

(51) International Patent Classification:
H04N 7/26 (2006.01)(21) International Application Number:
PCT/US2013/065754(22) International Filing Date:
18 October 2013 (18.10.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/717,097 22 October 2012 (22.10.2012) US
14/027,028 13 September 2013 (13.09.2013) US(71) Applicant: **MICROSOFT CORPORATION** [US/US];
One Microsoft Way, Redmond, WA 98052-6399 (US).(72) Inventors: **SULLIVAN, Gary, J.**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, WA 98052-6399 (US). **MALVAR, Henrique, Sarmiento**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, WA 98052-6399 (US). **WU, Yongjun**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, WA 98052-6399 (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, LT, 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, 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).

[Continued on next page]

(54) Title: BAND SEPARATION FILTERING / INVERSE FILTERING FOR FRAME PACKING / UNPACKING HIGHER-RESOLUTION CHROMA SAMPLING FORMATS

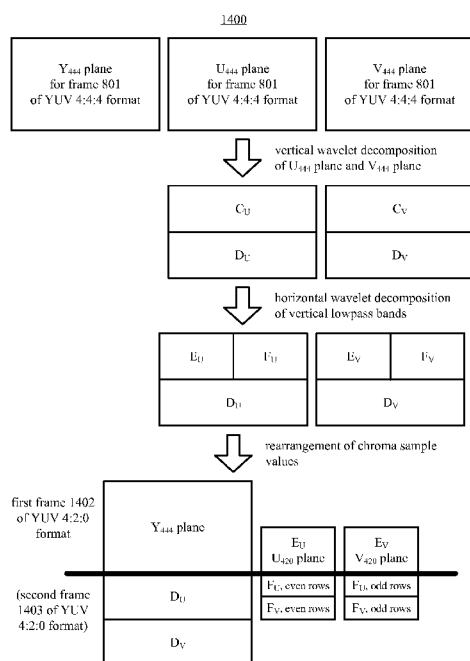


Figure 14

(57) Abstract: When packing a video frame of a higher-resolution chroma sampling format such as YUV 4:4:4 into frames of a lower-resolution chroma sampling format such as YUV 4:2:0, a computing device performs wavelet decomposition (or other band separation filtering) on sample values of chroma components of the higher-resolution frame, producing sample values of multiple bands. The device assigns the sample values of the bands to parts of the lower-resolution frames. During corresponding unpacking operations, a computing device assigns parts of the frames of the lower-resolution chroma sampling format to sample values of multiple bands. The device performs wavelet reconstruction (or other inverse band separation filtering) on the sample values of the bands, producing sample values of chroma components of the frame of the higher-resolution chroma sampling format. Band separation filtering can help improve quality of reconstruction when distortion has been introduced during encoding of the chroma components packed into low-resolution frames.

**Declarations under Rule 4.17:****Published:**

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
 - *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*
- *with international search report (Art. 21(3))*

BAND SEPARATION FILTERING / INVERSE FILTERING FOR FRAME PACKING / UNPACKING HIGHER-RESOLUTION CHROMA SAMPLING FORMATS

5 BACKGROUND

[001] Engineers use compression (also called source coding or source encoding) to reduce the bit rate of digital video. Compression decreases the cost of storing and transmitting video information by converting the information into a lower bit rate form. Decompression (also called decoding) reconstructs a version of the original information from the compressed form. A “codec” is an encoder/decoder system.

[002] Over the last two decades, various video codec standards have been adopted, including the ITU-T H.261, H.262 (MPEG-2 or ISO/IEC 13818-2), H.263 and H.264 (MPEG-4 AVC or ISO/IEC 14496-10) standards and the MPEG-1 (ISO/IEC 11172-2), MPEG-4 Visual (ISO/IEC 14496-2) and SMPTE 421M standards. More recently, the HEVC standard (ITU-T H.265 or ISO/IEC 23008-2) has been under development. *See, e.g.,* draft version JCTVC-I1003 of the HEVC standard – “High efficiency video coding (HEVC) text specification draft 7,” JCTVC-I1003_d5, 9th meeting, Geneva, April 2012.

A video codec standard typically defines options for the syntax of an encoded video bitstream, detailing parameters in the bitstream when particular features are used in encoding and decoding. In many cases, a video codec standard also provides details about the decoding operations a decoder should perform to achieve conformant results in decoding. Aside from codec standards, various proprietary codec formats define other options for the syntax of an encoded video bitstream and corresponding decoding operations.

[003] A video source such as a camera, animation output, screen capture module, etc. typically provides video that is converted to a format such as a YUV 4:4:4 chroma sampling format. A YUV format includes a luma (or Y) component with sample values representing approximate brightness values as well as multiple chroma (or U and V) components with sample values representing color difference values. In a YUV 4:4:4 format, chroma information is represented at the same spatial resolution as luma information.

[004] Many commercially available video encoders and decoders support only a YUV 4:2:0 chroma sampling format. A YUV 4:2:0 format is a format that sub-samples chroma information compared to a YUV 4:4:4 format, so that chroma resolution is half that of

luma resolution both horizontally and vertically. As a design principle, the decision to use a YUV 4:2:0 format for encoding/decoding is premised on the understanding that, for most use cases such as encoding/decoding of natural camera-captured video content, viewers do not ordinarily notice many visual differences between video encoded/decoded in a YUV 4:2:0 format and video encoded/decoded in a YUV 4:4:4 format. The compression advantages for the YUV 4:2:0 format, which has fewer samples per frame, are therefore compelling. There are some use cases, however, for which video has richer color information and higher color fidelity may be justified. In such use cases, the differences between YUV 4:4:4 and YUV 4:2:0 chroma sampling formats are more easily perceived by viewers. For example, for encoding/decoding of computer screen text content, animated video content with artificial hard-edged boundaries, or certain features of video content more generally (such as scrolling titles and hard-edged graphics, or video with information concentrated in chroma channels), a 4:4:4 format may be preferable to a 4:2:0 format. Although screen capture codecs supporting encoding and decoding in a 4:4:4 format are available, the lack of widespread support for codecs supporting 4:4:4 formats (especially with respect to hardware codec implementations) is a hindrance for these use cases.

SUMMARY

[005] In summary, the detailed description presents innovations in frame packing of video frames of a higher-resolution chroma sampling format into video frames of a lower-resolution chroma sampling format for purposes of encoding. For example, the higher-resolution chroma sampling format is a YUV 4:4:4 format, and the lower-resolution chroma sampling format is a YUV 4:2:0 format. After decoding, the video frames of the lower-resolution chroma sampling format can be unpacked to reconstruct the video frames of the higher-resolution chroma sampling format. In this way, available encoders and decoders operating at the lower-resolution chroma sampling format can be used, while still retaining higher resolution chroma information.

[006] For example, a computing device packs one or more frames of a higher-resolution chroma sampling format into one or more frames of a lower-resolution chroma sampling format. As part of the packing, the computing device performs wavelet decomposition (or other band separation filtering) on sample values of chroma components of the frame(s) of the higher-resolution chroma sampling format to produce sample values of multiple bands. The computing device then assigns the sample values of

the multiple bands to parts of the frame(s) of the lower-resolution chroma sampling format.

[007] As another example, a computing device unpacks one or more frames of a lower-resolution chroma sampling format into one or more frames of a higher-resolution chroma sampling format. As part of the unpacking, the computing device assigns parts of the frame(s) of the lower-resolution chroma sampling format to sample values of multiple bands. The computing device then performs wavelet reconstruction (or other inverse band separation filtering) on the sample values of the multiple bands to produce sample values of chroma components of the frame(s) of the higher-resolution chroma sampling format.

[008] The packing or unpacking can be implemented as part of a method, as part of a computing device adapted to perform the method or as part of a tangible computer-readable media storing computer-executable instructions for causing a computing device to perform the method.

[009] The foregoing and other objects, features, and advantages of the invention will become more apparent from the following detailed description, which proceeds with reference to the accompanying figures.

BRIEF DESCRIPTION OF THE DRAWINGS

[010] Figure 1 is a diagram of an example computing system in which some described embodiments can be implemented.

[011] Figures 2a and 2b are diagrams of example network environments in which some described embodiments can be implemented.

[012] Figure 3 is a diagram of a generalized frame packing/unpacking system in which some described embodiments can be implemented

[013] Figure 4 is a diagram of an example encoder system in conjunction with which some described embodiments can be implemented.

[014] Figure 5 is a diagram of an example decoder system in conjunction with which some described embodiments can be implemented.

[015] Figure 6 is a diagram illustrating an example video encoder in conjunction with which some described embodiments can be implemented.

[016] Figure 7 is a diagram illustrating an example video decoder in conjunction with which some described embodiments can be implemented.

[017] Figure 8 is a diagram illustrating an example approach to frame packing that uses spatial partitioning of frames.

[018] Figure 9 is a diagram illustrating an example approach to frame packing in which every second row of chroma component planes of frames of a higher-resolution chroma sampling format is copied.

5 [019] Figure 10 is a diagram illustrating example frames packed according to the approach of Figure 9.

[020] Figure 11 is a diagram illustrating an example approach to frame packing in which every second column of chroma component planes of frames of a higher-resolution chroma sampling format is copied.

10 [021] Figure 12 is a flow chart illustrating a generalized technique for frame packing for frames of a higher-resolution chroma sampling format.

[022] Figure 13 is a flow chart illustrating a generalized technique for frame unpacking for frames of a higher-resolution chroma sampling format.

15 [023] Figure 14 is a diagram illustrating an example approach to frame packing with vertical then horizontal filtering in three-band wavelet decomposition of chroma component planes of frames of a higher-resolution chroma sampling format.

[024] Figure 15 is a diagram illustrating an example approach to frame packing with horizontal then vertical filtering in three-band wavelet decomposition of chroma component planes of frames of a higher-resolution chroma sampling format.

20 [025] Figures 16a and 16b are diagrams illustrating example approaches to frame packing with vertical then horizontal then vertical filtering in four-band wavelet decomposition of chroma component planes of frames of a higher-resolution chroma sampling format.

25 [026] Figure 17 is a flow chart illustrating a generalized technique for frame packing for frames of a higher-resolution chroma sampling format, where the frame packing includes wavelet decomposition or other band separation filtering.

[027] Figure 18 is a flow chart illustrating a generalized technique for frame unpacking for frames of a higher-resolution chroma sampling format, where the frame packing includes wavelet reconstruction or other inverse band separation filtering.

DETAILED DESCRIPTION

30 [028] A video source such as a camera, animation output, screen capture module, etc. typically provides video that is converted to a format such as a YUV 4:4:4 chroma sampling format (an example of a 4:4:4 format, more generally). A YUV format includes a luma (or Y) component with sample values representing approximate brightness values as well as multiple chroma (or U and V) components with sample values representing

color difference values. The precise definitions of the color difference values (and conversion operations to/from a YUV color space to another color space such as RGB) depend on implementation. In general, as used herein, the term YUV indicates any color space with a luma (or luminance) component and one or more chroma (or chrominance) components, including Y'UV, YIQ, Y'IQ and YDbDr as well as variations such as YCbCr and YCoCg. The component signal measures that are used may be adjusted through the application of a non-linear transfer characteristics function (generally known as “gamma pre-compensation” and often denoted by the use of a prime symbol, although the prime symbol is often omitted for typographical convenience). Or, the component signal measures may be in a domain that has a linear relationship with light amplitude. The luma and chroma component signals may be well aligned with the perception of brightness and color for the human visual system, or the luma and chroma component signals may somewhat deviate from such measures (*e.g.*, as in the YCoCg variation, in which formulas are applied that simplify the computation of the color component values). Examples of YUV formats as described herein include those described in the international standards known as ITU-R BT.601, ITU-R BT.709, and ITU-R BT.2020. Examples of chroma sample types are shown in Figure E-1 of the H.264/AVC standard. A 4:4:4 format can be a YUV 4:4:4 format or format for another color space, such as RGB or GBR.

[029] Many commercially available video encoders and decoders support only a YUV 4:2:0 chroma sampling format (an example of a 4:2:0 format, more generally). YUV 4:2:0 is a format that sub-samples chroma information compared to a YUV 4:4:4 format, which preserves full-resolution chroma information (that is, chroma information is represented at the same resolution as luma information). As a design principle, the decision to use a YUV 4:2:0 format for encoding/ decoding is premised on the understanding that, for most use cases such as encoding/decoding of natural camera-captured video content, viewers do not ordinarily notice many visual differences between video encoded/decoded in a YUV 4:2:0 format and video encoded/decoded in a YUV 4:4:4 format. The compression advantages for the YUV 4:2:0 format, which has fewer samples per frame, are therefore compelling.

[030] There are some use cases, however, for which the differences between the two formats are more easily perceived by viewers. For example, for encoding/decoding of computer screen text content (especially text rendered using ClearType technology, animated video content with artificial hard-edged boundaries, or certain features of video content more generally (such as scrolling titles and hard-edged graphics, or video with

information concentrated in chroma channels), a 4:4:4 format may be preferable to a 4:2:0 format. The lack of widespread support for video codecs supporting 4:4:4 formats (especially with respect to hardware codec implementations) is a hindrance for these use cases.

5 **[031]** The detailed description presents various approaches to wavelet decomposition (or other band separation filtering) for packing frames of a higher-resolution chroma sampling format into frames of a lower-resolution chroma sampling format. The frames of the lower-resolution chroma sampling format can then be encoded using an encoder designed for the lower-resolution chroma sampling format. After decoding (using a
10 decoder designed for the lower-resolution chroma sampling format), the frames at the lower-resolution chroma sampling format can be output for further processing and display. Or, after such decoding, the frames of the higher-resolution chroma sampling format can be recovered through frame unpacking that includes wavelet reconstruction (or other inverse band separation filtering), for output and display. In many cases, these approaches
15 alleviate the shortcomings of the prior approaches by preserving chroma information from the frames in the higher-resolution chroma sampling format, while leveraging commercially available codecs adapted for the lower-resolution chroma sampling format. In particular, widely available codecs with specialized, dedicated hardware can provide faster encoding/decoding with lower power consumption for YUV 4:4:4 video frames
20 packed into YUV 4:2:0 video frames.

[032] The described approaches can be used to preserve chroma information for frames of one chroma sampling format when encoding/decoding uses another chroma sampling format. Some examples described herein involve frame packing/unpacking of frames of a YUV 4:4:4 format for encoding/decoding using a codec adapted for a YUV 4:2:0 format.

25 Other examples described herein involve frame packing/unpacking of frames of a YUV 4:2:2 format for encoding/decoding using a codec adapted for a YUV 4:2:0 format. More generally, the described approaches can be used for other chroma sampling formats. For example, in addition to variations of YUV color spaces such as Y'UV, YIQ, Y'IQ, YdbDr, YCbCr, YCoCg, etc. in sampling ratios such as 4:4:4, 4:2:2, 4:2:0, 4:1:1, 4:0:0, etc., the
30 described approaches can be used for color spaces such as RGB, GBR, etc. in sampling ratios such as 4:4:4, 4:2:2, 4:2:0, 4:1:1, 4:0:0, etc. as the chroma sampling formats.

[033] In example implementations, specific aspects of the innovations described herein include, but are not limited to, the following:

- Packing a 4:4:4 frame into two 4:2:0 frames, where the packing includes wavelet decomposition (or other band separation filtering), and encoding the two 4:2:0 frames using a video encoder designed for 4:2:0 format.
- 5 • Decoding the encoded frames using a video decoder designed for 4:2:0 format, and unpacking the two decoded 4:2:0 frames to form a decoded 4:4:4 frame, where the unpacking includes wavelet reconstruction (or other inverse band separation filtering).
- 10 • Performing the packing (including wavelet decomposition or other band separation filtering) for a YUV format such that a geometric correspondence is maintained between Y, U and V components for each of the two 4:2:0 frames.
- 15 • Performing the packing (including wavelet decomposition or other band separation filtering) for a YUV format such that one of the two 4:2:0 frames (a main view) represents the complete scene being represented by 4:4:4 frame, albeit with chroma components at a lower resolution, while the other 4:2:0 frame (an auxiliary view) packs the remaining chroma information.
- 20 • Signaling an indication of use wavelet decomposition (or other band separation filtering) in frame packing for the two 4:2:0 frames with a type of supplemental enhancement information (“SEI”) message or other metadata, such that a decoder that processes this SEI message can output the 4:4:4 frame (after wavelet reconstruction or other inverse band separation filtering) or the 4:2:0 frame that represents the scene.
- 25 • Pre-processing and post-processing operations (such as wavelet decomposition/reconstruction or other band separation filtering/inverse filtering) that can improve the quality of the final displayed frame for a YUV format when only one 4:2:0 frame (out of the two 4:2:0 frames) is used for final display. In conjunction with such pre-processing and post-processing operations, the 4:2:0 frames can have a higher bit depth for encoding/decoding, so as to avoid loss of chroma information in pre-processing and post-processing operations.
- 30 • Packing a 4:2:2 frame into (two or less) 4:2:0 frames, where the packing includes wavelet decomposition or other band separation filtering, and encoding the 4:2:0 frames using a video encoder designed for 4:2:0 format.
- Decoding the encoded frames using a video decoder designed for 4:2:0 format, and unpacking the decoded 4:2:0 frames to form a decoded 4:2:2 frame, where the

unpacking includes wavelet reconstruction or other inverse band separation filtering.

[034] In specific example implementations that use frame packing arrangement SEI messages, the definition of frame packing arrangement SEI message is extended to support representing 4:4:4 content in a nominally 4:2:0 bitstream. In some examples, one constituent frame (*e.g.*, in a top-bottom packing or alternating-frame coding scheme) can be decoded compatibly as an ordinary 4:2:0 image, or can be supplemented with the data from another constituent frame to form a complete 4:4:4 image representation. Since YUV 4:2:0 is the most widely supported format in products (especially with respect to hardware codec implementations), having an effective way of conveying YUV 4:4:4 content through such decoders can provide the substantial benefit of enabling widespread near-term deployment of YUV 4:4:4 capabilities (especially for screen content coding). In example implementations, the samples of a 4:4:4 frame are packed into two 4:2:0 frames, and the two 4:2:0 frames are encoded as the constituent frames of a frame packing arrangement. For implementations that use the frame packing arrangement SEI message, the semantics of the `content_interpretation_type` syntax element are extended to signal this usage. The `content_interpretation_type` syntax element signals how to interpret the data that are packed using a packing arrangement, and the frame configuration for the packing arrangement is signaled with a different syntax element. Some approaches described herein have high practical value for applications involving screen content. Also, relative to native 4:4:4 encoding, some approaches described herein can provide the advantage of compatibility with the ordinary 4:2:0 decoding process that is expected to be more widely supported in decoding products.

[035] Additional innovative aspects of frame packing and unpacking for higher-resolution chroma sampling formats are also described. The described techniques may be applied to additional applications other than video coding/decoding, such as still-image coding, medical scan content coding, multispectral imagery content coding, etc. Although operations described herein are in places described as being performed by an encoder (*e.g.*, video encoder) or decoder (*e.g.*, video decoder), in many cases the operations can alternatively be performed by another type of media processing tool.

[036] Some of the innovations described herein are illustrated with reference to syntax elements and operations specific to the HEVC standard. For example, reference is made to the draft version JCTVC-I1003 of the HEVC standard – “High efficiency video coding (HEVC) text specification draft 7,” JCTVC-I1003_d5, 9th meeting, Geneva, April 2012.

The innovations described herein can also be implemented for other standards or formats. For example, innovations described herein can be implemented for the H.264/AVC standard using frame packing arrangement SEI messages.

[037] More generally, various alternatives to the examples described herein are

5 possible. For example, any of the methods described herein can be altered by changing the ordering of the method acts described, by splitting, repeating, or omitting certain method acts, etc. The various aspects of the disclosed technology can be used in combination or separately. Different embodiments use one or more of the described innovations. Some of the innovations described herein address one or more of the
10 problems noted in the background. Typically, a given technique/tool does not solve all such problems.

I. Example Computing Systems.

[038] Figure 1 illustrates a generalized example of a suitable computing system (100)

15 in which several of the described innovations may be implemented. The computing system (100) is not intended to suggest any limitation as to scope of use or functionality, as the innovations may be implemented in diverse general-purpose or special-purpose computing systems.

[039] With reference to Figure 1, the computing system (100) includes one or more

20 processing units (110, 115) and memory (120, 125). In Figure 1, this most basic configuration (130) is included within a dashed line. The processing units (110, 115) execute computer-executable instructions. A processing unit can be a general-purpose central processing unit ("CPU"), processor in an application-specific integrated circuit or any other type of processor. In a multi-processing system, multiple processing units
25 execute computer-executable instructions to increase processing power. For example, Figure 1 shows a central processing unit (110) as well as a graphics processing unit or co-processing unit (115). The tangible memory (120, 125) may be volatile memory (*e.g.*, registers, cache, RAM), non-volatile memory (*e.g.*, ROM, EEPROM, flash memory, etc.), or some combination of the two, accessible by the processing unit(s). The memory (120,
30 125) stores software (180) implementing one or more innovations for frame packing with wavelet decomposition (or other band separation filtering) and/or unpacking with wavelet reconstruction (or other inverse band separation filtering) for higher-resolution chroma sampling formats, in the form of computer-executable instructions suitable for execution by the processing unit(s).

[040] A computing system may have additional features. For example, the computing system (100) includes storage (140), one or more input devices (150), one or more output devices (160), and one or more communication connections (170). An interconnection mechanism (not shown) such as a bus, controller, or network interconnects the components of the computing system (100). Typically, operating system software (not shown) provides an operating environment for other software executing in the computing system (100), and coordinates activities of the components of the computing system (100).

[041] The tangible storage (140) may be removable or non-removable, and includes magnetic disks, magnetic tapes or cassettes, CD-ROMs, DVDs, or any other medium which can be used to store information in a non-transitory way and which can be accessed within the computing system (100). The storage (140) stores instructions for the software (180) implementing one or more innovations for frame packing with wavelet decomposition (or other band separation filtering) and/or unpacking with wavelet reconstruction (or other inverse band separation filtering) for higher-resolution chroma sampling formats.

[042] The input device(s) (150) may be a touch input device such as a keyboard, mouse, pen, or trackball, a voice input device, a scanning device, or another device that provides input to the computing system (100). For video encoding, the input device(s) (150) may be a camera, video card, TV tuner card, or similar device that accepts video input in analog or digital form, or a CD-ROM or CD-RW that reads video samples into the computing system (100). The output device(s) (160) may be a display, printer, speaker, CD-writer, or another device that provides output from the computing system (100).

[043] The communication connection(s) (170) enable communication over a communication medium to another computing entity. The communication medium conveys information such as computer-executable instructions, audio or video input or output, or other data in a modulated data signal. A modulated data signal is a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal. By way of example, and not limitation, communication media can use an electrical, optical, RF, or other carrier.

[044] The innovations can be described in the general context of computer-readable media. Computer-readable media are any available tangible media that can be accessed within a computing environment. By way of example, and not limitation, with the computing system (100), computer-readable media include memory (120, 125), storage (140), and combinations of any of the above.

[045] The innovations can be described in the general context of computer-executable instructions, such as those included in program modules, being executed in a computing system on a target real or virtual processor. Generally, program modules include routines, programs, libraries, objects, classes, components, data structures, etc. that perform particular tasks or implement particular abstract data types. The functionality of the program modules may be combined or split between program modules as desired in various embodiments. Computer-executable instructions for program modules may be executed within a local or distributed computing system.

[046] The terms “system” and “device” are used interchangeably herein. Unless the context clearly indicates otherwise, neither term implies any limitation on a type of computing system or computing device. In general, a computing system or computing device can be local or distributed, and can include any combination of special-purpose hardware and/or general-purpose hardware with software implementing the functionality described herein.

[047] The disclosed methods can also be implemented using specialized computing hardware configured to perform any of the disclosed methods. For example, the disclosed methods can be implemented by an integrated circuit (*e.g.*, an application specific integrated circuit (“ASIC”) (such as an ASIC digital signal process unit (“DSP”), a graphics processing unit (“GPU”), or a programmable logic device (“PLD”), such as a field programmable gate array (“FPGA”)) specially designed or configured to implement any of the disclosed methods.

[048] For the sake of presentation, the detailed description uses terms like “determine” and “use” to describe computer operations in a computing system. These terms are high-level abstractions for operations performed by a computer, and should not be confused with acts performed by a human being. The actual computer operations corresponding to these terms vary depending on implementation.

II. Example Network Environments.

[049] Figures 2a and 2b show example network environments (201, 202) that include video encoders (220) and video decoders (270). The encoders (220) and decoders (270) are connected over a network (250) using an appropriate communication protocol. The network (250) can include the Internet or another computer network.

[050] In the network environment (201) shown in Figure 2a, each real-time communication (“RTC”) tool (210) includes both an encoder (220) and a decoder (270)

for bidirectional communication. A given encoder (220) can produce output compliant with the SMPTE 421M standard, ISO/IEC 14496-10 standard (also known as H.264/AVC), H.265/HEVC standard, another standard, or a proprietary format, with a corresponding decoder (270) accepting encoded data from the encoder (220). The

5 bidirectional communication can be part of a video conference, video telephone call, or other two-party communication scenario. Although the network environment (201) in Figure 2a includes two real-time communication tools (210), the network environment (201) can instead include three or more real-time communication tools (210) that participate in multi-party communication.

10 **[051]** A real-time communication tool (210) manages encoding by an encoder (220). Figure 4 shows an example encoder system (400) that can be included in the real-time communication tool (210). Alternatively, the real-time communication tool (210) uses another encoder system. A real-time communication tool (210) also manages decoding by a decoder (270). Figure 5 shows an example decoder system (500), which can be included
15 in the real-time communication tool (210). Alternatively, the real-time communication tool (210) uses another decoder system.

[052] In the network environment (202) shown in Figure 2b, an encoding tool (212) includes an encoder (220) that encodes video for delivery to multiple playback tools (214), which include decoders (270). The unidirectional communication can be provided for a
20 video surveillance system, web camera monitoring system, remote desktop conferencing presentation or other scenario in which video is encoded and sent from one location to one or more other locations. Although the network environment (202) in Figure 2b includes two playback tools (214), the network environment (202) can include more or fewer playback tools (214). In general, a playback tool (214) communicates with the encoding
25 tool (212) to determine a stream of video for the playback tool (214) to receive. The playback tool (214) receives the stream, buffers the received encoded data for an appropriate period, and begins decoding and playback.

[053] Figure 4 shows an example encoder system (400) that can be included in the encoding tool (212). Alternatively, the encoding tool (212) uses another encoder system.
30 The encoding tool (212) can also include server-side controller logic for managing connections with one or more playback tools (214). Figure 5 shows an example decoder system (500), which can be included in the playback tool (214). Alternatively, the playback tool (214) uses another decoder system. A playback tool (214) can also include client-side controller logic for managing connections with the encoding tool (212).

III. Example Frame Packing/Unpacking Systems.

[054] Figure 3 is a block diagram of a generalized frame packing/unpacking system (300) in conjunction with which some described embodiments may be implemented.

[055] The system (300) includes a video source (310), which produces source frames (311) of a higher-resolution chroma sampling format such as a 4:4:4 format. The video source (310) can be a camera, tuner card, storage media, or other digital video source.

[056] The frame packer (315) rearranges the frames (311) of the higher-resolution chroma sampling format to produce source frames (316) of a lower-resolution chroma sampling format such as a 4:2:0 format. Example approaches to frame packing (including wavelet decomposition or other band separation filtering) are described below. The frame packer (315) can signal metadata (317) that indicates whether and how frame packing was performed, for use by the frame unpacker (385) after decoding. Example approaches to signaling frame packing arrangement metadata are described below.

[057] The encoder (340) encodes the frames (316) of the lower-resolution chroma sampling format. Example encoders are described below with reference to Figures 4 and 6. The encoder (340) outputs coded data (341) over a channel (350), which represents storage, a communications connection, or another channel for the output.

[058] The decoder (360) receives the encoded data (341) and decodes the frames (316) of the lower-resolution chroma sampling format. Example decoders are described below with reference to Figures 5 and 7. The decoder outputs reconstructed frames (381) of the lower-resolution chroma sampling format.

[059] The frame unpacker (385) optionally rearranges the reconstructed frames (381) of the lower-resolution chroma sampling format to reconstruct the frames (386) of the higher-resolution chroma sampling format. Example approaches to frame unpacking (including wavelet reconstruction or other inverse band separation filtering) are described below. The frame unpacker (385) can receive the metadata (317) that indicates whether and how frame packing was performed, and use such metadata (317) to guide unpacking operations. The frame unpacker (385) outputs the reconstructed frames of the higher-resolution chroma sampling format to an output destination (390).

IV. Example Encoder Systems.

[060] Figure 4 is a block diagram of an example encoder system (400) in conjunction with which some described embodiments may be implemented. The encoder system (400) can be a general-purpose encoding tool capable of operating in any of multiple encoding

modes such as a low-latency encoding mode for real-time communication, transcoding mode, and regular encoding mode for media playback from a file or stream, or it can be a special-purpose encoding tool adapted for one such encoding mode. The encoder system (400) can be implemented as an operating system module, as part of an application library
5 or as a standalone application. Overall, the encoder system (400) receives a sequence of source video frames (411) (of a higher-resolution chroma sampling format such as a 4:4:4 format) from a video source (410), performs frame packing (including wavelet decomposition or other band separation filter) to a lower-resolution chroma sampling format such as a 4:2:0 format, encodes frames of the lower-resolution chroma sampling
10 format, and produces encoded data as output to a channel (490).

[061] The video source (410) can be a camera, tuner card, storage media, or other digital video source. The video source (410) produces a sequence of video frames at a frame rate of, for example, 30 frames per second. As used herein, the term “frame” generally refers to source, coded or reconstructed image data. For progressive-scan video,
15 a frame is a progressive-scan video frame. For interlaced video, in example embodiments, an interlaced video frame is de-interlaced prior to encoding. Alternatively, two complementary interlaced video fields are encoded as an interlaced video frame or as separate fields. Aside from indicating a progressive-scan video frame, the term “frame” can indicate a single non-paired video field, a complementary pair of video fields, a video
20 object plane that represents a video object at a given time, or a region of interest in a larger image. The video object plane or region can be part of a larger image that includes multiple objects or regions of a scene. After color space conversion from the capture format (e.g., an RGB format), the source frames (411) are in a higher-resolution chroma sampling format such as a 4:4:4 format.

[062] The frame packer (415) rearranges the frames (411) of the higher-resolution chroma sampling format to produce source frames (416) of a lower-resolution chroma sampling format such as a 4:2:0 format. Example approaches to frame packing are described below. The frame packer (415) can signal metadata (not shown) that indicates whether and how frame packing was performed, for use by a frame unpacker after
30 decoding. Example approaches to signaling frame packing arrangement metadata are described below. The frame packer (415) can perform pre-processing operations, for example, wavelet decomposition or other band separation filter, as described below.

[063] An arriving source frame (416) is stored in a source frame temporary memory storage area (420) that includes multiple frame buffer storage areas (421, 422, ... , 42n).

A frame buffer (421, 422, etc.) holds one source frame in the source frame storage area (420). After one or more of the source frames (416) have been stored in frame buffers (421, 422, etc.), a frame selector (430) periodically selects an individual source frame from the source frame storage area (420). The order in which frames are selected by the frame selector (430) for input to the encoder (440) may differ from the order in which the frames are produced by the video source (410), *e.g.*, a selected frame may be ahead in order, to facilitate temporally backward prediction.

[064] The order of the frame packer (415) and frame storage area (420) can be switched. Before the encoder (440), the encoder system (400) can include another pre-processor (not shown) that performs pre-processing (*e.g.*, filtering) of the selected frame (431) before encoding.

[065] The encoder (440) encodes the selected frame (431) (of the lower-resolution chroma sampling format) to produce a coded frame (441) and also produces memory management control operation (“MMCO”) signals (442) or reference picture set (“RPS”) information. If the current frame is not the first frame that has been encoded, when performing its encoding process, the encoder (440) may use one or more previously encoded/decoded frames (469) that have been stored in a decoded frame temporary memory storage area (460). Such stored decoded frames (469) are used as reference frames for inter-frame prediction of the content of the current source frame (431).

Generally, the encoder (440) includes multiple encoding modules that perform encoding tasks such as motion estimation and compensation, frequency transforms, quantization and entropy coding. The exact operations performed by the encoder (440) can vary depending on compression format. The format of the output encoded data can be a Windows Media Video format, VC-1 format, MPEG-x format (*e.g.*, MPEG-1, MPEG-2, or MPEG-4), H.26x format (*e.g.*, H.261, H.262, H.263, H.264), HEVC format or other format. In general, the encoder (440) is adapted for encoding frames of the lower-resolution chroma sampling format.

[066] For example, within the encoder (440), an inter-coded, predicted frame is represented in terms of prediction from reference frames. A motion estimator estimates motion of sets of samples of a source frame (441) with respect to one or more reference frames (469). A set of samples can be a macroblock, sub-macroblock or sub-macroblock partition (as in the H.264 standard), or it can be a coding tree unit or prediction unit (as in the HEVC standard). Generally, as used herein, the term “block” indicates a set of samples, which may be a single two-dimensional (“2D”) array or multiple 2D arrays (*e.g.*,

one array for a luma component and two arrays for chroma components). When multiple reference frames are used, the multiple reference frames can be from different temporal directions or the same temporal direction. The motion estimator outputs motion information such as motion vector information, which is entropy coded. A motion

5 compensator applies motion vectors to reference frames to determine motion-compensated prediction values. The encoder determines the differences (if any) between a block's motion-compensated prediction values and corresponding original values. These prediction residual values (*i.e.*, residuals, residue values) are further encoded using a frequency transform, quantization and entropy encoding. Similarly, for intra prediction, 10 the encoder (440) can determine intra-prediction values for a block, determine prediction residual values, and encode the prediction residual values. The entropy coder of the encoder (440) compresses quantized transform coefficient values as well as certain side information (*e.g.*, motion vector information, QP values, mode decisions, parameter choices). Typical entropy coding techniques include Exp-Golomb coding, arithmetic 15 coding, differential coding, Huffman coding, run length coding, variable-length-to-variable-length ("V2V") coding, variable-length-to-fixed-length ("V2F") coding, LZ coding, dictionary coding, probability interval partitioning entropy coding ("PIPE"), and combinations of the above. The entropy coder can use different coding techniques for different kinds of information, and can choose from among multiple code tables within a 20 particular coding technique.

[067] The coded frames (441) and MMCO/RPS information (442) are processed by a decoding process emulator (450). The decoding process emulator (450) implements some of the functionality of a decoder, for example, decoding tasks to reconstruct reference frames that are used by the encoder (440) in motion estimation and compensation. The 25 decoding process emulator (450) uses the MMCO/RPS information (442) to determine whether a given coded frame (441) needs to be stored for use as a reference frame in inter-frame prediction of subsequent frames to be encoded. If the MMCO/RPS information (442) indicates that a coded frame (441) needs to be stored, the decoding process emulator (450) models the decoding process that would be conducted by a decoder that receives the 30 coded frame (441) and produces a corresponding decoded frame (451). In doing so, when the encoder (440) has used decoded frame(s) (469) that have been stored in the decoded frame storage area (460), the decoding process emulator (450) also uses the decoded frame(s) (469) from the storage area (460) as part of the decoding process.

[068] The decoded frame temporary memory storage area (460) includes multiple frame buffer storage areas (461, 462, ..., 46 n). The decoding process emulator (450) uses the MMCO/RPS information (442) to manage the contents of the storage area (460) in order to identify any frame buffers (461, 462, etc.) with frames that are no longer needed by the encoder (440) for use as reference frames. After modeling the decoding process, the decoding process emulator (450) stores a newly decoded frame (451) in a frame buffer (461, 462, etc.) that has been identified in this manner.

[069] The coded frames (441) and MMCO/RPS information (442) are also buffered in a temporary coded data area (470). The coded data that is aggregated in the coded data area (470) can also include media metadata relating to the coded video data (*e.g.*, as one or more parameters in one or more SEI messages (such as frame packing arrangement SEI messages) or video usability information (“VUI”) messages).

[070] The aggregated data (471) from the temporary coded data area (470) are processed by a channel encoder (480). The channel encoder (480) can packetize the aggregated data for transmission as a media stream (*e.g.*, according to a media container format such as ISO/IEC 14496-12), in which case the channel encoder (480) can add syntax elements as part of the syntax of the media transmission stream. Or, the channel encoder (480) can organize the aggregated data for storage as a file (*e.g.*, according to a media container format such as ISO/IEC 14496-12), in which case the channel encoder (480) can add syntax elements as part of the syntax of the media storage file. Or, more generally, the channel encoder (480) can implement one or more media system multiplexing protocols or transport protocols, in which case the channel encoder (480) can add syntax elements as part of the syntax of the protocol(s). Such syntax elements for a media transmission stream, media storage stream, multiplexing protocols or transport protocols can include frame packing arrangement metadata. The channel encoder (480) provides output to a channel (490), which represents storage, a communications connection, or another channel for the output.

V. Example Decoder Systems.

[071] Figure 5 is a block diagram of an example decoder system (500) in conjunction with which some described embodiments may be implemented. The decoder system (500) can be a general-purpose decoding tool capable of operating in any of multiple decoding modes such as a low-latency decoding mode for real-time communication and regular decoding mode for media playback from a file or stream, or it can be a special-purpose

decoding tool adapted for one such decoding mode. The decoder system (500) can be implemented as an operating system module, as part of an application library or as a standalone application. Overall, the decoder system (500) receives coded data from a channel (510), decodes frames of a lower-resolution chroma sampling format such as a 4:2:0 format, optionally performs frame unpacking (including wavelet reconstruction or other inverse band separation filter) from the lower-resolution chroma sampling format to a higher-resolution chroma sampling format such as a 4:4:4 format, and produces reconstructed frames (of the higher-resolution chroma sampling format) as output for an output destination (590).

[072] The decoder system (500) includes a channel (510), which can represent storage, a communications connection, or another channel for coded data as input. The channel (510) produces coded data that has been channel coded. A channel decoder (520) can process the coded data. For example, the channel decoder (520) de-packetizes data that has been aggregated for transmission as a media stream (*e.g.*, according to a media container format such as ISO/IEC 14496-12), in which case the channel decoder (520) can parse syntax elements added as part of the syntax of the media transmission stream. Or, the channel decoder (520) separates coded video data that has been aggregated for storage as a file (*e.g.*, according to a media container format such as ISO/IEC 14496-12), in which case the channel decoder (520) can parse syntax elements added as part of the syntax of the media storage file. Or, more generally, the channel decoder (520) can implement one or more media system demultiplexing protocols or transport protocols, in which case the channel decoder (520) can parse syntax elements added as part of the syntax of the protocol(s). Such syntax elements for a media transmission stream, media storage stream, multiplexing protocols or transport protocols can include frame packing arrangement metadata.

[073] The coded data (521) that is output from the channel decoder (520) is stored in a temporary coded data area (530) until a sufficient quantity of such data has been received. The coded data (521) includes coded frames (531) (at the lower-resolution chroma sampling format) and MMCO/RPS information (532). The coded data (521) in the coded data area (530) can also include media metadata relating to the encoded video data (*e.g.*, as one or more parameters in one or more SEI messages such as frame packing arrangement SEI messages or VUI messages). In general, the coded data area (530) temporarily stores coded data (521) until such coded data (521) is used by the decoder (550). At that point, coded data for a coded frame (531) and MMCO/RPS information (532) are transferred

from the coded data area (530) to the decoder (550). As decoding continues, new coded data is added to the coded data area (530) and the oldest coded data remaining in the coded data area (530) is transferred to the decoder (550).

[074] The decoder (550) periodically decodes a coded frame (531) to produce a
5 corresponding decoded frame (551) of the lower-resolution chroma sampling format. As appropriate, when performing its decoding process, the decoder (550) may use one or more previously decoded frames (569) as reference frames for inter-frame prediction. The decoder (550) reads such previously decoded frames (569) from a decoded frame temporary memory storage area (560). Generally, the decoder (550) includes multiple
10 decoding modules that perform decoding tasks such as entropy decoding, inverse quantization, inverse frequency transforms and motion compensation. The exact operations performed by the decoder (550) can vary depending on compression format. In general, the decoder (550) is adapted for decoding frames of the lower-resolution chroma sampling format.

[075] For example, the decoder (550) receives encoded data for a compressed frame or
15 sequence of frames and produces output including decoded frame (551) of the lower-resolution chroma sampling format. In the decoder (550), a buffer receives encoded data for a compressed frame and makes the received encoded data available to an entropy decoder. The entropy decoder entropy decodes entropy-coded quantized data as well as
20 entropy-coded side information, typically applying the inverse of entropy encoding performed in the encoder. A motion compensator applies motion information to one or more reference frames to form motion-compensated predictions of blocks (*e.g.*, macroblocks, sub-macroblocks, sub-macroblock partitions, coding tree units, prediction units, or parts thereof) of the frame being reconstructed. An intra prediction module can
25 spatially predict sample values of a current block from neighboring, previously reconstructed sample values. The decoder (550) also reconstructs prediction residuals. An inverse quantizer inverse quantizes entropy-decoded data. An inverse frequency transformer converts the quantized, frequency domain data into spatial domain information. For a predicted frame, the decoder (550) combines reconstructed prediction
30 residuals with motion-compensated predictions to form a reconstructed frame. The decoder (550) can similarly combine prediction residuals with spatial predictions from intra prediction. A motion compensation loop in the video decoder (550) includes an adaptive de-blocking filter to smooth discontinuities across block boundary rows and/or columns in the decoded frame (551).

[076] The decoded frame temporary memory storage area (560) includes multiple frame buffer storage areas (561, 562, ..., 56n). The decoded frame storage area (560) is an example of a DPB. The decoder (550) uses the MMCO/RPS information (532) to identify a frame buffer (561, 562, etc.) in which it can store a decoded frame (551) of the lower-resolution chroma sampling format. The decoder (550) stores the decoded frame (551) in that frame buffer.

[077] An output sequencer (580) uses the MMCO/RPS information (532) to identify when the next frame to be produced in output order is available in the decoded frame storage area (560). When the next frame (581) of the lower-resolution chroma sampling format to be produced in output order is available in the decoded frame storage area (560), it is read by the output sequencer (580) and output to either (a) the output destination (590) (e.g., display) for display of the frame of the lower-resolution chroma sampling format, or (b) the frame unpacker (585). In general, the order in which frames are output from the decoded frame storage area (560) by the output sequencer (580) may differ from the order in which the frames are decoded by the decoder (550).

[078] The frame unpacker (585) rearranges the frames (581) of the lower-resolution chroma sampling format to produce output frames (586) of a higher-resolution chroma sampling format such as a 4:4:4 format. Example approaches to frame unpacking are described below. The frame packer (585) can use metadata (not shown) that indicates whether and how frame packing was performed, to guide frame unpacking operations. The frame unpacker (585) can perform post-processing operations, for example, wavelet reconstruction or other inverse band separation filter, as described below.

VI. Example Video Encoders.

[079] Figure 6 is a block diagram of a generalized video encoder (600) in conjunction with which some described embodiments may be implemented. The encoder (600) receives a sequence of video frames of a lower-resolution chroma sampling format such as a 4:2:0 format, including a current frame (605), and produces encoded data (695) as output.

[080] The encoder (600) is block-based and uses a macroblock format that depends on implementation. Blocks may be further sub-divided at different stages, e.g., at the frequency transform and entropy encoding stages. For example, a frame can be divided into 16x16 macroblocks, which can in turn be divided into 8x8 blocks and smaller sub-blocks of pixel values for coding and decoding.

[081] The encoder system (600) compresses predicted frames and intra-coded frames. For the sake of presentation, Figure 6 shows an “intra path” through the encoder (600) for intra-frame coding and an “inter path” for inter-frame coding. Many of the components of the encoder (600) are used for both intra-frame coding and inter-frame coding. The exact operations performed by those components can vary depending on the type of information being compressed.

[082] If the current frame (605) is a predicted frame, a motion estimator (610) estimates motion of blocks (*e.g.*, macroblocks, sub-macroblocks, sub-macroblock partitions, coding tree units, prediction units, or parts thereof) of the current frame (605) with respect to one or more reference frames. The frame store (620) buffers one or more reconstructed previous frames (625) for use as reference frames. When multiple reference frames are used, the multiple reference frames can be from different temporal directions or the same temporal direction. The motion estimator (610) outputs as side information motion information (615) such as differential motion vector information.

[083] The motion compensator (630) applies reconstructed motion vectors to the reconstructed reference frame(s) (625) when forming a motion-compensated current frame (635). The difference (if any) between block of the motion-compensated current frame (635) and corresponding part of the original current frame (605) is the prediction residual (645) for the block. During later reconstruction of the current frame, reconstructed prediction residuals are added to the motion-compensated current frame (635) to obtain a reconstructed frame that is closer to the original current frame (605). In lossy compression, however, some information is still lost from the original current frame (605). The intra path can include an intra prediction module (not shown) that spatially predicts pixel values of a current block from neighboring, previously reconstructed pixel values.

[084] A frequency transformer (660) converts spatial domain video information into frequency domain (*i.e.*, spectral, transform) data. For block-based video frames, the frequency transformer (660) applies a discrete cosine transform, an integer approximation thereof, or another type of forward block transform to blocks of pixel value data or prediction residual data, producing blocks of frequency transform coefficients. A quantizer (670) then quantizes the transform coefficients. For example, the quantizer (670) applies non-uniform, scalar quantization to the frequency domain data with a step size that varies on a frame-by-frame basis, macroblock-by-macroblock basis or other basis.

[085] When a reconstructed version of the current frame is needed for subsequent motion estimation/compensation, an inverse quantizer (676) performs inverse quantization on the quantized frequency coefficient data. The inverse frequency transformer (666) performs an inverse frequency transform, producing blocks of reconstructed prediction residuals or pixel values. For a predicted frame, the encoder (600) combines reconstructed prediction residuals (645) with motion-compensated predictions (635) to form the reconstructed frame (605). (Although not shown in Figure 6, in the intra path, the encoder (600) can combine prediction residuals with spatial predictions from intra prediction.) The frame store (620) buffers the reconstructed current frame for use in subsequent motion-compensated prediction.

[086] A motion compensation loop in the encoder (600) includes an adaptive in-loop deblock filter (610) before or after the frame store (620). The decoder (600) applies in-loop filtering to reconstructed frames to adaptively smooth discontinuities across boundaries in the frames. The adaptive in-loop deblock filter (610) can be disabled for some types of content. For example, in a frame packing approach with main and auxiliary views, the adaptive in-loop deblock filter (610) can be disabled when encoding an auxiliary view (including remaining chroma information that is not part of a main view) so as to not introduce artifacts such as blurring.

[087] The entropy coder (680) compresses the output of the quantizer (670) as well as motion information (615) and certain side information (*e.g.*, QP values). The entropy coder (680) provides encoded data (695) to the buffer (690), which multiplexes the encoded data into an output bitstream.

[088] A controller (not shown) receives inputs from various modules of the encoder. The controller evaluates intermediate results during encoding, for example, setting QP values and performing rate-distortion analysis. The controller works with other modules to set and change coding parameters during encoding. In particular, the controller can vary QP values and other control parameters to control quantization of luma components and chroma components during encoding. For example, the controller can vary QP values to dedicate more bits to luma content of a given frame (which could be a main view or auxiliary view in a frame packing approach) compared to chroma content of that frame. Or, in a frame packing approach with main and auxiliary views, the controller can vary QP values to dedicate more bits to the main view (including luma and sub-sampled chroma components) compared to the auxiliary view (including remaining chroma information).

[089] In some approaches to frame packing, even after chroma information from frames in a higher-resolution chroma sampling format has been packed into to-be-encoded frames of the lower-resolution chroma sampling format, the encoder can exploit geometric correspondence among sample values of the chroma components in several ways. The term geometric correspondence indicates a relationship between (1) chroma information at positions of a (nominally) luma component of a frame constructed from the lower-resolution chroma sampling format and (2) chroma information at corresponding scaled positions of chroma components of the frame of the lower-resolution chroma sampling format. A scaling factor applies between positions of the luma and chroma components. For example, for 4:2:0, the scaling factor is two both horizontally and vertically, and for 4:2:2, the scaling factor is two horizontally and one vertically.

[090] The encoder can use the geometric correspondence to guide motion estimation, QP selection, prediction mode selection or other decision-making processes from block-to-block, by first evaluating recent results of neighboring blocks when encoding a current block of the to-be-encoded frame. Or, the encoder can use the geometric correspondence to guide such decision-making processes for high-resolution chroma information packed into chroma components of the to-be-encoded frame, using results from encoding of high-resolution chroma information packed into a “luma” component of the to-be-encoded frame. Or, more directly, the encoder can use the geometric correspondence to improve compression performance, where motion vectors, prediction modes, or other decisions for high-resolution chroma information packed into a “luma” component of the to-be-encoded frame are also used for high-resolution chroma information packed into chroma components of the to-be-encoded frame. In particular, in some approaches described herein (*e.g.*, approach 2, below), when chroma information is packed into an auxiliary frame of the lower-resolution chroma sampling format, spatial correspondence and motion vector displacement relationships between the nominally luma component of the auxiliary frame and nominally chroma components of the auxiliary frame are preserved. Sample values at corresponding spatial positions in Y, U and V components of the auxiliary frame tend to be consistent, which is useful for such purposes as spatial block size segmentation and joint coding of coded block pattern information or other information that indicates presence/absence of non-zero coefficient values. Motion vectors for corresponding parts of Y, U and V components of the auxiliary frame tend to be consistent (*e.g.*, a vertical or horizontal displacement of two samples in Y corresponds to a displacement of 1 sample in U and V), which also helps coding efficiency.

[091] Depending on implementation and the type of compression desired, modules of the encoder can be added, omitted, split into multiple modules, combined with other modules, and/or replaced with like modules. In alternative embodiments, encoders with different modules and/or other configurations of modules perform one or more of the described techniques. Specific embodiments of encoders typically use a variation or supplemented version of the encoder (600). The relationships shown between modules within the encoder (600) indicate general flows of information in the encoder; other relationships are not shown for the sake of simplicity.

10 VII. Example Video Decoders.

[092] Figure 7 is a block diagram of a generalized decoder (700) in conjunction with which several described embodiments may be implemented. The decoder (700) receives encoded data (795) for a compressed frame or sequence of frames and produces output including a reconstructed frame (705) of a lower-resolution chroma sampling format such as a 4:2:0 format. For the sake of presentation, Figure 7 shows an “intra path” through the decoder (700) for intra-frame decoding and an “inter path” for inter-frame decoding. Many of the components of the decoder (700) are used for both intra-frame decoding and inter-frame decoding. The exact operations performed by those components can vary depending on the type of information being decompressed.

[093] A buffer (790) receives encoded data (795) for a compressed frame and makes the received encoded data available to the parser/entropy decoder (780). The parser/entropy decoder (780) entropy decodes entropy-coded quantized data as well as entropy-coded side information, typically applying the inverse of entropy encoding performed in the encoder.

[094] A motion compensator (730) applies motion information (715) to one or more reference frames (725) to form motion-compensated predictions (735) of blocks (*e.g.*, macroblocks, sub-macroblocks, sub-macroblock partitions, coding tree units, prediction units, or parts thereof) of the frame (705) being reconstructed. The frame store (720) stores one or more previously reconstructed frames for use as reference frames.

[095] The intra path can include an intra prediction module (not shown) that spatially predicts pixel values of a current block from neighboring, previously reconstructed pixel values. In the inter path, the decoder (700) reconstructs prediction residuals. An inverse quantizer (770) inverse quantizes entropy-decoded data. An inverse frequency transformer (760) converts the quantized, frequency domain data into spatial domain

information. For example, the inverse frequency transformer (760) applies an inverse block transform to frequency transform coefficients, producing pixel value data or prediction residual data. The inverse frequency transform can be an inverse discrete cosine transform, an integer approximation thereof, or another type of inverse frequency transform.

[096] For a predicted frame, the decoder (700) combines reconstructed prediction residuals (745) with motion-compensated predictions (735) to form the reconstructed frame (705). (Although not shown in Figure 7, in the intra path, the decoder (700) can combine prediction residuals with spatial predictions from intra prediction.) A motion compensation loop in the decoder (700) includes an adaptive in-loop deblock filter (710) before or after the frame store (720). The decoder (700) applies in-loop filtering to reconstructed frames to adaptively smooth discontinuities across boundaries in the frames. The adaptive in-loop deblock filter (710) can be disabled for some types of content, when it was disabled during encoding. For example, in a frame packing approach with main and auxiliary views, the adaptive in-loop deblock filter (710) can be disabled when decoding an auxiliary view (including remaining chroma information that is not part of a main view).

[097] In Figure 7, the decoder (700) also includes a post-processing deblock filter (708). The post-processing deblock filter (708) optionally smoothes discontinuities in reconstructed frames. Other filtering (such as de-ring filtering) can also be applied as part of the post-processing filtering. Typically, reconstructed frames that are subjected to later frame unpacking bypass the post-processing deblock filter (708).

[098] Depending on implementation and the type of decompression desired, modules of the decoder can be added, omitted, split into multiple modules, combined with other modules, and/or replaced with like modules. In alternative embodiments, decoders with different modules and/or other configurations of modules perform one or more of the described techniques. Specific embodiments of decoders typically use a variation or supplemented version of the decoder (700). The relationships shown between modules within the decoder (700) indicate general flows of information in the decoder; other relationships are not shown for the sake of simplicity.

VIII. Frame Packing/Unpacking for Higher-Resolution Chroma Sampling Formats.

[099] This section describes various approaches to packing frames of a higher-resolution chroma sampling format into frames of a lower-resolution chroma sampling format. The frames of the lower-resolution chroma sampling format can then be encoded using an encoder designed for the lower-resolution chroma sampling format. After decoding (using a decoder designed for the lower-resolution chroma sampling format), the frames at the lower-resolution chroma sampling format can be output for further processing and display. Or, after such decoding, the frames of the higher-resolution chroma sampling format can be recovered through frame unpacking for output and display.

A. Approaches to Frame Packing/Unpacking for YUV 4:4:4 Video.

[0100] Various approaches described herein can be used to preserve chroma information for frames of a 4:4:4 format when encoding/decoding uses a 4:2:0 format, as one specific example. In these approaches, for example, a YUV 4:4:4 frame is packed into two YUV 4:2:0 frames. A typical 4:4:4 frame contains 12 sample values for every 4 pixel positions, while a 4:2:0 frame contains only 6 sample values for every 4 pixel positions. So, all the sample values contained in a 4:4:4 frame can be packed into two 4:2:0 frames.

1. Approach 1.

[0101] In approach 1, a YUV 4:4:4 frame is packed into two YUV 4:2:0 frames using spatial partitioning. Figure 8 shows this approach (800) to frame packing that uses spatial partitioning of the YUV 4:4:4 frame.

[0102] A Y_{444} plane, U_{444} plane, and V_{444} plane are the three component planes for the YUV 4:4:4 frame (801). Each plane has the resolution of width W and height H . For convenience in describing the examples used herein, both W and H are divisible by 4, without implying that this is a limitation of the approach. The approach (800) to packing the YUV 4:4:4 frame into two YUV 4:2:0 frames splits the YUV 4:4:4 frame as shown in Figure 8. The U_{444} plane of the YUV 4:4:4 frame (801) is partitioned into a bottom half $H2-U_{444}$ and two upper quarters $Q1-U_{444}$ and $Q2-U_{444}$ using spatial partitioning. The V_{444} plane of the YUV 4:4:4 frame (801) is partitioned into a bottom half $H2-V_{444}$ and two upper quarters $Q1-V_{444}$ and $Q2-V_{444}$ using spatial partitioning.

[0103] The partitioned planes of the YUV 4:4:4 frame (801) are then reorganized as one or more YUV 4:2:0 frames. The Y_{444} plane for the YUV 4:4:4 frames becomes the luma component plane of a first frame (802) of the YUV 4:2:0 format. The bottom halves of the U_{444} plane and the V_{444} plane become the luma component plane of a second frame (803) of the YUV 4:2:0 format. The top quarters of the U_{444} plane and the V_{444} plane become the chroma component planes of the first frame (802) and second frame (803) of the YUV 4:2:0 format as shown in Figure 8.

[0104] The first frame (802) and second frame (803) of the YUV 4:2:0 format can be organized as separate frames (separated by the dark line in Figure 8). Or, the first frame (802) and second frame (803) of the YUV 4:2:0 format can be organized as a single frame having a height of $2 \times H$ (ignoring the dark line in Figure 8). Or, the first frame (802) and second frame (803) of the YUV 4:2:0 format can be organized as a single frame having a width of $2 \times W$. Or, the first frame (802) and second frame (803) of the YUV 4:2:0 format can be organized as a single frame or multiple frames using any of the methods defined for frame_packing_arrangement_type in the H.264/AVC standard or the HEVC standard.

[0105] Although this type of frame packing works, it does not result in geometric correspondence between Y, U and V components within each of the two YUV 4:2:0 frames. In particular, for the second frame (803) of the YUV 4:2:0 format, there is typically not a geometric correspondence between the luma component and chroma components. Other packing approaches described herein typically achieve much better geometric correspondence.

[0106] Alternatively, Approach 1 can be used for color spaces such as RGB, GBR, etc. in sampling ratios such as 4:4:4, 4:2:2, 4:2:0, etc., as the chroma sampling formats.

2. Approach 2.

[0107] In approach 2, a YUV 4:4:4 frame is packed into two YUV 4:2:0 frames while maintaining geometric correspondence for chroma information of the YUV 4:4:4 frame. YUV 4:2:0 frames with good geometric correspondence among their Y, U and V components can be compressed better because they fit the model expected by a typical encoder adapted to encode YUV 4:2:0 frames.

[0108] The packing can also be done such that one of the two YUV 4:2:0 frames represents the complete scene being represented by the YUV 4:4:4 frame, albeit with color components at a lower resolution. This provides options in decoding. A decoder that cannot perform frame unpacking, or chooses not to perform frame unpacking, can just take

a reconstructed version of the YUV 4:2:0 frame that represents the scene and directly feed it to the display.

[0109] Figure 9 illustrates one example approach (900) to frame packing that is consistent with these design constraints. In this approach (900), a YUV 4:4:4 frame (801) is packed into two YUV 4:2:0 frames (902, 903). The first frame (902) provides a “main view” in YUV 4:2:0 format – a lower chroma resolution version of the complete scene represented by the YUV 4:4:4 frame (801). The second frame (903) provides an “auxiliary view” in YUV 4:2:0 format and contains remaining chroma information.

[0110] In Figure 9, the areas B1 ... B9 are different areas within the respective frames (902, 903) of YUV 4:2:0 format. The sample values of odd rows of the U_{444} plane and V_{444} plane of the YUV 4:4:4 frame (801) are assigned to the areas B4 and B5, and the sample values of even rows of the U_{444} plane and V_{444} plane of the YUV 4:4:4 frame (801) are distributed between the areas B2, B3 and B6...B9. Specifically, sample values of the Y_{444} plane, U_{444} plane, and V_{444} plane of the YUV 4:4:4 frame (801) map to the areas B1 ... B9 as follows.

- For area B1, $Y_{420}^{\text{main}}(x, y) = Y_{444}(x, y)$, where the range of (x, y) is $[0, W - 1] \times [0, H - 1]$.
- For area B2, $U_{420}^{\text{main}}(x, y) = U_{444}(2x, 2y)$, where the range of (x, y) is $[0, \frac{W}{2} - 1] \times [0, \frac{H}{2} - 1]$.
- For area B3, $V_{420}^{\text{main}}(x, y) = V_{444}(2x, 2y)$, where the range of (x, y) is $[0, \frac{W}{2} - 1] \times [0, \frac{H}{2} - 1]$.
- For area B4, $Y_{420}^{\text{aux}}(x, y) = U_{444}(x, 2y + 1)$, where the range of (x, y) is $[0, W - 1] \times [0, \frac{H}{2} - 1]$.
- For area B5, $Y_{420}^{\text{aux}}(x, \frac{H}{2} + y) = V_{444}(x, 2y + 1)$, where the range of (x, y) is $[0, W - 1] \times [0, \frac{H}{2} - 1]$.
- For area B6, $U_{420}^{\text{aux}}(x, y) = U_{444}(2x + 1, 4y)$, where the range of (x, y) is $[0, \frac{W}{2} - 1] \times [0, \frac{H}{4} - 1]$.
- For area B7, $U_{420}^{\text{aux}}(x, \frac{H}{4} + y) = V_{444}(2x + 1, 4y)$, where the range of (x, y) is $[0, \frac{W}{2} - 1] \times [0, \frac{H}{4} - 1]$.

- For area B8, $V_{420}^{\text{aux}}(x, y) = U_{444}(2x + 1, 4y + 2)$, where the range of (x, y) is $[0, \frac{W}{2} - 1] \times [0, \frac{H}{4} - 1]$.
- For area B9, $V_{420}^{\text{aux}}(x, \frac{H}{4} + y) = V_{444}(2x + 1, 4y + 2)$, where the range of (x, y) is $[0, \frac{W}{2} - 1] \times [0, \frac{H}{4} - 1]$.

5 [0111] Alternatively, the sample values of the Y_{444} plane, U_{444} plane, and V_{444} plane of the YUV 4:4:4 frame (801) can be assigned to the areas B1 ... B9 in a different way. For example, the sample values of even rows of the U_{444} plane and V_{444} plane of the YUV 4:4:4 frame (801) are assigned to the areas B4 and B5, and the sample values of odd rows of the U_{444} plane and V_{444} plane of the YUV 4:4:4 frame (801) are distributed between the areas B2, B3 and B6...B9. Or, as another example, data from the original U plane of the YUV 4:4:4 frame can be arranged in the U plane of the auxiliary YUV 4:2:0 frame, and data from the original V plane of the YUV 4:4:4 frame can be arranged in the V plane of the auxiliary YUV 4:2:0 frame. In this example, compared to Figure 9, the sample values from $V_{444}(2x + 1, 4y)$ that are assigned to area B7 in the equations above can instead be assigned to area B8, and the sample values from $U_{444}(2x + 1, 4y + 2)$ that are assigned to area B8 in the equations above can instead be assigned to area B7. Or, the same sample values from U_{444} can be copied into a single area for B6 and B7 without separating every second row, and the same sample values from V_{444} can be copied into a single area for B8 and B9 without separating every second row. Either way, the U plane (or V plane) of the auxiliary YUV 4:2:0 frame is constructed from the U plane (or V plane) of the YUV 4:4:4 frame, without mixing content from different original U and V planes. (In contrast, in the example of Figure 9, the U plane (or V plane) of the auxiliary YUV 4:2:0 frame has a mixture of data from the U and V components of the YUV 4:4:4 frame. The upper half of the U plane (or V plane) of the auxiliary YUV 4:2:0 frame contains data from the original U plane, and the lower half contains data from the original V plane.)

25 [0112] The first frame (902) and second frame (903) of the YUV 4:2:0 format can be organized as separate frames (separated by the dark line in Figure 9). Or, the first frame (902) and second frame (903) of the YUV 4:2:0 format can be organized as a single frame having a height of $2 \times H$ (ignoring the dark line in Figure 9). Or, the first frame (902) and second frame (903) of the YUV 4:2:0 format can be organized as a single frame having a width of $2 \times W$. Or, the first frame (902) and second frame (903) of the YUV 4:2:0 format

can be organized as a single frame using any of the methods defined for frame_packing_arrangement_type in the H.264/AVC standard or the HEVC standard.

[0113] Figure 10 illustrates example frames packed according to the approach (900) of Figure 9. Figure 10 shows a YUV 4:4:4 frame (1001) that includes a Y₄₄₄ plane, U₄₄₄ plane, and V₄₄₄ plane.

[0114] After frame packing, the main view (1002) (first YUV 4:2:0 frame) is the YUV 4:2:0 equivalent of the original YUV 4:4:4 frame (1001). A decoding system can simply display a reconstructed version of the main view (1002) if YUV 4:4:4 is either not supported or considered not necessary.

[0115] The auxiliary view (1003) contains chroma information for the YUV 4:4:4 frame (1001). Even so, the auxiliary view (1003) fits the content model of a YUV 4:2:0 frame and is well suited for compression using a typical YUV 4:2:0 video encoder. Within the frame, the auxiliary view (1003) exhibits geometric correspondence across its Y, U and V components. Between frames, the auxiliary views are expected to show motion that is highly correlated across Y, U and V components.

[0116] Figure 11 illustrates another example approach (1100) to frame packing that is consistent with these design constraints. In this approach (1100), a YUV 4:4:4 frame (801) is packed into two YUV 4:2:0 frames (1102, 1103). Much like the approach (900) of Figure 9, in the approach (1100) of Figure 11, the first frame (1102) provides a “main view” in YUV 4:2:0 format – a lower chroma resolution version of the complete scene represented by the YUV 4:4:4 frame (801) – while the second frame (1103) provides an “auxiliary view” in YUV 4:2:0 format and contains remaining chroma information.

[0117] In Figure 11, the areas A1 ... A9 are different areas within the respective frames (1102, 1103) of YUV 4:2:0 format. The sample values of odd columns of the U₄₄₄ plane and V₄₄₄ plane of the YUV 4:4:4 frame (801) are assigned to the areas A4 and A5, and the sample values of even columns of the U₄₄₄ plane and V₄₄₄ plane of the YUV 4:4:4 frame (801) are distributed between the areas A2, A3 and A6...A9. Specifically, sample values of the Y₄₄₄ plane, U₄₄₄ plane, and V₄₄₄ plane of the YUV 4:4:4 frame (801) map to the areas A1 ... A9 as follows.

- For area A1, $Y_{420}^{\text{main}}(x, y) = Y_{444}(x, y)$, where the range of (x, y) is $[0, W - 1] \times [0, H - 1]$.
- For area A2, $U_{420}^{\text{main}}(x, y) = U_{444}(2x, 2y)$, where the range of (x, y) is $[0, \frac{W}{2} - 1] \times [0, \frac{H}{2} - 1]$.

- For area A3, $V_{420}^{\text{main}}(x, y) = V_{444}(2x, 2y)$, where the range of (x, y) is $[0, \frac{W}{2} - 1] \times [0, \frac{H}{2} - 1]$.
- For area A4, $Y_{420}^{\text{aux}}(x, y) = U_{444}(2x + 1, y)$, where the range of (x, y) is $[0, \frac{W}{2} - 1] \times [0, H - 1]$.
- 5 • For area A5, $Y_{420}^{\text{aux}}(\frac{W}{2} + x, y) = V_{444}(2x + 1, y)$, where the range of (x, y) is $[0, \frac{W}{2} - 1] \times [0, H - 1]$.
- For area A6, $U_{420}^{\text{aux}}(x, y) = U_{444}(4x, 2y + 1)$, where the range of (x, y) is $[0, \frac{W}{4} - 1] \times [0, \frac{H}{2} - 1]$.
- For area A7, $U_{420}^{\text{aux}}(\frac{W}{4} + x, y) = V_{444}(4x, 2y + 1)$, where the range of (x, y) is $[0, \frac{W}{4} - 1] \times [0, \frac{H}{2} - 1]$.
- 10 • For area A8, $V_{420}^{\text{aux}}(x, y) = U_{444}(4x + 2, 2y + 1)$, where the range of (x, y) is $[0, \frac{W}{4} - 1] \times [0, \frac{H}{2} - 1]$.
- For area A9, $V_{420}^{\text{aux}}(\frac{W}{4} + x, y) = U_{444}(4x + 2, 2y + 1)$, where the range of (x, y) is $[0, \frac{W}{4} - 1] \times [0, \frac{H}{2} - 1]$.

15 **[0118]** Alternatively, the sample values of the Y_{444} plane, U_{444} plane, and V_{444} plane of the YUV 4:4:4 frame (801) can be assigned to the areas A1 ... A9 in a different way. For example, the sample values of even columns of the U_{444} plane and V_{444} plane of the YUV 4:4:4 frame (801) are assigned to the areas A4 and A5, and the sample values of odd columns of the U_{444} plane and V_{444} plane of the YUV 4:4:4 frame (801) are distributed

20 between the areas A2, A3 and A6...A9. Or, as another example, data from the original U plane of the YUV 4:4:4 frame can be arranged in the U plane of the auxiliary YUV 4:2:0 frame, and data from the original V plane of the YUV 4:4:4 frame can be arranged in the V plane of the auxiliary YUV 4:2:0 frame. In this example, compared to Figure 11, the sample values from $V_{444}(4x, 2y + 1)$ that are assigned to area A7 in the equations above

25 are instead assigned to area A8, and the sample values from $U_{444}(4x + 2, 2y + 1)$ that are assigned to area A8 in the equations above are instead assigned to area A7. Or, the same sample values from U_{444} can be copied into a single area for A6 and A7 without separating every second column, and the same sample values from V_{444} can be copied into a single area for A8 and A9 without separating every second column. Either way, the U plane (or

V plane) of the auxiliary YUV 4:2:0 frame is constructed from the U plane (or V plane) of the YUV 4:4:4 frame, without mixing content from different original U and V planes.

[0119] The first frame (1102) and second frame (1103) of the YUV 4:2:0 format can be organized as separate frames (separated by the dark line in Figure 11). Or, the first frame

5 (1102) and second frame (1103) of the YUV 4:2:0 format can be organized as a single frame having a height of $2 \times H$ (ignoring the dark line in Figure 11). Or, the first frame (1102) and second frame (1103) of the YUV 4:2:0 format can be organized as a single frame having a width of $2 \times W$. Or, the first frame (1102) and second frame (1103) of the YUV 4:2:0 format can be organized as a single frame using any of the methods defined for
10 frame_packing_arrangement_type in the H.264/AVC standard or the HEVC standard.

[0120] Frame unpacking can simply mirror frame packing. Samples assigned to areas of with frames of YUV 4:2:0 format are assigned back to original locations in chroma components of frames of YUV 4:4:4 format. In one implementation, for example, during frame unpacking, samples in areas B2 ... B9 of frames of YUV 4:2:0 format are assigned
15 to reconstructed chroma components U'_{444} and V'_{444} of frame of YUV 4:4:4 format as shown in the following pseudocode..

```
for( x = 0; x < (W >> 1); x++ ) {
```

```
    for( y = 0; y < (H >> 1); y++ ) {
```

```
         $U'_{444}(2x, 2y + 1) = Y''_{420}^{aux}(2x, y)$ 
```

```
20         $V'_{444}(2x, 2y + 1) = Y''_{420}^{aux}(2x, (H \gg 1) + y)$ 
```

```
         $U'_{444}(2x + 1, 2y + 1) = Y''_{420}^{aux}(2x + 1, y)$ 
```

```
         $V'_{444}(2x + 1, 2y + 1) = Y''_{420}^{aux}(2x + 1, (H \gg 1) + y)$ 
```

```
        if( y % 2 == 0 ) {
```

```
             $U'_{444}(2x + 1, 2y) = U''_{420}^{aux}(x, y \gg 1)$ 
```

```
25             $V'_{444}(2x + 1, 2y) = U''_{420}^{aux}(x, (H \gg 2) + (y \gg 1))$ 
```

```
        } else {
```

```
             $U'_{444}(2x + 1, 2y) = V''_{420}^{aux}(x, y \gg 1)$ 
```

```
             $V'_{444}(2x + 1, 2y) = V''_{420}^{aux}(x, (H \gg 2) + (y \gg 1))$ 
```

```
        }
```

```
30         $U'_{444}(2x, 2y) = U''_{420}^{main}(x, y)$ 
```

```
         $V'_{444}(2x, 2y) = V''_{420}^{main}(x, y)$ 
```

```
    }
```

```
}
```

where the " mark indicates reconstruction from (possibly lossy) coding.

B. Syntax and Semantics of Values for Signaling Frame Packing

Information.

5 [0121] In example implementations, a frame packing arrangement SEI message is used to signal that two 4:2:0 frames include a packed 4:4:4 frame. Frame packing arrangement SEI messages are defined in the H.264/AVC standard as well as the HEVC standard, although such frame packing arrangement SEI messages have previously been used for a different purpose.

10 [0122] A frame packing arrangement SEI message was designed to send stereoscopic 3D video frames using a 2D video codec. In such a case, two 4:2:0 frames represent the left and right views of a stereoscopic 3D video scene. For the approaches described herein, the scope of the frame packing arrangement SEI message can be extended to instead support encoding/decoding of two 4:2:0 frames obtained from a single 4:4:4 frame,
15 followed by frame unpacking to recover the 4:4:4 frame. The two 4:2:0 frames represent a main view and an auxiliary view. Both the main and auxiliary views (frames) are in a format that is an equivalent of a 4:2:0 format. The main view (frame) may be independently useful, while the auxiliary view (frame) is useful when interpreted appropriately together with the main view. Thus, these approaches can use the frame
20 packing arrangement SEI message to effectively support encoding/decoding 4:4:4 frames using video codecs capable of coding/decoding 4:2:0 frames.

[0123] To this end, the SEI message is extended. For example, the semantics of the syntax element `content_interpretation_type` are extended as follows. In the relevant frame packing approaches, for a YUV 4:4:4 frame, there are two constituent YUV 4:2:0 frames –
25 a first frame for a main view, and a second frame for an auxiliary view. The `content_interpretation_type` indicates the intended interpretation of the constituent frames as specified in the following table. The values 0, 1 and 2 are interpreted as in the H.264/AVC standard and HEVC standard. New values for `content_interpretation_type` are defined to indicate that the constituent frames should be interpreted as containing data
30 from YUV 4:4:4 frames:

Value	Interpretation
0	Unspecified relationship between the frame packed constituent frames.
1	Indicates that the two constituent frames form the left and right views of a stereo view scene, with frame 0 being associated with the left view and frame 1 being associated with the right view.
2	Indicates that the two constituent frames form the right and left views of a stereo view scene, with frame 0 being associated with the right view and frame 1 being associated with the left view.
3	Indicates that the two constituent frames form the main and auxiliary YUV 4:2:0 frames representing a YUV 4:4:4 frame, with frame 0 being associated with the main view and frame 1 being associated with the auxiliary view. Indicates that the chroma samples of frame 0 should be interpreted as unfiltered samples of the 4:4:4 frame (without anti-alias filtering).
4	Indicates that the two constituent frames form the main and auxiliary YUV 4:2:0 frames representing a YUV 4:4:4 frame, with frame 0 being associated with the main view and frame 1 being associated with the auxiliary view. Indicates that the chroma samples of frame 0 should be interpreted as having been anti-alias filtered prior to frame packing.
5	Indicates that the two constituent frames form the main and auxiliary YUV 4:2:0 frames representing a YUV 4:4:4 frame, with frame 1 being associated with the main view and frame 0 being associated with the auxiliary view. Indicates that the chroma samples of frame 1 should be interpreted as unfiltered samples of the 4:4:4 frame (without anti-alias filtering).
6	Indicates that the two constituent frames form the main and auxiliary YUV 4:2:0 frames representing a YUV 4:4:4 frame, with frame 1 being associated with the main view and frame 0 being associated with the auxiliary view. Indicates that the chroma samples of frame 1 should be interpreted as having been anti-alias filtered prior to frame packing.

[0124] Alternatively, different values for the syntax element `content_interpretation_type` are associated with the interpretations shown in the preceding table. Or, other and/or additional interpretations for `content_interpretation_type` can be used to support

encoding/decoding of frames of a lower-resolution chroma sampling format obtained from one or more frames of a higher-resolution chroma sampling format by frame packing.

[0125] In addition, for the purpose of simplification, one or more of the following constraints may also be imposed for other syntax elements of a frame packing arrangement

5 SEI message. When `content_interpretation_type` has a value between 3 and 6 (that is, for cases involving frame packing of YUV 4:4:4 frames into YUV 4:2:0 frames), the values of the syntax elements `quincunx_sampling_flag`, `spatial_flipping_flag`, `frame0_grid_position_x`, `frame0_grid_position_y`, `frame1_grid_position_x`, and `frame1_grid_position_y` shall be 0. Furthermore, when `content_interpretation_type` is
10 equal to 3 or 5 (indicating absence of filtering in pre-processing), `chroma_loc_info_present_flag` shall be 1, and the values of `chroma_sample_loc_type_top_field` and `chroma_sample_loc_type_bottom_field` shall be 2.

[0126] In the H.264/AVC standard (and in the HEVC standard), the syntax element

15 `frame_packing_arrangement_type` indicates how two constituent frames of a stereoscopic view are arranged. For example, `frame_packing_arrangement_type` == 3 indicates side-by-side packing of the two constituent frames, `frame_packing_arrangement_type` == 4 indicates top-bottom packing of the two constituent frames, and

`frame_packing_arrangement_type` == 5 indicates temporal interleaving of the two

20 constituent frames. The syntax element `frame_packing_arrangement_type` can be used similarly in conjunction with values of `content_interpretation_type` that indicate packing of frames of a higher-resolution chroma sampling format. For example,

`frame_packing_arrangement_type` == 3 can indicate side-by-side packing of main and auxiliary frames, `frame_packing_arrangement_type` == 4 can indicate top-bottom packing
25 of main and auxiliary frames, and `frame_packing_arrangement_type` == 5 can indicate temporal interleaving of main and auxiliary frames. Or, frame packing arrangement metadata is signaled in some other way. Alternatively, instead of extending the semantics of the `content_interpretation_type` syntax element to indicate packing of frames of a higher-resolution chroma sampling format, the semantics of

30 `frame_packing_arrangement_type` can be extended to indicate packing of frames of a higher-resolution chroma sampling format. For example, frame packing arrangement metadata (such as values of `frame_packing_arrangement_type` higher than 5) can indicate whether frame packing/unpacking is used or not used, whether filtering or other pre-processing operations were used or not used (and hence whether corresponding post-

processing filtering or other post-processing operations should be used or not used), the type of post-processing operations to perform, or other information about frame packing/unpacking, in addition to indicating how the main and auxiliary views are arranged.

- 5 [0127] In these examples, the frame packing arrangement SEI message informs a decoder that the decoded pictures contain main and auxiliary views of a 4:4:4 frame as the constituent frames of the frame packing arrangement. This information can be used to process the main and auxiliary views appropriately for display or other purposes. For example, when the system at the decoding end desires the video in 4:4:4 format and is
- 10 capable of reconstructing the 4:4:4 frames from the main and auxiliary views, the system may do so, and the output format will be 4:4:4. Otherwise, only the main view is given as output, and the output format will then be 4:2:0.

C. First Set of Example Pre-processing and Post-processing Operations.

- 15 [0128] Simple sub-sampling of the chroma sample values of frames of a higher-resolution chroma sampling format can introduce aliasing artifacts in the downsampled chroma sample values. To mitigate aliasing, frame packing can include pre-processing operations to filter chroma sample values. Such filtering can be termed anti-alias filtering. Corresponding frame unpacking can then include post-processing operations to
- 20 compensate for the pre-processing filtering of the chroma sample values. For example, with reference to the preceding table, when the content_interpretation_type is 4 or 6, pre-processing operations can be used to filter the chroma sample values during frame packing, and frame unpacking can include corresponding post-processing operations. This section describes a first set of example pre-processing and post-processing operations.
- 25 Another set of example pre-processing and post-processing operations, involving wavelet decomposition/reconstruction or other band separation filtering/inverse filtering, are described below.

[0129] There are various reasons for pre-processing and post-processing adapted to frame packing/unpacking.

- 30 [0130] For example, pre-processing can help improve quality when only the YUV 4:2:0 frame representing the main view is used for display. This can permit a decoder to ignore the YUV 4:2:0 frame representing the auxiliary view without running the risk of aliasing artifacts caused by simple sub-sampling of chroma information. Without pre-processing (when the chroma signal for the YUV 4:2:0 frame representing the main view is obtained

by direct sub-sampling of the chroma signal from the YUV 4:4:4 frame), aliasing artifacts can be seen on some content, for example, ClearType text content, when only the main view is used to generate output.

[0131] As another example, pre-processing and post-processing can help

maintain/enforce consistency and smoothness of the compressed chroma signal in the YUV 4:4:4 domain. When frame packing is used to pack a YUV 4:4:4 frame into two YUV 4:2:0 frames, the chroma signal is split into multiple areas, and each area may get compressed differently (*e.g.*, with a different level of quantization) depending on its location. Because of this, when the chroma signal is assembled again by interleaving the data from multiple areas, artificial discontinuities and high-frequency noise may be introduced. A post-processing operation can help smooth the differences caused in these areas due to compression.

[0132] As another example, pre-processing can help enhance the compression of the YUV 4:2:0 frame representing the auxiliary view, which contains the remaining chroma information.

[0133] In some example implementations, the pre-processing operations and post-processing operations are limited such that they affect only the chroma signal that is part of the YUV 4:2:0 frame representing the main view. That is, the filtered sample values are part of the chroma components of the main view.

[0134] Additionally, for frame packing/unpacking in conjunction with AVC coding/decoding or HEVC coding/decoding, pre-processing operations and post-processing operations can be based on the chroma sample location type (indicating chroma sample grid alignment with luma sample grid). The chroma sample location type is determined from `chroma_sample_loc_type_top_field` and `chroma_sample_loc_type_bottom_field` syntax elements signaled as part of the compressed bitstream. (These two elements would ordinarily have equal values for progressive-scan source content.) For a given chroma sample location type, if the chroma sample aligns with the luma samples for a particular direction (horizontal or vertical), then an odd-tap symmetric filter (such as $[1\ 2\ 1]/4$, or $[0.25\ 0.5\ 0.25]$, along with a rounding operation) is used to filter chroma in that direction. On the other hand, if the chroma sample does not align with the luma samples for a particular direction (horizontal or vertical), and the chroma sample grid positions are centered between the luma sample positions for a particular direction (horizontal/vertical), then an even-tap symmetric filter (typically $[1\ 1]/2$, or $[0.5\ 0.5]$, along with a rounding operation) is used to filter chroma in

that direction. Another possible filter choice for the latter case is $[1 \ 3 \ 3 \ 1] / 8$, or $[0.125 \ 0.375 \ 0.375 \ 0.125]$, along with a rounding operation. The choice of post-processing operation is usually made such that the post-processing operation compensates for the pre-processing operation. In some cases post-processing directly inverts pre-processing, while in other cases post-processing only approximately inverts pre-processing, as explained below.

[0135] In implementations of frame packing/unpacking in conjunction with AVC coding/decoding or HEVC coding/decoding, if the chroma sample location type is 1 for chroma_sample_loc_type_top_field and chroma_sample_loc_type_bottom_field syntax elements, the chroma sample does not align with luma samples in either horizontal or vertical direction, and hence the filter $[0.5 \ 0.5]$ is applied in both the horizontal and vertical directions for the pre-processing operation. In such a case, for the approach (900) illustrated with reference to Figure 9, the equations for deriving the sample values areas B2 and B3 are as follows.

- For area B2: $U_{420}^{\text{main_filt}}(x, y) = [U_{444}(2x, 2y) + U_{444}(2x + 1, 2y) + U_{444}(2x, 2y + 1) + U_{444}(2x + 1, 2y + 1) + 2] / 4$, and
- For area B3: $V_{420}^{\text{main_filt}}(x, y) = [V_{444}(2x, 2y) + V_{444}(2x + 1, 2y) + V_{444}(2x, 2y + 1) + V_{444}(2x + 1, 2y + 1) + 2] / 4$,

where the range of (x, y) is $\left[0, \frac{W}{2} - 1\right] \times \left[0, \frac{H}{2} - 1\right]$ for both areas.

[0136] Due to this filtering, the sample values at positions $U_{444}(2x, 2y)$ and $V_{444}(2x, 2y)$ from the YUV 4:4:4 frame are not represented directly in the main view (902); instead, filtered sample values ($U_{420}^{\text{main_filt}}(x, y)$ and $V_{420}^{\text{main_filt}}(x, y)$) are at positions in the main view (902). The sample values at $U_{444}(2x + 1, 2y)$, $U_{444}(2x, 2y + 1)$, $U_{444}(2x + 1, 2y + 1)$, $V_{444}(2x + 1, 2y)$, $V_{444}(2x, 2y + 1)$ and $V_{444}(2x + 1, 2y + 1)$ from the YUV 4:4:4 frame are still represented directly in the auxiliary view (903) among the areas B4 ... B9.

[0137] In corresponding filtering as part of post-processing operations when frames in YUV 4:4:4 format are to be output, the sample values for position $U_{444}(2x, 2y)$ and $V_{444}(2x, 2y)$ of the YUV 4:4:4 frame can be calculated as $U'_{444}(2x, 2y)$ and $V'_{444}(2x, 2y)$, from values in the packed frame, as follows:

- $U'_{444}(2x, 2y) = (1 + 3\alpha) * U_{420}^{\text{main_filt}}(x, y) - \alpha * [U''_{444}(2x + 1, 2y) + U''_{444}(2x, 2y + 1) + U''_{444}(2x + 1, 2y + 1)]$, and

- $V'_{444}(2x, 2y) = (1 + 3\alpha) * V''_{420}^{\text{main_filt}}(x, y) - \alpha * [V''_{444}(2x + 1, 2y) + V''_{444}(2x, 2y + 1) + V''_{444}(2x + 1, 2y + 1)],$

where the range of (x, y) is $[0, \frac{W}{2} - 1] \times [0, \frac{H}{2} - 1]$, α is a weighting factor that depends on implementation, and the " mark indicates reconstruction from (possibly lossy) coding.

5 With chroma sample grid positions centered between luma sample positions both horizontally and vertically, with the suggested anti-alias filter of $[0.5 \ 0.5]$, the value $\alpha = 1$ would perfectly reconstruct the input values in the absence of quantization error and rounding error, directly inverting the filtering performed in pre-processing. For other values of α , filtering during post-processing only approximately inverts the filtering
10 performing in pre-processing. When considering quantization error, using a somewhat smaller value of α (e.g., $\alpha = 0.5$) may be advisable in order to reduce perceptible artifacts. In general, α should be in the range from 0.0 to 1.0, and α should be smaller when the quantization step size is larger. Using a high value of α may exacerbate artifacts introduced due to lossy compression.

15 **[0138]** Or, different weights can be assigned for different sample positions. The sample values for position $U_{444}(2x, 2y)$ and $V_{444}(2x, 2y)$ of the YUV 4:4:4 frame can be calculated as $U'_{444}(2x, 2y)$ and $V'_{444}(2x, 2y)$, from values in the packed frame, as follows:

- $U'_{444}(2x, 2y) = (1 + \alpha + \beta + \gamma) * U''_{420}^{\text{main_filt}}(x, y) - \alpha * U''_{444}(2x + 1, 2y) - \beta * U''_{444}(2x, 2y + 1) - \gamma * U''_{444}(2x + 1, 2y + 1),$
- $V'_{444}(2x, 2y) = (1 + \alpha + \beta + \gamma) * V''_{420}^{\text{main_filt}}(x, y) - \alpha * V''_{444}(2x + 1, 2y) - \beta * V''_{444}(2x, 2y + 1) - \gamma * V''_{444}(2x + 1, 2y + 1),$

where the range of (x, y) is $[0, \frac{W}{2} - 1] \times [0, \frac{H}{2} - 1]$, α , β and γ are weighting factors that depend on implementation, and the " mark indicates reconstruction from (possibly lossy)
25 coding. With chroma sample grid positions centered between luma sample positions both horizontally and vertically, with the suggested anti-alias filter of $[0.5 \ 0.5]$, the values $\alpha = \beta = \gamma = 1$ would perfectly reconstruct the input values in the absence of quantization error and rounding error, directly inverting the filtering performed in pre-processing. For other values of α , β and γ , filtering during post-processing only approximately inverts the
30 filtering performing in pre-processing. When considering quantization error, using somewhat smaller values of α , β and γ (e.g., $\alpha = \beta = \gamma = 0.5$) may be advisable in order to reduce perceptible artifacts. In general, α , β and γ should be in the range from 0.0 to 1.0,

and α , β and γ should be smaller when the quantization step size is larger. Using high values of α , β and γ may exacerbate artifacts introduced due to lossy compression. The values of α , β and γ can be designed for conditional optimality using cross-correlation analysis.

- 5 **[0139]** When $\alpha = \beta = \gamma = 1$, the sample values for position $U_{444}(2x, 2y)$ and $V_{444}(2x, 2y)$ of the YUV 4:4:4 frame can simply be calculated as $U'_{444}(2x, 2y)$ and $V'_{444}(2x, 2y)$, from values in the packed frame, as follows:

- $U'_{444}(2x, 2y) = 4 * U''_{420}^{\text{main_filt}}(x, y) - U''_{444}(2x + 1, 2y) - U''_{444}(2x, 2y + 1) - U''_{444}(2x + 1, 2y + 1)$, and

10 • $V'_{444}(2x, 2y) = 4 * V''_{420}^{\text{main_filt}}(x, y) - V''_{444}(2x + 1, 2y) - V''_{444}(2x, 2y + 1) - V''_{444}(2x + 1, 2y + 1)$,

where the range of (x, y) is $[0, \frac{W}{2} - 1] \times [0, \frac{H}{2} - 1]$.

- [0140]** For example, during pre-processing, the sample values 29, 15, 7, and 18 for locations $(2x, 2y)$, $(2x+1, 2y)$, $(2x, 2y+1)$ and $(2x+1, 2y+1)$ are filtered to produce a
 15 sample value 17.25, which is rounded to 17. The value filtered sample value of 17 is used in place of the original sample value of 29. During post-processing, the sample value for the position $(2x, 2y)$ is reconstructed to be $68 - 15 - 7 - 18 = 28$. The difference between the original sample value (29) and reconstructed sample value (28) shows loss of precision due to the filtering for the pre-processing operation.

- 20 **[0141]** Alternatively, a device can selectively skip filtering operations during post-processing, even when filtering was performed during pre-processing. For example, a device can skip filtering during post-processing to reduce the computational load of decoding and playback.

- [0142]** Alternatively, the pre-processing operations and post-processing operations are
 25 not limited to the chroma signal of the 4:4:4 frame that is part of the 4:2:0 frame representing the main view (for example, areas B2 and B3 for the frame 902 represented in Figure 9). Instead, the pre-processing operations and post-processing operations are also performed for the chroma signal of the 4:4:4 frame that is part of the 4:2:0 frame representing the auxiliary view (for example, areas B4 to B9 of the frame 903 represented
 30 in Figure 9). Such pre-processing and post-processing operations (for the chroma signal of the 4:4:4 frame that is part of the 4:2:0 frame representing the auxiliary view) can use different filtering operations than the pre-processing and post-processing of the chroma signal of the 4:4:4 frame that is made part of the 4:2:0 frame representing the main view.

[0143] In the foregoing examples of pre-processing operations and post-processing operations, an averaging filtering is used during pre-processing and corresponding filtering is used during post-processing. Alternatively, the pre-processing operations and post-processing operations can implement a transform/inverse transform pair. For example, the transform/inverse transform pair can be one of the class of wavelet transformations, lifting transformations and other transformations. Specific transforms can also be designed depending on use case scenarios, so as to satisfy the different design reasons mentioned above for the use of pre-processing operations and post-processing operations in the context of packing 4:4:4 frames. Or, the pre-processing and post-processing can use other filter structures, with other filter regions of support, or use filtering that is adaptive with respect to content and/or fidelity (*e.g.*, adaptive with respect to the quantization step sizes used for the encoding).

[0144] In some example implementations, the representation and/or compression of the frame-packed 4:2:0 content can use a higher sample bit depth than the original sample bit depth of the 4:4:4 content. For example, the sample bit depth of the 4:4:4 frames is 8 bits per sample, and the sample bit depth of the frame-packed 4:2:0 frames is 10 bits per sample. This can help reduce precision loss during the application of pre-processing operations and post-processing operations. Or, this can help achieve higher level of fidelity when 4:2:0 frames are encoded using lossy compression. For example, if the 4:4:4 content has a sample bit depth of 8 bits per sample, and the frame-packed 4:2:0 content has a sample bit depth of 10 bits per sample, the bit depth of 10 bits per sample can be maintained in all or most internal modules of the encoder and decoder. The sample bit depth can be reduced to 8 bits per sample, if necessary, after unpacking the content to 4:4:4 format at the receiving end. More generally, the sample values of frames of the higher-resolution chroma sampling format can have a first bit depth (such as 8, 10, 12 or 16 bits per sample), while the sample values of frames of the lower-resolution chroma sampling format (following frame packing) have a second bit depth higher than the first bit depth.

D. Alternatives for YUV 4:2:2 Video.

[0145] In many of the foregoing examples, YUV 4:4:4 frames are packed into YUV 4:2:0 frames for encoding and decoding. In other examples, YUV 4:2:2 frames are packed into YUV 4:2:0 frames for encoding and decoding. A typical 4:2:2 frame contains 8 sample values for every 4 pixel positions, while a 4:2:0 frame contains only 6 sample

values for every 4 pixel positions. So, the sample values contained in a 4:2:2 frame can be packed into 4/3 4:2:0 frames. That is, when packed efficiently, three 4:2:2 frames can be packed into four 4:2:0 frames.

[0146] In one approach, the frame packing for 4:2:2 frames is done in a simple manner similar to the simple approach (800) illustrated in Figure 8 for 4:4:4 to 4:2:0 frame packing.

[0147] In other approaches, a YUV 4:2:2 frame is packed into YUV 4:2:0 frames while maintaining geometric correspondence for chroma information of the YUV 4:2:2 frame. The resulting YUV 4:2:0 frames with good geometric correspondence among their Y, U and V components can be compressed better because they fit the model expected by a typical encoder adapted to encoded YUV 4:2:0 frames. At the same time, the packing can be done such that a YUV 4:2:0 frame represents the complete scene being represented by YUV 4:2:2 frame, albeit with color components at a lower resolution.

[0148] These design constraints can be satisfied while packing a YUV 4:2:2 into two YUV 4:2:0 frames (main view and auxiliary view). The auxiliary view will have “empty” areas, but these areas can be filled using a fixed value or by replicating chroma values. Or, the empty areas can be used to indicate other information such as depth of a scene. For example, for the packing approach (900) described with reference to Figure 9, the approach (900) can be used as is, except that the areas B4 and B5 will not have data. Or, for the packing approach (1100) described with reference to Figure 11, the approach (1100) can be used as is, except that the areas A4 and A5 will not have data.

[0149] In example implementations, new values for content_interpretation_type are defined to signal the packing of YUV 4:2:2 frames into the constituent YUV 4:2:0 frames, as shown in the following table.

Value	Interpretation
0	Unspecified relationship between the frame packed constituent frames.
1	Indicates that the two constituent frames form the left and right views of a stereo view scene, with frame 0 being associated with the left view and frame 1 being associated with the right view.
2	Indicates that the two constituent frames form the right and left views of a stereo view scene, with frame 0 being associated with the right view and frame 1 being associated with the left view.
...	...

7	Indicates that the two constituent frames form the main and auxiliary YUV 4:2:0 frames representing a YUV 4:2:2 frame, with frame 0 being associated with the main view and frame 1 being associated with the auxiliary view. Indicates that the chroma samples of frame 0 should be interpreted as unfiltered samples of the 4:2:2 frame (without anti-alias filtering).
8	Indicates that the two constituent frames form the main and auxiliary YUV 4:2:0 frames representing a YUV 4:2:2 frame, with frame 0 being associated with the main view and frame 1 being associated with the auxiliary view. Indicates that the chroma samples of frame 0 should be interpreted as having been anti-alias filtered prior to frame packing.
9	Indicates that the two constituent frames form the main and auxiliary YUV 4:2:0 frames representing a YUV 4:2:2 frame, with frame 1 being associated with the main view and frame 0 being associated with the auxiliary view. Indicates that the chroma samples of frame 1 should be interpreted as unfiltered samples of the 4:2:2 frame (without anti-alias filtering).
10	Indicates that the two constituent frames form the main and auxiliary YUV 4:2:0 frames representing a YUV 4:2:2 frame, with frame 1 being associated with the main view and frame 0 being associated with the auxiliary view. Indicates that the chroma samples of frame 1 should be interpreted as having been anti-alias filtered prior to frame packing.

[0150] Alternatively, different values for the syntax element `content_interpretation_type` are associated with the interpretations shown in the preceding table. Or, other and/or additional interpretations for `content_interpretation_type` can be used to support encoding/decoding of frames of a lower-resolution chroma sampling format obtained from one or more frames of a higher-resolution chroma sampling format by frame packing.

E. Other Chroma Sampling Formats.

[0151] Many of the examples described herein involve variations of YUV color spaces such as Y'UV, YIQ, Y'IQ, YdbDr, YCbCr, YCoCg, etc. in sampling ratios such as 4:4:4, 4:2:2, 4:2:0, etc., as the chroma sampling formats. Alternatively, the described approaches can be used for color spaces such as RGB, GBR, etc. in sampling ratios such as 4:4:4, 4:2:2, 4:2:0, etc., as the chroma sampling formats. For example, a device can pack frames

of a higher-resolution non-YUV chroma sampling format (such as RGB 4:4:4 or GBR 4:4:4) into frames of a lower resolution format (such as a 4:2:0 format), which may then be encoded. In the encoding, the nominally luma component and nominally chroma components represent sample values of the non-YUV components (rather than

5 approximate brightness and color-difference values). In corresponding unpacking, a device unpacks frames of the lower resolution format (such as a 4:2:0 format) into frames of the higher-resolution non-YUV chroma sampling format (such as RGB 4:4:4 or GBR 4:4:4).

[0152] Also, the described approaches can be used for frame packing of video content of
10 a 4:4:4 format, 4:2:2 format or 4:2:0 format into a 4:0:0 format, which is typically used for gray scale or monochrome video content. The chroma information from a frame of the 4:4:4 format, 4:2:2 format or 4:2:0 format can be packed into the primary component of one or more additional or auxiliary frames of 4:0:0 format.

15 **F. Generalized Techniques for Frame Packing/Unpacking.**

[0153] Figure 12 shows a generalized technique (1200) for frame packing. A computing device that implements a frame packer, for example, as described with reference to Figure 4, can perform the technique (1200).

[0154] The device packs (1210) one or more frames of a higher-resolution chroma
20 sampling format into one or more frames of a lower-resolution chroma sampling format. For example, the device packs frame(s) of 4:4:4 format (*e.g.*, YUV 4:4:4 format) into frame(s) of 4:2:0 format (*e.g.*, YUV 4:2:0 format). Or, the device packs frame(s) of 4:2:2 format (*e.g.*, YUV 4:2:2 format) into frame(s) of 4:2:0 format (*e.g.*, YUV 4:2:0 format). Or, the device packs frame(s) of 4:4:4 format (*e.g.*, YUV 4:4:4 format) into frame(s) of
25 4:2:2 format (*e.g.*, YUV 4:2:2 format).

[0155] For YUV formats, the device can perform the frame packing (1210) so as to maintain geometric correspondence between adjacent sample values of chroma components of the frame(s) of the higher-resolution chroma sampling format after the packing. For example, such sample values are maintained as adjacent samples and/or
30 collocated portions of luma and chroma components of the frame(s) of the lower-resolution chroma sampling format. Later encoding can exploit such geometric correspondence.

[0156] In some frame packing approaches, the device can embed a lower chroma resolution version of the frame(s) of the higher-resolution chroma sampling format as part

of the frame(s) of the lower-resolution chroma sampling format. Thus, part of the frame(s) of the lower-resolution chroma sampling format represents a lower chroma resolution version of the frame(s) of the higher-resolution chroma sampling format. The rest of the frame(s) of the lower-resolution chroma sampling format represents remaining chroma information from the frame(s) of the higher-resolution chroma sampling format. In other frame packing approaches, according to spatial partitioning of the frame(s) of the higher-resolution chroma sampling format, the device assigns sample values of chroma components of the frame(s) of the higher-resolution chroma sampling format to luma and chroma components of the frame(s) of the lower-resolution chroma sampling format.

[0157] During the frame packing, the sample values of chroma components of the frame(s) of the higher-resolution chroma sampling format can be filtered, and filtered sample values are assigned to parts of chroma components of the frame(s) of the lower-resolution chroma sampling format. In some implementations, the sample values of the chroma components of the frame(s) of the higher-resolution chroma sampling format have a lower bit depth (e.g., 8 bits per sample), and the filtered sample values have a higher bit depth (e.g., 10 bits per sample) for encoding at the higher bit depth.

[0158] The device can then encode (1220) the frame(s) of the lower-resolution chroma sampling format. Alternatively, a different device performs the encoding (1220). The device(s) can repeat the technique (1200) on a frame-by-frame basis or other basis.

[0159] The device can signal metadata about frame packing/unpacking. For example, the device signals metadata that indicates whether frame packing/unpacking is used or not used. Or, the device signals an indication that the sample values of the chroma components of the frame(s) of the higher-resolution chroma sampling format have been filtered during the frame packing, and should be filtered as part of post-processing. The metadata about frame packing/unpacking can be signaled as part of a supplemental enhancement information message or as some other type of metadata.

[0160] Figure 13 shows a generalized technique (1300) for frame unpacking. A computing device that implements a frame unpacker, for example, as described with reference to Figure 5, can perform the technique (1300).

[0161] Before the frame unpacking itself, the device can decode (1310) the frame(s) of a lower-resolution chroma sampling format. Alternatively, a different device performs the decoding (1310).

[0162] The device unpacks (1320) one or more frames of the lower-resolution chroma sampling format into one or more frames of a higher-resolution chroma sampling format.

For example, the device unpacks frame(s) of 4:2:0 format (*e.g.*, YUV 4:2:0 format) into frame(s) of 4:4:4 format (*e.g.*, YUV 4:4:4 format). Or, the device unpacks frame(s) of 4:2:0 format (*e.g.*, YUV 4:2:0 format) into frame(s) of 4:2:2 format (*e.g.*, YUV 4:2:2 format). Or, the device unpacks frame(s) of 4:2:2 format (*e.g.*, YUV 4:2:2 format) into frame(s) of 4:4:4 format (*e.g.*, YUV 4:4:4 format).

[0163] When a lower chroma resolution version of the frame(s) of the higher-resolution chroma sampling format is embedded as part of the frame(s) of the lower-resolution chroma sampling format, the device has options for display. The part of the frame(s) of the lower-resolution chroma sampling format that represents a lower chroma resolution version of the frame(s) of the higher-resolution chroma sampling format can be reconstructed for output and display. The rest of the frame(s) of the lower-resolution chroma sampling format represents remaining chroma information from the frame(s) of the higher-resolution chroma sampling format, and can be used as part of frame unpacking. In other frame unpacking approaches, to reverse spatial partitioning of the frame(s) of the higher-resolution chroma sampling format, the device assigns sample values of luma and chroma components of the frame(s) of the lower-resolution chroma sampling format to chroma components of the frame(s) of the higher-resolution chroma sampling format.

[0164] During the frame unpacking, the sample values of chroma components of the frame(s) of the higher-resolution chroma sampling format can be filtered as part of post-processing. In some implementations, at least some sample values of the chroma components of the frame(s) of the higher-resolution chroma sampling format have a higher bit depth (*e.g.*, 10 bits per sample) before the post-processing filtering, and such sample values have a lower bit depth (*e.g.*, 8 bits per sample) after the post-processing filtering.

[0165] The device can also receive metadata about frame packing/unpacking. For example, the device receives metadata that indicates whether frame packing/unpacking is used or not used. Or, the device receives an indication that the sample values of the chroma components of the frame(s) of the higher-resolution chroma sampling format have been filtered during the frame packing, and should be filtered as part of post-processing. The metadata about frame packing/unpacking can be signaled as part of a supplemental enhancement information message or as some other type of metadata.

[0166] The device(s) can repeat the technique (1300) on a frame-by-frame basis or other basis.

G. Band Separation Filtering/Inverse Filtering as Pre-processing/Post-processing Operations for Frame Packing/Unpacking.

[0167] According to a second set of example pre-processing and post-processing operations, band separation filtering and inverse filtering are used for pre-processing and post-processing, respectively. In many examples, the band separation filtering and inverse filtering are wavelet decomposition and reconstruction. Frame packing can include wavelet decomposition (analysis) as part of pre-processing operations to filter chroma sample values. Corresponding frame unpacking then includes wavelet reconstruction (synthesis) as part of post-processing to inverse filter the chroma sample values.

[0168] For example, in these approaches a YUV 4:4:4 frame can be packed into two YUV 4:2:0 frames while maintaining geometric correspondence for chroma information of the YUV 4:4:4 frame. YUV 4:2:0 frames with good geometric correspondence among their Y, U and V components can be compressed better because they fit the model expected by a typical encoder adapted to encode YUV 4:2:0 frames. As in some approaches described above, the packing can also be done such that one of the two YUV 4:2:0 frames represents the complete scene being represented by YUV 4:4:4 frame, albeit with color components at a lower resolution. This provides options in decoding. A decoder that cannot perform frame unpacking, or chooses not to perform frame unpacking, can just take a reconstructed version of the YUV 4:2:0 frame that represents the scene and directly feed it to the display.

[0169] At the same time, the auxiliary frame (with the remaining chroma information from the frames of the higher-resolution chroma sampling format) can be considered an enhancement layer signal to be combined with the main frame. Primary signal energy can be concentrated into the main frame (by dedicating more bits to the main frame during encoding), with the auxiliary frame consuming an arbitrarily low number of bits for the enhancement layer signal during encoding. When the auxiliary frame is encoded at a lower bit rate (lower quality relative to the main frame), the chroma information from the main frame sets the minimum level of quality for the chroma components after reconstruction, and any information from the auxiliary frame can be used to improve quality beyond that minimum quality level.

1. Framework for 3-Band Wavelet Decomposition for 4:4:4 to 4:2:0.

[0170] In one set of band separation filtering approaches, pre-processing operations include three-band wavelet decomposition for frame packing of video content of a 4:4:4 format into a 4:2:0 format. For example, suppose the height H of three original arrays of sample values in YUV 4:4:4 format is a multiple of 4, and suppose the width W of the three original arrays of sample values in YUV 4:4:4 format is also a multiple of 4.

Figure 14 illustrates an approach (1400) to frame packing with three-band wavelet decomposition as pre-processing. In this approach (1400), a YUV 4:4:4 frame (801) is packed into two YUV 4:2:0 frames (1402, 1403). The first frame (1402) provides a “main view” in YUV 4:2:0 format – a lower chroma resolution version of the complete scene represented by the YUV 4:4:4 frame (801). The second frame (1403) provides an “auxiliary view” in YUV 4:2:0 format and contains remaining chroma information.

[0171] For an array A (where A is U chroma sample values or V chroma sample values), the wavelet decomposition for a first stage is defined as:

$$\bullet \quad C_A(x, y) = A_{444}(x, 2y) + A_{444}(x, 2y + 1), \text{ for } x = 0, \dots, W - 1, y = 0, \dots, H / 2 - 1;$$

and

$$\bullet \quad D_A(x, y) = A_{444}(x, 2y) - A_{444}(x, 2y + 1), \text{ for } x = 0, \dots, W - 1, y = 0, \dots, H / 2 - 1.$$

[0172] In Figure 14, the plane U is decomposed into bands C_U and D_U . The plane V is decomposed into bands C_V and D_V . Thus, C and D are a vertically lowpass filtered version and vertically highpass filtered version, respectively, for the plane U or the plane V . After vertical filtering, the respective vertical lowpass bands (C_U and C_V) are further decomposed by horizontal wavelet decomposition.

$$\bullet \quad E_A(x, y) = C_A(2x, y) + C_A(2x + 1, y), \text{ for } x = 0, \dots, W / 2 - 1, y = 0, \dots, H / 2 - 1;$$

$$\bullet \quad F_A(x, y) = C_A(2x, y) - C_A(2x + 1, y), \text{ for } x = 0, \dots, W / 2 - 1, y = 0, \dots, H / 2 - 1.$$

[0173] In Figure 14, the lowpass band C_U is decomposed into bands E_U and F_U . The lowpass band C_V is decomposed into bands E_V and F_V . Thus, E and F are a horizontally lowpass filtered version and horizontally highpass filtered version, respectively, of a vertically lowpass filtered band. E and F can be referred to as LL and LH wavelet bands. After the wavelet decomposition (including normalization, as appropriate, as shown below), the respective bands (D_U , D_V , E_U , E_V , F_U and F_V) are arranged as sections of the YUV 4:2:0 frame(s).

$$\bullet \quad Y_{420}^{\text{main}}(x, y) = Y_{444}(x, y), \text{ for } x = 0, \dots, W - 1, y = 0, \dots, H - 1;$$

$$\bullet \quad U_{420}^{\text{main}}(x, y) = E_U(x, y)/4, \text{ for } x = 0, \dots, W / 2 - 1, y = 0, \dots, H / 2 - 1;$$

- $V_{420}^{\text{main}}(x, y) = E_V(x, y)/4$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/2 - 1$;
- $Y_{420}^{\text{aux}}(x, y) = D_U(x, y)/2 + 2^{B-1}$, for $x = 0, \dots, W - 1, y = 0, \dots, H/2 - 1$;
- $Y_{420}^{\text{aux}}(x, H/2 + y) = D_V(x, y)/2 + 2^{B-1}$, for $x = 0, \dots, W - 1, y = 0, \dots, H/2 - 1$;
- $U_{420}^{\text{aux}}(x, y) = F_U(x, 2y)/4 + 2^{B-1}$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/4 - 1$;
- 5 • $U_{420}^{\text{aux}}(x, H/4 + y) = F_V(x, 2y)/4 + 2^{B-1}$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/4 - 1$;
- $V_{420}^{\text{aux}}(x, y) = F_U(x, 2y + 1)/4 + 2^{B-1}$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/4 - 1$; and
- $V_{420}^{\text{aux}}(x, H/4 + y) = F_V(x, 2y + 1)/4 + 2^{B-1}$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/4 - 1$.

where B indicates bit depth. In this example, it is assumed that the bit depth is the same for the input sample values and output sample values of the pre-processing. The bit depth can vary between input sample values and the results of pre-processing, however, in some implementations. Generally, the sample values of frames of the higher-resolution chroma sampling format can have a first bit depth (such as 8, 10, 12 or 16 bits per sample), while the sample values of frames of the lower-resolution chroma sampling format (following frame packing) have a second bit depth higher than the first bit depth. Also, the pre-processing operations can include clipping of some signals to the range from 0 to $2^B - 1$ in some cases.

[0174] Alternatively, the sample values can be assigned to the areas in a different way. For example, data from the original U plane of the YUV 4:4:4 frame can be arranged in the U plane of the auxiliary YUV 4:2:0 frame, and data from the original V plane of the YUV 4:4:4 frame can be arranged in the V plane of the auxiliary YUV 4:2:0 frame. In this example, compared to Figure 14, the sample values from the odd rows of F_U are assigned to the lower half of U_{420}^{aux} , and the sample values from even rows of F_V are assigned to the top half of V_{420}^{aux} . Or, the sample values F_U are simply assigned to U_{420}^{aux} , and the sample values from F_V are simply assigned to V_{420}^{aux} . Either way, the U plane (or V plane) of the auxiliary YUV 4:2:0 frame is constructed from the U plane (or V plane) of the YUV 4:4:4 frame, without mixing content from different original U and V planes.

[0175] In summary, the bands D, E, and F form a three-band wavelet decomposition of the input array A, which may represent chroma sample values for plane U or V. Relative to more typical four-band wavelet decomposition into LL, LH, HL, and HH bands, the HL and HH bands are not created in the three-band decomposition. Instead, the vertically highpass signal D is kept at full horizontal resolution. As such, the bands D, E, and F can

be called the H, LL and LH bands. In this naming convention, the first letter indicates a vertical highpass (H) or lowpass (L) signal decimation, and the second letter (when present) indicates a horizontal highpass or lowpass signal decimation.

[0176] The first frame (1402) and second frame (1403) of the YUV 4:2:0 format can be

5 organized as separate frames (separated by the dark line in Figure 14). Or, the first frame (1402) and second frame (1403) of the YUV 4:2:0 format can be organized as a single frame having a height of $2 \times H$ (ignoring the dark line in Figure 14). Or, the first frame (1402) and second frame (1403) of the YUV 4:2:0 format can be organized as a single frame having a width of $2 \times W$. Or, the first frame (1402) and second frame (1403) of the
10 YUV 4:2:0 format can be organized as a single frame using any of the methods defined for frame_packing_arrangement_type in the H.264/AVC standard or the HEVC standard.

[0177] Figure 15 shows another example approach (1500) to frame packing with three-

band wavelet decomposition as pre-processing. In this approach (1500), a YUV 4:4:4 frame (801) is packed into two YUV 4:2:0 frames (1502, 1503). Again, the first frame

15 (1502) provides a “main view” in YUV 4:2:0 format, while the second frame (1503)

provides an “auxiliary view” in YUV 4:2:0 format and contains remaining chroma

information. Unlike the approach illustrated with reference to Figure 14 (which is

horizontally-oriented decomposition, such that Y_{420}^{aux} contains full-width, half-height U and

V regions), Figure 15 illustrates a corresponding vertically-oriented case in which Y_{420}^{aux}

20 contains half-width, full-height U and V regions. For the approach shown in Figure 15, the ordering of vertical decomposition and horizontal decomposition is switched.

[0178] Again, the sample values can be assigned to the areas of Figure 15 in a different way. For example, data from the original U plane of the YUV 4:4:4 frame can be

arranged in the U plane of the auxiliary YUV 4:2:0 frame, and data from the original V

25 plane of the YUV 4:4:4 frame can be arranged in the V plane of the auxiliary YUV 4:2:0

frame. In this example, compared to Figure 15, the sample values from the odd columns

of F_U are assigned to the right half of U_{420}^{aux} , and the sample values from even columns of

F_V are assigned to the left half of V_{420}^{aux} . Or, the sample values F_U are simply assigned to

U_{420}^{aux} , and the sample values from F_V are simply assigned to V_{420}^{aux} . Either way, the U

30 plane (or V plane) of the auxiliary YUV 4:2:0 frame is constructed from the U plane (or V

plane) of the YUV 4:4:4 frame, without mixing content from different original U and V

planes.

[0179] In this section, the wavelet decomposition uses a lowpass filter (“LPF”) based on the filter $LPF = [1 \ 1] / 2$, and uses a highpass filter (“HPF”) that is based on the filter $HPF = [1 \ -1] / 2$. This LPF-HPF filter pair is sometimes known as a Haar wavelet filter pair.

In this notation, the numbers in brackets indicate filter taps, and the denominator indicates a normalization factor. The filter pair is applied both horizontally and vertically, with a centered sampling phase horizontally and vertically (mid-point values derived by filtering), and with the division by two from each stage postponed to follow both the horizontal and vertical filtering. Alternatively, as described below, the wavelet decomposition uses a filter pair that differs in terms of filter taps, filter phase, normalization factor, timing of normalization, use of rounding, use of clipping, etc.

2. Framework for 4-Band Wavelet Decomposition for 4:4:4 to 4:2:0.

[0180] In another set of band separation filtering approaches, pre-processing operations include four-band wavelet decomposition for frame packing of video content of a 4:4:4 format into a 4:2:0 format. Again, suppose the height H of three original arrays of sample values in YUV 4:4:4 format is a multiple of 4, and suppose the width W of the three original arrays of sample values in YUV 4:4:4 format is also a multiple of 4.

[0181] In the approaches illustrated with reference to Figure 14, the vertically lowpass filtered/horizontally highpass filtered band $F_A(x, y)$ is vertically decimated when assigning sample values to U_{420}^{aux} and V_{420}^{aux} . Alternate rows are assigned to U_{420}^{aux} and V_{420}^{aux} from $F_A(x, y)$. This assignment happens without first applying anti-alias filtering.

[0182] Alternatively, additional filtering is applied to sample values of the band $F_A(x, y)$, as illustrated in Figure 16a, which continues the first part of Figure 14 but includes a different first frame (1602) and second frame (1603) of YUV 4:2:0 format. In particular, the band $F_A(x, y)$ is further decomposed by vertical wavelet decomposition into a lowpass band $G_A(x, y)$ and highpass band $H_A(x, y)$.

- $G_A(x, y) = F_A(x, 2y) + F_A(x, 2y + 1)$, for $x = 0, \dots, W / 2 - 1, y = 0, \dots, H / 4 - 1$;
- $H_A(x, y) = F_A(x, 2y) - F_A(x, 2y + 1)$, for $x = 0, \dots, W / 2 - 1, y = 0, \dots, H / 4 - 1$.

[0183] After the wavelet decomposition (including normalization, as appropriate), some of the respective bands (D_U, D_V, E_U and E_V) are arranged as sections of the YUV 4:2:0 frame(s) as described with reference to Figure 14. The remaining bands (G_U, G_V, H_U and H_V) are arranged as sections of the YUV 4:2:0 frame(s) as follows.

- $U_{420}^{aux}(x, y) = G_U(x, y) / 8 + 2^{B-1}$, for $x = 0, \dots, W / 2 - 1, y = 0, \dots, H / 4 - 1$;

- $U_{420}^{aux}(x, H/4 + y) = G_V(x, y)/8 + 2^{B-1}$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/4 - 1$;
- $V_{420}^{aux}(x, y) = H_U(x, y)/8 + 2^{B-1}$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/4 - 1$; and
- $V_{420}^{aux}(x, H/4 + y) = H_V(x, y)/8 + 2^{B-1}$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/4 - 1$.

[0184] Alternatively, the sample values can be assigned to the areas in a different way.

5 Figure 16b, for example, continues the first part of Figure 14 but includes a different first frame (1604) and second frame (1605) of YUV 4:2:0 format. As in the example of Figure 16a, some of the respective bands (D_U, D_V, E_U and E_V) are arranged as sections of the YUV 4:2:0 frame(s) as described with reference to Figure 14. The remaining bands (G_U, G_V, H_U and H_V) are arranged as sections of the YUV 4:2:0 frame(s) as follows:

- 10
- $U_{420}^{aux}(x, y) = G_U(x, y)/8 + 2^{B-1}$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/4 - 1$;
 - $U_{420}^{aux}(x, H/4 + y) = H_U(x, y)/8 + 2^{B-1}$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/4 - 1$;
 - $V_{420}^{aux}(x, y) = G_V(x, y)/8 + 2^{B-1}$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/4 - 1$; and
 - $V_{420}^{aux}(x, H/4 + y) = H_V(x, y)/8 + 2^{B-1}$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/4 - 1$.

In Figure 16b, data from the original U plane of the YUV 4:4:4 frame are arranged in the U plane of the auxiliary YUV 4:2:0 frame, and data from the original V plane of the YUV 4:4:4 frame are arranged in the V plane of the auxiliary YUV 4:2:0 frame. Thus, the U plane (or V plane) of the auxiliary YUV 4:2:0 frame is constructed from the U plane (or V plane) of the YUV 4:4:4 frame, without mixing content from different original U and V planes. (In contrast, in the example of Figure 16a, the U plane (or V plane) of the auxiliary YUV 4:2:0 frame has a mixture of data from the U and V components of the YUV 4:4:4 frame. The upper half of the U plane (or V plane) of the auxiliary YUV 4:2:0 frame contains data from the original U plane, and the lower half contains data from the original V plane.)

[0185] In summary, the bands D, E, G, and H form a four-band wavelet decomposition of the input array A, which may represent chroma sample values for plane U or V. The bands D, E, G, and H correspond to H, LL, LHL and LHH bands. In this naming convention, the first letter indicates a first vertical highpass (H) or lowpass (L) signal decimation, the second letter (when present) indicates a horizontal highpass or lowpass signal decimation, and the third letter (when present) indicates an additional vertical highpass or lowpass signal decimation of the LH band.

[0186] Similarly, in the approaches illustrated with reference to Figure 15, the horizontally lowpass filtered/vertically highpass filtered band $F_A(x, y)$ is horizontally

decimated when assigning sample values to U_{420}^{aux} and V_{420}^{aux} . Alternate columns are assigned to U_{420}^{aux} and V_{420}^{aux} from $F_A(x, y)$. This assignment happens without first applying anti-alias filtering. Alternatively, additional filtering can be applied to sample values of the band $F_A(x, y)$. In particular, the band $F_A(x, y)$ is further decomposed by horizontal wavelet decomposition into a lowpass band $G_A(x, y)$ and highpass band $H_A(x, y)$. Some of the respective bands (D_U , D_V , E_U and E_V) are arranged as sections of the YUV 4:2:0 frame(s) as described with reference to Figure 15. The remaining bands (G_U , G_V , H_U and H_V) are arranged in the U_{420}^{aux} and V_{420}^{aux} bands of the YUV 4:2:0 frame(s).

[0187] In this section, the wavelet decomposition uses a Haar wavelet filter pair. The filter pair is applied horizontally and vertically, with a centered sampling phase horizontally and vertically (mid-point values derived by filtering), and with division by two from each stage postponed to follow the final filtering stage. Alternatively, as described below, the wavelet decomposition uses a filter pair that differs in terms of filter taps, filter phase, normalization factor, timing of normalization, use of rounding, use of clipping, etc.

3. Framework for Wavelet Decomposition for 4:4:4 to 4:2:2.

[0188] In another set of band separation filtering approaches, pre-processing operations include wavelet decomposition for frame packing of video content of a 4:4:4 format into a 4:2:2 format. For example, for an array A (where A is U chroma sample values or V chroma sample values), the wavelet decomposition for a first stage is defined as:

- $C_A(x, y) = A_{444}(2x, y) + A_{444}(2x + 1, y)$, for $x = 0, \dots, W / 2 - 1, y = 0, \dots, H - 1$; and
- $D_A(x, y) = A_{444}(2x, y) - A_{444}(2x + 1, y)$, for $x = 0, \dots, W / 2 - 1, y = 0, \dots, H - 1$.

[0189] Thus, the plane U is decomposed into bands C_U and D_U , and the plane V is decomposed into bands C_V and D_V . C and D are a horizontally lowpass filtered version and horizontally highpass filtered version, respectively, for the plane U or the plane V . After horizontal filtering, the respective bands (C_U , C_V , D_U and D_V) are arranged as sections of the YUV 4:2:2 frame(s) as follows.

- $Y_{422}^{main}(x, y) = Y_{444}(x, y)$, for $x = 0, \dots, W - 1, y = 0, \dots, H - 1$;
- $U_{422}^{main}(x, y) = C_U(x, y)/2$, for $x = 0, \dots, W / 2 - 1, y = 0, \dots, H - 1$;
- $V_{422}^{main}(x, y) = C_V(x, y)/2$, for $x = 0, \dots, W / 2 - 1, y = 0, \dots, H - 1$;

- $Y_{422}^{aux}(x, y) = 2^{B-1}$, for $x = 0, \dots, W - 1, y = 0, \dots, H - 1$;
- $U_{422}^{aux}(x, y) = D_U(x, y)/2 + 2^{B-1}$, for $x = 0, \dots, W / 2 - 1, y = 0, \dots, H - 1$; and
- $V_{422}^{aux}(x, y) = D_V(x, y)/2 + 2^{B-1}$, for $x = 0, \dots, W / 2 - 1, y = 0, \dots, H - 1$.

[0190] In this approach, the "fabricated" signal $Y_{420}^{aux}(x, y)$ can simply be discarded after decoding. Alternatively, the sample values of the highpass filtered version can be assigned to the luma component $Y_{420}^{aux}(x, y)$ of the auxiliary frame, while the chroma components $U_{422}^{aux}(x, y)$ and $V_{422}^{aux}(x, y)$ of the auxiliary frame are assigned fabricated values that can be discarded after decoding.

- $Y_{422}^{aux}(x, y) = D_U(x, y)/2 + 2^{B-1}$, for $x = 0, \dots, W / 2 - 1, y = 0, \dots, H - 1$;
- $Y_{422}^{aux}(W/2 + x, y) = D_V(x, y)/2 + 2^{B-1}$, for $x = 0, \dots, W / 2 - 1, y = 0, \dots, H - 1$;
- $U_{422}^{aux}(x, y) = 2^{B-1}$, for $x = 0, \dots, W / 2 - 1, y = 0, \dots, H - 1$; and
- $V_{422}^{aux}(x, y) = 2^{B-1}$, for $x = 0, \dots, W / 2 - 1, y = 0, \dots, H - 1$.

[0191] In this section, the wavelet decomposition uses a Haar wavelet filter pair. The filter pair is applied horizontally, with a centered sampling phase (mid-point values derived by filtering). Alternatively, as described below, the wavelet decomposition uses a filter pair that differs in terms of filter taps, filter phase, normalization factor, timing of normalization, use of rounding, use of clipping, etc.

4. Framework for Wavelet Decomposition for 4:2:2 to 4:2:0.

[0192] In another set of band separation filtering approaches, pre-processing operations include wavelet decomposition for frame packing of video content of a 4:2:2 format into a 4:2:0 format. In this case, the width of the original U and V arrays is $W / 2$, and the height of all three original arrays is H. For example, for an array A (where A is U chroma sample values or V chroma sample values), the wavelet decomposition for a first stage is defined as:

- $C_A(x, y) = A_{422}(x, 2y) + A_{422}(x, 2y + 1)$, for $x = 0, \dots, W / 2 - 1, y = 0, \dots, H / 2 - 1$; and
- $D_A(x, y) = A_{422}(x, 2y) - A_{422}(x, 2y + 1)$, for $x = 0, \dots, W / 2 - 1, y = 0, \dots, H / 2 - 1$.

[0193] Thus, the plane U is decomposed into bands C_U and D_U , and the plane V is decomposed into bands C_V and D_V . C and D are a vertically lowpass filtered version and vertically highpass filtered version, respectively, for the plane U or the plane V. After

vertical filtering (including normalization, as appropriate), the respective bands (C_U , C_V , D_U and D_V) are arranged as sections of the YUV 4:2:0 frame(s), as follows.

- $Y_{420}^{\text{main}}(x, y) = Y_{422}(x, y)$, for $x = 0, \dots, W - 1, y = 0, \dots, H - 1$;
- $U_{420}^{\text{main}}(x, y) = C_U(x, y)/2$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/2 - 1$;
- 5 • $V_{420}^{\text{main}}(x, y) = C_V(x, y)/2$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/2 - 1$;
- $Y_{420}^{\text{aux}}(x, y) = 2^{B-1}$, for $x = 0, \dots, W - 1, y = 0, \dots, H - 1$;
- $U_{420}^{\text{aux}}(x, y) = D_U(x, y)/2 + 2^{B-1}$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/2 - 1$; and
- $V_{420}^{\text{aux}}(x, y) = D_V(x, y)/2 + 2^{B-1}$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/2 - 1$.

[0194] In this approach, the "fabricated" signal $Y_{420}^{\text{aux}}(x, y)$ can simply be discarded after
 10 decoding. Alternatively, the sample values of the highpass filtered version can be assigned to the upper half of the luma component $Y_{420}^{\text{aux}}(x, y)$ of the auxiliary frame, while the chroma components $U_{422}^{\text{aux}}(x, y)$ and $V_{422}^{\text{aux}}(x, y)$ of the auxiliary frame, and lower half of $Y_{420}^{\text{aux}}(x, y)$, are assigned fabricated values that can be discarded after decoding.

- $Y_{420}^{\text{aux}}(x, y) = D_U(x, y)/2 + 2^{B-1}$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/2 - 1$;
- 15 • $Y_{420}^{\text{aux}}(W/2 + x, y) = D_V(x, y)/2 + 2^{B-1}$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/2 - 1$;
- $Y_{420}^{\text{aux}}(x, H/2 + y) = 2^{B-1}$, for $x = 0, \dots, W - 1, y = 0, \dots, H/2 - 1$;
- $U_{420}^{\text{aux}}(x, y) = 2^{B-1}$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/2 - 1$; and
- $V_{420}^{\text{aux}}(x, y) = 2^{B-1}$, for $x = 0, \dots, W/2 - 1, y = 0, \dots, H/2 - 1$.

[0195] In this section, the wavelet decomposition uses a Haar wavelet filter pair. The
 20 filter pair is applied vertically, with a centered sampling phase (mid-point values derived by filtering). Alternatively, as described below, the wavelet decomposition uses a filter pair that differs in terms of filter taps, filter phase, normalization factor, timing of normalization, use of rounding, use of clipping, etc.

25 5. Example Implementations of Filters.

[0196] In some examples in the preceding sections, wavelet decomposition uses a Haar wavelet filter pair in which $\text{LPF} = [1 \ 1] / 2$, and $\text{HPF} = [1 \ -1] / 2$. The numbers in brackets indicate filter taps, and the denominator indicates a normalization factor. In the preceding sections, the filter pair is applied horizontally and/or vertically, with a centered

sampling phase where filtering is performed (mid-point values derived by filtering), and with the division by two from each stage postponed to follow the final filtering stage.

[0197] More generally, the filter used for wavelet decomposition or other band

separation filtering depends on implementation. The filters used can vary in terms of (a)

- 5 filter taps, (b) normalization factor, (c) how normalization occurs (*e.g.*, after each filtering stage, or deferred partially or completely for one or more filtering stages), (d) whether bit depth expansion is permitted (*e.g.*, by skipping or reducing normalization to provide scaling for one or more filtering stages), (e) how division is implemented (*e.g.*, with arithmetic right shift operations or integer division), (f) how rounding is applied (*e.g.*, no
- 10 rounding, with nearest-integer rounding, with dithered rounding), (g) whether clipping is used and/or (h) another factor. Ideally, a filter pair has low complexity but also provides good performances in terms of decorrelation and hence compression.

Implementation Options for Filter Taps, Normalization Factors and Timing of

Normalization

[0198] As shown in the preceding sections, the filter taps of a filter pair can be based on the Haar wavelet. Alternatively, a different filter is used for wavelet decomposition or other band separation filtering during pre-processing. For example, wavelet

decomposition can use a symmetric biorthogonal Daubechies wavelet filter pair. The filter

20 pair $\text{LPF} = [-1 \ 1 \ 8 \ 8 \ 1 \ -1] / 16$, $\text{HPF} = [-1 \ 1] / 2$ potentially facilitates better compression efficiency. This filter has the same sampling phase (mid-point values derived by filtering).

Or, as another example, for filtering in which a filtering stage should yield co-sited sample values vertically and/or horizontally (aligning with input sample values), wavelet

decomposition can use the filter pair $\text{LPF} = [-1 \ 2 \ 6 \ 2 \ -1] / 8$, $\text{HPF} = [-1 \ 2 \ -1] / 4$. Or, a

25 filter pair from another family of filters can be used. For example, the filter pair can be based on quadrature mirror filters, although this may involve filters that do not achieve the property of perfect signal reconstruction in the absence of quantization and rounding error.

Or, as another example, the filter pair can be based on polyphase quadrature filters, which are used in AAC audio coding/decoding. Or, as still another example, the filter pair can be

30 based on conjugate quadrature filters, which are the basis of many orthogonal filter banks including those using Daubechies orthogonal wavelets. Any of the filter pairs can be scaled or unscaled, depending on implementation.

[0199] The normalization factor generally depends on the filter taps and is conventionally expressed as a denominator in the definition of a filter. Depending on

implementation, the normalization factor can be implemented as a division after each filtering stage. Or, some of the normalization can be partially or completely deferred for one or more filtering stages, such that division at a later stage (e.g., final stage) compensates for expansion introduced at that stage and also accounts for deferred

5 normalization.

[0200] Or, the wavelet decomposition or other band separation filtering can permit some amount of bit depth expansion in the sample values of the bands, so as to reduce rounding/truncation error and potentially permit exact invertibility. For example, to allow scaling of the generated signals that result from application of a Haar filter pair, division

10 by 2 can be skipped when forming arrays from the sample values of the D_A arrays, or division by 2 can be used instead of division by 4 when forming arrays from the sample values of the F_A arrays. (In such cases, clipping can be applied to avoid violation of the 0 to $2^B - 1$ range, if appropriate.)

15 ***Implementation Options for Division Operations and Rounding***

[0201] For a lowpass filter or highpass filter, the division operation may include nearest-integer rounding. When using integer arithmetic, a typical way to do this, with a numerator equal to some value n (ordinarily represented as a two's complement binary number) and a denominator that is equal to 2^k for some $k > 1$, is to perform the operation

20 $(n + 2^{(k-1)}) \gg k$, where " $\gg$ " denotes an arithmetic right shift operation.

[0202] When performing division operations for filtering, dithered rounding or other optimized rounding can be applied. For example, a rounding factor can be included when dividing by the k^{th} power of two (i.e., division by 2^k), taking the form $(n + 2^{(k-1)} - p) \gg k$, where the value p is a rounding factor. Different patterns of values can be used for p . For

25 example, the value of p alternates between 0 and 1 following a 2D periodic pattern such as a repeating $M \times M$ block pattern. Or, the value of p alternates between 0 and 1 following a pseudo-random basis. Or, the value of p alternates between 0 and 1 following a "blue noise dither" signal or other dither signal.

[0203] In terms of timing, as noted above, normalization (and hence division and

30 rounding) can be deferred until a final stage. Alternatively, the filtering stages could be cascaded with rounding applied between the two stages, e.g., such that intermediate rounding happens between vertical and horizontal filtering stages.

Implementation Options for Clipping, Offsets and Range of Values

[0204] In some implementations, the pre-processing operations include clipping to prevent dynamic range expansion. Overall, clipping can simplify implementation, but it may introduce distortion when processing some pathological input signal values. For example, the LPF kernel $[-1 \ 1 \ 8 \ 8 \ 1 \ -1] / 16$ could produce negative output values when processing unsigned input data. Since image and video signals are typically highly statistically correlated, however, it is usually very rare for such a negative output value to occur in practice. In fact, it may be so rare that simply clipping the output of the LPF to not allow negative output values is a sensible design choice.

[0205] In many of the preceding examples, the sample values of Y, U, and V signals have a range from 0 to $2^B - 1$. Alternatively, sample values can have some other range. For example, sample values have a range from $-2^{(B-1)}$ to $2^{(B-1)} - 1$. With this example alternative signal range, the offsets of $2^{(B-1)}$ are not used in the above equations.

Implementation Options for Edge Handling

[0206] For any of the approaches that use wavelet decomposition or other band separation filtering, the pre-processing operations can include edge-handling stages. For example, edge values can be padded outward from an edge. Or, edge values can be mirrored at an edge. Or, values can be repeated using modulo/circular wrapping.

Implementation Options for Lifting, with or without Clipping

[0207] For the approaches that use wavelet decomposition, lifting may be used to reduce rounding/truncation error and potentially permit exact invertibility. Lifting can be used in combination with clipping.

[0208] In some implementations, for example, the filters of a Haar wavelet filter pair are adapted to use lifting and optionally clipping. In the example described with reference to Figure 14, band splitting is first performed vertically (in the y dimension) to create the lowpass signal $C_A(x, y)$ and highpass signal $D_A(x, y)$. If the input values $A_{444}(x, 2y)$ and $A_{444}(x, 2y + 1)$ have a dynamic range of B bits (where A can be sample values of the U or V chroma component), the dynamic range of the values of $C_A(x, y)$ and $D_A(x, y)$ will be (approximately) $B + 1$ bits each. In many cases, such expansion of dynamic range is undesirable. Lifting can be applied to eliminate the dynamic range expansion for one of the two signals, without sacrificing exact invertibility, as follows.

- $D_A(x, y) = A_{444}(x, 2y) - A_{444}(x, 2y + 1)$; and
- $C_A^{\text{lifted}}(x, y) = A_{444}(x, 2y + 1) + (D_A(x, y) \gg 1)$.

[0209] In this variation, the bands $C_A^{\text{lifted}}(x, y)$ and $D_A(x, y)$ serve the same purpose (lowpass and highpass representations) as the bands $C_A(x, y)$ and $D_A(x, y)$ in the example described with reference to Figure 14. The dynamic range of $C_A^{\text{lifted}}(x, y)$, however, is only B bits rather than B + 1 bits, as dynamic range of $C_A^{\text{lifted}}(x, y)$ is approximately equal to only half the value of the $C_A(x, y)$ band. Invertibility is not sacrificed by this lifting operation (despite the rounding/truncation error introduced by the right shift operation), as the original signals can be recovered exactly from the bands $C_A^{\text{lifted}}(x, y)$ and $D_A(x, y)$ by simply reversing these equations algebraically as follows.

- $A_{444}(x, 2y + 1) = C_A^{\text{lifted}}(x, y) - (D_A(x, y) \gg 1)$, and
- $A_{444}(x, 2y) = D_A(x, y) + A_{444}(x, 2y + 1)$.

[0210] The dynamic range for sample values of one of the bands in this transformation (*i.e.*, $D_A(x, y)$) is still expanded to B + 1 bits, rather than having the same dynamic range as the input signal. One way to address this difference is to apply clipping. Image and video content is typically "lowpass" in nature – extremely large-magnitude differences between neighboring sample values are very rare in practice. The band $D_A(x, y)$ represents differences between neighboring sample values. For clipping, after computing $D_A(x, y)$ and $C_A^{\text{lifted}}(x, y)$ in the encoder, the encoder computes (and encodes)

$D_A^{\text{clipped}}(x, y) = \text{Min}(\text{Max}(-2^{(B-1)}, D_A(x, y)), 2^{(B-1)} - 1)$, instead of simply encoding $D_A(x, y)$ as the highpass signal. This operation clips the sample values of $D_A(x, y)$ to be in the range of $-2^{(B-1)}$ to $2^{(B-1)} - 1$. A corresponding decoder inverts the band separation filtering using $D_A^{\text{clipped}}(x, y)$ (or a lossy-decoded approximation thereof), instead of using $D_A(x, y)$ (or a lossy-decoded approximation thereof). For typical video content, cases where this clipping introduces distortion are rare, and the result will still be an improvement over using only a 4:2:0 representation – as at least part of the amplitude of the highpass signal will be retained. Of course, whenever $D_A(x, y)$ is within the clipping range, the clipping operation will not introduce any distortion. This enables the use of the bands $C_A^{\text{lifted}}(x, y)$ and $D_A^{\text{clipped}}(x, y)$ to represent the input values without requiring any expansion of dynamic range, and without sacrificing invertibility so long as $D_A(x, y)$ is within the clipping range, which is nearly always the case.

[0211] In some implementations, a sample value of the band $D^{\text{clipped}}_A(x, y)$ is offset by adding the constant offset $2^{(B-1)}$ to its value, to produce a signal with a range from 0 to $2^B - 1$, if this was the original input signal range.

- $D_A(x, y) = A_{444}(x, 2y) - A_{444}(x, 2y + 1)$;
- 5 • $C^{\text{lifted}}_A(x, y) = A_{444}(x, 2y + 1) + (D_A(x, y) >> 1)$; and
- $D^{\text{clipped_offset}}_A(x, y) = \text{Min}(\text{Max}(0, D_A(x, y) + 2^{(B-1)}), 2^B - 1)$.

The operation is fully reversible except when the clipping affects the signal, and the lowpass signal ($C^{\text{lifted}}_A(x, y)$) has the same dynamic range as the input signal. Expansion of the dynamic range of the highpass signal ($D^{\text{clipped_offset}}_A(x, y)$) is prevented by the clipping. The clipping can introduce distortion, but cases where clipping actually limits the highpass signal are expected to be rare, and in any case a clipped highpass signal still provides an enhancement signal. The constant offset can be subtracted when inverting the transformation after decoding.

- $D^{\text{clipped}'}_A(x, y) = D^{\text{clipped_offset}''}_A(x, y) - 2^{(B-1)}$;
- 15 • $A'_{444}(x, 2y + 1) = \text{Min}(\text{Max}(0, C^{\text{lifted}''}_A(x, y) - (D^{\text{clipped}'}_A >> 1)), 2^B - 1)$; and
- $A'_{444}(x, 2y) = \text{Min}(\text{Max}(0, ((D^{\text{clipped}'}_A) + A'_{444}(x, 2y + 1))), 2^B - 1)$.

where the " mark indicates reconstruction from (possibly lossy) coding. This example shows vertical processing steps. The same processing steps can be applied in conjunction with horizontal filtering as well.

20 [0212] Using clipping to limit the dynamic range of the highpass signal can also be applied without using lifting, although this may involve sacrificing exact invertibility. Also, the use of a half-magnitude lowpass signal can also be applied without using lifting, although this may involve sacrificing exact invertibility.

[0213] In this section, lifting and clipping operations are described for a first-step vertical filtering stage using Haar wavelet filters as LPF and HPF, as described with reference to Figure 14. Alternatively, the same lifting and clipping techniques can be applied for other stages (e.g., horizontally) and with other LPF and HPF kernels.

[0214] In this section, there are several references to exact invertibility as applied to wavelet decomposition and wavelet reconstruction. When lossy compression is performed on sample values of bands during subsequent encoding, the decoder operations may be performed on approximations of the encoded values rather than on their actual original values. Exact invertibility is not possible, in general, when lossy compression is performed. In any case, a decoder may perform the same (or similar) inverse conversion

operations with lossy compressed values as the decoder would perform for lossless representations of values.

Other Alternatives for Implementation Options

5 [0215] In the examples of three-band wavelet decomposition described herein, the H band is not subjected to additional wavelet decomposition after the first filtering stage. Alternatively, the H band is subjected to additional wavelet decomposition after the first filtering stage, so as to yield LL, LH, HL and HH bands, whose values are then assigned to main and auxiliary frames in the lower-resolution chroma sampling format.

10 [0216] Depending on implementation, intermediate and/or final results of encoding of the main frame of the lower-resolution chroma sampling format can be used during encoding of the auxiliary frame of the lower-resolution chroma sampling format. For example, an encoder can use motion vectors (and potentially macroblock/sub-macroblock/sub-macroblock-partition information and/or CU/PU/merge/TU segmentation
15 information) from encoding of the main frame to assist in searching for motion vectors (and corresponding segmentation information) for use in encoding the auxiliary frame. Or, an encoder can select quantization step size values to apply in various areas of an auxiliary frame based on results of encoding of the main frame.

[0217] In other band separation filtering approaches, pre-processing operations include
20 wavelet decomposition for frame packing of video content of a 4:4:4 format, 4:2:2 format or 4:2:0 format into a 4:0:0 format, which is typically used for gray scale or monochrome video content. The chroma information from a frame of the 4:4:4 format, 4:2:2 format or 4:2:0 format can be packed into the primary component of one or more additional or auxiliary frames of 4:0:0 format.

25 [0218] For additional details about variations of filters that can be used for wavelet decomposition or other band separation filtering in examples in the preceding sections, and for additional details about implementation options for such filters, see the following references:

- R. C. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, “Wavelet
30 Transforms That Map Integers to Integers,” *Appl. Comput. Harmon. Anal.*, vol. 5, no. 3, pp. 332–369 (1998).
- W. Sweldens, “The Lifting Scheme: A Construction of Second Generation Wavelets,” *SIAM Journal on Mathematical Analysis* (1998).

- A. Cohen, I. Daubechies, J.-C. Feauveau, “Biorthogonal Bases of Compactly Supported Wavelets,” *Communications on Pure and Applied Mathematics* 45 (5): 485–560 (1992).
- Rec. ITU-T T.800 | ISO/IEC 15444-1 *Information technology – JPEG 2000 image coding system: Core coding system* (2002).
- D. Le Gall and A. Tabatabai, “Sub-band Coding of Digital Images Using Symmetric Short Kernel Filters and Arithmetic Coding Techniques,” *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 761-764 (1988).
- J. D. Villasenor, B. Belzer, and J. Liao, “Wavelet Filter Evaluation for Image Compression,” *IEEE Trans on Image Processing*, pp. 1053-1057 (Aug. 1995).
- G. Uytterhoeven, D. Roose, and A. Bultheel, *Wavelet-Based Interactive Video Communication and Image Database Consulting – Wavelet Transforms using the Lifting Scheme*, Technical Report ITA-Wavelets-WP1.1, Katholieke Universiteit Leuven, Revised version (April 28, 1997).
- J. Rothweiler, “Polyphase Quadrature Filters - A New Subband Coding Technique,” *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 8, pp. 1280-1283 (1983).
- M. J. T. Smith and T. P. Barnwell, III, “Exact Reconstruction Techniques for Tree-structured Subband Coders,” *IEEE Trans. on Signal Processing*, vol. 34, pp. 434-441 (1986).

6. Syntax Elements for Band Separation Filtering/Inverse Filtering.

[0219] In example implementations, syntax indications in the bitstream can be used to identify the type of filtering applied in the encoder and/or indicate the type of filtering that should be applied when decoding. For example, the `content_interpretation_type` syntax element of a frame packing arrangement SEI message can indicate the filtering type. Extending the table shown in Section VIII.B, for example, additional values of `content_interpretation_type` can indicate the application of wavelet decomposition or other band separation filtering. The range of values 1-6 is defined as in the table. A value of 7 can indicate a first band separation/wavelet filtering scheme, a value of 8 can indicate a second band separation/wavelet filtering scheme, a value of 9 can indicate a third band separation/wavelet filtering scheme, and so on. Different band separation/wavelet filtering

schemes can differ in terms of pattern of decomposition (*e.g.*, three-band decomposition versus four-band decomposition, for different ways of rearranging sample values), filters used and/or implementation choices for filters.

[0220] Also, syntax indications in the bitstream can be used to identify phase of the filtering applied in the encoder and/or indicate the phase of filtering that should be applied when decoding. For example, syntax elements such as the `chroma_sample_loc_type_top_field` and `chroma_sample_loc_type_bottom_field` elements of the video usability information (“VUI”) can be used to indicate the filtering phase (*e.g.*, filtering that yields mid-point values versus filtering that yields co-sited values). When band separation filtering and inverse band separation filtering use a Haar wavelet pair for horizontal and vertical filtering, `chroma_loc_info_present_flag`, `chroma_sample_loc_type_top_field` and `chroma_sample_loc_type_bottom_field` syntax elements have a value of 1.

[0221] In example implementations for encoding/decoding for the H.264/AVC standard or HEVC standard, additional constraints on values of certain syntax elements may apply. In a frame-sequential approach, the syntax element `pic_output_flag` could be set to 1 for main video frames and set to 0 for auxiliary video frames. For example, in H.264/AVC, the `pic_output_flag` syntax element is set equal to 0 when `frame_packing_arrangement` is equal to 5, and the syntax element `current_frame_is_frame0_flag` is set equal to 0. This contrasts with an approach in which `pic_output_flag` is set to the same value in both the main and auxiliary constituent frames.

[0222] As another example, in a side-by-side arrangement for the main and auxiliary frames (in H.264/AVC, `frame_packing_arrangement` equal to 3) or top-bottom arrangement for the main and auxiliary frames (`frame_packing_arrangement` equal to 4), the cropping rectangle can be set to enclose only the main constituent frame without enclosing the auxiliary constituent frame. Alternatively, the cropping rectangle can be set to enclose both the main and auxiliary constituent frames.

7. Example Wavelet Reconstruction or Other Inverse Band Separation

Filtering.

[0223] Using wavelet reconstruction or other inverse band separation filtering, a decoder reconstructs an approximation of the input signal from values of bands. The values of bands may be approximations of original values (after lossy encoding and corresponding decoding) or exact representations of the original values (*e.g.*, after lossless intermediate

processing). For example, the decoder reconstructs the values of the input signal (*e.g.*, sample values for A) from the bandpass filtered values (*e.g.*, for the bands D, E, and F). In general, the lowpass/highpass subband filtering techniques and associated inverse processing operations applied by the decoder mirror the lowpass/highpass subband filtering techniques and associated operations applied by the encoder.

[0224] In more detail, a decoder decodes values of the main and auxiliary frames of the lower-resolution chroma sampling format. The decoder then assigns reconstructed sample values from the main and auxiliary frames of the lower-resolution chroma sampling format to appropriate bands of the frame in the higher-resolution chroma sampling format. In Figures 14, 15 and 16, for example, reconstructed values are assigned from the frame(s) in the lower-resolution chroma sampling format to the bands C, D, E, F, G and/or H, as appropriate for the approach used for wavelet decomposition. The decoder then performs wavelet reconstruction or other inverse band separation filtering on the values in the bands. If values for some bands are not available or otherwise not used in reconstruction, the decoder can perform wavelet reconstruction or other inverse band separation filtering on the values in the remaining bands. The details of the wavelet reconstruction or other inverse band separation filtering typically match or mirror the details of corresponding wavelet decomposition or other band separation filtering.

[0225] When wavelet decomposition included multiple stages of filtering, the wavelet reconstruction typically performs the multiple stages in reverse order for the affected bands. For example, in the example of Figure 14, a vertical filtering stage for the entire component plane A to produce bands $C_A(x, y)$ and $D_A(x, y)$, is followed by a horizontal filtering stage for the vertical lowpass band $C_A(x, y)$ to produce bands $E_A(x, y)$ and $F_A(x, y)$. In corresponding wavelet reconstruction, a horizontal inverse filtering stage for reconstructed versions of the two bands $E_A(x, y)$ and $F_A(x, y)$ produces a reconstructed version of the band $C_A(x, y)$. Then, vertical inverse filtering of the reconstructed versions of the bands $C_A(x, y)$ and $D_A(x, y)$ yields a reconstructed version of the component plane A.

[0226] For some implementation options (such as filter taps, corresponding normalization factor, filter phase, adjusting for bit depth expansion), the details of inverse filtering depend on how corresponding filtering was performed. For example, when wavelet decomposition uses a Haar wavelet filter pair, wavelet reconstruction can use a Haar wavelet filter pair to invert the wavelet decomposition. When the wavelet decomposition uses another type of filter pair (*e.g.*, a symmetric biorthogonal Daubechies

wavelet filter pair), the wavelet reconstruction uses a filter pair with filter taps and normalization factor defined to invert the filtering. If bit depth expanded when producing values of bands from input values, bit depth is restored when reconstructing values of the component planes.

- 5 [0227] Other implementation options (such as how normalization occurs, whether lifting is used, how division is implemented, how rounding is applied, whether clipping is used) may also be set to match or mirror operations performed during wavelet decomposition or other band separation filtering. Or, in some cases, a decoder may implement inverse filtering operations in ways that do not exactly mirror corresponding filtering operations performed before encoding.

[0228] In one implementation, for example, during frame unpacking, samples in areas of frames of YUV 4:2:0 format are inverse band separation filtered and assigned to reconstructed chroma components U'_{444} and V'_{444} of frame of YUV 4:4:4 format as shown in the following pseudocode.

```

15      T = 1 << ( B - 1 )
      for( x = 0; x < (W >> 1); x++ ) {
          for( y = 0; y < (H >> 1); y++ ) {
              E'_{U\_TEMP} = U''^{main}_{420}(x, y)
              E'_{V\_TEMP} = V''^{main}_{420}(x, y)
20      D'_{U\_TEMP1} = Y''^{aux}_{420}(2 * x, y) - T
              D'_{V\_TEMP1} = Y''^{aux}_{420}(2 * x, (H >> 1) + y) - T
              D'_{U\_TEMP2} = Y''^{aux}_{420}(2 * x + 1, y) - T
              D'_{V\_TEMP2} = Y''^{aux}_{420}(2 * x + 1, (H >> 1) + y) - T
              if( y % 2 == 0 ) {
25      F'_{U\_TEMP} = U''^{aux}_{420}(x, (y >> 1)) - T
              F'_{V\_TEMP} = U''^{aux}_{420}(x, (H >> 2) + (y >> 1)) - T
              } else {
              F'_{U\_TEMP} = V''^{aux}_{420}(x, (y >> 1)) - T
              F'_{V\_TEMP} = V''^{aux}_{420}(x, (H >> 2) + (y >> 1)) - T
30      }
      C'_{U\_TEMP1} = (E'_{U\_TEMP} + F'_{U\_TEMP}) << 1

```

$$C'_{V_TEMP1} = (E'_{V_TEMP} + F'_{V_TEMP}) \ll 1$$

$$C'_{U_TEMP2} = (E'_{U_TEMP} - F'_{U_TEMP}) \ll 1$$

$$C'_{V_TEMP2} = (E'_{V_TEMP} - F'_{V_TEMP}) \ll 1$$

$$U'_{444}(2 * x, 2 * y) = (C'_{U_TEMP1} + D'_{U_TEMP1}) \gg 1$$
5
$$U'_{444}(2 * x, 2 * y + 1) = (C'_{U_TEMP1} - D'_{U_TEMP1}) \gg 1$$

$$U'_{444}(2 * x + 1, 2 * y) = (C'_{U_TEMP2} + D'_{U_TEMP2}) \gg 1$$

$$U'_{444}(2 * x + 1, 2 * y + 1) = (C'_{U_TEMP2} - D'_{U_TEMP2}) \gg 1$$

$$V'_{444}(2 * x, 2 * y) = (C'_{V_TEMP1} + D'_{V_TEMP1}) \gg 1$$

$$V'_{444}(2 * x, 2 * y + 1) = (C'_{V_TEMP1} - D'_{V_TEMP1}) \gg 1$$
10
$$V'_{444}(2 * x + 1, 2 * y) = (C'_{V_TEMP2} + D'_{V_TEMP2}) \gg 1$$

$$V'_{444}(2 * x + 1, 2 * y + 1) = (C'_{V_TEMP2} - D'_{V_TEMP2}) \gg 1$$

$$\}$$

$$\}$$

where the " mark indicates reconstruction from (possibly lossy) coding.

15

8. Generalized Techniques for Frame Packing/Unpacking with Band Separation Filtering/Inverse Filtering.

[0229] Figure 17 shows a generalized technique (1700) for frame packing with wavelet decomposition or other band separation filtering. A computing device that implements a frame packer, for example as described with reference to Figure 4, can perform the technique (1700).

[0230] In general, according to the technique (1700), the device packs one or more frames of a higher-resolution chroma sampling format into one or more frames of a lower-resolution chroma sampling format. For example, the device packs frame(s) of 4:4:4 format into frame(s) of 4:2:0 format. Or, the device packs frame(s) of 4:2:2 format into frame(s) of 4:2:0 format. Or, the device packs frame(s) of 4:4:4 format into frame(s) of 4:2:2 format. Or, the device packs frame(s) of a 4:4:4, 4:2:2 or 4:2:0 format into frame(s) of a 4:0:0 format.

[0231] In particular, as part of the packing, the device performs (1711) wavelet decomposition or other band separation filtering on sample values of chroma components of the frame(s) of the higher-resolution chroma sampling format to produce sample values

of multiple bands. The device then assigns (1712) the sample values of the multiple bands to parts of the frame(s) of the lower-resolution chroma sampling format.

[0232] For example, the wavelet decomposition can be a three-band wavelet

decomposition as shown in Figure 14 or 15, four-band wavelet decomposition as shown in

5 Figure 16a or 16b, or other wavelet decomposition. Typically, the wavelet decomposition or other band separation filtering uses a filter pair with a LPF and HPF. The filter pair can be scaled or unscaled (a) $LPF = [1 \ 1] / 2$ and $HPF = [1 \ -1] / 2$; (b) $LPF = [-1 \ 1 \ 8 \ 8 \ 1 \ -1] / 16$ and $HPF = [1 \ -1] / 2$; (c) $LPF = [-1 \ 2 \ 6 \ 2 \ -1] / 8$ and $HPF = [-1 \ 2 \ -1] / 2$; (d) LPF and HPF based on quadrature mirror filters; (e) LPF and HPF based on polyphase quadrature

10 filters; (f) LPF and HPF based on conjugate quadrature filters; or (g) some other filter pair. For the filter pair, division operations can be implemented with arithmetic right shift operations, implemented with integer division operations or implemented in another way.

Division operations can include no rounding, include nearest-integer rounding or include dithered rounding. Depending on implementation, wavelet decomposition or other band

15 separation filtering may use lifting. Also, depending on implementation, wavelet decomposition or other band separation filtering may include clipping for at least some of the sample values of the bands.

[0233] In many examples, the wavelet decomposition or other band separation filtering includes multiple filtering stages. Depending on implementation, the multiple filtering

20 stages may include a vertical filtering stage followed by a horizontal filtering stage. Or, the multiple filtering stages may include a horizontal filtering stage followed by a vertical filtering stage. For such multi-stage filtering, depending on implementation,

normalization to compensate for expansion (a) can occur after each filtering stage of the multiple filtering stages, (b) can be at least partially deferred for one or more of the

25 multiple filtering stages, or (c) can be at least partially ignored for one or more filtering stages, so as to provide scaling of the sample values of the bands.

[0234] The device can then encode (1720) the frame(s) of the lower-resolution chroma sampling format. Alternatively, a different device performs the encoding (1720). The device(s) can repeat the technique (1700) on a frame-by-frame basis or other basis.

30 [0235] The device can signal metadata about frame packing/unpacking. For example, the device signals metadata that indicates type of filtering applied and/or filtering phase of filtering applied.

[0236] Figure 18 shows a generalized technique (1800) for frame unpacking with wavelet reconstruction or other inverse band separation filtering. A computing device that

implements a frame unpacker, for example as described with reference to Figure 5, can perform the technique (1800).

[0237] Before the frame unpacking itself, the device can decode (1810) one or more frames of a lower-resolution chroma sampling format. Alternatively, a different device
5 performs the decoding (1810).

[0238] In general, according to the technique (1800), the device unpacks the frame(s) of the lower-resolution chroma sampling format into one or more frames of a higher-resolution chroma sampling format. For example, the device unpacks frame(s) of 4:2:0 format into frame(s) of 4:4:4 format. Or, the device unpacks frame(s) of 4:2:0 format into
10 frame(s) of 4:2:2 format. Or, the device unpacks frame(s) of 4:2:2 format into frame(s) of 4:4:4 format. Or, the device unpacks frame(s) of a 4:0:0 format into frame(s) of a 4:4:4, 4:2:2 or 4:2:0 format.

[0239] In particular, as part of the unpacking, the device assigns (1821) parts of the frame(s) of the lower-resolution chroma sampling format to sample values of multiple
15 bands. The device then performs (1822) wavelet reconstruction or other inverse band separation filtering on the sample values of the bands to produce sample values of chroma components of the frame(s) of the higher-resolution chroma sampling format.

[0240] For example, the wavelet reconstruction can be a three-band wavelet reconstruction, four-band wavelet reconstruction or other wavelet reconstruction.
20 Typically, the wavelet reconstruction or other band separation filtering uses a filter pair with a LPF and HPF. For the filter pair, division operations can be implemented with arithmetic right shift operations, implemented with integer division operations, or implemented in another way. Division operations can include no rounding, include nearest-integer rounding or include dithered rounding. Depending on implementation,
25 wavelet reconstruction or other inverse band separation filtering may use lifting. Also, depending on implementation, wavelet reconstruction or other inverse band separation filtering may include clipping for at least some of the sample values of the bands.

[0241] In many examples, the wavelet reconstruction or other inverse band separation filtering includes multiple filtering stages. Depending on implementation, the multiple
30 filtering stages may include a vertical filtering stage followed by a horizontal filtering stage. Or, the multiple filtering stages may include a horizontal filtering stage followed by a vertical filtering stage. For such multi-stage filtering, depending on implementation, normalization to compensate for expansion (a) can occur after each filtering stage of the multiple filtering stages, (b) can be at least partially deferred for one or more of the

multiple filtering stages, or (c) can compensate for scaling of the sample values of the bands in pre-processing.

[0242] The device can also receive metadata about frame packing/unpacking. For example, the device receives metadata that indicates type of filtering applied and/or
5 filtering phase of filtering applied.

[0243] The device(s) can repeat the technique (1800) on a frame-by-frame basis or other basis.

[0244] In view of the many possible embodiments to which the principles of the disclosed invention may be applied, it should be recognized that the illustrated
10 embodiments are only preferred examples of the invention and should not be taken as limiting the scope of the invention. Rather, the scope of the invention is defined by the following claims. We therefore claim as our invention all that comes within the scope and spirit of these claims.

CLAIMS

1. A method comprising:
packing one or more frames of a higher-resolution chroma sampling format into one or more frames of a lower-resolution chroma sampling format, wherein the packing includes:
performing band separation filtering on sample values of chroma components of the one or more frames of the higher-resolution chroma sampling format to produce sample values of plural bands; and
assigning the sample values of the plural bands to parts of the one or more frames of the lower-resolution chroma sampling format.
2. The method of claim 1 further comprising, after the packing:
encoding the one or more frames of the lower-resolution chroma sampling format.
3. The method of claim 1 wherein the band separation filtering is a three-band wavelet decomposition or four-band wavelet decomposition.
4. The method of claim 1 wherein the band separation filtering uses lifting and/or includes clipping at least some of the sample values of the plural bands.
5. The method of claim 1 further comprising signaling metadata that indicates one or more of:
type of filtering applied; and
filtering phase of filtering applied.
6. A computing device adapted to perform a method comprising:
unpacking one or more frames of a lower-resolution chroma sampling format into one or more frames of a higher-resolution chroma sampling format, wherein the unpacking includes:
assigning parts of the one or more frames of the lower-resolution chroma sampling format to sample values of plural bands; and
performing inverse band separation filtering on the sample values of the plural bands to produce sample values of chroma components of the one or more frames of the higher-resolution chroma sampling format.

7. The computing device of claim 6 wherein the method further comprises, before the unpacking:

decoding the one or more frames of the lower-resolution chroma sampling format.

8. The computing device of claim 6 wherein the inverse band separation filtering is a three-band wavelet reconstruction or four-band wavelet reconstruction.

9. The computing device of claim 6 wherein the method further comprises receiving metadata that indicates one or more of:

type of filtering applied; and

filtering phase of filtering applied.

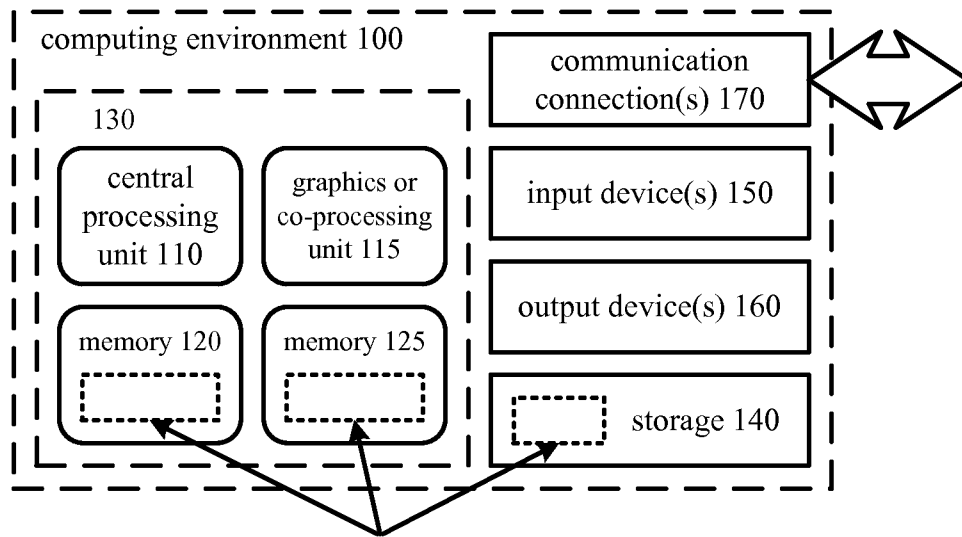
10. A computing device adapted to perform a method comprising:

packing one or more frames of a higher-resolution chroma sampling format into one or more frames of a lower-resolution chroma sampling format, wherein the higher-resolution chroma sampling format is a 4:4:4 format, 4:2:2 format or 4:2:0 format, wherein the lower-resolution chroma sampling format is a 4:2:2 format, 4:2:0 format or 4:0:0 format lower than the higher-resolution chroma sampling format, and wherein the packing includes:

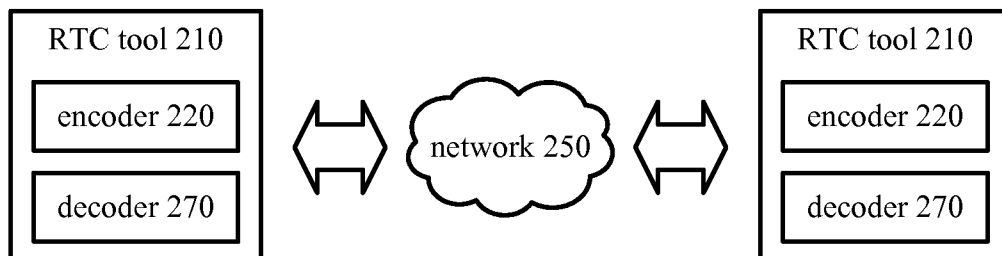
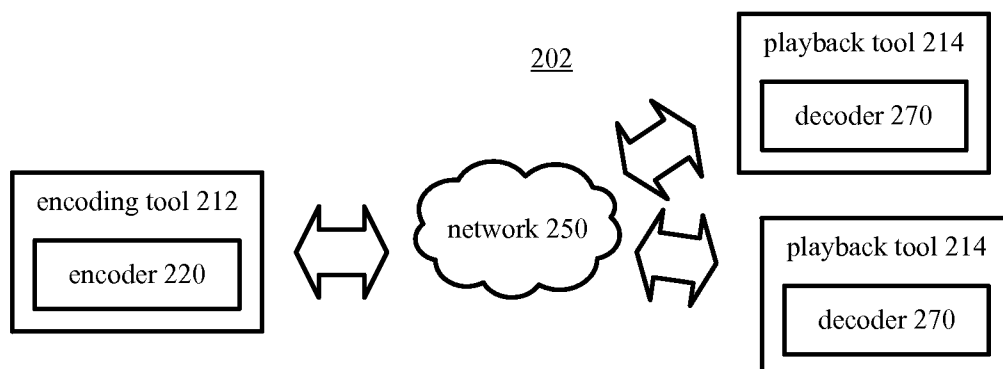
performing wavelet decomposition on sample values of chroma components of the one or more frames of the higher-resolution chroma sampling format to produce sample values of plural bands; and

assigning the sample values of the plural bands to parts of the one or more frames of the lower-resolution chroma sampling format.

1/16



software 180 implementing one or more innovations for frame packing with band separation filtering and/or unpacking with inverse band separation filtering for higher-resolution chroma sampling formats

Figure 1201**Figure 2a**202**Figure 2b**

300

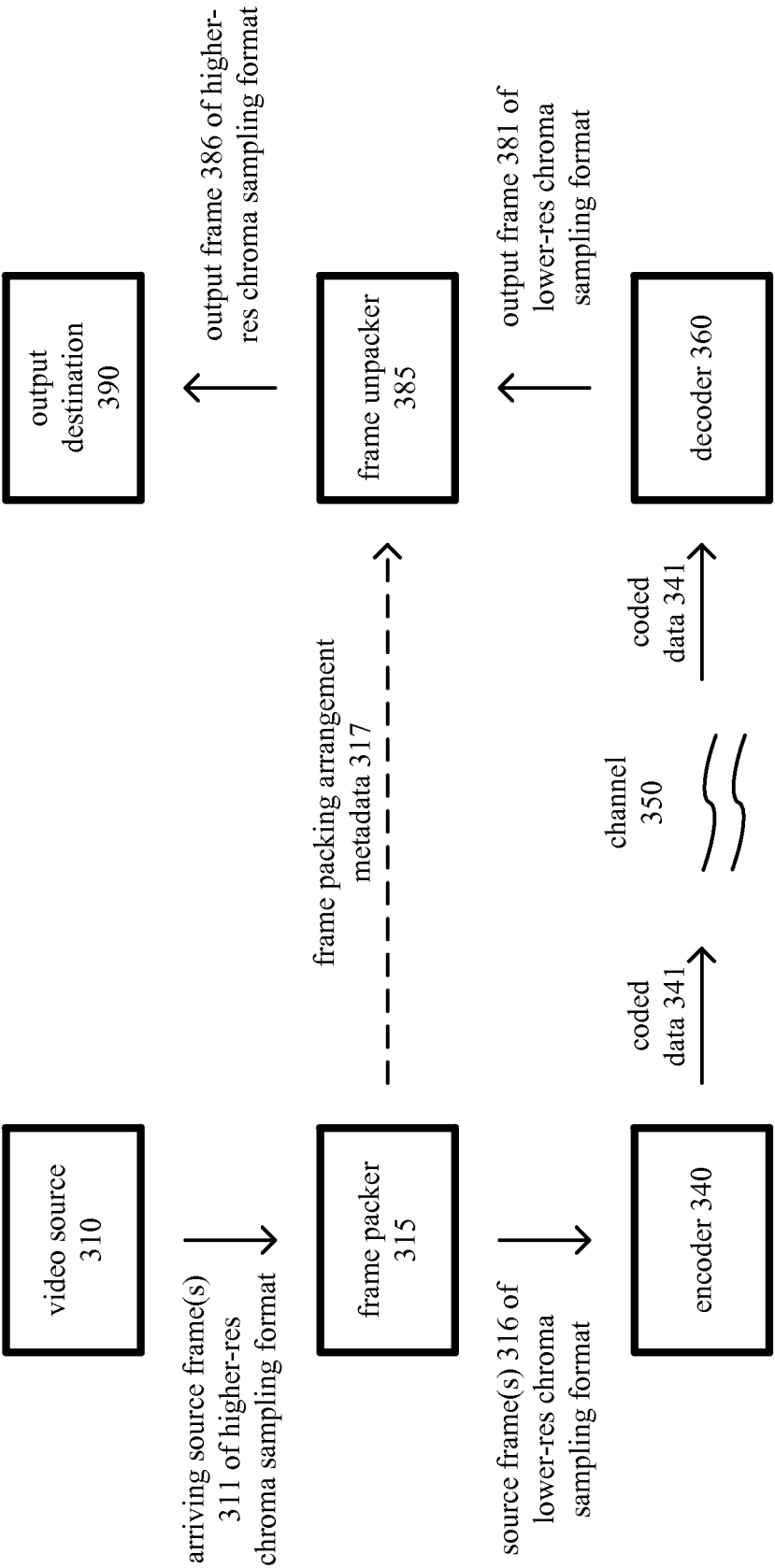


Figure 3

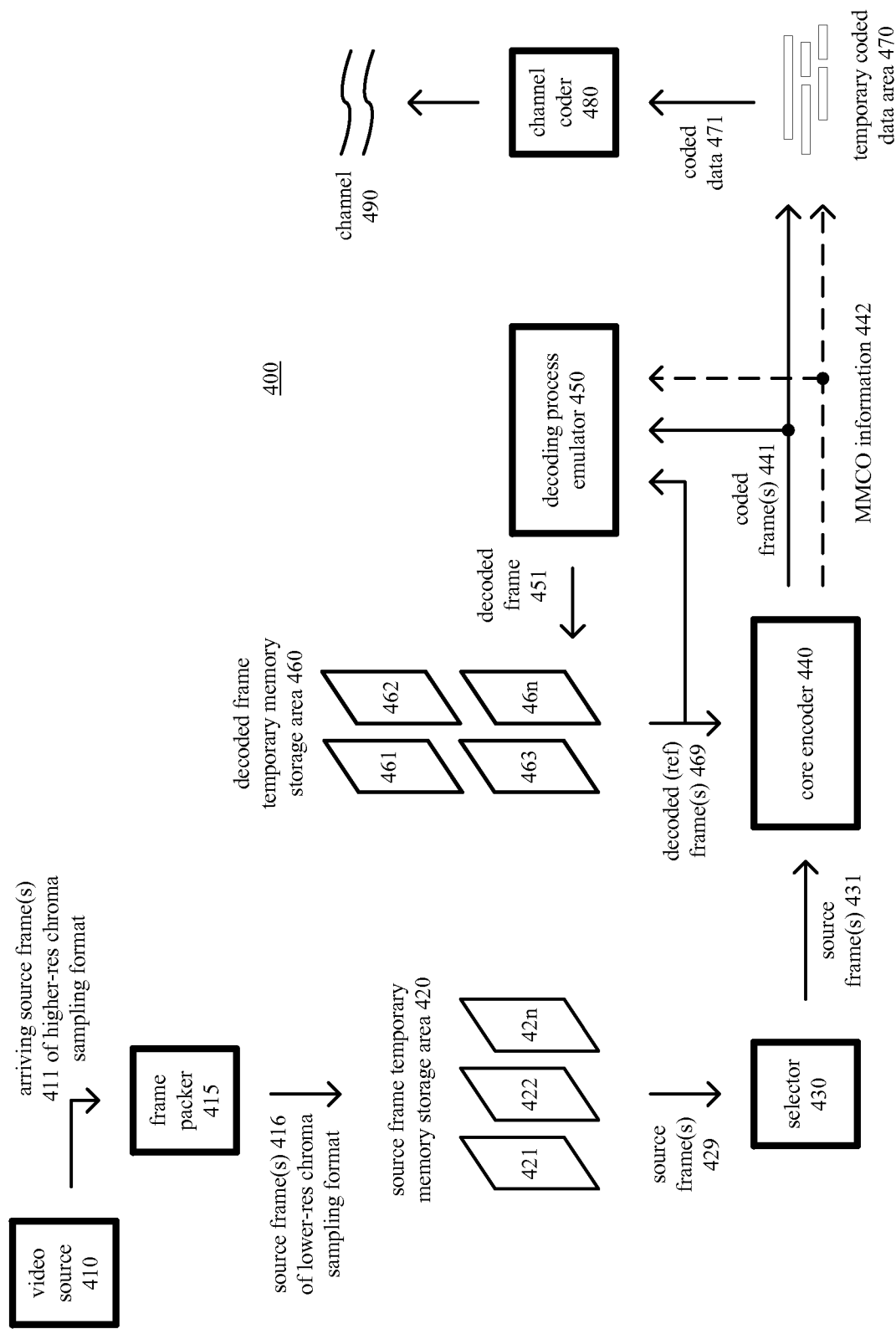


Figure 4

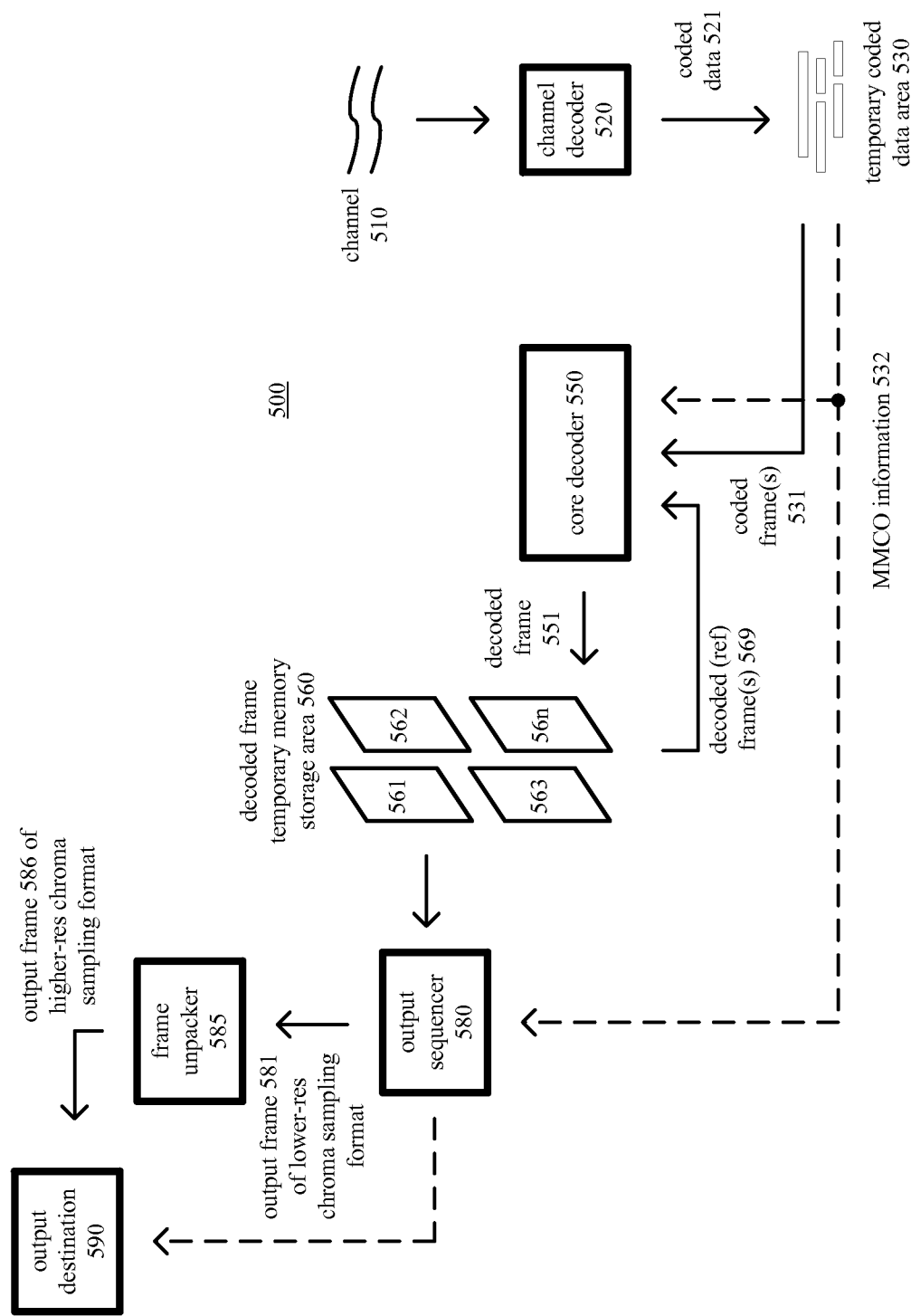


Figure 5

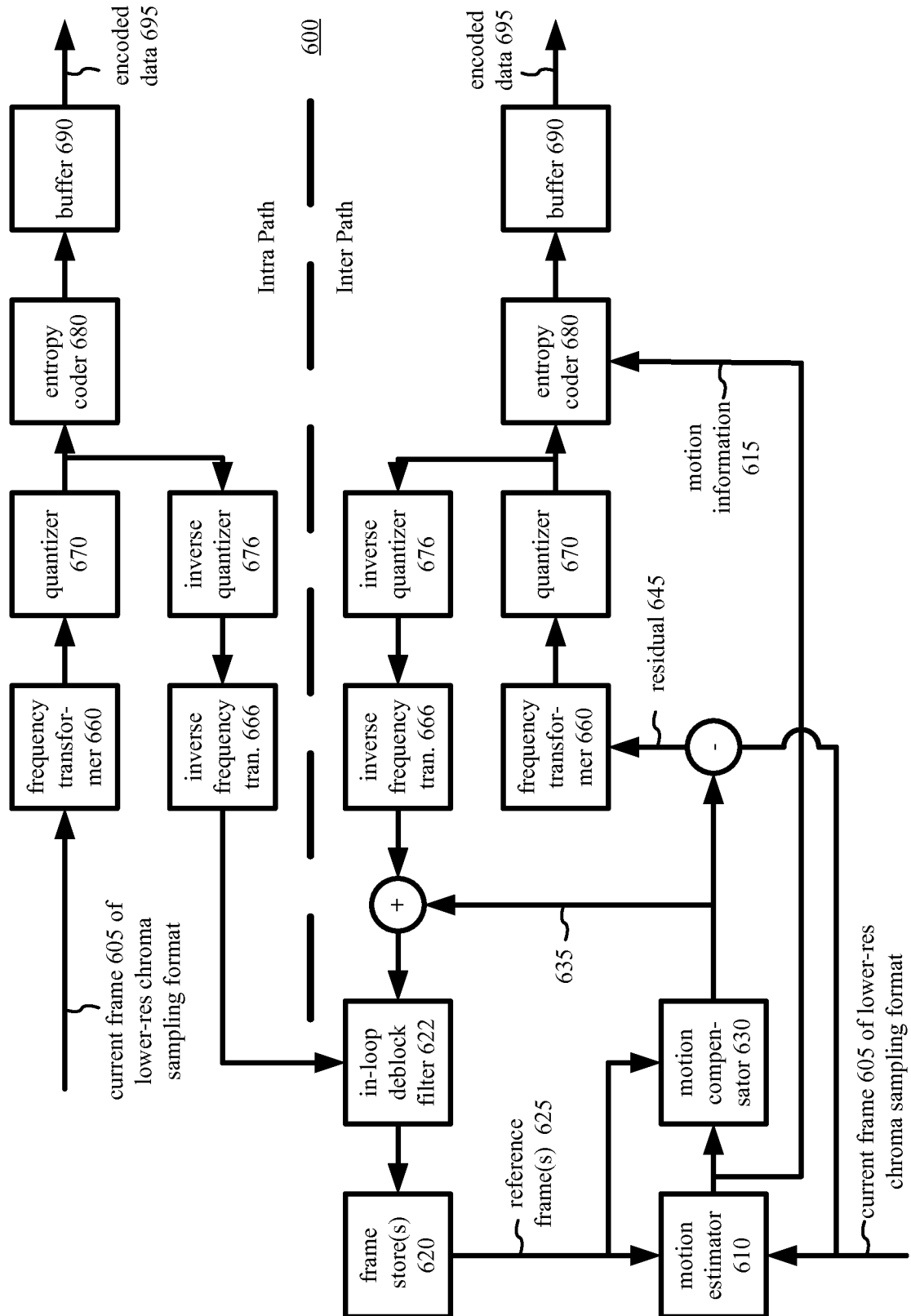


Figure 6

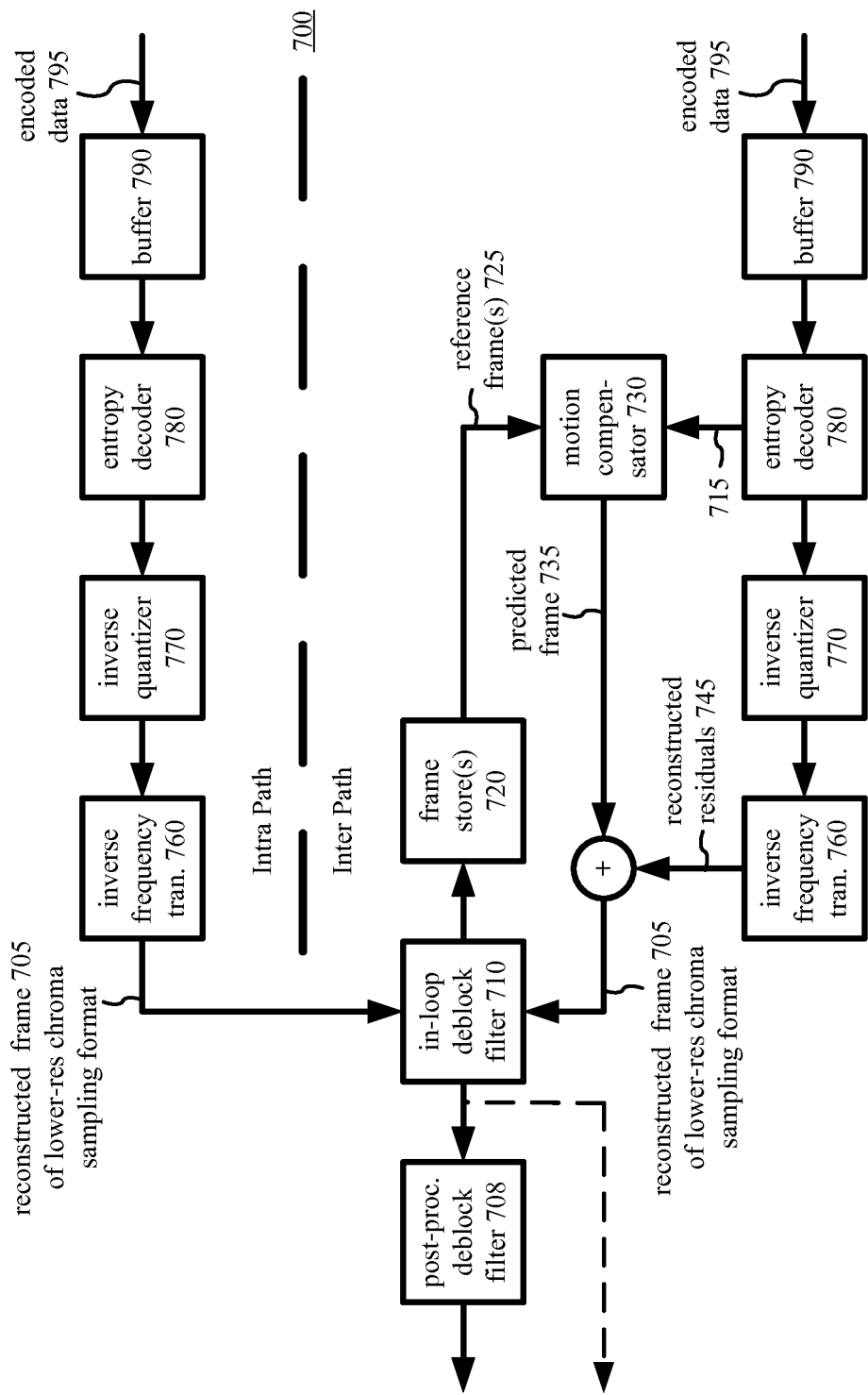


Figure 7

7/16

800

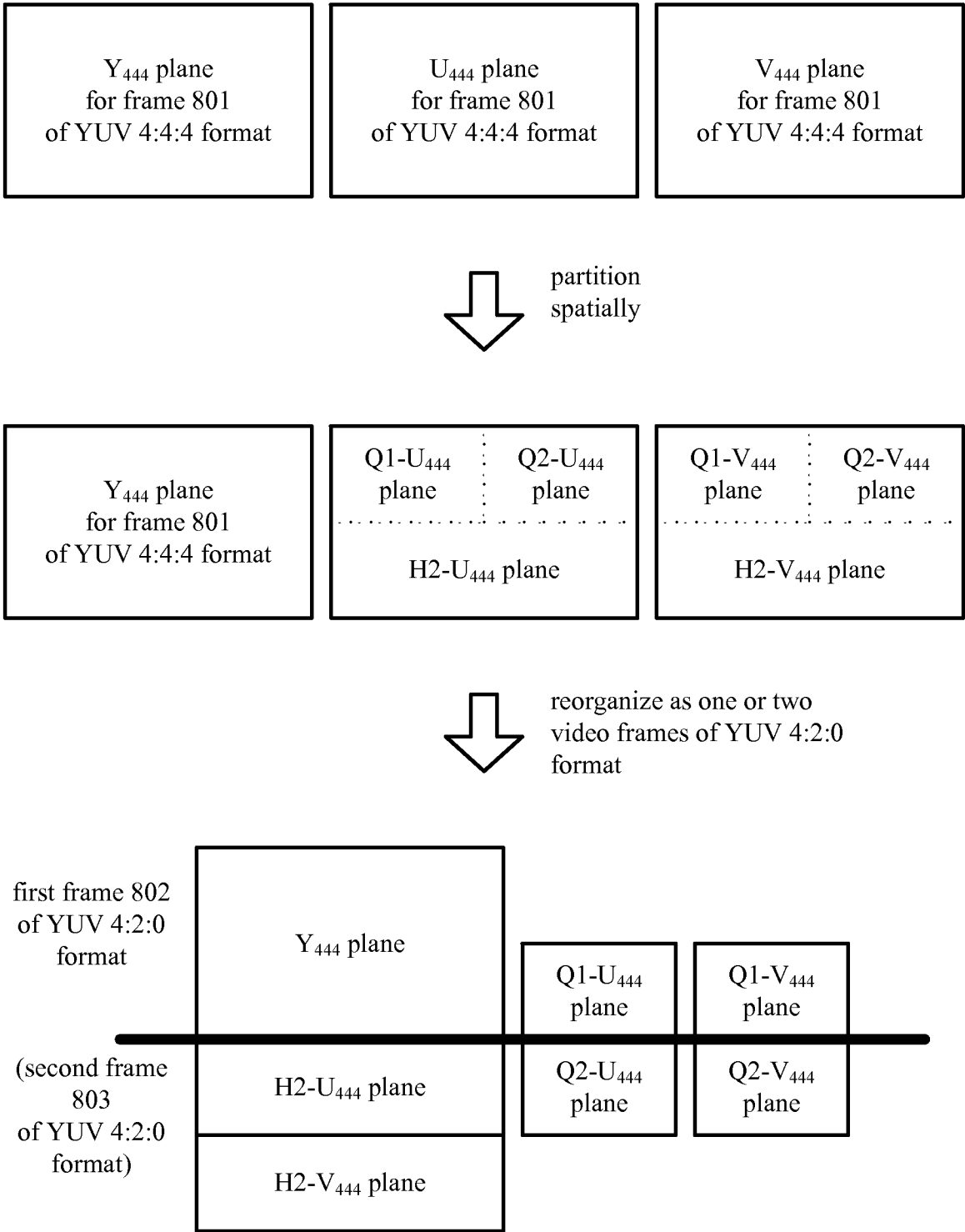


Figure 8

8/16

900

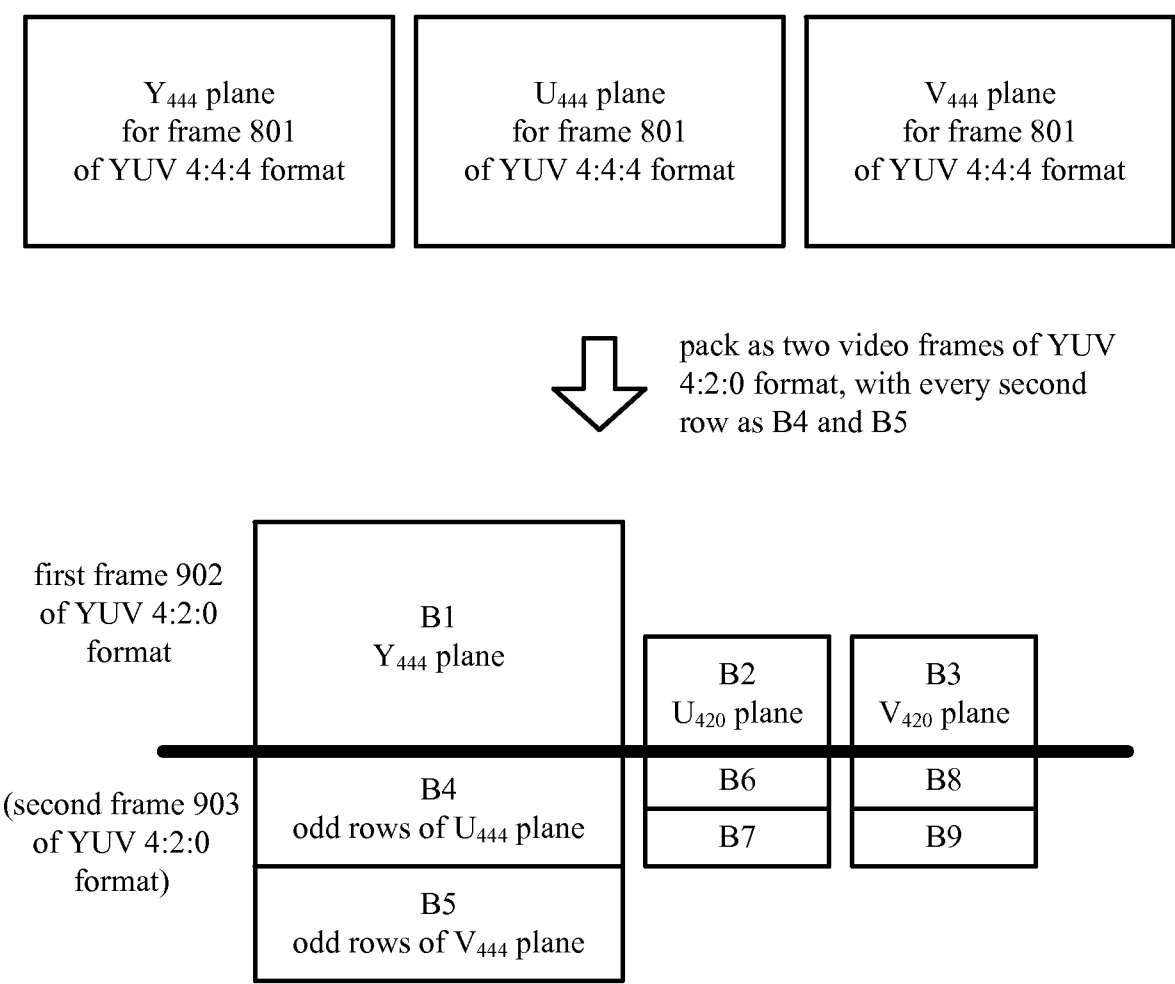


Figure 9

9/16

1000

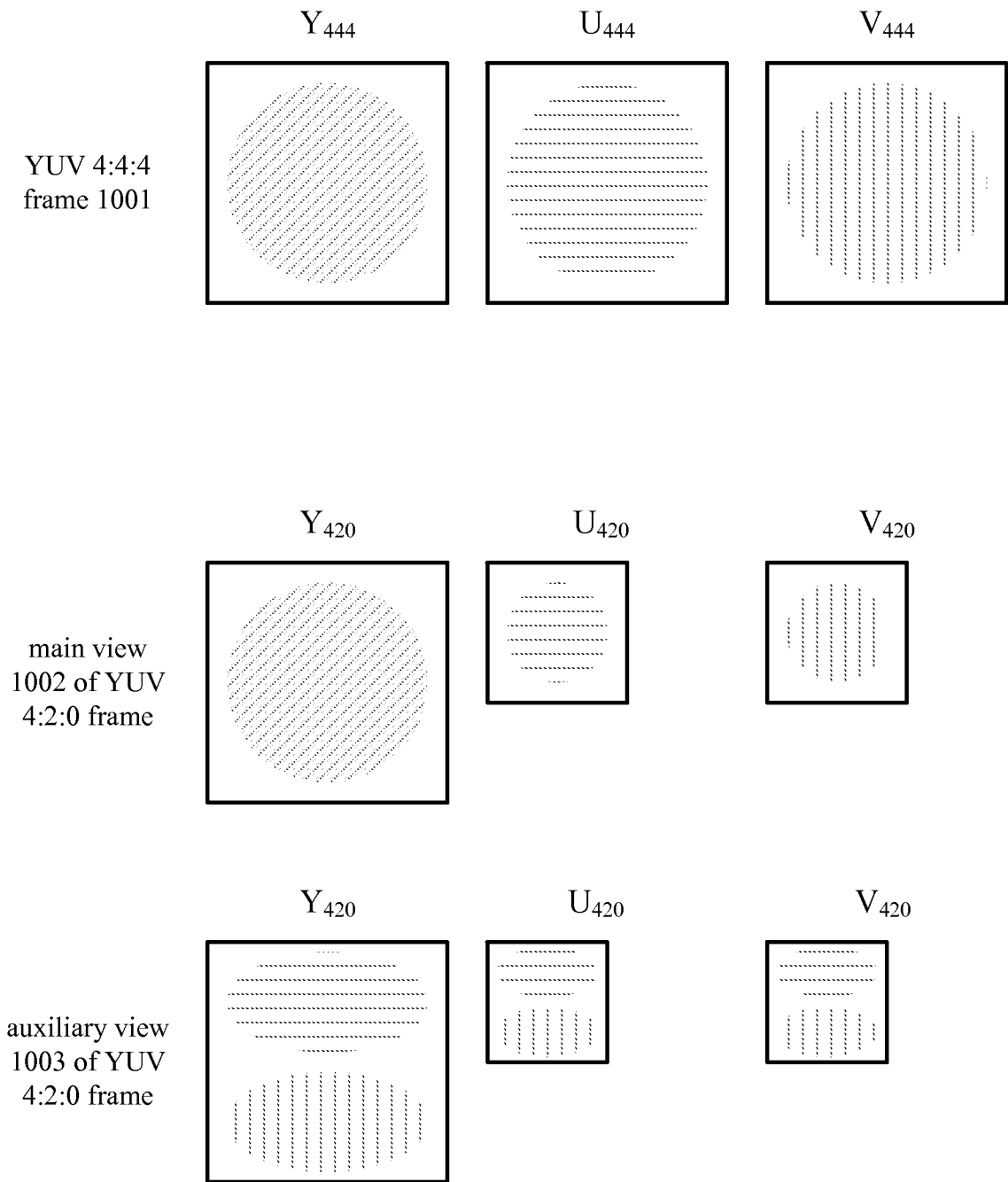
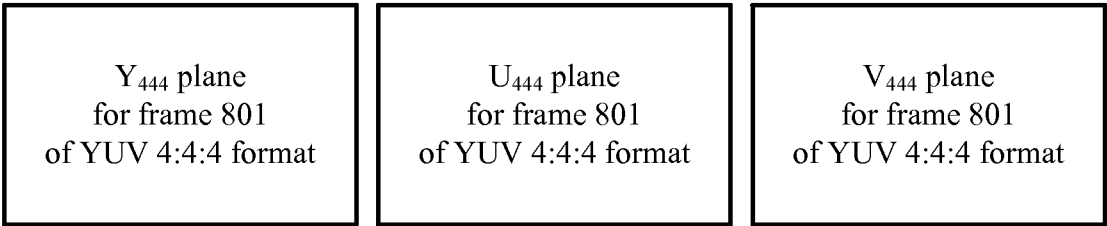


Figure 10

10/16

1100



pack as two video frames of YUV 4:2:0 format, with every second column as A4 and A5

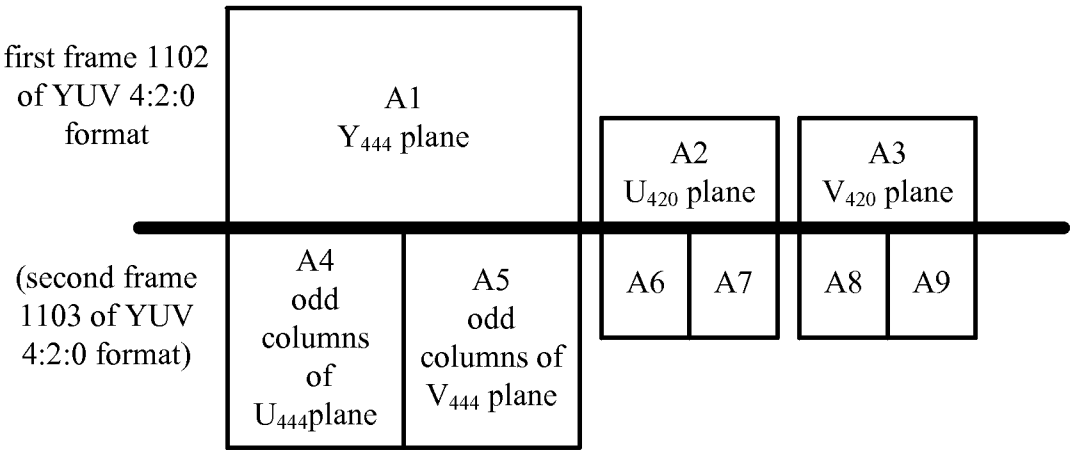
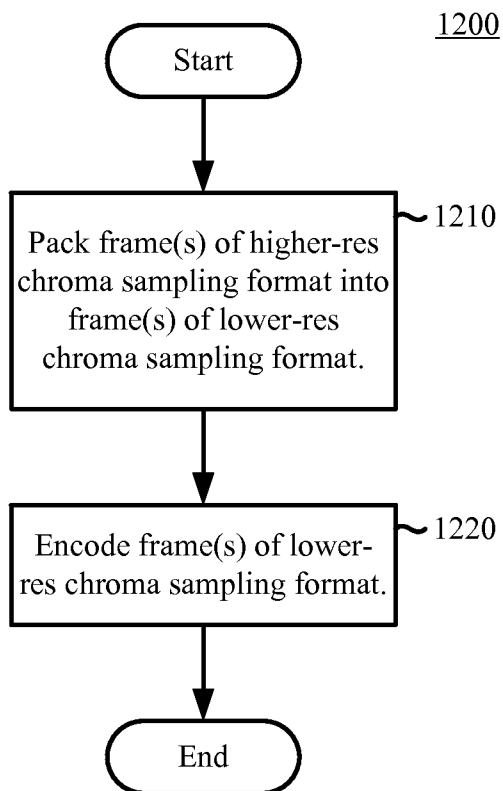
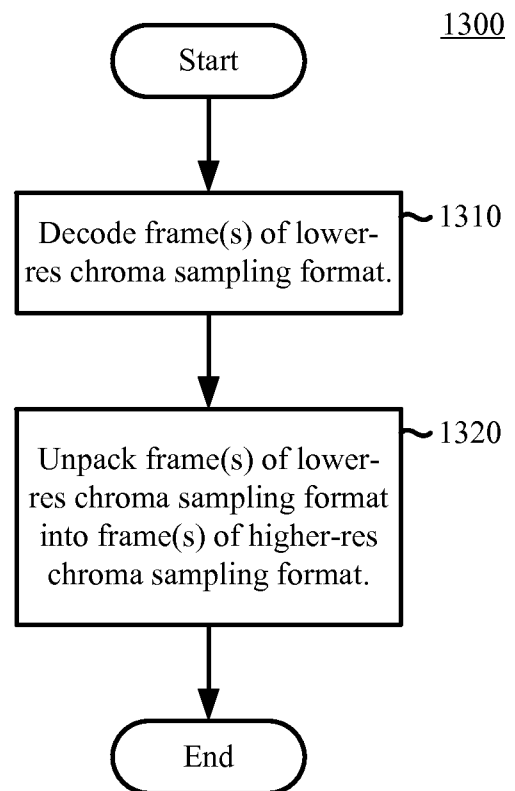


Figure 11

**Figure 12****Figure 13**

12/16

1400

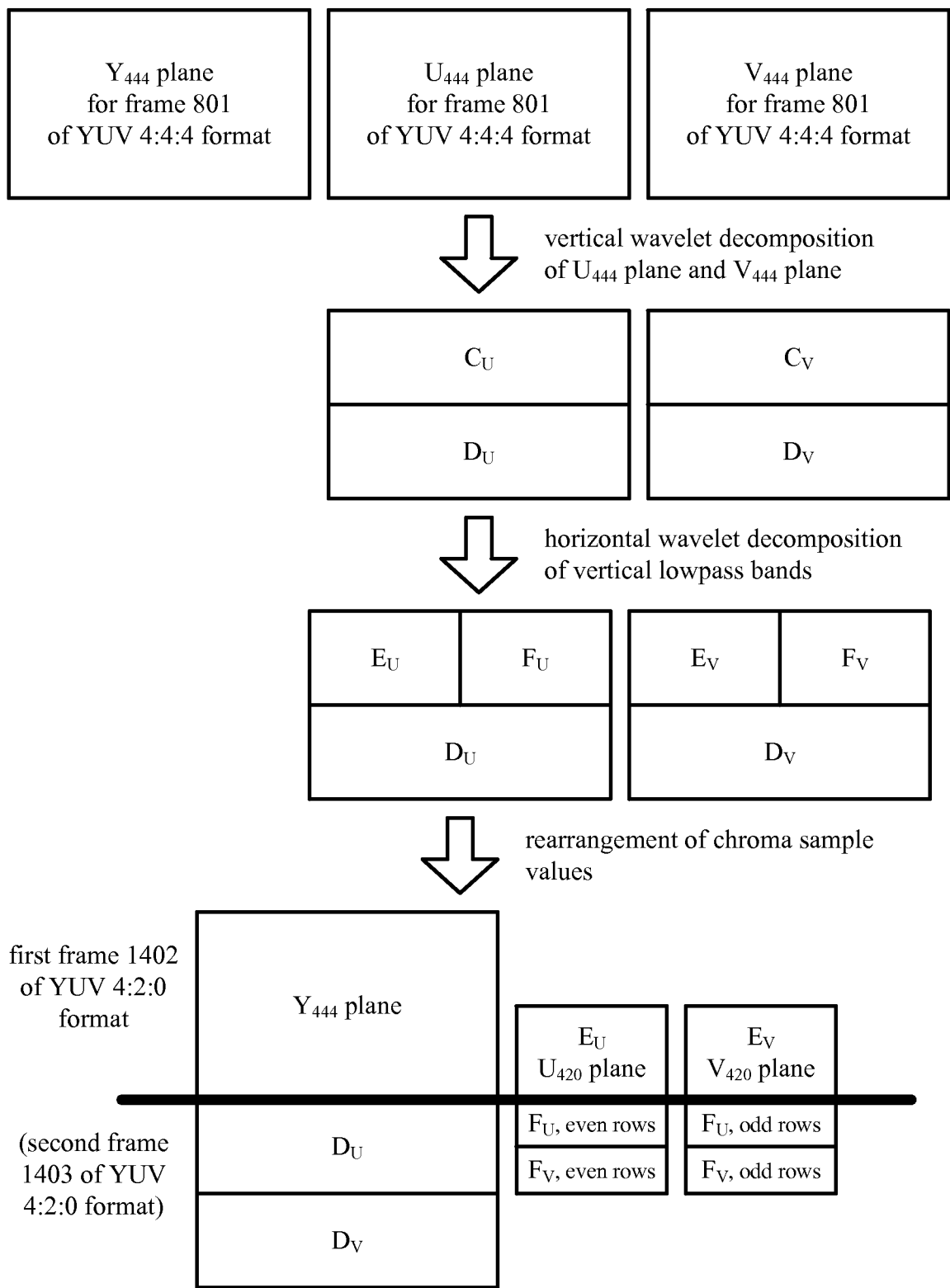


Figure 14

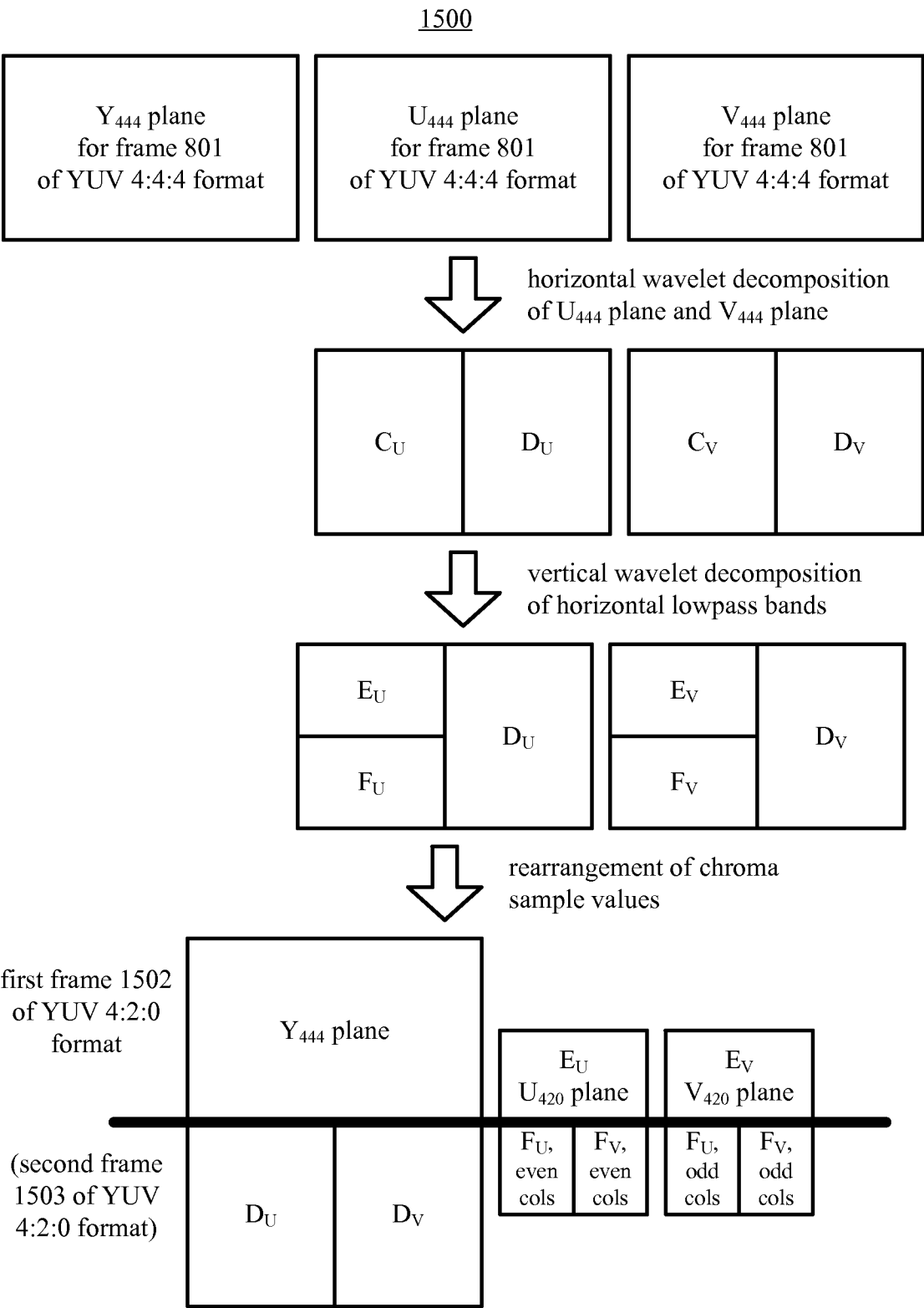


Figure 15

14/16

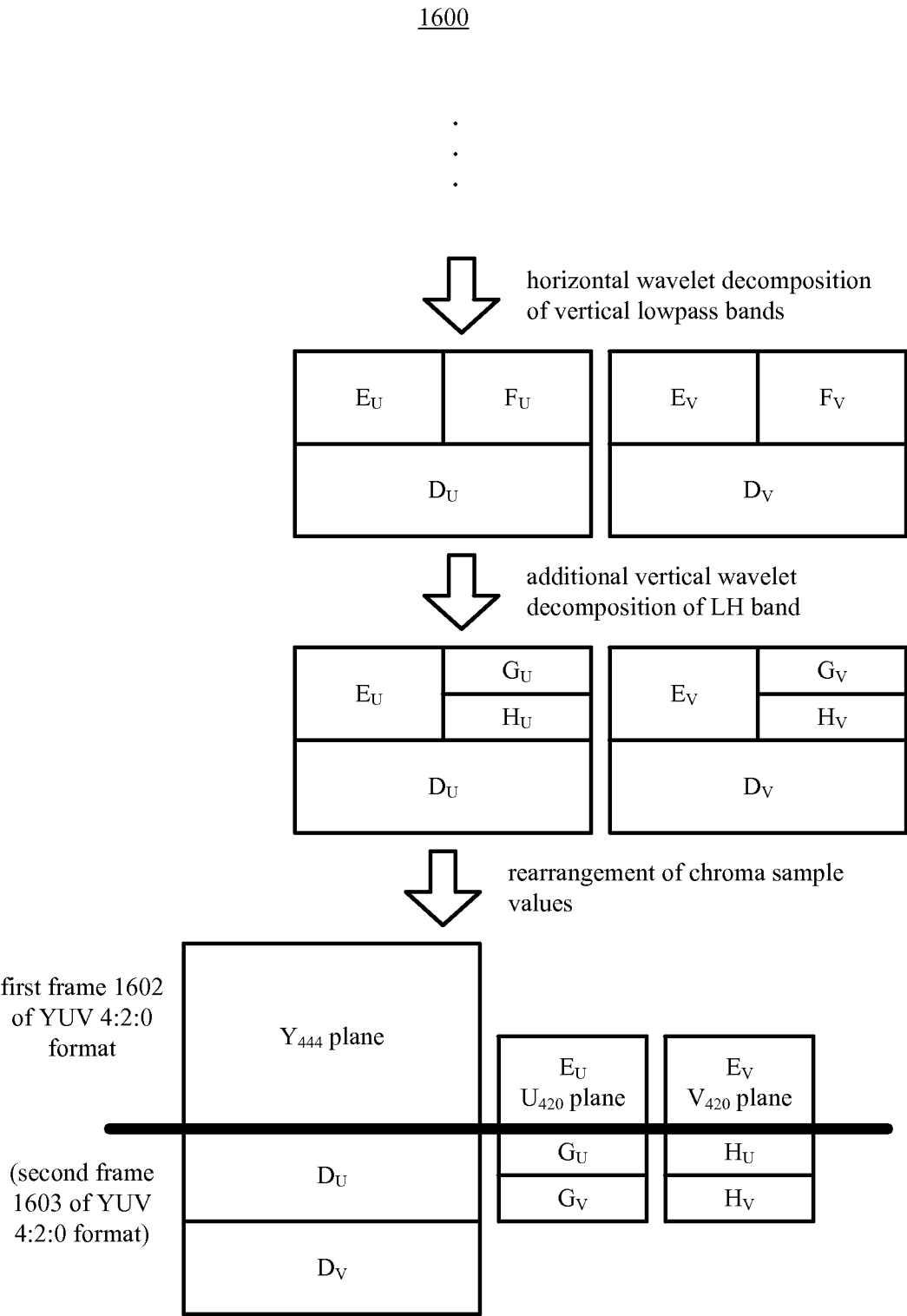


Figure 16a

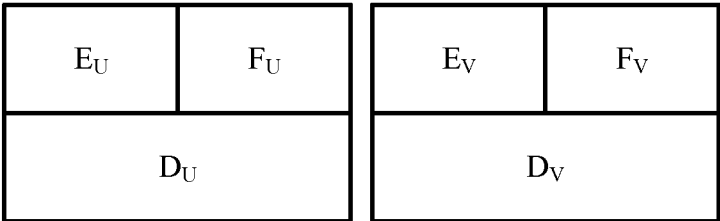
15/16

1650

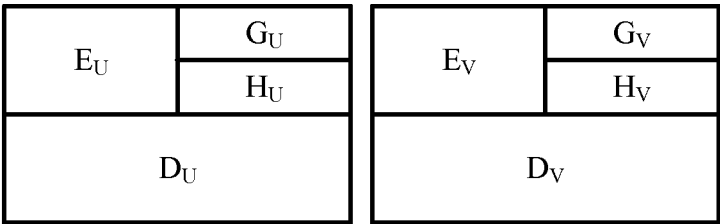
•
•
•



horizontal wavelet decomposition
of vertical lowpass bands



additional vertical wavelet
decomposition of LH band



rearrangement of chroma
sample values

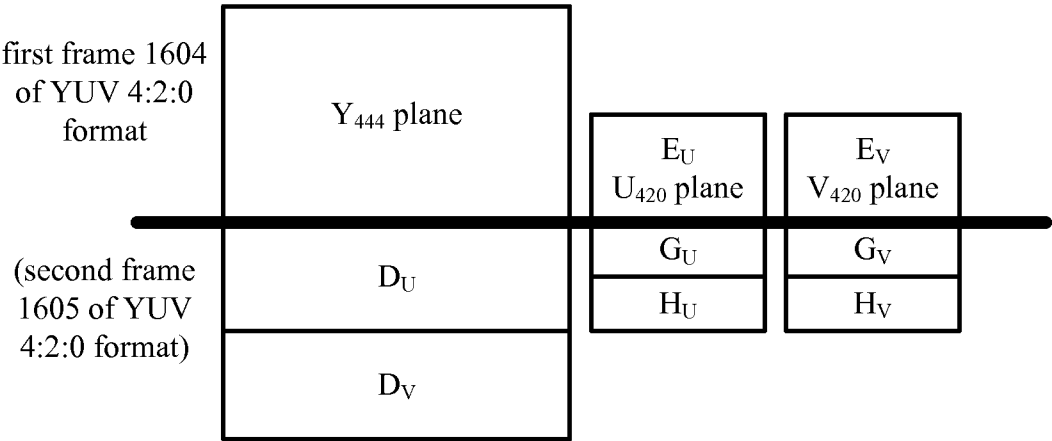
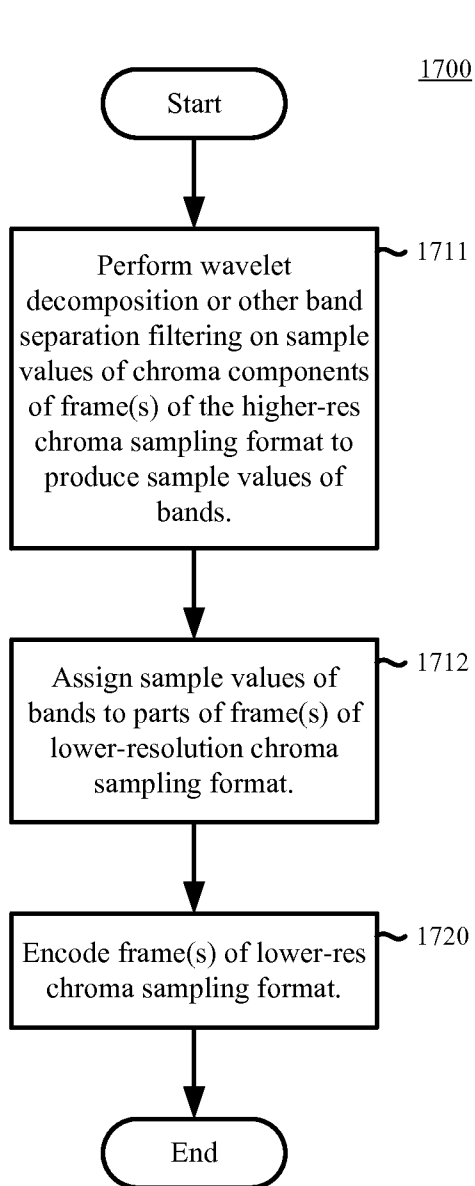
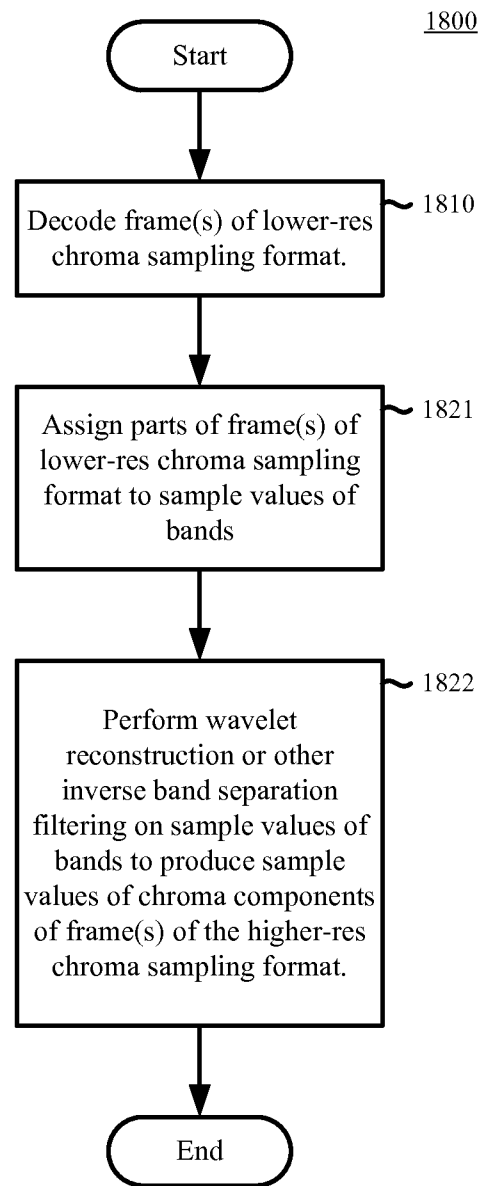


Figure 16b

**Figure 17****Figure 18**

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2013/065754

A. CLASSIFICATION OF SUBJECT MATTER

INV. H04N7/26

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WU Y ET AL: "Frame packing arrangement SEI for 4:4:4 content in 4:2:0 bitstreams", 11. JCT-VC MEETING; 102. MPEG MEETING; 10-10-2012 - 19-10-2012; SHANGHAI; (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-K0240, 1 October 2012 (2012-10-01), XP030113122, the whole document -----	1-10



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

13 January 2014

Date of mailing of the international search report

22/01/2014

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

Oelbaum, Tobias