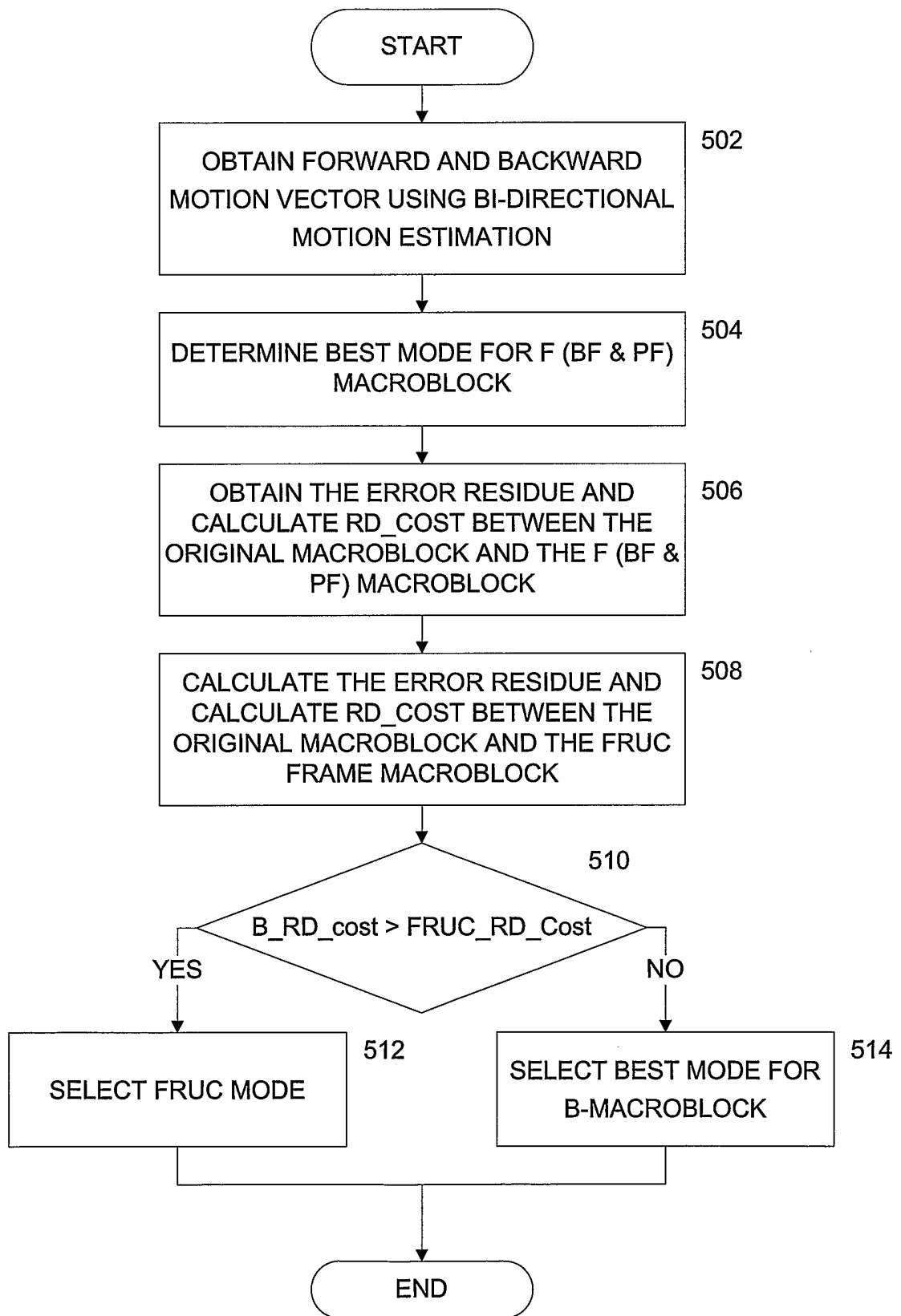


SINGLE LAYER ENCODING**FIG. 3**

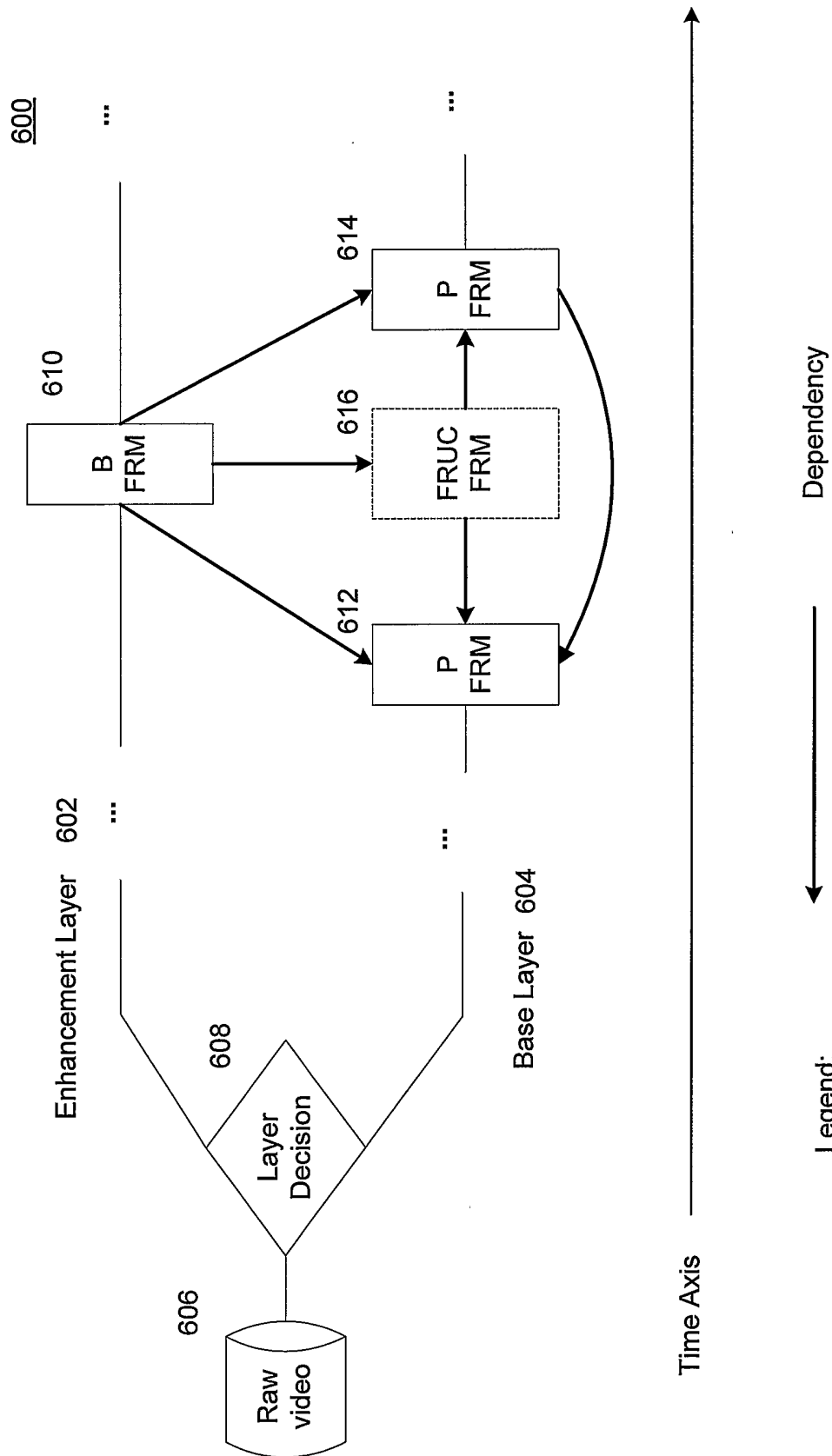


**FIG. 4**

5/8

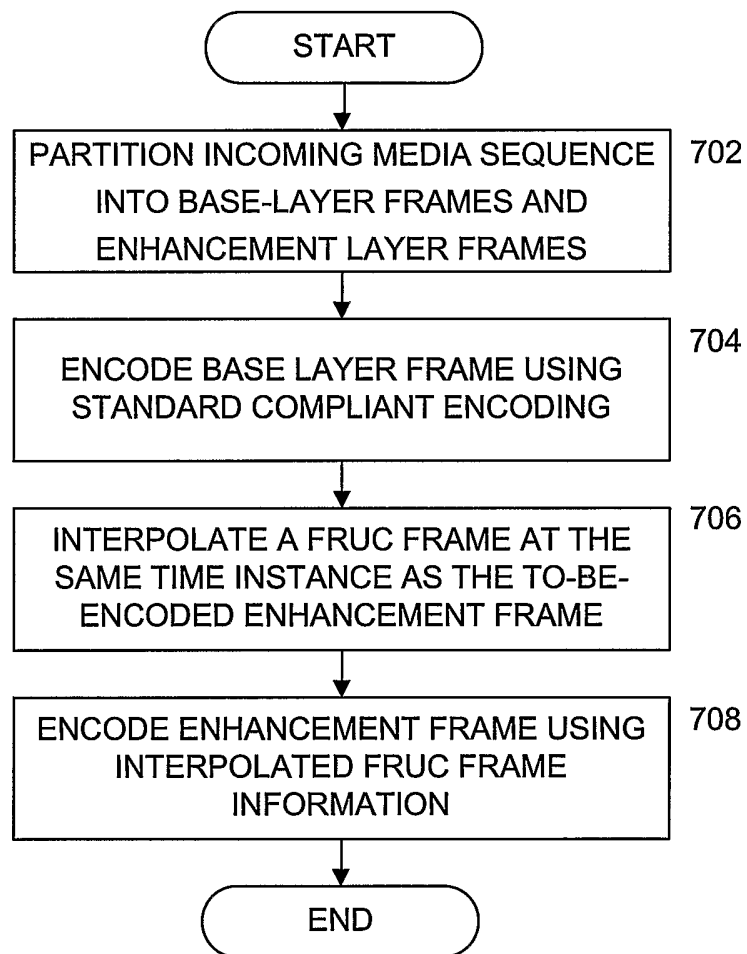


**FIG. 5**



**FIG. 6**

7/8

TWO-LAYER ENCODING**FIG. 7**

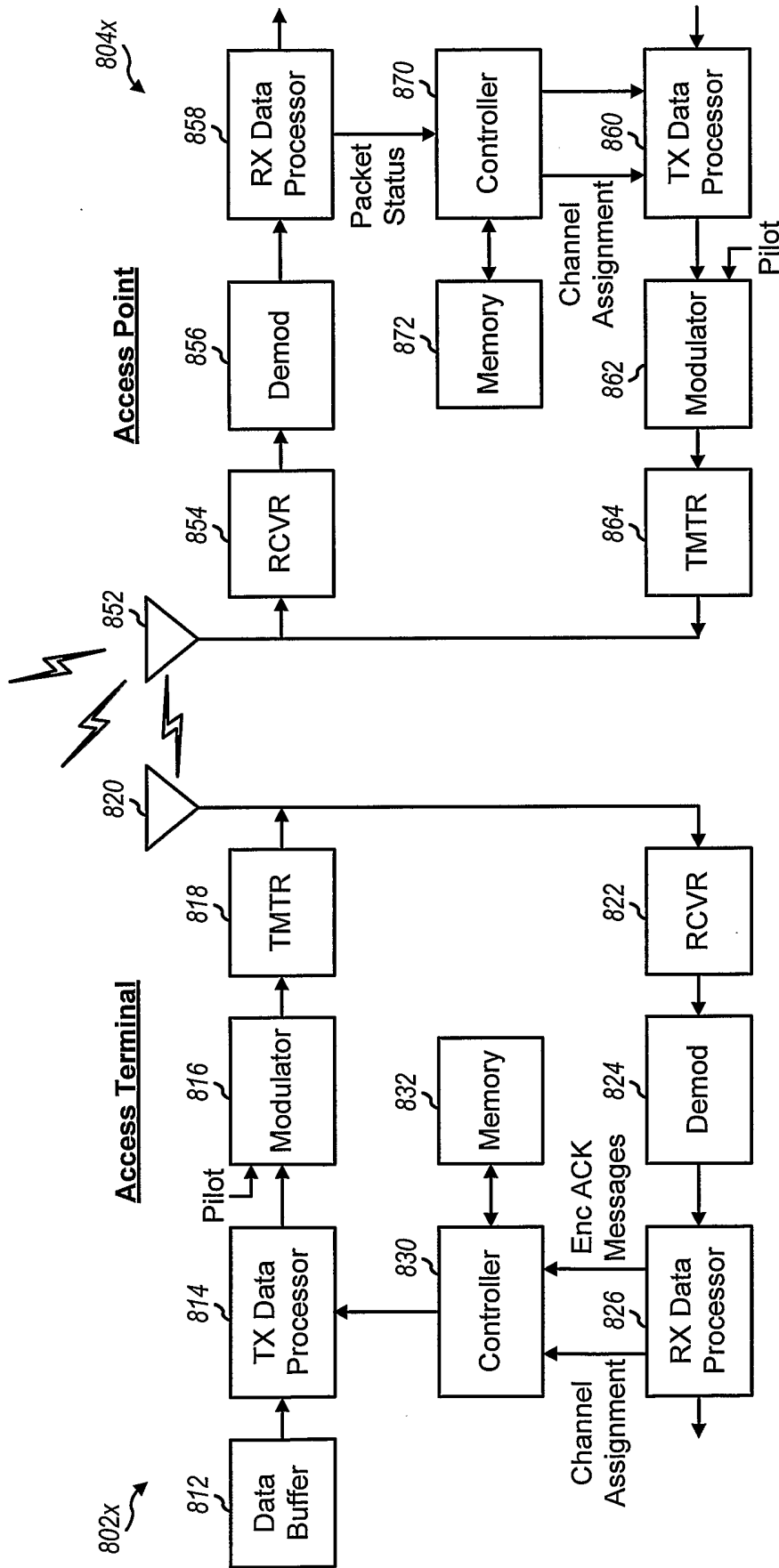


FIG. 8