



- (51) **International Patent Classification:**
H04N 21/236 (2011.01) *H04N 19/30* (2014.01)
- (21) **International Application Number:**
PCT/US2014/061985
- (22) **International Filing Date:**
23 October 2014 (23.10.2014)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
61/896,367 28 October 2013 (28.10.2013) US
14/521,763 23 October 2014 (23.10.2014) US
- (71) **Applicant:** ARRIS ENTERPRISES, INC. [US/US];
3871 Lakefield Drive, Suwanee, Georgia 30024 (US).
- (72) **Inventors:** NARASIMHAN, Mandayam; 13002 Walking
Path Place, San Diego, California 92130 (US). LUTHRA,
Ajay K.; 13695 Winstanley Way, San Diego, California
92130 (US).
- (74) **Agents:** WARD, Thomas A. et al.; 3871 Lakefield Drive,
Suwanee, Georgia 30024 (US).

- (81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

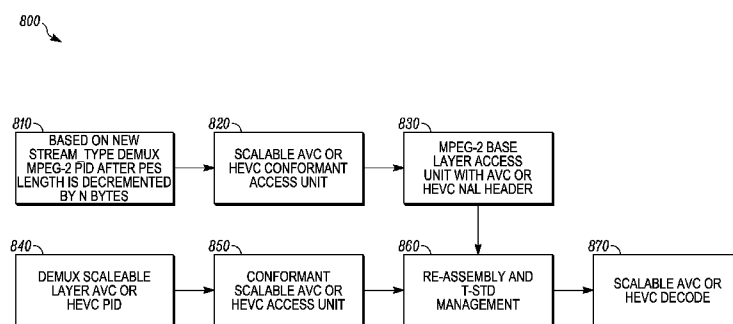
(54) **Title:** METHOD AND APPARATUS FOR DECODING AN ENHANCED VIDEO STREAM

FIG. 8

(57) **Abstract:** Disclosed are herein are methods and systems for encoding and decoding enhanced video streams using a legacy and scalable video components. In one aspect, a method of encoding video data in a packetized elementary stream (PES), is disclosed. The method includes receiving a signal including frames of image data, wherein the frames include base layer and enhancement layer components; encoding the frames of image data to form video access units (AUs), wherein base layer AUs are formed for base layer components and enhancement layers AUs are formed for enhancement layer components, each AU including: an elementary stream header (ELSM) including image data metadata parameters associated with decoding and displaying the image data; a PES header including a presentation time stamp (PTS) and a PES packet length indicating the length of the PES header; one or more codestreams associated with a frame of the image data; adding a network abstraction layer (NAL) unit header defined by an enhancement layer video standard to each base layer AU, the NAL unit header being a constant value and length; and forming a modified PES header by increasing the PES packet length to include the length of the NAL unit header in it

METHOD AND APPARATUS FOR DECODING AN ENHANCED VIDEO STREAM

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims priority to U.S. Provisional Patent Application 61/896,367, entitled "Proposal for PES encapsulation of AVC base layer HEVC NAL units" filed on October 28, 2013, which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

[0002] The present disclosure is related generally to systems and methods for decoding an enhanced video stream.

BACKGROUND

[0003] Digital video capabilities can be incorporated into a wide range of devices, including digital televisions, digital direct broadcast systems, wireless broadcast systems, personal digital assistants (PDAs), laptop or desktop computers, tablet computers, e-book readers, digital cameras, digital recording devices, digital media players, video gaming devices, video game consoles, cellular or satellite radio telephones, "smart phones," video conferencing devices, video streaming devices, etc. Digital video devices implement video coding techniques, such as those described in the standards defined by MPEG-2, MPEG-4, H.264/MPEG-4, Part 10, Advanced Video Coding (AVC), High Efficiency Video Coding (HEVC), and extensions of such standards, such as Scalable Video Coding (SVC). The video devices may transmit, receive, encode, decode, and/or store digital video information more efficiently by implementing such video coding techniques.

[0004] Currently the HEVC scalable specification supports an application model where the base layer is AVC video with enhancements using scalable HEVC. However,

the transport of such scalable streams requires the AVC base layer to be available to a legacy decoder without the scalable HEVC enhancements in the video stream.

SUMMARY

[0005] Described herein are methods and systems for encoding and decoding enhanced video streams using a legacy and scalable video components.

[0006] In a first aspect, a method of encoding video data in a packetized elementary stream (PES), is disclosed. The method includes receiving a signal including frames of image data, wherein the frames include base layer and enhancement layer components; encoding the frames of image data to form video access units (AUs), wherein base layer AUs are formed for base layer components and enhancement layers AUs are formed for enhancement layer components, each AU including: an elementary stream header (ELSM) including image data metadata parameters associated with decoding and displaying the image data; a PES header including a presentation time stamp (PTS) and a PES packet length indicating the length of the PES header; one or more codestreams associated with a frame of the image data; adding a network abstraction layer (NAL) unit header defined by an enhancement layer video standard to each base layer AU, the NAL unit header being a constant value and length; and forming a modified PES header by increasing the PES packet length to include the length of the NAL unit header in it. In an embodiment of the first aspect, the method further includes mapping the AUs into PES packets; and ordering the PES packets to form a PES in a transport stream, wherein the transport stream includes a base layer formed from base layer AUs and an enhancement layer formed from enhancement layer AUs. In an embodiment of the first aspect, the base layer AUs include AVC or MPEG-2 video data. In an embodiment of the first aspect, the enhancement layer AUs include HEVC or AVC video data. In an embodiment of the first aspect, the base layer AUs are AVC video data and the NAL unit header is HEVC video data. In an embodiment of the first aspect, a NAL unit header directly precedes each base layer AU. In an embodiment of the first aspect, each HEVC

NAL unit header is configured to signal and carry each AVC AU. In an embodiment of the first aspect, each AVC AU includes at least one of: AU delimiter, SEI, SPS, PPS, and video slices. In an embodiment of the first aspect, the method further includes mapping the AUs into PES packets; and ordering the PES packets to form a PES in a transport stream, wherein the transport stream includes a base layer formed from base layer AUs and an enhancement layer formed from enhancement layer AUs, wherein the transport stream includes at least two packet identifiers (PIDs) and a program map table (PMT), and wherein the AVC base layer PES packets are carried in a first dedicated PID. In an embodiment of the first aspect, the PMT signals the base layer AVC with MPEG defined stream type of 0x1B. In an embodiment of the first aspect, the NAL unit header is an HEVC enhancement layer and wherein the HEVC enhancement layer is carried in a second dedicated PID. In an embodiment of the first aspect, the PMT signals the HVC enhancement layer with a new stream type value defined by MPEG. In an embodiment of the first aspect, the base layer AUs are MPEG-2 video data and the NAL unit header is defined in AVC video data. In an embodiment of the first aspect, the base layer AUs are MPEG-2 video data and the NAL unit header is defined in HEVC video data.\

[0007] In a second aspect, a method of decoding video data in a packetized elementary stream (PES) is disclosed. The method includes receiving a transport stream including the PES, wherein the PES includes frames on image data in video access units (AUs) including both base layer AUs and enhancement layer AUs, wherein base layer AUs are formed from base layer components and enhancement layer AUs are formed from enhancement layer components, each AU including: an elementary stream header (ELSM) including image data metadata parameters associated with decoding and displaying the image data; a network abstraction layer (NAL) unit header for each base layer AU; a modified PES header including a presentation time stamp (PTS) and a PES packet length indicating the length of the PES header plus a constant value, wherein the constant value represents the length of the NAL unit header; one or more codestreams associated with a frame of the image data; removing the modified PES header and NAL unit for each base layer AU; and decoding each base layer AU to form a signal including

frames of image data. In an embodiment of the second aspect, the decoding is performed by a legacy decoder.

[0008] In a third aspect, a method of decoding video data in a packetized elementary stream (PES) is disclosed. The method includes receiving a transport stream including the PES, wherein the PES includes frames of image data in video access units (AUs) including both base layer AUs and enhancement layer AUs, wherein base layer AUs are formed from base layer components and enhancement layer AUs are formed from enhancement layer components, each AU including: an elementary stream header (ELSM) including image data metadata parameters associated with decoding and displaying the image data; a network abstraction layer (NAL) unit header for each base layer AU; a modified PES header including a presentation time stamp (PTS) and a PES packet length indicating the length of the PES header plus a constant value, wherein the constant value represents the length of the NAL unit header; one or more codestreams associated with a frame of the image data; forming an original PES header by decreasing the modified PES packet length by the constant value that represents the NAL unit header length; removing the original PES header for each base layer AU; removing the original PES header for each enhancement layer AU; and decoding each base layer AU and each enhancement layer AU to form a signal including frames of image data. In an embodiment of the third aspect, the method further includes receiving a signal indicating that the modified PES header will be decremented by the constant value.

[0008] In a fourth aspect, a system configured to encode video data in a packetized elementary stream (PES) is disclosed. The system includes a processor configured to receive a signal including frames of image data, wherein the frames include base layer and enhancement layer components; encode the frames of image data to form video access units (AUs), wherein base layer AUs are formed for base layer components and enhancement layers AUs are formed for enhancement layer components, each AU including: an elementary stream header (ELSM) including image data metadata parameters associated with decoding and displaying the image data; a PES header

including a presentation time stamp (PTS) and a PES packet length indicating the length of the PES header; one or more codestreams associated with a frame of the image data; add a network abstraction layer (NAL) unit header defined by an enhancement layer video standard to each base layer AU, the NAL unit header being a constant value and length; form a modified PES header by increasing the PES packet length to include the length of the NAL unit header in it; and a storage device configured to store the frames of image data, and executable instructions for controlling the processor.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[0009] While the appended claims set forth the features of the present techniques with particularity, these techniques, together with their objects and advantages, may be best understood from the following detailed description taken in conjunction with the accompanying drawings of which:

[0010] FIG. 1 is a system context diagram illustrating a content distribution system, according to an embodiment;

[0011] FIG. 2 is a block schematic illustration of the architecture of a first system for supplying compressed video material for presentation;

[0012] FIG. 3 is a block schematic illustration of the architecture and use case for transmission of compressed scalable video coded material for decode and presentation, according to an embodiment;

[0013] FIG. 4 is a breakout frame of an MPEG-2 transport stream, illustrating encoding packets carried in the MPEG-2 transport stream, according to an embodiment;

[0014] FIG. 5 is a flowchart of a representative method for delivering only an AVC base layer to a legacy decoder, according to an embodiment;

[0015] FIG. 6 is a flowchart of a representative method for assembling an AVC base layer and HEVC enhancement layer by a scalable HEVC decoder, according to an embodiment;

[0016] FIG. 7 is a flowchart of a representative method for delivering only MPEG-2 base layer to a legacy decoder, according to an embodiment;

[0017] FIG. 8 is a flowchart of a representative method for assembling an MPEG-2 base layer and HEVC or AVC enhancement layer by a scalable HEVC or AVC decoder, according to an embodiment;

[0018] FIG. 9A is a block schematic illustrating a method of AVC encoding with a new NAL header, according to an embodiment;

[0019] FIG. 9B is a block schematic illustrating a method of decoding an AVC encoded stream with a new NAL header, according to an embodiment; and

[0020] FIG. 10 is a block schematic diagram of a computing machine that may be used to implement parts of the processes provided in the disclosure.

DETAILED DESCRIPTION

[0021] Turning to the drawings, wherein like reference numerals refer to like elements, techniques of the present disclosure are illustrated as being implemented in a suitable environment. The following description is based on embodiments of the claims and should not be taken as limiting the claims with regard to alternative embodiments that are not explicitly described herein.

[0022] As provided above, current HEVC scalable specification supports an application model where the base layer is AVC video with enhancements using scalable HEVC. In order to simplify the access unit (AU) structure and hypothetical reference decoder (HRD) model in scalable HEVC, there is a proposal to add an HEVC network abstraction layer (NAL) header to base layer AVC NAL units so that the bit stream looks

like an HEVC video stream. This function could also be extended to use of MPEG-2 video as the base layer for an AVC or HEVC enhancement, where the MPEG-2 video access unit is encapsulated in an AVC or HEVC NAL unit.

[0023] However, the transport of such scalable streams requires the (AVC or MPEG-2 video) base layer to be available to a legacy decoder without the special NAL header in the video stream. Provided herein are systems and methods that achieve this – for MPEG-2 video base layer and for AVC base layer. Additionally, the disclosed systems and methods may also be applicable to HEVC base layer with enhancements using scalable AVC.

[0024] FIG. 1 depicts a system context diagram illustrating a content distribution system (CDS) 100, according to an embodiment. In CDS 100, a content distribution facility, such as a headend 101, may be used to package and distribute protected content, such as video-based content or video content. The video content may be packaged and transmitted, for example, via transport stream (TS) 102, such as an MPEG-2 transport stream, AVC transport stream or HEVC transport stream. TS 102 may include a plurality of video and audio streams associated with different data formats and programs provided through the CDS 100 to various types of client premises equipment (CPE) and client devices such as, for example, a cell phone 103, a set-top box 104 and a computer 105. Other CPEs and client devices may include multimedia devices (e.g., digital cameras, personal digital assistants, mobile phones, color facsimile, printers, scanners, etc.).

[0025] Before reaching headend 101, or as part of headend 101, the video content is encoded. Referring now to FIG. 2, a video encoder 210 receives raw video data, from a source such as a camera. The video encoder utilizes the video data to generate a video elementary stream and supplies the video elementary stream to a video packetizer 214, which produces a video packetized elementary stream (PES) composed of variable length packets. Typically, each packet of the video PES contains one or more video frames. Similarly, an audio encoder (not shown) receives raw audio data from, for example, a microphone and supplies an audio elementary stream to an audio packetizer (not shown),

which creates an audio PES composed of variable length packets. For simplification, only video encoding/decoding is shown and audio encoding/decoding is not discussed further.

[0026] The video packetizer supplies the video PES to a transport stream multiplexer 218, which assigns program identifiers (PIDs) to the video PES and organizes the variable-length packets of the video PES as fixed-length transport stream (TS) packets having a header that includes the PID of the PES and a payload containing the PES video data. The length of PES packets are provided or indicated by a PES_packet_length field.

[0027] The single program transport stream (SPTS) that is output by the transport stream multiplexer may be supplied to a program multiplexer 222 that combines the SPTS with other transport streams, conveying other programs, to produce a multi-program transport stream (MPTS). The MPTS is transmitted over a channel to a receiver at which a program demultiplexer 226 separates a selected SPTS from the MPTS and supplies it to a transport stream demultiplexer 230. It will be appreciated by those skilled in the art that the SPTS that is output by the transport stream multiplexer may be transmitted directly to the transport stream demultiplexer without first being combined with other transport streams to create the MPTS but in either case the transport stream demultiplexer receives the transport stream packets of the selected SPTS and separates them on the basis of PID, depacketizes the transport stream packets to recreate the PES packets, and directs the video PES to a so-called video system target decoder (T-STD) 234.

[0028] The video T-STD 234 includes a system target decoder buffer 240 and a video decoder 242. The STD buffer 240 is functionally equivalent to a transport buffer Tb, a multiplexing buffer Mb, and an elementary stream buffer Eb. The transport buffer Tb receives the video PES at a variable bit rate and outputs the data at a constant bit rate to the multiplexing buffer Mb, which depacketizes the video PES and supplies an encoded bit stream at a constant bit rate to the elementary stream buffer Eb. The elementary

stream buffer, which is sometimes referred to as the decoder buffer or as the coded picture buffer (CPB), receives the CBR bitstream and holds the bits for decoding a picture until they are all removed instantaneously by the video decoder at the picture decode time.

[0029] The bitstream produced by the video encoder 210 may comply with the video compression standard that is specified in ISO/IEC 14496-10 (MPEG-4 part 10) Advanced Video Coding (AVC), commonly referred to as H.264/AVC. H.264/AVC uses picture as a collective term for a frame or field. H.264/AVC defines an access unit (AU) as a set of network abstraction layer (NAL) units and specifies that the decoding of an access unit always results in a decoded picture. A NAL unit of an AU produced by an AVC encoder may be a video coding layer (VCL) unit, which includes picture information, or a non-VCL unit, which includes other information, such as closed captioning and timing.

[0030] A decoder having the capability to decode both a base layer and one or more enhancement layers is referred to herein as a scalable video decoder, whereas a decoder that cannot recognize an enhancement layer and is able to decode only the base layer access units, and therefore does not have scalable decoding capability, is referred to herein as a legacy base layer decoder. The legacy base layer decoder may be configured to decode AVC, HEVC, or MPEG-2 base layer access units, for example. The scalable video decoder may be configured to decode AVC, HEVC, or MPEG-2 base layer access units and/or AVC and HEVC enhancement access units.

[0031] In one example, an access unit produced by a scalable video encoder comprises not only the base layer NAL units mentioned above, which may include AVC NAL units, but also HEVC NAL units. A scalable video decoder that extracts the base layer NAL units from the access unit selects only the AVC NAL units. The scalable video decoder that extracts the enhancement layer NAL units from the access units selects both the AVC NAL units and HEVC NAL units.

[0032] Referring to FIG. 3, a scalable video encoder 310 generates a unitary bitstream that conveys the base layer and, for example, two enhancement layers ENH1 and ENH2. Depending on its capabilities, a decoder might expect to receive, and decode, the base layer only, or the base layer and enhancement layer ENH1, or the base layer and both enhancement layer ENH1 and enhancement layer ENH2. Under an MPEG-2 systems standard and use case for SVC, the encoder may not provide three bitstreams, conveying respectively the base layer only, the base layer and enhancement layer ENH1, and the base layer and both enhancement layer ENH1 and enhancement layer ENH2 and allow the decoder to select whichever bitstream it is able to decode. The encoder must provide the base layer access units and parts of each enhancement layer in separate bitstreams. It would be possible in principle to comply with the MPEG-2 systems standard by using a NAL separator 348 to separate the unitary bitstream into three sub-bitstreams. One sub-bitstream would convey the base layer NAL units and the other two sub-bitstreams would convey the NAL units for the two enhancement layers respectively. The three sub-bitstreams would pass to respective video packetizers (generally designated 314), which create respective video PESs. The three video PESs would be supplied to a transport stream multiplexer 318 including a T-STD buffer equivalent to the buffer that is included in a scalable video T-STD, for the purpose of multiplexing together the outputs of the three packetizers. The multiplexer 318 would assign different PIDs to the three PESs and output a transport stream conveying the three layers.

[0033] The video T-STD 234 shown in FIG. 1 is unable to decode the bitstream conveyed by the transport stream produced by the transport stream multiplexer 318' shown in FIG. 3 because it provides no capability to reassemble the base layer and enhancement layer access units to produce a complete scalable video coded access unit that can be decoded by a scalable video decoder and a legacy decoder. Therefore, the architecture shown in FIG. 3 has hitherto lacked practical application.

[0034] FIG. 4 is a breakout frame of an MPEG-2 transport stream, illustrating encoding packets carried in the MPEG-2 transport stream, such as transport stream 102 of FIG. 1. MPEG-2 transport stream 402 may include a PES which are the video packets 470*a* to 470*d* and the program information packet 460, according to an embodiment. The program information packet (PIP) 460 includes control information to reference one or a plurality of video packets, such as video packets 470*a* to 470*d*. The program information packet 460 includes control information, such a program map table (PMT) 461 including a video descriptor 462 to reference the video packets 470*a* to 470*d*. The video descriptor 462 includes information describing what types of video data, such as video data 475*a*, may be included in a video packet, such as video packet 470*a*, and how the video data is used, such as use in a still picture, a slide show or as frames of video in a video sequence used in a program specified within PMT 461. In another embodiment, a PES may include only video packets which may be referenced by control information, such as the information in the video descriptor 462, which is not located in a program information packet associated with a PES in a transport stream but is instead located elsewhere in the transport stream including the PES. It should be appreciated that the video data 475*a* encoded in video packets 470*a* to 470*d* may comply with various industry compression standards, including AVC and HEVC.

[0035] The video packets in the PES illustrated in FIG. 4 include video packets 470*a* to 470*d*. Video packet 470*a* is representative of the other video packets 470*b* to 470*d*. It includes a NAL unit 471*a* and a video access unit 473*a*. The NAL unit 471*a* includes a NAL header 472*a*. As provided above, the NAL header 472*a* may be an HEVC NAL header so that the bit stream looks like an HEVC video stream.

[0036] The program information packet 460 includes control information for managing and/or using the video packets 470*a* to 470*d*. The program information packet 460 associated with video packet 470*a* includes a program map table (PMT) 461 which includes video descriptor 462. The information in video descriptor 462 may be used to direct usage of access unit 473*a* in the video packet 470*a*. The video

packets 470a to 470d may be included in a video PES which is multiplexed with other elementary streams and program streams in the MPEG-2 transport stream 402.

[0037] The video access unit 473a includes encoded video data 475a and an elementary stream (ELSM) header 474a. The video data 475a in the access unit 473a, before it is encoded, may be referred to as image data. The image data encoded in the video access unit 473a may be associated with a single frame or a still picture of image data. A video access unit, such as video access unit 473a, is a unit of encoded image data for a complete frame in a video or a still picture. The video access unit 473a also includes parameters in the ELSM header 474a which may be used to decode the access unit 473a and to display the decoded image data. The information in the video descriptor 462 and the ELSM header 474a both reference the video data 475a, but in different ways. In comparison with the information in the video descriptor 462, the ELSM header 474a in the access unit 473a contains relatively low level video-centric data for decoding and displaying the video data 475a as decoded image data. As noted above, the information in the video descriptor 462 is used to direct usage of the access unit 473a in the video packet 470a. This is a higher level referencing than in the ELSM header 472a.

[0038] The MPEG-2 transport stream 402 including video packets, such as video packets 470a to 470, may be used in client/server communications (e.g., the Internet, Image database, video streaming, video server, etc.) distributing cinema and for the exchange of content as well as the contribution of content to service providers. It may also be used in entertainment services and over broadband services utilizing satellites, cable system, 3DTV systems and IPTV systems.

[0039] As provided above, there is a proposal to add an HEVC network abstraction layer (NAL) header to base layer AVC NAL units so that the bit stream looks like an HEVC video stream. FIG. 5 is a flowchart of a representative method 500 for delivering only an AVC base layer to a legacy decoder, according to an embodiment.

[0040] In block 510, for each AVC access unit, an HEVC NAL unit is provided as part of the video packet. In some embodiments, an HEVC NAL header is placed proximate to the AVC NAL header (e.g., so that it is immediately followed by the AVC NAL header and AVC access unit). The AVC NAL header may be included as part of an ELSM header.

[0041] Referring briefly to FIG. 9A, a general schematic illustrating a method 900 of AVC encoding with a new NAL header is provided. In block 910, an AVC access unit (AU) is provided. At block 915, an additional NAL header (e.g., HEVC NAL header) is added to the AVC AU. At block 920, each AU with the additional NAL header is PES packetized, where the PES packet length includes all the PES header data. The PES header includes an element PES packet length indicating the length of the PES header, so that demultiplexers can remove the header and access the video data when decoding.

[0042] In some embodiments, a new HEVC NAL unit type is used to signal and carry an entire AVC access unit. This may be specified in MPEG-2 transport, where this includes AU delimiter, supplemental enhancement information (SEI), sequence parameter set (SPS), picture parameter set (PPS), pictures or video slices, etc.

[0043] In block 520, the AVC base layer stream including an HEVC NAL header for each AVC access unit, is encapsulated in PES. In the PES, the HEVC NAL unit header start code, followed by AVC access unit, is aligned to the PES header for each AVC access unit.

[0044] In some embodiments, during PES encapsulation, the PES_packet_length field in the PES header is incremented by 'n' bytes where n = number of bytes of new HEVC NAL unit header, e.g., including the byte stream start code. As provided herein, 'n' is usually less than 10 bytes and represents a constant value. This incrementing of the PES_packet_length field may be referred to as modifying the PES header.

[0045] In block 530, PID encapsulation is performed. The AVC base layer is carried in a first separate or dedicated PID with the PMT signaling AVC legacy stream_type of

0x1B. The scalable HEVC enhancements are carried in second dedicated PID with a stream_type to be allocated by MPEG.

[0046] In block 540, the transport streams are transmitted over a channel to a receiver at which a demultiplexer separates them on the basis of PID and depacketizes the transport stream packets to recreate the PES packets in block 560. In block 570, PES packets that have an AVC conformant elementary stream are directed to a video system target decoder (T-STD).

[0047] In block 580, the video T-STD performs normal T-STD management, including depacketizing the PES into an encoded bit stream, such as described with reference to FIG. 2. In block 590, a legacy AVC decoder removes the bit stream at the picture decode time.

[0048] Turning briefly to FIG. 9B, a general schematic illustrating a method 925 of decoding an AVC encoded stream with a new NAL header is provided. At block 930, the transport stream having packetized AUs with the additional NAL header is received. At block 935, a demultiplexer for the base layer depacketizes the transport stream packets to recreate the AVC access units in block 940. As is apparent from FIGS. 5 and 9B, only the AVC access units are provided to the legacy decoders because the new NAL headers has been treated as part of the modified PES, which have been removed by the demultiplexers.

[0049] FIG. 6 is a flowchart of a representative method 600 for assembling an AVC base layer and HEVC enhancement layer by a scalable HEVC decoder, according to an embodiment. In block 610, the PES_packet_length field in the PES header is decremented by 'n' bytes where n = number of bytes of new HEVC NAL unit header, e.g., including the byte stream start code. As provided herein, 'n' is usually less than 10 bytes and represents a constant value. This decrementing of the PES_packet_length field may be referred to as returning to the original PES header.

[0050] The transport streams are transmitted over a channel to a receiver at which a demultiplexer separates them on the basis of PID and depacketizes the transport stream packets to recreate the PES packets. In block 620, PES packets that have AVC conformant access units are directed to a block 630, where the AVC base layer access units with HEVC NAL header are provided.

[0051] Turning back to FIG. 9B, at block 945, a demultiplexer for the enhancement layer depacketizes the transport stream packets and decrements the PES header length to provide an original PES header in block 950. As shown in block 955, the AVC access units with the new HEVC NAL header is passed onto a scalable video decoder.

[0052] As provided in the discussion of FIG. 5, the scalable HEVC enhancements are carried in second dedicated PID with a stream_type to be allocated by MPEG. In block 640, the scalable HEVC enhancements are separated on the basis of PID. In block 650, PES packets that have scalable HEVC conformant access units are provided.

[0053] Thereafter, in block 660, the AVC base layer access units with HEVC NAL header from block 630 and the scalable HEVC conformant access units from block 650 are reassembled and directed to a T-STD that performs normal T-STD management, including depacketizing the PES into an encoded bit stream, such as described with reference to FIG. 2. In block 670, a scalable HEVC decoder removes the bit stream at the picture decode time.

[0054] While FIGS. 5 and 6 illustrate methods for encoding and decoding an AVC base layer with HEVC enhancements, similar methods are also contemplated for encoding and decoding an HEVC base layer with AVC enhancements. Additionally, these methods can be extended to use of MPEG-2 video as the base layer for an AVC or HEVC enhancement, where the MPEG-2 video access unit is encapsulated in an AVC or HEVC NAL unit. In such embodiments, the transport of such scalable streams requires the (AVC or MPEG-2 video) base layer to be available to a legacy decoder without the special or modified NAL header in the video stream.

[0055] FIG. 7 is a flowchart of a representative method 700 for delivering only MPEG-2 base layer to a legacy decoder, according to an embodiment. In block 710, for each MPEG-2 access unit, an AVC or HEVC NAL unit is provided as part of the video packet. For simplification, an HEVC NAL unit will be used as the enhancement layer, although it could be substituted with an AVC NAL unit. In some embodiments, an HEVC NAL header is placed proximate to the MPEG-2 NAL header (e.g., so that it is immediately followed by the MPEG-2 NAL header and MPEG-2 access unit). In some embodiments, the MPEG-2 video access unit includes a sequence header, GOP header, user data and picture header. The MPEG-2 NAL header may be included as part of an ELSM header.

[0056] In some embodiments, an HEVC NAL unit type may be used to signal the MPEG-2 video access unit that follows. This signaling may be specified in MPEG-2 transport, by use of a new stream_type for the HEVC enhancements.

[0057] In block 720, the MPEG-2 base layer stream including an HEVC NAL header for each MPEG-2 access unit, is encapsulated in PES. In the PES, the HEVC NAL unit header start code, followed by MPEG-2 access unit, is aligned to the PES header for each MPEG-2 access unit.

[0058] In some embodiments, during PES encapsulation, the PES_packet_length field in the PES header is incremented by 'n' bytes where n = number of bytes of new HEVC NAL unit header, e.g., including the byte stream start code. As provided herein, 'n' is usually less than 10 bytes and represents a constant value. This incrementing of the PES_packet_length field may be referred to as modifying the PES header.

[0059] In block 730, PID encapsulation is performed. The MPEG-2 base layer is carried in a first separate or dedicated PID with the PMT signaling AVC legacy stream_type of 0x02. The scalable HEVC enhancements are carried in second dedicated PID with a stream_type to be allocated by MPEG.

[0060] In block 740, the transport streams are transmitted over a channel to a receiver at which a demultiplexer separates them on the basis of PID and depacketizes the transport stream packets to recreate the PES packets in block 760. In block 770, PES packets that have an MPEG-2 conformant elementary stream are directed to a video system target decoder (T-STD).

[0061] In block 780, the video T-STD performs normal T-STD management, including depacketizing the PES into an encoded bit stream, such as described with reference to FIG. 2. In block 790, a legacy MPEG-2 decoder removes the bit stream at the picture decode time.

[0062] FIG. 8 is a flowchart of a representative method 800 for assembling an MPEG-2 base layer and HEVC or AVC enhancement layer by a scalable HEVC or AVC decoder, according to an embodiment. For simplification, an HEVC NAL unit will be used as the enhancement layer, although it could be substituted with an AVC NAL unit. In block 810, the PES_packet_length field in the PES header is decremented by 'n' bytes where n = number of bytes of new HEVC NAL unit header, e.g., including the byte stream start code. As provided herein, 'n' is usually less than 10 bytes and represents a constant value. This decrementing of the PES_packet_length field may be referred to as returning to the original PES header.

[0063] The transport streams are transmitted over a channel to a receiver at which a demultiplexer separates them on the basis of PID and depacketizes the transport stream packets to recreate the PES packets. In block 820, PES packets that have MPEG-2 conformant access units are directed to a block 830, where the MPEG-2 base layer access units with HEVC NAL header are provided.

[0064] As provided in the discussion of FIG. 7, the scalable HEVC enhancements are carried in second dedicated PID with a stream_type to be allocated by MPEG. In block 840, the scalable HEVC enhancements are separated on the basis of PID. In block 850, PES packets that have scalable HEVC conformant access units are provided.

[0065] Thereafter, in block 860, the MPEG-2 base layer access units with HEVC NAL header from block 830 and the scalable HEVC conformant access units from block 850 are reassembled and directed to a T-STD that performs normal T-STD management, including depacketizing the PES into an encoded bit stream, such as described with reference to FIG. 2. In block 870, a scalable HEVC decoder removes the bit stream at the picture decode time.

[0066] FIG. 10 is a diagram illustrating an example computer system 1000 that could be used to implement elements of the disclosure including, encoders, NAL separators, packetizers, demultiplexers, and decoders. The computer system 1000 includes a processing device such as a computer 1002, which comprises a general purpose hardware processor 1004A and/or a special purpose hardware processor 1004B (hereinafter alternatively collectively referred to as processor 1004) and a memory 1006, such as random access memory (RAM). The computer 1002 may be coupled to other devices, including input/output (I/O) devices such as a keyboard 1014, a mouse device 1016 and a printer 1028.

[0067] In one embodiment, the computer 1002 operates by the general purpose processor 1004A performing instructions defined by the computer program 1010 under control of an operating system 1008. The computer program 1010 and/or the operating system 1008 may be stored in the memory 1006 and may interface with the user and/or other devices to accept input and commands and, based on such input and commands and the instructions defined by the computer program 1010 and operating system 1008 to provide output and results.

[0068] In some embodiments, the computer program 1010 is composed of instructions such that when the computer 1002 receives a bitstream conveying the particular input signal involved, by way of an input adaptor, the computer 1010 allocates memory 1006 to appropriate buffers and utilizes other suitable resources and functions to perform the various operations that are described above with reference to FIGS. 1-9 and transmits, by way of an output adaptor, a bitstream conveying the particular output signal.

[0069] Output/results may be presented on the display 1022 or provided to another device for presentation or further processing or action. In one embodiment, the display 1022 comprises a liquid crystal display (LCD) having a plurality of separately addressable pixels formed by liquid crystals. Each pixel of the display 1022 changes to an opaque or translucent state to form a part of the image on the display in response to the data or information generated by the processor 1004 from the application of the instructions of the computer program 1010 and/or operating system 1008 to the input and commands. Other display 1022 types also include picture elements that change state in order to create the image presented on the display 1022. The image may be provided through a graphical user interface (GUI) module 1018A. Although the GUI module 1018A is depicted as a separate module, the instructions performing the GUI functions can be resident or distributed in the operating system 1008, the computer program 1010, or implemented with special purpose memory and processors.

[0070] Some or all of the operations performed by the computer 1002 according to the computer program 1010 instructions may be implemented in a special purpose processor 1004B. In this embodiment, some or all of the computer program 1010 instructions may be implemented via firmware instructions stored in a read only memory (ROM), a programmable read only memory (PROM) or flash memory within the special purpose processor 1004B or in memory 1006. The special purpose processor 1004B may also be hardwired through circuit design to perform some or all of the operations to implement embodiments of the present disclosure. Further, the special purpose processor 1004B may be a hybrid processor, which includes dedicated circuitry for performing a subset of functions, and other circuits for performing more general functions such as responding to computer program instructions. In one embodiment, the special purpose processor is an application specific integrated circuit (ASIC).

[0071] The computer 1002 may also implement a compiler 1012 which allows an application program 1010 written in a programming language such as COBOL, C++, FORTRAN, or other language to be translated into processor 1004 readable code. After

completion, the application or computer program 1010 accesses and manipulates data accepted from I/O devices and stored in the memory 1006 of the computer 1002 using the relationships and logic that was generated using the compiler 1012.

[0072] The computer 1002 also optionally comprises an external communication device such as a modem, satellite link, Ethernet card, or other device for accepting input from and providing output to other computers. In one embodiment, instructions implementing the operating system 1008, the computer program 1010, and/or the compiler 1012 are tangibly embodied in a computer-readable medium, e.g., data storage device, which could include an internal storage device 1020 or external storage device 1024 in the form of one or more fixed or removable data storage devices, such as a zip drive, floppy disc drive, hard drive, CD-ROM drive, tape drive, or a flash drive. Further, the operating system 1008 and the computer program 1010 are comprised of computer program instructions which, when accessed, read and executed by the computer 1002, causes the computer 1002 to perform the steps necessary to implement and/or use embodiments of the present disclosure or to load the program of instructions into a memory, thus creating a special purpose data structure causing the computer to operate as a specially programmed computer executing the method steps described herein. Computer program 1010 and/or operating instructions may also be tangibly embodied in memory 1006 and/or data communications devices 1030, thereby making a computer program product or article of manufacture according to the disclosure. As such, the terms “article of manufacture,” “program storage device” and “computer program product” or “computer readable storage device” as used herein are intended to encompass a computer program accessible from any computer readable device or media.

[0073] Of course, those skilled in the art will recognize that any combination of the above components, or any number of different components, peripherals, and other devices, may be used with the computer 1002.

[0074] Although the term “computer” is referred to herein, it is understood that the computer may include portable devices such as cellphones, portable MP3 players, video

game consoles, notebook computers, pocket computers, or any other device with suitable processing, communication, and input/output capability.

[0075] In view of the many possible embodiments to which the principles of the present discussion may be applied, it should be recognized that the embodiments described herein with respect to the drawing figures are meant to be illustrative only and should not be taken as limiting the scope of the claims. Therefore, the techniques as described herein contemplate all such embodiments as may come within the scope of the following claims and equivalents thereof.

CLAIMS

We claim:

1. A method of encoding video data in a packetized elementary stream (PES), the method comprising:

receiving a signal including frames of image data, wherein the frames include base layer and enhancement layer components;

encoding the frames of image data to form video access units (AUs), wherein base layer AUs are formed for base layer components and enhancement layers AUs are formed for enhancement layer components, each AU including:

an elementary stream header (ELSM) including image data metadata parameters associated with decoding and displaying the image data;

a PES header including a presentation time stamp (PTS) and a PES packet length indicating the length of the PES header;

one or more codestreams associated with a frame of the image data;

adding a network abstraction layer (NAL) unit header defined by an enhancement layer video standard to each base layer AU, the NAL unit header being a constant value and length; and

forming a modified PES header by increasing the PES packet length to include the length of the NAL unit header in it.

2. The method of claim 1, further comprising:

mapping the AUs into PES packets; and

ordering the PES packets to form a PES in a transport stream,

wherein the transport stream includes a base layer formed from base layer AUs and an enhancement layer formed from enhancement layer AUs.

3. The method of claim 1, wherein the base layer AUs comprise AVC or MPEG-2 video data.
4. The method of claim 1, wherein the enhancement layer AUs comprise HEVC or AVC video data.
5. The method of claim 1, wherein the base layer AUs are AVC video data and the NAL unit header is HEVC video data.
6. The method of claim 5, wherein a NAL unit header directly precedes each base layer AU.
7. The method of claim 5, wherein each HEVC NAL unit header is configured to signal and carry each AVC AU.
8. The method of claim 7, wherein each AVC AU includes at least one of: AU delimiter, SEI, SPS, PPS, and video slices.
9. The method of claim 5, further comprising:
 - mapping the AUs into PES packets; and
 - ordering the PES packets to form a PES in a transport stream,wherein the transport stream includes a base layer formed from base layer AUs and an enhancement layer formed from enhancement layer AUs,
 - wherein the transport stream includes at least two packet identifiers (PIDs) and a program map table (PMT), and
 - wherein the AVC base layer PES packets are carried in a first dedicated PID.
10. The method of claim 9, wherein the PMT signals the base layer AVC with MPEG defined stream type of 0x1B.

11. The method of claim 9, wherein the NAL unit header an HEVC enhancement layer and wherein the HEVC enhancement layer is carried in a second dedicated PID.
12. The method of claim 11, wherein the PMT signals the HVC enhancement layer with a new stream type value defined by MPEG.
13. The method of claim 1, wherein the base layer AUs are MPEG-2 video data and the NAL unit header is defined in AVC video data.
14. The method of claim 1, wherein the base layer AUs are MPEG-2 video data and the NAL unit header is defined in HEVC video data.
15. A method of decoding video data in a packetized elementary stream (PES), the method comprising:
 - receiving a transport stream including
 - the PES, wherein the PES includes frames on image data in video access units (AUs) including both base layer AUs and enhancement layer AUs, wherein base layer AUs are formed from base layer components and enhancement layer AUs are formed from enhancement layer components, each AU including:
 - an elementary stream header (ELSM) including image data metadata parameters associated with decoding and displaying the image data;
 - a network abstraction layer (NAL) unit header for each base layer AU;
 - a modified PES header including a presentation time stamp (PTS) and a PES packet length indicating the length of the PES header plus a constant value, wherein the constant value represents the length of the NAL unit header;

- one or more codestreams associated with a frame of the image data;
 - removing the modified PES header and NAL unit for each base layer AU; and
 - decoding each base layer AU to form a signal including frames of image data.
16. The method of claim 15, wherein the decoding is performed by a legacy decoder.
17. A method of decoding video data in a packetized elementary stream (PES), the method comprising:
- receiving a transport stream including
 - the PES, wherein the PES includes frames on image data in video access units (AUs) including both base layer AUs and enhancement layer AUs, wherein base layer AUs are formed from base layer components and enhancement layer AUs are formed from enhancement layer components, each AU including:
 - an elementary stream header (ELSM) including image data metadata parameters associated with decoding and displaying the image data;
 - a network abstraction layer (NAL) unit header for each base layer AU;
 - a modified PES header including a presentation time stamp (PTS) and a PES packet length indicating the length of the PES header plus a constant value, wherein the constant value represents the length of the NAL unit header;
 - one or more codestreams associated with a frame of the image data;
 - forming an original PES header by decreasing the modified PES packet length by the constant value that represents the NAL unit header length;
 - removing the original PES header for each base layer AU;
 - removing the original PES header for each enhancement layer AU; and

decoding each base layer AU and each enhancement layer AU to form a signal including frames of image data.

18. The method of claim 17, further comprising:

receiving a signal indicating that the modified PES header will be decremented by the constant value.

19. A system configured to encode video data in a packetized elementary stream (PES), the system comprising:

a processor configured to

receive a signal including frames of image data, wherein the frames include base layer and enhancement layer components;

encode the frames of image data to form video access units (AUs), wherein base layer AUs are formed for base layer components and enhancement layers AUs are formed for enhancement layer components, each AU including:

an elementary stream header (ELSM) including image data metadata parameters associated with decoding and displaying the image data;

a PES header including a presentation time stamp (PTS) and a PES packet length indicating the length of the PES header;

one or more codestreams associated with a frame of the image data;

add a network abstraction layer (NAL) unit header defined by an enhancement layer video standard to each base layer AU, the NAL unit header being a constant value and length;

form a modified PES header by increasing the PES packet length to include the length of the NAL unit header in it; and

a storage device configured to store

the frames of image data, and

executable instructions for controlling the processor.

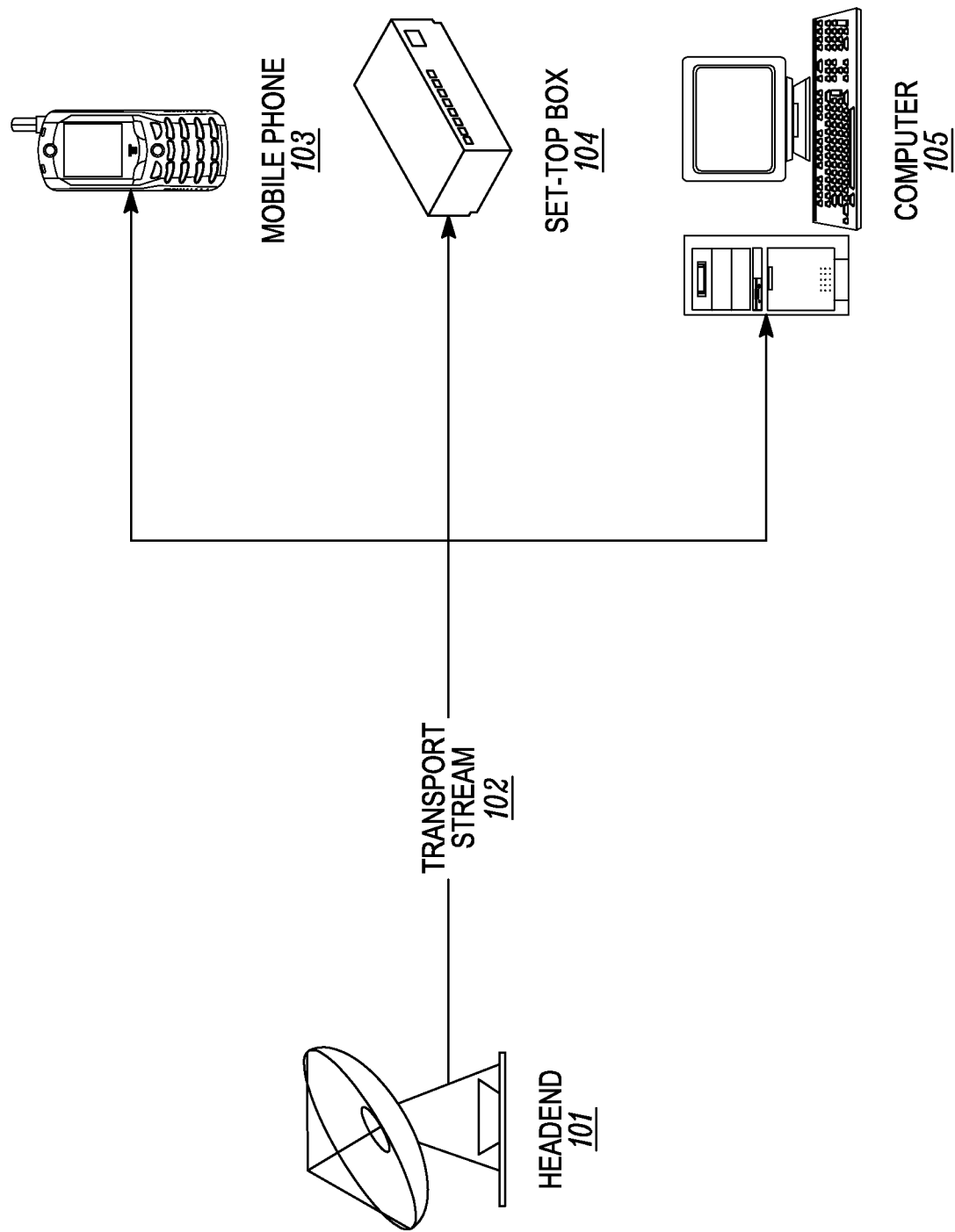


FIG. 1

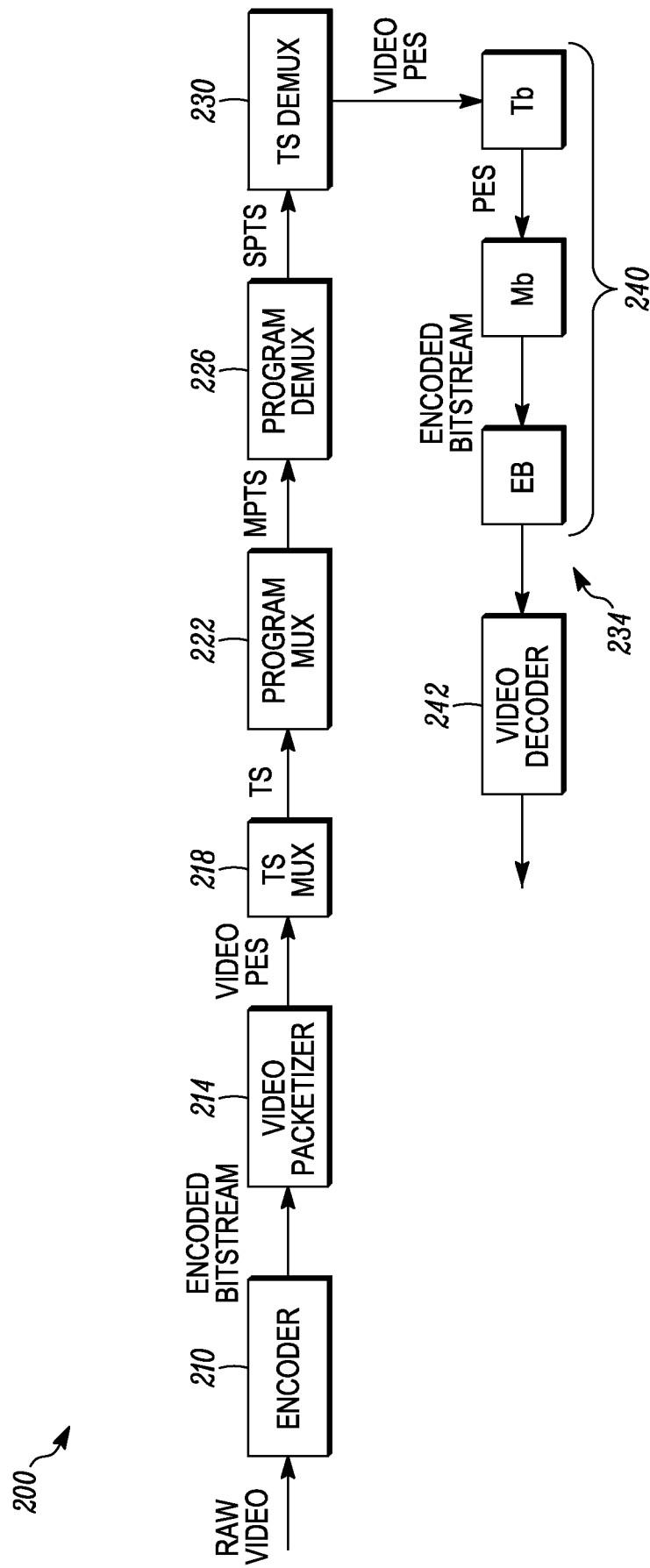


FIG. 2

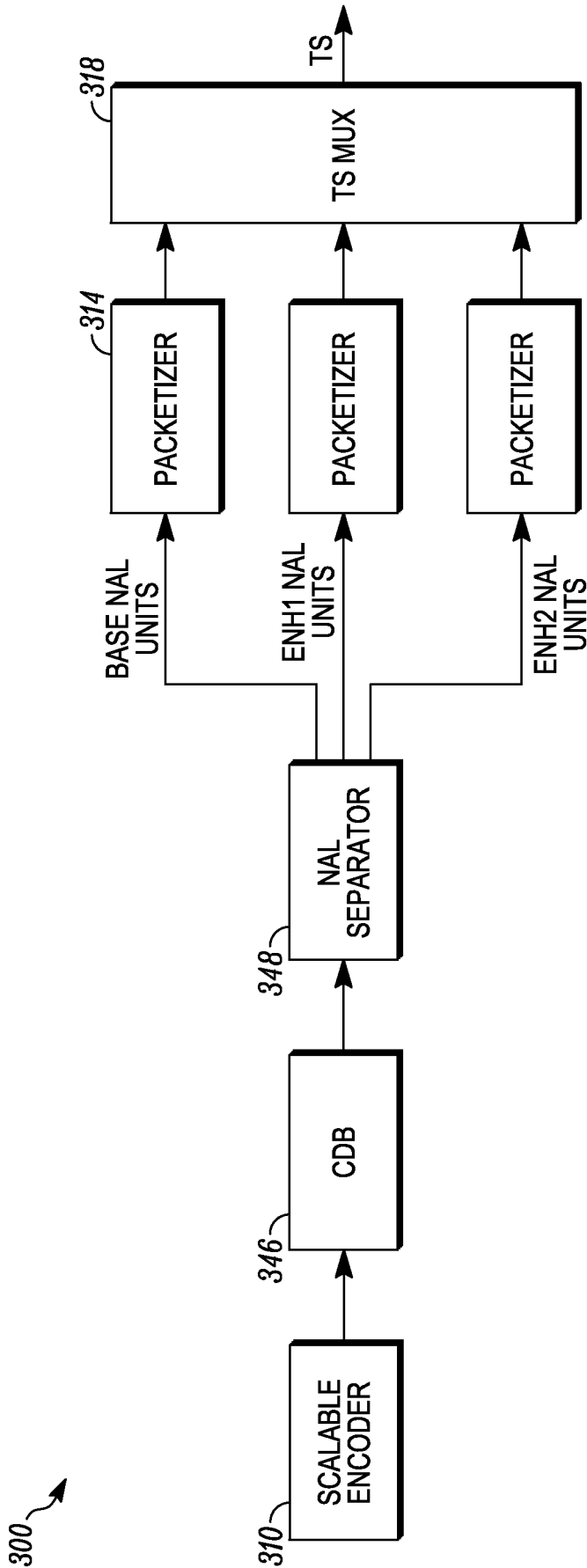


FIG. 3

MPEG-2 TRANSPORT STREAM CUT-OUT 150

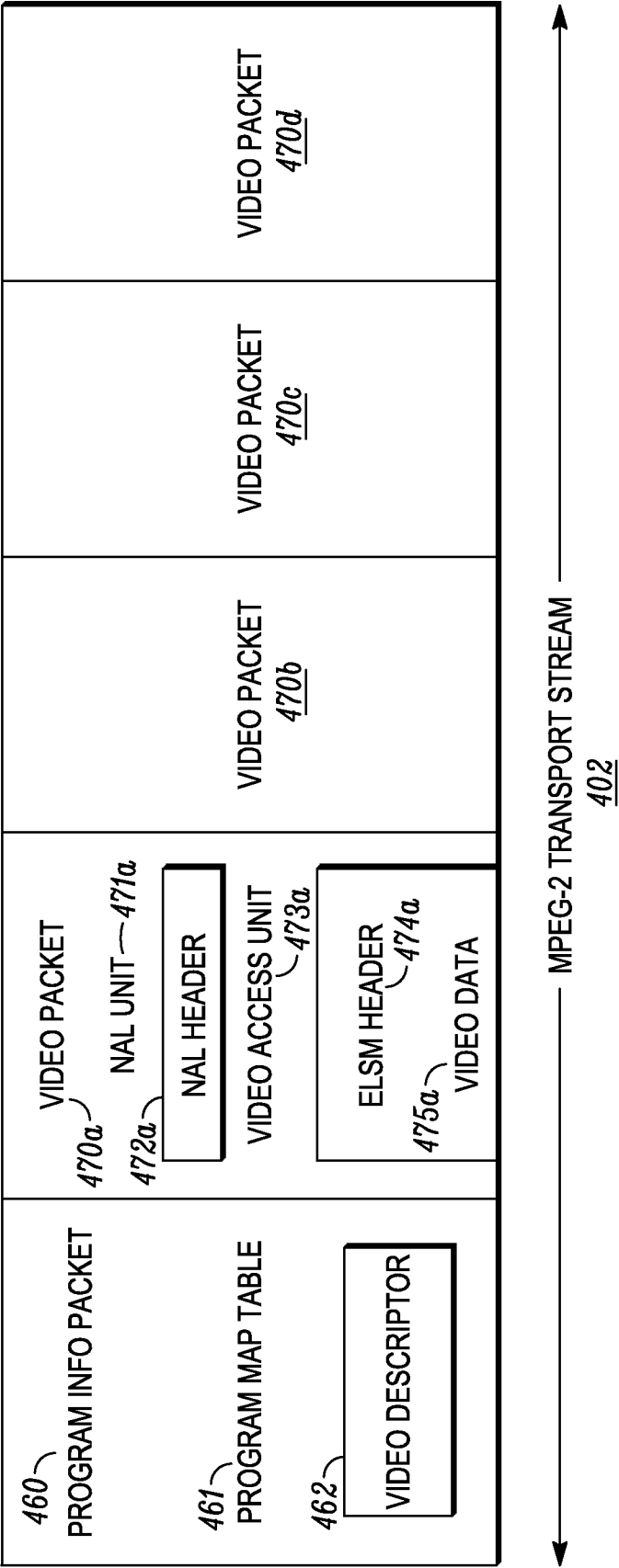


FIG. 4

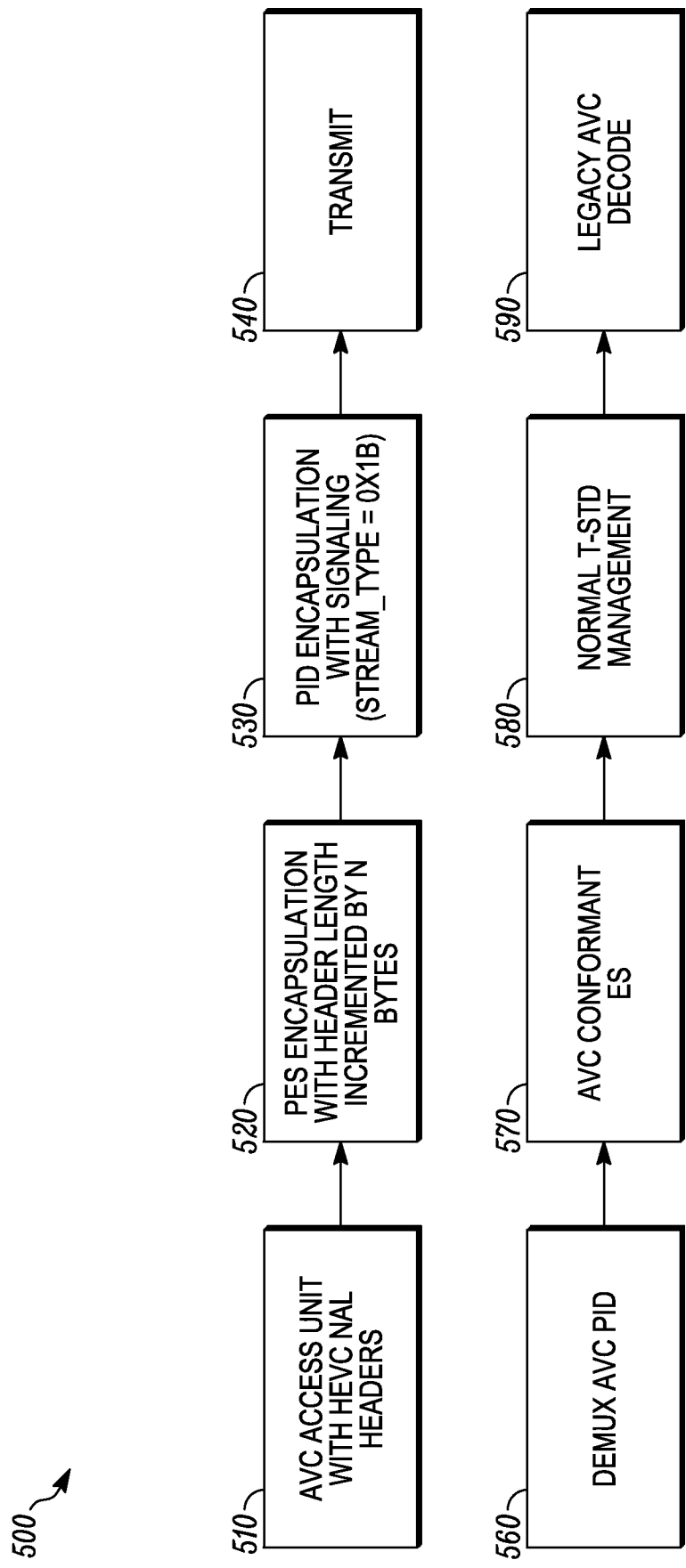


FIG. 5

600 ↗

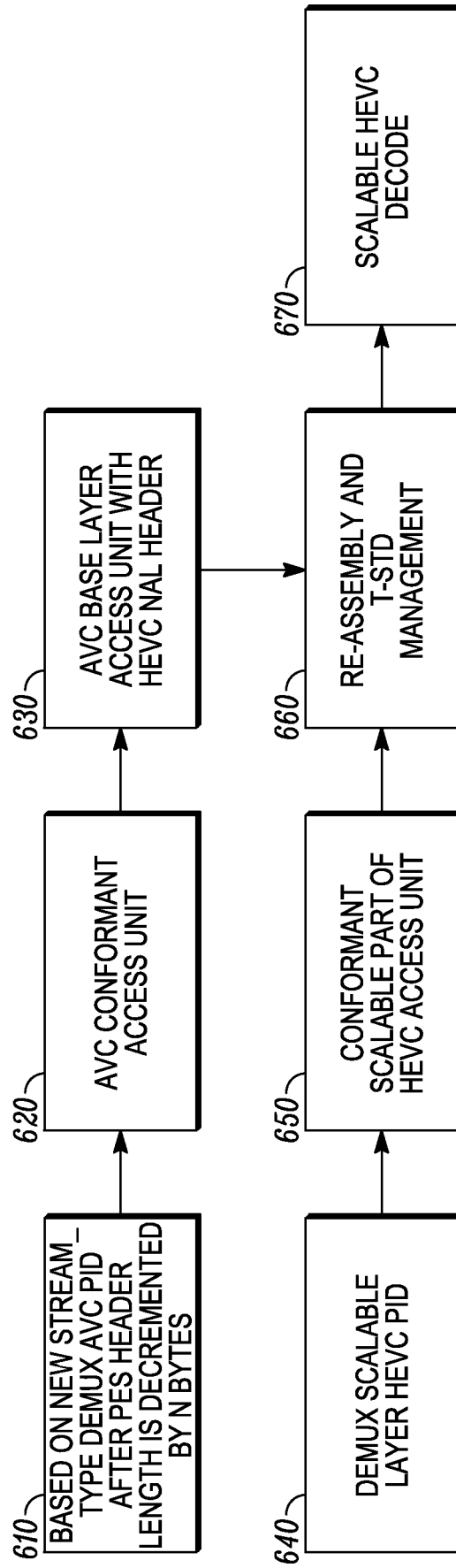


FIG. 6

700 ↗

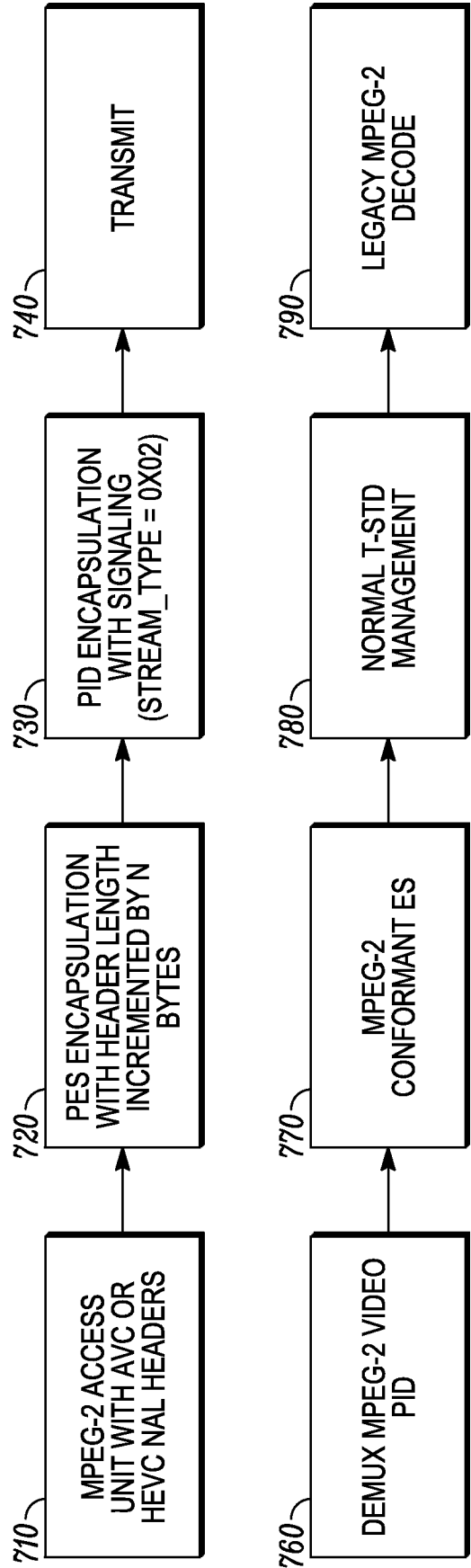


FIG. 7

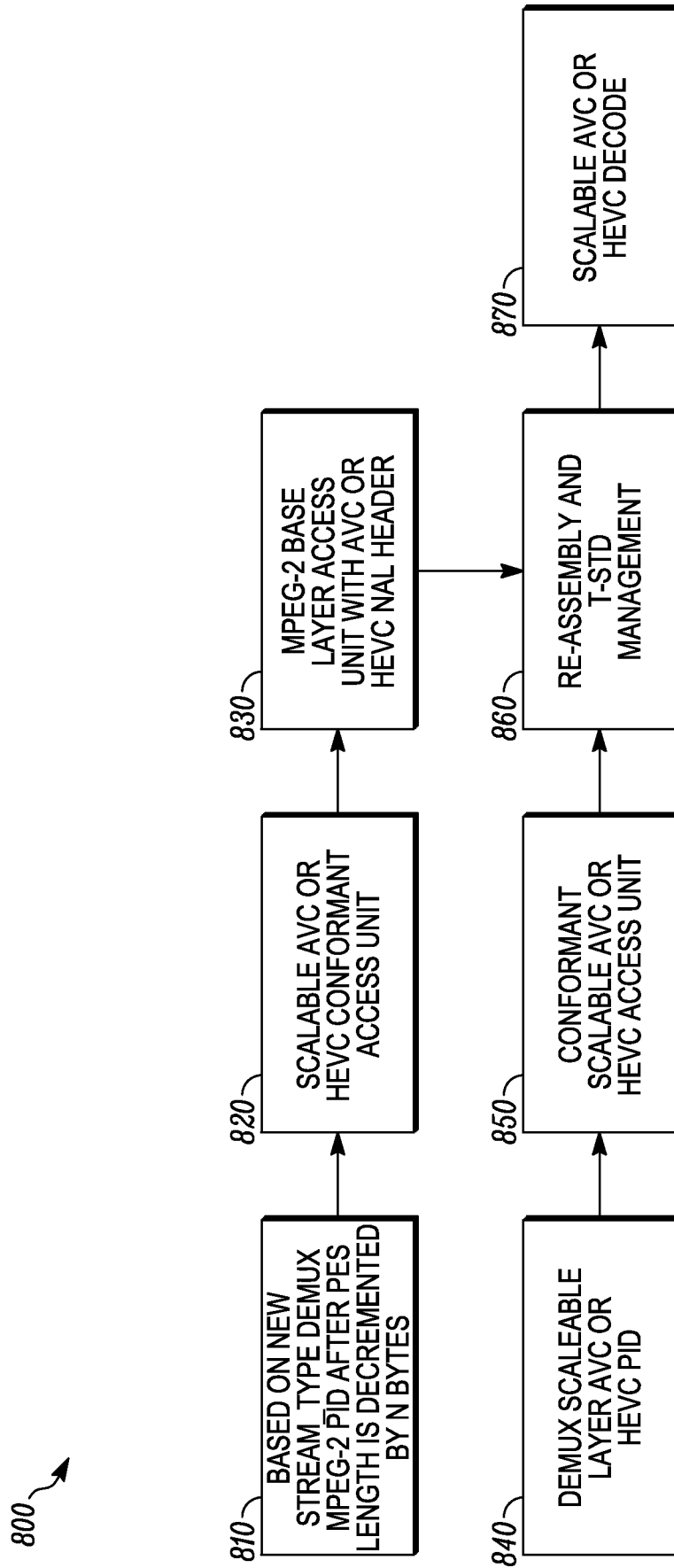


FIG. 8

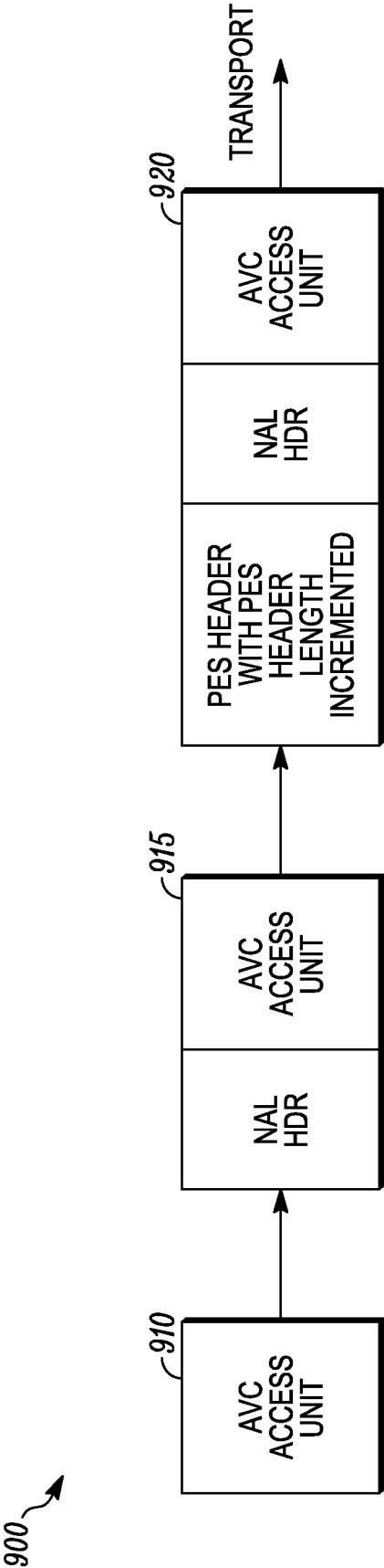


FIG. 9A

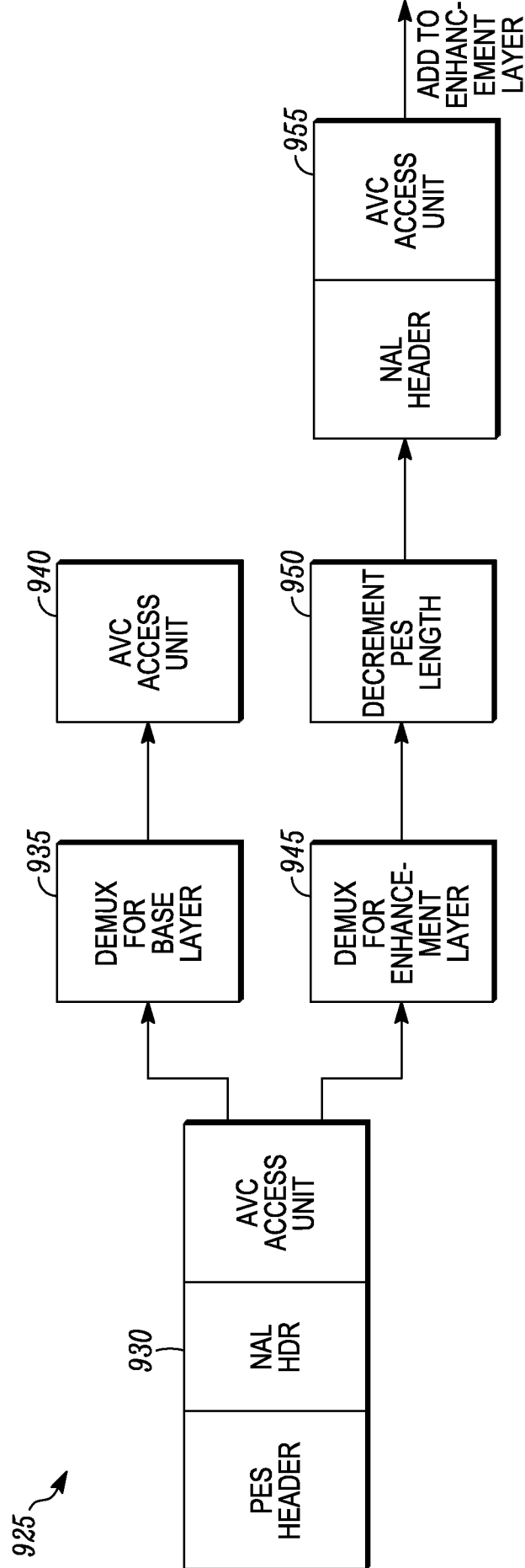
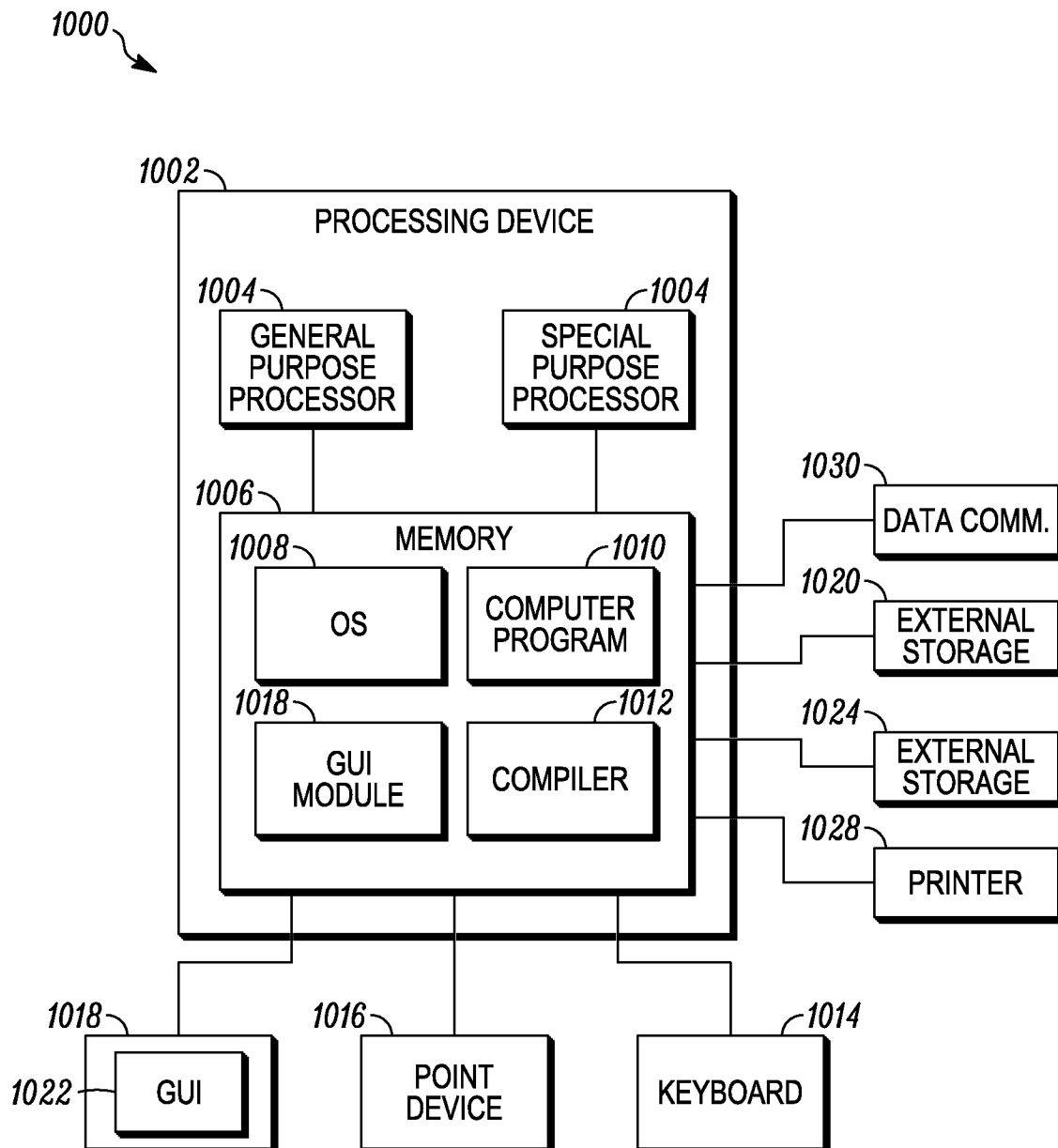


FIG. 9B

*FIG. 10*

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2014/061985

A. CLASSIFICATION OF SUBJECT MATTER

INV. H04N21/236 H04N19/30
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04N

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

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

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SAMUELSSON J ET AL: "AHG15: AVC and HEVC encapsulation for hybrid codec scalability", 15. JCT-VC MEETING; 23-10-2013 - 1-11-2013; GENEVA; (JOINT COLLABORATIVE TEAM ON VIDEO CODING OF ISO/IEC JTC1/SC29/WG11 AND ITU-T SG.16); URL: HTTP://WFTP3.ITU.INT/AV-ARCH/JCTVC-SITE/,, no. JCTVC-00190, 15 October 2013 (2013-10-15), XP030115229, the whole document ----- -/--	1-19

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

14 January 2015

Date of mailing of the international search report

22/01/2015

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Belardinelli, Carlo

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2014/061985

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CHOI B ET AL: "Multi-standard extension design", 12. JCT-VC MEETING; 103. MPEG MEETING; 14-1-2013 - 23-1-2013; GENEVA; (JOINT COLLABORATIVE TEAM ON VIDEO CODING OF ISO/IEC JTC1/SC29/WG11 AND ITU-T SG.16); URL: HTTP://WFTP3.ITU.INT/AV-ARCH/JCTVC-SITE/,, no. JCTVC-L0131, 8 January 2013 (2013-01-08), XP030113619, the whole document -----	1-19
A	BOYCE (VIDYO) J: "Specification text to support AVC base layer in HEVC layered extensions", 5. JCT-3V MEETING; 27-7-2013 - 2-8-2013; VIENNA; (THE JOINT COLLABORATIVE TEAM ON 3D VIDEO CODING EXTENSION DEVELOPMENT OF ISO/IEC JTC1/SC29/WG11 AND ITU-T SG.16); URL: HTTP://PHENIX.INT-EVRY.FR/JCT2/,, no. JCT3V-E0037, 17 July 2013 (2013-07-17), XP030131036, the whole document -----	1-19
A	EP 2 194 717 A2 (KOREA ELECTRONICS TELECOMM [KR]; IND ACADEMIC COOP [KR]) 9 June 2010 (2010-06-09) paragraph [0003] - paragraph [0034]; figures 3, 4 -----	1-19
A	BENJAMIN BROSS ET AL: "High Efficiency Video Coding (HEVC) text specification draft 10 (for FDIS & Last Call), JCTVC-L1003_v32", JOINT COLLABORATIVE TEAM ON VIDEO CODING (JCT-VC) OF ITU-T SG 16 WP 3 AND ISO/IEC JTC 1/SC 29/WG 11 12TH MEETING, vol. JCTVC-L1003_v32, 16 March 2013 (2013-03-16), XP055140558, paragraph [07.3] - paragraph [07.4] Annex B -----	1-19
A	Anonymous: "Packetized elementary stream - Wikipedia, the free encyclopedia", Wikipedia, 3 April 2012 (2012-04-03), pages 1-4, XP055161849, Retrieved from the Internet: URL:http://web.archive.org/web/20120403092609/http://en.wikipedia.org/wiki/Packetized_Elementary_Stream [retrieved on 2015-01-13] the whole document ----- -/--	1,15,17, 19

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2014/061985

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Anonymous: "Elementary stream - Wikipedia, the free encyclopedia", Wikipedia, 15 August 2013 (2013-08-15), pages 1-3, XP055161850, Retrieved from the Internet: URL:http://web.archive.org/web/20130815170903/http://en.wikipedia.org/wiki/Elementary_stream [retrieved on 2015-01-13] the whole document -----	1,15,17, 19

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2014/061985

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 2194717 A2	09-06-2010	EP 2194717 A2	09-06-2010
		US 2010142625 A1	10-06-2010
